A Survey of Image Style Transfer and Generation Techniques

www.surveyx.cn

Abstract

This survey explores the transformative impact of image style transfer and generation technologies on creative and artistic domains, emphasizing the role of deep learning and neural networks in advancing these fields. Key advancements include the development of neural style transfer methods, diffusion models, and generative adversarial networks, which have enhanced the quality, efficiency, and flexibility of image synthesis. Despite significant progress, challenges persist in computational complexity, generalization, and maintaining ethical standards, particularly concerning bias and copyright issues. The survey systematically examines the evolution of these techniques, highlighting innovative approaches like DiffStyler and Style Projection, which address localized style transfer and feature alignment. The integration of transformers and multimodal inputs further expands creative possibilities, enabling more nuanced and diverse artistic expressions. Ethical considerations are paramount, with ongoing efforts to mitigate biases and ensure responsible use. This survey concludes by identifying future research directions, including optimizing computational efficiency and enhancing control over style transfer, to continue advancing the field and its applications in art generation.

1 Introduction

1.1 Significance in Creative and Artistic Domains

Image style transfer and generation technologies have become essential in the creative industries, enhancing artistic expression by transforming content with distinct stylistic attributes while preserving semantic integrity [1]. This practice of applying various painting styles to input images not only fosters innovative artistic expressions but also addresses the need for accurate correspondences between content and style regions, ensuring coherent artistic vision [2, 3].

The integration of style transfer in digital imagery has gained popularity in both academia and industry, allowing for the infusion of artistic styles into everyday photographs, thus enhancing user engagement and visual appeal [4, 5]. This technique opens new avenues for artistic expression, expanding creative possibilities for artists and designers [6].

Furthermore, the application of style transfer in videos is emerging, providing solutions for universal style transfer while maintaining visual consistency across frames, which is crucial for the creative industries that rely on video content [7, 8]. Traditional methods often depend on pre-selected style images, which can constrain creativity [9]. However, advancements in technology now enable artists to synthesize a diverse range of stylized outputs from a single content and style image pair, enhancing their creative toolkit [1].

The proposed method of transferring style to arbitrary instances addresses significant challenges in fields like fashion design and augmented reality, where visual content customization is vital [10]. Additionally, image style transfer techniques are pivotal in generating high-resolution, photo-realistic images, significantly impacting the quality and realism of digital art [11].

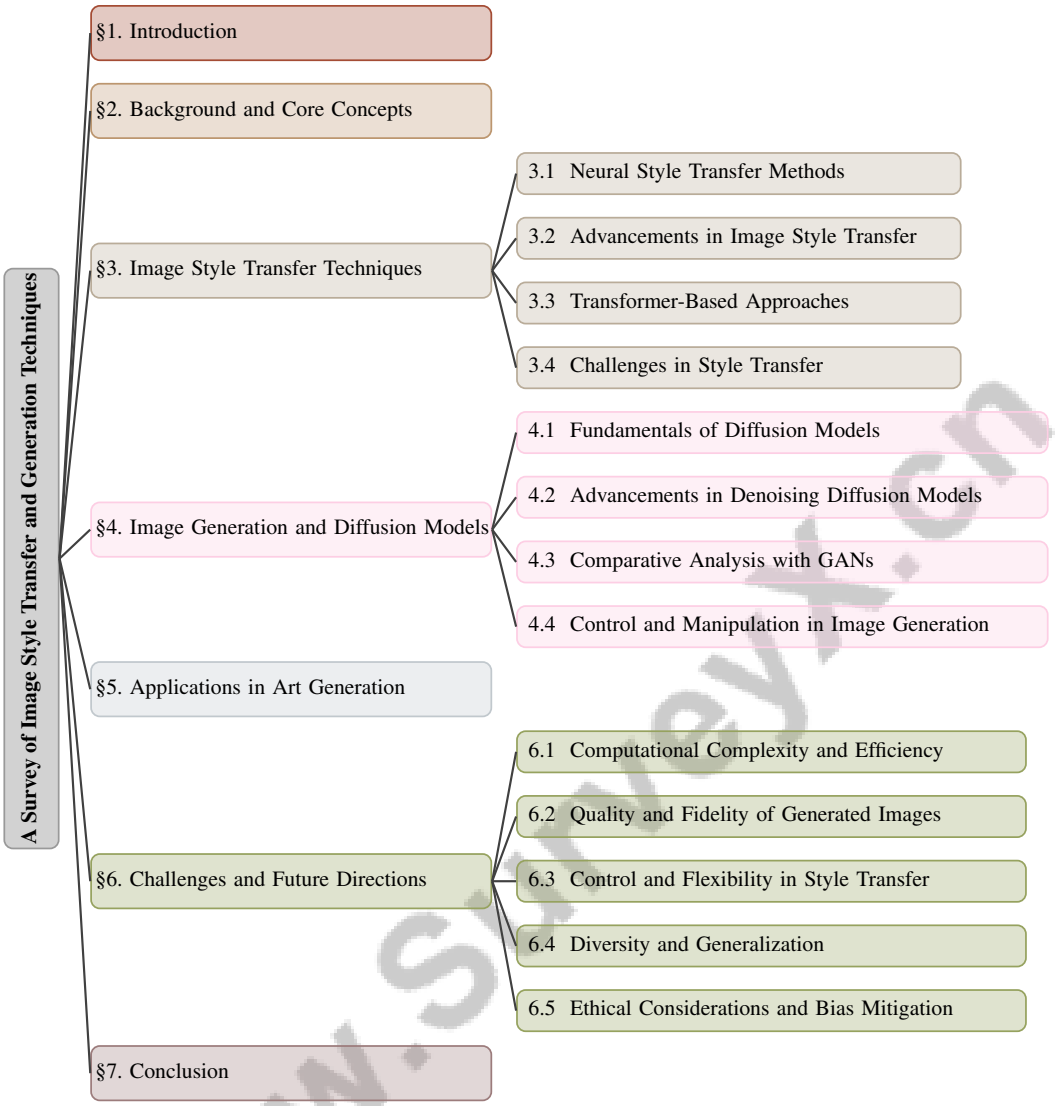


Figure 1: chapter structure

1.2 Role of Deep Learning in Image Generation

Deep learning and neural networks have revolutionized image style transfer and generation, enhancing efficiency, quality, and creative potential. Convolutional Neural Networks (CNNs) facilitate the transfer of artistic styles between images by learning transformation matrices that maintain semantic content while adapting stylistic features [5]. The introduction of perceptual loss functions, which utilize high-level features from pretrained networks, has further improved the quality of real-time transformations, ensuring high fidelity in stylized outputs [12].

Advanced models such as the Spatiotemporal Style Transfer (STST) algorithm showcase deep learning's capability to synthesize dynamic stimuli while retaining low-level features, particularly in video content generation [8]. The Content-Style Disentanglement (CSD) method exemplifies the separation of content and style representations, enhancing the flexibility and effectiveness of style transfer techniques [6].

Innovative approaches like the edge-enhanced image style transfer method (STT) incorporate novel edge loss functions to improve content details in stylized images [13]. Meanwhile, SC-StyleGAN introduces spatial constraints into the StyleGAN generation process, allowing for precise control over artistic outputs [14].

AI-generated, for reference only.

The emergence of text-to-image diffusion models and language-driven style transfer methods further broadens creative possibilities, utilizing human language and advanced image-text encoders to enhance controllability in visual effects. Additionally, integrating spatial constraints into generative adversarial networks (GANs) enables more nuanced and personalized image generation, reflecting deep learning's adaptability to diverse artistic preferences [15].

Despite advancements, optimizing neural networks for style transfer remains challenging, as traditional designs have primarily focused on classification tasks. However, ongoing developments, such as post-processing image enhancement models, continue to enhance visual quality and artistic innovation [16]. The StyleBank approach allows for explicit style representation, improving flexibility and efficiency in style transfer [17]. The Plug and Play Generative Networks (PPGNs) leverage deep learning to enhance the quality and diversity of generated images [11].

The Forward Stretching method employs topological insights for direct style transfer to arbitrary shapes, highlighting deep learning's role in advancing image style transfer [10]. The multi-style transfer (MST) issue in neural image processing underscores limitations in existing methods like Neural Style (NS) and Fast Neural Style (FNS), which either lack flexibility or are computationally expensive [18]. As these technologies evolve, they are set to redefine image style transfer and generation, offering unprecedented opportunities for creative exploration.

1.3 Structure of the Survey

This survey systematically presents a comprehensive overview of image style transfer and generation techniques, emphasizing their significance in creative and artistic domains. It begins with an introduction highlighting the transformative impact of these technologies, followed by a detailed discussion on deep learning's role in enhancing image generation capabilities. Subsequent sections define key terms such as image style transfer, image generation, diffusion models, and generative models, establishing a foundational understanding.

The survey explores various image style transfer techniques, starting with neural style transfer methods and their evolution, discussing recent advancements and challenges in achieving high-quality style transfer, and examining transformer-based approaches and their impact on the field. Challenges such as computational complexity and control over style and content are also identified.

The intricacies of image generation and diffusion models are discussed, focusing on the foundational principles of diffusion models and their advantages over traditional methods like GANs and autoregressive Transformers. The survey provides a comprehensive overview of these models' evolution, their applications in generating images from text, and their performance in tasks such as image editing and video generation, while addressing current challenges and future directions [19, 20, 21]. Recent advancements in denoising diffusion models and their applications are compared with other generative models like GANs, exploring techniques for controlling and manipulating generated images.

The application of image style transfer and generation techniques in art creation is discussed through innovative techniques and case studies. The exploration of large-scale text-to-image generation models and style transfer techniques reveals significant implications for artistic expression and creativity, alongside ethical considerations regarding automation, preservation of artistic integrity, and the potential alteration of traditional practices [22, 23, 24, 25].

Finally, the survey identifies current challenges and potential future research directions, addressing issues related to computational complexity, quality, fidelity, control, flexibility, diversity, generalization, and ethical considerations. The conclusion emphasizes the significance of the discussed techniques in advancing artistic style transfer, summarizing key insights that enhance the separation of content and style, improve personalization in generative models, and address compositional challenges in text-to-image generation [22, 26, 27, 28].The following sections are organized as shown in Figure 1.

2 Background and Core Concepts

2.1 Definitions and Key Terms

Image style transfer involves the application of a reference image's artistic style to a target image, modifying its visual aesthetics while retaining its semantic content [9]. This is primarily executed

through neural networks, notably Neural Style Transfer (NST), which has been pivotal in artistic style applications [5]. Traditional methods, often constrained by predefined style images, have limited flexibility, but recent advancements allow for style transfer on segmented objects, ensuring structural integrity across multiple views [9]. Applications have expanded to include makeup transfer, where cosmetic styles are applied while preserving the subject's identity.

Image generation refers to the creation of new visual content from various inputs like sketches and textual descriptions, essential for producing diverse images that capture both stylistic and conceptual elements, such as generating images from hand-drawn sketches [29]. Text-to-image generation, a specific subset, uses textual descriptions to create images reflecting complex prompts, integrating multiple objects and their interrelations [30]. The integration of aesthetic gradients and CLIP-conditioned diffusion models enhances personalization and aesthetic quality in generated images.

Diffusion models are generative models that iteratively refine noisy data into coherent outputs, providing distinct advantages over traditional methods in producing high-fidelity images [11]. These models are particularly effective in scenarios requiring recursive refinement, such as improving low-quality images to achieve outputs with natural visual characteristics and intricate semantic attributes. Style-Extracting Diffusion Models (STEDM) illustrate the capability of diffusion models to generate images with novel styles while preserving known content, emphasizing their relevance in creative domains.

Generative models, including Generative Adversarial Networks (GANs) and diffusion probabilistic models, are foundational to image synthesis. They facilitate the generation of photorealistic images and manipulation of specific attributes, such as disentangling spatial content and styles [15]. The development of multi-content GANs (MC-GAN) demonstrates the ability to generate unobserved glyphs from limited examples, highlighting the versatility and robustness of generative models in creating diverse outputs [29]. Conditional generative adversarial networks (cGAN) have been effectively employed in applications like handwriting imitation and style transfer, showcasing the adaptability of generative models across various domains [15].

2.2 Interrelation of Core Concepts

The interrelation of core concepts in image style transfer and generation is crucial for understanding technological advancements and their creative applications. Generative models like GANs and diffusion models are foundational for synthesizing realistic and diverse images. The complexity of disentangling latent spaces in GANs is vital for enhancing image quality and enabling precise modifications in editing tasks [31]. This underscores the need for interpretative methods that allow users to customize and personalize generated content effectively [32].

Diffusion models, with their iterative refinement capabilities, offer significant advantages in producing high-fidelity images, particularly in recursive enhancement scenarios [33]. These models maintain semantic attributes while addressing challenges related to color and texture distortions, thus ensuring the naturalness of restored results [33]. The permutation invariance and complex intrinsic dependencies in graph structures further highlight the discrete nature of these models, enabling effective handling of diverse and complex prompts [34].

Personalization within generative models significantly influences the interrelation of core concepts. Effective representation and editing of specific visual attributes are crucial for tailoring generated images to user preferences [26]. This personalization is particularly relevant in large-scale text-to-image generation models, where user roles and collaborative exploration enhance creative outputs [24].

Moreover, the synthesis of images that seamlessly integrate foreground objects into backgrounds while preserving their identity underscores the importance of compositional harmony in image generation [35]. This compositional approach is vital for generating high-quality, diverse, and realistic synthetic images that align with user prompts and demographic attributes [36]. Ongoing challenges related to high computational costs and the necessity for large-scale datasets continue to drive research and development, emphasizing the interconnected nature of these core concepts [21].

In recent years, the field of image style transfer has witnessed significant advancements, particularly with the introduction of transformer-based approaches. These developments have not only enhanced

the quality of style transfer but have also presented new challenges that researchers are striving to overcome. To illustrate this progression, Figure 2 provides a comprehensive overview of the hierarchical categorization of image style transfer techniques. This figure highlights the key methods and innovations within each area, effectively summarizing the current landscape of advancements and challenges in the domain. By examining this categorization, one can gain a clearer understanding of the evolution of image style transfer techniques and the impact of emerging methodologies on the field.

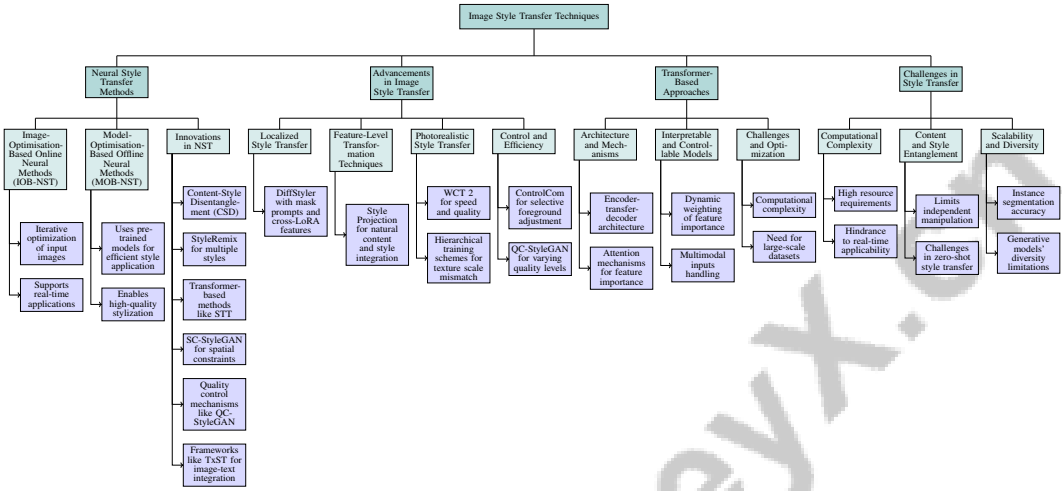


Figure 2: This figure shows the hierarchical categorization of image style transfer techniques, advancements, transformer-based approaches, and challenges, highlighting the key methods and innovations in each area.

3 Image Style Transfer Techniques

3.1 Neural Style Transfer Methods

Neural Style Transfer (NST) has advanced significantly, enhancing artistic stylization by balancing content preservation with style application. NST methods are divided into Image-Optimisation-Based Online Neural Methods (IOB-NST) and Model-Optimisation-Based Offline Neural Methods (MOB-NST). IOB-NST involves iterative optimization of input images, while MOB-NST uses pre-trained models for efficient style application [5]. This division supports both real-time applications and high-quality stylization. Figure 3 illustrates the categorization of Neural Style Transfer (NST) methods into Image-Optimisation-Based and Model-Optimisation-Based approaches, highlighting key techniques and innovations in each category.

Innovations such as Content-Style Disentanglement (CSD) leverage techniques like triplet and cycle-consistency losses to improve content and style separation, enhancing NST adaptability [6]. StyleRemix integrates multiple styles within a single network, facilitating smooth transitions and manipulations [18]. These developments reflect ongoing efforts to refine NST for nuanced artistic outputs.

Transformer-based methods like STT use an encoder-transfer-decoder architecture to merge content and style features, highlighting the role of advanced neural architectures in enhancing NST outputs [13]. SC-StyleGAN incorporates spatial constraints into the StyleGAN framework, enhancing sketch-based portrait generation by integrating spatial awareness [14].

Quality control mechanisms such as QC-StyleGAN enable controllable quality in image generation, meeting diverse artistic and practical requirements [37]. Frameworks like TxST use image-text encoders for style transfer guided by textual descriptions, broadening NST's creative possibilities [2].

In video style transfer, Spatiotemporal Style Transfer (STST) extends traditional techniques to dynamic content, maintaining spatiotemporal features across frames [8]. Photorealistic methods using

wavelet transforms enhance structural preservation and realism [38]. Learning linear transformations for rapid style transfer introduces a learnable matrix for efficient application [7].

The Forward Stretching method applies style to irregularly shaped instances through tensor transformations, while StyleBank offers a feed-forward network with convolution filter banks for each style, allowing style transfers without retraining the entire network [17]. Recent advancements in NST methodologies highlight a dynamic evolution, broadening their applicability in artistic and creative fields. Gatys et al.'s foundational work using Convolutional Neural Networks (CNNs) established NST principles, prompting extensive research aimed at refining and extending these algorithms [39, 5].

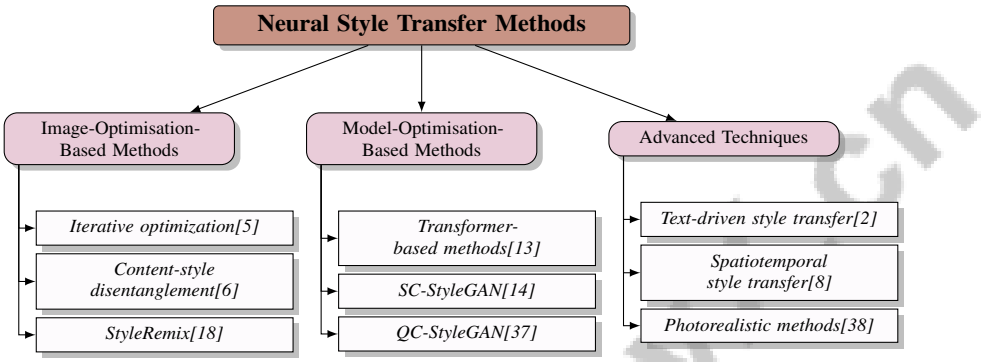


Figure 3: This figure illustrates the categorization of Neural Style Transfer (NST) methods into Image-Optimisation-Based and Model-Optimisation-Based approaches, highlighting key techniques and innovations in each category.

3.2 Advancements in Image Style Transfer

Method Name	Methodological Innovations	Application Scenarios	Challenges and Limitations
DS[1]	Mask Prompts	Creative Domains	Semantic Integrity
SP[40]	Style Projection	Real-time Applications	Computational Complexity
MT[4]	Hierarchical Architecture	Real-time Stylization	Scale Mismatches
CC[35]	Two-stage Fusion	Realistic Composite Images	Low-quality Input
QC-StyleGAN[37]	Qc-StyleGAN	Image Restoration	Computational Complexity
LLT[7]	Linear Transformation Matrix	Real-time Applications	Computationally Intensive
STT[13]	Edge Loss	Creative Domains	Content Leak
SPNST[3]	Graph Neural Networks	Creative Domains	Content-style Mismatching
SB[17]	Stylebank Filter	Region-specific Transfer	Network Retraining Requirement
TxST[2]	Contrastive Training Strategy	Artist-aware Style	Extensive Training Data

Table 1: This table provides a comprehensive overview of various image style transfer methods, highlighting their methodological innovations, application scenarios, and associated challenges and limitations. It serves as a valuable resource for understanding the advancements and ongoing challenges in the field of image stylization, emphasizing both creative and practical applications.

Recent advancements in image style transfer have enhanced the quality, efficiency, and flexibility of stylization techniques, expanding their applicability in creative domains. Table 1 presents a detailed comparison of recent advancements in image style transfer methods, illustrating their unique innovations, practical applications, and the challenges they face. DiffStyler exemplifies innovation by using mask prompts and cross-LoRA features for localized style transfer, allowing precise control over stylistic modifications within specific image regions [1]. This approach addresses the challenge of achieving high-quality localized style applications while maintaining overall image coherence.

Style Projection, a parameter-free feature-level transformation technique, facilitates natural content and style integration by reordering style features to align with content features [40]. This method enhances blending, ensuring stylized outputs retain artistic integrity and content fidelity.

In photorealistic style transfer, WCT 2 delivers superior speed and quality without requiring post-processing [38]. This method effectively preserves photorealism while transferring style, maintaining structural and semantic attributes of the original image.

Hierarchical training schemes in multimodal transfer networks address texture scale mismatch, generating visually appealing results on high-resolution images [4]. ControlCom employs a conditional diffusion model to selectively adjust foreground attributes, enhancing user control over stylistic outcomes [35]. Such advancements are essential for tailoring style transfer applications to specific artistic or practical needs.

QC-StyleGAN has improved image generation and manipulation across varying quality levels, addressing existing GAN model limitations [37]. Recent methods have also achieved computational efficiency, reducing overhead while preserving content fidelity, pivotal for real-time applications [7].

Challenges persist in achieving consistent style transfer quality across diverse inputs and managing computational complexity. The STT method captures long-range dependencies and preserves structural information through a transformer architecture, illustrating efforts to address these challenges [13]. However, techniques often struggle with content-style mismatching, leading to distorted local patterns or artifacts [3]. Continued research is essential to overcome these limitations, ensuring style transfer technologies evolve to meet diverse artistic and practical demands.

StyleBank demonstrates improvements in style transfer efficiency and flexibility, enabling easy addition of new styles and region-specific transfer capabilities [17]. The CSGO framework achieves state-of-the-art results in style control and content retention, marking a notable advancement in the field [41]. The TxST approach, utilizing a contrastive training strategy and novel attention module, effectively aligns stylization with text descriptions, representing a significant advancement over existing methods [2].

3.3 Transformer-Based Approaches

Transformers have emerged as a transformative force in image style transfer, offering novel architectures that enhance the flexibility, efficiency, and quality of stylization processes. The encoder-transfer-decoder architecture characteristic of transformer-based approaches effectively merges content and style features by capturing long-range dependencies, addressing traditional convolutional neural networks' limitations in managing global context and complex transformations [13].

A key innovation in transformer-based style transfer is the integration of attention mechanisms, allowing dynamic weighting of feature importance across input images. This capability enables models to focus on salient regions, resulting in coherent and visually appealing outputs. The STT method exemplifies this by incorporating edge loss functions that enhance content clarity and detail, demonstrating transformers' potential to improve both artistic quality and structural integrity of stylized images [13].

Transformers also facilitate the development of interpretable and controllable models. By leveraging attention mechanisms, these models provide insights into the decision-making processes of style transfer, allowing users to manipulate specific stylization aspects, such as color, texture, and spatial arrangement. This level of control is crucial for applications requiring precise customization of artistic outputs, such as digital art creation and design [2].

The application of transformers is further enhanced by their ability to handle multimodal inputs, such as text and images. The TxST approach, for instance, employs image-text encoders to align stylization with textual descriptions, expanding creative possibilities and practical applications of style transfer technologies. This integration of textual guidance enhances the expressiveness and diversity of generated images, facilitating personalized artistic content creation [2].

Despite significant advancements, challenges remain in computational complexity and the need for large-scale datasets for effective model training. The substantial resource requirements of transformer models underscore the urgent need for continued research focused on enhancing efficiency and scalability. This optimization is crucial for enabling widespread adoption of transformers across diverse applications, particularly in advanced image style transfer techniques that leverage long-range dependencies to maintain content integrity while applying artistic styles. By refining these models, we ensure their advantages are accessible across various technological contexts [42, 43, 13]. The impact of transformers on image style transfer is undeniable, offering a powerful framework for advancing the state of the art in artistic and creative domains.

3.4 Challenges in Style Transfer

The domain of image style transfer faces challenges arising from computational complexity, the intricate balance of style and content, and limitations in scalability and efficiency. A primary obstacle is the high computational demand associated with neural style transfer (NST) methods, which often require substantial processing resources and time, hindering real-time applicability [18]. This issue is compounded by the need to maintain coherence and fidelity of fine structures within images, frequently compromised in the pursuit of stylization.

The entanglement of content and style presents another significant challenge, limiting independent manipulation of these elements and restricting the flexibility and effectiveness of style transfer techniques [18]. Traditional methods often struggle to adequately model the relationship between content and style, resulting in incomplete or distorted stylization outcomes. This challenge is exacerbated by existing methods' inability to effectively separate and recombine style and content information, crucial for scenarios requiring zero-shot style transfer.

Effective instance segmentation is vital for methods transferring style to arbitrary instances within images. The quality of style transfer heavily depends on instance segmentation accuracy, which can vary significantly and affect overall stylization quality [10]. Additionally, the necessity to retrain the entire network for each new style limits the scalability and efficiency of current methods, particularly when multiple styles are needed [17].

Moreover, current generative models' failure to capture full diversity leads to biased outputs and amplifies biases present in training datasets [29]. This limitation critically affects the fairness and inclusivity of style transfer applications, necessitating the development of models that better represent the diversity inherent in artistic styles.

Addressing the challenges in image style transfer technologies is crucial for enhancing capabilities, particularly in enabling users to transfer artistic styles based solely on textual descriptions instead of requiring reference images. This advancement broadens application ranges, allowing for diverse artistic expressions and practical uses while ensuring high standards of quality and efficiency, preserving content details, and minimizing issues such as content leakage [44, 13].

4 Image Generation and Diffusion Models

Understanding diffusion models' transformative impact in generative modeling requires examining their core principles. This section delves into diffusion models' foundational concepts, emphasizing their iterative processes that distinguish them from traditional generative techniques and enable advancements and applications in the field.

4.1 Fundamentals of Diffusion Models

Diffusion models represent a significant leap in generative modeling, characterized by iterative refinement processes that transform noisy inputs into coherent, high-quality outputs. This denoising process sets diffusion models apart from traditional techniques like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), offering enhanced image quality and diversity [29]. By systematically adding noise and reversing it through denoising steps, diffusion models maintain semantic integrity and improve visual quality.

A key strength of diffusion models is their operation in both pixel and latent spaces, providing greater flexibility. Classifier-Guided Diffusion Generation (CGDG) employs calibration techniques to refine classifier guidance, enhancing image quality [45]. Cross-attention mechanisms in Diff-Text further improve object placement and scene coherence, demonstrating adaptability in generating complex visual content [30].

Diffusion models excel in handling multimodal inputs, such as text and images, enabling the generation of images that accurately reflect intricate prompts. Techniques like Sketch-Guided Scene Image Generation (SGSIG) leverage diffusion models to create detailed scene images from sketches [46]. Additionally, Diversity-Aware Diffusion Models (DiADM) enhance unconditional diffusion models by integrating a diversity-aware module, disentangling diversity from image fidelity to improve output range and quality [29].

However, diffusion models face computational inefficiencies due to extensive iterations required for image generation. Innovative strategies, such as similarity-guided training and adversarial noise selection in TAN, enhance transfer learning capabilities in diffusion probabilistic models [47]. This not only improves computational efficiency but also extends diffusion models' applicability to video generation, evidenced by promising results in high-quality image synthesis [48].

The rise of diffusion models has prompted discussions on societal risks, including potential for generating harmful content and copyright infringement, emphasizing the need for ethical considerations in their deployment [49]. Nevertheless, ongoing developments continue to push generative modeling boundaries, with innovations like latent code optimization in Plug and Play Generative Networks (PPGNs) enhancing image quality and diversity, showcasing diffusion models' potential to redefine image synthesis [11].

4.2 Advancements in Denoising Diffusion Models

Recent advancements in denoising diffusion models have significantly improved image synthesis capabilities, enhancing both quality and efficiency. Accurate Post-training Quantization for Diffusion Models (APQ-DM) reduces quantization errors using distribution-aware quantization functions and optimizing timestep selection for calibration image generation [50], addressing computational challenges and enhancing applicability in resource-constrained environments.

The Ditail framework exemplifies diffusion models' ability to mix domain-specific models without additional training, leveraging a broader array of diffusion models to create novel images [51], highlighting their flexibility in generating diverse outputs across various domains.

In image processing, FreeEnhance has shown superior performance in enhancing image quality and human preference compared to existing state-of-the-art methods [16], demonstrating diffusion models' potential to improve visual quality across applications.

Scalable tokenization-free diffusion models by Palit eliminate complexities associated with tokenization and positional embeddings, employing fixed-size core structures for image processing [52], simplifying the diffusion process and reducing computational overhead while facilitating high-quality image generation.

In medical imaging, diffusion-based approaches generate synthetic medical images for training convolutional neural networks (CNNs), addressing data scarcity in medical image analysis [53], underscoring diffusion models' potential to contribute to critical fields by providing high-quality training data.

The UNLEARNCANVAS framework standardizes evaluation for assessing artistic style unlearning in diffusion models, facilitating comparisons of different machine unlearning methods [49], emphasizing ethical considerations in deploying diffusion models.

Advancements in transfer learning efficiency have been achieved through the TAN method, which estimates divergence between source and target domains using similarity-guided training and employs adversarial noise selection [47], enhancing adaptability to new domains and expanding utility in diverse applications.

Recent advancements in denoising diffusion models exemplify a significant evolution in image synthesis technology, marked by improvements in quality, diversity, and control. These models now handle sophisticated tasks such as zero-shot interpolation between images, enhancing creative applications. Innovations like the Diffusion^2 framework leverage strengths of video and multi-view diffusion models to produce seamless 4D content efficiently, while latent diffusion models optimize computational resources without sacrificing fidelity. Collectively, these developments push image generation boundaries, paving the way for versatile applications across modalities, including text-guided generation and high-resolution synthesis [21, 54, 55, 56].

As illustrated in Figure 4, these advancements can be categorized into three main areas: image synthesis, model efficiency, and specialized applications. The figure highlights significant methods such as APQ-DM, Ditail, and FreeEnhance, which focus on improving image generation quality, alongside efficiency-oriented models like STOIC and TAN. Furthermore, it presents specialized applications in medical imaging and artistic style unlearning, showcasing the diverse utility of diffusion models across various domains.

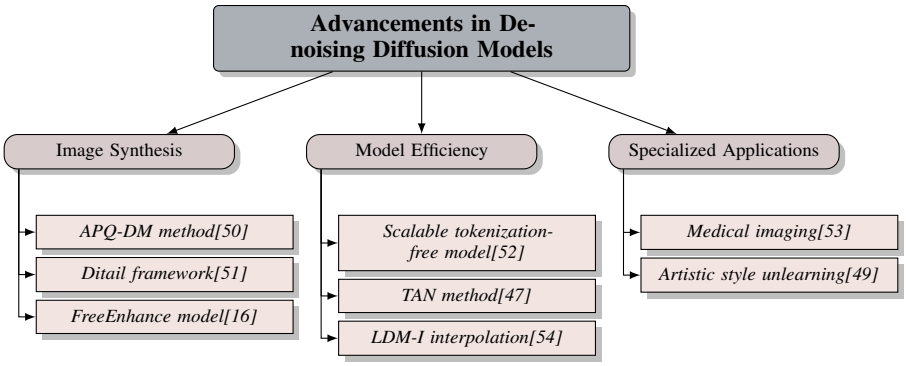


Figure 4: This figure illustrates the recent advancements in denoising diffusion models, categorized into image synthesis, model efficiency, and specialized applications. It highlights significant methods such as APQ-DM, Ditail, and FreeEnhance for improving image generation quality, alongside efficiency-focused models like STOIC and TAN. Additionally, specialized applications in medical imaging and artistic style unlearning are presented, showcasing the diverse utility of diffusion models across domains.

4.3 Comparative Analysis with GANs

Diffusion models and Generative Adversarial Networks (GANs) represent two prominent paradigms in generative modeling, each with distinct methodologies and capabilities. Diffusion models are defined by their iterative denoising process, refining noisy inputs into high-fidelity outputs, ensuring robust semantic integrity and visual quality. This iterative approach contrasts with GANs' generator-discriminator framework, which, while capable of producing sharp images, often encounters challenges like mode collapse and training instability [15].

Advancements in diffusion models, such as the Residual Video Diffusion (RVD) method, demonstrate their superiority in video generation by correcting deterministic predictions with stochastic residuals, enhancing video dynamics representation [48]. This capability highlights diffusion models' versatility in handling dynamic content, surpassing traditional GANs in this domain.

The Sketch-Guided Scene Image Generation (SGSIG) method further exemplifies diffusion models' strengths, achieving superior performance in generating coherent scene images from sketches, effectively balancing object fidelity with background integration [46]. This adaptability allows diffusion models to integrate features from multiple inputs without extensive retraining, unlike GANs, which typically require significant adjustments for new modalities.

Benchmark evaluations of state-of-the-art diffusion models, including DALL-E 2 and Stable Diffusion 2, reveal superior performance across various image generation tasks, evidenced by favorable FID scores and human evaluations. These advanced models excel in high-quality text-image alignment and synthesis, particularly through integrating large vision-language models and innovative techniques like aesthetic gradients and prompt spectrum representation. This multifaceted approach enhances accuracy in generating images related to complex textual descriptions and allows personalized aesthetic adjustments, resulting in diverse and intricately detailed outputs [57, 58, 27, 59, 26].

While both diffusion models and GANs offer unique advantages, the choice between them often depends on specific application requirements. Diffusion models provide a robust framework for generating high-quality images with intricate prompts, while GANs are favored for their rapid generation capabilities and realistic outputs [15]. Continuous advancements in both fields enhance their strengths, ensuring ongoing relevance and application in generative modeling research.

4.4 Control and Manipulation in Image Generation

Control and manipulation of generated images are critical aspects of image synthesis, determining the extent of user influence over creative outputs. A significant challenge is the limited user intervention in text-to-image (T2I) models that operate end-to-end, restricting creative control and customization essential for artistic applications [60].

Recent advancements have introduced methods for localized control over image generation. Techniques enabling precise placements of objects and styles within specified regions enhance user flexibility in composing generated scenes [61]. Additionally, Filter Style Transfer (FST) allows real-time filter application while maintaining high quality, even with unseen filters, expanding creative possibilities [62].

Dynamic attention mechanisms, exemplified by StyleMaster, enhance control capabilities by allowing users to adjust attention across different image regions, facilitating better management of stylistic attributes and addressing fidelity and consistency challenges in image synthesis [63].

Adaptive patch-based manipulation techniques effectively reduce artifacts and improve manipulated image quality. By enabling fine-grained control over specific regions, these techniques maintain the integrity of generated content while allowing creative modifications [3].

In large-scale text-to-image generation models (LTGMs), challenges remain in incorporating these models into creative workflows and overcoming biases that may hinder creativity [24]. Existing models often struggle to align generated images with input segmentation maps, resulting in imprecise outputs. Addressing these challenges is vital for enhancing the precision and accuracy of generated images, improving their applicability in artistic and commercial domains [64].

Moreover, the risk of copyright infringement due to style imitation by generative models, particularly in artwork generated by diffusion models, highlights the need for robust control mechanisms to protect artistic styles from unauthorized replication [65]. Techniques that harmonize layer sequences while minimizing changes to specific properties, such as color and texture, are essential for maintaining input layer integrity and ensuring authenticity in generated outputs [66].

Recent advancements in image generation technologies, including aesthetic gradients and the Prompt Spectrum approach, reflect a concerted effort to enhance user control and customization. These innovations allow users to guide the generative process toward desired aesthetics and manipulate specific visual attributes—such as material, style, and layout—effectively overcoming previous limitations in personalization. Such methods pave the way for sophisticated applications in the field, enabling tailored image outputs from minimal input without extensive fine-tuning of the underlying models [26, 57].

5 Applications in Art Generation

5.1 Innovative Techniques in Art Generation

Advancements in generative models and deep learning have significantly reshaped art generation, offering new creative possibilities. VisioBlend exemplifies the transformation by converting sketches into realistic images through sketch-guided denoising, highlighting generative models' capacity to produce detailed visuals from minimal input [67]. LayerDiff extends this capability with text-guided multilayered style transfer, enabling intricate modifications while maintaining stylistic coherence [68].

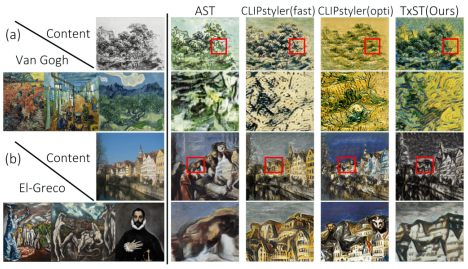
Zero-shot style transfer methods like Z-STAR facilitate nuanced artistic style generation without extensive retraining, offering efficiency for artists with limited computational resources [69]. Diffusion models further advance high-quality image generation from text, pushing traditional art boundaries and achieving rapid stylization of high-resolution images [21, 38].

Cross-modal style transfer frameworks convert auditory inputs, such as music, into visual art using conditional generative adversarial networks and contrastive learning, enhancing multimedia expressiveness [70, 71, 72, 73, 44]. Diffusion models also generate synthetic images for training, notably in medical imaging, demonstrating their versatility across domains [53].

Advanced computational models, including large-scale text-to-image generation and personalized aesthetic gradients, automate the creative process, expanding artists' conceptual horizons. Dynamic Memory Generative Adversarial Networks and content-style disentanglement approaches provide unprecedented tools for artists to articulate their visions [22, 23, 24, 57].

As shown in Figure 5, modern digital art techniques are redefining creativity. The first image presents a stylized landscape, demonstrating reinterpretation of natural beauty, while the second compares synthesis techniques on iconic paintings, expanding artistic appreciation and expression [74, 2].

(a) Stylized Image of a Mountain Landscape with Yellow Flowers and Pine Trees[74]



(b) Comparison of Image Synthesis Techniques for Van Gogh and El Greco Paintings[2]

Figure 5: Examples of Innovative Techniques in Art Generation

5.2 Impact on the Art World

The integration of image style transfer and generation technologies has democratized art creation, fostering inclusivity by allowing artists of all skill levels to explore diverse styles without traditional training [2]. Generative models have influenced the commercial art market, enabling artists to create and sell digital art, thus enhancing visibility and commercial potential [38, 37].

Diffusion models enable novel artistic expressions like text-to-image synthesis, enhancing storytelling by translating complex narratives into visual art [21]. This capability accelerates contemporary art evolution, blending traditional and digital mediums to enrich the cultural landscape [69].

Despite these advancements, ethical considerations arise regarding originality and authorship. The ease of style replication challenges traditional notions of ownership, necessitating measures to protect artists' rights while fostering innovation [65].

5.3 Ethical Considerations and Challenges

The rise of image style transfer and generation technologies presents ethical challenges concerning authenticity, originality, and potential misuse. Style imitation by generative models risks copyright infringements, compromising originality [65]. While adversarial perturbations offer some protection, they often degrade visual quality, underscoring the need for balanced solutions.

AI's role in art generation raises issues of misinformation and bias, as AI-generated images can perpetuate harmful stereotypes from biased training data [75]. Establishing ethical guidelines is crucial to ensure AI-generated content remains accurate and unbiased.

Ensuring authenticity and accuracy in AI-generated images is vital, particularly in sensitive contexts like journalism. Verification mechanisms and ethical standards should guide AI art creation and dissemination, acknowledging human contributions while respecting authenticity. Techniques like Dynamic Memory Generative Adversarial Networks can enhance artistic output while recognizing human artists' roles [23, 24, 76].

The democratization of art through AI tools prompts discussions about human creativity and traditional skills' value. As AI-generated art becomes prevalent, concerns about undermining human artistry's intrinsic value arise, necessitating a reassessment of artistic merit in the digital era [22, 23, 24].

6 Challenges and Future Directions

In image generation, computational efficiency is crucial for deploying advanced models in practical applications, especially as demand for real-time processing grows. This section examines the challenges of computational demands, focusing on diffusion models and their implications for efficiency in image and video generation tasks.

AI-generated, for reference only.

6.1 Computational Complexity and Efficiency

Diffusion models, while capable of high-quality output, face challenges in real-time application due to extensive sampling steps, limiting scalability [47]. This complexity is exacerbated in video generation, where temporal coherence across frames demands sophisticated methods to maintain consistency without excessive computational costs [48]. The TAN framework and StyleBank address these challenges by enhancing transfer learning capabilities and supporting incremental learning, respectively, allowing style transfer without complete network retraining [47, 17].

Despite these advancements, achieving efficient real-time performance remains difficult, particularly in complex style transfers or high-resolution outputs. Current Neural Style Transfer (NST) algorithms often struggle with preserving fine details or managing complex styles effectively [18]. Scene text generation also highlights challenges such as generating small-scale text and controlling text color, necessitating more efficient methods [30]. GAN models further complicate efficiency enhancement due to training instability and diversity output issues [15]. The complexity of generating images from intricate sketches can lead to identity loss and semantic blending, compromising quality [46].

The pursuit of computational efficiency in image generation necessitates advancements that enhance processing capabilities while addressing quality, diversity, and style consistency challenges. Although generative models can produce visually realistic images, they often fail to capture the full data distribution, limiting output diversity. The Image Retrieval Score (IRS) highlights that current diffusion models achieve only about 77

6.2 Quality and Fidelity of Generated Images

The quality and fidelity of generated images are critical for realism and aesthetic appeal in image synthesis. A significant challenge is the entanglement of stylistic attributes with structural elements, which can compromise semantic integrity during stylization [1]. The scarcity of high-quality, labeled datasets, especially in specialized domains like medical imaging, poses a core obstacle, potentially leading to model overfitting and limiting generalizability [53].

Advancements such as QC-StyleGAN enable direct manipulation of low-quality images while maintaining content consistency, offering advantages over traditional high-quality GANs [37]. This capability is vital for applications requiring enhancement of lower-quality inputs without sacrificing original content integrity. The performance and fidelity of generated images heavily depend on the quality of training datasets, underscoring a persistent challenge in the field [6].

In video generation, techniques like Residual Video Diffusion (RVD) excel at producing sharp, high-resolution frames and effectively modeling complex dynamics, outperforming traditional methods [48]. Despite these advancements, challenges persist in achieving perfect accuracy and fidelity, particularly in complex scenes where style and content intertwining can introduce artifacts or distortions. Ongoing research is vital for developing robust methodologies that can reliably produce high-quality, realistic outputs in diverse contexts, ensuring that generative models, especially GANs, continue to advance the boundaries of image synthesis. This includes addressing challenges such as generating high-resolution images with multiple objects, establishing reliable evaluation metrics aligned with human judgment, and improving architectural designs and training processes. Continuous exploration and innovation are necessary to enhance visual realism, diversity, and semantic alignment in generated content [58, 15, 32, 77].

6.3 Control and Flexibility in Style Transfer

Control and flexibility in style transfer require balancing content integrity preservation with effective stylistic attribute application. Region-Controlled Style Transfer (RCST) enhances control by introducing weighted Mean Squared Error (MSE) content loss, reducing color leakage and allowing controlled texture intensity based on region smoothness [78]. ArtFlow prevents content leakage and preserves content image integrity, offering a robust framework for achieving control and flexibility in style transfer [79]. The Wavelet Diffusion method enhances control by leveraging frequency information, improving detail and achieving faster convergence, addressing traditional diffusion models' limitations [80].

LayerDiff allows detailed control over individual layers in generated images, essential for applications in graphic design and digital artistry requiring multi-layered compositions [68]. Kim's method

simplifies feature alignment by producing independent channels, reducing computational complexity and improving processing speeds, enhancing flexibility in style transfer [81]. Challenges related to achieving visual consistency and coherence in video style transfer are addressed by methods focusing on maintaining consistency during motion [82].

Despite advancements, control and flexibility in style transfer remain ongoing challenges. The TxST method, while effective in many scenarios, can struggle with highly specific or nuanced styles requiring extensive training data, indicating the need for further development [2]. The integration of dual supervision approaches, as discussed in the UNLEARNCANVAS framework, highlights the potential for nuanced assessments of style transfer performance, particularly in unlearning targets combining artistic styles and object classes [49].

The continuous pursuit of enhanced control and flexibility in artistic style transfer technologies necessitates ongoing innovation to address inherent challenges, such as the need for pre-selected style images and limitations in creativity and accessibility. Recent advancements, including language-driven artistic style transfer (LDAST) and arbitrary style transfer through multi-adaptation networks, showcase the potential of using natural language instructions to enhance controllability and accessibility. These innovative approaches leverage advanced techniques like contrastive learning and feature disentanglement to enable nuanced style manipulation, allowing detailed content structure preservation while adapting style patterns based on user-defined criteria. As these technologies evolve, they promise to expand creative possibilities for artists and designers, making style transfer more intuitive and effective [70, 78, 71]. Addressing these challenges can enhance the precision, adaptability, and creative potential of style transfer applications, paving the way for more sophisticated artistic expressions.

6.4 Diversity and Generalization

Achieving diversity and generalization in generated content is a critical challenge in image synthesis, as current methodologies often struggle with constraints imposed by their training datasets. The pursuit of diversity is exemplified by approaches such as the Diversified Feature Projection (DFP) method, which generates a wide array of outputs without additional learning processes, distinguishing it from style-specific methodologies [83]. However, diversity assessment effectiveness is often compromised by inadequate existing metrics, which rely heavily on feature extractors that may introduce variability in performance, thus affecting reliability [29].

Generalization presents its own challenges, particularly regarding models' ability to extend learned capabilities to novel and unseen data. Many models struggle when applied to content classes not included in their training datasets, as observed in the context of content transformation blocks [28]. This limitation underscores the necessity for robust neural network designs capable of accommodating a broader spectrum of styles and content variations [74].

Exploring multimodal priors in image generation adds complexity, necessitating future research to enhance method robustness in handling intricate interdependencies among different modalities [84]. Dataset size constraints further restrict generalizability, as evidenced by benchmarks like CSGO [41].

Moreover, generalizing models to diverse poses and backgrounds remains a substantial hurdle, particularly in applications like pose-guided fashion image generation, where complex poses or unrepresented background variations can impede performance [85]. Similarly, models like StyleNAS face challenges in maintaining complexity and performance when applied to diverse styles or larger datasets, highlighting the need for further optimization [86].

Future research should focus on optimizing fine-tuning processes and extending methods to additional categories and tasks, thereby enhancing both diversity and generalization in image generation [87]. Developing more stable training algorithms and addressing challenges in generating high-quality outputs in real-time applications are crucial for advancing the field, ensuring generative models can produce varied and representative outputs across a wide array of applications and domains [15].

6.5 Ethical Considerations and Bias Mitigation

The integration of image style transfer and generation technologies into various applications necessitates careful consideration of ethical challenges and bias mitigation strategies. A primary ethical concern is the potential for these models to perpetuate biases present in their training data, leading to

outputs that reinforce stereotypes or provide unfair representations [29]. Addressing this requires diversifying training datasets and implementing bias-aware training protocols to ensure models reflect a broad spectrum of perspectives and experiences. Ethical considerations in image generation also encompass dataset quality and its impact on stylized output realism, which is crucial for responsible technology use [10].

Reliance on pre-trained models, as seen in methods like StyleRemix, may limit adaptability to new styles not represented in the training data, highlighting the need for ongoing research to enhance model adaptability and ensure they can generalize beyond biases in initial datasets [18]. Moreover, the aesthetic evaluation of stylized images remains an open question, necessitating standardized metrics to assess algorithm performance and address potential biases [30].

The potential misuse of generative technologies, such as creating misleading or deceptive content, presents significant ethical challenges. Future research will explore applying perceptual loss functions to various image transformation tasks, addressing ethical considerations and potential bias mitigation [11]. Additionally, a robust evaluation framework for future developments in machine unlearning methods is crucial for addressing ethical challenges in deploying these technologies [46].

Future research could further enhance the edge loss mechanism and investigate additional architectural modifications to improve performance across various artistic styles, thereby addressing ethical considerations [17]. Additionally, optimizations in the noising and denoising stages of FreeEnhance could tackle challenges related to the quality and fidelity of generated images, contributing to ethical advancements in the field [47].

7 Conclusion

The exploration of image style transfer and generation techniques underscores their transformative impact within creative and artistic sectors, highlighting both significant progress and ongoing challenges. Advances in generative models, such as Plug and Play Generative Networks (PPGNs), have substantially enhanced the capacity to produce high-quality and diverse images, facilitating new creative possibilities for artists and designers. These technological strides enable the creation of photorealistic and stylistically varied outputs, expanding the scope of artistic innovation.

Despite these advancements, challenges in computational efficiency and style generalization persist, posing barriers to real-time application and broader adaptability. The complexity inherent in neural style transfer methods often results in considerable computational demands, limiting their practical implementation. Furthermore, the ability to generalize effectively across diverse styles while maintaining image quality is a critical area for further research, essential for improving the robustness and versatility of these technologies.

Ethical considerations, including issues of bias and copyright infringement, require ongoing attention to ensure the responsible use of generative models. Addressing these ethical challenges is vital for safeguarding the originality and integrity of artistic works, while fostering inclusive and equitable creative practices.

References

[1] Shaoxu Li. Diffstyler: Diffusion-based localized image style transfer, 2024.

[2] Zhi-Song Liu, Li-Wen Wang, Wan-Chi Siu, and Vicky Kalogeiton. Name your style: An arbitrary artist-aware image style transfer, 2024.

[3] Yongcheng Jing, Yining Mao, Yiding Yang, Yibing Zhan, Mingli Song, Xinchao Wang, and Dacheng Tao. Learning graph neural networks for image style transfer, 2023.

[4] Xin Wang, Geoffrey Oxholm, Da Zhang, and Yuan-Fang Wang. Multimodal transfer: A hierarchical deep convolutional neural network for fast artistic style transfer. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5239–5247, 2017.

[5] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. Neural style transfer: A review. *IEEE transactions on visualization and computer graphics*, 26(11):3365–3385, 2019.

[6] Sailun Xu, Jiazhi Zhang, and Jiamei Liu. Image style transfer and content-style disentanglement, 2021.

[7] Xueting Li, Sifei Liu, Jan Kautz, and Ming-Hsuan Yang. Learning linear transformations for fast image and video style transfer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3809–3817, 2019.

[8] Antonino Greco and Markus Siegel. A spatiotemporal style transfer algorithm for dynamic visual stimulus generation, 2024.

[9] Tsu-Jui Fu, Xin Eric Wang, and William Yang Wang. Language-driven image style transfer, 2022.

[10] Zhifeng Yu, Yusheng Wu, and Tianyou Wang. A method for arbitrary instance style transfer, 2019.

[11] Anh Nguyen, Jeff Clune, Yoshua Bengio, Alexey Dosovitskiy, and Jason Yosinski. Plug & play generative networks: Conditional iterative generation of images in latent space. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4467–4477, 2017.

[12] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution, 2016.

[13] Chiyu Zhang, Jun Yang, Zaiyan Dai, and Peng Cao. Edge enhanced image style transfer via transformers, 2023.

[14] Wanchao Su, Hui Ye, Shu-Yu Chen, Lin Gao, and Hongbo Fu. Drawinginstyles: Portrait image generation and editing with spatially conditioned stylegan, 2022.

[15] Ming-Yu Liu, Xun Huang, Jiahui Yu, Ting-Chun Wang, and Arun Mallya. Generative adversarial networks for image and video synthesis: Algorithms and applications. *Proceedings of the IEEE*, 109(5):839–862, 2021.

[16] Yang Luo, Yiheng Zhang, Zhaofan Qiu, Ting Yao, Zhineng Chen, Yu-Gang Jiang, and Tao Mei. Freeenhance: Tuning-free image enhancement via content-consistent noising-and-denoising process, 2024.

[17] Dongdong Chen, Lu Yuan, Jing Liao, Nenghai Yu, and Gang Hua. Stylebank: An explicit representation for neural image style transfer. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1897–1906, 2017.

[18] Hongmin Xu, Qiang Li, Wenbo Zhang, and Wen Zheng. Styleremix: An interpretable representation for neural image style transfer, 2019.

[19] Zhen Xing, Qijun Feng, Haoran Chen, Qi Dai, Han Hu, Hang Xu, Zuxuan Wu, and Yu-Gang Jiang. A survey on video diffusion models, 2024.

[20] Tianyi Zhang, Zheng Wang, Jing Huang, Mohiuddin Muhammad Tasnim, and Wei Shi. A survey of diffusion based image generation models: Issues and their solutions, 2023.

[21] Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, In So Kweon, and Junmo Kim. Text-to-image diffusion models in generative ai: A survey, 2024.

[22] Dmytro Kotovenko, Artsiom Sanakoyeu, Sabine Lang, and Bjorn Ommer. Content and style disentanglement for artistic style transfer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4422–4431, 2019.

[23] Qinghe Tian and Jean-Claude Franchitti. Text to artistic image generation, 2022.

[24] Hyung-Kwon Ko, Gwanmo Park, Hyeon Jeon, Jaemin Jo, Juho Kim, and Jinwook Seo. Large-scale text-to-image generation models for visual artists' creative works, 2023.

[25] Vladimir Arkhipkin, Andrei Filatov, Viacheslav Vasilev, Anastasia Maltseva, Said Azizov, Igor Pavlov, Julia Agafonova, Andrey Kuznetsov, and Denis Dimitrov. Kandinsky 3.0 technical report, 2024.

[26] Yuxin Zhang, Weiming Dong, Fan Tang, Nisha Huang, Haibin Huang, Chongyang Ma, Tong-Yee Lee, Oliver Deussen, and Changsheng Xu. Prospect: Prompt spectrum for attribute-aware personalization of diffusion models, 2023.

[27] Song Wen, Guian Fang, Renrui Zhang, Peng Gao, Hao Dong, and Dimitris Metaxas. Improving compositional text-to-image generation with large vision-language models, 2023.

[28] Dmytro Kotovenko, Artsiom Sanakoyeu, Pingchuan Ma, Sabine Lang, and Björn Ommer. A content transformation block for image style transfer, 2020.

[29] Mischa Dombrowski, Weitong Zhang, Sarah Cechnicka, Hadrien Reynaud, and Bernhard Kainz. Image generation diversity issues and how to tame them, 2024.

[30] Lingjun Zhang, Xinyuan Chen, Yaohui Wang, Yue Lu, and Yu Qiao. Brush your text: Synthesize any scene text on images via diffusion model, 2023.

[31] Ming Liu, Yuxiang Wei, Xiaohe Wu, Wangmeng Zuo, and Lei Zhang. A survey on leveraging pre-trained generative adversarial networks for image editing and restoration, 2022.

[32] Bolei Zhou. Interpreting generative adversarial networks for interactive image generation, 2022.

[33] Jiangtong Tan and Feng Zhao. Diffloss: unleashing diffusion model as constraint for training image restoration network, 2024.

[34] Chengyi Liu, Wenqi Fan, Yunqing Liu, Jiatong Li, Hang Li, Hui Liu, Jiliang Tang, and Qing Li. Generative diffusion models on graphs: Methods and applications, 2023.

[35] Bo Zhang, Yuxuan Duan, Jun Lan, Yan Hong, Huijia Zhu, Weiqiang Wang, and Li Niu. Controlcom: Controllable image composition using diffusion model, 2023.

[36] Harrison Rosenberg, Shimaa Ahmed, Guruprasad V Ramesh, Ramya Korlakai Vinayak, and Kassem Fawaz. Limitations of face image generation, 2023.

[37] Dat Viet Thanh Nguyen, Phong Tran The, Tan M. Dinh, Cuong Pham, and Anh Tuan Tran. Qc-stylegan – quality controllable image generation and manipulation, 2022.

[38] Jaejun Yoo, Youngjung Uh, Sanghyuk Chun, Byeongkyu Kang, and Jung-Woo Ha. Photorealistic style transfer via wavelet transforms. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9036–9045, 2019.

[39] Dan Ruta, Gemma Canet Tarrés, Andrew Gilbert, Eli Shechtman, Nicholas Kolkin, and John Collomosse. Diff-nst: Diffusion interleaving for deformable neural style transfer, 2023.

[40] Siyu Huang, Haoyi Xiong, Tianyang Wang, Bihan Wen, Qingzhong Wang, Zeyu Chen, Jun Huan, and Dejing Dou. Parameter-free style projection for arbitrary style transfer, 2022.

[41] Peng Xing, Haofan Wang, Yanpeng Sun, Qixun Wang, Xu Bai, Hao Ai, Renyuan Huang, and Zechao Li. Csgo: Content-style composition in text-to-image generation, 2024.

[42] Yingying Deng, Fan Tang, Weiming Dong, Chongyang Ma, Xingjia Pan, Lei Wang, and Changsheng Xu. Stytr2: Image style transfer with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11326–11336, 2022.

[43] Yingying Deng, Fan Tang, Weiming Dong, Chongyang Ma, Xingjia Pan, Lei Wang, and Changsheng Xu. Stytr^2: Image style transfer with transformers, 2022.

[44] Gihyun Kwon and Jong Chul Ye. Clipstyler: Image style transfer with a single text condition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 18062–18071, 2022.

[45] Jiajun Ma, Tianyang Hu, Wenjia Wang, and Jiacheng Sun. Elucidating the design space of classifier-guided diffusion generation, 2023.

[46] Tianyu Zhang, Xiaoxuan Xie, Xusheng Du, and Haoran Xie. Sketch-guided scene image generation, 2024.

[47] Xiyu Wang, Baijiong Lin, Daochang Liu, and Chang Xu. Efficient transfer learning in diffusion models via adversarial noise, 2023.

[48] Ruihan Yang, Prakhar Srivastava, and Stephan Mandt. Diffusion probabilistic modeling for video generation, 2022.

[49] Yihua Zhang, Chongyu Fan, Yimeng Zhang, Yuguang Yao, Jinghan Jia, Jiancheng Liu, Gaoyuan Zhang, Gaowen Liu, Ramana Rao Kompella, Xiaoming Liu, and Sijia Liu. Unlearncanvas: Stylized image dataset for enhanced machine unlearning evaluation in diffusion models, 2024.

[50] Changyuan Wang, Ziwei Wang, Xiuwei Xu, Yansong Tang, Jie Zhou, and Jiwen Lu. Towards accurate post-training quantization for diffusion models, 2024.

[51] Haoming Liu, Yuanhe Guo, Shengjie Wang, and Hongyi Wen. Diffusion cocktail: Mixing domain-specific diffusion models for diversified image generations, 2024.

[52] Sanchar Palit, Sathya Veera Reddy Dendi, Mallikarjuna Talluri, and Raj Narayana Gadde. Scalable, tokenization-free diffusion model architectures with efficient initial convolution and fixed-size reusable structures for on-device image generation, 2024.

[53] Abdullah al Nomaan Nafi, Md. Alamgir Hossain, Rakib Hossain Rifat, Md Mahabub Uz Zaman, Md Manjurul Ahsan, and Shivakumar Raman. Diffusion-based approaches in medical image generation and analysis, 2024.

[54] Clinton J. Wang and Polina Golland. Interpolating between images with diffusion models, 2023.

[55] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[56] Zeyu Yang, Zijie Pan, Chun Gu, and Li Zhang. Diffusion^2: Dynamic 3d content generation via score composition of video and multi-view diffusion models, 2024.

[57] Victor Gallego. Personalizing text-to-image generation via aesthetic gradients, 2022.

[58] Stanislav Frolov, Tobias Hinz, Federico Raue, Jörn Hees, and Andreas Dengel. Adversarial text-to-image synthesis: A review, 2021.

[59] Amir Hertz, Andrey Voynov, Shlomi Fruchter, and Daniel Cohen-Or. Style aligned image generation via shared attention, 2024.

[60] John Joon Young Chung and Eytan Adar. Promptpaint: Steering text-to-image generation through paint medium-like interactions, 2023.

[61] Álvaro Barbero Jiménez. Mixture of diffusers for scene composition and high resolution image generation, 2023.

[62] Jonghwa Yim, Jisung Yoo, Won joon Do, Beomsu Kim, and Jihwan Choe. Filter style transfer between photos, 2020.

[63] Chengming Xu, Kai Hu, Qilin Wang, Donghao Luo, Jiangning Zhang, Xiaobin Hu, Yanwei Fu, and Chengjie Wang. Artweaver: Advanced dynamic style integration via diffusion model, 2024.

[64] Guillaume Couairon, Marlène Careil, Matthieu Cord, Stéphane Lathuilière, and Jakob Verbeek. Zero-shot spatial layout conditioning for text-to-image diffusion models, 2023.

[65] Namhyuk Ahn, Wonhyuk Ahn, KiYoon Yoo, Daesik Kim, and Seung-Hun Nam. Imperceptible protection against style imitation from diffusion models, 2024.

[66] Vishnu Sarukkai, Linden Li, Arden Ma, Christopher Ré, and Kayvon Fatahalian. Collage diffusion, 2023.

[67] Harshkumar Devmurari, Gautham Kuckian, Prajjwal Vishwakarma, and Krunali Vartak. Visioblend: Sketch and stroke-guided denoising diffusion probabilistic model for realistic image generation, 2024.

[68] Runhui Huang, Kaixin Cai, Jianhua Han, Xiaodan Liang, Renjing Pei, Guansong Lu, Songcen Xu, Wei Zhang, and Hang Xu. Layerdiff: Exploring text-guided multi-layered composable image synthesis via layer-collaborative diffusion model, 2024.

[69] Yingying Deng, Xiangyu He, Fan Tang, and Weiming Dong. z^*: Zero-shot style transfer via attention rearrangement, 2023.

[70] Tsu-Jui Fu, Xin Eric Wang, and William Yang Wang. Language-driven artistic style transfer. In *European Conference on Computer Vision*, pages 717–734. Springer, 2022.

[71] Yingying Deng, Fan Tang, Weiming Dong, Wen Sun, Feiyue Huang, and Changsheng Xu. Arbitrary style transfer via multi-adaptation network. In *Proceedings of the 28th ACM international conference on multimedia*, pages 2719–2727, 2020.

[72] Cheng-Che Lee, Wan-Yi Lin, Yen-Ting Shih, Pei-Yi Patricia Kuo, and Li Su. Crossing you in style: Cross-modal style transfer from music to visual arts, 2020.

[73] Zhi-Song Liu, Li-Wen Wang, Jun Xiao, and Vicky Kalogeiton. Bridging text and image for artist style transfer via contrastive learning, 2024.

[74] Chenggui Sun and Li Bin Song. Image style transfer: from artistic to photorealistic, 2022.

[75] Michael Cahyadi, Muhammad Rafi, William Shan, Jurike Moniaga, and Henry Lucky. Accuracy and fidelity comparison of luna and dall-e 2 diffusion-based image generation systems, 2023.

[76] Zhizhong Wang, Zhanjie Zhang, Lei Zhao, Zhiwen Zuo, Ailin Li, Wei Xing, and Dongming Lu. Aesust: towards aesthetic-enhanced universal style transfer. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1095–1106, 2022.

[77] Xiaosheng He, Fan Yang, Fayao Liu, and Guosheng Lin. Few-shot image generation via style adaptation and content preservation, 2023.

[78] Junjie Kang, Jinsong Wu, and Shiqi Jiang. Region-controlled style transfer, 2023.

[79] Jie An, Siyu Huang, Yibing Song, Dejing Dou, Wei Liu, and Jiebo Luo. Artflow: Unbiased image style transfer via reversible neural flows, 2021.

[80] Hao Phung, Quan Dao, and Anh Tran. Wavelet diffusion models are fast and scalable image generators, 2023.

[81] Minseong Kim, Jongju Shin, Myung-Cheol Roh, and Hyun-Chul Choi. Uncorrelated feature encoding for faster image style transfer, 2018.

[82] Dongdong Chen, Jing Liao, Lu Yuan, Nenghai Yu, and Gang Hua. Coherent online video style transfer. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1105–1114, 2017.

[83] Zhizhong Wang, Lei Zhao, Haibo Chen, Lihong Qiu, Qihang Mo, Sihuan Lin, Wei Xing, and Dongming Lu. Diversified arbitrary style transfer via deep feature perturbation, 2020.

[84] Nithin Gopalakrishnan Nair, Wele Gedara Chaminda Bandara, and Vishal M Patel. Image generation with multimodal priors using denoising diffusion probabilistic models, 2022.

[85] Wei Sun, Jawadul H. Bappy, Shanglin Yang, Yi Xu, Tianfu Wu, and Hui Zhou. Pose guided fashion image synthesis using deep generative model, 2019.

[86] Jie An, Haoyi Xiong, Jinwen Ma, Jiebo Luo, and Jun Huan. Stylenas: An empirical study of neural architecture search to uncover surprisingly fast end-to-end universal style transfer networks, 2019.

[87] Ziying Pan, Kun Wang, Gang Li, Feihong He, and Yongxuan Lai. Finediffusion: Scaling up diffusion models for fine-grained image generation with 10,000 classes, 2024.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.