Diffusion Models and Their Role in Enhancing Computer Vision: A Survey

www.surveyx.cn

Abstract

Diffusion models have emerged as transformative tools in computer vision, advancing generative modeling and representation learning across diverse applications. Their ability to generate high-fidelity samples through iterative denoising processes positions them as pivotal in synthesizing high-quality images and enhancing the diversity and quality of training datasets. This survey explores the multifaceted applications of diffusion models, including image generation, data augmentation, and object detection enhancement. By leveraging sophisticated probabilistic frameworks, diffusion models address challenges such as data scarcity and imbalance, improving model robustness and accuracy. The survey highlights significant advancements in medical imaging, where diffusion models enhance diagnostic precision and anomaly detection. In autonomous driving, these models contribute to improved perception and decision-making, while in creative industries, they facilitate innovative image manipulation and video generation. Despite their strengths, diffusion models face challenges in computational efficiency and robustness against adversarial attacks. Ongoing research aims to optimize these models for broader applications, ensuring ethical deployment and addressing potential biases. The survey underscores the transformative potential of diffusion models in advancing computer vision, emphasizing their role in generating high-quality data and enhancing diverse tasks. Future research should focus on refining diffusion processes, developing generalized frameworks, and exploring their integration with other AI advancements to unlock further potential across various domains.

1 Introduction

1.1 Significance of Diffusion Models in Computer Vision

Diffusion models have revolutionized computer vision, enhancing generative modeling and representation learning across a range of applications [1]. Notably, denoising diffusion probabilistic models (DDPMs) excel in generating high-fidelity samples through iterative denoising, showcasing their ability to synthesize quality images. Their flexibility in addressing complex tasks establishes them as foundational elements in both theoretical and applied research.

Beyond image synthesis, diffusion models play a crucial role in generating high-quality labeled datasets, essential for training robust machine learning models. By improving dataset diversity and quality, these models advance various computer vision tasks, bridging existing knowledge gaps and underscoring the importance of generative tasks. Their capabilities extend to generating synthetic media with high photorealism, raising challenges for media forensics in distinguishing real from AI-generated images [2].

Moreover, diffusion models have achieved state-of-the-art results in generative learning and unsupervised anomaly detection, particularly in identifying pixel-level anomalies in medical imaging [3]. However, further research is required to assess their robustness against out-of-distribution data and address potential biases, ensuring ethical applications [4].

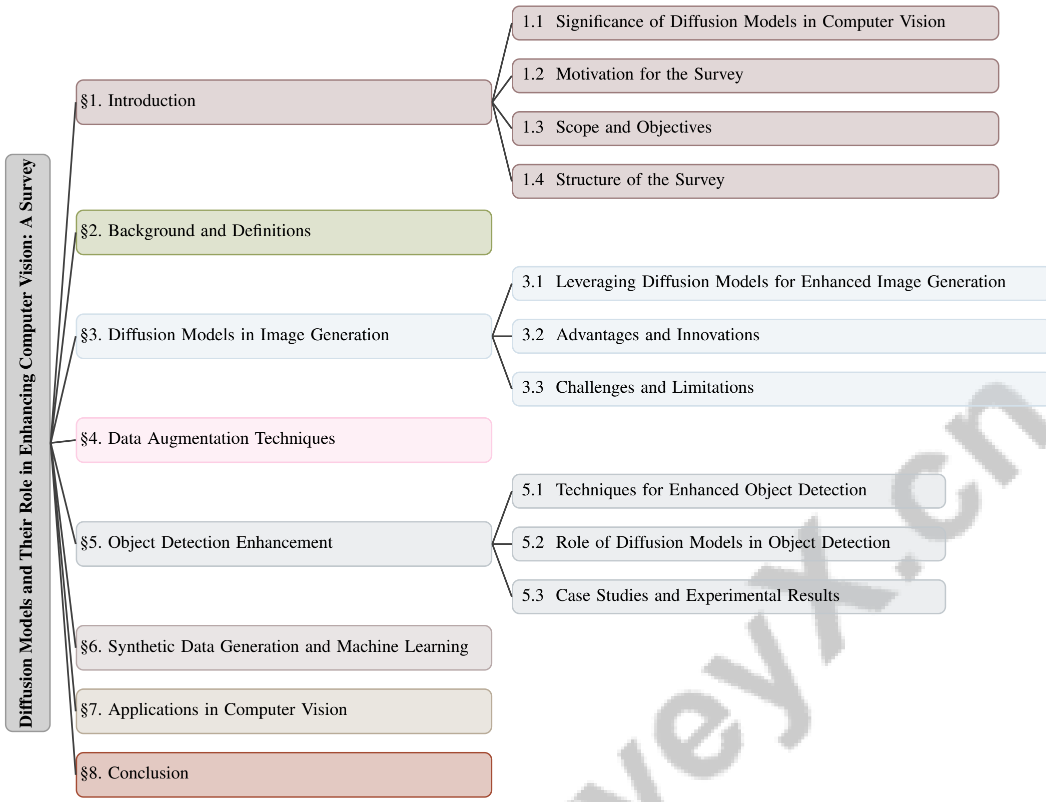


Figure 1: chapter structure

These models also enhance computational efficiency in adversarial robustness defense strategies, signifying the importance of maintaining performance while improving efficiency [5]. Their ability to improve the resolution of low-resolution microscopy images further highlights their advantages over traditional super-resolution techniques [6].

The versatility of diffusion models extends to planning tasks across various domains [7]. Additionally, as deepfake technology evolves, effective detection methods have become crucial, with diffusion models facilitating the differentiation between real and AI-generated images [8]. Their capacity to manipulate concepts within multi-dimensional latent spaces enhances visual outputs [9].

The MULTIFUSION approach further enriches image generation, enabling more nuanced outputs compared to traditional text-to-image models [10]. In forgery detection, diffusion models offer transformative advantages over conventional image generation methods [11].

As a class of generative models trained with an approximation to the log-likelihood objective, diffusion models significantly impact human-perceived image quality [12]. They have demonstrated remarkable improvements in generating photorealistic face images from multi-modal inputs, surpassing existing methods in both 2D and 3D contexts [13]. The hierarchical structures embedded in diffusion models through multi-stage processes further enhance high-quality generation with progressive transformations [14].

1.2 Motivation for the Survey

This survey is motivated by the emerging potential of diffusion models as essential tools in representation learning, providing a comprehensive framework for advancing computer vision applications [1]. These models overcome limitations associated with traditional generative adversarial networks (GANs), enabling more robust and efficient training and sampling. By systematically exploring diffusion models' applications in planning, this survey aims to highlight recent advancements and their transformative effects on computational tasks [7].

Furthermore, this survey addresses challenges in deepfake detection technologies, vital for maintaining multimedia content integrity. Developing effective benchmarks is crucial for advancing these technologies and ensuring their reliability [8]. The exploration of MULTIFUSION techniques is also driven by the need to address the limitations of current text-to-image diffusion models, particularly in managing single-language inputs and generating complex images [10].

Additionally, the survey tackles challenges in mapping multi-modal inputs into the latent space of pre-trained GANs, essential for generating realistic face images [13]. It also addresses conventional diffusion models' limitations, which often struggle with fixed Gaussian noising processes that impede their ability to learn abstract representations and adapt latent spaces [14].

By synthesizing motivations highlighted in recent literature, this survey aims to provide an in-depth overview that broadens the applicability of diffusion models across various domains, including natural language processing, structured data, and image generation. It aspires to contribute meaningfully to artificial intelligence and machine learning advancements in computer vision by addressing existing challenges, exploring innovative applications, and identifying future research directions within these rapidly evolving fields [15, 16, 17, 18].

1.3 Scope and Objectives

This survey meticulously explores the application of diffusion models within computer vision, emphasizing their impact on vision-specific tasks such as image data augmentation, object detection, and synthetic data generation [17]. By focusing on diffusion models, the survey intentionally excludes broader generative models like GANs and VAEs, concentrating on the unique contributions and enhancements provided by diffusion processes [6]. This targeted approach allows for a detailed examination of their role and efficacy in enhancing computer vision methodologies.

Specifically, the survey addresses the application of diffusion models in low-level vision tasks, including image super-resolution, deblurring, dehazing, inpainting, low-light enhancement, and remote sensing [19]. This comprehensive coverage ensures that both theoretical and practical contributions of diffusion models in computer vision are addressed, particularly in generative tasks, while deliberately excluding non-vision applications [20].

The survey also investigates advancements in text-to-image diffusion models, exploring state-of-the-art methods for text-conditioned image synthesis and their applications in creative generation and image editing [21]. By examining these specific applications, the survey aims to highlight innovations and efficiencies achieved through diffusion models in generating high-quality outputs.

The primary objectives include analyzing how diffusion models capture the hierarchical nature of data features and their implications in computer vision [22]. Additionally, the survey seeks to explore a proposed multi-stage framework with a customized multi-decoder U-net architecture, enhancing training and sampling efficiency [23]. By focusing on efficient sampling, improved likelihood estimation, and methods for handling data with unique structures, the survey aims to provide insights into the operational improvements and potential of diffusion models in complex computer vision tasks [24].

1.4 Structure of the Survey

This survey is structured to provide a comprehensive exploration of diffusion models and their transformative role in enhancing computer vision tasks. It is organized into key sections, each focusing on distinct aspects of diffusion models and their applications.

The introductory section establishes the significance of diffusion models in computer vision, outlines the motivation for the survey, and defines its scope and objectives, setting the stage for a detailed examination of core concepts and applications.

Following the introduction, the survey delves into the background and definitions necessary for understanding diffusion models, including foundational principles and interconnections with machine learning models. This section aims to clarify the theoretical underpinnings and practical implications of diffusion models in computer vision.

Subsequent sections focus on specific applications and advancements in diffusion models, exploring methodologies, advantages, and challenges associated with realistic image generation. The survey

AI-generated, for reference only.

examines data augmentation techniques leveraging diffusion models, discussing their impact on enhancing object detection and improving machine learning model training.

The analysis extends to how diffusion models enhance object detection capabilities, providing insights into improved training methods and the use of probabilistic models. It highlights the generation of synthetic data using diffusion models, discussing their advantages and drawbacks across various computer vision applications, including enhancing the realism of synthetic images and addressing challenges regarding data privacy and synthetic media detection. The effectiveness of diffusion models is compared against traditional techniques, such as random texture application, in producing high-quality object-centric representations essential for tasks like keypoint detection and segmentation [25, 26, 27].

In the applications section, the survey provides an overview of the diverse applications of diffusion models in computer vision, including medical imaging, autonomous driving, anomaly detection, creative industries, and video generation, with each subsection highlighting their transformative impacts.

Finally, the conclusion summarizes the survey's key findings, reflecting on the current state and future directions of diffusion models in computer vision while discussing challenges and opportunities for further research and development in this area.

This structured approach facilitates an in-depth examination of diffusion models, underscoring their diverse applications across various domains, including natural language processing and structured data, while addressing existing limitations and proposing future research directions, ultimately highlighting their significant potential to advance the field of computer vision and beyond [15, 18].The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Diffusion Models: Foundations and Applications

Diffusion models have revolutionized generative modeling in computer vision through their stochastic processes that iteratively add and remove noise to produce high-quality data. Denoising Diffusion Probabilistic Models (DDPMs) are a prime example, excelling in generating samples closely aligned with the training distribution while mitigating common issues like mode collapse seen in GANs [13, 14]. Despite their reliance on U-Net architectures, which may limit scalability, diffusion models have advanced image synthesis, particularly when integrated with alternative architectures like transformers [28, 29].

These models are pivotal in data augmentation, enhancing model training by generating high-quality synthetic data across modalities. Their integration with pre-trained GANs improves multi-modal face image generation by linking features to GANs' latent spaces, demonstrating diffusion models' capacity to augment existing generative frameworks [13]. However, challenges remain in computational efficiency due to their iterative nature, necessitating optimization for practical application [14]. The development of frameworks like f-DM, which facilitate diffusion on transformed signals, broadens their applicability across tasks [14].

2.2 Interconnections with Machine Learning Models

Integrating diffusion models with machine learning frameworks enhances generative capabilities and efficiency, yielding significant benefits across various computer vision applications. The GMCD method exemplifies this, preserving categorical structures during diffusion for high-quality sample generation and improved training efficiency [30]. Managing biases in large datasets is crucial for generating outputs that reflect historical and prospective contexts, enhancing interpretability and reliability. Addressing biases and employing innovative frameworks can improve performance on underrepresented data scenarios, fostering robust applications in fields like audiovisual arts and medical diagnostics [31, 32, 33, 34, 35].

Diffusion models have shown promise in table-to-text generation, outperforming traditional auto-regressive models while balancing quality and diversity in generated text. Studies on diffusion model training, sampling strategies, and prediction aggregation reveal their potential in natural language processing, offering advantages in parallel generation and token-level control [36, 37, 25, 18, 38]. This

AI-generated, for reference only.

versatility extends to analyzing training sample influence on outputs, enhancing model explainability and addressing privacy concerns [39, 26].

In medical imaging, diffusion models excel in anomaly detection using weakly supervised learning, requiring only image-level annotations. They capture healthy tissue distributions and identify anomalies through denoising, improving detection without complex training. Innovations like implicit guidance with temporal anomaly maps enhance detail preservation and reduce false positives, as demonstrated in evaluations on datasets like BRATS2020 and CheXpert [40, 41]. Additionally, diffusion models improve high-quality diffusion MRI (dMRI) image generation from lower-quality inputs, employing latent diffusion models and transfer learning to maintain structural integrity.

In ensemble learning, diffusion models address class imbalance in classification tasks, incorporating diverse generative data to improve accuracy and promote equitable performance across datasets. This approach mitigates overfitting and leverages category and prompt diversity for enhanced outcomes, evidencing performance gains on benchmark datasets [32, 42]. These capabilities highlight diffusion models' potential to enhance fairness and effectiveness in machine learning applications.

In recent years, diffusion models have emerged as a pivotal technology in the field of image generation, offering significant advantages in terms of quality and versatility. Figure 2 illustrates the hierarchical structure of these models, highlighting the key areas where diffusion models can be leveraged for enhanced image generation. This figure delineates not only the innovations brought forth by architectural advancements such as Diffusion Transformers and multimodal applications but also the challenges and limitations that practitioners face, including computational demands and output quality. By examining both the strengths and weaknesses of diffusion models, we can better understand their role in advancing image generation technology.

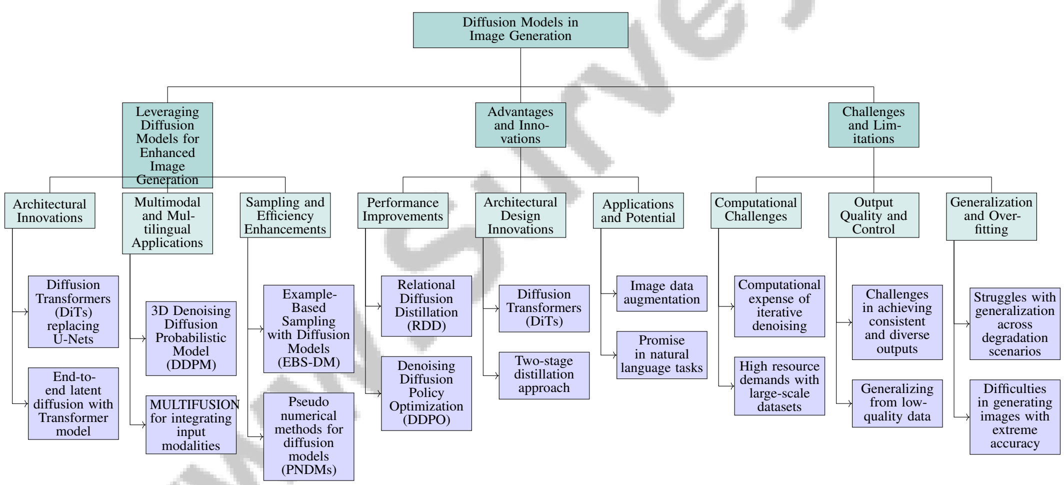


Figure 2: This figure illustrates the hierarchical structure of diffusion models in image generation, highlighting the key areas of leveraging diffusion models for enhanced image generation, their advantages and innovations, and the challenges and limitations faced. Key innovations include architectural advancements like Diffusion Transformers and multimodal applications, while challenges primarily revolve around computational demands and output quality.

3 Diffusion Models in Image Generation

3.1 Leveraging Diffusion Models for Enhanced Image Generation

Diffusion models have significantly advanced image generation, improving both quality and efficiency. A notable development is the Diffusion Transformers (DiTs), which replace U-Nets, enhancing performance and scalability [28]. This architectural evolution highlights the potential of transformers in optimizing image synthesis. Furthermore, using a Transformer model for end-to-end latent diffusion facilitates interactions between text and image modalities, boosting synthesis capabilities [29].

AI-generated, for reference only.

The 3D Denoising Diffusion Probabilistic Model (DDPM) generates diverse, realistic training data, addressing data scarcity and variability in manual segmentations. This versatility extends to natural language processing tasks like text generation and sentiment analysis, and multimodal contexts such as text-to-image generation [37, 39, 18, 43]. MULTIFUSION enhances multimodal and multilingual image generation by integrating various input modalities and languages, facilitating complex idea articulation. It effectively fuses distinct components using pre-trained models, even when trained on monomodal data [44, 10, 37, 45].

Example-Based Sampling with Diffusion Models (EBS-DM) enhances image generation by learning distributions from various samplers, improving fidelity and diversity. This adaptability is evident across architectures and training procedures [46, 47, 18]. Recent innovations include integrating graph representations as conditioning signals, enabling direct synthesis from scene graphs without intermediate layouts, leveraging pre-trained text-to-image diffusion models and CLIP guidance [48, 49].

Relational Diffusion Distillation (RDD) introduces cross-sample relationship interactions, enhancing knowledge distillation and efficiency in generated images [36, 15, 50, 26, 18]. In medical imaging, latent diffusion models generate high-quality 7T-like dMRI data from low-quality 3T data, addressing resolution differences and improving diagnostic capabilities [51, 52].

Pseudo numerical methods for diffusion models (PNDMs) enhance efficiency, reducing the required iterations for sample generation, achieving speedups while maintaining quality [24, 53, 54, 55]. The Diffusion-driven GAN Inversion (DGI) method translates feature representations into GAN latent codes, generating high-quality facial images, exemplifying the synergy between diffusion models and other generative frameworks [13, 56].

The f-DM approach combines diffusion models' expressivity with hierarchical features, improving sampling quality in video synthesis and structured data applications [50, 57, 15, 18]. Sliding window guidance (SWG) enhances image quality by constraining the receptive field, improving perceptual quality while aligning with human preferences [58, 59, 60].

As illustrated in Figure 3, the hierarchical categorization of enhanced image generation techniques using diffusion models emphasizes transformer-based methods, multimodal integration, and innovative techniques. These methodologies highlight diffusion models' advancements in image generation, synthesizing high-quality images from textual descriptions, and supporting applications like text-guided image editing and video generation. They also raise important considerations regarding data privacy and model interpretability [29, 43, 26, 16]. As these models evolve, they offer new opportunities for creative and practical applications across diverse domains, enhancing computer vision technologies.

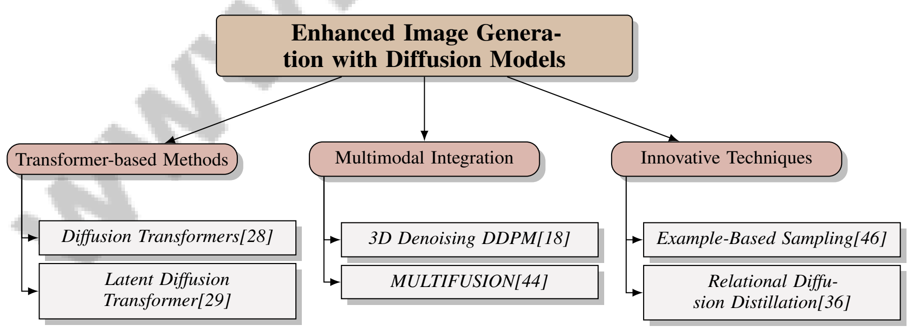


Figure 3: This figure illustrates the hierarchical categorization of enhanced image generation techniques using diffusion models, highlighting transformer-based methods, multimodal integration, and innovative techniques.

3.2 Advantages and Innovations

Diffusion models have introduced significant advancements in image generation, improving quality and efficiency. Relational Diffusion Distillation (RDD) achieves superior performance with fewer sampling steps, balancing speed and image quality, ideal for resource-constrained environments [61]. Denoising Diffusion Policy Optimization (DDPO) optimizes diffusion models for downstream tasks, enhancing applicability in complex environments [12].

Efforts to reduce computational costs without sacrificing performance have led to methods achieving competitive sample quality with smaller diffusion times [62]. Pseudo numerical methods for diffusion models (PNDMs) provide a 20x speedup in sample generation without compromising quality [55]. The f-DM approach employs a progressive transformation strategy that outperforms baseline models in visual results while requiring fewer computational resources [14].

Innovations in architectural design, such as Diffusion Transformers (DiTs), improve sample quality and scalability compared to traditional U-Net architectures [28]. The two-stage distillation approach allows a single distilled model to handle various guidance strengths, improving sampling efficiency [63]. Integrating transformer backbones simplifies interaction modeling between text and image features [29]. Sliding Window Guidance (SWG) achieves competitive generative performance while aligning better with human preferences [60].

These innovations underscore diffusion models' transformative potential in image generation, paving the way for sophisticated applications in computer vision and beyond. They improve image data augmentation by generating realistic and diverse images, enhancing training datasets and machine learning model performance. Diffusion models show promise in natural language tasks, offering advantages over traditional autoregressive models through parallel generation and improved syntactic and semantic control. This evolution is expected to drive advancements in multimodal applications and few-shot learning [17, 18].

3.3 Challenges and Limitations

Despite advancements, diffusion models in image generation face challenges and limitations. A primary challenge is the computational expense of the iterative denoising process, resulting in slow inference speeds and high resource demands, particularly with large-scale datasets [28]. Achieving high fidelity and temporal coherence in video generation remains complex, hindering high-definition and temporally coherent video sequences [63].

Conditioning mechanisms like Object Saliency Noise show promise in improving control over generated images, but challenges persist in achieving consistent and diverse outputs. Many diffusion methods struggle with managing syntactic relationships between multiple objects, leading to vague compositions [29]. Generalizing from low-quality data affects the model's ability to produce high-quality outputs, compounded by the need for additional tuning steps for optimization, particularly for unconditional images. Robust classifiers are essential for guidance, yet performance may degrade without resilience against certain noise or data distributions [64, 65, 3].

Current studies often struggle with generalization across various degradation scenarios, highlighting overfitting challenges, especially in instance segmentation tasks [46, 66, 32]. Managing non-trainable components during training remains a challenge, as methods do not accommodate unique training procedures and dependencies. Diffusion models may encounter difficulties generating images with extreme accuracy in complex scenarios. The M2M method struggles with high-fidelity human faces and experiences quality decline during extended sequences [67, 51, 68, 60, 3].

The Attention-driven Training-free Efficient Diffusion Model (AT-EDM) framework addresses computational challenges by implementing runtime pruning of redundant tokens based on attention maps, optimizing pruning budgets across denoising timesteps. This approach achieves significant reductions in floating-point operations and speed-ups while maintaining performance metrics [69, 64, 70]. However, challenges remain in suppressing issues related to fine-grain object repetition and composition.

Addressing these challenges is essential for enhancing diffusion models' capabilities in generating high-quality and diverse images, broadening applicability across computer vision domains like image data augmentation, semantic manipulation, and personalization. Future research must focus on optimizing computational efficiency, improving generalization across diverse scenarios, and managing complexities in high-dimensional data generation [17, 18, 71].

4 Data Augmentation Techniques

4.1 Innovative Approaches to Data Augmentation

Recent advancements in data augmentation have prominently featured diffusion models, which significantly enhance the quality and diversity of training datasets, thereby improving model performance

AI-generated, for reference only.

across various domains. A comprehensive review of 65 distinct techniques, particularly in medical imaging, underscores the applicability of these methods in bolstering model efficacy [72]. AeroGen exemplifies the innovative use of diffusion processes, generating high-quality remote sensing images tailored to specific layout conditions [73]. This demonstrates the versatility of diffusion models in producing nuanced outputs based on spatial constraints.

Integrating synthetic data generation techniques, facilitated by diffusion models, is crucial for augmenting training datasets, thereby enhancing representativeness and robustness [8]. Example-Based Sampling with Diffusion Models (EBS-DM) effectively addresses data scarcity by generating diverse point sets, optimizing the network's differentiability to enrich generated samples [74]. Additionally, Relational Diffusion Distillation (RDD) improves image generation by utilizing spatial relationships and memory-efficient interactions, enhancing the quality of augmented datasets [61].

In medical imaging, the THOR approach refines de-noising by reintegrating healthy tissue information while suppressing anomalies, thus improving augmented data quality [40]. These methodologies underscore the transformative impact of diffusion models on data augmentation, advancing fields such as image classification and natural language processing. Traditional methods, including transformations and GANs, have enhanced model performance through better generalization to out-of-distribution data, while diffusion models have emerged as powerful tools for generating high-quality outputs from structured data [67, 37, 75].

4.2 Addressing Data Scarcity and Imbalance

Data scarcity and imbalance pose significant challenges in machine learning, often leading to biased models and suboptimal performance. Diffusion models offer a promising solution by generating high-quality synthetic data to enhance dataset diversity and representativeness. In medical imaging, these models have been pivotal in producing synthetic images that alleviate data sharing dilemmas, providing a framework for evaluating synthetic image efficacy in healthcare contexts [76].

In scenarios with limited positive samples, such as defect detection, diffusion models generate complementary synthetic data, reducing reliance on scarce original samples. Similarly, in facial emotion recognition, where class representation is often skewed, diffusion models have proven effective in creating balanced datasets, thereby improving model performance across all classes. Recent studies suggest that these models can be tailored for specific applications, such as table-to-text generation, where they surpass traditional auto-regressive models in maintaining high-quality outputs and diversity [37, 39].

A notable advantage of diffusion models is their capacity to generate high-quality synthetic data without necessitating domain-specific knowledge or explicit rules, making them particularly useful in contexts where expertise is limited. Their generalizability, evidenced by consistent outputs across various frameworks despite differing training procedures, enhances the robustness and flexibility of training datasets. Furthermore, diffusion models have shown substantial benefits in natural language processing tasks, such as generation and sentiment analysis, by enabling parallel generation and token-level controls [46, 18]. However, many existing methods require exhaustive searches over transformation parameters, which can be computationally prohibitive, especially with large datasets.

In federated learning contexts, diffusion models facilitate privacy-preserving data sharing at the synthetic data level, enabling collaborative learning without compromising data privacy. This is particularly relevant in medical applications, where privacy concerns often hinder data sharing. Despite these advancements, challenges remain in balancing the diversity of synthetic images with their distributional fidelity. Increasing diversity in training data can enhance model robustness by exposing it to a broader range of features, yet it may also result in samples that deviate from the target distribution, negatively impacting model performance. This highlights the necessity of implementing effective strategies to ensure that introduced diversity aligns with desired data characteristics [31, 37, 39, 32, 77].

The potential of data augmentation techniques, such as Reinforcement Diffusion Model Augmentation, which achieves high accuracy with reduced computational costs, exemplifies how diffusion models can optimize data augmentation processes. Overfitting remains a critical issue in machine learning, particularly for models trained on limited datasets, often leading to inadequate generalization performance on validation and test sets. This is especially pronounced in instance segmentation tasks, where a scarcity of annotated data results in models overly conforming to training data. Recent advancements have focused on leveraging generative models for data augmentation, expanding training dataset diversity and scale. Strategies that incorporate diverse generative data can enhance model robustness, particularly for rare categories, as evidenced by substantial gains from methods like DiverGen. Moreover, the choice of data augmentation techniques is crucial, as specific augmentations have been shown to significantly bolster a model's resilience against out-of-distribution (OOD) data, underscoring the importance of effective data strategies in combating overfitting and enhancing model generalization [67, 37, 32, 42, 75].

Ongoing innovation and refinement of diffusion models are essential to fully leverage their potential in overcoming data scarcity and imbalance. Approaches like DAGAN have demonstrated significant improvements in classification performance under low-data settings, highlighting the effectiveness of data augmentation strategies. The integration of learned augmentation policies has also proven effective in enhancing classification accuracy on domain-specific datasets, addressing challenges of data scarcity and imbalance [42]. By framing data augmentation within a group-theoretic framework, significant improvements in variance reduction and model performance have been achieved.

Despite the promise of diffusion models, the computational demands of training and the necessity for extensive datasets remain limitations in achieving optimal performance. Current studies often rely on large datasets and face challenges in generalization to rare or out-of-distribution entities, constraining their practical applicability. Reducing the size of the diffusion model without compromising image generation quality is critical for effective classification in adversarial scenarios. As research progresses, prioritizing computational efficiency, expanding generalization capabilities across diverse scenarios, and managing the complexities associated with high-dimensional data generation will be essential. This includes leveraging diverse generative datasets to mitigate overfitting, particularly in instance segmentation tasks, and optimizing diffusion models for improved text generation that balances quality and diversity. Furthermore, addressing limitations in generative output diversity through robust evaluation metrics and innovative approaches can significantly enhance model performance across various applications, including image classification and data augmentation strategies [78, 37, 77, 32].

5 Object Detection Enhancement

5.1 Techniques for Enhanced Object Detection

Diffusion models have advanced object detection by significantly improving accuracy and robustness. InstaGen exemplifies this by employing diffusion models fine-tuned on object detection datasets, thus generating images with instance-level annotations to enhance training data quality and model performance [79]. This highlights diffusion models' capacity to produce detailed annotations essential for effective detection. Conditional diffusion models like ControlCom facilitate controllable image composition, allowing precise manipulation of foreground attributes to synthesize composite images, thereby improving object differentiation in complex scenes [80].

Methods like DiffPure enhance model robustness against adversarial attacks by leveraging diffusion models' denoising capabilities, thus improving object detection systems' resilience without requiring classifier retraining [65]. The DRIVE system utilizes a diffusion model pipeline to boost visual fidelity in driving simulations, crucial for training object detection models in autonomous driving [81]. In medical imaging, Denoising Diffusion Implicit Models (DDIM) generate anomaly maps by translating diseased images to healthy representations, enhancing diagnostic accuracy [41].

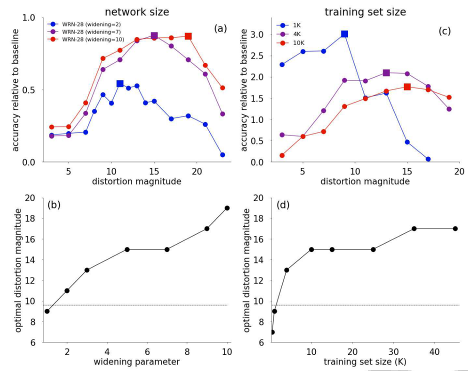
Optimal Diffusion Process Analysis evaluates score function perturbations' impact on sample generation quality, optimizing diffusion processes for better detection fidelity [82]. Reinforcement Diffusion Model Augmentation (DRDM) enhances fine-grained feature discrimination in few-shot conditions, aiding object detection with limited labeled data [83]. Prompt Diffusion demonstrates diffusion models' versatility by integrating multiple vision tasks into a single model capable of in-context learning, enhancing detection across diverse tasks [84]. Utilizing U-Net feature maps for classification tasks further demonstrates diffusion models' utility in extracting rich feature representations [85].

Diffusion models offer a suite of techniques that enhance object detection by improving data quality, robustness, and accuracy. Recent research highlights their transformative potential, particularly through frameworks like DiffusionDet, which redefine object detection as a denoising process to

AI-generated, for reference only.

iteratively refine bounding boxes, improving performance metrics such as average precision (AP) over traditional methods [86, 26].

(a) Salient Object Inpainting and Stable Inpainting Comparison[87]



(b) Impact of network size and training set size on neural network accuracy.[88]

Figure 4: Examples of Techniques for Enhanced Object Detection

As depicted in Figure 4, enhancing detection systems' accuracy and reliability is crucial, with techniques like "Salient Object Inpainting and Stable Inpainting Comparison" improving visual completeness for occluded objects, and optimizing network and training set size for higher accuracy [87, 88].

5.2 Role of Diffusion Models in Object Detection

Diffusion models are pivotal in object detection, utilizing generative capabilities to enhance data quality and robustness. Challenges in semantic segmentation and object detection, such as employing generative models like Denoising Diffusion Probabilistic Models (DDPMs) for discriminative tasks, are addressed by diffusion models through a probabilistic framework that iteratively refines predictions, exemplified by DiffusionDet's distributional approach to bounding box predictions [89, 86].

Integrating diffusion models into detection workflows enhances synthetic image generation, crucial for dataset augmentation and model robustness. The Stable Diffusion Data Augmentation benchmark highlights synthetic images' potential to boost classification and detection performance [90]. DIffusion Feature Fusion (DIFF) further illustrates this by integrating multi-step diffusion process features to enhance accuracy across domains [91].

Despite their strengths, diffusion models face challenges like susceptibility to data poisoning attacks, necessitating defensive strategies to maintain model integrity [92]. Retaining benign concepts related to unlearned concepts is another challenge, as diffusion models might relearn these even when fine-tuned on unrelated data [66].

Focusing on feature rather than image generation enhances detection accuracy through synthetic feature creation. Notable models like DALL-E 2 and Stable Diffusion generate high-quality synthetic data that bolster training datasets and detection capabilities. A generate-and-filter pipeline has extracted numerous training examples from these models, revealing their potential to produce novel images while raising concerns about data memorization and privacy. Understanding training data's influence on model outputs through ensemble methods underscores the importance of the relationship between training datasets and generated content, enhancing detection system effectiveness [39, 26].

Diffusion models also enhance robustness against adversarial attacks. The Cross-Attention Attack (CAAT) method targets cross-attention layers of latent diffusion models, improving detection accuracy and robustness against adversarial perturbations [3]. This highlights diffusion models' potential to fortify detection systems against adversarial threats.

Olearo et al. [9] demonstrate how U-Net improves detection accuracy by blending concepts, leveraging the U-Net architecture's strengths in processing latent representations, enhancing diffusion models' capabilities in capturing complex feature interactions. In medical imaging, diffusion models improve segmentation accuracy through implicit guidance mechanisms, as shown in Saillard et al.'s study on bone metastasis segmentation in CT scans [76].

Example-Based Sampling with Diffusion Models (EBS-DM) enhances point set generation, improving detection capabilities [74]. The Trinity Detector method leverages frequency domain analysis to enhance detection accuracy through advanced signal processing techniques [11].

5.3 Case Studies and Experimental Results

Benchmark	Size	Domain	Task Format	Metric
BadNets-DM[92]	1,000	Image Classification	Image Generation	FID
3D-LDM[93]	1,000	Medical Imaging	Image Synthesis	Mean Square Distance
SYN-BT[94]	100,000	Medical Imaging	Segmentation	Dice, Hausdorff
Diffusion-PLM[38]	50,000	Text Classification	Reconstruction And Ood Detection	AUROC, FAR95
DMimageDetection[27]	1,000,000	Media Forensics	Image Classification	AUC, Accuracy
SD-DA[90]	2,505,648	Weed Detection	Object Detection	mAP@50
Diffusion-TT[37]	120,000	Table-to-Text Generation	Text Generation	BLEU, PARENT
D-TRAK[70]	3,000	Artistic Creation	Data Attribution	Linear Datamodeling Score

Table 1: Table ef presents a comprehensive overview of various benchmarks used to evaluate diffusion models across different domains and tasks. It includes details on the size of each benchmark, the domain it pertains to, the specific task format, and the metrics used for evaluation. This table serves as a critical resource for understanding the diverse applications and performance measures of diffusion models in contemporary research.

The effectiveness of diffusion models in object detection is substantiated through various case studies and experimental evaluations. Table 1 provides a detailed overview of the representative benchmarks utilized in the evaluation of diffusion models, highlighting their application across various domains and tasks. The DiffusionDet framework was rigorously tested on benchmarks like COCO and CrowdHuman, demonstrating superior detection accuracy and robustness compared to established detectors across different configurations of box numbers and iteration steps [86].

Another significant study focused on data augmentation policies learned through diffusion models, which substantially improved detection accuracy, achieving state-of-the-art results adaptable across datasets and architectures [95]. However, diffusion models faced challenges in experiments involving BadNets-like data poisoning attacks, revealing performance degradation, misalignment, and trigger amplification in generated images, underscoring the need for robust defensive strategies [92].

Insights from case studies and experimental results illuminate diffusion models' transformative potential in object detection, excelling in generating high-quality synthetic images and representation learning across multiple vision tasks. However, findings also reveal critical areas for further research, particularly in enhancing diffusion models' robustness and applicability in diverse environments, addressing data memorization, privacy, and model generalization concerns [1, 26, 85].

6 Synthetic Data Generation and Machine Learning

6.1 The Role of Diffusion Models in Synthetic Data Generation

Diffusion models play a crucial role in synthetic data generation, offering sophisticated methods that enhance machine learning by producing datasets that closely resemble real-world distributions. The Diffusion-driven GAN Inversion (DGI) method exemplifies this by integrating multi-modal features into pre-trained GANs' latent space, significantly enhancing synthetic data realism, especially in complex tasks like face image synthesis [13]. The Generalized Mixture of Categorical Diffusion (GMCD) model is notable for efficiently generating high-quality categorical data, addressing data scarcity and imbalance by producing diverse synthetic datasets [30].

Pseudo Numerical Methods for diffusion models (PNDMs) enhance sample quality and achieve a 20x speedup in generation, highlighting efforts to improve diffusion models' practicality for large-scale applications [55]. The f-DM approach refines data representation through multi-stage transformations, essential for generating high-quality samples across various domains [14]. Innovations like Diffusion Transformers (DiTs) leverage latent patches within diffusion frameworks to facilitate synthetic data generation for training purposes [28]. The Latent Diffusion Transformer improves data generation quality by employing a Transformer in the latent space during the reverse diffusion process, enhancing modality interactions [29].

Despite their advantages, diffusion models face challenges such as vulnerability to adversarial attacks, which can compromise data integrity. Addressing these vulnerabilities is crucial for developing resilient strategies against malicious exploitation, thereby enhancing synthetic data applications' performance in fields like medical imaging and data augmentation. Issues like data poisoning and models' memorization tendencies must be addressed to improve robustness and reliability in real-world applications [27, 96, 3, 97, 92].

Diffusion models significantly enhance synthetic data generation, offering advanced solutions that improve training datasets' quality and utility. Their ability to generate diverse, high-quality synthetic data positions them as vital resources for advancing machine learning applications. Current research focuses on optimizing these generative models, with initiatives like DiverGen improving instance segmentation accuracy by expanding data distribution and mitigating overfitting for rare categories. BrainSPADE demonstrates the effectiveness of fully synthetic data in medical image segmentation, achieving performance comparable to models trained on real data. Furthermore, diffusion models are increasingly recognized for their potential in text generation tasks, balancing quality and diversity effectively, thereby addressing data scarcity and variability challenges across various domains [98, 37, 75, 32].

6.2 Benefits and Limitations of Synthetic Data

Synthetic data generation offers substantial benefits in enhancing machine learning models, particularly in fields like remote sensing and medical imaging, where data scarcity poses significant challenges. By providing diverse and representative training samples, synthetic data improves detection performance, especially for rare object classes [73]. This enrichment allows models to generalize more effectively across applications, enhancing robustness and accuracy.

The GMCD model exemplifies synthetic data's benefits, offering faster training and sampling times, improved sample quality, and effective categorical data modeling without arbitrary ordering [30]. Such advancements highlight synthetic data's potential to streamline training processes and improve outcomes. The MULTIFUSION framework further enhances efficiency by reducing computational resource requirements for synthetic data generation [10]. Similarly, PNDMs facilitate faster convergence and sampling, significantly improving speed and quality [55].

However, limitations persist. Synthetic images' quality can vary, and segmentation errors indicate a need for further refinement in the synthesis process [76]. The complexity of relational interactions in methods like Relational Diffusion Distillation may pose scalability challenges when applied to large datasets or when extreme memory efficiency is required [61]. Additionally, reliance on high-quality data for training, such as the need for high-quality 7T data in diffusion MRI, may restrict model performance in scenarios with limited data availability [99]. Overoptimization issues may lead models to diverge from the original distribution, potentially degrading performance on specific tasks [12]. Synthetic data generation methods may encounter overfitting challenges when applied to highly imbalanced datasets, impacting result generalizability [42]. The choice of transformations in multi-stage diffusion models like f-DM can also influence performance, making the search for optimal stage schedules an empirical challenge [14].

6.3 Applications of Synthetic Data in Computer Vision

Synthetic data is invaluable in computer vision, offering innovative solutions that enhance model training and performance across various applications. In object detection, synthetic data addresses challenges such as data scarcity and imbalance. The InstaGen framework generates high-quality synthetic images with instance-level annotations, significantly improving object detection models' robustness and mitigating long-tail issues in category representation. Future research could optimize this method by balancing injection and loss guidance or extending it to other generative tasks [52].

In style transfer and image segmentation, synthetic data enhances model generalization. Style Extracting Diffusion Models (STEDM) generate images with unseen styles, significantly boosting segmentation performance in histopathology [100]. This illustrates synthetic data's potential to leverage unannotated data, facilitating advancements in medical imaging applications. Future research should focus on improving data quality assessment metrics and exploring multimodal augmentation in other vision-language tasks [101].

Synthetic data generation is also instrumental in augmenting datasets for training GANs. The Data Augmentation with Generative Models (DAG) framework has demonstrated significant improvements in GAN performance across multiple datasets, achieving state-of-the-art Fréchet Inception Distance (FID) scores and addressing classical data augmentation limitations [102]. This highlights synthetic data's effectiveness in enhancing training datasets' quality and diversity, thereby improving generative models' performance.

In medical imaging, synthetic data generation addresses the scarcity of annotated datasets. For example, synthetic data generated for rare cataract surgery procedures has been made publicly available, facilitating research and development in this critical area [103]. Such initiatives demonstrate synthetic data's transformative impact in enabling research and innovation in fields where data collection is challenging. Future research should explore the implications of conditioning in diffusion models beyond those studied and investigate other continuous normalizing flows [104].

Moreover, synthetic data is leveraged in video generation and editing applications, where models trained on extensive video-text and image-text datasets are evaluated using metrics like FID and Fréchet Video Distance (FVD) to ensure high-definition output quality [105]. This showcases synthetic data's potential in enhancing multimedia content creation, providing high-quality training samples that improve model performance in generating realistic video sequences. Experiments on the Laion-5B dataset for image synthesis and Webvid-10M for video generation further highlight synthetic data's effectiveness in these domains [106].

The effectiveness of synthetic data in improving classification tasks has been demonstrated through experiments on datasets such as Tiny ImageNet and MNIST, where various augmentation strategies have been compared [75]. These experiments highlight synthetic data's role in enhancing model accuracy and robustness across different classification challenges. Notably, synthetic data augmentation using diffusion models significantly improves facial emotion recognition systems, with ResEmoteNet achieving accuracy increases from 79.79% to 96.47% on FER2013 and from 94.76% to 99.23% on RAF-DB [107].

7 Applications in Computer Vision

Diffusion models are pivotal in advancing generative capabilities across diverse technological domains. This section explores their transformative applications in medical imaging, autonomous driving, anomaly detection, creative industries, and video generation.

7.1 Medical Imaging

In medical imaging, diffusion models significantly enhance imaging techniques and diagnostic outcomes. The recycling training strategy exemplifies this by iteratively refining model predictions for improved segmentation performance, thereby increasing accuracy and reliability [108]. Pseudo Numerical Methods (PNDMs) for diffusion models improve efficiency by generating high-quality images with fewer computational steps, crucial for rapid processing in medical settings [55]. These models also augment training datasets, enhancing machine learning robustness in segmentation tasks. For instance, the Denoising Diffusion Medical Model (DDMM) generates realistic radiographic images from limited annotated datasets, facilitating effective biomedical analysis [109, 17]. Continued research is expected to further enhance healthcare capabilities, leading to more precise and timely diagnoses.

7.2 Autonomous Driving and Robotics

Diffusion models enhance perception and decision-making in autonomous driving and robotics by generating high-fidelity synthetic data for training perception systems, addressing data scarcity and imbalance [81]. The DRIVE system, for instance, uses these models to enhance visual fidelity in driving simulations, improving object detection in complex scenarios [81]. They also support probabilistic decision-making frameworks, enabling navigation in uncertain environments through rare event simulation [84]. In robotics, diffusion models facilitate multi-modal data integration, enhancing task performance and adaptability in dynamic settings [73]. Their application across various domains, including natural language processing and computer vision, underscores their role in advancing intelligent systems [17, 24, 18, 43, 38].

7.3 Anomaly Detection and Security

Diffusion models provide innovative solutions for anomaly detection and security by modeling normal data distributions to identify anomalies, crucial in applications like medical imaging [3]. High-fidelity synthetic data enhances anomaly detection systems' robustness, improving generalization and detection accuracy, especially for rare anomalies [41]. In security, these models enhance detection systems' ability to differentiate between normal and malicious activities, vital for early threat detection in cybersecurity [8]. They also contribute to developing robust defense mechanisms against adversarial attacks, leveraging denoising capabilities to mitigate perturbations, as seen in methods like DiffPure [65]. The application of diffusion models in anomaly detection and security enhances detection systems across domains, including manufacturing and natural language processing [18, 40, 17, 110].

7.4 Creative Industries and Image Manipulation

In creative industries, diffusion models are vital for image manipulation and artistic expression, enabling sophisticated style transfer techniques for new aesthetic exploration. The UNLEARNCANVAS project highlights ethical considerations in image manipulation [111]. These models generate stylized images that retain structural integrity while incorporating new stylistic elements, broadening creative applications. They allow precise manipulation of image attributes for tasks like object resizing, enhanced through attention mechanisms focusing on specific regions [112, 52]. Moreover, they synthesize high-resolution images from textual descriptions, opening new creative avenues and allowing artists to visualize concepts innovatively [17, 47, 113, 58, 9]. However, concerns about content replication from training datasets challenge originality in generated artworks, raising questions about data privacy and digital forgery in art [47, 17, 26, 58]. Ethical guidelines and responsible use are crucial to ensuring diffusion models positively impact creative industries.

7.5 Video Generation and Editing

Diffusion models have advanced video generation and editing significantly. The Sector-Shaped Diffusion Model (S2DM) employs a two-stage generation strategy for text-to-video tasks, allowing precise control over video dynamics and visual coherence [114]. These models produce high-definition sequences by capturing complex temporal dependencies through joint training on image and video data, innovative conditional sampling techniques, and architectural enhancements. Systems like Imagen Video integrate text conditioning and super-resolution techniques for diverse, artistically styled outputs [105, 115, 57]. This capability is valuable in film production and virtual reality, where realistic content demand is increasing. Diffusion models also enhance video editing by enabling intuitive manipulation of video attributes, such as lighting and color grading, maintaining high quality without extensive retraining [105, 57, 115]. Their potential in video generation and editing is exemplified by synthesizing high-fidelity video from textual descriptions, enhancing user experience and creative expression [57, 44, 115, 16, 105]. As these models evolve, their integration into video production is expected to revolutionize content creation and manipulation, enhancing quality and diversity in visual outputs and unlocking innovative storytelling and communication avenues [34, 57, 105]. Optimizing diffusion models for video applications will be crucial in enhancing their industry impact as research progresses.

8 Conclusion

Diffusion models have emerged as pivotal tools in computer vision, revolutionizing image generation and enhancing application performance across diverse domains. Their capability to generate intricate scenes and improve object detection has significantly advanced classification tasks. The fusion of synthetic and real data via transfer learning has particularly benefitted critical sectors like healthcare, enhancing image segmentation models. Furthermore, diffusion models have demonstrated superior proficiency in identifying out-of-distribution inputs, thus enhancing robustness across varied applications. Innovations such as DDPO have optimized diffusion models for complex objectives, improving image quality and user alignment beyond traditional methods. The Trinity Detector's exceptional performance in recognizing diffusion-generated images underscores the potential for further exploration in this field.

Despite these strides, challenges persist in ensuring the generated data's quality and diversity, optimizing computational efficiency, and addressing privacy concerns. Current efforts focus on refining text-to-image models to control image composition and manage GPU memory efficiently. Future research should aim to develop generalized models that can adapt to diverse degradation scenarios and enhance the interpretability of diffusion models in low-level vision tasks. Advancing the denoising process, exploring novel encoding strategies, and applying GMCD to a wider array of categorical data generation tasks are promising directions. Improving the mapping of diffusion features into GAN's latent space for enhanced style transfer capabilities also presents a valuable opportunity. In medical imaging, future endeavors will focus on overcoming existing limitations, improving diagnostic accuracy, and broadening the scope of unsupervised anomaly detection across various anomalies and imaging modalities.

References

[1] Michael Fuest, Pingchuan Ma, Ming Gui, Johannes Schusterbauer, Vincent Tao Hu, and Bjorn Ommer. Diffusion models and representation learning: A survey. *arXiv preprint arXiv:2407.00783*, 2024.

[2] Yunzhuo Chen, Nur Al Hasan Haldar, Naveed Akhtar, and Ajmal Mian. Text-image guided diffusion model for generating deepfake celebrity interactions, 2023.

[3] Jingyao Xu, Yuetong Lu, Yandong Li, Siyang Lu, Dongdong Wang, and Xiang Wei. Perturbing attention gives you more bang for the buck: Subtle imaging perturbations that efficiently fool customized diffusion models, 2024.

[4] Chenshuang Zhang, Chaoning Zhang, Sheng Zheng, Mengchun Zhang, Maryam Qamar, Sung-Ho Bae, and In So Kweon. A survey on audio diffusion models: Text to speech synthesis and enhancement in generative ai, 2023.

[5] Hefei Mei, Minjing Dong, and Chang Xu. Efficient image-to-image diffusion classifier for adversarial robustness, 2025.

[6] Harshith Bachimanchi and Giovanni Volpe. Diffusion models to enhance the resolution of microscopy images: A tutorial, 2024.

[7] Toshihide Ubukata, Jialong Li, and Kenji Tei. Diffusion model for planning: A systematic literature review, 2024.

[8] Luca Guarnera, Oliver Giudice, and Sebastiano Battiato. Level up the deepfake detection: a method to effectively discriminate images generated by gan architectures and diffusion models, 2023.

[9] Lorenzo Olearo, Giorgio Longari, Simone Melzi, Alessandro Raganato, and Rafael Peñaloza. How to blend concepts in diffusion models, 2024.

[10] Marco Bellagente, Manuel Brack, Hannah Teufel, Felix Friedrich, Björn Deiseroth, Constantin Eichenberg, Andrew Dai, Robert Baldock, Souradeep Nanda, Koen Oostermeijer, Andres Felipe Cruz-Salinas, Patrick Schramowski, Kristian Kersting, and Samuel Weinbach. Multifusion: Fusing pre-trained models for multi-lingual, multi-modal image generation, 2023.

[11] Jiawei Song, Dengpan Ye, and Yunming Zhang. Trinity detector:text-assisted and attention mechanisms based spectral fusion for diffusion generation image detection, 2024.

[12] Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning, 2024.

[13] Jihyun Kim, Changjae Oh, Hoseok Do, Soohyun Kim, and Kwanghoon Sohn. Diffusion-driven gan inversion for multi-modal face image generation, 2024.

[14] Jiatao Gu, Shuangfei Zhai, Yizhe Zhang, Miguel Angel Bautista, and Josh Susskind. f-dm: A multi-stage diffusion model via progressive signal transformation, 2022.

[15] Heejoon Koo and To Eun Kim. A comprehensive survey on generative diffusion models for structured data, 2023.

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

[17] Panagiotis Alimisis, Ioannis Mademlis, Panagiotis Radoglou-Grammatikis, Panagiotis Sarigiannidis, and Georgios Th. Papadopoulos. Advances in diffusion models for image data augmentation: A review of methods, models, evaluation metrics and future research directions, 2025.

[18] Hao Zou, Zae Myung Kim, and Dongyeop Kang. A survey of diffusion models in natural language processing, 2023.

[19] Chunming He, Yuqi Shen, Chengyu Fang, Fengyang Xiao, Longxiang Tang, Yulun Zhang, Wangmeng Zuo, Zhenhua Guo, and Xiu Li. Diffusion models in low-level vision: A survey, 2025.

[20] Florinel-Alin Croitoru, Vlad Hondru, Radu Tudor Ionescu, and Mubarak Shah. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10850–10869, 2023.

[21] Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, and In So Kweon. Text-to-image diffusion models in generative ai: A survey. *arXiv preprint arXiv:2303.07909*, 2023.

[22] Antonio Sclocchi, Alessandro Favero, and Matthieu Wyart. A phase transition in diffusion models reveals the hierarchical nature of data, 2024.

[23] Huijie Zhang, Yifu Lu, Ismail Alkhouri, Saiprasad Ravishankar, Dogyoon Song, and Qing Qu. Improving efficiency of diffusion models via multi-stage framework and tailored multi-decoder architectures, 2024.

[24] Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys*, 56(4):1–39, 2023.

[25] Thomas Lips and Francis wyffels. Evaluating text-to-image diffusion models for texturing synthetic data, 2024.

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

[27] Riccardo Corvi, Davide Cozzolino, Giada Zingarini, Giovanni Poggi, Koki Nagano, and Luisa Verdoliva. On the detection of synthetic images generated by diffusion models, 2022.

[28] William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4195–4205, 2023.

[29] Princy Chahal. Exploring transformer backbones for image diffusion models, 2022.

[30] Florence Regol and Mark Coates. Diffusing gaussian mixtures for generating categorical data, 2023.

[31] Atoosa Chegini and Soheil Feizi. Identifying and mitigating model failures through few-shot clip-aided diffusion generation, 2023.

[32] Chengxiang Fan, Muzhi Zhu, Hao Chen, Yang Liu, Weijia Wu, Huaqi Zhang, and Chunhua Shen. Divergen: Improving instance segmentation by learning wider data distribution with more diverse generative data, 2024.

[33] Bahman Mirheidari, Yilin Pan, Daniel Blackburn, Ronan O'Malley, Traci Walker, Annalena Venneri, Markus Reuber, and Heidi Christensen. Data augmentation using generative networks to identify dementia, 2020.

[34] Luís Arandas, Mick Grierson, and Miguel Carvalhais. Antagonising explanation and revealing bias directly through sequencing and multimodal inference, 2023.

[35] Shi Fu, Sen Zhang, Yingjie Wang, Xinmei Tian, and Dacheng Tao. Towards theoretical understandings of self-consuming generative models, 2024.

[36] Melike Nur Yeğin and Mehmet Fatih Amasyalı. Theoretical research on generative diffusion models: an overview, 2024.

[37] Aleksei S. Krylov and Oleg D. Somov. Table-to-text generation with pretrained diffusion models, 2024.

[38] Huazheng Wang, Daixuan Cheng, Haifeng Sun, Jingyu Wang, Qi Qi, Jianxin Liao, Jing Wang, and Cong Liu. How does diffusion influence pretrained language models on out-of-distribution data?, 2023.

[39] Zheng Dai and David K Gifford. Training data attribution for diffusion models, 2023.

[40] Cosmin I. Bercea, Benedikt Wiestler, Daniel Rueckert, and Julia A. Schnabel. Diffusion models with implicit guidance for medical anomaly detection, 2024.

[41] Julia Wolleb, Florentin Bieder, Robin Sandkühler, and Philippe C. Cattin. Diffusion models for medical anomaly detection, 2022.

[42] Azal Ahmad Khan, Omkar Chaudhari, and Rohitash Chandra. A review of ensemble learning and data augmentation models for class imbalanced problems: combination, implementation and evaluation, 2023.

[43] Rushikesh Zawar, Shaurya Dewan, Prakanshul Saxena, Yingshan Chang, Andrew Luo, and Yonatan Bisk. Diffusionpid: Interpreting diffusion via partial information decomposition, 2024.

[44] Wang Lin, Jingyuan Chen, Jiaxin Shi, Yichen Zhu, Chen Liang, Junzhong Miao, Tao Jin, Zhou Zhao, Fei Wu, Shuicheng Yan, and Hanwang Zhang. Non-confusing generation of customized concepts in diffusion models, 2024.

[45] Cusuh Ham, James Hays, Jingwan Lu, Krishna Kumar Singh, Zhifei Zhang, and Tobias Hinz. Modulating pretrained diffusion models for multimodal image synthesis, 2023.

[46] Huijie Zhang, Jinfan Zhou, Yifu Lu, Minzhe Guo, Peng Wang, Liyue Shen, and Qing Qu. The emergence of reproducibility and generalizability in diffusion models, 2024.

[47] Diffusion art or digital forgery? investigating data replication in diffusion models.

[48] Nan Liu, Shuang Li, Yilun Du, Antonio Torralba, and Joshua B Tenenbaum. Compositional visual generation with composable diffusion models. In *European Conference on Computer Vision*, pages 423–439. Springer, 2022.

[49] Rameshwar Mishra and A V Subramanyam. Image synthesis with graph conditioning: Clip-guided diffusion models for scene graphs, 2024.

[50] Hanqun Cao, Cheng Tan, Zhangyang Gao, Yilun Xu, Guangyong Chen, Pheng-Ann Heng, and Stan Z Li. A survey on generative diffusion models. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[51] Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Null-text inversion for editing real images using guided diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6038–6047, 2023.

[52] Zakaria Patel and Kirill Serkh. Enhancing image layout control with loss-guided diffusion models, 2024.

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

[54] Kaiwen Zheng, Cheng Lu, Jianfei Chen, and Jun Zhu. Improved techniques for maximum likelihood estimation for diffusion odes, 2024.

[55] Luping Liu, Yi Ren, Zhijie Lin, and Zhou Zhao. Pseudo numerical methods for diffusion models on manifolds. *arXiv preprint arXiv:2202.09778*, 2022.

[56] Unifying diffusion models' latent space with applications to cyclediffusion and guidance.

[57] Patrick Esser, Johnathan Chiu, Parmida Atighehchian, Jonathan Granskog, and Anastasis Germanidis. Structure and content-guided video synthesis with diffusion models. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7346–7356, 2023.

[58] Gwanghyun Kim, Taesung Kwon, and Jong Chul Ye. Diffusionclip: Text-guided diffusion models for robust image manipulation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2426–2435, 2022.

[59] Vincent Tao Hu, David W Zhang, Yuki M. Asano, Gertjan J. Burghouts, and Cees G. M. Snoek. Self-guided diffusion models, 2023.

[60] Tim Kaiser, Nikolas Adaloglou, and Markus Kollmann. The unreasonable effectiveness of guidance for diffusion models, 2024.

[61] Weilun Feng, Chuanguang Yang, Zhulin An, Libo Huang, Boyu Diao, Fei Wang, and Yongjun Xu. Relational diffusion distillation for efficient image generation, 2024.

[62] Giulio Franzese, Simone Rossi, Lixuan Yang, Alessandro Finamore, Dario Rossi, Maurizio Filippone, and Pietro Michiardi. How much is enough? a study on diffusion times in score-based generative models, 2022.

[63] Chenlin Meng, Robin Rombach, Ruiqi Gao, Diederik Kingma, Stefano Ermon, Jonathan Ho, and Tim Salimans. On distillation of guided diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14297–14306, 2023.

[64] Hongjie Wang, Difan Liu, Yan Kang, Yijun Li, Zhe Lin, Niraj K. Jha, and Yuchen Liu. Attention-driven training-free efficiency enhancement of diffusion models, 2024.

[65] Weili Nie, Brandon Guo, Yujia Huang, Chaowei Xiao, Arash Vahdat, and Anima Anandkumar. Diffusion models for adversarial purification. *arXiv preprint arXiv:2205.07460*, 2022.

[66] Hongcheng Gao, Tianyu Pang, Chao Du, Taihang Hu, Zhijie Deng, and Min Lin. Meta-unlearning on diffusion models: Preventing relearning unlearned concepts, 2024.

[67] Zahra Golpayegani, Patrick St-Amant, and Nizar Bouguila. Clarifying myths about the relationship between shape bias, accuracy, and robustness, 2024.

[68] Vikash Sehwag, Caner Hazirbas, Albert Gordo, Firat Ozgenel, and Cristian Canton Ferrer. Generating high fidelity data from low-density regions using diffusion models, 2022.

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

[70] Xiaosen Zheng, Tianyu Pang, Chao Du, Jing Jiang, and Min Lin. Intriguing properties of data attribution on diffusion models, 2024.

[71] Hadas Orgad, Bahjat Kawar, and Yonatan Belinkov. Editing implicit assumptions in text-to-image diffusion models, 2023.

[72] Manuel Cossio. Augmenting medical imaging: A comprehensive catalogue of 65 techniques for enhanced data analysis, 2023.

[73] Datao Tang, Xiangyong Cao, Xuan Wu, Jialin Li, Jing Yao, Xueru Bai, Dongsheng Jiang, Yin Li, and Deyu Meng. Aerogen: Enhancing remote sensing object detection with diffusion-driven data generation, 2025.

[74] Bastien Doignies, Nicolas Bonneel, David Coeurjolly, Julie Digne, Loïs Paulin, Jean-Claude Iehl, and Victor Ostromoukhov. Example-based sampling with diffusion models, 2023.

[75] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.

[76] Emile Saillard, Aurélie Levillain, David Mitton, Jean-Baptiste Pialat, Cyrille Confavreux, Hélène Follet, and Thomas Grenier. Enhanced segmentation of femoral bone metastasis in ct scans of patients using synthetic data generation with 3d diffusion models, 2024.

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

[78] Tao Gui, Lizhi Qing, Qi Zhang, Jiacheng Ye, Hang Yan, Zichu Fei, and Xuanjing Huang. Constructing multiple tasks for augmentation: Improving neural image classification with k-means features, 2019.

[79] Chengjian Feng, Yujie Zhong, Zequn Jie, Weidi Xie, and Lin Ma. Instagen: Enhancing object detection by training on synthetic dataset, 2024.

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

[81] Fanjun Bu and Hiroshi Yasuda. Boosting visual fidelity in driving simulations through diffusion models, 2024.

[82] Yu Cao, Jingrun Chen, Yixin Luo, and Xiang Zhou. Exploring the optimal choice for generative processes in diffusion models: Ordinary vs stochastic differential equations, 2023.

[83] Tianxu Wu, Shuo Ye, Shuhuang Chen, Qinmu Peng, and Xinge You. Detail reinforcement diffusion model: Augmentation fine-grained visual categorization in few-shot conditions, 2024.

[84] Zhendong Wang, Yifan Jiang, Yadong Lu, Yelong Shen, Pengcheng He, Weizhu Chen, Zhangyang Wang, and Mingyuan Zhou. In-context learning unlocked for diffusion models, 2023.

[85] Soumik Mukhopadhyay, Matthew Gwilliam, Vatsal Agarwal, Namitha Padmanabhan, Archana Swaminathan, Srinidhi Hegde, Tianyi Zhou, and Abhinav Shrivastava. Diffusion models beat gans on image classification, 2023.

[86] Shoufa Chen, Peize Sun, Yibing Song, and Ping Luo. Diffusiondet: Diffusion model for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 19830–19843, 2023.

[87] Amir Erfan Eshratifar, Joao V. B. Soares, Kapil Thadani, Shaunak Mishra, Mikhail Kuznetsov, Yueh-Ning Ku, and Paloma de Juan. Salient object-aware background generation using text-guided diffusion models, 2024.

[88] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703, 2020.

[89] Dmitry Baranchuk, Ivan Rubachev, Andrey Voynov, Valentin Khrulkov, and Artem Babenko. Label-efficient semantic segmentation with diffusion models, 2022.

[90] Boyang Deng. Stable diffusion for data augmentation in coco and weed datasets, 2025.

[91] Yuxiang Ji, Boyong He, Chenyuan Qu, Zhuoyue Tan, Chuan Qin, and Liaoni Wu. Diffusion features to bridge domain gap for semantic segmentation, 2024.

[92] Zhuoshi Pan, Yuguang Yao, Gaowen Liu, Bingquan Shen, H. Vicky Zhao, Ramana Rao Kompella, and Sijia Liu. From trojan horses to castle walls: Unveiling bilateral data poisoning effects in diffusion models, 2024.

[93] Salman Ul Hassan Dar, Arman Ghanaat, Jannik Kahmann, Isabelle Ayx, Theano Papavassiliu, Stefan O. Schoenberg, and Sandy Engelhardt. Investigating data memorization in 3d latent diffusion models for medical image synthesis, 2023.

[94] Muhammad Usman Akbar, Måns Larsson, and Anders Eklund. Brain tumor segmentation using synthetic mr images – a comparison of gans and diffusion models, 2024.

[95] Barret Zoph, Ekin D Cubuk, Golnaz Ghiasi, Tsung-Yi Lin, Jonathon Shlens, and Quoc V Le. Learning data augmentation strategies for object detection. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16*, pages 566–583. Springer, 2020.

[96] Hoo-Chang Shin, Neil A Tenenholtz, Jameson K Rogers, Christopher G Schwarz, Matthew L Senjem, Jeffrey L Gunter, Katherine Andriole, and Mark Michalski. Medical image synthesis for data augmentation and anonymization using generative adversarial networks, 2018.

[97] Fabio Henrique Kiyoiti Dos Santos Tanaka and Claus Aranha. Data augmentation using gans. *arXiv preprint arXiv:1904.09135*, 2019.

[98] Virginia Fernandez, Walter Hugo Lopez Pinaya, Pedro Borges, Petru-Daniel Tudosiu, Mark S Graham, Tom Vercauteren, and M Jorge Cardoso. Can segmentation models be trained with fully synthetically generated data?, 2022.

[99] Xi Zhu, Wei Zhang, Yijie Li, Lauren J. O'Donnell, and Fan Zhang. When diffusion mri meets diffusion model: A novel deep generative model for diffusion mri generation, 2024.

[100] Mathias Öttl, Frauke Wilm, Jana Steenpass, Jingna Qiu, Matthias Rübner, Arndt Hartmann, Matthias Beckmann, Peter Fasching, Andreas Maier, Ramona Erber, Bernhard Kainz, and Katharina Breininger. Style-extracting diffusion models for semi-supervised histopathology segmentation, 2024.

[101] Changrong Xiao, Sean Xin Xu, and Kunpeng Zhang. Multimodal data augmentation for image captioning using diffusion models, 2023.

[102] Ngoc-Trung Tran, Viet-Hung Tran, Ngoc-Bao Nguyen, Trung-Kien Nguyen, and Ngai-Man Cheung. On data augmentation for gan training. *IEEE Transactions on Image Processing*, 30:1882–1897, 2021.

[103] Yannik Frisch, Moritz Fuchs, Antoine Sanner, Felix Anton Ucar, Marius Frenzel, Joana Wasielica-Poslednik, Adrian Gericke, Felix Mathias Wagner, Thomas Dratsch, and Anirban Mukhopadhyay. Synthesising rare cataract surgery samples with guided diffusion models, 2023.

[104] Mattias Cross and Anton Ragni. What happens to diffusion model likelihood when your model is conditional?, 2024.

[105] Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P. Kingma, Ben Poole, Mohammad Norouzi, David J. Fleet, and Tim Salimans. Imagen video: High definition video generation with diffusion models, 2022.

[106] Lanqing Guo, Yingqing He, Haoxin Chen, Menghan Xia, Xiaodong Cun, Yufei Wang, Siyu Huang, Yong Zhang, Xintao Wang, Qifeng Chen, Ying Shan, and Bihan Wen. Make a cheap scaling: A self-cascade diffusion model for higher-resolution adaptation, 2024.

[107] Arnab Kumar Roy, Hemant Kumar Kathania, and Adhitiya Sharma. Improvement in facial emotion recognition using synthetic data generated by diffusion model, 2024.

[108] Yunguan Fu, Yiwen Li, Shaheer U Saeed, Matthew J Clarkson, and Yipeng Hu. A recycling training strategy for medical image segmentation with diffusion denoising models, 2023.

[109] Pham Ngoc Huy and Tran Minh Quan. Denoising diffusion medical models, 2023.

[110] Haonan Yin, Guanlong Jiao, Qianhui Wu, Borje F. Karlsson, Biqing Huang, and Chin Yew Lin. Lafite: Latent diffusion model with feature editing for unsupervised multi-class anomaly detection, 2023.

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

[112] Jiteng Mu, Michaël Gharbi, Richard Zhang, Eli Shechtman, Nuno Vasconcelos, Xiaolong Wang, and Taesung Park. Editable image elements for controllable synthesis, 2024.

[113] Shounak Chatterjee. Diffmorph: Text-less image morphing with diffusion models, 2024.

[114] Haoran Lang, Yuxuan Ge, and Zheng Tian. S2dm: Sector-shaped diffusion models for video generation, 2024.

[115] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. Video diffusion models. *Advances in Neural Information Processing Systems*, 35:8633–8646, 2022.

Disclaimer:

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