Artificial Intelligence Review (2025) 58:99
https://doi.org/10.1007/s10462-025-11110-3

Comprehensive exploration of diffusion models in image generation: a survey

Hang Chen[1] · Qian Xiang[2,3,4] · Jiaxin Hu[1] · Meilin Ye[1] · Chao Yu[1] · Hao Cheng[1] · Lei Zhang[1]

Accepted: 7 January 2025 / Published online: 25 January 2025

Abstract

The rapid development of deep learning technology has led to the emergence of diffusion models as a promising generative model with diverse applications. These include image generation, audio and video synthesis, molecular design, and text generation. The distinctive generation mechanism and exceptional generation quality of diffusion models have made them a valuable tool in these diverse fields. However, with the extensive deployment of diffusion models in the domain of image generation, concerns pertaining to data privacy, data security, and artistic ethics have emerged with increasing prominence. Given the accelerated pace of development in the field of diffusion models, the majority of extant surveys are deficient in two respects: firstly, they fail to encompass the latest advances in diffusion-based image synthesis; and secondly, they seldom consider the potential social implications of diffusion models. In order to address these issues, this paper presents a comprehensive survey of the most recent applications of diffusion models in the field of image generation. Furthermore, it provides an in-depth analysis of the potential social impacts that may result from their use. Firstly, this paper presents a systematic survey of the background principles and theoretical foundations of diffusion models. Subsequently, this paper provides a detailed examination of the most recent applications of diffusion models across a range of image generation subfields, including style transfer, image completion, image editing, super-resolution, and beyond. Finally, we present a comprehensive examination of these social issues, addressing data privacy concerns, such as the potential for data leakage and the implementation of protective measures during model training. We also analyse the risk of malicious exploitation of the model and the defensive strategies employed to mitigate such risks. Additionally, we examine the implications of the authenticity and originality of generated images on artistic creativity and copyright protection.

Keywords Image generation · Diffusion models · Generative models · Data privacy · Data security

Extended author information available on the last page of the article

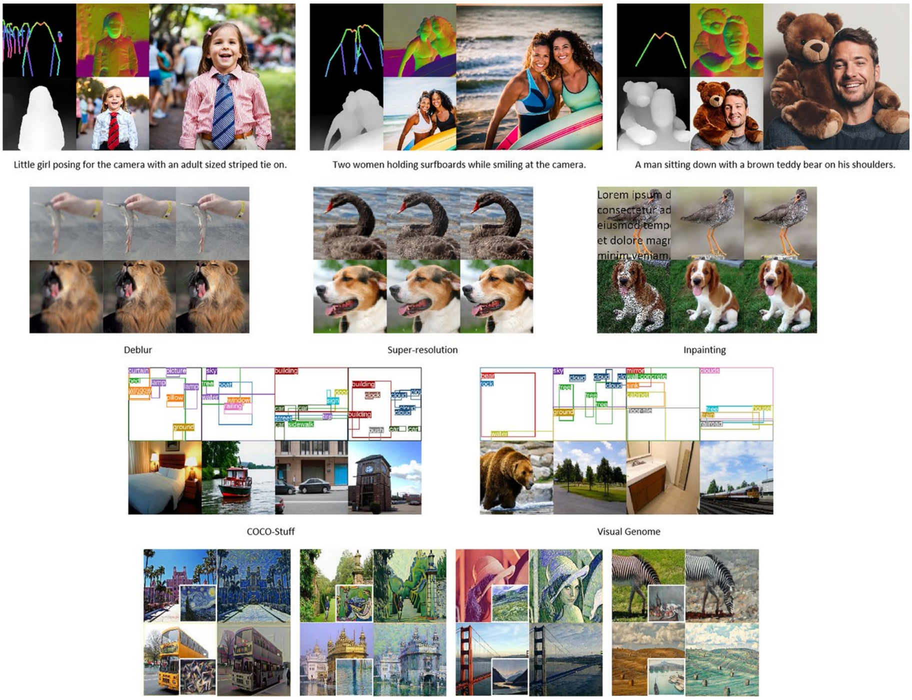

Little girl posing for the camera with an adult sized striped tie on.
Two women holding surfboards while smiling at the camera.
A man sitting down with a brown teddy bear on his shoulders.
Deblur
Super-resolution
Inpainting
COCO-Stuff
Visual Genome

Fig. 1 Images generated by diffusion models in different visual tasks. The first row of images shows a combined coarse RGB, depth, normal, and high-resolution images generated conditioned on text and skeleton (Liu et al. 2024). The second row of images shows the generation results of several image restoration subtasks (Fei et al. 2023). The third row of images shows examples of the 256 × 256 images generated by layout-to-image methods on COCO-Stuff and Visual Genome (Yang et al. 2023). The last row of the images shows the style transfer results generated by the StyleDiffusion (Wang et al. 2023)

1 Introduction

Image generation task is one of the most important research topics in the field of artificial intelligence, which aims to generate realistic image content through algorithms and models. As shown in Fig. 1, this task ranges from simple image restoration (Bao et al. 2024; Fei et al. 2023; et al. 2023) and style transfer (Huang et al. 2024; Li et al. 2024; Wang et al. 2023) to complex scene generation (Liu and Liu 2024; Scene generation with hierarchical latent diffusion models 2023; Yang et al. 2023) and human generation (Ju et al. 2023; Wang et al. 2024; Liu et al. 2024). The advent of sophisticated technologies, including Variational Autoencoders (VAE) (Kingma et al. 2014), Normalizing Flows (Papamakarios et al. 2021), Generative Adversarial Networks (GANs) (Goceri 2024), and wavelet-based augmentation

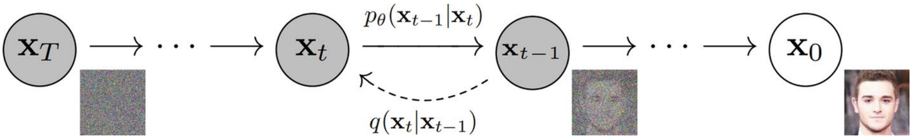

$\mathbf{x}_T$
$\mathbf{x}_t$
$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$
$q(\mathbf{x}_t|\mathbf{x}_{t-1})$
$\mathbf{x}_{t-1}$
$\mathbf{x}_0$

Fig. 2 The directed graphical model of DDPM. Ho et al. (2020)

Table 1 Summary of the application of diffusion models in the field of image generation

Models (Papers)	Coference and year	Methods and Insights	Target tasks	Datasets
Stylediffusion Wang et al. (2023)	ICCV 2023	Framework for style transfer	Style Transfer	COCO-Stuff, Visual Genome
ZeCon Yang et al. (2023)	ICCV 2023	Zero-shot contrastive loss for text-guided diffusion image style transfer	Style Transfer	MS-COCO, WikiArt
Z-STAR Deng et al. (2023)	Arxiv 2023	Zero-shot style transfer via attention rearrangement	Style Transfer	
Style Injection Chung et al. (2023)	Arxiv 2023	Style transfer using large-scale pre-trained diffusion models	Style Transfer	LAION
InST Zhang et al. (2023)	CVPR 2023	Style transfer method based on inversion	Style Transfer	
Stable Artist Brack et al. (2022)	Arxiv 2022	Iterative approach for guiding generated images to desired output	Style Transfer	Laion-5b
Diffstyler Huang et al. (2022)	Arxiv 2022	Text-driven image stylization framework based on dual diffusion	Style Transfer	
Style Guidance Pan (2023)	WACV 2023	Style guidance approach for text-to-image diffusion models	Style Transfer	
Hicast Wang et al. (2024)	Arxiv 2024	Highly Controllable Arbitrary Style Transfer	Style Transfer	MS-COCO, WikiArt
Artfusion Chen et al. (2023)	Arxiv 2023	Controllable arbitrary style transfer using dual conditional latent diffusion models	Style Transfer	MS-COCO, WikiArt
Diffbir Lin et al. (2023)	Arxiv 2023	Applies pre-trained text-to-image diffusion model to blind image restoration	Image Restoration	
Diffbfr Qiu et al. (2023)	ACM-ICM 2023	Bootstrap diffusion model for blind face recovery	Image Restoration	MINST, BFR, FFFQ, CelebA-Test
Repaint Lugmayr et al. (2022)	CVPR 2022	Redrawing method based on DDPM that does not require specific mask training	Image Restoration	CelebA-HQ, Imagenet
Multiscale Structure Guide Ren et al. (2023)	ICCV 2023	Introduces multiscale structure guide to inform icDPM about the coarse structure of sharp images	Image Restoration	Realblur-J, REDS, HIDE

Table 1 (continued)

Models (Papers)	Coference and year	Methods and Insights	Target tasks	Datasets
C2F-DFT Wang et al. (2023)	Arxiv 2023	Coarse-to-Fine Diffusion Transformer for image recovery	Image Restoration	Rain13K, Rain100H, Rain100L, Test100, Test2800, GoPro, RealBlur-R, RealBlur-J, RealBlur,SIDD
HI-Diff Chen et al. (2024)	NeurIPS 2024	Hierarchical integrated diffusion model for real-world image deblurring	Image Restoration	GoPro, HIDE, RealBlur, RWBI
RDDM Liu et al. (2023)	Arxiv 2023	Residual denoising diffusion model for image recovery	Image Restoration	ISTD, LOL, RainDrop, CelebA
Cross Initialization Pang et al. (2023)	Arxiv 2023	New initialization method for improving image reconstruction quality and editability	Image Editing	
Custom-Edit Choi et al. (2023)	Arxiv 2023	Text-guided editing method with customized diffusion models	Image Editing	LAION, Imagen, eDiffI, Imagic, Muse
Dreamedit Li et al. (2023)	Arxiv 2023	Iterative generation method for high-quality theme replacement and addition	Image Editing	DreamEditBench
ProxNPI Han et al. (2024)	WACA 2024	Proximal Negative-Prompt Inversion for tuning-free real image editing	Image Editing	
Mdp Wang et al. (2023)	Arxiv 2023	General framework for explaining various operations suitable for editing in diffusion models	Image Editing	
Kv inversion Huang et al. (2023)	PRCV 2023	High-quality image editing without fine-tuning	Image Editing	Tuning-free
Learnable Regions Lin et al. (2023)	Arxiv 2023	Image editing method based on learnable regions	Image Editing	
FEC-kv-reuse Chen et al. (2023)	ICICML 2023	Three sampling methods for different editing types and settings	Image Editing	
Visual Instruction Inversion Nguyen et al. (2023)	Arxiv 2023	Guides text-to-image diffusion model through visual prompts	Image Editing	Clean-InstructPix2Pix
Photoswap Gu et al. (2024)	ANIP 2024	Personalized subject swapping in images	Image Editing	
SR3 Saharia et al. (2022)	PAMI 2022	Image super-resolution using denoising diffusion probability model and U-Net	Image Super-Resolution	FFHQ, CelebA-HQ

Table 1 (continued)

Models (Papers)	Coference and year	Methods and Insights	Target tasks	Datasets
Srdiff Li et al. (2022)	Neurocomputing 2022	Single image super-resolution method based on diffusion models	Image Super-Resolution	CelebA, DIV2K
Partdiff Zhao et al. (2023)	Arxiv 2023	Image super-resolution method based on diffusion probability models	Image Super-Resolution	FFHQ, ImageNet, Prostate MRI
Dream Zhou et al. (2023)	Arxiv 2023	Diffusion Rectification and Estimation-Adaptive Models for super-resolution	Image Super-Resolution	CeleA-HQ, DIV2K, CAT, LSUN
Sinsr Wang et al. (2024)	CVPR 2024	Single-step SR generation method	Image Super-Resolution	
GLIDE Nichol et al. (2022)	ICML 2022	Guided diffusion for textual conditional image synthesis	Text-to-Image Generation	MS-COCO
Imagen Saharia et al. (2022)	NeurIPS 2022	Text-to-image generation using large language models as text encoders	Text-to-Image Generation	LAION-400 M, FIT400M
DALL-E2 Ramesh et al. (2022)	Arxiv 2022	Two-stage model for text-to-image generation	Text-to-Image Generation	AVA, CLIP, DALL-E
Low-Light Image Shang et al. (2024)	CAI 2024	Multi-domain multi-scale diffusion model for low-light image enhancement	Low-Light Image Enhancement	LOL, LOLv2
Diff-Unmix Zeng et al. (2024)	CVPR 2024	Self-supervised denoising using denoising diffusion model for hyperspectral image denoising	Image Denoising	HIS, KAIST, CAVE, CAVE-Toy
LRed Zhao et al. (2024)	CVPR 2024	Latent Background Knowledge Retrieval-Augmented Diffusion for camouflage image generation	Camouflaged Vision Perception	COD10K, COCO-Stuff, CAMO, NC4K, DUTS, DUT-OM-RON, COCO2017
Medm2g Chenlu et al. (2024)	CVPR 2024	Unified model for multimodal medical image generation	Medical Generative Modeling	MIMIC-CXR, MedI-Cat, MRI, CT scan
Monocular Depth Estimation Ke et al. (2024)	CVPR 2024	Affine-invariant monocular depth estimation using stable diffusion	Monocular Depth Estimation	Hypersim, Virtual KITTI, NYUv2, ScanNet, ETH3D, DIODE
Soda Hudson et al. (2024)	CVPR 2024	Bottleneck diffusion models for representation learning	Representation Learning	ImageNet, FID, FID, LPIPS, GSO
Texturedreamer Yeh et al. (2024)	CVPR 2024	Image-guided texture synthesis using diffusion models	Texture Synthesis	

Table 1 (continued)

Models (Papers)	Coference and year	Methods and Insights	Target tasks	Datasets
Promark Asnani et al. (2024)	CVPR 2024	Proactive diffusion watermarking for causal attribution	Causal Attribution	Adobe Stock, ImageNet, LSUN, Wikiar, BAM
Diff-Mix Wang et al. (2024)	CVPR 2024	Inter-class data augmentation method for image classification	Image Classification	Diff-Mix
Diffmorpher Zhang et al. (2024)	CVPR 2024	Method for smoothing and interpolating images using prior knowledge of pre-trained diffusion models	Image Morphing	
StrDiffusion Liu et al. (2024)	CVPR 2024	Image inpainting diffusion model that reconstructs texture denoising under structural guidance	Image Inpainting	PSV, CelebA, Places2
3D Point Cloud Nunes et al. (2024)	CVPR 2024	Diffusion model operating directly on points for 3D point cloud generation	Scene Completion	SemanticKITTI, KITTI-360
Intrinsic Image Decomposition Kocsis et al. (2024)	CVPR 2024	Sampling from the solution space of the conditional generation model for appearance decomposition	Intrinsic Image Decomposition	InteriorVerse
Id-blau et al. (2024)	CVPR 2024	ReBLurring AUgmentation for diversified generation of blurred images	Image Deblurring	GoPro, HIDE, RealBlur

methods (Goceri 2023), has enabled the creation of unprecedented and imaginative visual effects in image generation.

Diffusion models, originally inspired by the process of diffusion in physics, have been adapted to describe the stochastic process of gradually transforming simple noise distributions into complex data distributions, such as images (Neal 2001; Jarzynski 1997). Diffusion models offer a distinctive approach to image generation, whereby the data distribution is conceptualised as a sequence of conditional distributions (Ho et al. 2020; Song and Ermon 2020; Song et al. 2021). In contrast to traditional generative models, such as GANs and VAEs, which sample directly from a learned latent space, diffusion models capture intricate data dependencies through the diffusion process (Yang et al. 2024; Luo 2022, 2023). Despite their considerable successes, diffusion models remain constrained by certain limitations, including prolonged training periods, elevated computational requirements, and difficulties in scaling to high-resolution imagery (Cao et al. 2024; Yang et al. 2024; Chang et al. 2023).

Diffusion models have recently become a prominent area of interest within the field of artificial intelligence, due to their remarkable ability to generate images of exceptional quality, which often rival those created by humans (Fan et al. 2023; Zhang et al. 2023; et al. 2024; Yang et al. 2024; Croitoru et al. 2023a; Lin et al. 2024). Diffusion models have also been applied to various areas of image generation, including Style Transfer (Zhang et al. 2023; Wang et al. 2023, 2024), Image Restoration (Luo et al. 2023; Qiu et al. 2023; Ren et al. 2023), Image Editing (Pang et al. 2023; Wang et al. 2023; Gu et al. 2024), Super-Resolution (Zhao et al. 2023; Wang et al. 2024; Gandikota et al. 2024), Text-to-image Generation (Nichol et al. 2022; Saharia et al. 2022; Structured prediction for efficient text-to-image generation 2024), and other tasks (Zeng et al. 2024; Zhao et al. 2024; Hudson et al. 2024). The application of diffusion modelling in the field of image generation has profoundly impacted many aspects of society, driving content innovation in industries such as advertising, gaming, film and television, reducing creation costs and increasing productivity by generating high-quality, photorealistic images. However, it also raises issues such as copyright and originality (Wang et al. 2024; Dubinski et al. 2024; Gu et al. 2024; Zhang et al. 2023), and poses challenges to traditional artistic creation and market patterns. Meanwhile, at the social and ethical level, images generated by diffusion models may be used for improper purposes such as misleading and false propaganda, posing a potential threat to public perception and information authenticity (Hu et al. 2023; Zhu et al. 2023; Linet al. 2023; Carlini et al. 2023; Ni et al. 2023; Seunghoo and Juhun 2024; Qu et al. 2023; Brack et al. 2023).

In light of the above, this paper focuses on recent applications of diffusion models in the field of image generation and the socio-ethical implications of diffusion model-based image generation. The following key aspects are addressed in this survey: firstly, in-depth research has been conducted on the historical background and theoretical basis of diffusion models (Ho et al. 2020; Alexander et al. 2021; Song and Ermon 2019), thereby providing a solid foundation for subsequent discussions. Secondly, the practical applications of diffusion models in image generation tasks, including image inpainting (Zhang et al. 2023; Anciukevicius et al. 2023; Liu et al. 2024; Lee et al. 2024), style transfer (Huang et al. 2024; Li et al. 2024; Wang et al. 2023) and super-resolution (Luo et al. 2024; Ma et al. 2024; Gao et al. 2023; Metzger et al. 2023), will be investigated. Thirdly, this study examines the challenges and potential solutions to the social problems encountered by diffusion models when utilised for image generation.

The aim of this work is to synthesise the extensive body of knowledge scattered across a wide range of publications, distil the key findings and present them in a coherent manner that will facilitate future research efforts. By providing a comprehensive overview of diffusion models in image generation, this survey aims to serve as a valuable resource for both novice researchers and experienced professionals wishing to deepen their knowledge or explore new avenues of research in this dynamic and evolving field.

The paper is divided into seven sections to provide a structured exploration of diffusion models in image generation. The introduction (Sect. 1) outlines the aims and significance of the study. Sect. 2, Related Works, places our research in the context of previous studies. Sect. 3, Background on diffusion models, explains the theoretical basis of diffusion models. Sect. 4, Diffusion Models in Image Generation, describes recent applications of these models. Ethical and social implications (Sect. 5) discusses the potential impact on society. Sect. 6, challenges and future directions, identifies current obstacles and suggests avenues for future research. The conclusion (Sect. 7) summarises the main findings and contributions of the paper.

2 Related works

In recent years, diffusion models have emerged as a promising approach to image generation. These models, inspired by the principles of non-equilibrium thermodynamics, employ an iterative process to refine the noise, ultimately resulting in the generation of coherent images. Notably, studies such as those conducted by Sohl-Dickstein et al. (2015) and Song and Ermon (2019) have illustrated the efficacy of diffusion models in producing high-quality images. The advent of diffusion models has significantly accelerated the advancement of the field of image generation. A considerable number of researchers have made significant contributions to the advancement of knowledge in the various subfields of image generation. Additionally, comprehensive survey papers on the application of diffusion generation models have also emerged, providing valuable synthesis and summarisation for the continuous progress in this field (Po et al. 2024; Cao et al. 2024; Zhang et al. 2023; Fan et al. 2024; et al. 2024; Yang et al. 2024; Li et al. 2023; Croitoru et al. 2023b; Moser et al. 2024; Ulhaq et al. 2022; Fan et al. 2024).

The objective of text-to-image (T2I) generation is to transform natural language text descriptions into corresponding visual images. This task requires the model to have both language understanding and visual representation in order to generate images that match the textual description. The previous survey literature (Cao et al. 2024; Zhang et al. 2023) provides a comprehensive review of T2I using the diffusion model, covering both the theoretical foundations and practical progress in the field. In addition to natural images, magnetic resonance imaging (MRI), as an important medical imaging modality, also offers unique application opportunities for diffusion models (Fan et al. 2024). Other survey literature (Cao et al. 2024; Yang et al. 2024) aims to provide a comprehensive and in-depth understanding of diffusion models, from basic formulae and algorithmic improvements to a variety of applications, revealing their development history and future trends.

In the field of computer vision, there is a substantial body of survey literature (Ulhaq et al. 2022; Croitoru et al. 2023b) that provides a comprehensive review of denoising diffusion models. This includes theoretical and practical contributions to the field. Additionally, there is a significant corpus of survey literature that addresses more specific visual subfields, such as image restoration and enhancement (Li et al. 2023), video generation (et al. 2024), and image super-resolution (Moser et al. 2024).

The field of diffusion models is undergoing a period of rapid development, with a considerable amount of new research emerging in the area of image generation. It is therefore of great importance to conduct a comprehensive literature survey on the application of the latest diffusion model in image generation. However, the majority of existing surveys are flawed in two ways. Firstly, due to time constraints, they fail to cover the latest advances in diffusion-based image generation (Ulhaq et al. 2022; Li et al. 2023). Secondly, they rarely consider the potential social impact of diffusion models in the field of image generation (Croitoru et al. 2023b; Moser et al. 2024). Therefore, the aim of this study is to provide insight into the development of diffusion models in the field of image generation by conducting a large number of surveys in the field of image generation and related social ethics literature.

In order to guarantee that the research is both pioneering and pertinent, it is essential to document the most recent developments and discourses within the field. Consequently, our work has concentrated primarily on diffusion model methodologies for image generation over the past three years. In addition, we have investigated the ethical implications and potential mitigating strategies associated with these approaches. The selected papers not only demonstrate significant technological advances, but also address the broader impact of these technologies on society, thereby ensuring a comprehensive analysis of the technical and ethical dimensions of diffusion models in image generation.

3 Background on diffusion models

The core idea of the diffusion model comes from sequential Monte Carlo (Neal 2001) and non-equilibrium statistical physics (Jarzynski 1997), which uses a Markov chain to transform one distribution into another. The diffusion model has two main components: the forward diffusion process and the backward denoising process. The forward process gradually adds noise to the original image and eventually converts the image into a pure noise image that conforms to a Gaussian distribution. The corresponding backward process is exactly the opposite, converting a pure noise image into a realistic image that conforms to the original distribution through several steps of denoising operations. To further demonstrate the intuition of diffusion models, we will discuss the three main formulaic representations of diffusion models currently being studied: denoised diffusion probability models (Ho et al. 2020; Alexander et al. 2021), score-based generation models (Song and Ermon 2019, 2020), and stochastic differential equations (Song et al. 2020, 2021). The following three subsections will independently elaborate on each formulaic representation, while discussing their connections and differences.

3.1 Denoising diffusion probilistic models (DDPMs)

Suppose we sample the initial data $x_0 \sim q(x)$ from a real data distribution $q(x)$. By using the forward diffusion process to gradually add noise to the initial data x_0, a series of noisy data

$x_1, ..., x_T$ are obtained. According to the properties of Markov processes and the chain rule of probability, we can represent the joint distribution of all data as $q(x_1, ..., x_T|x_0)$,

$$q(x_1, ..., x_T|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1}). \tag{1}$$

The transition kernel $q(x_t|x_{t-1})$ is designed manually in DDPM (Ho et al. 2020), and noise is gradually added to the initial data at each step of the transition. Additionally, hyperparameters β can be set to establish a schedule for noise introduction as a variance control term.

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I), \tag{2}$$

where $\beta_t \in (0, 1)$.
By leveraging the properties of the Gaussian distribution and Eq.(2), we can streamline the calculation of the forward diffusion process, thereby directly obtaining the analytical form of $q(x_t|x_0)$,

$$\begin{aligned} \alpha_t = 1 - \beta_t, \bar{\alpha}_t = \prod_{i=1}^{t} \alpha_t \\ q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I) \end{aligned}. \tag{3}$$

As *t* approaches infinity and $\bar{\alpha}_t$ approaches 0, there is $x_T \sim \mathcal{N}(0, 1)$. By employing initial data x_0 and the sampled Gaussian vector ϵ, it is a straightforward process to utilise Eq.(3) to calculate sample x_t,

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + (1-\bar{\alpha}_t)\epsilon. \tag{4}$$

A new sample can be generated from the initial distribution by starting from a sample $x_T \sim \mathcal{N}(0, I)$ and employing the reverse denoising process. A learnable transfer kernel based on x_0 for a reverse step can be expressed in the following form:

$$\begin{aligned} q(x_{t-1}|x_t, x_0) = \frac{q(x_t|x_{t-1}, x_0)q(x_{t-1}|x_0))}{q(x_t|x_0))} \\ = \frac{\mathcal{N}(x_t; \sqrt{\alpha}_t x_{t-1}, (1-\alpha_t)I)\mathcal{N}(x_{t-1}; \sqrt{\bar{\alpha}}_{t-1}x_0, (1-\bar{\alpha}_{t-1})I)}{\mathcal{N}(x_t; \sqrt{\bar{\alpha}}_t x_0, (1-\bar{\alpha}_t)I)} \\ = \mathcal{N}(x_{t-1}; \mu_q(x_t, x_0), \Sigma_q(t)I), \end{aligned} \tag{5}$$

It can be observed that the x_{t-1} obtained from each reverse step follows a normal distribution, with the mean $\mu_q(x_t, x_0)$ being a function of x_t and x_0, and the variance $\Sigma_q(t)$ being a function of the coefficient α. The coefficient α is fixed and known at each step. Therefore, the key to learning the transfer kernel is to fit the distribution, which is to predict the mean $\mu_q(x_t, x_0)$ and variance $\Sigma_q(t)$,

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_q(t)I), \tag{6}$$

where θ is a learnable neural network parameter.

During the training process, we need to make the approximate denoising transformation $p_\theta(x_{t-1}|x_t)$ as closely as possible to the actual denoising transformation $q(x_{t-1}|x_t, x_0)$. According to Eq.(5) and Eq.(6), it is known that their variance terms are fixed and can be accurately matched. Therefore, when using KL-divergence to calculate the difference between two distributions, we only need to consider their means. Due to the difficulty in calculating $p_\theta(x_0)$, it can be processed by minimizing the variational lower bound of negative log-likelihood. For the matching of the entire trajectory, the objective function can be formulated as follows:

$$L_{vlb} = L_0 + \sum_{1<t<T} L_t + L_T, \tag{7}$$

where $L_0 = -\log p_\theta(x_0|x_1)$, $L_T = KL(q(x_T|x_0)||\pi(x_T))$, and $L_t = KL(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t))$. *KL*() represents the KL-divergence of two distributions. Furthermore, since $q(x_{t-1}|x_t, x_0)$ follows a normal distribution, we can calculate the closed form of KL-divergence.

As proposed in DDPM (Ho et al. 2020), the optimization of the function L_{vlb} can be achieved by minimizing the deviation between the model's true noise and approximate noise at random time steps in the trajectory:

$$L_s = \mathbb{E}_{t\sim[1,T], x_0\sim q(x_0), \epsilon_t\sim\mathcal{N}(0,I)}[||\epsilon_t - \epsilon_\theta(x_t, t)||^2]. \tag{8}$$

where $\mathbb{E}$ is the expected value calculation, $\epsilon_\theta(x_t, t)$ is predict noise in step *t* by neural network, x_0 is the initial data.

3.2 Score-based generative models (SGMs)

The objective of an explicit generative model is to model the probability distribution of data, and then use this distribution to generate samples through sampling. Score-Based Models (Song and Ermon 2019) do not directly learn the probability distribution *p*(*x*) of the data, but rather learn score function, which is the logarithmic gradient of the probability distribution. It is defined as follows:

$$\nabla_x \log p(x) = \frac{\partial \log p(x)}{\partial x}. \tag{9}$$

The score function learned by Score-Based Models can be denoted as $s_\theta(x)$, and by learning to approximate the logarithmic gradient of the probability distribution, $s_\theta(x) \approx \nabla_x \log p(x)$. Furthermore, $s_\theta(x)$ can be parameterized using an energy-based model,

$$\begin{aligned} p_\theta(x) &= \frac{e^{-f_\theta(x)}}{Z_\theta} \\ s_\theta(x) = \nabla_x \log p_\theta(x) &= -\nabla_x f_\theta(x) - \nabla_x \log Z_\theta, \end{aligned} \tag{10}$$

where $Z_\theta = \int e^{-f_\theta(x)} \mathrm{d}x$ is normalizing constant. This also ensures that $p_\theta(x)$ is a density function. Since Z_θ being a constant, thus $\nabla_x \log Z_\theta$ equals zero. Therefore, Score-Based Models are not related to normalization terms, which greatly expands the range of models or architectures that can be used.
Similar to other explicit generative models, Score-Based Models can calculate the Fisher divergence between the predicted distribution of the model and the ground-truth data distribution to measure the difference between the two distributions. Then, the model is trained by minimizing Fisher divergence. Therefore, the objective function can be defined as:

$$\mathbb{E}_{p(x)}[||\nabla_x \log p(x) - s_\theta(x)||_2^2] \tag{11}$$

However, it is difficult to calculate Fisher's divergence directly because the score of real data cannot be calculated. Fortunately, we can use the score matching method (Hyvarinen and Dayan 2005; Vincent 2011; Song et al. 2020) to bypass the calculation of $p(x)$ and minimize divergence. The optimization objective function can be rewritten as

$$\mathbb{E}_{p(x)}[2tr(\nabla_x s_\theta(x)) + ||s_\theta(x)||_2^2] \tag{12}$$

Train the model using the above equation to make $s_\theta(x) \approx \nabla_x \log p(x)$, thus obtaining the trained Score-Based Models. Then, we can use Langevin dynamics (Parisi 1981; Ulf and Miller 1994) to iteratively generate samples from the predicted score model. The process of sampling through Langevin dynamics can be expressed as the following iterations:

$$x_{i+1} \leftarrow x_i + \eta \nabla_x \log p(x) + \sqrt{2\eta} z_i, i = 0, 1, ..., K, \tag{13}$$

where $z_i \sim \mathcal{N}(0, I)$, η is a sufficiently small quantity. As K approaches infinity, x_K converges to a sample that follows the data distribution $p(x)$. Furthermore, it can be seen from the above equation that the unknown variable $\nabla_x \log p(x)$ during the iterative calculation process has also been approximated through the prediction of the trained model $s_\theta(x)$.
We have discussed the training and sampling of Score-Based Models. However, the performance in practical applications is not ideal, mainly due to the limited amount of data used for score matching in low-density areas, which leads to inaccurate estimation of the learned model in low-density areas, thereby limiting the model's ability of the model to generate high-quality samples. Noise Conditional Score-Based Models (NCSMs) (Song and Ermon 2020; Song et al. 2020) propose the use of multi-scale noise to perturb data points and fill low-density areas, thereby improving the accuracy of estimated scores. The distribution after noise disturbance can be expressed as

$$p_{\sigma_i}(x) = \int p(y) \mathcal{N}(x; y, \sigma_i^2 I) \mathrm{d}y, \tag{14}$$

where $\sigma_1 < \sigma_2 < ... < \sigma_L$ are a set of increasing standard deviations. As a result, the goal of Score-Based Models also becomes to compute the score function $s_\theta(x, i)$ of each noise disturbance distribution. Therefore, we can use the score matching to train the parameterized NCSN. The overall optimization objective function is a weighted Fisher divergence:

$$\sum_{i=1}^{L} \lambda(i)\mathbb{E}_{p_{\sigma_i}(x)}[||\nabla_x \log p_{\sigma_i}(x) - s_\theta(x,i)||_2^2], \tag{15}$$

where $\lambda(i) = \sigma_i^2$ is weighting term. The training process is consistent with $s_\theta(x)$ without disturbance. After training the model $s_\theta(x,i)$, use the annealed Langevin dynamics (Song and Ermon 2019, 2020; Jolicoeur-Martineau et al. 2020) to iteratively generate samples, where $i = L, L-1, ..., 1$.

3.3 Stochastic differential equations (SDEs)

As the noise scale and time step approach infinity, the noise disturbance process of Score-Based Models and the noise addition process of diffusion models (i.e. the forward diffusion process) can be summarized as a continuous time stochastic process. Many stochastic processes can be represented as solutions of stochastic differential equations (SDEs), and a stochastic process represented by SDEs can be expressed in the following form:

$$\mathrm{d}x = f(x,t)\mathrm{d}t + g(t)\mathrm{d}w, \tag{16}$$

where $f(\cdot, t) : \mathcal{R}^d \to \mathcal{R}^d$ and $g(t) \in \mathcal{R}$ are vector-valued function and real-valued function respectively. $\mathrm{d}w$ is an infinitesimal Gaussian white noise, and w is a standard Brownian motion. It should be noted that Anderson (Anderson 1982) proposes that there exists a corresponding reverse SDEs for any SDEs, and it has the following closed form:

$$\mathrm{d}x = [f(x,t) - g^2(t)\nabla_x \log p_t(x)]\mathrm{d}t + g(t)\mathrm{d}w, \tag{17}$$

where $\mathrm{d}t$ is an negative infinitesimal quantity of time. To compute the inverse SDEs, we need to estimate the score function of $p_t(x)$.
To train $s_\theta(x,i)$, the Fisher divergence under continuous time can be calculated as follows:

$$\mathbb{E}_{t\in\mathcal{U}(0,T)}\mathbb{E}_{p_t(x)}[\lambda(t)||\nabla_x \log p_t(x) - s_\theta(x,t)||_2^2] \tag{18}$$

where $\mathcal{U}(0,T)$ represents a uniform distribution over time [0, *T*]. λ is consistent with Score-Based Models and is a weight function.
Specifically, DDPMs and SGMs correspond to two special SDEs, and their ways of adding noise in the forward process are different. As proposed by Song et al. (2020) (Song et al. 2020), the DDPM corresponds to the Variance Preserving SDE (VP SDE) and has the following form:

$$\mathrm{d}x = -\frac{1}{2}\beta(t)x\mathrm{d}t + \sqrt{\beta(t)}\mathrm{d}w, \tag{19}$$

where $\beta(t)$ is a predefined schedule function. SGMs correspond to Variance Exploring SDE (VE SDE) and have the following form:

$$\mathrm{d}x = \sqrt{\frac{\mathrm{d}[\sigma^2(t)]}{\mathrm{d}t}}\mathrm{d}w, \tag{20}$$

where $\sigma(t)$ represents the disturbance noise corresponding to time T approaching infinity. Both SGMs and DDPMs can be regarded as discretization of stochastic differential equations determined by fractional functions. Therefore, the score based generative model and the diffusion probability model can be summarized into a unified framework of SDEs.

3.4 Controllable generation for diffusion models

The main purpose of the generative models we discussed earlier is to fit the data distribution $p(x)$, to generate samples with the same distribution as the initial data. These models can be summarized as unconditional generative models. However, in practical applications, we prefer to generate samples with certain characteristics according to our ideas, and these corresponding models are called conditional controlled diffusion models. Therefore, at this point, the fitting target of the model also correspondingly becomes the conditional data distribution $p(x|y)$. According to the Bayes theorem, we can express it as

$$p(x|y) = \frac{p(x)p(y|x)}{p(y)} = \frac{p(x)p(y|x)}{\int p(x)p(y|x)\mathrm{d}x}. \tag{21}$$

Similar to Score-Based Models, by taking the gradients of x on both sides of this expression, we can obtain the score function of the conditional data distribution

$$\nabla_x \log p(x|y) = \nabla_x \log p(x) + \nabla_x \log p(y|x). \tag{22}$$

From the above expression, it can be seen that the score function of the conditional data distribution consists of the score function of the initial data and the known forward process. Similarly, conditional controlled diffusion models can be divided into Classifier Guidance and Classifier Free based on whether existing unconditional generation models are used.
For Classifier Guidance (Dhariwal and Nichol 2021; Liu et al. 2023), the conditional control generation model does not require retraining the diffusion model, and simple control can be achieved at a low cost by training a classifier. Therefore, in a conditional sampling process, the probability of state transition can be rewritten as (Dhariwal and Nichol 2021):

$$p_{\theta,\phi}(x_t|x_{t+1}, y) = Zp_\theta(x_t|x_{t+1})p_\phi(y|x_t), \tag{23}$$

where Z denotes the normalization term, θ is the diffusion model parameter, and ϕ is the classifier parameter.
By taking the logarithm of both sides of the above equation and expanding it, we can obtain

$$\begin{aligned} \log p_{\theta,\phi}(x_t|x_{t+1}, y) &= \log p_\theta(x_t|x_{t+1}) + \log p_\phi(y|x_t) + C \\ &= -\frac{1}{2}(x_t - \mu - \Sigma g)^\top \Sigma^{-1}(x_t - \mu - \Sigma g) + \frac{1}{2}g^\top \Sigma g + C \end{aligned} \tag{24}$$

where $g = \nabla_{x_t} \log p_\phi(y|x_t)|_{x_t=\mu}$. It can be seen that compared to the unconditional transition distribution, the conditional transition distribution also follows a Gaussian distribution, and the variance Σ is the same as the unconditional transition distribution $p_\theta(x_t|x_{t+1})$. At the same time, the mean has a shift of Σg, which also includes gradient information from the classifier.

By using a hyperparameter s to control the degree of classifier guidance, the sampling algorithm with classifier guidance is represented as

$$x_{t-1} \sim \mathcal{N}(\mu + s\Sigma\nabla_{x_t} \log p_\phi(y|x_t), \Sigma), \tag{25}$$

where $\mu = \mu_\theta(x_t), \Sigma = \Sigma_\theta(x_t)$.

In the guided diffusion model, the diffusion model was not retrained, but the guidance information of the classifier was added during the sampling process. This type of method makes the model training process cumbersome and complex, and cannot fully exploit the performance of the diffusion model. Jonathan and Salimans (2022) has improved DDPM and proposed a classifier-free guidance technique, also known as Classifier Free Guidance (CFG). Its core idea is very simple: the conditional input is fed into the training process of the diffusion model and fitted directly to the model. There is no need for a classifier. This is fundamentally different from Classifier Guidance, which did not involve the training process of dynamic diffusion models, but only guided the sampling process.

In the training process of the CFG model, the conditional input y is introduced, and the input parameters for the noise prediction are x_t, t, and y. Therefore, the loss function becomes as follows:

$$L_s(\theta) = \mathbb{E}_{t\sim[1,T],x_0\sim q(x_0),\epsilon\sim\mathcal{N}(0,I)}[||\epsilon - \epsilon_\theta(x_t, t, y)||^2]. \tag{26}$$

3.5 Improvements of diffusion models

Compared to VAEs (Higgins et al. 2017; Kingma et al. 2014; van et al. 2017) and GANs (Goodfellow et al. 2014; Arjovsky et al. 2017; Reed et al. 2016; Zhang et al. 2017), diffusion models theoretically require more time in the sampling process and may require thousands of evaluation steps to extract a single sample. This is because when using SDE or Markov processes to iteratively transform prior distributions into complex data distributions, a large number of function evaluations are involved in the reverse process. In addition, diffusion models also face the instability of the reverse process, as well as the computational requirements and constraint challenges required for training models in high-dimensional Euclidean space. Maximum likelihood estimation is not comparable to likelihood-based models, which is also a challenge for diffusion models.

Researchers have proposed various methods to address these challenges. For example, to improve sampling efficiency, many advanced SDE solution methods have been applied to diffusion models (Lu et al. 2022; Zhang et al. 2023; et al. 2023; Huang et al. 2024; Zheng et al. 2023). Meanwhile, diffusion distillation can also be used, which trains a large model but uses a small model to accelerate the sampling algorithm (Salimans et al. 2022; Li et al. 2023; Wizadwongsa et al. 2023; Wenliang et al. 2023). In summary, these efficient sampling methods can be divided into two main categories: learning-free sampling and learning-based sampling, with the difference between them being whether additional learning processes are required after the training of the diffusion model has been completed (Yang et al. 2024; Luo 2023). In addition, the design of new forward processes can also be used to improve sampling stability and reduce dimensionality (Yilun et al. 2023; Vahdat et al. 2021; Rombach et al. 2022).

When training a diffusion model, the (negative) variational lower bound (VLB) on the log-likelihood is used as the target, which may not be tight in many cases. This can lead to suboptimal log-likelihood in the diffusion model (Kingma et al. 2021; Maggiora et al. 2024). Zheng et al. (2023) proposed an improved diffusion ODE maximum likelihood estimation technique from both training and evaluation perspectives. In addition, many methods have been proposed to further maximise VLB and log-likelihood values from different aspects, including Noise Schedule Optimization (Kingma et al. 2021; Nichol et al. 2022), Reverse Variance Learning (Bao et al. 2022), Exact Likelihood Computation (Song et al. 2020; Cheng et al. 2022). To overcome the likelihood optimization ignored in diffusion models due to the intractability of log-likelihood, MLE training (Song et al. 2021; Kingma et al. 2021; Huang et al. 2021) and hybrid loss (Nichol et al. 2022; Cheng et al. 2022) are proposed to improve likelihood training.

Almost all diffusion models use the Convolutional U-Net (Ronneberger et al. 2015) as their backbone, but Peebles and Xie (2023); Ma et al. (2024) introduced Transformer to diffusion models, further enhancing their generation capabilities and resulting in the recent high-performance video generation model Sora Liuet al. (2024). The diffusion model assumes that the data exists in Euclidean space, which initially only handles continuous data such as images. To improve performance on discrete data or other data types, Feature Space Unification (Vahdat et al. 2021) and Data Dependent Transition Kernels (Austin et al. 2021) have been proposed to extend the application scope of diffusion models.

4 Diffusion models in image generation

4.1 Style transfer

The task of image style transfer has received widespread attention in the research community, to transform images of one style (source style) into another style (target style) (Gatys et al. 2015; Johnson et al. 2016; Dumoulin et al. 2017; Xun et al. 2017), as shown in Fig. 3. This transformation can be achieved by training a model to learn the image features of the source and target styles and then generating new images based on these features. Recently, diffusion models have been introduced into this area, achieving better performance and high-fidelity style image generation (Wang et al. 2023; Yang et al. 2023; Wang et al. 2023; Qi et al. 2024).

To solve the problem of preserving image content in the diffusion model, Huang et al. (2022) propose a text-driven image stylization framework based on dual diffusion to control the balance between content and style. This method integrates the multimodal style information as a guide into the step-by-step diffusion process, and performs the reverse denoising process on this basis, so that the styled results can better retain the structural information of the content image.

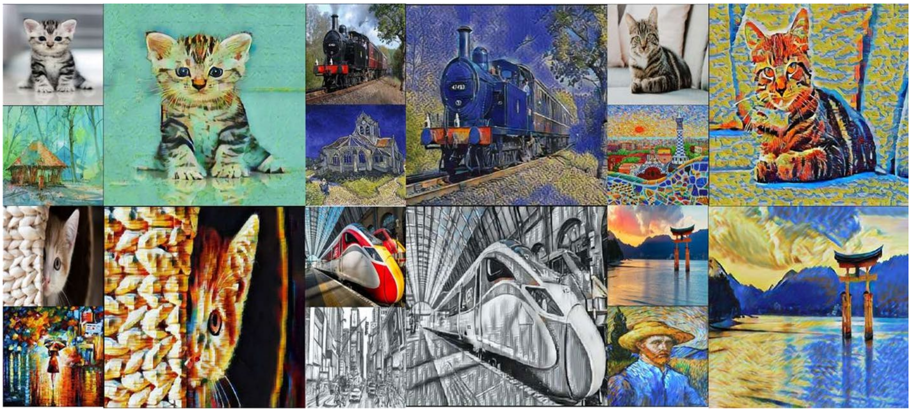


Fig. 3 Stylized images generated by Puff-Net (Zheng et al. 2024)

Brack et al. (2022) propose the Stable Artist, an iterative approach to guiding the generated images to the desired output. It achieves control by allowing the artist to guide the diffusion process along a variable number of semantic directions. This semantic guidance (SEGA) provides fine-grained control over the image generation process by exploiting complex operations in the underlying space of the model, allowing for tiny edits to the image, changes in composition and style, and optimization of the overall artistic concept.

Deng et al. (2023) propose a zero-shot (i.e. no training) training method through attention rearrangement, namely Z-STAR. This is a zero-shot image style conversion method that uses generative prior knowledge to transform image styles without retraining or adaptation. This approach can better solve the problem that the text prompt is too coarse to effectively express the required style details.

Chung et al. (2023) propose a style transfer approach that uses large-scale pre-trained diffusion models to simply manipulate features of self-attention, replacing the key and value of content with style, without the need for optimization or supervision (such as text). This method can effectively solve the problem that the existing style transfer methods based on diffusion model need to optimize the inference stage (such as fine-tuning or style text inversion), or cannot take advantage of the generation ability of large-scale diffusion model.

Zhang et al. (2023) propose a style transfer method based on inversion, namely InST. This method can effectively and accurately learn the key information of images to capture and transfer the style of painting art. This method is a good solution to the problems that specific artistic elements are difficult to transfer, the textual prompts of the target style can only be misdescribed, and it is difficult to reproduce the key ideas of specific paintings in the result.

To address the high cost of fine-tuning the diffusion model or additional neural network, Yang et al. (2023) propose a zero-shot contrastive (ZeCon) loss of the diffusion model without additional fine-tuning or auxiliary network to transfer the style of a given image and preserve its semantic content in the way of zero-shot. In addition, this method not only preserves the content but also realizes texture modification.

Wang et al. (2023) propose a new C-S disentanglement framework for style transfer. This framework can explicitly extract content information and implicitly learn complementary style information, which achieves interpretable and controllable C-S disentanglement and high-quality stylized results. They also further introduced the diffusion model into the C-S disentanglement framework, making the C-S disentanglement framework achieve SOTA results.

Pan (2023) propose an innovative style guidance approach that can improve the existing text-to-image diffusion model, while also supporting the use of reference images to guide arbitrary styles of generated images. This method optimizes the style guidance function to reduce the influence of noise input and maximize the guidance efficiency. As a result, the supervised style guidance and self-style guidance achieve effective results in generating the desired style images while maintaining a high correlation between the generated images and the text input.

Chen et al. (2023) propose a model called ArtFusion, which aims to provide a flexible balance between content and style for AST. The model utilizes a dual-condition latent diffusion probability model, which breaks the limitation of paired data in cDM training and promoting the progress of other multi-condition generation tasks. During the model training phase, the model transforms the style transfer task into a self-reconstruction task while maintaining robust stylization ability during inference.

Wang et al. (2024) propose a novel AST method called Highly Controllable Arbitrary Style Transfer (HiAST) to address the demand for flexible and customized stylized results. This model introduces a Style Adapter that allows users to flexibly manipulate the output stylized results by aligning multi-level style information and intrinsic knowledge in LDM.

4.2 Image restoration

Image Restoration (IR) is a long-standing problem due to its broad applicability and ill-defined nature. The goal of IR is to recover a high-quality (HQ) image from its low-quality (LQ) counterpart, which has been damaged by various degradation factors (e.g., blur, mask, downsampling), as shown in Fig. 4. The role of different image restoration methods is shown below:

Lugmayr et al. (2022) propose a redrawing method based on the denoising diffusion probability model (DDPM) that does not require specific mask training. The reverse diffusion iteration is modified by sampling the unmasked region with the given image information. Instead of learning the mask condition generation model, this model samples the condition generation process from a given pixel in the reverse diffusion iteration and is not trained for the internal spray-painting task itself.

Luo et al. (2023) propose to use a latent diffusion model to achieve realistic image recovery at large scale. A U-net-based latent diffusion strategy is proposed, which allows image restoration to be performed in a compressed and low-resolution latent space, thus speeding up training and inference.

Lin et al. (2023) propose DiffBIR, which applies a pre-trained text-to-image diffusion model to the problem of blind image restoration. This method outperforms state-of-the-art approaches in blind image super-resolution and face restoration tasks on both synthetic and real-world datasets. Furthermore, DiffBIR can effectively handle severe degradation and restore both realistic and vivid semantic content.

Qiu et al. (2023) propose a bootstrap diffusion model, DiffBFR, for blind face recovery. DiffBRF effectively employs the diffusion model to solve the problem of blind face recovery, which not only reduces the training difficulty and training time of the whole model

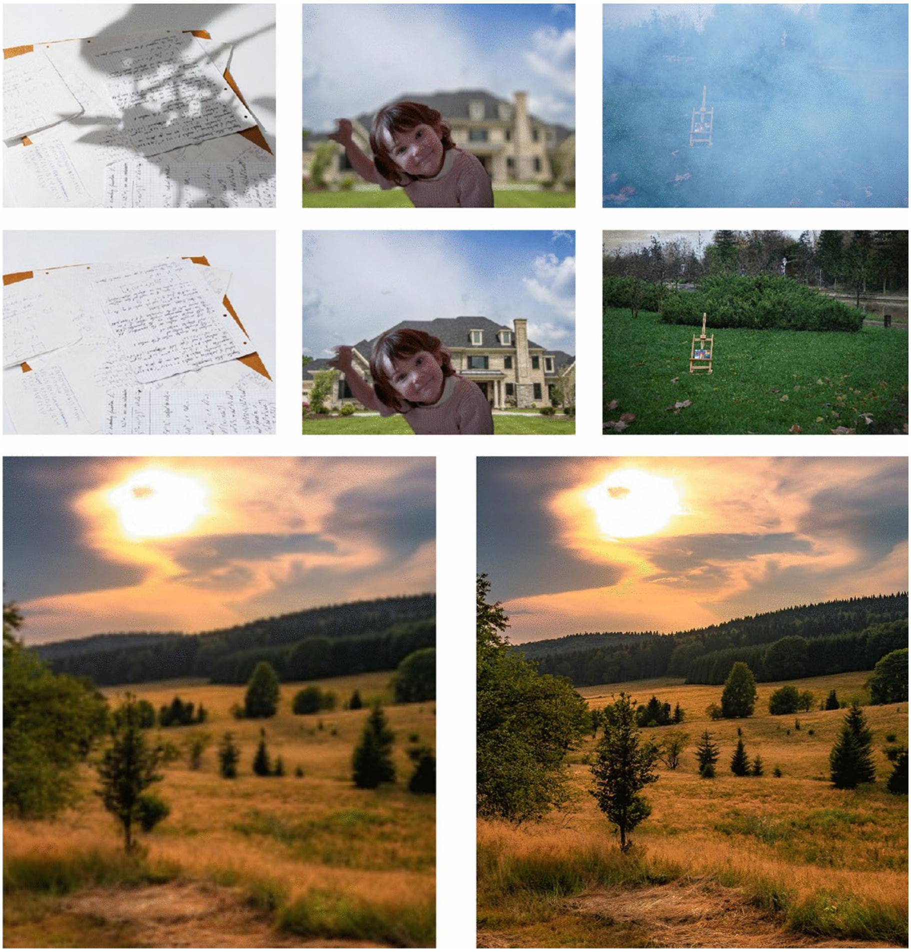


Fig. 4 Visual images of a variety of synthetic and real-world restoration tasks (Luo et al. 2023)

but also provides a less degraded input truncated sampling module with severe conditions. Moreover, this method outperforms GANs in terms of avoiding training collapse and generating long-tailed distributions.

Ren et al. (2023) introduce a simple and effective multiscale structure guide as an implicit bias to inform the icDPM about the coarse structure of sharp images in the middle layer. This guide leads to significant improvements in deblurring results, especially in invisible regions. The model can recover clean images more accurately and effectively.

Wang et al. (2023) propose a Coarse-to-Fine Diffusion Transformer (C2F-DFT). The C2F-DFT is built from a diffusion transformer block containing Diffusion Self-Attention (DFSA) and Diffusion Feedforward Network (DFN).C2F-DFT can well embed diffusion in transformers, allowing them not only to model long dependencies, but also to take full advantage of the generative power of diffusion models to facilitate better image recovery.

Liu et al. (2023) propose a residual denoising diffusion model (RDDM). It decouples the traditional single denoising diffusion process into residual diffusion and noise diffusion.

Residual diffusion represents directional diffusion from the target image to the degraded input image and explicitly guides the reverse generation process of image recovery, while noise diffusion represents random disturbances in the diffusion process. RDDM can solve the uninterpretability problem of a single denoising process and unify different tasks that require different levels of certainty or diversity.

Chen et al. (2024) propose a hierarchical integrated diffusion model (HI-Diff) for real-world image deblurring. HI-Diff exploits the power of diffusion models to generate informative priors for better results and generalization in complex blurred scenes.

4.3 Image editing

As shown in Fig. 5, image editing is a technology that modifies, enhances, or synthesizes images, which can be used in a variety of application scenarios, such as artistic creation, entertainment, education, medical treatment, etc. The goal of image editing is to produce images of high quality, variety, and control while maintaining the naturalness and semantic correctness of the images. In recent years, image editing methods based on generative models have made remarkable progress, especially those based on diffusion models.

Pang et al. (2023) propose a new initialization method called cross initialization, which can significantly reduce the gap between initial and learned embeddings, thereby improving

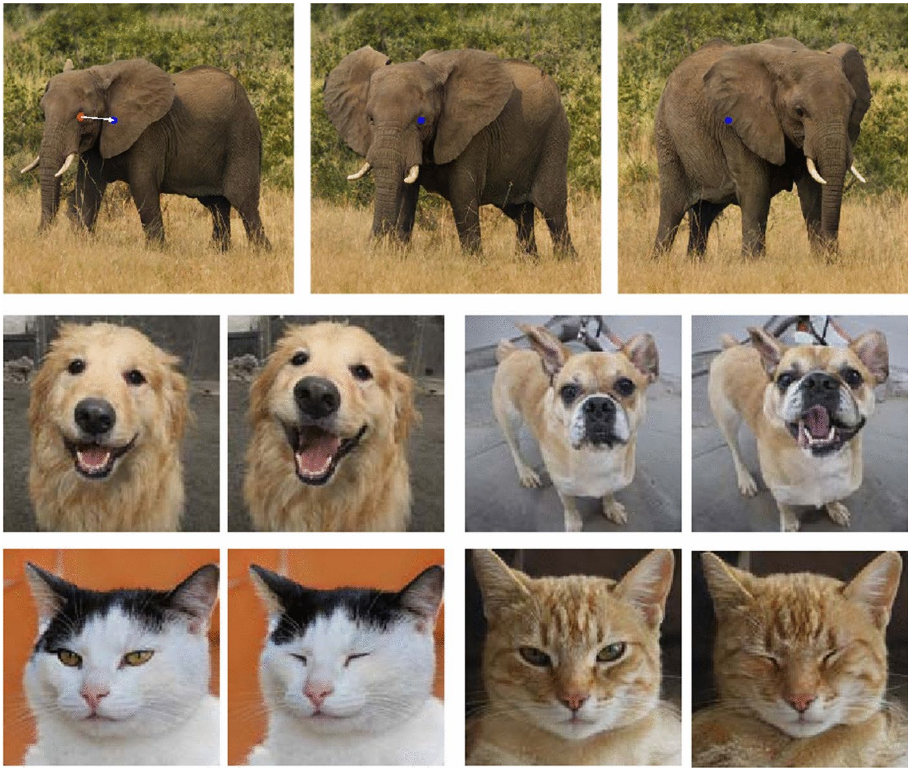


Fig. 5 The resulting images obtained by editing the animal images (Nguyen et al. 2024; Ling et al. 2024)

the reconstruction quality and editability of images. It also introduces a regularization term to make the learned embedding stay close to the initial one, further improving editability.

Choi et al. (2023) propose a new editing method called Custom-Edit, which consists of two steps: (i) using a small number of reference images to customize the diffusion model, and (ii) using effective text-guided editing methods to edit the images. They found that customizing only language-related parameters and using enhanced text prompts significantly improved the similarity of the reference image while maintaining the similarity of the source image.

Li et al. (2023) propose two novel subject-driven image editing subtasks, namely subject replacement and subject addition, which can realize more refined and flexible image editing. A new iterative generation method, called DreamEditor, is designed to achieve high-quality theme replacement and addition by gradually adapting to changes from the source theme to the target theme.

Wang et al. (2023) propose a general framework called MDP, which can explain various operations suitable for editing in diffusion models. They found five different operations, including intermediate latent variables, conditional embeddings, cross-attention graphs, guided and predictive noise, and analyzed the parameters and operation plans corresponding to these operations. They also demonstrated a new control method that can achieve higher-quality local and global editing than previous work by manipulating predictive noise.

Huang et al. (2023) propose a new sampling method called KV Inversion, which enables high-quality image editing without the need for fine-tuning. This method can change the action of the object in the image according to the textual prompt while maintaining the texture and identity of the object in the image. The advantage of KV inversion is that there is no need to train the diffusion model itself, nor does it need to scan large data sets for time-consuming training, but only to use the pre-trained diffusion model and text encoder to achieve image action editing.

Lin et al. (2023) propose an image editing method based on learnable regions that can specify the editing target and content through text prompts without requiring the user to provide masks or sketches. This method uses a pre-trained text-to-image model and introduces a bounding box generator to find the editing region aligned with the text prompt. Therefore, the mask-based text-to-image model can perform local image editing without masks or other user-provided guidance.

Aiming at the problem of reconstruction failure in real image editing, Chen et al. (2023) propose three sampling methods: FEC-ref, FEC-noise, and FEC-kv-reuse for different editing types and settings. The goal of these three methods is to ensure the success of reconstruction, that is, the result of sampling can retain the texture and features of the original real image, and cooperate with multiple editing methods to improve the performance of the editing task. All three methods do not require fine-tuning of diffusion models or training on large datasets, thus saving time and computational resources.

Nguyen et al. (2023) propose a new image editing method, Visual Instruction Inversion, which can guide the text-to-image diffusion model through visual prompts. A new visual cue generator is designed to generate a short text instruction to describe the changes between images based on the given source image and target image.

Kim et al. (2023) propose a simple and effective way to make the process of writing text prompts more user-friendly by incorporating a text generation framework. Specifically, they first classify text prompts into three categories based on the level of semantic detail:

simple, medium, and complex. They then use existing text generation frameworks, such as T5 (Raffel et al. 2020) and DALL-E (Ramesh et al. 2021), to generate medium or complex text prompts based on the target words entered by the user.

Dong et al. (2023) propose a method for image editing by text prompt. Given an original image and a target text prompt, the goal is to generate an edited image that is similar to the original image but conforms to the text prompt. The method uses a pre-trained text-to-image diffusion model and achieves editing through the technique of prompting adjustment inversion. It can realize flexible and diverse image editing without losing the details and quality of the original image.

Gu et al. (2024) propose a novel approach to an immersive image editing experience through personalized subject swapping. Photoswap first learns the visual concept of the subject from the reference image and then replaces it untrained into the target image using a pre-trained diffusion model. The effectiveness and controllability of Photoswap in personalized subject replacement show its wide application potential in entertainment and professional editing.

Han et al. (2024) propose a new editing method called Proximal Negative-Prompt Inversion (ProxNPI), It is an extension of the concepts of Negative-prompt inversion (NPI) Miyake et al. (2023) and Null-text inversion (NTI) (Mokady et al. 2023). ProxNPI reduces artifacts by introducing a regularized term and reconstructing the boot while preserving the untrained property.

4.4 Super-resolution

As shown in Fig. 6, the super-resolution task aims to enhance a low-resolution image or video to a high resolution through algorithms and models, while preserving and recovering as much detail information as possible in the image and reducing blurring and distortion.

Saharia et al. (2022) introduce a method called SR3 for image super-resolution. This method aims to achieve efficient and realistic image super-resolution processing by combin-

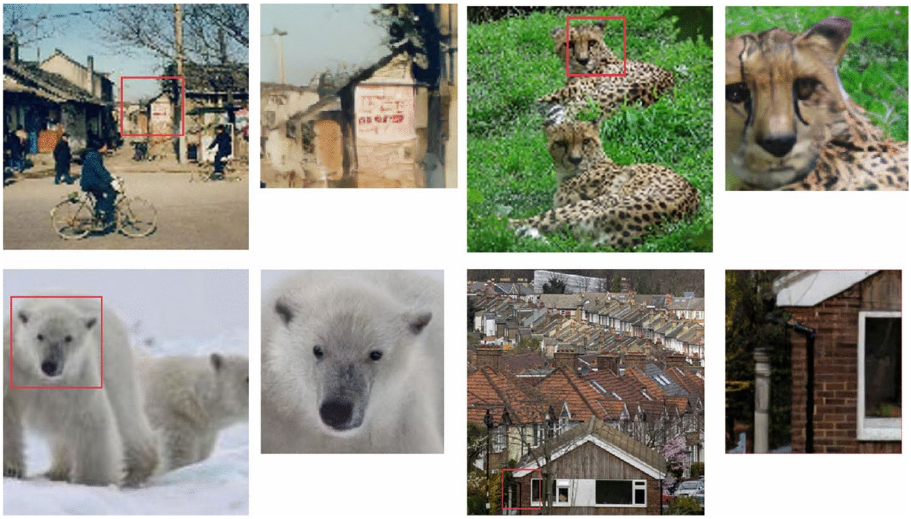


Fig. 6 Super-resolution results on real-world sample images (Wang et al. 2024)

ing a denoising diffusion probability model and a U-Net model. The innovation of SR3 lies in the use of the diffusion probability model and the implementation of the reverse process using the U-Net architecture, effectively avoiding the complex intrinsic optimization problems of traditional autoregressive models.

Li et al. (2022) investigate a single image super-resolution method based on diffusion models - SRDiff. This method adopts the U-Net structure and achieves a stable and reliable training process by introducing multi-scale skip connections and diffusion models, thus generating high-quality and diverse super-resolution images. SRDiff performs excellently on various datasets, avoiding the excessive smoothness and mode collapse problems of traditional methods, while also supporting flexible image operations.

Zhao et al. (2023) introduce a novel image super-resolution method based on diffusion probability models - Partial Diffusion Models (PartDiff). This method gradually diffuses the low-resolution input image into an intermediate latent state of the high-resolution image and performs partial denoising operations, to achieve high-quality image super-resolution. PartDiff achieves good performance on both magnetic resonance imaging and natural images.

The Diffusion Rectification and Estimation-Adaptive Models (DREAM) framework, proposed by Zhou et al. (2023), aims to address the problem of training and sampling inconsistency in conditional diffusion models for super-resolution tasks. By introducing two key components, diffusion rectification, and estimation adaptation, the DREAM framework effectively improves the quality of generated images while accelerating the training convergence speed and sampling efficiency.

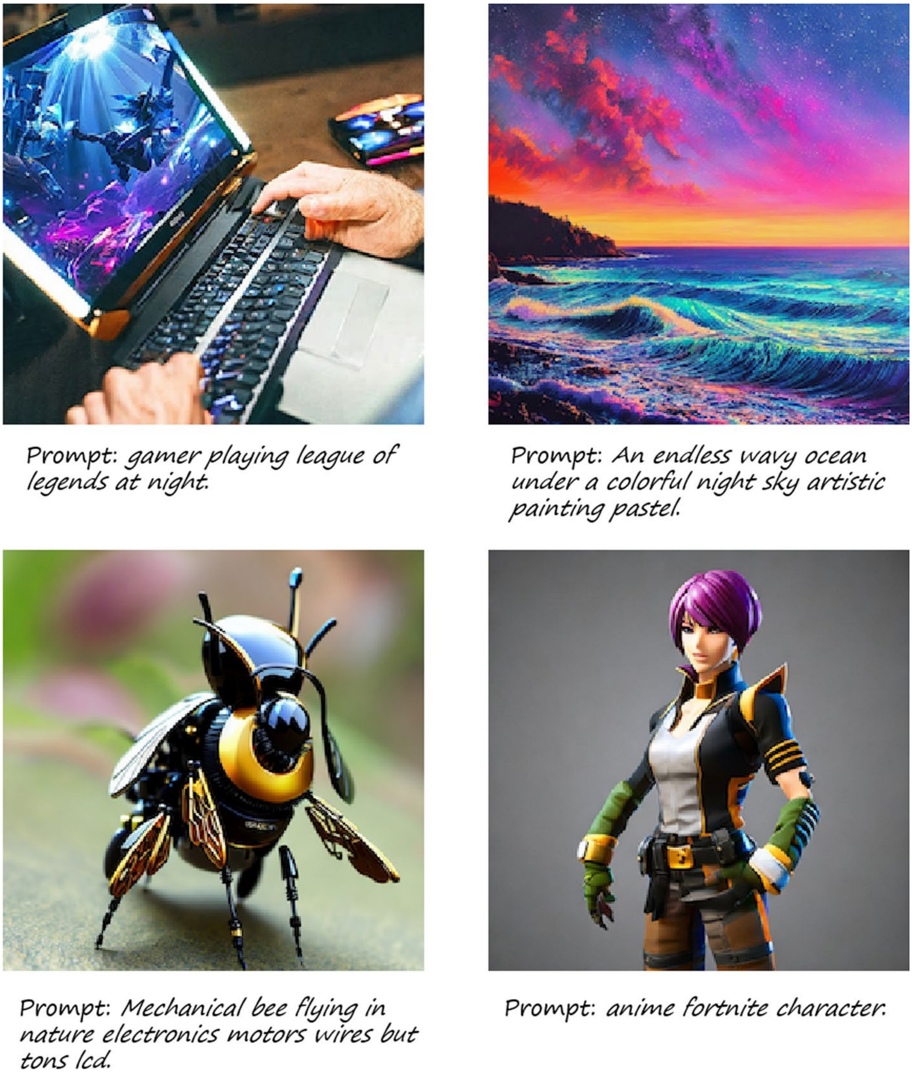
Wang et al. (2024) propose a simple and effective single-step SR generation method, SinSR, to overcome the limitation of the number of inferences faced by recent methods to improve the inference speed. This method reduces the number of inference steps for mapping between random noise at training time and generating high-resolution images by deriving deterministic sampling and proves that this deterministic mapping can use only one inference step to perform SR's student model.

Gandikota et al. (2024) introduce the zero-shot text guidance problem of an open-domain image super-resolution solution. This approach allows users to explore diverse, semantically accurate reconstructions to maintain data consistency with low-resolution inputs with different large downsampling factors, without the need for explicit training for these specific downgrades.

To eliminate the problem of artifacts in the iterative process of traditional diffusion-based SR techniques,Qingping et al. (2024) propose a training-free method, Adaptive Reality-Guided Diffusion (SARGD), which can effectively identify and mitigate the propagation of artifacts in latent spaces. The method first uses an artifact detector to identify untrustworthy pixels to create a binary mask that highlights artifacts and then uses Reality Guided Optimization (RGR) to integrate this mask with a realistic latent representation to optimize artifacts to improve alignment with the original image.

4.5 Text-to-image generation

As shown in Fig. 7, the text-to-image generation task refers to the process of automatically converting the input text description into the corresponding visual image through natural language processing technology, aiming to achieve seamless conversion and fusion between language and image (Müller et al. 2013; Yarom et al. 2023; Zhang et al. 2023). With the



Prompt: *gamer playing league of legends at night.*

Prompt: *An endless wavy ocean under a colorful night sky artistic painting pastel.*

Prompt: *Mechanical bee flying in nature electronics motors wires but tons lcd.*

Prompt: *anime fortnite character.*

Fig. 7 Several examples of text-to-image generation (Liang et al. 2024)

current development of deep learning technology (LeCun et al. 2015; Bengio et al. 2021), especially diffusion models, the quality of generated images has been greatly improved, and text-to-image generation has become the most attractive application in the field of computer vision (Reed et al. 2016; et al. 2018; Yu et al. 2022; Nichol et al. 2022; Jiayi et al. 2024).

Nichol et al. (2022) explored the application of guided diffusion to textual conditional image synthesis problems and compared two different guidance strategies: CLIP guidance (Radford et al. 2021) and classifier-free guidance (Jonathan and Salimans 2022). This method is the first attempt to apply a diffusion model to text-to-image generation, and it intuitively replaces the class labels in the class-conditioned diffusion model (i.e., ADM (Dhariwal et al. 2021)) with text, so that the sample generation is limited by text conditions.

Table 2 Summary of the ethical and social implications of image generation based on diffusion models and corresponding countermeasures

Models (Papers)	Conference and year	Methods and Insights	Target tasks	Datasets
MIDM Hu et al. (2023)	ICIS 2023	Differential privacy as a defense measure	Data Privacy	FFHQ, DRD
DFDM Zhu et al. (2023)	Arxiv 2023	Reconstruction-based member inference attack	Data Privacy	CelebA, ImageNet
DP-Foundation Model Images Linet al. (2023)	Arxiv 2023	Generating privacy-protected synthetic data via API	Data Privacy	CIFAR10, Camelyon17, Cat Images
ETDM Carlini et al. (2023)	USENIX 2023	Extracting training data from diffusion models	Data Privacy	CelebA, CIFAR-10
Degeneration-Tuning Ni et al. (2023)	ACM MM 2023	Degeneration-Tuning (DT) for concept protection	Data Privacy	COCO-30K
RMIA Dubinski et al. (2024)	CVPR 2024	Evaluation framework for membership inference attacks	Copyright Issues	
UnlearnDiffAtk Zhang et al. (2023)	Arxiv 2023	UnlearnDiff	safety-driven unlearned DMs, adversarial attacks	Unlearncanvas
Unified Concept Editing Gandikota et al. (2024)	CVPR 2024	Unified Concept Editing for copyright protection	Copyright Issues	
Concept Ablation Kumari et al. (2023)	CVPR 2023	Concept ablation method based on diffusion model	Copyright Issues	
CIDM Zhang et al. (2023)	Arxiv 2023	Data generation pipeline for copyright investigation	Copyright Issues	WikiArt, Dreambooth
Success of Diffusion Models in Artistic Imitation [20]	Arxiv 2023	Measuring model performance in imitating artists	Copyright Issues	LAION
Stable Bias Luccioni et al. (2023)	Arxiv 2023	Evaluation method based on social attributes	Bias and Fairnesss	Professions
Social Biases in T2I Generation Naik and Nushi (2023)	AAAI 2023	Systematic investigation of social biases in text-to-image models	Bias and Fairnesss	
RS-Corrector Jiang et al. (2023)	Arxiv 2023	RS-corrector framework to correct racial stereotypes	Bias Correction	
Fair Diffusion Model De Simone et al. (2023)	Arxiv 2023	Fair Diffusion model for fairness and transparency	Fairness in Text-to-Image Diffusion	LAION-5B
ENTIGEN Bansal et al. (2022)	Arxiv 2022	Ethical natural language interventions on text-to-image models	Bias and Fairness	ENTIGEN
Fair Diffusion Friedrich et al. (2023)	Arxiv 2023	Fair Diffusion strategy to reduce bias	Bias and Fairness	
Safe Latent Diffusion Schramowski et al. (2023)	CVPR 2023	Secure Latent Diffusion approach	Inappropriate Image Generationmodels	

Table 2 (continued)

Models (Papers)	Conference and year	Methods and Insights	Target tasks	Datasets
Finetuning for Fairness Shen et al. (2023)	Arxiv 2023	Distribution alignment loss function and fine-tuning	Fairness in Text-to-Image Diffusion biases in diffusion models	LAION-AESTHETICS
Fair Mapping Li et al. (2023)	Arxiv 2023	Fair Mapping method for fairer image generation	Fairness in Text-to-Image Diffusion	
All but One Seunghoo and Juhun (2024)	AAAI 2024	Concept Eraser algorithm	Inappropriate Image Generation	
Unsafe Diffusion Qu et al. (2023)	ACM CCS 2023	Study of potential risks in text-to-image models	Inappropriate Image Generation	4chan, Lexica, COCO
Adversarial Nibbler Report Brack et al. (2023)	Arxiv 2023	Distilling adversarial prompts from safety benchmarks	Inappropriate Image Generation	I2P
Mitigating Inappropriateness Brack et al. (2023)	Arxiv 2023	Negative cuing and semantic guidance	Inappropriate Image Generation	I2P
Red-teaming Stable Diffusion Rando et al. (2022)	Arxiv 2022	Security problem of stable diffusion model	Inappropriate Image Generation	
Prompting4Debugging Chin et al. (2023)	Arxiv 2023	Prompting4Debugging	Inappropriate Image Generation	I2P
Backdooring TI for Censorship Zhang et al. (2023)	Arxiv 2023	Backdooring textual inversion for concept censorship	Inappropriate Image Generation	
Forget-me-not Zhang et al. (2023)	Arxiv 2023	“Selflessness” solution for privacy, copyright, and security	Inappropriate Image Generation	M-Score, ConceptBench
SA Heng et al. (2024)	NeurIPS 2024	Selective Amnesia for selective forgetting	Inappropriate Image Generation	
Exposing FI Qiang et al. (2023)	PRL 2023	Developed a hybrid neural network model integrating AGFE and ViTFE modules to enhance fake trace feature representation	Fake image detection	
Mmganguard Raza et al. (2024)	IEEE Access 2024	Proposed MMGANGuard framework using transfer learning and multi-model fusion for automatic fake image identification	Fake image detection	RealandFakeFaces, Nvidia Flickr, FAKE

Table 2 (continued)

Models (Papers)	Conference and year	Methods and Insights	Target tasks	Datasets
CFMD Tassone et al. (2024)	CVIU 2024	Proposed analysis of continual learning techniques for fake media detection, focusing on short and long sequence fake media	Deepfake detection	CelebA, ILSVRC12
Famsec et al. (2023)	Arxiv 2024	Developed FAMSeC, a method to train a general-purpose detector with limited samples while maintaining generalization	AI-generated image detection	GenImage, ProGAN, ForenSynths, UniversalFakeDetect
Detecting FI Chen et al. (2024)	PRCV 2024	Proposed a CNN and Transformer-based detection model to identify artificially generated images from SDMs	Fake image detection	
Fakeinversion Cazenavette et al. (2024)	CVPR 2024	Proposed FakeInversion technique using inverted features from pre-trained Stable Diffusion models to detect synthetic images	Fake image detection	LAION, magenet, COCO, UCID

Compared to CLIP guidance, classifier-free guidance is preferred by human evaluators in terms of realism and text similarity.

Saharia et al. (2022) proposed Imagen, which abandons the cumbersome steps of GLIDE that require pre-trained visual-language models, and directly uses large language models such as T5 (Raffel et al. 2020) as text encoders, combined with diffusion models, to complete the direct association mapping from text to images. More importantly, they found that the general large language model is a very efficient text encoder for text-to-image generation, and increasing the size of the text encoder can effectively improve the quality of the generated samples and the alignment between text and image compared to increasing the size of the image diffusion model. In addition, the latest Imagen 3 (Jason et al. 2024), based on the latent diffusion model (Rombach et al. 2022), not only greatly improves text-image alignment to produce high-quality images, but also discusses security and representation issues.

Ramesh et al. (2022) proposes a two-stage model, in which a text is first given, an image embedding similar to CLIP is generated by the prior model, and then the image is generated by a decoder under the condition of image embedding. Both the prior and the decoder use the diffusion model. This method develops a method for training the diffusion prior in latent space and shows that it has comparable performance to the autoregressive Prior but with higher computational efficiency. [13] proposed DALL-E3 to address the difficulty of text-to-image models in following detailed image descriptions. By training the highly descriptive generated image titles, this method can greatly improve the prompt following ability of the text-to-image model.

In addition, diffusion-based text-to-image technologies, such as Stable Diffusion and Midjourney [1], show great potential for commercial applications. These models can transform simple text descriptions into high-quality images, greatly accelerating up the content creation process.

To solve the problem of high computational cost required by modern text-to-image models to generate high quality images, Jayasumana et al. propose a lightweight method Structured prediction for efficient text-to-image generation (2024) to optimize image region compatibility, reduce computational cost, and improve image quality by using Markov Random Field (MRF). Based on the Muse latent marker text-to-image model, the MarkovGen model combined with MRF speeds up the generation process and reduces artifacts to improve image quality.

Due to the shortcomings of the current diffusion model used in image generation to identify abstract continuous attributes, Cheng et al. (2024) propose a continuous 3D words technology to realize the fine control of multiple attributes in the image by users of the text-to-image model and realize efficient and burden-free image generation adjustment.

Inspired by the success of reinforcement learning with human feedback (RLHF) in large language models, Liang et al. (2024) proposed a variety of ways to enrich human feedback information and train a multimodal transformer to automatically predict these feedbacks to improve image generation.

4.6 Other tasks

A considerable number of notable works have employed diffusion models to address a range of subtasks associated with image generation. This paper will examine these applications and contributions in detail.

Low-light Image Enhancement Shang et al. (2024) propose a multi-domain multi-scale (MDMS) diffusion model for low-light image enhancement to address the limitations of the diffusion model and thus improve the quality of the generated images. Yi et al. (2023) proposed a physically interpretable diffusion model for low-light image enhancement, and solved various degradation problems in the image generation process by designing a multipath generative diffusion network, including noise, color bias, and dark illumination.

Image Denoising Zeng et al. (2024) introduce the diffusion model into the hyperspectral image denoising scene, and proposed a method Diff-Unmix that uses the denoising diffusion model to perform self-supervised denoising, which solves the problem that the current supervised denoising method is limited by the dataset.

Camouflaged Vision Perception Fan et al. (2022, 2023). To solve the problem that the current camouflage image generation methods require humans to specify the background and lead to high cost, Zhao et al. (2024) propose a Latent Background Knowledge Retrieval-Augmented Diffusion (LAKE-RED) to generate camouflage images, to expand the diversity of camouflage image samples at low cost.

Medical Generative Modeling Chenlu et al. (2024) unify the medical generation task and the unified generation task of the medical model, and proposed to align, extract and generate the multimodal medical model MedM2G in the unified model, which greatly enhanced the comprehensive diagnostic ability of the multimodal medical model.

Monocular Depth Estimation Ke et al. (2024) explore the role of the diffusion model's ability to capture a wide range of priors in the depth estimation task, and proposed an affine-invariant monocular depth estimation method based on the use of stable diffusion to retain prior knowledge, which improved the model's ability to understand new scenarios. Patni et al. (2024) explored the use of a global prior of images generated by a pre-trained ViT-based diffusion model to provide richer contextual information.

Representation Learning Hudson et al. (2024) propose a self-supervised diffusion model for representation learning. This method uses the diffusion model as a powerful representation learner to realize semantic learning in an unsupervised manner, and clarifies the potential of the diffusion model for learning rich representations.

Texture Synthesis Yeh et al. (2024) introduce a new image-guided texture synthesis method based on the diffusion model, which overcomes the limitations of traditional methods by using densely sampled views and precisely aligned geometric images, and greatly improves the quality of texture generation.

Causal Attribution Asnani et al. (2024) propose a causal attribution technique, ProMark, that allows the images generated by the generative model to be attributed to the model's training data, such as images, objects, artists, etc.. This approach can transform creative workflows by enabling creators to generate relevant content for model training in order to earn rewards.

Image Classification Wang et al. (2024) propose a new inter-class data augmentation method, Diff-Mix, which can generate images that conform to the diversity of foreground objects and backgrounds for specific concepts, thereby improving image classification performance.

Image Morphing (Wolberg 1998) Zhang et al. (2024) propose DiffMorpher, a method for smoothing and interpolating images using prior knowledge of pre-trained diffusion models. This method uses two LoRA (Hu et al. 2022) fitting images to capture the semantics, and then interpolates between the LoRA parameters and the latent noise to achieve a smooth transition of the semantics, which solves the problem that the latent space of the current diffusion model is highly unstructured.

Image Inpainting Liu et al. (2024) explore the problem of semantic differences between masked and unmasked regions in the diffusion model for image inpainting, and proposed StrDiffusion, an image inpainting diffusion model that reconstructs texture denoising under structural guidance.

Scene Completion Nunes et al. (2024) explore the application of the diffusion model in 3D point cloud generation, and instead of the previous work directly using image-based diffusion methods Lee et al. (2023); Luo and Wei (2021); Lyu et al. (2022), it proposes to operate directly on the points, and redesigns the forward and backward processes of the diffusion model to make it work effectively within the 3D point cloud scene.

Intrinsic Image Decomposition The ambiguity between illumination and material properties, as well as the lack of real-world data sets, make the appearance decomposition task quite challenging. Kocsis et al. (2024)use the powerful prior knowledge of the latest diffusion models to sample from the solution space of the conditional generation model, which greatly improves the generalization of the model to the real image.

Image Deblurring To construct an efficient training dataset based on the generated realistic blurred images, et al. (2024) propose reBLurring AUgmentation (ID-Blau), which uses clear images and controllable blur conditions to pair to generate corresponding blurred images, to realize the diversified generation of blurred images.

From the above discussion, it can be seen that diffusion models are widely used in all corners of the field of image generation, and the above discussion is only a part of the application of tasks. In addition, the diffusion model is also applied to Image Rectangling (Zhou et al. 2024), Image Segmentation (Baranchuk et al. 2022; et al. 2023), Semantic Matching (Li et al. 2024), Visual Emotion Analysis (Yang et al. 2024), Face Recognition (Boutros et al. 2023), Anomaly Detection (Zhang et al. 2023), etc.

5 Ethical and social implications

As technology progresses, it is imperative to contemplate the harmonization of artistic innovation with ethical responsibility. For instance, in the case of diffusion models used for image generation, it is imperative to adhere to the principles of copyright and refrain from encroaching upon the intellectual property rights of others. Concurrently, it is imperative to consider the influence of technology on societal culture and values in order to attain sustainable and responsible technological advancement. The generation of images by diffusion models has the potential to result in the dissemination of misleading information. Such occurrences have the potential to influence public perception and decision-making processes. It is, therefore, incumbent upon developers and users to ensure that generated images are accompanied by clear labels and instructions, in order to obviate any potential for misdirection on the part of the user.

5.1 User privacy data leakage

Although diffusion model-based image generation technology has shown great potential in the fields of creativity, entertainment, and design, the problem of user privacy leakage cannot be ignored. If this technology is used improperly, it may unknowingly leak sensitive information such as personal identity and living habits by collecting and analyzing image data uploaded by users, posing a serious threat to individual privacy (Hu et al. 2023; Zhu et al. 2023; Linet al. 2023; Carlini et al. 2023; Ni et al. 2023).In addition, privacy leakage may also lead to public distrust of new technologies, hinder the healthy development of technological innovation and application, and affect the reputation and prospects of the entire industry. Therefore, strengthening data protection and ensuring users' privacy security are key issues that must be taken seriously in the promotion and application of such technologies.

Carlini et al. (2023) study the memory capacity of image diffusion models (such as stable diffusion) in the training data and the privacy problems it causes. By proposing a new data extraction method, the paper successfully extracts a large number of training examples from the diffusion model and shows how to identify these memory-generating samples by generating and filtering them. The researchers also thoroughly analyzed the effects of different types of models and parameter settings on the degree of privacy leakage, and found that the diffusion model has a higher risk of privacy leakage than other generative models. Zhu et al. (2023) investigate the privacy leakage problem of diffusion model in image generation, propose a reconstruction-based member inference attack method, and conduct experimental verification on several pre-trained diffusion models. This method realizes member inference by reconstructing images and calculating reconstruction errors. Compared with traditional gradient-based attack methods, this method has higher efficiency and stability, and is more difficult to defend. The experimental results show that the attack method can effectively infer the information of the members in the training dataset, and reveal the potential risks of the diffusion model in privacy protection.

Currently, many researchers have also conducted in-depth research and proposed solutions. Hu et al. (2023) explores the effectiveness of differential privacy technology as a potential defense measure, and points out that future research should focus on how to strike a balance between protecting privacy and improving model quality. Linet al. (2023) explores a solution for generating privacy-protected synthetic data via API interfaces using generative adversarial networks such as Stable Diffusion. They propose a framework called Private Evolution (PE), which uses evolutionary algorithms and models supported by existing APIs to generate synthetic data similar to the distribution of private data through an iterative process that ensures differential privacy protection. Ni et al. (2023) proposes a new method called Degeneration-Tuning (DT). The core idea of the method is to create a degraded dataset by disrupting low-frequency visual content and retuning the stable diffusion model to mask unwanted concepts when generating images, thus protecting certain concepts from attacks or leaks.

5.2 Copyright issues

The rapid development of image generation technology based on diffusion models has greatly enriched the opportunities for creative expression and visual content production, but it also raises serious copyright issues (Wang et al. 2024; Dubinski et al. 2024). These technologies can imitate or even create original works with high fidelity, making copyright ownership ambiguous and affecting the protection of rights for original creators (Gu et al. 2024; Zhang et al. 2023). Unauthorized use or reproduction of other people's work styles for creation may infringe the copyrights of original creators, dampen their enthusiasm for creativity and hinder the healthy development of cultural industries. At the same time, it also increases the difficulty of copyright supervision and rights protection, and poses new challenges to the legal system.

In this regard, the solutions not only include improving laws and regulations, strengthening industry self-discipline, improving public awareness of copyright, and exploring new copyright protection modes, but researchers have also explored a large number of technical solutions from the technical aspect. For the copyright challenges of the diffusion model, Gandikota et al. (2024) proposes a method called Unified Concept Editing (UCE), which precisely edits models through closed-form solutions to eliminate copyrighted content, correct bias, and control inappropriate concepts. Kumari et al. (2023) propose a concept ablation method based on a diffusion model, including two schemes based on noise and anchor point, which avoid generating target concepts by adjusting the KL divergence between distributions. This method can effectively remove the target concept while retaining the related concept, which proves its practicality in copyright protection and privacy enhancement. Zhang et al. (2023) addressed a part of the infringement problem, namely the generation of infringing content using queries not directly related to the copyrighted subject matter. The authors develop a data generation pipeline for generating copyright investigation datasets for diffusion models, and through this pipeline generate datasets containing infringement examples for different diffusion models. Casper et al. (2023) proposed and implemented a simple and quantitative method to measure the performance of the model when imitating a particular artist by combining CLIP encoders and standard techniques. The results of the experiment showed that Stable Diffusion was able to successfully mimic the style of most professional digital artists, a finding that has important implications for addressing the correlation between AI-generated images and copyright law. In addition, the paper discusses how image classification techniques can be used to analyze legal claims and test defense strategies against AI imitations of copyrighted works.

In addition, for the commercial application of the diffusion model, researchers also proposed an evaluation framework and an attack method to evaluate the copyright security of the model. Wang et al. (2024) proposed a data pollution attack method called SilentBadDiffusion. This method uses multimodal large language models and text-guided image-filling techniques to generate images with specific prompts, and then injects this "poisoned" data into the training process of the diffusion model. The experimental results show that only a small amount of poisoning data is needed to enable the fine-tuned diffusion model to generate infringing content under specific trigger prompts. Dubinski et al. (2024) propose a new evaluation framework and attack method to evaluate the effectiveness of members' inferred attacks. By designing a new dataset, LAION-mi, the authors find that previous evaluation schemes fail to fully reflect the true impact of members' inferred attacks, revealing the serious privacy and copyright issues that large diffusion models face when processing copyrighted images. At the same time, this paper also highlights the challenges of shadow model attacks, such as high computational cost and difficult sampling.

5.3 Bias and fairness problem

The rapid development of the diffusion model in the field of image generation has greatly enriched the possibilities of content creation and visual presentation, but has also brought with it problems of prejudice and fairness that cannot be ignored (Luccioni et al. 2023; Naik and Nushi 2023; Jiang et al. 2023; De Simone et al. 2023). These issues can not only exacerbate existing inequalities in society, such as the automatic reproduction and propagation of gender, racial, or cultural stereotypes, but also limit the diversity of innovation and affect the inclusiveness and ethical acceptability of technology. In the long run, without effective governance, it will hinder the healthy development of technology, damage public trust, and pose a threat to the cultural diversity and inclusiveness of society.

Bansal et al. (2022) study the effect of moral natural language interventions on text-to-image generation models, specifically the performance of the stable diffusion model. Using the ENTIGEN dataset, the authors evaluated the impact of ethical interventions on image-generating diversity across three social axes: gender, skin color, and culture.

Naik and Nushi (2023) aim to systematically investigate and quantify the social biases in text-to-image generation models. Using the stable diffusion model as an example, this paper examines its performance in terms of gender, race, age, and geographic location. By designing a series of experiments, including the use of different prompt words and automated and human scoring methods, the study found that the Stable Diffusion model has a significant bias in image generation.

Luccioni et al. (2023) aim to explore the problem of social bias in machine learning driven text-to-image (TTI) systems, specifically for the performance of stable diffusion models. By proposing an evaluation method based on social attributes, combined with an analysis of occupational and social attributes, this paper reveals the gender and racial biases that exist in TTI systems when generating images.

Friedrich et al. (2023) mainly discuss the bias problem of artificial intelligence in text-to-image generation and proposes a new strategy called Fair Diffusion. It is designed to reduce or eliminate bias by controlling the direction and proportion of model output. The research focuses on the Stable Diffusion model, and reveals the model's gender bias in image generation through a series of experiments.

Schramowski et al. (2023) focus on solving the problems of bias and misbehavior in real-world applications of image generation models under textual conditions. Specifically for stable diffusion models. By introducing a "secure latent diffusion" (SLD) approach, the paper aims to filter and balance the training data to eliminate or suppress inappropriate parts of the image.

Shen et al. (2023) focus on solving the fairness problem of text-to-image diffusion models, especially the bias of stable diffusion models. By introducing the distribution alignment loss function and fine-tuning the sampling process, this paper aims to control the distribution of certain attributes of the generated image to achieve fairness and diversity.

Jiang et al. (2023) aim to correct for racial stereotyping in image generation models such as stable diffusion. By introducing a framework called RS-corrector, this method adjusts the hidden code in latent space to eliminate racial bias while maintaining the integrity of the original model.

Li et al. (2023) focus on solving the fairness problem in text-to-image diffusion models, especially for the bias that can arise when generating human-related descriptions. To this end, the authors propose the Fair Mapping method, which is a lightweight, general-purpose solution that does not depend on a specific model. With well-designed prompts to control sensitive attributes and adjust offsets in embedded spaces to correct semantic features of the original language, Fair Mapping aims to achieve fairer image generation.

De Simone et al. (2023) aim to solve the problem of bias in the generation of text-to-image (GTTI) models by designing and implementing a tool called the fair diffusion model to improve the fairness and transparency of the model. Through the interactive interface and editing options, the tool allows users to analyze and adjust the worldview of the model to ensure that the generated image meets the fairness standards expected by users.

Gandikota et al. (2024) mainly study various security problems in text-to-image models and propose a solution called "Unified concept editing (UCE)". The method uses a closed-form solution to accurately edit the model and supports simultaneous processing of multiple concepts such as bias, copyright, and offensive content without retraining the model. UCE enables targeted bias correction for multiple attributes while removing potentially copyrighted content and controlling inappropriate concepts.

5.4 Inappropriate image generation

The problem of inappropriate image generation faced by image generation technology based on diffusion models cannot be ignored. This problem may not only lead to the generation of violent, pornographic, or discriminatory images, which seriously violate social ethics, laws, and regulations, but also mislead the public and affect the healthy dissemination of information (Seunghoo and Juhun 2024; Qu et al. 2023; Brack et al. 2023, 2023; Rando et al. 2022). In addition, it can damage personal reputation and privacy, and aggravate the sense of insecurity and distrust in cyberspace. Therefore, how to effectively identify and prevent the generation of inappropriate images has become a key challenge to be solved in this field.

Rando et al. (2022) mainly study the security problem of the stable diffusion model in natural language processing tasks. Through the simulation of attacker behavior and reverse engineering analysis, the vulnerabilities of the model's security filter are revealed and the corresponding improvement measures are proposed. The researchers generated several types of images to test the performance of the security filter and successfully bypassed the filter to generate images with pornographic content. The conclusion points out that there are loopholes in the security filter of the current stable diffusion model, and it is necessary to strengthen the security through open documents and disclosure channels.

Chin et al. (2023) propose an automated tool called Prompting4Debugging (P4D) to detect security vulnerabilities that can lead to inappropriate image generation by optimizing prompt words. P4D uses prompt engineering technology to find modified prompts that bypass security mechanisms through continuous/soft embedding optimization and discrete/hard embedding projection. The experiments show that even seemingly secure prompts can be vulnerable to manipulation, underscoring the importance of fully testing the security of T2I diffusion models.

Qu et al. (2023) aim to comprehensively study the potential risks of text-to-image models in generating unsafe images and hateful messages. By constructing five kinds of unsafe image classification systems and using four advanced text-to-image models, it is found that these models have the risk of generating unsafe images, and the stable diffusion model is particularly prominent.

Brack et al. (2023) conduct an in-depth study of the security of image generation models under text conditions in the application. The goal of the study was to uncover systemic security issues in existing imaging models and to assess the impact of counterattacks.

Pham et al. (2023) conduct an in-depth study of concept erasure methods in text-to-image generation models. Seven different concept erasure methods are described in detail and their effects on pre-trained diffusion models are shown. The experimental results show that these concept erasure methods do not eliminate sensitive concepts, but reintroduce the "erased" concepts by adjusting the embedding of input words.

Zhang et al. (2023) study the problem of text inversion for concept censorship in the text-to-image generation model and proposed a solution based on backdoor technology. By selecting sensitive words as triggers during training and using these triggers in combination with personalized embedding in the generation phase, the model outputs predefined target images instead of images containing malicious concepts. Experimental results show that this method can effectively prevent the cooperation between text inversion technology and censored words, while maintaining the original function of the model.

Brack et al. (2023) focus on the problem of text-to-image models in generating inappropriate content and propose two solutions: negative cuing and semantic guidance (SEGA). The study aims to make the images generated by the model consistent with human preferences by evaluating and guiding strategies. In practice, negative cuing reduces the generation of inappropriate content by avoiding specific cues, while SEGA manipulates the image generation process by adding additional cues while minimizing changes to the original image. These guidance methods can effectively reduce the probability of generating inappropriate content, and SEGA performs better.

Zhang et al. (2023) focus on solving the privacy, copyright, and security problems in text-to-image generation models. In particular, the models can learn and generate unauthorized personal information, content, and potentially harmful content. To this end, the authors propose an efficient and cost-effective solution called "selflessness," which aims to remove a particular identity, object, or style from the model without affecting the model's ability to generate other content.

Heng et al. (2024) propose Selective Amnesia (SA) to solve the problem of selective forgetting in deep generation models. Combining Bayesian Continual Learning (BCL), the method integrates Elastic Weight Consolidation (EWC) and Generative Replay (GR) into a training loss function. It allows forgetting of specific concepts without access to the original training data set. This research provides a new solution to the problem that large text-to-image models can generate harmful, misleading, and inappropriate content.

Seunghoo and Juhun (2024) propose a new algorithm called "Concept Eraser". This method achieves the goal of removing or replacing specific concepts in the pre-trained model by modifying the drift of the classifier guide term and the unconditional score term. This algorithm can not only effectively erase the object concept, but also maintain the generative ability of the model.

5.5 Fake images

The proliferation of fake images, with their deceptive nature, has the potential to mislead the public, distort the truth, and contribute to misunderstanding and panic. It can harm trust, disrupt social stability, and have negative effects on individual reputation and mental health (Shen et al. 2019; Nash et al. 2009). Researchers have developed a number of detection and identification methods to address the threat of fake images. Among these, the method combining generative adversarial networks (GANs) and convolutional neural networks (CNNs) has demonstrated particularly promising results (Neves et al. 2020; Raza et al. 2024; Bhandari et al. 2023).

The advent of diffusion models in image generation has introduced new challenges to the authenticity and integrity of digital images. Qiang et al. (2023) conducted a comprehensive investigation into the collection mechanism of images generated by diffusion models and developed a hybrid neural network model that integrates attention-guided feature extraction (AGFE) and vision transformers (ViTs) based feature extraction (ViTFE) modules to enhance the representation of fake trace features.

Raza et al. (2024) have proposed an innovative framework, designated as Multi-Model GAN Guard (MMGANGuard), which employs transfer learning and multi-model fusion techniques to facilitate the automated identification of PAN-generated fake images, thereby enhancing the precision and scalability of the detection process.

Tassone et al. (2024) have proposed an in-depth analysis of the application of two continual learning techniques in addressing the generalization challenges faced by deepfake detection technology. This research involves a comprehensive examination of continual learning techniques for both short and long sequence fake media.

The efficacy of existing AI-generated image detection methods is contingent upon the availability of extensive training data, which often proves challenging to obtain when the number of samples is limited. et al. (2023) developed FAMSeC, a novel AI-generated image detection method that aims to train a general-purpose detector using a limited number of training samples while avoiding overfitting and maintaining the generalization capabilities of the pre-trained model.

Chen et al. (2024) explored the intricacies of differentiating genuine images from those generated by the Stable Diffusion Model (SDM). They devised a novel approach, comprising a convolutional neural network (CNN) and a Transformer-based detection model, which effectively identifies artificially created images from SDMs. Furthermore, they were the first to assess the generalisation capacity of these detection models across diverse scenarios.

Cazenavette et al. (2024) put forth a technique, designated as “FakeInversion, ” which employs features derived from open-source pre-trained stable diffusion models to identify synthetic images. A salient attribute of this technique is its capacity to generalise effectively to high-visual-fidelity invisible generators, even when trained exclusively on low-fidelity images generated by Stable Diffusion.

6 Challenges and future directions

Dataset limitation The accelerated advancement of image generation techniques based on diffusion models can be attributed to the accessibility of extensive, high-quality datasets. For instance, the current text-to-image synthesis employs billions of high-quality (text, image) pairs (Ramesh et al. 2022; Nichol et al. 2022). However, some other subtasks continue to grapple with data scarcity. Furthermore, datasets also confront challenges related to data bias, encompassing aspects such as language, ethnicity, and gender. These issues can give rise to substantial biases and fairness concerns.

High computational cost The principal challenges to diffusion models include the high cost of training and the number of steps in inference, which serve to exacerbate the disparity in access to resources between industry and academia (Blattmann et al. 2023; Ganguli et al. 2022). Despite efforts to reduce training costs, dataset size and time complexity remain significant obstacles. Furthermore, the model has difficulty generating readable text, and the computational requirements limit its deployment in real-world applications. Therefore, research should focus on improving the efficiency of the model, reducing the computational cost, and exploring further improvements in wavelet-based methods. Additionally, the success of deep learning relies on large amounts of labelled data, which poses a challenge for small companies and edge devices.

Image evaluation The current methods for evaluating image generation are limited in their ability to comprehensively assess quality, rely on user testing and subjective scoring, and are susceptible to bias (Saharia et al. 2022; Parmar et al. 2022; Radford et al. 2021). To address these shortcomings, it is essential to develop more targeted evaluation benchmarks and indicators, as well as more reliable and diverse automatic evaluation criteria.

Multimodal framework The generation of content from text to image, otherwise known as Artificial Intelligence Generated Content (AIGC), has attracted considerable interest from both academic and industrial perspectives. The current popular large language models (OpenAI 2023), based on autoregressive models, have achieved considerable success, particularly in terms of their capacity to generalise across legal domains and zero-shot tasks. Meanwhile, in the field of image generation, represented by models such as Stable Diffusion (Esser et al. 2024) and Sora (OpenAI 2023), diffusion models are widely adopted. As a crucial step towards general artificial intelligence, current research aims to integrate multiple tasks into a single model, thereby constructing multimodal models. Consequently, research interest has shifted towards analysing the emergence capability of diffusion models and developing versatile models capable of generating diverse outputs and handling various data types. This is considered a major challenge and research direction in image generation.

Data Security and Social Ethics The advent of models such as Stable Diffusion and Midjourney has precipitated a period of rapid development in the field of image generation, resulting in a notable increase in the stylistic and diverse range of image creation. However, this technological advancement has also been accompanied by data privacy violations, copyright disputes, and the potential for misinformation and disinformation. The existence of these problems not only threatens the rights and interests of individuals, but also poses a challenge to social trust and moral standards. It is therefore recommended that future research should focus on strengthening data ethics and privacy protection research, developing more transparent and explainable models, and improving the controllability of generated content.

7 Conclusion

We provide a multi-perspective for observing the development and impact of diffusion models in the field of image generation. We first introduce the development background of diffusion models from three basic theories: DDPM, SGMs, and SDEs, and explain some improvement methods of diffusion models in image generation. Second, we explore the wide application and high performance of diffusion models in various subfields of computer vision, including style transfer, image completion, image processing, super-resolution, 3D image generation, etc. Finally, we conduct a comprehensive analysis of the potential social and ethical implications and challenges of diffusion model-based image generation techniques. In summary, this paper provides an in-depth analysis and discussion of the application and potential social impact of diffusion models in the field of image generation. We hope that this survey can provide some guidance and inspiration for the future development of diffusion models in this field.

Acknowledgements The work is supported by Talent Introduction Program of Hubei Polytechnic University (23xjz14R) and 2023 Joint Open Fund for Hubei Key Laboratory of Mine Environmental Pollution Control & Remedition, Hubei Polytechnic University and Huangshi Daye Lake National Hi-tech Development Zone Science Park (23xjz08AK) and 2023 Hubei Provincial Department of Education Scientific Research Plan Guiding Project (B2023234) and Hubei Provincial Science and Technology Plan Project (KJRQ2023000110) and National Natural Science Foundation of China (61871178) and 2023 Hubei Polytechnic University-Wuhan Nanhua Industrial Equipments Engineering Co., Ltd. Joint Laboratory Open Fund.

References

Alexander Quinn Nichol and Prafulla Dhariwal (2021) Improved denoising diffusion probabilistic models. In International conference on machine learning, pp 8162–8171

Ali Raza Syed, Usman Habib, Muhammad Usman, Ashraf Cheema Adeel, Sajid Khan Muhammad (2024) MMGANGUARD: a robust approach for detecting fake images generated by GANS using multi-model techniques. IEEE Access 12:104153–104164

Anciukevicius Titas, Xu Zexiang, Fisher Matthew, Henderson Paul, Bilen Hakan, Mitra Niloy J, Guerrero Paul (2023) Renderdiffusion: Image diffusion for 3D reconstruction, inpainting and generation. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 12608–12618

Anderson Brian DO (1982) Reverse-time diffusion equation models. Stochast Process Appl 12(3):313–326

Arjovsky Martín, Chintala Soumith, Bottou Léon (2017) Wasserstein generative adversarial networks. In Proceedings of the 34th International Conference on Machine Learning, vol 70, pp 214–223

Asnani Vishal, Collomosse John, Bui Tu, Liu Xiaoming, Agarwal Shruti (2024) Promark: Proactive diffusion watermarking for causal attribution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 10802–10811

Austin Jacob, Johnson Daniel D., Ho Jonathan, Tarlow Daniel, Berg Rianne van den (2021) Structured denoising diffusion models in discrete state-spaces. In Advances in Neural Information Processing Systems, pp 17981–17993

Bansal Hritik, Yin Da, Monajatipoor Masoud, Chang Kai-Wei (2022) How well can text-to-image generative models understand ethical natural language interventions? Preprint at arXiv:2210.15230

Bao Qiqi, Hui Zheng, Zhu Rui, Ren Peiran, Xie Xuansong, Yang Wenming (2024) Improving diffusion-based image restoration with error contraction and error correction. In Thirty-Eighth AAAI Conference on Artificial Intelligence, pp 756–764

Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, Bo Dai (2023) Generative diffusion prior for unified image restoration and enhancement. In IEEE/CVF Conference on Computer Vision and Pattern Recognition 55:9935–9946

Betker James, Goh Gabriel, Jing Li, et. al. Improving image generation with better captions. https://api.semanticscholar.org/CorpusID:264403242

Blattmann Andreas, Rombach Robin, Ling Huan, Dockhorn Tim, Kim Seung Wook, Fidler Sanja, Kreis Karsten (2023) Align your latents: High-resolution video synthesis with latent diffusion models. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023, pp 22563–22575

Boutros Fadi, Grebe Jonas Henry, Kuijper Arjan, Damer Naser (2023) Idiff-face: Synthetic-based face recognition through fizzy identity-conditioned diffusion model. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp 19650–19661

Brack Manuel, Friedrich Felix, Schramowski Patrick, Kersting Kristian (2023) Mitigating inappropriateness in image generation: Can there be value in reflecting the world's ugliness? Preprint at arXiv:2305.18398

Brack Manuel, Schramowski Patrick, Friedrich Felix, Hintersdorf Dominik, Kersting Kristian (2022) The stable artist: Steering semantics in diffusion latent space. Preprint at arXiv:2212.06013

Brack Manuel, Schramowski Patrick, Kersting Kristian (2023) Distilling adversarial prompts from safety benchmarks: Report for the adversarial nibbler challenge. Preprint at arXiv:2309.11575

Cao Pu, Zhou Feng, Song Qing, Yang Lu (2024) Controllable generation with text-to-image diffusion models: A survey. Preprint at arXiv:2403.04279

Carlini Nicolas, Hayes Jamie, Nasr Milad, Jagielski Matthew, Sehwag Vikash, Tramer Florian, Balle Borja, Ippolito Daphne, Wallace Eric (2023) Extracting training data from diffusion models. In 32nd USENIX Security Symposium (USENIX Security 23), pp 5253–5270

Casper Stephen, Guo Zifan, Mogulothu Shreya, Marinov Zachary, Deshpande Chinmay, Yew Rui-Jie, Dai Zheng, Hadfield-Menell Dylan (2023) Measuring the success of diffusion models at imitating human artists. Preprint at arXiv:2307.04028

Cazenavette George, Sud Avneesh, Leung Thomas, Usman Ben (2024) Fakeinversion: Learning to detect images from unseen text-to-image models by inverting stable diffusion. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024, pp 10759–10769

Chang Ziyi, Koulieris George Alex, Shum Hubert P. H (2023) On the design fundamentals of diffusion models: A survey. Preprint at arXiv:2306.04542

Chen Dar-Yen (2023) Artfusion: Controllable arbitrary style transfer using dual conditional latent diffusion models. Preprint at arXiv:2306.09330

Chen Jingyi , Wang Xiaolong, He Zhijian, Peng Xiaojiang (2024) A comprehensive exploration on detecting fake images generated by stable diffusion. In *Pattern Recognition and Computer Vision - 7th Chinese Conference, PRCV 2024, Urumqi, China, October 18-20, 2024, Proceedings, Part I*, volume 15031 of Lecture Notes in Computer Science, pp 461–475

Chen Songyan, Huang Jiancheng (2023) Fec: Three finetuning-free methods to enhance consistency for real image editing. In 2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML), pp 76–87

Chen Zheng, Zhang Yulun, Liu Ding, Gu Jinjin, Kong Linghe, Yuan Xin et al (2024) Hierarchical integration diffusion model for realistic image deblurring. Adv Neural Inform Process Syst 36

Cheng Lu, Kaiwen Zheng, Fan Bao, Jianfei Chen, Chongxuan Li, Jun Zhu (2022) Maximum likelihood training for score-based diffusion odes by high order denoising score matching. In International Conference on Machine Learning vol 162, pp 14429–14460

Cheng Ta-Ying, Gadelha Matheus, Groueix Thibault, Fisher Matthew, Mech Radomir, Markham Andrew, Trigoni Niki (2024) Learning continuous 3d words for text-to-image generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 6753–6762

Chenlu Zhan Yu, Lin Gaoang Wang, Wang Hongwei, Jian Wu (2024) Medm2g: Unifying medical multimodal generation via cross-guided diffusion with visual invariant. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 11502–11512

Chin Zhi-Yi, Jiang Chieh-Ming, Huang Ching-Chun, Chen Pin-Yu, Chiu Wei-Chen (2023) Prompting4debugging: Red-teaming text-to-image diffusion models by finding problematic prompts. Preprint at arXiv:2309.06135

Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, Fleet David J, Mohammad Norouzi (2022) Image super-resolution via iterative refinement. IEEE Trans Pattern Anal Mach Intell 45(4):4713–4726

Choi Jooyoung, Choi Yunjey, Kim Yunji, Kim Junho, Yoon Sungroh (2023) Custom-edit: Text-guided image editing with customized diffusion models. Preprint at arXiv:2305.15779

Christopher Jarzynski (1997) Equilibrium free-energy differences from nonequilibrium measurements: a master-equation approach. Phys Rev E 56(5):5018

Chung Jiwoo, Hyun Sangeek, Heo Jae-Pil (2023) Style injection in diffusion: A training-free approach for adapting large-scale diffusion models for style transfer. Preprint at arXiv:2312.09008

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Liu Peter J (2020) Exploring the limits of transfer learning with a unified text-to-text transformer. J Mach Learn Res 21(140):1–67

De Simone Zoe, Boggust Angie, Satyanarayan Arvind, Wilson Ashia (2023) What is a fair diffusion model? designing generative text-to-image models to incorporate various worldviews. Preprint at arXiv:2309.09944

Deng Yingying, He Xiangyu, Tang Fan, Dong Weiming (2023) Z^*: Zero-shot style transfer via attention rearrangement. Preprint at arXiv:2311.16491

Deng-Ping Fan, Ge-Peng Ji, Ming-Ming Cheng, Ling Shao (2022) Concealed object detection. IEEE Trans Pattern Anal Mach Intell 44(10):6024–6042

Dhariwal Prafulla, Nichol Alexander Quinn (2021) Diffusion models beat gans on image synthesis. In Advances in Neural Information Processing Systems, pp 8780–8794

Dmitry Baranchuk, Andrey Voynov, Ivan Rubachev, Valentin Khrulkov, Artem Babenko (2022) Label-efficient semantic segmentation with diffusion models

Dong Wenkai, Xue Song, Duan Xiaoyue, Han Shumin (2023) Prompt tuning inversion for text-driven image editing using diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 7430–7440

Dubinski Jan, Kowalczuk Antoni, Pawlak Stanisaw, Rokita Przemyslaw, Trzciski Tomasz, Morawiecki Pawe (2024) Towards more realistic membership inference attacks on large diffusion models. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp 4860–4869

Dumoulin Vincent, Shlens Jonathon, Kudlur Manjunath (2017) A learned representation for artistic style. In 5th International Conference on Learning Representations

Esser Patrick, Kulal Sumith, Blattmann Andreas, Entezari Rahim,üller Jonas M, Saini Harry, Levi Yam, Lorenz Dominik, Sauer Axel, Boesel Frederic, Podell Dustin, Dockhorn Tim, English Zion, Rombach Robin (2024) 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

Evgin Goceri (2023) Comparison of the impacts of dermoscopy image augmentation methods on skin cancer classification and a new augmentation method with wavelet packets. Int J Imaging Syst Technol 33(5):1727–1744

Evgin Goceri (2024) GAN based augmentation using a hybrid loss function for dermoscopy images. Artif Intell Rev 57(9):234

Fan Bao, Chongxuan Li, Jun Zhu, Bo Zhang (2022) Analytic-DPM: an analytic estimate of the optimal reverse variance in diffusion probabilistic models

Fan Deng-Ping, Ji Ge-Peng, Xu Peng, Cheng Ming-Ming, Sakaridis Christos, Gool Luc Van (2023) Advances in deep concealed scene understanding. Vis Intell 1(1)

Fan Mingyuan, Wang Chengyu, Chen Cen, Liu Yang, Huang Jun (2024) On the trustworthiness landscape of state-of-the-art generative models: a survey and outlook. Preprint at arXiv:2307.16680

Fan Yuheng, Liao Hanxi, Huang Shiqi, Luo Yimin, Fu Huazhu, Qi Haikun (2023) A survey of emerging applications of diffusion probabilistic models in MRI. Preprint at arXiv:2311.11383

Florinel-Alin Croitoru, Vlad Hondru, Tudor Ionescu Radu, Mubarak Shah (2023) Diffusion models in vision: a survey. IEEE Trans Pattern Anal Mach Intell 45(9):10850–10869

Florinel-Alin Croitoru, Vlad Hondru, Tudor Ionescu Radu, Mubarak Shah (2023) Diffusion models in vision: a survey. IEEE Trans Pattern Anal Mach Intell 45(9):10850–10869

Francesco Tassone, Luca Maiano, Irene Amerini (2024) Continuous fake media detection: adapting deepfake detectors to new generative techniques. Comput Vis Image Underst 249:104143

Friedrich Felix, Brack Manuel, Struppek Lukas, Hintersdorf Dominik, Schramowski Patrick, Luccioni Sasha, Kersting Kristian (2023) Fair diffusion: Instructing text-to-image generation models on fairness. Preprint at arXiv:2302.10893

Gandikota Kanchana Vaishnavi,Chandramouli Paramanand (2024) Text-guided explorable image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 25900–25911

Gandikota Rohit, Orgad Hadas, Belinkov Yonatan, Materzyńska Joanna, Bau David (2024) Unified concept editing in diffusion models. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp 5111–5120

Ganguli Deep, Hernandez Danny, Lovitt Liane, Askell Amanda, Bai Yuntao, Chen Anna, Conerly Tom, DasSarma Nova, Drain Dawn, Elhage Nelson, Showk Sheer El, Fort Stanislav, Hatfield-Dodds Zac, Henighan Tom, Johnston Scott, Jones Andy, Joseph Nicholas, Kernian Jackson, Kravec Shauna, Mann Ben, Nanda Neel, Ndousse Kamal, Olsson Catherine, Amodei Daniela, Brown Tom B., Kaplan Jared, McCandlish Sam, Olah Christopher, Amodei Dario, Clark Jack (2022) Predictability and surprise in large generative models. In FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, June 21 - 24, 2022, pp 1747–1764

Gao Sicheng, Liu Xuhui, Zeng Bohan, Xu Sheng, Li Yanjing, Luo Xiaoyan, Liu Jianzhuang, Zhen Xiantong, Zhang Baochang (2023) Implicit diffusion models for continuous super-resolution. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 10021–10030

Gatys Leon A, Ecker Alexander S, Bethge Matthias (c) A neural algorithm of artistic style. Preprint at arXiv:1508.06576

George Wolberg (1998) Image morphing: a survey. Vis Comput 14(8/9):360–372

George Papamakarios, Nalisnick Eric T, Jimenez Rezende Danilo, Shakir Mohamed, Balaji Lakshminarayanan (2021) Normalizing flows for probabilistic modeling and inference. J Mach Learn Res 57:1–57

Giorgio Parisi (1981) Correlation functions and computer simulations. Nucl Phys B 180(3):378–384

Goodfellow Ian J, Pouget-Abadie Jean, Mirza Mehdi, Xu Bing, Warde-Farley David, Ozair Sherjil, Courville Aaron C, Bengio Yoshua (2014) Generative adversarial nets. In Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems, pp 2672–2680

Grenander Ulf, Miller Michael I (1994) Representations of knowledge in complex systems. J Roy Stat Soc: Ser B (Methodol) 56(4):549–581

Gu Jing, Wang Yilin, Zhao Nanxuan, Fu Tsu-Jui, Xiong Wei, Liu Qing, Zhang Zhifei, Zhang He, Zhang Jianming, Jung HyunJoon et al (2024) Photoswap: Personalized subject swapping in images. Adv Neural Inform Process Sys 36

Gu Xiangming, Du Chao, Pang Tianyu, Li Chongxuan, Lin Min, Wang Ye (2024) On memorization in diffusion models. Preprint at arXiv:2310.02664

Guanhua Zhang, Jiabao Ji, Yang Zhang, Mo Yu, Jaakkola Tommi S, Shiyu Chang (2023) Towards coherent image inpainting using denoising diffusion implicit models. In International Conference on Machine Learning vol 202, pp 41164–41193

Han Ligong, Wen Song, Chen Qi, Zhang Zhixing, Song Kunpeng, Ren Mengwei, Gao Ruijiang, Stathopoulos Anastasis, He Xiaoxiao, Chen Yuxiao et al (2024) Proxedit: Improving tuning-free real image editing with proximal guidance. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp 4291–4301

Hanqun Cao, Cheng Tan, Zhangyang Gao, Yilun Xu, Guangyong Chen, Pheng-Ann Heng, Stan Z. Li. (2024) A survey on generative diffusion models. IEEE Trans Knowl Data Eng 36(7):2814–2830

Haoying Li, Yifan Yang, Meng Chang, Shiqi Chen, Huajun Feng, Zhihai Xu, Qi Li, Yueting Chen (2022) SRDIFF: Single image super-resolution with diffusion probabilistic models. Neurocomputing 479:47–59

Heng Alvin, Soh Harold (2024) Selective amnesia: A continual learning approach to forgetting in deep generative models. Adv Neural Inform Process Syst 36

Higgins Irina, Matthey Loïc, Pal Arka, Burgess Christopher P., Glorot Xavier, Botvinick Matthew M, Mohamed Shakir, Lerchner Alexander (2017) beta-vae: Learning basic visual concepts with a constrained variational framework. In 5th International Conference on Learning Representations

Ho Jonathan, Salimans Tim (2022) Classifier-free diffusion guidance. Preprint at arXiv:2207.12598

Ho Jonathan, Salimans Tim (2022) Classifier-free diffusion guidance. Preprint at arXiv:2207.12598

Hong Seunghoo, Lee Juhun, Woo Simon S (2024) All but one: Surgical concept erasing with model preservation in text-to-image diffusion models. In Proceedings of the AAAI Conference on Artificial Intelligence 38:21143–21151

Hongkai Zheng, Weili Nie, Arash Vahdat, Kamyar Azizzadenesheli, Anima Anandkumar (2023) Fast sampling of diffusion models via operator learning. In International Conference on Machine Learning vol 202, pp 42390–42402

Hu Edward J, Shen Yelong, Wallis Phillip, Allen-Zhu Zeyuan, Li Yuanzhi, Wang Shean, Wang Lu, Chen Weizhu (2022) Lora: Low-rank adaptation of large language models. In The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022

Hu Hailong, Pang Jun (2023) Loss and likelihood based membership inference of diffusion models. In International Conference on Information Security, pp 121–141

Huang Chin-Wei, Lim Jae Hyun, Aaron C (2021) Courville. A variational perspective on diffusion-based generative models and score matching. In Advances in Neural Information Processing Systems, pp 22863–22876

Huang Jiancheng, Liu Yifan, Qin Jin, Chen Shifeng (2023). Kv inversion: Kv embeddings learning for text-conditioned real image action editing. In Chinese Conference on Pattern Recognition and Computer Vision (PRCV), pp 172–184. Springer

Huang Nisha, Zhang Yuxin, Tang Fan, Ma Chongyang, Huang Haibin, Zhang Yong, Dong Weiming, Xu Changsheng (2022) Diffstyler: Controllable dual diffusion for text-driven image stylization. Preprint at arXiv:2211.10682

Hudson Drew A, Zoran Daniel, Malinowski Mateusz, Lampinen Andrew K, Jaegle Andrew, McClelland James L, Matthey Loic, Hill Felix, Lerchner Alexander (2024) Soda: Bottleneck diffusion models for representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 23115–23127

Hyvarinen Aapo, Dayan Peter (2005) Estimation of non-normalized statistical models by score matching. J Mach Learn Res 6(4)

Imagen 3 Team: Jason Baldridge, Jakob Bauer, Mukul Bhutani, and et. al. (2024) Imagen 3. Preprint at arXiv:2408.07009

Jia-Hao Wu, Tsai Fu-Jen, Peng Yan-Tsung, Tsai Chung-Chi, Lin Chia-Wen, Lin Yen-Yu (2024) Id-blau: Image deblurring by implicit diffusion-based reblurring augmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 25847–25856

Jiang Yue, Lyu Yueming, Ma Tianxiang, Peng Bo, Dong Jing (2023) Rs-corrector: Correcting the racial stereotypes in latent diffusion models. Preprint at arXiv:2312.04810

Jiayi Liao Xu, Chen Qiang Fu, Lun Du, He Xiangnan, Wang Xiang, Han Shi, Zhang Dongmei (2024) Text-to-image generation for abstract concepts. In Thirty-Eighth AAAI Conference on Artificial Intelligence, pp 3360–3368

Jolicoeur-Martineau Alexia, Piche-Taillefer Remi, Tachet des Combes Remi, Mitliagkas Ioannis (2020) Adversarial score matching and improved sampling for image generation. Preprint at arXiv:2009.05475

Jonathan Ho, Ajay Jain, Pieter Abbeel (2020) Denoising diffusion probabilistic models. Adv Neural Inf Process Syst 33:6840–6851

Ju Xuan, Zeng Ailing, Zhao Chenchen, Wang Jianan, Zhang Lei, Xu Qiang (2023) Humansd: A native skeleton-guided diffusion model for human image generation. In IEEE/CVF International Conference on Computer Vision, pp 15942–15952

Justin Johnson, Alexandre Alahi, Li Fei-Fei (2016) Perceptual losses for real-time style transfer and super-resolution. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, Computer Vision - ECCV 2016–14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016. Proceedings, Part II vol 9906, pp 694–711

Kaiwen Zheng, Cheng Lu, Jianfei Chen, Jun Zhu (2023) Improved techniques for maximum likelihood estimation for diffusion odes. In International Conference on Machine Learning vol 202, pp 42363–42389

Ke Bingxin, Obukhov Anton, Huang Shengyu, Metzger Nando, Daudt Rodrigo Caye, Schindler Konrad (2024) Repurposing diffusion-based image generators for monocular depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 9492–9502

Kim Sunwoo, Jang Wooseok, Kim Hyunsu, Kim Junho, Choi Yunjey, Kim Seungryong, Lee Gayeong (2023) User-friendly image editing with minimal text input: Leveraging captioning and injection techniques. Preprint at arXiv:2306.02717,

Kingma Diederik P, Salimans Tim, Poole Ben, Ho Jonathan (2021) Variational diffusion models. Preprint at arXiv:2107.00630

Kingma Diederik P, Welling Max (2014) Auto-encoding variational bayes. In 2nd International Conference on Learning Representations

Kingma Diederik P, Welling Max (2014) Auto-encoding variational bayes. In 2nd International Conference on Learning Representations

Kocsis Peter, Sitzmann Vincent, Nießner Matthias (2024) Intrinsic image diffusion for indoor single-view material estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 5198–5208

Kumari Nupur, Zhang Bingliang, Wang Sheng-Yu, Shechtman Eli, Zhang Richard, Zhu Jun-Yan (2023) Ablating concepts in text-to-image diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 22691–22702

Lee Jumin, Im Woobin, Lee Sebin, Yoon Sung-Eui (2023) Diffusion probabilistic models for scene-scale 3d categorical data. Preprint at arXiv:2301.00527

Lequan Lin, Zhengkun Li, Ruikun Li, Xuliang Li, Junbin Gao (2024) Diffusion models for time-series applications: a survey. Front Inf Technol Electron Eng 25(1):19–41

Li Jia, Hu Lijie, Zhang Jingfeng, Zheng Tianhang, Zhang Hua, Wang Di (2023) Fair text-to-image diffusion via fair mapping. Preprint at arXiv:2311.17695

Li Shengmeng, Liu Luping, Chai Zenghao, Li Runnan, Tan Xu (2023) Era-solver: Error-robust adams solver for fast sampling of diffusion probabilistic models. Preprint at arXiv:2301.12935

Li Sifei, Zhang Yuxin, Tang Fan, Ma Chongyang, Dong Weiming, Xu Changsheng (2024) Music style transfer with time-varying inversion of diffusion models. In Thirty-Eighth AAAI Conference on Artificial Intelligence, pp 547–555

Li Tianle, Ku Max, Wei Cong, Chen Wenhu (2023) Dreamedit: Subject-driven image editing. Preprint at arXiv:2306.12624

Li Xin, Ren Yulin, Jin Xin, Lan Cuiling, Wang Xingrui, Zeng Wenjun, Wang Xinchao, Chen Zhibo (2023) Diffusion models for image restoration and enhancement—a comprehensive survey. Preprint at arXiv:2308.09388

Li Xinghui, Jingyi Lu, Han Kai, Prisacariu Victor Adrian (2024) Sd4match: Learning to prompt stable diffusion model for semantic matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 27558–27568

Liang Youwei, He Junfeng, Li Gang, Li Peizhao, Klimovskiy Arseniy, Carolan Nicholas, Sun Jiao, Pont-Tuset Jordi, Young Sarah, Yang Feng, Ke Junjie, Dvijotham Krishnamurthy Dj, Collins Katherine M, Luo Yiwen, Li Yang, Kohlhoff Kai J, Ramachandran Deepak, Navalpakkam Vidhya (2024) Rich human feedback for text-to-image generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 19401–19411

Lin Xinqi, He Jingwen, Chen Ziyan, Lyu Zhaoyang, Fei Ben, Dai Bo, Ouyang Wanli, Qiao Yu, Dong Chao (2023) Diffbir: Towards blind image restoration with generative diffusion prior. Preprint at arXiv:2308.15070

Lin Yuanze, Chen Yi-Wen, Tsai Yi-Hsuan, Jiang Lu, Yang Ming-Hsuan (2023) Text-driven image editing via learnable regions. Preprint at arXiv:2311.16432

Lin Zinan, Gopi Sivakanth, Kulkarni Janardhan, Nori Harsha, Yekhanin Sergey (2023) Differentially private synthetic data via foundation model apis 1: Images. Preprint at arXiv:2305.15560

Ling Pengyang, Chen Lin, Zhang Pan, Chen Huaian, Jin Yi, Zheng Jinjin (2024) Freedrag: Feature dragging for reliable point-based image editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 6860–6870

Ling Yang, Zhilong Zhang, Yang Song, Hong Shenda Xu, Runsheng Zhao Yue, Wentao Zhang, Bin Cui, Ming-Hsuan Yang (2024) Diffusion models: a comprehensive survey of methods and applications. ACM Comput Surv 56(4):105–39

Liu Anji, Niepert Mathias, den Broeck Guy Van (2024) Image inpainting via tractable steering of diffusion models. In The Twelfth International Conference on Learning Representations

Liu Haipeng, Wang Yang, Qian Biao, Wang Meng, Rui Yong (2024) Structure matters: Tackling the semantic discrepancy in diffusion models for image inpainting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 8038–8047

Liu Jiawei, Wang Qiang, Fan Huijie, Wang Yinong, Tang Yandong, Qu Liangqiong (2023) Residual denoising diffusion models. Preprint at arXiv:2308.13712

Liu Jinxiu, Liu Qi (2024) R3CD: scene graph to image generation with relation-aware compositional contrastive control diffusion. In Thirty-Eighth AAAI Conference on Artificial Intelligence, pp 3657–3665

Liu Xian, Ren Jian, Siarohin Aliaksandr, Skorokhodov Ivan, Li Yanyu, Lin Dahua, Liu Xihui, Liu Ziwei, Tulyakov Sergey (2024) Hyperhuman: Hyper-realistic human generation with latent structural diffusion. In The Twelfth International Conference on Learning Representations

Liu Xihui, Park Dong Huk, Azadi Samaneh, Zhang Gong, Chopikyan Arman, Yuxiao Hu, Shi Humphrey, Rohrbach Anna, Darrell Trevor (2023) More control for free! image synthesis with semantic diffusion guidance. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp 289–299

Liu Yixin, Zhang Kai, Li Yuan, Yan Zhiling, Gao Chujie, Chen Ruoxi, Yuan Zhengqing, Huang Yue, Sun Hanchi, Gao Jianfeng, He Lifang, Sun Lichao (2024) Sora: A review on background, technology, limitations, and opportunities of large vision models. Preprint at arXiv:2402.17177

Lu Cheng, Zhou Yuhao, Bao Fan, Chen Jianfei, Li Chongxuan, Zhu Jun (2022) Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. Preprint at arXiv:2211.01095

Luccioni Alexandra Sasha, Akiki Christopher, Mitchell Margaret, Jernite Yacine (2023) Stable bias: Analyzing societal representations in diffusion models. Preprint at arXiv:2303.11408

Lugmayr Andreas, Danelljan Martin, Romero Andres, Fisher Yu, Timofte Radu, Van Gool Luc (2022) Repaint: Inpainting using denoising diffusion probabilistic models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 11461–11471

Luo Calvin (2022) Understanding diffusion models: a unified perspective. Preprint at arXiv:2208.11970

Luo Shitong, Wei Hu (2021) Diffusion probabilistic models for 3D point cloud generation. In IEEE Conference on Computer Vision and Pattern Recognition, pp 2837–2845

Luo Weijian (2023) A comprehensive survey on knowledge distillation of diffusion models. Preprint at arXiv:2304.04262

Luo Xiaotong, Xie Yuan, Yanyun Qu, Yun Fu (2024) Skipdiff: Adaptive skip diffusion model for high-fidelity perceptual image super-resolution. In Thirty-Eighth AAAI Conference on Artificial Intelligence, pp 4017–4025

Luo Ziwei, Gustafsson Fredrik K, Zhao Zheng, Sjölund Jens, Schön Thomas B (2023) Refusion: Enabling large-size realistic image restoration with latent-space diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 1680–1691

Ma Nanye, Goldstein Mark, Albergo Michael S, BoffiNicholas M, Vanden-Eijnden Eric, Xie Saining (2024) Sit: Exploring flow and diffusion-based generative models with scalable interpolant transformers. Preprint at arXiv:2401.08740

Maggiora Gabriel della, Croquevielle Luis Alberto, Deshpande Nikita, Horsley Harry, Heinis Thomas, Yakimovich Artur (2024) Conditional variational diffusion models. In The Twelfth International Conference on Learning Representations

Metzger Nando, Daudt Rodrigo Caye, Schindler Konrad (2023) Guided depth super-resolution by deep anisotropic diffusion. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 18237–18246

Midjourney. https://www.midjourney.com/home

Miyake Daiki, Iohara Akihiro, Saito Yu, Tanaka Toshiyuki (2023) Negative-prompt inversion: Fast image inversion for editing with text-guided diffusion models. Preprint at arXiv:2305.16807

Mohan Bhandari, Arjun Neupane, Saurav Mallik, Loveleen Gaur, Hong Qin (2023) Auguring fake face images using dual input convolution neural network. J Imaging 9(1):3

Mokady Ron, Hertz Amir, Aberman Kfir, Pritch Yael, Cohen-Or Daniel (2023) Null-text inversion for editing real images using guided diffusion models. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 6038–6047

Moser Brian B, Shanbhag Arundhati S, Raue Federico, Frolov Stanislav, Palacio Sebastian, Dengel Andreas (2024) Diffusion models, image super-resolution and everything: A survey. IEEE Transactions on Neural Networks and Learning Systems, pp 1–21

Müller Vincent C, Bostrom Nick (2013) Future progress in artificial intelligence: A survey of expert opinion. In Vincent C. Müller, (ed) Fundamental Issues of Artificial Intelligence - 2nd Conference on Philosophy and Theory of Artificial Intelligence, PT-AI 2013, Oxford, UK, September 21-22, 2013, selected and invited papers, vol 376 of Synthese Library, pp 555–572

Naik Ranjita, Nushi Besmira (2023) Social biases through the text-to-image generation lens. In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society, pp 786–808

Nash Robert A, Wade Kimberley A, Brewer Rebecca J (2009) Why do doctored images distort memory? Conscious Cogn 18(3):773–780

Neal Radford M (2001) Annealed importance sampling. Stat Comput 11:125–139

Neves João C, Ruben Tolosana, Rubén Vera-Rodríguez, Vasco Lopes, Hugo Proença, Julian Fiérrez (2020) Ganprintr: Improved fakes and evaluation of the state of the art in face manipulation detection. IEEE J Sel Top Signal Process 14(5):1038–1048

Nguyen Thao, Li Yuheng, Ojha Utkarsh, Lee Yong Jae (2023) Visual instruction inversion: Image editing via visual prompting. Preprint at arXiv:2307.14331

Nguyen Thao, Ojha Utkarsh, Li Yuheng, Liu Haotian, Lee Yong Jae (2024) Edit one for all: Interactive batch image editing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 8271–8280

Ni Zixuan, Wei Longhui, Li Jiacheng, Tang Siliang, Zhuang Yueting, Tian Qi (2023) Degeneration-tuning: Using scrambled grid shield unwanted concepts from stable diffusion. In Proceedings of the 31st ACM International Conference on Multimedia, pp 8900–8909

Nichol Alexander Quinn, Dhariwal Prafulla (2022) Improved denoising diffusion probabilistic models. In Proceedings of the 38th International Conference on Machine Learning, vol 139, pp 8162–8171

Nisha Huang, Yuxin Zhang, Weiming Dong (2024) Style-a-video: Agile diffusion for arbitrary text-based video style transfer. IEEE Signal Process Lett 31:1494–1498

Nunes Lucas, Marcuzzi Rodrigo, Mersch Benedikt, Behley Jens, Stachniss Cyrill (2024) Scaling diffusion models to real-world 3d lidar scene completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 14770–14780

Olaf Ronneberger, Philipp Fischer, Thomas Brox (2015) U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention vol 9351, pp 234–241

OpenAI (2023) GPT-4 technical report. Preprint at arXiv:2303.08774

OpenAI (2023) Video generation models as world simulators. https://openai.com/index/video-generation-models-as-world-simulators/

Pan Zhihong, Zhou Xin, Tian Hao (2023) Arbitrary style guidance for enhanced diffusion-based text-to-image generation. In IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2023, Waikoloa, HI, USA, January 2-7, 2023, pp 4450–4460

Pang Lianyu, Yin Jian, Xie Haoran, Wang Qiping, Li Qing, Mao Xudong (2023) Cross initialization for personalized text-to-image generation. Preprint at arXiv:2312.15905

Parmar Gaurav, Zhang Richard, Zhu Jun-Yan (2022) On aliased resizing and surprising subtleties in GAN evaluation. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, pp 11400–11410

Pascal Vincent (2011) A connection between score matching and denoising autoencoders. Neural Comput 23(7):1661–1674

Patni Suraj, Agarwal Aradhye, Arora Chetan (2024) Ecodepth: Effective conditioning of diffusion models for monocular depth estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June, pp 28285–28295

Peebles William, Xie Saining (2023) Scalable diffusion models with transformers. In IEEE/CVF International Conference on Computer Vision, pp 4172–4182

Pham Minh, Marshall Kelly O, Cohen Niv, Mittal Govind, Hegde Chinmay (2023) Circumventing concept erasure methods for text-to-image generative models. In The Twelfth International Conference on Learning Representations

PNVR Koutilya, Singh Bharat, Ghosh Pallabi, Siddiquie Behjat, Jacobs David (2023) Ld-znet: A latent diffusion approach for text-based image segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp 4157–4168

Po R, Yifan W, Golyanik V, Aberman K, Barron JT, Bermano A, Chan E, Dekel T, Holynski A, Kanazawa A, Liu CK, Liu L, Mildenhall B, NieÃŷner M, Ommer B, Theobalt C, Wonka P, Wetzstein G (2024) State of the art on diffusion models for visual computing. Comput Graph Forum 43(2):e15063

Prafulla Dhariwal, Alexander Nichol (2021) Diffusion models beat GANS on image synthesis. Adv Neural Inf Process Syst 34:8780–8794

Qi Tianhao, Fang Shancheng, Wu Yanze, Xie Hongtao, Liu Jiawei, Chen Lang, He Qian, Zhang Yongdong (2024) Deadiff: An efficient stylization diffusion model with disentangled representations. Preprint at arXiv:2403.06951

Qiang Xu, Hao Wang, Laijin Meng, Zhongjie Mi, Jianye Yuan, Hong Yan (2023) Exposing fake images generated by text-to-image diffusion models. Pattern Recognit Lett 176:76–82

Qingping Zheng, Ling Zheng, Yuanfan Guo, Ying Li, Songcen Xu, Jiankang Deng, Hang Xu (2024) Self-adaptive reality-guided diffusion for artifact-free super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 25806–25816

Qinsheng Zhang, Yongxin Chen (2023) Fast sampling of diffusion models with exponential integrator

Qiu Xinmin, Han Congying, Zhang Zicheng, Li Bonan, Guo Tiande, Nie Xuecheng (2023) Diffbfr: Bootstrapping diffusion model for blind face restoration. In Proceedings of the 31st ACM International Conference on Multimedia, pp 7785–7795

Qu Yiting, Shen Xinyue, He Xinlei, Backes Michael, Zannettou Savvas, Zhang Yang (2023) Unsafe diffusion: On the generation of unsafe images and hateful memes from text-to-image models. In Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security, pp 3403–3417

Quinn Nichol Alexander, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, Mark Chen (2022) GLIDE: towards photorealistic image generation and editing with text-guided diffusion models. In International Conference on Machine Learning vol 162, pp 16784–16804

Radford Alec, Kim Jong Wook, Hallacy Chris, Ramesh Aditya, Goh Gabriel, Agarwal Sandhini, Sastry Girish, Askell Amanda, Mishkin Pamela, Clark Jack, Krueger Gretchen, Sutskever Ilya (2021) Learning transferable visual models from natural language supervision. In Proceedings of the 38th International Conference on Machine Learning, vol 139, pp 8748–8763

Ramesh Aditya, Dhariwal Prafulla, Nichol Alex, Chu Casey, Chen Mark (2022) Hierarchical text-conditional image generation with CLIP latents. Preprint at arXiv:2204.06125

Ramesh Aditya, Dhariwal Prafulla, Nichol Alex, Chu Casey, Chen Mark (2022) Hierarchical text-conditional image generation with clip latents. Preprint at arXiv:2204.06125

Ramesh Aditya, Pavlov Mikhail, Goh Gabriel, Gray Scott, Voss Chelsea, Radford Alec, Chen Mark, Sutskever Ilya (2021) Zero-shot text-to-image generation. In Marina Meila and Tong Zhang (eds), Proceedings of the 38th International Conference on Machine Learning, vol 139, pp 8821–8831

Rando Javier, Paleka Daniel, Lindner David, Heim Lennart, Tramèr Florian (2022) Red-teaming the stable diffusion safety filter. Preprint at arXiv:2210.04610

Reed Scott E, Akata Zeynep, Yan Xinchen, Logeswaran Lajanugen, Schiele Bernt, Lee Honglak (2016) Generative adversarial text to image synthesis. In Proceedings of the 33nd International Conference on Machine Learning, vol 48, pp 1060–1069

Ren Mengwei, Delbracio Mauricio, Talebi Hossein, Gerig Guido, Milanfar Peyman (2023) Multiscale structure guided diffusion for image deblurring. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp 10721–10733

Rombach Robin, Blattmann Andreas, Lorenz Dominik, Esser Patrick, Ommer Björn (2022) High-resolution image synthesis with latent diffusion models. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 10674–10685

Saharia Chitwan, Chan William, Saxena Saurabh, Li Lala, Whang Jay, Denton Emily L., Kamyar Seyed Ghasemipour Seyed, Lopes Raphael Gontijo, Ayan Burcu Karagol, Salimans Tim, Ho Jonathan, Fleet David J, Norouzi Mohammad (2022) Photorealistic text-to-image diffusion models with deep language understanding. In Advances in Neural Information Processing Systems

Scene generation with hierarchical latent diffusion models (2023) Seung Wook Kim, Bradley Brown, Kangxue Yin, Karsten Kreis, Katja Schwarz, Daiqing Li, Robin Rombach, Antonio Torralba, and Sanja Fidler. Neuralfield-ldm. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 8496–8506

Schramowski Patrick, Brack Manuel, Deiseroth Bjorn, Kersting Kristian (2023) Safe latent diffusion: Mitigating inappropriate degeneration in diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 22522–22531

Seung-Lee Lee, Minjae Kang, Jong-Uk Hou (2024) Localization of diffusion model-based inpainting through the inter-intra similarity of frequency features. Image Vis Comput 148:105138

Shang Kai, Shao Mingwen, Wang Chao, Cheng Yuanshuo, Wang Shuigen (2024) Multi-domain multi-scale diffusion model for low-light image enhancement. In Thirty-Eighth AAAI Conference on Artificial Intelligence, pp 4722–4730

Shen Xudong, Du Chao, Pang Tianyu, Lin Min, Wong Yongkang, Kankanhalli Mohan (2023) Finetuning text-to-image diffusion models for fairness. Preprint at arXiv:2311.07604

Shen Cuihua, Kasra Mona, Pan Wenjing, Bassett Grace A, Malloch Yining, O'Brien James F (2019) Fake images: The effects of source, intermediary, and digital media literacy on contextual assessment of image credibility online. New Media Soc 21(2):438–463. https://doi.org/10.1177/1461444818799526

Sohl-Dickstein Jascha, Weiss Eric A, Maheswaranathan Niru, Ganguli Surya (2015) Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, vol 37, pp 2256–2265

Song Yang, Ermon Stefano (2019) Generative modeling by estimating gradients of the data distribution. Adv Neural Inform Process Syst 32

Song Yang, Garg Sahaj, Shi Jiaxin, Ermon Stefano (2020) Sliced score matching: A scalable approach to density and score estimation. In Uncertainty in Artificial Intelligence, pp 574–584

Song Yang, Sohl-Dickstein Jascha, Kingma Diederik P, Kumar Abhishek, Ermon Stefano, Poole Ben (2020) Score-based generative modeling through stochastic differential equations. Preprint at arXiv:2011.13456,

Structured prediction for efficient text-to-image generation (2024) Jayasumana, Daniel Glasner, Srikumar Ramalingam, Andreas Veit, Ayan Chakrabarti, and Sanjiv Kumar. Markovgen. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 9316–9325

Suttisak Wizadwongsa, Supasorn Suwajanakorn (2023) Accelerating guided diffusion sampling with splitting numerical methods

Tao Xu, Zhang Pengchuan, Huang Qiuyuan, Zhang Han, Gan Zhe, Huang Xiaolei, He Xiaodong (2018) Attngan: Fine-grained text to image generation with attentional generative adversarial networks. In IEEE Conference on Computer Vision and Pattern Recognition, pp 1316–1324

Tim Salimans, Jonathan Ho (2022) Progressive distillation for fast sampling of diffusion models

Ulhaq Anwaar, Akhtar Naveed, Pogrebna Ganna (2022) Efficient diffusion models for vision: a survey. Preprint at arXiv:2210.09292

Vahdat Arash, Kreis Karsten, Kautz Jan (2021) Score-based generative modeling in latent space. In Advances in Neural Information Processing Systems, pp 11287–11302

van den Oord Aäron, Vinyals Oriol, Kavukcuoglu Koray (2017) Neural discrete representation learning. In Advances in Neural Information Processing Systems, pp 6306–6315

Wang Hanzhang, Wang Haoran, Yang Jinze, Yu Zhongrui, Xie Zeke, Tian Lei, Xiao Xinyan, Jiang Junjun, Liu Xianming, Sun Mingming (2024) Hicast: Highly customized arbitrary style transfer with adapter enhanced diffusion models. Preprint at arXiv:2401.05870

Wang Haonan, Shen Qianli, Tong Yao, Zhang Yang, Kawaguchi Kenji (2024) The stronger the diffusion model, the easier the backdoor: Data poisoning to induce copyright breaches without adjusting finetuning pipeline. Preprint at arXiv:2401.04136

Wang Liyan, Yang Qinyu, Wang Cong, Wang Wei, Pan Jinshan, Su Zhixun (2023) Learning a coarse-to-fine diffusion transformer for image restoration. Preprint at arXiv:2308.08730

Wang Qian, Zhang Biao, Birsak Michael, Wonka Peter (2023) Mdp: A generalized framework for text-guided image editing by manipulating the diffusion path. Preprint at arXiv:2303.16765

Wang Yibin, Zhang Weizhong, Zheng Jianwei, Jin Cheng (2023) High-fidelity person-centric subject-to-image synthesis. Preprint at arXiv:2311.10329

Wang Yufei, Yang Wenhan, Chen Xinyuan, Wang Yaohui, Guo Lanqing, Chau Lap-Pui, Ziwei Liu Yu, Qiao Alex C, Kot, and Bihan Wen. (2024) Sinsr: Diffusion-based image super-resolution in a single step. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 25796–25805

Wang Zhicai, Wei Longhui, Wang Tan, Chen Heyu, Hao Yanbin, Wang Xiang, He Xiangnan, Tian Qi (2024) Enhance image classification via inter-class image mixup with diffusion model. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 17223–17233

Wang Zhizhong, Zhao Lei, Xing Wei (2023) Stylediffusion: Controllable disentangled style transfer via diffusion models. In IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1–6, 2023, pp 7643–7655

Wang Zhizhong, Zhao Lei, Xing Wei (2023) Stylediffusion: Controllable disentangled style transfer via diffusion models. In IEEE/CVF International Conference on Computer Vision, pp 7643–7655

Weng Zhenzhen, Sánchez Laura Bravo, Yeung-Levy Serena (2024) Diffusion-hpc: Synthetic data generation for human mesh recovery in challenging domains. In International Conference on 3D Vision, pp 257–267

Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, Jiwen Lu (2023) Unipc: A unified predictor-corrector framework for fast sampling of diffusion models. In Advances in Neural Information Processing Systems

Xia Bin, Zhang Yulun, Wang Shiyin, Wang Yitong, Xinglong Wu, Tian Yapeng, Yang Wenming, Van Gool Luc (2023) Diffir: Efficient diffusion model for image restoration. In IEEE/CVF International Conference on Computer Vision, pp 13049–13059

Xing Zhen, Feng Qijun, Chen Haoran, Dai Qi, Hu Han, Xu Hang, Wu Zuxuan, Jiang Yu-Gang (2024) A survey on video diffusion models. ACM Comput Surv 57(2)

Xu Juncong, Yang Yang, Fang Han, Liu Honggu, Zhang Weiming (2024) Famsec: A few-shot-sample-based general ai-generated image detection method. Preprint at arXiv:2410.13156

Xu Yilun, Deng Mingyang, Cheng Xiang, Tian Yonglong, Liu Ziming, Jaakkola Tommi S (2023) Restart sampling for improving generative processes. In Advances in Neural Information Processing Systems

Xun Huang, Belongie Serge J (2017) Arbitrary style transfer in real-time with adaptive instance normalization. In IEEE International Conference on Computer Vision, pp 1510–1519

Xunpeng Huang, Difan Zou, Hanze Dong, Yi-An Ma, Tong Zhang (2024) Faster sampling without isoperimetry via diffusion-based Monte Carlo. In The Thirty Seventh Annual Conference on Learning Theory 247:2438–2493

Yang Binbin, Luo Yi, Chen Ziliang, Wang Guangrun, Liang Xiaodan, Lin Liang (2023) Law-diffusion: Complex scene generation by diffusion with layouts. In IEEE/CVF International Conference on Computer Vision, pp 22612–22622

Yang Jingyuan, Feng Jiawei, Huang Hui (2024) Emogen: Emotional image content generation with text-to-image diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 6358–6368

Yang Serin, Hwang Hyunmin, Ye Jong Chul (2023) Zero-shot contrastive loss for text-guided diffusion image style transfer. In IEEE/CVF International Conference on Computer Vision, pp 22816–22825

Yang Song, Stefano Ermon (2020) Improved techniques for training score-based generative models. Adv Neural Inf Process Syst 33:12438–12448

Yang Song, Conor Durkan, Iain Murray, Stefano Ermon (2021) Maximum likelihood training of score-based diffusion models. Adv Neural Inf Process Syst 34:1415–1428

Yann LeCun, Yoshua Bengio, Hinton Geoffrey E (2015) Deep learning. Nat 521(7553):436–444

Yarom Michal, Bitton Yonatan, Changpinyo Soravit, Aharoni Roee, Herzig Jonathan, Lang Oran, Ofek Eran, Szpektor Idan (2023) What you see is what you read? improving text-image alignment evaluation. 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

Yeh Yu-Ying, Huang Jia-Bin, Kim Changil, Xiao Lei, Nguyen-Phuoc Thu, Khan Numair, Zhang Cheng, Chandraker Manmohan, Marshall Carl S, Dong Zhao, Li Zhengqin (2024) Texturedreamer: Image-guided texture synthesis through geometry-aware diffusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 4304–4314

Yi Xunpeng, Han Xu, Zhang Hao, Tang Linfeng, Ma Jiayi (2023) Diff-retinex: Rethinking low-light image enhancement with A generative diffusion model. In IEEE/CVF International Conference on Computer Vision, pp 12268–12277

Yilun Xu, Ziming Liu, Yonglong Tian, Shangyuan Tong, Max Tegmark, Jaakkola Tommi S (2023) PFGM++: unlocking the potential of physics-inspired generative models. In International Conference on Machine Learning vol 202, pp 38566–38591

Yiyang Ma, Huan Yang, Yang Wenhan Fu, Jianlong Liu Jiaying (2024) Solving diffusion odes with optimal boundary conditions for better image super-resolution

Yoshua Bengio, Yann LeCun, Hinton Geoffrey E (2021) Deep learning for AI. Commun ACM 64(7):58–65

Yu Jiahui Xu, Yuanzhong Koh Jing, Yu Luong Thang, Gunjan Baid, Zirui Wang, Vasudevan Vijay Ku, Alexander Yang Yinfei, Karagol Ayan Burcu, Ben Hutchinson, Wei Han, Zarana Parekh, Xin Li, Han Zhang, Baldridge Jason Wu, Yonghui, (2022) Scaling autoregressive models for content-rich text-to-image generation. Trans Mach Learn Res 2022

Yuheng Fan, Hanxi Liao, Shiqi Huang, Yimin Luo, Huazhu Fu, Haikun Qi (2024) A survey of emerging applications of diffusion probabilistic models in MRI. Meta-Radiol 2(2):100082

Zeng Haijin, Cao Jiezhang, Zhang Kai, Chen Yongyong, Luong Hiep, Philips Wilfried (2024) Unmixing diffusion for self-supervised hyperspectral image denoising. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 27820–27830

Zhang Chenshuang, Zhang Chaoning, Zhang Mengchun, Kweon In So (2023) Text-to-image diffusion models in generative AI: A survey. Preprint at arXiv:2303.07909

Zhang Chenshuang, Zhang Chaoning, Zhang Mengchun, Kweon In So (2023) Text-to-image diffusion models in generative AI: A survey. Preprint at arXiv:2303.07909

Zhang Chenshuang, Zhang Chaoning, Zheng Sheng, Zhang Mengchun, Qamar Maryam, Bae Sung-Ho, Kweon In So (2023) A survey on audio diffusion models: Text to speech synthesis and enhancement in generative AI. Preprint at arXiv:2303.13336

Zhang Eric, Wang Kai, Xu Xingqian, Wang Zhangyang, Shi Humphrey (2023) Forget-me-not: Learning to forget in text-to-image diffusion models. Preprint at arXiv:2303.17591

Zhang Han, Xu Tao, Li Hongsheng (2017) Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *IEEE International Conference on Computer Vision*, pp 5908–5916

Zhang Jie, Kerschbaum Florian, Zhang Tianwei, et al (2023) Backdooring textual inversion for concept censorship. Preprint at arXiv:2308.10718

Zhang Kaiwen, Zhou Yifan, Xudong Xu, Dai Bo, Pan Xingang (2024) Diffmorpher: Unleashing the capability of diffusion models for image morphing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 7912–7921

Zhang Xinyi, Li Naiqi, Li Jiawei, Dai Tao, Jiang Yong, Xia Shu-Tao (2023) Unsupervised surface anomaly detection with diffusion probabilistic model. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp 6782–6791

Zhang Yang, Tzun Teoh Tze, Hern Lim Wei, Wang Haonan, Kawaguchi Kenji (2023) Investigating copyright issues of diffusion models under practical scenarios. Preprint at arXiv:2311.12803

Zhang Yimeng, Jia Jinghan, Chen Xin, Chen Aochuan, Zhang Yihua, Liu Jiancheng, Ding Ke, Liu Sijia (2023) To generate or not? safety-driven unlearned diffusion models are still easy to generate unsafe images... for now. Preprint at arXiv:2310.11868

Zhang Yuxin, Huang Nisha, Tang Fan, Huang Haibin, Ma Chongyang, Dong Weiming, Xu Changsheng (2023) Inversion-based style transfer with diffusion models. In IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17–24, 2023, pp 10146–10156

Zhao Kai, Hung Alex Ling Yu, Pang Kaifeng, Zheng Haoxin, Sung Kyunghyun (2023) Partdiff: Image super-resolution with partial diffusion models. Preprint at arXiv:2307.11926

Zhao Pancheng, Peng Xu, Qin Pengda, Fan Deng-Ping, Zhang Zhicheng, Jia Guoli, Zhou Bowen, Yang Jufeng (2024) Lake-red: Camouflaged images generation by latent background knowledge retrieval-augmented diffusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp 4092–4101

Zhaoyang Lyu, Kong Zhifeng Xu, Xudong Pan Liang, Dahua Lin (2022) A conditional point diffusion-refinement paradigm for 3D point cloud completion

Zheng Sizhe, Gao Pan, Zhou Peng, Qin Jie (2024) Puff-net: Efficient style transfer with pure content and style feature fusion network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 8059–8068

Zhou Jinxin, Ding Tianyu, Chen Tianyi, Jiang Jiachen, Zharkov Ilya, Zhu Zhihui, Liang Luming (2023) Dream: Diffusion rectification and estimation-adaptive models. Preprint at arXiv:2312.00210

Zhou Tianhao, Li Haipeng, Wang Ziyi, Luo Ao, Zhang Chen-Lin, Li Jiajun, Zeng Bing, Liu Shuaicheng (2024) Recdiffusion: Rectangling for image stitching with diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June, p 2692–2701

Zhu Derui, Chen Dingfan, Grossklags Jens, Fritz Mario (2023) Data forensics in diffusion models: A systematic analysis of membership privacy. Preprint at arXiv:2302.07801

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Hang Chen[1] · Qian Xiang[2,3,4] · Jiaxin Hu[1] · Meilin Ye[1] · Chao Yu[1] · Hao Cheng[1] · Lei Zhang[1]

✉ Lei Zhang
zhanglei@hbpu.edu.cn

[1] School of Electrical and Electronic Information Engineering, Hubei Polytechnic University, Huangshi 435003, China

[2] Wuchang Shouyi University College Information Science and Engineering, Wuhan 430064, China

[3] GongQing Institute of Science and Technology, Jiujiang 332020, China

[4] Wuhan Nanhua Industrial Equipments Engineering Co.,Ltd, Wuhan 430200, China