

Anomaly Detection and Generation with Diffusion Models: A Survey

Yang Liu , Jing Liu , Member, IEEE, Chengfang Li , Rui Xi , Wenchao Li , Liang Cao , Jin Wang , Laurence T. Yang , Fellow, IEEE, Junsong Yuan , Fellow, IEEE, Wei Zhou , Senior Member, IEEE

Abstract—Anomaly detection (AD) plays a pivotal role across diverse domains, including cybersecurity, finance, healthcare, and industrial manufacturing, by identifying unexpected patterns that deviate from established norms in real-world data. Recent advancements in deep learning, specifically diffusion models (DMs), have sparked significant interest due to their ability to learn complex data distributions and generate high-fidelity samples, offering a robust framework for unsupervised AD. In this survey, we comprehensively review anomaly detection and generation with diffusion models (ADGDM), presenting a tutorial-style analysis of the theoretical foundations and practical implementations and spanning images, videos, time series, tabular, and multimodal data. Crucially, unlike existing surveys that often treat anomaly detection and generation as separate problems, we highlight their inherent synergistic relationship. We reveal how DMs enable a reinforcing cycle where generation techniques directly address the fundamental challenge of anomaly data scarcity, while detection methods provide critical feedback to improve generation fidelity and relevance, advancing both capabilities beyond their individual potential. A detailed taxonomy categorizes ADGDM methods based on anomaly scoring mechanisms, conditioning strategies, and architectural designs, analyzing their strengths and limitations. We finally discuss key challenges including scalability and computational efficiency, and outline promising future directions such as efficient architectures, conditioning strategies, and integration with foundation models (e.g., visual-language models and large language models). By synthesizing recent advances and outlining open research questions, this survey (QADGDM) aims to guide researchers and practitioners in leveraging DMs for innovative AD solutions across diverse applications.

Index Terms—Anomaly Detection, Anomaly Generation, Diffusion Models, Generative AI, Unsupervised Learning, Survey

1 INTRODUCTION

Anomaly detection (AD) is a crucial task in various domains, from cybersecurity [1] and finance [2] to healthcare [3] and industrial manufacturing [4], aiming to identify instances deviating significantly from established normal patterns [5]. However, the unsupervised nature of

- Detailed biographical information for all authors is provided in the supplementary material "Author Biographies".
- Yang Liu is with the Department of Computer Science, The University of Toronto, ON M5S 1A1, Canada (e-mail: yangliu@cs.toronto.edu).
- Jing Liu is with the College of Future Information Technology, Fudan University, Shanghai 200433, China, and also with the Department of Electrical and Computer Engineering, The University of British Columbia, BC V6T 1Z4, Canada (e-mail: jing.liu@ieee.org).
- Chengfang Li is with the Academy for Engineering & Technology, Fudan University, Shanghai 200433, China (e-mail: cfl20@fudan.edu.cn).
- Rui Xi is with the Department of Electrical and Computer Engineering, The University of British Columbia, BC V6T 1Z4, Canada (e-mail: rui.xi@ece.ubc.ca).
- Wenchao Li is with the School of Computer Science, University of Sydney, NSW 2006, Australia (e-mail: li58843972@163.com).
- Liang Cao is with the Department of Chemical Engineering, Massachusetts Institute of Technology, MA 02139, United States (e-mail: liangcao@mit.edu).
- Jin Wang is with the School of Future Science and Engineering, Soochow University, Suzhou 215006, China (e-mail: wjin1985@suda.edu.cn).
- Laurence T. Yang is with School of Computer Science and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China, and also with the Department of Computer Science, St. Francis Xavier University, Antigonish, NS B2G 2W5, Canada (e-mail: ltyang@ieee.org).
- Junsong Yuan is with Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, NY 14260, USA (e-mail: jsyuan@buffalo.edu).
- Wei Zhou is with the School of Computer Science and Informatics, Cardiff University, CF24 4AG Cardiff, U.K. (e-mail: zhoulw26@cardiff.ac.uk).

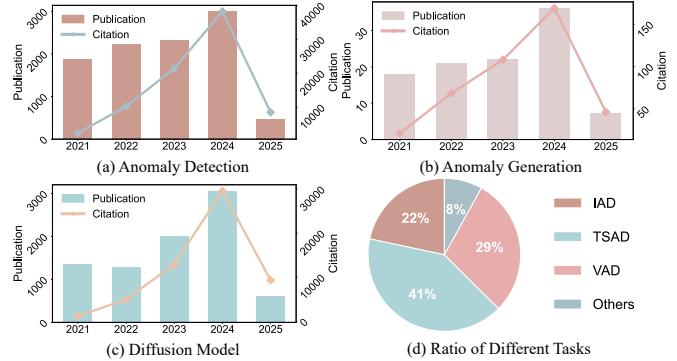


Fig. 1: Publication and citation trends in anomaly-related research topic from 2021 to 2025 (statistical time 2025/03/20).

AD presents substantial challenges, particularly the lack of labeled anomalous data, which makes defining precise boundaries between normal and abnormal behavior difficult [6]. Traditional AD methods, often relying on predefined thresholds or assuming specific data distributions, are sensitive to noise and struggle to adapt to complex real-world scenarios [7]. The increasing complexity and dimensionality of data in modern applications further compound these challenges, demanding more robust and adaptable techniques [8]. Consequently, innovative approaches are needed to learn from the inherent structure of normal data and effectively identify deviations without explicit supervision [9]. The scarcity of labeled anomalies coupled with data heterogeneity complicates the development of effective unsupervised AD methods [10]. In addition, the need for interpretability in anomaly detection adds another layer of complexity, as

understanding the reasons for flagged anomalies is crucial for effective decision-making [11].

Recent years have witnessed the emergence of DMs as a powerful class of generative models that have rapidly gained prominence across various computer vision tasks [12], [13]. At their core, DMs operate through a distinctive two-stage process: forward diffusion and reverse denoising. During forward diffusion, Gaussian noise progressively corrupts input data across multiple steps, transforming complex data distributions into simple isotropic Gaussian structures [14]. The neural network-parameterized reverse denoising process then learns to iteratively remove the added noise, reconstructing original data from noisy samples while capturing intricate data dependencies [15]. Compared to traditional generative approaches such as generative adversarial networks (GANs) and variational autoencoders (VAEs), DMs offer significant advantages in training stability by avoiding mode collapse and vanishing gradient challenges [16]. With superior sample quality and improved mode coverage, DMs generate diverse, high-fidelity samples that accurately represent underlying data distributions, positioning them at the forefront of generative modeling research for applications including anomaly detection [17]. Ongoing research addresses computational limitations through techniques like patching, which employs ViT-style transformations to reduce sampling time and memory requirements, while flexibility in the forward process enables multi-scale training that enhances overall performance [1].

Considering the inherent challenges of anomaly scarcity and data distribution complexity, identifying instances that deviate significantly from expected behavior poses particular challenges in unsupervised settings with limited labeled anomalies [18], [19], where DMs provide a promising solution by learning complex data distributions and generating high-fidelity samples for effective reconstruction-based anomaly scoring [20], [21]. Specifically, a DM trained on normal data learns to progressively add and then reverse noise, effectively reconstructing data from a noise distribution [22]. Consequently, when presented with an anomalous input, the DM attempts to reconstruct a "normal" version, and the discrepancy between the input and its reconstruction, quantified as the reconstruction error, serves as an anomaly score [23], [24]. A larger reconstruction error thus implies a higher likelihood of an anomaly, allowing DMs to capture subtle deviations, particularly in high-dimensional image data [25], [26]. As shown in Fig. 1, research interest in anomaly detection and diffusion models has grown substantially from 2021 to 2025, with both publication counts and citation metrics exhibiting steep upward trajectories. Time series anomaly detection (TSAD) currently dominates the research landscape, accounting for 41% of all the AD publications, followed by video anomaly detection (VAD, 29%) and image anomaly detection (IAD, 22%). Such a distribution reflects both the versatility of diffusion models across different data modalities and the particularly challenging nature of detecting anomalies in sequential data [27]–[29]. Due to incomplete data for the current year, the overall trend indicates sustained interest in applying DMs to anomaly detection and generation tasks across various domains.

Related Surveys. Despite the growing interest in anomaly detection research, existing surveys typically focus on specific

modalities or methodological aspects rather than offering a unified perspective across different data types and generative approaches. For example, Cook et al. [30] concentrate solely on IoT time-series data anomaly detection without addressing generative approaches, while Pang et al. [31] provide a general overview of deep learning for anomaly detection but omit specific coverage of diffusion models. In the medical domain, Fernando et al. [32] thoroughly examine anomaly detection techniques but predominantly focus on discriminative rather than generative approaches. Regarding VAD, Ramachandra et al. [7] and Chandrakala et al. [33] offer valuable insights into single-scene surveillance applications but lack exploration of generative approaches such as diffusion models. Furthermore, while Liu et al. [36] establish a comprehensive taxonomy for VAD and Liu et al. [37] investigate networking systems for VAD, neither addresses the potential of DMs. Similarly, Ma et al. [34] extensively explore graph anomaly detection with deep learning but without investigating diffusion-based approaches. More recently, Jin et al. [6] and Zamanzadeh Darban et al. [8] have advanced TSAD research, yet their coverage of DMs remains limited. A detailed comparison between our survey and existing works is presented in Table 1, which demonstrates our unique contribution in comprehensively addressing both anomaly detection and generation across multiple modalities through the lens of diffusion models.

Scope and Contributions. In this survey¹, we provide a comprehensive and up-to-date examination of anomaly detection and generation with diffusion models (ADGDM), spanning multiple data modalities including images, videos, time series, and tabular data. A distinctive aspect is our integrated view of detection and generation as complementary processes, wherein DMs enable controllable generation of synthetic anomalies to address data scarcity, while detection insights guide targeted generation. Consequently, a synergistic feedback loop emerges, transforming the fundamental limitation of anomaly data scarcity into an opportunity for continuous improvement. As DMs continue to demonstrate remarkable capabilities in generative AI tasks across various domains, their application to anomaly detection represents a rapidly evolving frontier with significant potential. The key contributions of our survey are summarized as follows:

- To the best of our knowledge, this is the first comprehensive survey specifically dedicated to examining the intersection of diffusion models with anomaly detection and generation across diverse data modalities including images, videos, time series, and tabular data. Our integrative perspective reveals common principles and modality-specific challenges, fostering knowledge transfer between previously siloed research areas.
- We establish a taxonomy that categorizes ADGDM methods based on their anomaly scoring mechanisms (reconstruction-based, density-based, score-based) and conditioning strategies (unconditional, conditional). Such structured framework provides researchers with a clear understanding of the methodological landscape and facilitates the positioning of new contributions.
- Unlike previous surveys focusing solely on detection, we address both anomaly detection and generation with

1. Resources at Github: <https://github.com/fudanyliu/ADGDM>

TABLE 1: Summary of our survey with recently AD surveys, highlighting the main focus, domain, scope, and resources.

Ref.	Year	Venue	Main Focus	Domain		Scope						Resources			
				Specific	General	IAD	VAD	TSAD	TAD	MAD	AG	DM	Dataset	Metric	Page
[30]	2020	IEEE IoTJ	Time series AD in IoT	✓		○	○	●	○	○	○	○	○	○	-
[31]	2021	ACM CSUR	General AD with DL		✓	●	○	●	●	●	○	○	●	○	-
[32]	2022	ACM CSUR	Medical AD	✓		●	○	●	●	○	○	○	○	●	-
[7]	2022	IEEE TPAMI	Single-scene video AD	✓		○	●	○	○	○	○	○	●	●	-
[33]	2023	AI Rev.	DL for surveillance VAD	✓		○	●	○	○	○	○	○	●	○	-
[34]	2023	IEEE TKDE	Graph AD with DL		✓	○	○	○	●	●	○	●	●	●	○
[35]	2023	IEEE COMST	Blockchain AD	✓		○	○	●	●	○	○	○	○	●	-
[6]	2024	IEEE TPAMI	GNN for time series analytics		✓	○	○	●	●	○	○	○	●	●	○
[36]	2024	ACM CSUR	Generalized VAD taxonomy		✓	●	●	○	○	●	○	○	●	●	○
[37]	2024	ACM CSUR	Networking systems for VAD	✓		●	●	○	○	●	○	○	●	●	○
[8]	2025	ACM CSUR	Time series AD with DL	✓		○	○	●	●	○	●	○	●	●	○
Ours	2025	-	AD & AG with DMs	✓		●	●	●	●	●	●	●	●	●	●

Notes: Symbol definitions used throughout the comparison: ○: "Not Covered", ●: "Partially Covered", ○: "Not applicable", and ●: "Fully Covered". Abbreviations include image AD (IAD), video AD (VAD), time series AD (TSAD), tabular AD (TAD), multimodal AD (MAD), anomaly generation (AG), and diffusion models (DMs).

diffusion models, where the processes form a synergistic cycle—generation creates diverse synthetic anomalies to overcome data scarcity, while detection techniques inform targeted generation approaches. We highlight how generative capabilities enhance data augmentation, model robustness testing, and self-supervised learning.

- We identify key challenges in the field, such as computational efficiency and identity shortcut problem, and propose promising research directions including efficient architectures, novel conditioning strategies, hybrid approaches integrating traditional techniques, and LLMs with diffusion-based frameworks. This roadmap aims to accelerate progress in this rapidly evolving field.

Organization. As illustrated in Fig. 2, the organization reflects the multifaceted nature of ADGDM research, highlighting both the core theoretical underpinnings and their practical applications across diverse domains. The remainder of this survey is structured as follows. Sec. 2 establishes the foundations of ADGDM, covering diffusion models, anomaly scoring mechanisms, and conditioning strategies. Sec. 3 and Sec. 4 examine image and video AD respectively, addressing modality-specific challenges and methodological advances. Sec. 5 and Sec. 6 explore anomaly detection in time series and tabular data, respectively, while Sec. 7 investigates multimodal anomaly detection. Sec. 8 investigates anomaly generation with diffusion models. Sec. 9 details evaluation metrics and benchmark datasets across different modalities. Sec. 10 discusses open challenges and future research directions, and Sec. 11 concludes this survey.

2 FOUNDATIONS

2.1 DMs Primer

DMs constitute a powerful class of generative models capable of synthesizing high-quality data across various modalities [38]. The underlying principle of DMs involves a two-stage process: a forward diffusion process that gradually corrupts data by adding noise, and a reverse denoising process that learns to recover the original data from this noisy version. The forward process can be mathematically formulated as a stochastic differential equation (SDE) $d\mathbf{x} = f(\mathbf{x}, t)dt + g(t)d\mathbf{w}$ [39] or its discrete-time counterpart, where Gaussian noise is iteratively added to an initial data sample \mathbf{x}_0 according to a predefined schedule β_t [40]. In the discrete-time formulation, the forward process transforms \mathbf{x}_0 through

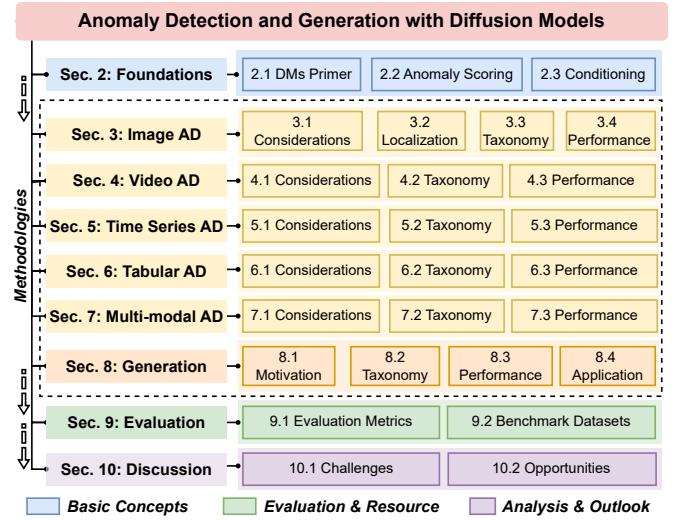


Fig. 2: Organization outline of this survey, which illustrates the article structure, covering basic concepts (Sec. 2), methodologies for AD in different modalities (including image (Sec. 3), video (Sec. 4), time series (Sec. 5), tabular (Sec. 6), and multimodal data (Sec. 7)) and anomaly generation (Sec. 8), evaluation (Sec. 9), and future research outlook (Sec. 10).

a series of steps $q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$, eventually approximating a standard Gaussian distribution \mathcal{N} . Sampling from DMs can be viewed as solving the corresponding SDEs or ordinary differential equations (ODEs) [40], and recent work such as DynGMA [41] explores robustly learning SDEs from data, potentially enhancing the forward process. In the reverse process, a neural network, typically a U-Net [14], is trained to predict either the noise $\epsilon_\theta(\mathbf{x}_t, t)$ or estimate the score function $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$ at each time step, effectively reversing the noise addition. The network learns to predict either the original data or the added noise, guided by maximizing a variational lower bound on the data likelihood [38]. Several DM variants exist, including denoising diffusion probabilistic models (DDPMs) [42] which employ a Markov chain, denoising diffusion implicit models (DDIMs) [40] which offer a non-Markovian formulation for faster sampling, and SDE-based models [39], [43] for a more flexible framework. Additionally, alternative frameworks based on phase space dynamics [44] and methods using neural operators for accelerated sampling [45] have been

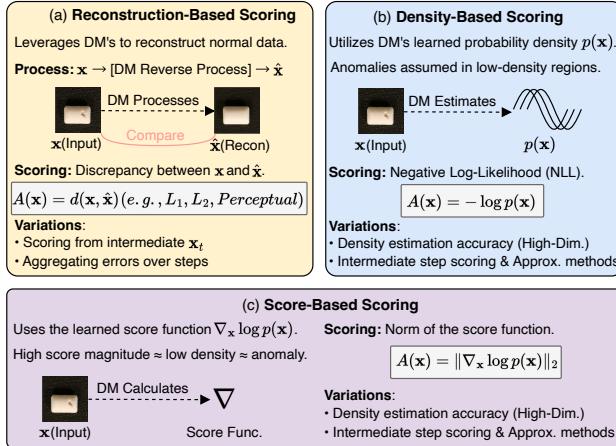


Fig. 3: Overview of anomaly scoring mechanisms with DMs. The diagram illustrates three primary paradigms for calculating anomaly scores: (a) Reconstruction-based scoring computes $A(x) = d(x, \hat{x})$ between input x and its reconstruction \hat{x} ; (b) Density-based scoring uses negative log-likelihood $A(x) = -\log p(x)$ as the anomaly score; (c) Score-based scoring utilizes $A(x) = \|\nabla_x \log p(x)\|_2$, reflecting the data manifold's geometry.

proposed. A mathematical analysis of singularities in DMs provides further insights into the challenges associated with learning complex data distributions, particularly when data lies on lower-dimensional manifolds [46].

2.2 Anomaly Scoring Mechanisms with DMs

Anomaly scoring with DMs leverages their learned understanding of data distributions to identify deviations from normality. As illustrated in Fig. 3, DMs provide three primary paradigms for anomaly scoring.

2.2.1 Reconstruction-Based Scoring

Leveraging the reconstruction capability inherent in diffusion models, anomaly scoring can quantify the discrepancy between an input x and its reconstructed version \hat{x} . The underlying principle posits that a larger reconstruction error signifies a higher likelihood of x being anomalous, stemming from the model's training predominantly on normal data which hinders accurate reconstruction of out-of-distribution samples [47]. As depicted in Fig. 3(a), the process involves passing x through the reverse process to produce \hat{x} , then comparing them. The reconstruction error can be formally expressed as $A(x) = d(x, \hat{x})$, where common distance metrics d include the L_1 norm, L_2 norm, or perceptual loss [48]. Perceptual loss uniquely compares high-level features extracted via pre-trained networks, capturing perceptual similarity rather than pixel-wise differences. Further refinements involve utilizing intermediate diffusion states x_t for scoring rather than just the final reconstruction. To refine scoring, intermediate diffusion states x_t are utilized, capturing diverse data characteristics at varying noise levels [49]. A generalized multi-step scoring approach aggregates errors over different timesteps as $S(x) = \sum t=1^T w_t \mathcal{L}(x, \hat{x}_t)$, where weights w_t prioritize contributions from specific steps [16]. Additionally, methods like dynamic denoising and latent space projections aim to improve reconstruction fidelity, particularly for anomalies of varying scales [25].

2.2.2 Density-Based Scoring

Density-based ADGDM exploits the model's capacity to learn the probability density function of the training data, operating under the premise that anomalies reside in low-density regions. Specifically, an anomaly score is assigned based on the negative log-likelihood: $A(x) = -\log p(x)$, where $A(x)$ represents the anomaly score and $p(x)$ is the estimated probability density of x [50]. A primary advantage of this approach is its interpretability, as the score directly corresponds to the probability density [51]. However, accurately estimating this density, especially in high-dimensional spaces, poses a considerable challenge due to the curse of dimensionality, which necessitates a large number of samples for accurate estimation [52]. In addition, diffusion models often operate in a transformed space, and the estimated density in this space may not accurately reflect the original data space density. Another key challenge stems from the fact that diffusion models primarily learn the score function (the gradient of the log density) rather than the density itself [53]. Estimating the density from the score function can be computationally intensive and error-prone, particularly in high dimensions. Several studies have explored methods to address these challenges, including joint dimensionality reduction and density estimation using low-rank tensor models [54], analyzing score approximation and distribution recovery in low-dimensional data [55], estimating the data manifold dimension [56], and improving density estimation through maximum likelihood training [57].

2.2.3 Score-Based Scoring

Score-based anomaly detection utilizes the learned score function, $\nabla_x \log p(x)$, representing the gradient of the log-probability density. $\nabla_x \log p(x)$ provides crucial information about the local geometry of the probability density in data space [24]. Importantly, a higher score magnitude indicates steeper probability density change rather than directly indicating the probability value itself. The anomaly score is defined as $A(x) = \|\nabla_x \log p(x)\|_2$. The effectiveness stems from geometric properties of normal data manifolds, which typically form regions with smooth probability transitions, while anomalies often lie where density functions exhibit sharp changes. This occurs when: (1) data points fall between normal clusters, causing gradients to point strongly toward nearby modes, or (2) out-of-distribution samples prompt stronger "pulling" toward the learned manifold. During training, the model estimates scores at different noise levels [58] through score matching. At each reverse diffusion step, the score function $s_\theta(x_t, t)$ guides denoising as $x_{t-1} = x_t - \epsilon s_\theta(x_t, t) + \sqrt{2\epsilon}z$, pulling samples toward higher density regions [59]. In AD, this represents restoring potentially anomalous inputs toward normality [22], where persistent high score magnitudes signal anomalous behavior. Empirical studies confirm that while score magnitude and probability density aren't strictly monotonically related, score-based methods effectively identify anomalies through gradient field behavior near anomalous regions.

2.3 Conditioning Strategies

Conditioning strategies are crucial for adapting diffusion models to anomaly detection by shaping the learned distribution and influencing reconstruction behavior. We first

examine unconditional diffusion models (UDMs), noting their simplicity but acknowledging potential limitations with complex data. Subsequently, we explore conditional diffusion models (CDMs), investigating how incorporating additional information, such as class labels, image content, or textual descriptions, improves anomaly detection and enables controllable anomaly generation.

2.3.1 Unconditional DMs

UDMs learn the underlying data distribution without external labels, simplifying training and making them readily applicable to anomaly detection [26]. Specifically, an unconditional DM learns a generative process mapping a standard Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$ to the complex data distribution $p(\mathbf{x})$, where \mathbf{x} represents a data sample. The forward diffusion process gradually adds noise to the data according to a schedule parameterized by β_t , which can be formulated as a Markov chain: $q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I})$, where \mathbf{x}_t is the data sample at time step t . Conversely, the reverse process, also a Markov chain, learns the denoising process as $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t))$, where μ_θ and Σ_θ , parameterized by neural networks with parameters θ , are trained to reverse this noise addition, effectively reconstructing data from noise.

For anomaly detection, unconditional DMs leverage this reconstruction capability [25]. Given a test sample, the model reconstructs its "normal" counterpart by reversing the diffusion process, and the reconstruction error, often measured using L_1 or L_2 norms, serves as the anomaly score. A high reconstruction error indicates a potential anomaly, suggesting the input deviates significantly from the learned distribution of normal data. However, this approach has limitations in capturing complex data structures with intricate dependencies or multimodal distributions [60]. The assumption of a single, unified distribution for normal data may not accurately represent real-world dataset diversity, potentially reducing anomaly detection performance, especially with subtle deviations from normality within specific sub-populations. Additionally, reliance on a global distribution can lead to the "identity shortcut" problem, where the model simply reconstructs the input, even if anomalous, leading to false negatives [22]. Despite these limitations, the simplicity of unconditional DMs makes them a valuable starting point, particularly for scenarios with relatively homogeneous normal data [61].

2.3.2 Conditional DMs

CDMs incorporate external information, denoted as y , to guide the denoising and thus control generated samples. The core modification lies in the reverse process, where the neural network parameterizing the denoising process takes both the noisy input \mathbf{x}_t and the conditioning signal y as inputs, specifically $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t, y) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t, y), \Sigma_\theta(\mathbf{x}_t, t, y))$, where μ_θ and Σ_θ represent the learned mean and covariance, respectively. The conditioning signal y can represent various data types depending on the application. For example, Zhan et al. [62] use class labels as the conditioning signal for multi-class anomaly detection, allowing the model to learn class-specific distributions. Similarly, for image restoration, Mousakhan et al. [63] condition the denoising process on the input image itself to reconstruct a defect-free version. In

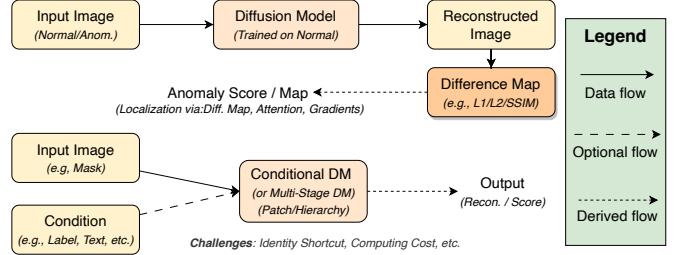


Fig. 4: IAD with (a) reconstruction-based methods and (b) conditional/multi-stage variants.

another approach, Tebbe and Tayyub [25] utilize an initial anomaly prediction as a conditioning signal to guide the dynamic addition of noise during the forward diffusion process. Additionally, Singh et al. [64] explore conditioning on crafted input noise artifacts for controlled image generation with semantic attributes. For controllable anomaly generation, text-conditioned diffusion models, as explored by Sadat et al. [65], leverage textual descriptions to improve the diversity of generated samples. In the context of time series analysis, Pintilie et al. [26] investigate the use of diffusion models for anomaly detection, while Tebbe and Tayyub [22] explore conditioning strategies for improved anomaly localization.

3 IMAGE ANOMALY DETECTION

3.1 Considerations

Applying diffusion models to IAD presents key challenges, particularly the "identity shortcut" problem and substantial computational costs. As illustrated in Fig. 4, the identity shortcut phenomenon occurs when diffusion models trained for high-fidelity reconstruction inadvertently reconstruct anomalous regions in the input, thus masking the very anomalies they should detect. Additionally, the high computational cost, especially during inference due to the iterative nature of the reverse diffusion process, necessitates mitigation strategies such as conditional approaches and multi-stage designs for practical deployment.

The "identity shortcut" in ADGDM manifests when models reconstruct anomalous regions during high-fidelity training, obscuring anomalies and yielding low reconstruction errors despite their presence [23]. Such behavior risks false negatives in anomaly detection. To address this, masked reconstruction leverages learned normal data distributions to enhance sensitivity to deviations [66]. Alternatively, feature editing manipulates latent space representations to reconstruct normal counterparts of anomalous inputs [49]. Adversarial training further strengthens robustness against perturbations, potentially improving anomaly differentiation [67]. Nevertheless, these methods face challenges: masked reconstruction demands precise masking strategies, feature editing requires latent space expertise, and adversarial training incurs high computational costs.

High computational demands hinder DM deployment in real-time anomaly detection, driven by iterative reverse diffusion requiring extensive neural network evaluations [17]. For example, processing high-resolution medical images (2048×2048 pixels) necessitates 50-1000 denoising steps, consuming 15-30 seconds per image on advanced GPUs.

Acceleration strategies address this bottleneck. Progressive distillation optimizes models for faster sampling with fewer steps, balancing speed, and reconstruction quality [68]. Similarly, SLAM [69] and FDM [70] employ linear ODE approximations and momentum-based techniques, respectively. Efficient ODE solvers, such as DPM-Solver [71], streamline the reverse diffusion process. Operating in lower-dimensional latent spaces, LDMs [72] reduce complexity, while LoRA [73] and SparseDM [20] enhance efficiency through model compression and sparse layers.

3.2 Anomaly Localization with DMs

Precise localization of anomalous regions is crucial for applications like industrial defect detection [16] and medical diagnosis [17]. While diffusion models excel at capturing global data distributions, pinpointing anomalies requires specialized techniques, such as generating a difference map between the original and reconstructed image [74]. The magnitude of this difference then serves as the anomaly likelihood measure. In addition, incorporating attention mechanisms within the diffusion model architecture offers finer-grained localization [3], where attention weights highlight anomalous regions during reconstruction. Gradient-based methods provide another strategy [75], identifying anomalous pixels by computing the gradient of the reconstruction loss with respect to the input. Specifically, methods like AnoFPDM [75] leverage the forward diffusion process for brain MRI anomaly segmentation, while D3AD [22] uses dynamic denoising and latent projections to localize multi-scale anomalies, and AnoDDPM [74] employs multi-scale simplex noise diffusion for improved localization accuracy, particularly in medical images.

3.3 Taxonomy and Advances

IAD employing diffusion models can be categorized based on their scoring mechanisms, conditioning strategies, and architectural designs. As summarized in Table 2, we distinguish methods by reconstruction error, density estimation, or score function analysis, analyze trade-offs between unconditional vs. conditional diffusion models, and examine single-stage vs. multi-stage reconstruction approaches.

3.3.1 Reconstruction-based vs. Density/Score-based

Diffusion models for IAD employ reconstruction-based, density-based, and score-based approaches. Reconstruction-based methods, such as THOR [76] and Image-Conditioned Diffusion Models [77], leverage the model's ability to reconstruct a "normal" version of the input, and the discrepancy between the input and its reconstruction, measured by metrics like L_1 , L_2 , or perceptual loss, serves as the anomaly score. Variations include using different noise levels or intermediate reconstructions for scoring like FNDM [17], as well as masked reconstruction to address the identity shortcut problem [78]. Density-based methods, in contrast, utilize the learned probability density function, identifying anomalies by their low probability under this distribution. However, accurately estimating this density in high-dimensional image spaces remains challenging. Consequently, score-based methods offer an alternative by directly using the score function, representing the gradient of the log density, to

efficiently assess anomaly likelihood. In addition, methods like Cold Diffusion [79] explore deterministic degradations, while ensemble methods like Cold-Diffusion Restorations [80] combine multiple restorations for improved scoring.

3.3.2 Unconditional vs. Conditional

Training exclusively on normal data, UDMs [3] identify anomalies through elevated reconstruction errors or diminished likelihood estimates during inference [81]. While offering streamlined training procedures and implementation frameworks, UDMs frequently encounter limitations when processing complex datasets with significant intra-class variation, potentially compromising detection sensitivity for subtle anomalous patterns [82]. In contrast, CDMs incorporate additional information, such as class labels [83], image masks [78], or other modalities [84], during both training and inference [77]. Consequently, CDMs can better capture normal data structure and generate more accurate reconstructions, improving anomaly detection performance [76]. For example, image-conditioned models can be explicitly trained to correct synthetic anomalies, improving localization [85]. Such conditioning also enables targeted anomaly generation [86]. The choice between UDMs and CDMs depends on the specific dataset and application. In complex scenarios like medical imaging, where subtle anomalies are common, CDMs conditioned on image features or anatomical priors might be preferred [17], [87]. Conversely, UDMs may suffice for simpler tasks with less intra-class variation. The selection of appropriate conditioning strategies [88], [89] and consideration of computational costs, especially for high-dimensional data [80], [90], are crucial for CDM effectiveness. Methods like Cold Diffusion may offer computational advantages in certain scenarios [79].

3.3.3 Single-stage vs. Multi-stage

Diffusion models for anomaly detection are primarily categorized into single-stage and multi-stage approaches. Single-stage methods reconstruct the input image directly through a single reverse diffusion process [3], and the resulting reconstruction error, calculated between the input and its reconstruction, serves as the anomaly score [90]. While computationally efficient, such an approach may not effectively capture subtle anomalies or accurately reconstruct complex image structures [82]. In contrast, multi-stage methods, incorporating multiple reconstruction stages, offer a more nuanced approach. One such approach uses patch-based processing, where the image is divided into smaller patches, each independently reconstructed [80], and the individual patch reconstruction errors are aggregated into a final anomaly score, enabling finer-grained anomaly localization [78]. Another multi-stage approach utilizes hierarchical reconstruction at different scales, from coarse to fine [84], which allows the model to capture both global and local anomalies. Additionally, some methods combine different stages or incorporate steps like feature editing or adversarial training to enhance performance [76], [77], including masked reconstruction where a mask forces the model to reconstruct only masked regions [87], thus preventing the "identity shortcut" [83]. Generally, multi-stage approaches achieve better performance, especially for complex anomalies, at the

TABLE 2: Summary of IAD methods across various imaging domains with implementations.

Method	Year	Venue	Imaging Domains	DM Code
Latent DDPM [3]	2022	MICCAI	Brain CT and MRI	✓ -
LDM-OOD [81]	2023	MICCAI	3D medical data	✓ ⚡
FNDM [17]	2023	MICCAI	Brain MR images	✓ -
Cold Diffusion [79]	2023	NeurIPS	General image	✓ -
Mao et al. [88]	2023	ACM MM	Text-image pairs	✓ -
THOR [76]	2024	MICCAI	Brain MRI, Wrist X-rays	✓ ⚡
DDPM-DDIM [87]	2024	IEEE TMI	Brain images	✓ -
mDDPM [78]	2024	MLMI	Brain MRI	✓ ⚡
MMCCD [84]	2024	MICCAIW	Multimodal MRI	✓ -
DAG [80]	2024	MICCAI	Brain MRI	✓ -
IgCONDA-PET [83]	2024	ArXiv	Medical imaging (PET)	✓ ⚡
ODEED [90]	2024	CVPRW	Remote sensing	✓ -
AnomalyDiffusion [85]	2024	AAAI	Industrial images	✓ ⚡
DualAnoDiff [86]	2024	ArXiv	Industrial images	✓ -
Babaei et al. [82]	2024	ArXiv	Medical imaging	✓ -
UNIMO-G [89]	2024	ACL	Text-image pairs	✓ -
ICDM [77]	2025	USMLMI	Medical imaging	✓ ⚡

cost of higher computational overhead compared to single-stage methods [85]. The optimal choice depends on specific application requirements and the trade-off between accuracy and computational efficiency [17].

3.4 Performance Comparison

Diffusion-based approaches for IAD have demonstrated significant performance improvements on standard benchmarks in recent years. As evidenced in Table 3, methods from 2021 to 2025 show a clear upward trajectory in detection metrics across challenging datasets like MVTec AD [91] and VisA [23]. Early approaches such as DRAEM [92] established strong foundations through discriminatively trained reconstruction embeddings for surface anomaly detection, while PaDiM [93] advanced the field through patch distribution modeling techniques. Subsequently, PatchCore [94] achieved competitive results by leveraging memory banks of nominal patch features for efficient inference. Moving beyond traditional approaches, AnoDDPM [74] introduced partial diffusion with simplex noise to outperform Gaussian diffusion models, particularly in medical image applications. Recent innovations include AutoDDPM [95], which enhances robustness through masking and resampling techniques, and Lu et al.’s [91] anomaly removal approach that treats irregularities as noise for improved localization. Notably, GLAD [23] represents the current state-of-the-art through its global-local adaptive diffusion framework, achieving nearly perfect detection scores on the VisA [96]. Throughout the evolution of these methods, evaluation metrics have remained consistent, focusing on AUROC, AUPR, and pixel-level metrics like Dice coefficient and IoU for comprehensive assessment. Comparative analysis across these approaches reveals inherent trade-offs between detection accuracy, localization precision, and computational demands, with certain methods excelling in image-level anomaly detection (e.g., PatchCore on MVTec AD [97]) while others demonstrate superior anomaly localization capabilities (e.g., AnoDDPM [74] on medical images).

4 VIDEO ANOMALY DETECTION

4.1 Considerations

VAD faces distinct challenges compared to image-based methods due to the inherent temporal dimension and

TABLE 3: Performance comparison of key IAD methods on VisA and BTAD datasets.

Method	Year	Venue	Dataset (Metric)	
			VisA (I-AUROC)	BTAD (PRO)
DRAEM [92]	2021	ICCV	88.7, 73.1	-
SPADE [98]	2021	Arxiv	82.1, 65.9	-
PaDiM [93]	2021	ICPR	89.1, 85.9	-
RD4AD [99]	2022	CVPR	96.0, 70.9	94.3, 77.1
PatchCore [94]	2022	CVPR	95.1, 91.2	92.7, 77.3
RAN [91]	2023	ICCV	-	93.4, 78.7
Tebbe et al. [25]	2024	CVPRW	96.0, 94.1	95.2, 83.2
GLAD [47]	2025	ECCV	99.5, 98.6	-

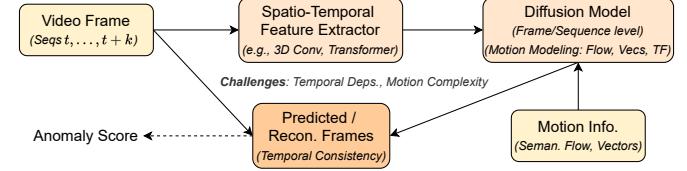


Fig. 5: VAD incorporating spatio-temporal features and motion modeling.

complex motion patterns [100]–[102]. As illustrated in Fig. 5, effective VAD frameworks must process sequential frames while addressing temporal dependencies in motion, where anomalies can manifest as unusual action sequences or deviations from established patterns. The pipeline integrates spatio-temporal feature extractors with specialized diffusion models that incorporate motion information through flow vectors or transformers, enabling the detection of irregularities in both spatial appearance and temporal evolution [103]. Motion’s complexity, encompassing variations in speed, direction, and acceleration, requires accurate capture and representation for effective VAD, with sudden changes or unexpected trajectories potentially indicating anomalies [12]. The interplay between spatial and temporal information further complicates matters, where an action considered normal in isolation might be anomalous given surrounding events. To address these challenges, diffusion models for VAD incorporate key adaptations, including explicitly modeling temporal Dependencies using spatio-temporal transformers, such as in VDT [104], to capture long-range dependencies and temporal context. Additionally, incorporating motion information directly into the diffusion process, exemplified by spectral motion alignment (SMA) [105] and MoVideo [106], refines motion dynamics and preserves temporal consistency. Conditioning on past frames or motion representations allows the model to learn expected temporal evolution and identify deviations [13]. Similarly, techniques like sampling space truncation and robustness penalty in VIDM [107] improve video quality and anomaly detection robustness.

4.2 Taxonomy and Advances

As shown in Table 4, VAD methods employ diffusion models across surveillance to disaster detection domains, categorized by processing granularity (frame vs. sequence-level), motion modeling approaches, and conditioning mechanisms. We examine frame-level analysis treating videos as image collections versus sequence-level approaches modeling tem-

TABLE 4: Summary of VAD with learning paradigms.

Method	Year	Venue	Type	Domain	DM
Tur et al. [103]	2023	IEEE ICIP	F	Video surveillance	✓
Awasthi et al. [108]	2024	IEEE MMSP	F	Satellite imagery, disaster detection	✓
Vide LDM [109]	2023	CVPR	S, M	Driving simulation	✓
VidRD [110]	2023	Arxiv	S	Text-to-video generation	✓
FPDM [111]	2023	ICCV	F	Video surveillance	✓
Stable Video Diffusion [112]	2023	Arxiv	S, M	Text/Image-to-video generation	✓
DMAD [113]	2023	CVPR	F	Surveillance VAD	✗
Wang et al. [114]	2023	Neurocomput.	F	Video surveillance	✓
Masked Diffusion [115]	2023	Arxiv	C	Video procedure planning	✓
GD-VDM [116]	2023	Arxiv	S	Complex scene video generation	✓
AADiff [117]	2023	Arxiv	C	Audio-aligned video synthesis	✓
Basak et al. [118]	2024	ESWA	F	Video surveillance	✓
DHVAD [119]	2024	IJCAI	F	Security and surveillance	✓
DiffVAD [21]	2024	IJCAI	F	Sustainable cities management	✓
GICISAD [120]	2024	Arxiv	M	Skeleton-based VAD	✓
VADiffusion [121]	2024	Arxiv	F	Security surveillance	✓
Cai et al. [122]	2024	Sensors	F	Dust pollution monitoring	✓
FedDiff [123]	2024	IEEE TCSVT	C	Multi-modal remote sensing	✓

Notes: "Type" refers to learning paradigm: F=Frame-level, S=Sequence-level, M=Motion Modeling, C=conditioning Strategies.

poral dynamics, investigate motion incorporation through optical flow and spatio-temporal architectures, and analyze conditioning strategies using temporal context.

4.2.1 Frame-level Methods

VAD methods analyze video data at different granularities, primarily frame-level and sequence-level. Frame-level methods analyze individual frames independently, effectively treating videos as image collections and enabling direct application of image-based anomaly detection techniques using diffusion models, but consequently disregarding temporal dependencies. In contrast, sequence-level methods explicitly consider temporal relationships between frames, facilitating a more comprehensive understanding of normal behavior and the detection of anomalies evolving over time. Specifically, some approaches leverage recurrent networks [124] or spatio-temporal transformers within the diffusion model framework to capture temporal dependencies [125]. Other methods reconstruct entire video sequences or segments, using the temporal coherence of the generated output as a normality measure [126]. While offering improved accuracy by incorporating temporal context, sequence-level methods often incur increased computational complexity compared to frame-level approaches [26]. The choice between these methods depends on the specific application and the trade-off between accuracy and computational cost [127], as well as the nature of the anomalies themselves; for example, sudden object appearances might be detectable at the frame level, while unusual motion patterns require sequence-level analysis [60]. Researchers have also explored the concept of a "normal gathering latent space" for enhanced anomaly detection in time series data using diffusion models and incorporated decontamination techniques within the diffusion model framework to address contaminated training data [60], [128]. Converting time-series data into images before applying diffusion models has also been proposed for network traffic generation and anomaly detection [129].

4.2.2 Motion Modeling

Motion cues are crucial for distinguishing anomalies in video data, thus integrating motion information into diffusion models is essential for robust VAD [130]–[132]. Optical flow, representing the apparent motion of objects or surfaces, can condition diffusion models [133], enabling them to learn normal motion patterns and identify deviations. Similarly, motion vectors, capturing pixel displacement between

frames, can be utilized for motion-aware anomaly detection. Researchers have also explored spatio-temporal transformers [134], [135] to capture long-range dependencies and complex motion dynamics, leading to more nuanced representations of normal behavior. Integrating structured state space models (SSSMs) with DMs [61], [136] offers another promising approach to capturing long-term temporal dependencies. In addition, incorporating collector entity ID embedding [61] can enhance pattern learning by distinguishing different collection signals, while adaptive dynamic neighbor masking mechanisms, used with transformers and denoising diffusion models [126], can mitigate information leakage during reconstruction, improving anomaly identification.

4.2.3 Conditioning Strategies

Conditioning strategies are crucial for adapting diffusion models to VAD, particularly for leveraging temporal context and motion information. One common approach is conditioning on past frames, which enables the model to learn temporal dependencies and predict future frames, thereby identifying deviations from expected motion patterns, as explored in TimeDiT [134]. Incorporating motion representations, such as optical flow or motion vectors, can further enhance the model's sensitivity to subtle anomalies related to unusual movements. Alternatively, conditioning on semantic features extracted from the video content can provide higher-level contextual information, helping differentiate between normal and anomalous events based on scene understanding. Hybrid conditioning strategies, combining past frames with motion representations or semantic features, leverage both low-level motion cues and high-level context. For example, ProDiffAD [137] uses inter-cloud network conditions for enhanced anomaly detection, while DDMT [126] employs an adaptive dynamic neighbor mask (ADNM) to mitigate information leakage. Approaches like NGLS-Diff [60] operate within a normal gathering latent space for improved anomaly detection in normal time series data, and TSAD-C [128] utilizes spatio-temporal graph conditional diffusion models to address contaminated training data. Similarly, SSSD [136] leverages structured state space models for improved imputation and forecasting by capturing long-term dependencies.

4.3 Performance Comparison

Emerging research in VAD increasingly leverages diffusion models to capture complex temporal dynamics and generate high-fidelity video reconstructions. As shown in Table 5, contemporary methods demonstrate varying performance across standard benchmark datasets, with approaches like DMAD [113] achieving exceptional results on UCSD Ped2 (99.7% AUC) and CUHK Avenue (92.8% AUC). Despite the promising capabilities of diffusion-based techniques, directly comparing model performance remains challenging due to inconsistent experimental protocols and reporting standards across studies. While traditional VAD research has predominantly employed autoencoders and generative adversarial networks, diffusion-based approaches like VADiffusion [121] show competitive performance by employing a dual-branch architecture that combines motion vector reconstruction with I-frame prediction guided by compressed domain information. Such integration of diffusion principles enhances

TABLE 5: Performance (AUC) comparison of key VAD methods on UCSD Ped2, CUHK Avenue, and ShanghaiTech.

Method	Year	Venue	Dataset		
			UCSD Ped2	CUHK Avenue	ShanghaiTech
RTFM [141]	2021	ICCV	96.3	85.1	73
DMAD [113]	2023	CVPR	99.7	92.8	78.8
LSH [142]	2023	IEEE TCSV	91.3	87.4	77.6
VADNet [143]	2023	IEEE TCSV	-	87.3	75.2
UMIL [144]	2023	CVPR	-	88.3	-
UR-DMU [27]	2023	AAAI	-	88.9	97.92
VADiffusion [121]	2024	IEEE TCSV	98.2	87.2	71.7
PLOVAD [145]	2025	IEEE TCSV	-	-	97.98

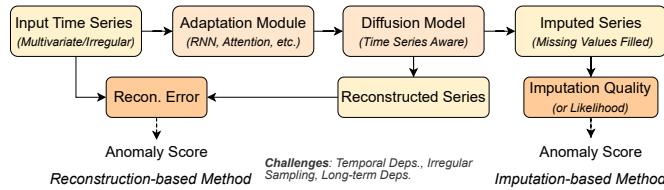


Fig. 6: TSAD with reconstruction and imputation paths.

framework stability and detection accuracy, particularly when addressing both sudden and persistent anomalies. Nevertheless, computational overhead presents a significant concern for high-resolution video processing, creating an inevitable trade-off between detection accuracy and operational efficiency. Methods such as ReFLIP-VAD [138] and VadCLIP [19], though not directly via DMs, further highlight challenges in adapting complex vision-language architectures for video analysis despite their strong performance on datasets like XD-Violence [139] and UCF-Crime [140].

5 TIME SERIES ANOMALY DETECTION

TSAD with DMs faces distinct challenges related to the inherent temporal structure of sequential data. As shown in Fig. 6, effective TSAD frameworks must address temporal dependencies, irregular sampling rates, and long-term relationships through specialized adaptation modules incorporating recurrent networks and attention mechanisms. The figure highlights two primary approaches in diffusion-based TSAD: reconstruction-based methods that compute anomaly scores by comparing original and reconstructed series, and imputation-based methods that evaluate anomaly likelihood through the quality of imputed missing values. Both approaches utilize specialized time series-aware diffusion models that capture the complex dynamics of sequential data, enabling more nuanced anomaly detection compared to traditional methods.

5.1 Considerations

Due to inherent temporal dependencies and often irregular sampling, TSAD sometimes exhibits long-term dependencies requiring models to capture relationships across extended periods [146]. While traditional methods often struggle with these characteristics, diffusion models offer a promising approach with appropriate adaptations. One key adaptation incorporates recurrent neural networks (RNNs), such as LSTMs, designed to process sequential data and capture temporal dependencies [35]. Another crucial adaptation leverages attention mechanisms [3], allowing models to

TABLE 6: Summary of TSAD with learning paradigms.

Method	Year	Venue	Learing Paradigm	DM
D ³ R [125]	2023	NeurIPS	Decompo. and recon.	✓
DDMT [126]	2023	ArXiv	Mask-based	✓
DiffAD [127]	2023	KDD	Imputation-based	✓
SSSD [136]	2023	ArXiv	Imputation and forecasting	✓
Diffusion+ [151]	2023	ESEC / FSE	Imputation-based	✓
SaSDim [133]	2023	ArXiv	Noise-scaling diffusion	✓
Pintilie et al. [26]	2023	ICDMW	Diffusion-based	✓
NetDiffus [129]	2024	COMNET	Time-series imaging	✓
NGLS-Diff [60]	2024	ECML PKDD	Latent space diffusion	✓
TimeADMM [124]	2024	ICASSP	Diffusion-based	✓
Zuo et al. [61]	2024	APIN	Reconstruction-based	✓
TimeDiT [134]	2024	ICMLW	Foundation model	✓
TSDE [135]	2024	ArXiv	Self-supervised learning	✓
Chen et al. [146]	2024	IEEE TCC	Reinforcement learning	✗
ProDiffAD [137]	2024	IJCNN	Progressive distillation	✓
TSAD-C [128]	2024	ArXiv	Graph conditional diffusion	✓

focus on relevant parts of the time series when reconstructing or scoring anomalies, effectively handling long-term dependencies and irregular sampling. For example, TimeDiT [134] utilizes a transformer architecture with attention and a diffusion process for generating high-quality samples. Similarly, CAT [147] employs a moment network with reinforcement learning to address irregular sampling by seeking relevant moments within the continuous timeline. Progressive learning paradigms [14] offer an additional strategy for complex time series by gradually increasing data complexity and network capacity to capture long-term trends and short-term variations. Addressing missing data, a common issue in real-world time series, is also crucial, with methods like those in [148] demonstrating the capability of handling such data without explicit structural priors. Incorporating temporal priors into self-attention [149] can further enhance a model's ability to learn from the arrow of time, as exemplified by Triformer [150], which introduces a triangular attention mechanism with linear complexity for efficient processing of long sequences and capturing distinct temporal dynamics.

5.2 Taxonomy and Advances

Diffusion models for TSAD are primarily categorized into reconstruction-based and imputation-based methods, as shown in Table 6. Reconstruction-based methods exploit the generative capabilities of diffusion models to reconstruct input time series, identifying anomalies as instances with high reconstruction errors [124]. For example, DDTAD [28] uses a 1-D U-Net within a DDPM framework for time series reconstruction and anomaly detection. Approaches like D³R [125] and NGLS-Diff [60] further enhance this paradigm; D³R incorporates dynamic decomposition to handle drifting time series, while NGLS-Diff operates within a learned latent space of normal temporal patterns. DDMT [126] integrates a denoising diffusion model with a Transformer architecture and an adaptive dynamic neighbor mask mechanism.

In contrast, imputation-based approaches use anomaly detection as a framework for the missing value imputation task [148]. DiffAD [127] leverages a density ratio strategy and a conditional weight-incremental DMs for improved anomaly detection, particularly with concentrated anomalous points. Similarly, SSSD [136] combines conditional diffusion models with structured state space models for imputation and

TABLE 7: Performance comparison of key TSAD methods on SWaT, WADI, MSL, and SMD datasets.

Method	Year	Venue	Dataset (Metrics)			
			SWaT (P, R, F1, F1_PA)	WADI (P, R, F1, F1_PA)	MSL (P, R, F1, F1_PA)	SMD (P, R, F1, F1_PA)
DAGMM [153]	2018	ICLR	27.46, 69.52, 39.37, 85.33	54.44, 26.99, 36.09, 61.65	25.91, 62.86, 36.69, 70.09	42.59, 50.46, 26.85, 72.29
OmniAnomaly [156]	2019	ACM KDD	98.25, 64.97, 78.22, 86.61	49.47, 12.98, 22.08, 41.72	16.19, 84.66, 27.18, 89.94	20.61, 46.73, 28.20, 75.29
AnomalyTran [156]	2022	ICLR	12.00, 100.00, 21.43, 94.07	~7.79, 43.43, 10.21, 89.10	~, 2.10, 93.58	~, 2.11, 92.33
TransAD [149]	2022	VLDL	97.94, 60.52, 74.16, 91.04	86.68, 15.50, 26.00, 42.04	29.06, 75.96, 42.04, 94.94	26.95, 57.36, 37.16, 76.35
D ³ R [125]	2023	NIPS	11.72, 10.04, 89.09, 21.49	12.23, 11.75, 46.11, 11.04	93.01, 19.74, 87.44, 23.37	52.20, 24.71, 45.49
PUDAD [154]	2023	ICML	97.94, 60.52, 74.16, 91.04	86.68, 15.50, 26.00, 42.04	29.06, 75.96, 42.04, 94.94	26.95, 57.36, 37.16, 76.35
NPSR [155]	2023	NeurIPS	93.42, 75.52, 83.52, 91.07	78.43, 50.53, 61.31, 75.22	24.03, 83.92, 37.37, 96.45	26.58, 62.36, 37.27, 76.35
SensitiveHUE [157]	2024	ACM KDD	94.68, 87.74, 91.08, 96.75	85.51, 58.73, 69.96, 92.25	33.05, 71.26, 45.16, 98.42	29.54, 60.80, 39.76, 96.33

Notes: P, R, F1, and F1_PA refer to Precision, Recall, F1-score, and Point-Adjusted F1-score.

forecasting. In practical applications, Diffusion+ [151] focuses on efficient imputation for cloud failure prediction, and SaS-Dim [133] introduces a self-adaptive noise scaling diffusion model for spatial time series imputation. General-purpose diffusion transformer models like TimeDiT [134] and TSDE [135] demonstrate the versatility of diffusion models for various time series tasks, including imputation and anomaly detection. Addressing contaminated training data, TSAD-C [128] incorporates a decontamination module within a spatio-temporal graph conditional diffusion model framework.

5.3 Performance Comparison

Originally developed for image generation, DMs are increasingly being adapted for TSAD, leveraging their capability to capture complex temporal dependencies and generate high-fidelity synthetic data [152]. As evidenced in Table 7, methods demonstrate significant performance improvements across challenging benchmarks such as SWaT, WADI, MSL, and SMD datasets. Early approaches like DAGMM [153] established foundational techniques using deep generative models, while subsequent methods such as OmniAnomaly [152] advanced the field through more sophisticated probabilistic modeling. Architectural adaptations addressing specific challenges of time series data have emerged, with researchers incorporating recurrent networks and attention mechanisms to handle irregular sampling and long-term dependencies [125]. Various conditioning strategies have been explored, including prototype-based methods [154] that achieve notable precision-recall balance on industrial datasets, and approaches conditioning on "nominality scores" [155] that demonstrate improved performance in point-adjusted metrics. Applications span diverse domains from network monitoring [156] to industrial fault detection [50], with generative capabilities being leveraged for data augmentation and robustness enhancement [149]. Recent innovations such as D³R [125] incorporate dynamic decomposition to handle unstable multivariate time series with data drift, while SensitiveHUE [157] represents the current state-of-the-art by enhancing sensitivity to normal patterns, achieving remarkable F1 scores (91.08% on SWaT) and point-adjusted metrics (98.42% on MSL). Comparative analysis across these approaches reveals an inherent trade-off between precision and recall, with certain methods excelling in different application domains based on their specific architectural characteristics and conditioning strategies.

6 TABULAR ANOMALY DETECTION

Tabular data presents unique challenges for anomaly detection, primarily due to the inherent nature of mixed data types (numerical, categorical, ordinal) and the frequent presence of missing values. As illustrated in Fig. 7, effective TAD



Fig. 7: TAD handling mixed data types.

TABLE 8: Summary of TAD methods with type and metrics.

Method	Year	Venue	Type	Performance Metrics	DM
TabADM [164]	2023	ArXiv	Diffusion-based	Detection accuracy	✓
MPDR [165]	2023	NeurIPS	Energy-based	AUPR, AUROC	✗
Tritscher et al. [166]	2024	xAI	Generative inpainting	Explanation quality	✓
TimeAutoDiff [162]	2024	ArXiv	VAE + DDPM hybrid	Fidelity and utility metrics	✓
Dai et al. [167]	2025	AAAI	Noise evaluation	AUC score	✗
CoDi [168]	2023	ICML	Co-evolving diffusion	Synthesis quality	✓
SimpDM [169]	2024	CIKM	Self-supervised diffusion	Imputation accuracy	✓
FinDiff [170]	2023	ICAF	Diffusion-based	Fidelity, privacy, utility	✓
PHAD [171]	2025	IPM	Hypergraph representation	Detection accuracy	✓
Thimonier et al. [172]	2024	CIKM	Retrieval-augmented	Detection performance	✗

frameworks must process heterogeneous input data through specialized preprocessing and embedding techniques before applying tabular-adapted diffusion models, which generate reconstructions that can be compared with the original input to compute anomaly scores. The figure highlights how reconstruction or generation loss serves as the primary mechanism for anomaly detection, while addressing key challenges specific to tabular data throughout the pipeline. Consequently, researchers have developed specific data preprocessing techniques and specialized diffusion model architectures to address these issues.

6.1 Considerations

Mixed data types in tabular datasets require specialized modeling frameworks since conventional diffusion approaches poorly handle categorical variables. DiSK [158] employs Gaussian mixture modeling for simultaneous processing of text, categorical, and continuous data, while TabSynDex [159] converts diverse data types into unified latent representations. Missing values further complicate modeling, with MissDiff [160] addressing this limitation by masking denoising score matching regression loss during training. DiffImpute [161] leverages harmonization techniques for enhanced coherence between observed and imputed data. Computational efficiency constraints have led researchers toward using gradient-boosted trees for score function learning and implementing accelerated sampling methods like TimeAutoDiff's parallel generation approach [162]. For imbalanced feature distributions, researchers have developed fair DMs generating balanced data while maintaining sample quality [29]. Architectural adaptations include specialized denoising networks with Transformers showing superior performance [161], while latent diffusion approaches combine VAEs with diffusion models to optimize tabular data representation [163]. For time series tabular data, structured state space models effectively capture temporal dependencies, complemented by appropriate preprocessing techniques tailored to domain-specific patterns [136].

6.1.1 Taxonomy and Advances

Modern approaches to TAD with generative models exhibit distinct methodological divisions, as illustrated in Table 8, which categorizes recent techniques across multiple

TABLE 9: Performance comparison of key TAD models on 47 real-world tabular datasets, including domains such as healthcare, image processing, and finance.

Method	Year	Venue	Avg. performance on 47-dataset		
			AUC (%)	Mean Rank	p-value
NeuTraLAD [173]	2021	ICML	71.45 ± 22.6	9.16	0
SCAD [174]	2022	ICLR	69.15 ± 15.3	6	0.0002
ECOD [175]	2023	IEEE TKDE	52.77 ± 11.7	10.12	0
PLAD [176]	2022	NeurIPS	67.42 ± 19.6	9.16	0
DPAD [177]	2024	Arxiv	88.05 ± 12.4	4.96	0.0003
AutoUAD [178]	2024	Arxiv	92.68 ± 11.8	2.04	0.47
DNN [167]	2025	AAAI	92.27 ± 11.1	1.68	0.47

dimensions including architectural paradigms and evaluation frameworks. Among reconstruction-based methods, TabADM [164] leverages diffusion principles to model normal data distributions and identify anomalies through reconstruction errors, establishing a foundational approach for diffusion-based tabular analysis. Complementary to this paradigm, generation-based techniques such as FinDiff [170] and CoDi [168] focus on synthesizing realistic tabular data, enabling broader applications spanning data augmentation and comparative anomaly detection between generated and real instances. Beyond standard diffusion approaches, self-supervised paradigms exemplified by SimpDM [169] enhance imputation accuracy and robustness when handling sparse or missing values—a prevalent challenge in tabular domains. Integration with complementary techniques further diversifies the methodological landscape, with PHAD [171] merging diffusion-based augmentation with hypergraph representation learning to capture higher-order relationships, while retrieval-augmented methods [172] leverage retrieved samples for enhanced anomaly identification. Architectural innovations continue to emerge through hybrid approaches like TimeAutoDiff [162], which synergistically combines autoencoder frameworks with DMs to enhance data synthesis by capitalizing on the strengths of both paradigms.

6.1.2 Performance Comparison

Direct application of image-based DMs to tabular data presents challenges due to inherent structural differences, including mixed data types and missing values—obstacles that necessitate specialized adaptations and preprocessing techniques. As evidenced in Table 9, TAD approaches demonstrate considerable performance variations across 47 real-world datasets spanning healthcare, finance, and image processing domains. Earlier methods like NeuTraLAD [173] and SCAD [174] established foundational techniques with AUC scores of 71.45% and 69.15% respectively, while recent innovations including DPAD [177] and AutoUAD [178] show substantial improvements, achieving 88.05% and 92.68% AUC scores. Related work in multivariate time series demonstrates competitive performance through specialized architectures that capture complex dependencies in structured data [26]. Advanced frameworks incorporating dynamic step size computation and latent space projection [25] enhance both detection accuracy and anomaly localization capabilities. Applications span diverse domains from fraud detection to intrusion monitoring, leveraging diffusion models' ability to learn complex distributions for

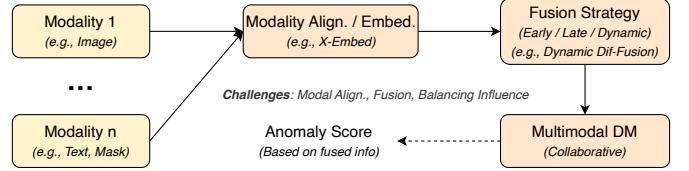


Fig. 8: MAD with early/late/dynamic fusion strategies.

identifying unusual patterns in tabular structures. Notably, few-shot frameworks like AnomalySD [179] utilizing Stable Diffusion have demonstrated particular effectiveness in industrial settings, achieving high performance with minimal labeled examples. Comparative analysis across these approaches reveals the importance of domain-specific adaptations, with methods like DNN [167] achieving state-of-the-art performance (92.27% AUC) through specialized architectures tailored to tabular data characteristics.

7 MULTIMODAL ANOMALY DETECTION

By leveraging the complementary nature of diverse data sources, multimodal approaches significantly enhance the accuracy and robustness of anomaly detection systems. Fig. 8 illustrates how these frameworks process multiple input modalities through alignment and embedding techniques before applying fusion strategies (early, late, or dynamic) within collaborative diffusion models. The diagram demonstrates data flow while highlighting critical challenges including modal alignment, fusion techniques, and influence balancing.

7.1 Considerations

MAD introduces challenges due to diverse data sources. Modal alignment is challenging as different modalities occupy distinct feature spaces with varying scales and distributions [180]. Effective fusion techniques are crucial for capturing modality interactions, as simple concatenation may not exploit their complementary nature [181]. Additionally, balancing the influence of modalities is essential, as some may be more informative for specific anomaly types [182]. To address these challenges, collaborative diffusion frameworks enable joint learning of representations across modalities, promoting alignment and facilitating information exchange [183]. Multimodal embedding spaces provide a common ground for representing and comparing information, mitigating the alignment issue, as exemplified by AnomalyXFusion's X-embedding approach, which integrates image, text, and mask features [184]. Dynamic fusion mechanisms, such as AnomalyXFusion's Dynamic Dif-Fusion module, offer adaptive weighting of modalities based on their relevance, adjusting the contribution of multimodal features during diffusion steps for context-aware generation [185]. Ongoing research on cross-modal feature mapping and quality-aware fusion further enhances the robustness and accuracy of these systems [180], [182]. Deep structured anomaly detection frameworks [186] and the incorporation of semantic guidance in diffusion models [187] also contribute to handling the complexities of multimodal data.

7.2 Taxonomy and Advances

Recent research in anomaly detection increasingly exploits complementary information from diverse data sources to

TABLE 10: Summary of MAD methods on different datasets.

Method	Year	Venue	Datasets	DM Code
Multimodal LDM [190]	2024	Ital-IA	KSDD2	✓ -
Ano-cDiff [188]	2024	ESWA	Brain glioma datasets	✓ ✓ 
Flaboreau et al. [191]	2023	ICCV	UBnormal, HR-UBNormal, etc.,	✓ -
AnomalyXFusion [183]	2024	ArXiv	MVTec AD, LOCO, MVTec Caption	✓ 
Collaborative Diffusion [189]	2023	CVPR	CelebAMask-HQ, CelebA-Dialog	✓ -

enhance detection accuracy and robustness. As shown in Table 10, recent multimodal approaches leverage various fusion strategies tailored to specific application domains and data characteristics. Early fusion methods, exemplified by FedDiff [123], combine modalities at the input level, integrating features before model processing. In contrast, late fusion approaches like Ano-cDiff [188] process each modality independently before combining outputs, preserving modality-specific characteristics. Dynamic fusion strategies, represented by Collaborative Diffusion [189], adaptively weight modality contributions based on input characteristics, optimizing information extraction from heterogeneous sources. Multiple modality combinations have demonstrated effectiveness across application domains, with image-text pairings [190] enhancing industrial defect detection through expert knowledge integration, video-audio combinations [117], [191] improving surveillance anomaly recognition, and image-mask pairings [183] facilitating precise region localization. Notable architectural innovations include X-embeddings [183] that unify features from multiple modalities into coherent embedding spaces, and human-in-the-loop systems [190] that incorporate expert knowledge through text descriptions and region localization, substantially enhancing model performance on datasets like KSDD2. The growing availability of open-source implementations for methods like AnomalyXFusion and Ano-cDiff further accelerates progress in this rapidly evolving research area.

7.3 Applications

MAD with DMs offers substantial advancements in Industry 5.0, especially for defect detection. This approach leverages diffusion models to integrate diverse data modalities like images and text, providing a more comprehensive understanding of industrial processes and facilitating accurate anomaly identification. Approaches combining expert knowledge with visual data show particular promise. For example, multimodal LDMs incorporate text descriptions and region localization [190], enabling more targeted anomaly detection. In these models, experts provide textual descriptions and localize potential anomalies within images, guiding the model towards learning relevant and interpretable defect representations [190]. Consequently, the integration of human expertise enhances the model's ability to detect subtle anomalies often missed by purely data-driven methods, and facilitates a robust feedback loop for iterative refinement and improved accuracy [190]. Similarly, the Myriad model [192] uses an Expert Perception module to embed prior knowledge as tokens understandable by LLMs [193]. Adopting this neuro-symbolic strategy facilitates the generation of granular anomaly descriptions encompassing attributes like color, shape, and category [192], which is crucial for effective decision-making in industrial settings.

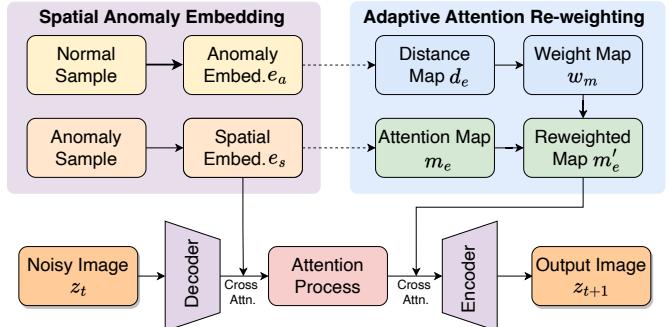


Fig. 9: Overview of the AnomalyDiffusion framework.

8 ANOMALY GENERATION

8.1 Motivation

A key motivation for generating synthetic anomalies with DMs stems from the need to augment limited anomaly datasets. As shown in Fig. 10, AG methods typically take normal data as input seed, along with optional conditioning information (e.g., text descriptions or masks), and employ guided DMs to produce synthetic anomalies across various data modalities. The figure highlights how models like CUT [194] use latent space manipulation and controlled generation to create realistic anomalies for multiple purposes. Anomalies are inherently infrequent in real-world scenarios, resulting in imbalanced datasets that hinder the training of robust anomaly detection models. Consequently, DMs with their ability to learn complex data distributions, offer a promising avenue for generating realistic synthetic anomalies, thus addressing the scarcity of real-world examples.

In addition, data augmentation can improve the performance and generalization of anomaly detection models, especially in few-shot learning scenarios. Generating synthetic anomalies with varied characteristics and severities also allows researchers to stress-test anomaly detection models and identify potential vulnerabilities, such as limitations in detecting anomalies of different sizes, shapes, and locations [22], [25]. As illustrated in Fig. 9, frameworks like AnomalyDiffusion [85] implement sophisticated architectures that integrate spatial anomaly embedding modules and adaptive attention re-weighting mechanisms to precisely control anomaly generation. Moreover, systematically varying generated anomalies enables researchers to assess model performance under different anomaly conditions, which can inform the development of more robust anomaly detection systems, particularly in critical applications like medical imaging [195]. Finally, the generation of synthetic anomalies facilitates the development of self-supervised anomaly detection methods [16], [196] by training models on a mixture of normal data and synthetic anomalies to learn discriminative features that distinguish normal from anomalous patterns without explicit labels.

8.2 Taxonomy and Advances

Various diffusion-based approaches for anomaly generation have emerged in recent years, each addressing different aspects of the generation challenge with distinct methodological foundations. As summarized in Table 11, these methods can be classified according to their underlying techniques and application domains. Manipulation-based approaches

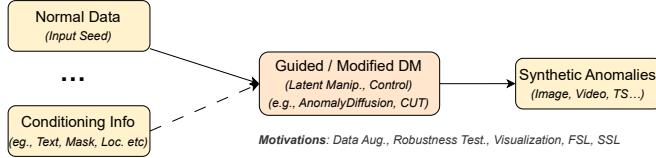


Fig. 10: AG techniques with guided DMs for data augmentation, robustness testing, and self-supervised learning.

TABLE 11: Summary of AG methods across various domains with implementations.

Method	Year	Venue	Domain	DM Code
HumanRefiner [199]	2024	ECCV	Human image generation	✓ ⚡
AnomalyDiffusion [85]	2024	AAAI	Industrial inspection	✓ ⚡
DualAnoDiff [86]	2024	ArXiv	Industrial inspection	✓ ⚡
AdaBLDM [197]	2024	ArXiv	Industrial defect generation	✓ ⚡
CUT [194]	2024	ArXiv	Visual anomaly detection	✓ ⚡
Rai et al. [200]	2024	CVPR	Video anomaly detection	✓ -
NSA [201]	2022	ECCV	Manufacturing, Medical imaging	✗ ⚡
CutPaste [198]	2021	CVPR	Industrial inspection	✗ -
PRN [202]	2023	CVPR	Industrial manufacturing	✗ -
Fontanella et al. [87]	2024	IEEE TMI	Medical imaging (Brain)	✓ -
FinDiff [170]	2023	ACM ICAIF	Financial data	✓ -
NetDiffus [129]	2024	COMNET	Network traffic analysis	✓ -

directly alter the diffusion process to produce anomalous samples, while conditional generation methods leverage additional information to guide the generation toward specific anomaly types. For example, AnomalyDiffusion [85] and DualAnoDiff [86] utilize specialized conditioning strategies for industrial inspection applications, whereas CUT [194] offers a more general framework for visual anomaly detection. Latent space modification represents another significant category, where techniques modify the underlying representation space to create anomalies, contrasting with data-space approaches that operate directly on raw input. Furthermore, the integration of diffusion models with other generative or discriminative architectures has gained traction, as exemplified by AdaBLDM [197], which enhances the realism and diversity of generated anomalies. Beyond visual domains, applications have expanded to specialized fields such as financial data analysis with FinDiff [170] and network traffic analysis through NetDiffus [129], demonstrating the versatility of diffusion models for anomaly generation across diverse domains. The evolution from earlier techniques like CutPaste [198] to sophisticated diffusion-based frameworks indicates significant progression in generating increasingly realistic and controllable anomalies.

8.3 Performance Comparison

Leveraging the powerful capabilities of DMs, several frameworks have emerged as effective tools for generating synthetic anomalies with remarkable realism and diversity. As shown in Table 12, AnomalyDiffusion [85] demonstrates superior performance over traditional GAN-based methods like DefectGAN [203] and DFMGAN [204], achieving the highest Inception Score (1.8) and IC-LPIPS (0.32) on the MVTec AD dataset. Beyond AnomalyDiffusion, methods like CUT [194] further advance the field by leveraging Stable Diffusion for generating realistic anomalies without retraining, demonstrating universality across unseen data and novel anomaly types, which facilitates targeted anomaly creation for applications like data augmentation. Additionally, conditioning strategies play a crucial role in guiding

TABLE 12: Performance comparison of key AG models on the MVTec dataset using IS and IC-LPIPS metrics.

Method	Year	Venue	Metric	
			IS	IC-L
DefectGAN [203]	2021	WACV	1.69	0.15
SDGAN [206]	2020	IEEE TASE	1.71	0.13
DiffAug [207]	2020	NeurIPS	1.58	0.09
Crop&Paste [208]	2021	IEEE ICME	1.51	0.14
DFMGAN [204]	2023	AAAI	1.72	0.2
CDC [209]	2021	CVPR	1.65	0.07
AnomalyDiffusion [85]	2024	AAAI	1.8	0.32

the generation process. For example, SDAS in RealNet [205] employ a diffusion-based synthesis strategy specifically designed to mimic the distribution of real anomalies, creating a diverse set of synthetic examples for training robust anomaly detection models. Moreover, vision-language models [138] can effectively leverage textual descriptions to condition the generation process, enabling fine-grained control over the characteristics of generated anomalies.

8.4 Applications

Anomaly generation with DMs has found diverse applications, notably in data augmentation [85]. In scenarios with limited anomaly samples, synthetically generated anomalies can augment training datasets, consequently improving the performance and robustness of AD models, which is particularly valuable in areas like industrial inspection [23] where acquiring real-world anomalies is challenging. Additionally, generating a wide range of potential anomalies facilitates model robustness testing [95] and allows researchers to evaluate detection models under diverse conditions. Such a process helps identify vulnerabilities and enhances the reliability of anomaly detection systems. Anomaly generation also enables anomaly visualization [26], aiding in understanding anomalous event characteristics and interpreting model predictions. For example, in medical imaging, visualizing generated anomalies can assist clinicians in identifying subtle abnormalities. Furthermore, some methods [22], [25] leverage dynamic noise addition during diffusion, guided by initial anomaly predictions, to improve anomaly localization, particularly for anomalies of varying scales.

9 MODEL EVALUATION

9.1 Evaluation Metrics

Effective ADGDM relies on robust evaluation metrics tailored to specific data modalities. As shown in Fig. 11, evaluation frameworks must address the distinct characteristics of each data type through specialized metrics.

9.1.1 Image Anomaly Detection Metrics

IAD employs evaluation metrics reflecting the task's multifaceted nature, categorized into pixel-level metrics for anomaly localization and image-level metrics for classification [210]. Pixel-level localization relies on the Dice coefficient and Intersection over Union (IoU) to measure overlap between predicted and ground truth anomaly masks [211], while the Hausdorff Distance quantifies maximum boundary distances, proving particularly valuable in medical imaging where precise localization is crucial [212]. For classification

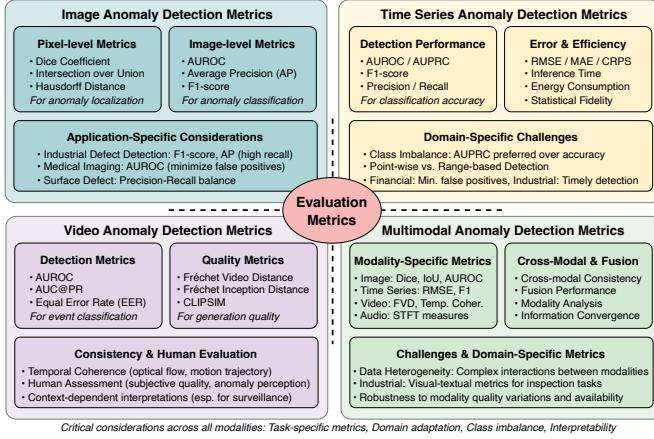


Fig. 11: Comprehensive evaluation metrics for ADGDM across different data modalities.

tasks, AUROC measures the model’s ability to distinguish between normal and anomalous images [213], while Average Precision addresses class imbalance inherent in anomaly datasets [214], and F1-score balances precision and recall to reflect both correct identification and false positive avoidance [215]. Metric selection depends on application requirements, with industrial defect detection prioritizing high recall through F1-score or AP, whereas medical applications favor AUROC to minimize false positives.

9.1.2 Video Anomaly Detection Metrics

VAD evaluation requires assessing both detection accuracy and video quality while maintaining temporal consistency [36]. Detection metrics including AUROC, AUC@PR, and Equal Error Rate quantify distinguishing capabilities between normal and anomalous events [216], with AUC@PR particularly relevant given class imbalance challenges. Video quality assessment employs Fréchet Video Distance, Fréchet Inception Distance, and CLIPSIM to evaluate generated video fidelity through feature distribution comparisons and semantic similarity measures [200]. Temporal coherence metrics ensure smooth frame transitions by measuring optical flow consistency and motion trajectory smoothness [103], while human evaluation provides subjective assessment for subtle or context-dependent anomalies.

9.1.3 Time Series Anomaly Detection Metrics

TSAD metrics encompass detection performance, error measurement, computational efficiency, and statistical fidelity [217]. Detection capabilities utilize AUROC, AUPRC, F1-score, precision, and recall [218], while error metrics like RMSE, MAE, and CRPS quantify prediction accuracy for forecasting-based methods [219]. Computational considerations including inference time and energy consumption prove critical for deployment [105], alongside statistical fidelity metrics ensuring synthetic data reflects underlying distributions [26]. Unique challenges include severe class imbalance where normal data vastly outnumbers anomalies [220], distinctions between point-wise and range-based detection requiring different evaluation approaches [221], and domain-specific requirements where financial applications prioritize false positive minimization while industrial monitoring emphasizes timely detection [222].

TABLE 13: Summary of benchmark datasets for IAD, TSAD, TAD, and VAD across industrial, medical, surveillance, and cybersecurity domains.

Name	Task	Year	Venue	Type	#Samples	#Subjects	Domain	Site
MVTec-AD [97]	IAD	2019	CVPR	Real	5,354	15	Industrial anomaly detection	■
BraTS [228]	IAD	2020	-	Real	5,000+	1,250+	Medical Imaging (Brain Tumor)	■
MVTec 3D-AD [229]	IAD	2021	IJCVIPA	Real	4,433	10	Industrial anomaly detection	■
MPDD [230]	IAD	2021	ICUMT	Real	4,566	3	Metal part defect detection	■
BTAD [231]	IAD	2021	ISIE	Real	2,830	3	Industrial Inspection	■
KSDD2 [232]	IAD	2021	Comput. Ind.	Real	3,420	1	Industrial (Surface) Inspection	■
MVTec LOCO [233]	IAD	2022	IJCV	Real	1,772	5	Industrial anomaly detection	■
VisA [96]	IAD	2022	ECCV	Real	10,821	12	Visual anomaly detection	■
PCB-Bank [47]	IAD	2024	ECCV	Real	-	-	Circuit board defect detection	■
SWAT [234]	TSAD	2016	CRITIS	Real	950,000	1	Industrial Control Systems	■
MSL [235]	TSAD	2018	KDD	Real	73,729	27	Aerospace Telemetry	■
SMAP [235]	TSAD	2018	KDD	Real	427,617	55	Satellite Monitoring	■
SMD [152]	TSAD	2019	KDD	Real	20M	28	AIOps Server Monitoring	■
UCSD Ped2 [236]	VAD	2010	CVPR	Real	28	-	Video Surveillance	■
CUHK Avenue [237]	VAD	2013	ICCV	Real	30,652	16	Crowd Behavior	■
ShanghaiTech [238]	VAD	2016	CVPR	Real	1,198	13	Crowd Counting	■
UCF-Crime [140]	VAD	2018	CVPR	Real	1,900	13	Surveillance Video	■
CrossTask [239]	VAD	2019	CVPR	Real	4,700	83	Weakly Supervised AD	■
COIN [240]	VAD	2019	CVPR	Real	11,827	180	Instructional Video Analysis	■
Unormal [241]	VAD	2022	CVPR	Synth	236,902	29	Open-set Video	■
TEP [242]	TAD	1993	Springer	Synth	-	21	Industrial Process	■
BatchBenchmark [243]	TAD	2016	CILS	Synth	500	5	Chemical Process	■
ADBench [244]	TAD	2022	NeurIPS	Real	57	-	Tabular AD	■

9.1.4 Multimodal Anomaly Detection Metrics

MAD evaluation addresses both modality-specific and cross-modal performance aspects through comprehensive metric frameworks [223]. Individual modalities employ standard metrics like Dice coefficient, IoU, and Hausdorff Distance for localization alongside AUROC, AP, and F1-score for classification. Cross-modal assessment introduces consistency metrics quantifying information alignment between sources [186] and fusion performance metrics evaluating multimodal integration benefits over single-modality approaches [224]. Application-specific adaptations combine visual-textual data for industrial inspection [225] or specialized audio-visual metrics for surveillance, while addressing challenges from data heterogeneity, lack of standardized benchmarks [226], and the need for interpretable metrics capturing complementary information across modalities [227].

9.2 Benchmark Datasets

Evaluating AD requires access to well-established benchmark datasets, as summarized in Table 13, which presents key datasets categorized by modality and application domain.

9.2.1 Image Anomaly Detection Datasets

MVTec AD [97] serves as the de facto standard for industrial anomaly detection, encompassing 15 object categories with diverse defect types. Its extension, *MVTec 3D-AD* [229], incorporates depth information, facilitating research on 3D anomaly detection. Additionally, *MVTec LOCO* [233] focuses on logical anomalies characterized by structural or compositional irregularities. Other industrial datasets include *BTAD* [231], *VisA* [96], *KSDD2* [232], *MPDD* [230], and *PCB-Bank* [47], each presenting unique characteristics for evaluating different aspects of diffusion model performance. Finally, the medical dataset *BraTS* [228] provides MRI scans for brain tumor segmentation, introducing challenges related to subtle anomalies and the critical need for accurate detection in medical applications.

9.2.2 Video Anomaly Detection Datasets

Industrial surveillance datasets including *UCSD Ped2* [236] and *CUHK Avenue* [237] provide foundational benchmarks for non-pedestrian detection and unusual pedestrian behaviors, while comprehensive datasets *ShanghaiTech* [140]

and *UCF-Crime* [140] offer diverse evaluation scenarios across 13 scenes and crime categories respectively. The most challenging benchmark, *NWPU Campus* [245], [246], features 43 scenes with scene-dependent anomaly definitions where identical events may be normal or abnormal based on context. Complementing surveillance data, *UBnormal* [241] provides densely annotated videos for pixel-level localization, while instructional datasets *CrossTask* [239] and *COIN* [240] evaluate procedural anomaly detection through deviations from standard procedures.

9.2.3 Time Series Anomaly Detection Datasets

Industrial monitoring drives TSAD benchmark development, with *SMD* [152] providing server metrics for multivariate dependency evaluation, while spacecraft datasets *SMAP* [235] and *MSL* [235] offer real-world telemetry scenarios. Moreover, industrial control contributions include *SWaT* [234] featuring real-world water purification testbed data with simulated attacks, specialized datasets *TEP* [242] with chemical process environments, *Batchbenchmark* [243] for manufacturing processes, and comprehensive *ADBench* [244] containing 57 real-world datasets across domains. The diversity in temporal characteristics and anomaly prevalence provides comprehensive evaluation settings for diffusion model performance in capturing temporal correlations essential for robust TSAD applications.

10 DISCUSSION

10.1 Open Challenges

The application of ADGDM has shown promising results, yet several critical challenges remain to be addressed to achieve robust, scalable, and generalizable solutions.

Scalability. Scaling DMs to high-dimensional data, such as high-resolution images or long video sequences, remains challenging due to significant memory and computational demands [247]. Iterative denoising processes exacerbate scalability issues, limiting applicability in resource-intensive settings [73]. Developing efficient architectures, such as latent diffusion or optimized sampling techniques, is crucial to handling complex multimodal inputs effectively while maintaining detection accuracy [89].

Generalization. Achieving robust generalization for diffusion models across domains, e.g., from industrial to medical imaging, is hindered by domain shifts and varying anomaly types [6], [11]. Models trained on specific datasets often fail to adapt to new contexts, reducing their practical utility [32], [33]. Transfer learning and domain-adaptive techniques are essential to enhance cross-domain performance, ensuring DMs can handle diverse data distributions and anomaly characteristics effectively [212].

Class Imbalance. Detecting rare anomalies in imbalanced datasets, where normal samples dominate, poses a significant challenge for diffusion models [210]. Overfitting to normal data distributions reduces sensitivity to subtle anomalies, compromising detection accuracy [195]. Techniques like anomaly augmentation or robust scoring mechanisms are needed to improve model performance on rare events, ensuring reliable detection in critical applications such as industrial monitoring [86].

Computational Efficiency. High computational cost of DMs during inference, driven by iterative denoising steps, hinders deployment in real-time or resource-constrained environments [248], thereby limiting their practicality for applications like industrial monitoring [93]. Consequently, techniques such as model distillation, fewer-step sampling, or hardware-aware optimizations are essential to reduce inference latency while preserving AD accuracy [94].

Robustness. DMs are vulnerable to adversarial perturbations and noisy inputs, which can distort anomaly scoring and compromise reliability [91], particularly in domains like medical diagnostics, where robustness is paramount [17]. Given these vulnerabilities, developing adversarial training strategies or noise-robust scoring mechanisms is necessary to enhance model stability, ensuring consistent performance under challenging conditions and safeguarding against malicious or erroneous inputs [67].

10.2 Future Opportunities

ADGDM presents several promising opportunities for future research and development, with prospects to enhance efficiency, integrate novel conditioning, combine with other techniques, explore new modalities, and leverage LLMs for intelligent detection, that could significantly advance the capabilities and practical applications of this field.

Efficient DM architectures. High computational costs limit DM deployment in real-time anomaly detection applications, particularly for high-resolution images and videos. Dynamic step size computation [22] addresses this through initial anomaly prediction guidance, while latent space projections and quantization techniques reduce memory demands. Moreover, alternatives to full-length Markov chain diffusion [74] and multi-scale simplex noise diffusion offer pathways to scaling DMs to high-resolution imagery without compromising detection quality.

New Conditioning Strategies. Enhanced conditioning strategies incorporate prior knowledge and contextual information to improve anomaly detection performance [248]. Multimodal conditioning through text or audio provides complementary detection cues [249], while domain-specific knowledge integration improves detection specificity. Temporal conditioning benefits time-series and video analysis [26], while feedback loops using anomaly scores [22], [25] enhance localization accuracy. Adaptive conditioning strategies that respond to input characteristics offer promising solutions for heterogeneous datasets.

DMs with Other AD techniques. Hybrid approaches combining DMs with existing methods yield superior performance through complementary strengths. Clustering integration leverages generative capabilities for representative normal data synthesis [23], while one-class classifiers benefit from DM-generated samples [241]. Notable methods include ODD [18] combining similarity networks with outlier exposure, and dual conditioning frameworks for multi-class detection through category-specific reconstructions [62].

New Modalities and Applications. Current ADGDM focuses primarily on images, videos, and time series, leaving opportunities in broader modalities and applications. Tabular data applications in finance and healthcare show promise despite mixed data type challenges [2]. Multimodal detection [25]

and extensions to audio or 3D point clouds could enable new applications, while time series advances address irregular sampling and long-term dependencies [26], combined with self-supervised learning [196].

AD with LLMs. Large language models enhance diffusion-based detection through advanced pattern understanding. LLMs provide interpretable explanations for anomalies [250] while analyzing multimodal data for subtle patterns [251], [252]. Complex temporal pattern identification benefits time series detection [253], while synthetic anomaly generation addresses data scarcity [254], [255]. Integration enables intelligent adaptive monitoring for medical diagnosis and cybersecurity [256], [257].

11 CONCLUSION

In this survey, we comprehensively examine anomaly detection and generation with diffusion models (ADGDM) across diverse data modalities. We explore various conditioning strategies and their impact on performance while highlighting adaptations for image, video, time series, and tabular data. We also explore emerging multimodal approaches and assess the potential of DMs for synthetic anomaly generation. Researchers currently face challenges including scalability, generalization, class imbalance, computational efficiency, and robustness across applications. Future research opportunities encompass developing efficient architectures, novel conditioning strategies, hybrid approaches integrating traditional techniques, and leveraging large language models.

REFERENCES

- [1] C. Qiu, M. Kloft, S. Mandt, and M. Rudolph, "Self-supervised anomaly detection with neural transformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 3, pp. 2170–2185, Mar. 2025.
- [2] V. Livernoche, V. Jain, Y. Hezaveh, and S. Ravanbakhsh, "On diffusion modeling for anomaly detection," in *ICLR*, 2023.
- [3] W. H. L. Pinaya, M. S. Graham, R. Gray, P. F. da Costa, P.-D. Tudosi, P. Wright, Y. H. Mah, A. D. MacKinnon, J. T. Teo, R. Jager, D. Werring, G. Rees, P. Nachev, S. Ourselin, and M. J. Cardoso, "Fast unsupervised brain anomaly detection and segmentation with diffusion models," in *MICCAI*, 2022, pp. 705–714.
- [4] X. Yang, T. Ye, X. Yuan, W. Zhu, X. Mei, and F. Zhou, "A novel data augmentation method based on denoising diffusion probabilistic model for fault diagnosis under imbalanced data," *IEEE Trans. Ind. Informat.*, vol. 20, no. 5, pp. 7820–7831, 2024.
- [5] T. K. K. Ho, A. Karami, and N. Armanfard, "Graph anomaly detection in time series: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2025.
- [6] M. Jin, H. Y. Koh, Q. Wen, D. Zambon, C. Alippi, G. I. Webb, I. King, and S. Pan, "A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 1–20, 2024.
- [7] B. Ramachandra, M. J. Jones, and R. R. Vatsavai, "A survey of single-scene video anomaly detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 5, pp. 2293–2312, 2022.
- [8] Z. Zamanzadeh Darban, G. I. Webb, S. Pan, C. Aggarwal, and M. Salehi, "Deep learning for time series anomaly detection: A survey," *ACM Comput. Surv.*, vol. 57, no. 1, pp. 1–42, 2025.
- [9] H. Hojjati, T. K. K. Ho, and N. Armanfard, "Self-supervised anomaly detection in computer vision and beyond: A survey and outlook," *Neural Netw.*, vol. 172, p. 106106, 2024.
- [10] C.-H. Chang, J. Yoon, S. Ö. Arik, M. Uddel, and T. Pfister, "Data-efficient and interpretable tabular anomaly detection," in *KDD*, 2023, pp. 190–201.
- [11] J. Carvalho, M. Zhang, R. Geyer, C. Cotrini, and J. M. Buhmann, "Invariant anomaly detection under distribution shifts: A causal perspective," *NeurIPS*, vol. 36, pp. 56310–56337, 2023.
- [12] D. Liu, Q. Li, A.-D. Dinh, T. Jiang, M. Shah, and C. Xu, "Diffact++: Diffusion action segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 3, pp. 1644–1659, 2025.
- [13] M. Awais, M. Naseer, S. Khan, R. M. Anwer, H. Cholakkal, M. Shah, M.-H. Yang, and F. S. Khan, "Foundation models defining a new era in vision: A survey and outlook," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 4, pp. 2245–2264, 2025.
- [14] J. E. Santos, Z. R. Fox, N. Lubbers, and Y. T. Lin, "Blackout diffusion: Generative diffusion models in discrete-state spaces," in *ICML*, 2023, pp. 9034–9059.
- [15] C. Duan, F. Yang, P. Zhao, L. Zheng, Y. Dagli, Y. Liu, Q. Lin, and D. Zhang, "Soil: Score conditioned diffusion model for imbalanced cloud failure prediction," in *WebConf*, 2024, pp. 65–72.
- [16] K. Kumar, S. Chakraborty, and S. Roy, "Self-supervised diffusion model for anomaly segmentation in medical imaging," in *PREMI*, vol. 14301, 2023, pp. 359–368.
- [17] J. Li, H. Cao, J. Wang, F. Liu, Q. Dou, G. Chen, and P.-A. Heng, "Fast non-markovian diffusion model for weakly supervised anomaly detection in brain mr images," in *MICCAI*, vol. 14224, 2023, pp. 579–589.
- [18] H. Wang, L. Dai, J. Tong, and Y. Zhai, "Odd: One-class anomaly detection via the diffusion model," in *ICIP*, 2023, pp. 3000–3004.
- [19] P. Wu, X. Zhou, G. Pang, L. Zhou, Q. Yan, P. Wang, and Y. Zhang, "Vadclip: Adapting vision-language models for weakly supervised video anomaly detection," *AAAI*, vol. 38, no. 6, pp. 6074–6082, 2024.
- [20] K. Wang, J. Chen, H. Li, Z. Mi, and J. Zhu, "Sparsedm: Toward sparse efficient diffusion models," 2024.
- [21] M. Zhang, J. Wang, Q. Qi, P. Ren, H. Sun, Z. Zhuang, L. Zhang, and J. Liao, "Safeguarding sustainable cities: Unsupervised video anomaly detection through diffusion-based latent pattern learning," in *IJCAI*, vol. 8, 2024, pp. 7572–7580.
- [22] J. Tebbe and J. Tayyub, "D3ad: Dynamic denoising diffusion probabilistic model for anomaly detection," 2023.
- [23] H. Yao, M. Liu, H. Wang, Z. Yin, Z. Yan, X. Hong, and W. Zuo, "Glad: Towards better reconstruction with global and local adaptive diffusion models for unsupervised anomaly detection," 2024.
- [24] S. N. Marimont, M. Baugh, V. Siomos, C. Tzelepis, B. Kainz, and G. Tarroni, "Disyre: Diffusion-inspired synthetic restoration for unsupervised anomaly detection," 2024.
- [25] J. Tebbe and J. Tayyub, "Dynamic addition of noise in a diffusion model for anomaly detection," in *CVPR*, 2024, pp. 3940–3949.
- [26] I. Pintilie, A. Manolache, and F. Brad, "Time series anomaly detection using diffusion-based models," in *ICDM*, 2023, pp. 570–578.
- [27] H. Zhou, J. Yu, and W. Yang, "Dual memory units with uncertainty regulation for weakly supervised video anomaly detection," *AAAI*, vol. 37, no. 3, pp. 3769–3777, 2023.
- [28] J. Sui, J. Yu, Y. Song, and J. Zhang, "Anomaly detection for telemetry time series using a denoising diffusion probabilistic model," *IEEE Sensors J.*, vol. 24, no. 10, pp. 16429–16439, 2024.
- [29] Z. Yang, H. Yu, P. Guo, K. Zarna, X. Yang, and A. Sano, "Balanced mixed-type tabular data synthesis with diffusion models," 2024.
- [30] A. A. Cook, G. Misirli, and Z. Fan, "Anomaly detection for iot time-series data: A survey," *IEEE Internet Things J.*, vol. 7, no. 7, pp. 6481–6494, 2020.
- [31] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 38:1–38:38, 2021.
- [32] T. Fernando, H. Gammulle, S. Denman, S. Sridharan, and C. Fookes, "Deep learning for medical anomaly detection – a survey," *ACM Comput. Surv.*, vol. 54, no. 7, pp. 1–37, 2022.
- [33] S. Chandrakala, K. Deepak, and G. Revathy, "Anomaly detection in surveillance videos: A thematic taxonomy of deep models, review and performance analysis," *Artif. Intell. Rev.*, vol. 56, no. 4, pp. 3319–3368, 2023.
- [34] X. Ma, J. Wu, S. Xue, J. Yang, C. Zhou, Q. Z. Sheng, H. Xiong, and L. Akoglu, "A comprehensive survey on graph anomaly detection with deep learning," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12012–12038, 2023.
- [35] M. Ul Hassan, M. H. Rehmani, and J. Chen, "Anomaly detection in blockchain networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 289–318, 2023.
- [36] Y. Liu, D. Yang, Y. Wang, J. Liu, J. Liu, A. Boukerche, P. Sun, and L. Song, "Generalized video anomaly event detection: Systematic taxonomy and comparison of deep models," *ACM Comput. Surv.*, vol. 56, no. 7, pp. 189:1–189:38, 2024.
- [37] J. Liu, Y. Liu, J. Lin, J. Li, L. Cao, P. Sun, B. Hu, L. Song, A. Boukerche, and V. C. Leung, "Networking systems for video anomaly detection: A tutorial and survey," *ACM Comput. Surv.*, Apr. 2025.
- [38] D. McAllester, "On the mathematics of diffusion models," 2023.
- [39] P. Kidger, J. Foster, X. Li, and T. J. Lyons, "Neural sdes as infinite-dimensional gans," in *ICML*, 2021, pp. 5453–5463.

- [40] Q. Cui, X. Zhang, Z. Lu, and Q. Liao, "Elucidating the solution space of extended reverse-time sde for diffusion models," 2023.
- [41] A. Zhu and Q. Li, "Dyngma: A robust approach for learning stochastic differential equations from data," *J. Comput. Phys.*, vol. 513, p. 113200, 2024.
- [42] K. Ghanem and D. Bzdok, "The uncanny valley: A comprehensive analysis of diffusion models," 2024.
- [43] N. Dridi, L. Drumetz, and R. Fablet, "Learning stochastic dynamical systems with neural networks mimicking the euler-maruyama scheme," in *EUSIPCO*, 2021, pp. 1990–1994.
- [44] T. Chen, J. Gu, L. Dinh, E. A. Theodorou, J. Susskind, and S. Zhai, "Generative modeling with phase stochastic bridges," 2024.
- [45] H. Zheng, W. Nie, A. Vahdat, K. Azizzadenesheli, and A. Anandkumar, "Fast sampling of diffusion models via operator learning," in *ICML*, 2023, pp. 42390–42402.
- [46] Y. Lu, Z. Wang, and G. Bal, "Mathematical analysis of singularities in the diffusion model under the submanifold assumption," 2024.
- [47] H. Yao, M. Liu, Z. Yin, Z. Yan, X. Hong, and W. Zuo, "Glad: Towards better reconstruction with global and local adaptive diffusion models for unsupervised anomaly detection," in *ECCV*, 2025, pp. 1–17.
- [48] J. Choi, J. Lee, C. Shin, S. Kim, H. Kim, and S. Yoon, "Perception prioritized training of diffusion models," in *CVPR*, 2022, pp. 11472–11481.
- [49] C. Hou, G. Wei, and Z. Chen, "High-fidelity diffusion-based image editing," *AAAI*, vol. 38, no. 3, pp. 2184–2192, 2024.
- [50] H. Xu, S. Xu, and W. Yang, "Unsupervised industrial anomaly detection with diffusion models," *J. Vis. Commun. Image Represent.*, vol. 97, p. 103983, 2023.
- [51] T. Iwata and Y. Yamanaka, "Supervised anomaly detection based on deep autoregressive density estimators," 2019.
- [52] V. Mikuni and B. Nachman, "High-dimensional and permutation invariant anomaly detection," *SciPost Physics*, vol. 16, no. 3, p. 062, 2024.
- [53] S. Saremi, A. Mehrjou, B. Schölkopf, and A. Hyvärinen, "Deep energy estimator networks," 2018.
- [54] M. Amiridi, N. Kargas, and N. D. Sidiropoulos, "Low-rank characteristic tensor density estimation part ii: Compression and latent density estimation," *IEEE Trans. Signal Process.*, vol. 70, pp. 2669–2680, 2022.
- [55] M. Chen, K. Huang, T. Zhao, and M. Wang, "Score approximation, estimation and distribution recovery of diffusion models on low-dimensional data," in *ICML*, 2023, pp. 4672–4712.
- [56] J. Stanczuk, G. Batzolis, T. Devaney, and C.-B. Schörnlieb, "Your diffusion model secretly knows the dimension of the data manifold," 2023.
- [57] Y. Song, C. Durkan, I. Murray, and S. Ermon, "Maximum likelihood training of score-based diffusion models," in *NeurIPS*, vol. 34, 2021, pp. 1415–1428.
- [58] Y. Han, M. Razaviyayn, and R. Xu, "Neural network-based score estimation in diffusion models: Optimization and generalization," 2024.
- [59] F. Permenter and C. Yuan, "Interpreting and improving diffusion models using the euclidean distance function," 2023.
- [60] J. Han, S. Feng, M. Zhou, X. Zhang, Y. S. Ong, and X. Li, "Diffusion model in normal gathering latent space for time series anomaly detection," in *ECML PKDD*, vol. 14943, 2024, pp. 284–300.
- [61] H. Zuo, A. Zhu, Y. Zhu, Y. Liao, S. Li, and Y. Chen, "Unsupervised diffusion based anomaly detection for time series," *Appl. Intell.*, vol. 54, no. 19, pp. 8968–8981, 2024.
- [62] J. Zhan, J. Lai, B.-B. Gao, J. Liu, X. Chen, and C. Wang, "Enhancing multi-class anomaly detection via diffusion refinement with dual conditioning," 2024.
- [63] A. Mousakhan, T. Brox, and J. Tayyub, "Anomaly detection with conditioned denoising diffusion models," 2023.
- [64] V. Singh, S. Jandial, A. Chopra, S. Ramesh, B. Krishnamurthy, and V. N. Balasubramanian, "On conditioning the input noise for controlled image generation with diffusion models," 2022.
- [65] S. Sadat, J. Buhmann, D. Bradley, O. Hilliges, and R. M. Weber, "Cads: Unleashing the diversity of diffusion models through condition-annealed sampling," 2024.
- [66] C. Wei, K. Mangalam, P.-Y. Huang, Y. Li, H. Fan, H. Xu, H. Wang, C. Xie, A. Yuille, and C. Feichtenhofer, "Diffusion models as masked autoencoders," in *ICCV*, 2023, pp. 16238–16248.
- [67] G. Lin, Z. Tao, J. Zhang, T. Tanaka, and Q. Zhao, "Adversarial guided diffusion models for adversarial purification," 2025.
- [68] T. Salimans and J. Ho, "Progressive distillation for fast sampling of diffusion models," in *ICLR*, 2021.
- [69] C. Xu, T. Song, W. Feng, X. Li, T. Ge, B. Zheng, and L. Wang, "Accelerating image generation with sub-path linear approximation model," in *ECCV*, vol. 15111, 2025, pp. 323–339.
- [70] Z. Wu, P. Zhou, K. Kawaguchi, and H. Zhang, "Fast diffusion model," 2023.
- [71] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li, and J. Zhu, "Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps," *NeurIPS*, vol. 35, pp. 5775–5787, 2022.
- [72] M. F. Liu, S. Lyu, M. Vinaroz, and M. Park, "Dp-lldms: Differentiably private latent diffusion models," 2025.
- [73] P. A. Golnari, "Lora-enhanced distillation on guided diffusion models," 2023.
- [74] J. Wyatt, A. Leach, S. M. Schmon, and C. G. Willcocks, "Anodddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise," in *CVPRW*, 2022, pp. 650–656.
- [75] Y. Che, F. Rafsanji, J. Shah, M. M. R. Siddiquee, and T. Wu, "Anofpdm: Anomaly segmentation with forward process of diffusion models for brain mri," 2024.
- [76] C. I. Bercea, B. Wiestler, D. Rueckert, and J. A. Schnabel, "Diffusion models with implicit guidance for medical anomaly detection," in *MICCAI*, vol. 15011, 2024, pp. 211–220.
- [77] M. Baugh, H. Reynaud, S. N. Marimont, S. Cechnicka, J. P. Müller, G. Tarroni, and B. Kainz, "Image-conditioned diffusion models for medical anomaly detection," in *MLMI*, vol. 15167, 2025, pp. 117–127.
- [78] H. Iqbal, U. Khalid, C. Chen, and J. Hua, "Unsupervised anomaly detection in medical images using masked diffusion model," in *MLMI*, 2024, pp. 372–381.
- [79] A. Bansal, E. Borgnia, H.-M. Chu, J. Li, H. Kazemi, F. Huang, M. Goldblum, J. Geiping, and T. Goldstein, "Cold diffusion: Inverting arbitrary image transforms without noise," in *NeurIPS*, vol. 36, 2023, pp. 41259–41282.
- [80] S. Naval Marimont, V. Siomos, M. Baugh, C. Tzelepis, B. Kainz, and G. Tarroni, "Ensembled cold-diffusion restorations for unsupervised anomaly detection," in *MICCAI*, 2024, pp. 243–253.
- [81] M. S. Graham, W. H. L. Pinaya, P. Wright, P.-D. Tudosiu, Y. H. Mah, J. T. Teo, H. R. Jäger, D. Werring, P. Nachev, S. Ourselin, and M. J. Cardoso, "Unsupervised 3d out-of-distribution detection with latent diffusion models," in *MICCAI*, 2023, pp. 446–456.
- [82] R. Babaei, S. Cheng, T. Thai, and S. Zhao, "Pancreatic tumor segmentation as anomaly detection in ct images using denoising diffusion models," 2024.
- [83] S. Ahamed, Y. Xu, and A. Rahmim, "Igconda-pet: Implicitly-guided counterfactual diffusion for detecting anomalies in pet images," 2024.
- [84] Z. Liang, H. Anthony, F. Wagner, and K. Kamnitsas, "Modality cycles with masked conditional diffusion for unsupervised anomaly segmentation in mri," in *MICCAI*, 2024, pp. 168–181.
- [85] T. Hu, J. Zhang, R. Yi, Y. Du, X. Chen, L. Liu, Y. Wang, and C. Wang, "Anomalydiffusion: Few-shot anomaly image generation with diffusion model," in *AAAI*, vol. 38, 2024, pp. 8526–8534.
- [86] Y. Jin, J. Peng, Q. He, T. Hu, H. Chen, J. Wu, W. Zhu, M. Chi, J. Liu, Y. Wang, and C. Wang, "Dualanodiff: Dual-interrelated diffusion model for few-shot anomaly image generation," 2024.
- [87] A. Fontanella, G. Mair, J. Wardlaw, E. Trucco, and A. Storkey, "Diffusion models for counterfactual generation and anomaly detection in brain images," *IEEE Trans. Med. Imag.*, pp. 1–1, 2024.
- [88] J. Mao, X. Wang, and K. Aizawa, "Guided image synthesis via initial image editing in diffusion model," in *ACM MM*, 2023, pp. 5321–5329.
- [89] W. Li, X. Xu, J. Liu, and X. Xiao, "Unimo-g: Unified image generation through multimodal conditional diffusion," in *ACL*, 2024, pp. 6173–6188.
- [90] G. L. Bellier and N. Audebert, "Detecting out-of-distribution earth observation images with diffusion models," in *CVPRW*, 2024, pp. 481–491.
- [91] F. Lu, X. Yao, C.-W. Fu, and J. Jia, "Removing anomalies as noises for industrial defect localization," in *ICCV*, 2023, pp. 16166–16175.
- [92] V. Zavrtanik, M. Kristan, and D. Skočaj, "Draem - a discriminatively trained reconstruction embedding for surface anomaly detection," in *ICCV*, 2021, pp. 8330–8339.
- [93] T. Defard, A. Setkov, A. Loesch, and R. Audigier, "Padim: A patch distribution modeling framework for anomaly detection and localization," in *Pattern Recognit.*, 2021, pp. 475–489.
- [94] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards total recall in industrial anomaly detection," in *CVPR*, 2022, pp. 14318–14328.
- [95] C. I. Bercea, M. Neumayr, D. Rueckert, and J. A. Schnabel, "Mask, stitch, and re-sample: Enhancing robustness and generalizability in anomaly detection through automatic diffusion models," 2023.
- [96] Y. Zou, J. Jeong, L. Pemula, D. Zhang, and O. Dabeer, "Spot-the-difference self-supervised pre-training for anomaly detection and segmentation," in *ECCV*, 2022, pp. 392–408.

- [97] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Mvttec ad—a comprehensive real-world dataset for unsupervised anomaly detection," in *CVPR*, 2019, pp. 9584–9592.
- [98] N. Cohen and Y. Hoshen, "Sub-image anomaly detection with deep pyramid correspondences," 2021.
- [99] H. Deng and X. Li, "Anomaly detection via reverse distillation from one-class embedding," in *CVPR*, 2022, pp. 9737–9746.
- [100] J. Zhang, Z. Wang, J. Meng, Y.-P. Tan, and J. Yuan, "Boosting positive and unlabeled learning for anomaly detection with multi-features," *IEEE Trans. Multimedia*, vol. 21, no. 5, pp. 1332–1344, May 2019.
- [101] Y. Cong, J. Yuan, and J. Liu, "Abnormal event detection in crowded scenes using sparse representation," *Pattern Recognit.*, vol. 46, no. 7, pp. 1851–1864, Jul. 2013.
- [102] F. Jiang, J. Yuan, S. A. Tsafaris, and A. K. Katsaggelos, "Anomalous video event detection using spatiotemporal context," *Comput. Vis. Image Underst.*, vol. 115, no. 3, pp. 323–333, Mar. 2011.
- [103] A. O. Tur, N. Dall'Asen, C. Beyan, and E. Ricci, "Exploring diffusion models for unsupervised video anomaly detection," in *ICIP*, 2023, pp. 2540–2544.
- [104] H. Lu, G. Yang, N. Fei, Y. Huo, Z. Lu, P. Luo, and M. Ding, "Vdt: General-purpose video diffusion transformers via mask modeling," 2023.
- [105] G. Y. Park, H. Jeong, S. W. Lee, and J. C. Ye, "Spectral motion alignment for video motion transfer using diffusion models," 2024.
- [106] J. Liang, Y. Fan, K. Zhang, R. Timofte, L. Van Gool, and R. Ranjan, "Movideo: Motion-aware video generation with diffusion model," in *ECCV*, vol. 15102, 2025, pp. 56–74.
- [107] K. Mei and V. Patel, "Vidm: Video implicit diffusion models," *AAAI*, vol. 37, no. 8, pp. 9117–9125, 2023.
- [108] A. Awasthi, S. T. Ly, J. Nizam, V. Mehta, S. Ahmad, R. Nemani, S. Prasad, and H. Van Nguyen, "Anomaly detection in satellite videos using diffusion models," in *IEEE MMSP*. IEEE, 2024, pp. 1–6.
- [109] A. Blattmann, R. Rombach, H. Ling, T. Dockhorn, S. W. Kim, S. Fidler, and K. Kreis, "Align your latents: High-resolution video synthesis with latent diffusion models," in *CVPR*, 2023, pp. 22 563–22 575.
- [110] J. Gu, S. Wang, H. Zhao, T. Lu, X. Zhang, Z. Wu, S. Xu, W. Zhang, Y.-G. Jiang, and H. Xu, "Reuse and diffuse: Iterative denoising for text-to-video generation," 2023.
- [111] C. Yan, S. Zhang, Y. Liu, G. Pang, and W. Wang, "Feature prediction diffusion model for video anomaly detection," in *ICCV*, 2023, pp. 5504–5514.
- [112] A. Blattmann, T. Dockhorn, S. Kulal, D. Mendelevitch, M. Kilian, D. Lorenz, Y. Levi, Z. English, V. Voleti, A. Letts, V. Jampani, and R. Rombach, "Stable video diffusion: Scaling latent video diffusion models to large datasets," 2023.
- [113] W. Liu, H. Chang, B. Ma, S. Shan, and X. Chen, "Diversity-measurable anomaly detection," in *CVPR*, 2023, pp. 12 147–12 156.
- [114] Z. Wang, X. Gu, J. Hu, and X. Gu, "Ensemble anomaly score for video anomaly detection using denoise diffusion model and motion filters," *Neurocomputing*, vol. 553, p. 126589, 2023.
- [115] F. Fang, Y. Liu, A. Koksal, Q. Xu, and J.-H. Lim, "Masked diffusion with task-awareness for procedure planning in instructional videos," 2023.
- [116] A. Lapid, I. Achituve, L. Bracha, and E. Fetaya, "Gd-vdm: Generated depth for better diffusion-based video generation," 2023.
- [117] S. Lee, C. Kong, D. Jeon, and N. Kwak, "Aadiff: Audio-aligned video synthesis with text-to-image diffusion," 2023.
- [118] S. Basak and A. Gautam, "Diffusion-based normality pre-training for weakly supervised video anomaly detection," *Expert Syst. Appl.*, vol. 251, p. 124013, 2024.
- [119] K. Cheng, Y. Pan, Y. Liu, X. Zeng, and R. Feng, "Denoising diffusion-augmented hybrid video anomaly detection via reconstructing noised frames," in *IJCAI*, vol. 2, 2024, pp. 695–703.
- [120] A. Karami, T. K. K. Ho, and N. Armanfard, "Graph-jigsaw conditioned diffusion model for skeleton-based video anomaly detection," 2024.
- [121] H. Liu, L. He, M. Zhang, and F. Li, "Vadiffusion: Compressed domain information guided conditional diffusion for video anomaly detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 9, pp. 8398–8411, 2024.
- [122] L. Cai, M. Li, and D. Wang, "Unsupervised conditional diffusion models in video anomaly detection for monitoring dust pollution," *Sensors*, vol. 24, no. 5, p. 1464, 2024.
- [123] D. Li, W. Xie, Z. Wang, Y. Lu, Y. Li, and L. Fang, "Feddifff: Diffusion model driven federated learning for multi-modal and multi-clients," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 10, pp. 10 353–10 367, 2024.
- [124] R. Hu, X. Yuan, Y. Qiao, B. Zhang, and P. Zhao, "Unsupervised anomaly detection for multivariate time series using diffusion model," in *ICASSP*, 2024, pp. 9606–9610.
- [125] C. Wang, Z. Zhuang, Q. Qi, J. Wang, X. Wang, H. Sun, and J. Liao, "Drift doesn't matter: Dynamic decomposition with diffusion reconstruction for unstable multivariate time series anomaly detection," in *NeurIPS*, vol. 36, 2023, pp. 10 758–10 774.
- [126] C. Yang, T. Wang, and X. Yan, "Ddmt: Denoising diffusion mask transformer models for multivariate time series anomaly detection," 2023.
- [127] C. Xiao, Z. Gou, W. Tai, K. Zhang, and F. Zhou, "Imputation-based time-series anomaly detection with conditional weight-incremental diffusion models," in *KDD*, 2023, pp. 2742–2751.
- [128] T. K. K. Ho and N. Armanfard, "Contaminated multivariate time-series anomaly detection with spatio-temporal graph conditional diffusion models," 2024.
- [129] N. Sivaroopan, D. Bandara, C. Madarasingha, G. Jourjon, A. Jayasumana, and K. Thilakarathna, "Netdiffus: Network traffic generation by diffusion models through time-series imaging," *Computer Networks*, vol. 251, p. 110616, 2024.
- [130] Y. Chang, Z. Tu, W. Xie, B. Luo, S. Zhang, H. Sui, and J. Yuan, "Video anomaly detection with spatio-temporal dissociation," *Pattern Recognit.*, vol. 122, p. 108213, Feb. 2022.
- [131] Y. Chang, Z. Tu, W. Xie, and J. Yuan, "Clustering driven deep autoencoder for video anomaly detection," in *ECCV*, 2020, pp. 329–345.
- [132] Y. Cong, J. Yuan, and Y. Tang, "Video anomaly search in crowded scenes via spatio-temporal motion context," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 10, pp. 1590–1599, Oct. 2013.
- [133] S. Zhang, S. Wang, X. Tan, R. Liu, J. Zhang, and J. Wang, "Sasdim: Self-adaptive noise scaling diffusion model for spatial time series imputation," 2023.
- [134] D. Cao, W. Ye, and Y. Liu, "Timedit: General-purpose diffusion transformers for time series foundation model," in *ICMLW*, 2024.
- [135] Z. Senane, L. Cao, V. L. Buchner, Y. Tashiro, L. You, P. A. Herman, M. Nordahl, R. Tu, and V. von Ehrenheim, "Self-supervised learning of time series representation via diffusion process and imputation-interpolation-forecasting mask," in *KDD*, Aug. 2024, pp. 2560–2571.
- [136] J. M. L. Alcaraz and N. Strodthoff, "Diffusion-based time series imputation and forecasting with structured state space models," 2023.
- [137] F. Tian, X. Shi, L. Zhou, L. Chen, C. Ma, and W. Zhu, "Prodiffad: Progressively distilled diffusion models for multivariate time series anomaly detection in jointcloud environment," in *IJCNN*, 2024, pp. 1–8.
- [138] P. P. Dev, R. Hazari, and P. Das, "Reflip-vad: Towards weakly supervised video anomaly detection via vision-language model," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1–1, 2024.
- [139] P. Wu, J. Liu, Y. Shi, Y. Sun, F. Shao, Z. Wu, and Z. Yang, "Not only look, but also listen: Learning multimodal violence detection under weak supervision," in *ECCV*, 2020, pp. 322–339.
- [140] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *CVPR*, 2018, pp. 6479–6488.
- [141] Y. Tian, G. Pang, Y. Chen, R. Singh, J. W. Verjans, and G. Carneiro, "Weakly-supervised video anomaly detection with robust temporal feature magnitude learning," in *ICCV*, 2021, pp. 4955–4966.
- [142] Y. Lu, C. Cao, Y. Zhang, and Y. Zhang, "Learnable locality-sensitive hashing for video anomaly detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 2, pp. 963–976, 2023.
- [143] X. Huang, Y. Hu, X. Luo, J. Han, B. Zhang, and X. Cao, "Boosting variational inference with margin learning for few-shot scene-adaptive anomaly detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 6, pp. 2813–2825, 2023.
- [144] H. Lv, Z. Yue, Q. Sun, B. Luo, Z. Cui, and H. Zhang, "Unbiased multiple instance learning for weakly supervised video anomaly detection," in *CVPR*, 2023, pp. 8022–8031.
- [145] C. Xu, K. Xu, X. Jiang, and T. Sun, "Plovad: Prompting vision-language models for open vocabulary video anomaly detection," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1–1, 2025.
- [146] L. Chen, X. Shi, L. Zhou, Y. Wang, C. Ma, and W. Zhu, "Dynamic splitting of diffusion models for multivariate time series anomaly detection in a jointcloud environment," in *KSEM*, vol. 14886, 2024, pp. 28–40.
- [147] T. Hartvigsen, J. Thadjarassiri, X. Kong, and E. Rundensteiner, "Finding short signals in long irregular time series with continuous-time attention policy networks," 2023.
- [148] D. Drakulic and J.-M. Andreoli, "Structured time series prediction without structural prior," 2022.
- [149] S. Tuli, G. Casale, and N. R. Jennings, "Tranad: Deep transformer networks for anomaly detection in multivariate time series data," *Proc. VLDB Endow.*, vol. 15, no. 6, pp. 1201–1214, 2022.

- [150] R.-G. Cirstea, C. Guo, B. Yang, T. Kieu, X. Dong, and S. Pan, "Triforner: Triangular, variable-specific attentions for long sequence multivariate time series forecasting—full version," 2022.
- [151] F. Yang, W. Yin, L. Wang, T. Li, P. Zhao, B. Liu, P. Wang, B. Qiao, Y. Liu, M. Björkman, S. Rajmohan, Q. Lin, and D. Zhang, "Diffusion-based time series data imputation for cloud failure prediction at microsoft 365," in *ESEC/FSE*, 2023, pp. 2050–2055.
- [152] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust anomaly detection for multivariate time series through stochastic recurrent neural network," in *KDD*, 2019, pp. 2828–2837.
- [153] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep autoencoding gaussian mixture model for unsupervised anomaly detection," in *ICLR*, 2018.
- [154] Y. Li, W. Chen, B. Chen, D. Wang, L. Tian, and M. Zhou, "Prototype-oriented unsupervised anomaly detection for multivariate time series," in *ICML*, 2023, pp. 19407–19424.
- [155] C.-Y. A. Lai, F.-K. Sun, Z. Gao, J. H. Lang, and D. Boning, "Nominality score conditioned time series anomaly detection by point/sequential reconstruction," in *NeurIPS*, vol. 36, pp. 76 637–76 655, 2023.
- [156] J. Xu, H. Wu, J. Wang, and M. Long, "Anomaly transformer: Time series anomaly detection with association discrepancy," in *ICLR*, 2021.
- [157] Y. Feng, W. Zhang, Y. Fu, W. Jiang, J. Zhu, and W. Ren, "Sensitive-hue: Multivariate time series anomaly detection by enhancing the sensitivity to normal patterns," in *KDD*, 2024, pp. 782–793.
- [158] O. Kitouni, N. Nolte, J. Hensman, and B. Mitra, "Disk: A diffusion model for structured knowledge," 2023.
- [159] V. S. Chundawat, A. K. Tarun, M. Mandal, M. Lahoti, and P. Narang, "Tabsyndex: A universal metric for robust evaluation of synthetic tabular data," 2024.
- [160] Y. Ouyang, L. Xie, C. Li, and G. Cheng, "Missdiff: Training diffusion models on tabular data with missing values," in *ICMLW*, 2023.
- [161] Y. Wen, Y. Wang, K. Yi, J. Ke, and Y. Shen, "Diffimpute: Tabular data imputation with denoising diffusion probabilistic model," in *ICME*, 2024, pp. 1–6.
- [162] N. Suh, Y. Yang, D.-Y. Hsieh, Q. Luan, S. Xu, S. Zhu, and G. Cheng, "Timeautodiff: Combining autoencoder and diffusion model for time series tabular data synthesizing," 2024.
- [163] H. Zhang, J. Zhang, Z. Shen, B. Srinivasan, X. Qin, C. Faloutsos, H. Rangwala, and G. Karypis, "Mixed-type tabular data synthesis with score-based diffusion in latent space," in *ICLR*, 2023.
- [164] G. Zamberg, M. Salhov, O. Lindenbaum, and A. Averbuch, "Tabadm: Unsupervised tabular anomaly detection with diffusion models," 2023.
- [165] S. Yoon, Y.-U. Jin, Y.-K. Noh, and F. Park, "Energy-based models for anomaly detection: A manifold diffusion recovery approach," in *NeurIPS*, vol. 36, 2023, pp. 49 445–49 466.
- [166] J. Tritscher, P. Lissmann, M. Wolf, A. Krause, A. Hotho, and D. Schlör, "Generative inpainting for shapley-value-based anomaly explanation," in *Explain. Artif. Intell.*, 2024, pp. 230–243.
- [167] W. Dai, K. Hwang, and J. Fan, "Unsupervised anomaly detection for tabular data using noise evaluation," in *AAAI*, 2025.
- [168] C. Lee, J. Kim, and N. Park, "Codi: Co-evolving contrastive diffusion models for mixed-type tabular synthesis," in *ICML*, 2023, pp. 18 940–18 956.
- [169] Y. Liu, T. Ajanthan, H. Husain, and V. Nguyen, "Self-supervision improves diffusion models for tabular data imputation," in *CIKM*, 2024, pp. 1513–1522.
- [170] T. Sattarov, M. Schreyer, and D. Borth, "Findiff: Diffusion models for financial tabular data generation," in *ICAIF*, 2023, pp. 64–72.
- [171] S. Li, Y. Lu, S. Jiu, H. Huang, G. Yang, and J. Yu, "Prototype-oriented hypergraph representation learning for anomaly detection in tabular data," *Inf. Process. Manag.*, vol. 62, no. 1, p. 103877, 2025.
- [172] H. Thimonier, F. Popineau, A. Rimmel, and B.-L. Doan, "Retrieval augmented deep anomaly detection for tabular data," in *CIKM*, 2024, pp. 2250–2259.
- [173] C. Qiu, T. Pfommer, M. Kloft, S. Mandt, and M. Rudolph, "Neural transformation learning for deep anomaly detection beyond images," in *ICML*, 2021, pp. 8703–8714.
- [174] T. Shenkar and L. Wolf, "Anomaly detection for tabular data with internal contrastive learning," in *ICLR*, 2021.
- [175] Z. Li, Y. Zhao, X. Hu, N. Botta, C. Ionescu, and G. H. Chen, "Ecod: Unsupervised outlier detection using empirical cumulative distribution functions," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 12, pp. 12 181–12 193, 2023.
- [176] J. Cai and J. Fan, "Perturbation learning based anomaly detection," in *NeurIPS*, vol. 35, pp. 14 317–14 330, 2022.
- [177] X. Yuan and L. Chen, "D-pad: Deep-shallow multi-frequency patterns disentangling for time series forecasting," 2024.
- [178] W. Dai and J. Fan, "Autouad: Hyper-parameter optimization for unsupervised anomaly detection," in *ICLR*, 2024.
- [179] Z. Yan, Q. Fang, W. Lv, and Q. Su, "Anomalysd: Few-shot multi-class anomaly detection with stable diffusion model," 2024.
- [180] A. Costanzino, P. Z. Ramirez, G. Lisanti, and L. Di Stefano, "Multimodal industrial anomaly detection by crossmodal feature mapping," in *CVPR*, 2024, pp. 17 234–17 243.
- [181] S. Roheda, H. Krim, and B. S. Riggan, "Robust multi-modal sensor fusion: An adversarial approach," *IEEE Sensors J.*, vol. 21, no. 2, pp. 1885–1896, 2021.
- [182] Q. Zhang, H. Wu, C. Zhang, Q. Hu, H. Fu, J. T. Zhou, and X. Peng, "Provable dynamic fusion for low-quality multimodal data," in *ICML*, vol. 202, 2023, pp. 41 753–41 769.
- [183] J. Hu, Y. Huang, Y. Lu, G. Xie, G. Jiang, Y. Zheng, and Z. Lu, "Anomalyxfusion: Multi-modal anomaly synthesis with diffusion," 2024.
- [184] D. Chen, K. Pan, G. Dai, G. Wang, Y. Zhuang, S. Tang, and M. Xu, "Improving vision anomaly detection with the guidance of language modality," *IEEE Trans. Multimedia*, vol. 27, pp. 1410–1419, 2025.
- [185] H. Chen, H. Zhou, and Y. Deng, "Interpretation on multi-modal visual fusion," 2023.
- [186] Y. Li, N. Liu, J. Li, M. Du, and X. Hu, "Deep structured cross-modal anomaly detection," in *IJCNN*, 2019, pp. 1–8.
- [187] H. He, J. Zhang, H. Chen, X. Chen, Z. Li, X. Chen, Y. Wang, C. Wang, and L. Xie, "Diad: A diffusion-based framework for multi-class anomaly detection," 2023.
- [188] X. Chen and Y. Peng, "Counterfactual condition diffusion with continuous prior adaptive correction for anomaly detection in multimodal brain mri," *Expert Syst. Appl.*, vol. 254, p. 124295, 2024.
- [189] Z. Huang, K. C. K. Chan, Y. Jiang, and Z. Liu, "Collaborative diffusion for multi-modal face generation and editing," in *CVPR*, 2023, pp. 6080–6090.
- [190] L. Capogrosso, A. Vivenza, A. Chiariini, F. Setti, and M. Cristani, "Exploiting multimodal latent diffusion models for accurate anomaly detection in industry 5.0," in *Ital-IA*, vol. 3762, 2024, pp. 230–235.
- [191] A. Flaborea, L. Collorone, G. M. D'Amely Di Melendugno, S. D'Arrigo, B. Prenkaj, and F. Galasso, "Multimodal motion conditioned diffusion model for skeleton-based video anomaly detection," in *ICCV*, 2023, pp. 10 284–10 295.
- [192] Y. Li, H. Wang, S. Yuan, M. Liu, D. Zhao, Y. Guo, C. Xu, G. Shi, and W. Zuo, "Myriad: Large multimodal model by applying vision experts for industrial anomaly detection," 2025.
- [193] L. Capogrosso, A. Mascolini, F. Girella, G. Skenderi, S. Gaiardelli, N. Dall'Ora, F. Ponzio, E. Fraccaroli, S. Di Cataldo, S. Vinco, E. Macii, F. Fummi, and M. Cristani, "Neuro-symbolic empowered denoising diffusion probabilistic models for real-time anomaly detection in industry 4.0: Wild-and-crazy-idea paper," in *FDL*, 2023, pp. 1–4.
- [194] H. Sun, Y. Cao, and O. Fink, "Cut: A controllable, universal, and training-free visual anomaly generation framework," 2024.
- [195] A. Celaya, B. Riviere, and D. Fuentes, "A generalized surface loss for reducing the hausdorff distance in medical imaging segmentation," 2024.
- [196] S. Li, J. Yu, Y. Lu, G. Yang, X. Du, and S. Liu, "Self-supervised enhanced denoising diffusion for anomaly detection," *Inf. Sci.*, vol. 669, p. 120612, 2024.
- [197] H. Li, Z. Zhang, H. Chen, L. Wu, B. Li, D. Liu, and M. Wang, "A novel approach to industrial defect generation through blended latent diffusion model with online adaptation," 2024.
- [198] C.-L. Li, K. Sohn, J. Yoon, and T. Pfister, "Cutpaste: Self-supervised learning for anomaly detection and localization," in *CVPR*, 2021, pp. 9664–9674.
- [199] G. Fang, W. Yan, Y. Guo, J. Han, Z. Jiang, H. Xu, S. Liao, and X. Liang, "Humanrefiner: Benchmarking abnormal human generation and refining with coarse-to-fine pose-reversible guidance," in *ECCV*, 2024.
- [200] A. K. Rai, T. Krishna, F. Hu, A. Drimbarean, K. Mcguinness, A. F. Smeaton, and N. E. O'connor, "Video anomaly detection via spatio-temporal pseudo-anomaly generation : A unified approach," in *CVPR*, 2024, pp. 3887–3899.
- [201] H. M. Schlüter, J. Tan, B. Hou, and B. Kainz, "Natural synthetic anomalies for self-supervised anomaly detection and localization," in *ECCV*, 2022, pp. 474–489.
- [202] H. Zhang, Z. Wu, Z. Wang, Z. Chen, and Y.-G. Jiang, "Prototypical residual networks for anomaly detection and localization," in *CVPR*, 2023, pp. 16 281–16 291.
- [203] G. Zhang, K. Cui, T.-Y. Hung, and S. Lu, "Defect-gan: High-fidelity defect synthesis for automated defect inspection," in *WACV*, 2021, pp. 2524–2534.

- [204] Y. Duan, Y. Hong, L. Niu, and L. Zhang, "Few-shot defect image generation via defect-aware feature manipulation," *AAAI*, vol. 37, no. 1, pp. 571–578, 2023.
- [205] X. Zhang, M. Xu, and X. Zhou, "Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection," in *CVPR*, 2024, pp. 16 699–16 708.
- [206] S. Niu, B. Li, X. Wang, and H. Lin, "Defect image sample generation with gan for improving defect recognition," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 3, pp. 1611–1622, 2020.
- [207] S. Zhao, Z. Liu, J. Lin, J.-Y. Zhu, and S. Han, "Differentiable augmentation for data-efficient gan training," in *NeurIPS*, vol. 33, 2020, pp. 7559–7570.
- [208] D. Lin, Y. Cao, W. Zhu, and Y. Li, "Few-shot defect segmentation leveraging abundant defect-free training samples through normal background regularization and crop-and-paste operation," in *ICME*, 2021, pp. 1–6.
- [209] U. Ojha, Y. Li, J. Lu, A. A. Efros, Y. J. Lee, E. Shechtman, and R. Zhang, "Few-shot image generation via cross-domain correspondence," in *CVPR*, 2021, pp. 10 743–10 752.
- [210] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Res. Notes*, vol. 15, no. 1, p. 210, 2022.
- [211] W. Yan, Q. Yang, T. Syer, Z. Min, S. Punwani, M. Emberton, D. Barratt, B. Chiu, and Y. Hu, "The impact of using voxel-level segmentation metrics on evaluating multifocal prostate cancer localisation," in *Appl. Med. Artif. Intell.*, 2022, pp. 128–138.
- [212] X. Hu and C. Jin, "Anodode: Anomaly detection with diffusion ode," 2023.
- [213] X. Sheng, S. Tuo, and L. Wang, "Surface anomaly detection and localization with diffusion-based reconstruction," in *IJCNN*, 2024, pp. 1–8.
- [214] H. Zhang, Z. Wang, Z. Wu, and Y.-G. Jiang, "Diffusionad: Norm-guided one-step denoising diffusion for anomaly detection," 2023.
- [215] M. Tailanian, Á. Pardo, and P. Musé, "U-flow: A u-shaped normalizing flow for anomaly detection with unsupervised threshold," *J. Math. Imaging Vis.*, vol. 66, no. 4, pp. 678–696, 2024.
- [216] M. Z. Zaheer, A. Mahmood, M. Astrid, and S.-I. Lee, "Clustering aided weakly supervised training to detect anomalous events in surveillance videos," *Neural Netw.*, vol. 35, no. 10, pp. 14 085–14 098, 2024.
- [217] S. Sørøbø and M. Ruocco, "Navigating the metric maze: A taxonomy of evaluation metrics for anomaly detection in time series," *Data Min. Knowl. Discov.*, vol. 38, no. 3, pp. 1027–1068, 2024.
- [218] M. Goswami, C. Challu, L. Callot, L. Minorics, and A. Kan, "Unsupervised model selection for time-series anomaly detection," 2023.
- [219] W. Wu, L. He, W. Lin, Y. Su, Y. Cui, C. Maple, and S. Jarvis, "Developing an unsupervised real-time anomaly detection scheme for time series with multi-seasonality," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 9, pp. 4147–4160, 2022.
- [220] Y. Zhao, L. Deng, X. Chen, C. Guo, B. Yang, T. Kieu, F. Huang, T. B. Pedersen, K. Zheng, and C. S. Jensen, "A comparative study on unsupervised anomaly detection for time series: Experiments and analysis," 2022.
- [221] C. Zhang, W. Zuo, and X. Wang, "Reconstruct anomaly to normal: Adversarial learned and latent vector-constrained autoencoder for time-series anomaly detection," 2020.
- [222] C. Li, G. Feng, Y. Li, R. Liu, Q. Miao, and L. Chang, "Difftad: Denoising diffusion probabilistic models for vehicle trajectory anomaly detection," *Knowl. Based Syst.*, vol. 286, p. 111387, 2024.
- [223] D. Wei, Y. Liu, X. Zhu, J. Liu, and X. Zeng, "Msaf: Multimodal supervise-attention enhanced fusion for video anomaly detection," *IEEE Signal Process. Lett.*, vol. 29, pp. 2178–2182, 2022.
- [224] S. Shankar, "Neural dependency coding inspired multimodal fusion," 2021.
- [225] Z. Hu, S. Jia, and M. Rostami, "An intermediate fusion vit enables efficient text-image alignment in diffusion models," 2024.
- [226] P. P. Liang, Y. Lyu, X. Fan, J. Tsaw, Y. Liu, S. Mo, D. Yogatama, L.-P. Morency, and R. Salakhutdinov, "High-modality multimodal transformer: Quantifying modality & interaction heterogeneity for high-modality representation learning," 2023.
- [227] W. Liu, H. Zhan, H. Chen, and F. Lv, "Multimodal sentiment analysis with missing modality: A knowledge-transfer approach," 2025.
- [228] A. F. Kazerooni, N. Khalili, X. Liu, D. Haldar, Z. Jiang, A. Resnick, S. Bakas, and M. G. Linguraru, "The brain tumor segmentation (brats) challenge 2023: Focus on pediatrics (cbtn-connect-dipgr-asnr-miccai brats-peds)," 2024.
- [229] P. Bergmann, X. Jin, D. Sattlegger, and C. Steger, "The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization," in *VISAPP*, 2022, pp. 202–213.
- [230] S. Jezek, M. Jonak, R. Burget, P. Dvorak, and M. Skotak, "Deep learning-based defect detection of metal parts: Evaluating current methods in complex conditions," in *ICUIMT*, 2021, pp. 66–71.
- [231] P. Mishra, R. Verk, D. Fornasier, C. Piciarelli, and G. L. Foresti, "Vt-adl: A vision transformer network for image anomaly detection and localization," in *ISIE*, 2021, pp. 01–06.
- [232] J. Božič, D. Tabernik, and D. Škočaj, "Mixed supervision for surface-defect detection: From weakly to fully supervised learning," *Comput. Ind.*, vol. 129, p. 103459, 2021.
- [233] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger, and C. Steger, "Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization," *Int. J. Comput. Vis.*, vol. 130, no. 4, pp. 947–969, 2022.
- [234] J. Goh, S. Adepu, K. N. Junejo, and A. Mathur, "A dataset to support research in the design of secure water treatment systems," in *Crit. Inf. Infrastruct. Secur.*, 2017, pp. 88–99.
- [235] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding," in *KDD*, 2018, pp. 387–395.
- [236] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *CVPR*, 2010, pp. 1975–1981.
- [237] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 fps in matlab," in *ICCV*, 2013, pp. 2720–2727.
- [238] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in *CVPR*, 2016, pp. 589–597.
- [239] D. Zhukov, J.-B. Alayrac, R. G. Cinbis, D. Fouhey, I. Laptev, and J. Sivic, "Cross-task weakly supervised learning from instructional videos," in *CVPR*, 2019, pp. 3532–3540.
- [240] Y. Tang, D. Ding, Y. Rao, Y. Zheng, D. Zhang, L. Zhao, J. Lu, and J. Zhou, "Coin: A large-scale dataset for comprehensive instructional video analysis," in *CVPR*, 2019, pp. 1207–1216.
- [241] A. Acsintoa, A. Florescu, M.-I. Georgescu, T. Mare, P. Sumedrea, R. T. Ionescu, F. S. Khan, and M. Shah, "Ubnormal: New benchmark for supervised open-set video anomaly detection," in *CVPR*, 2022, pp. 20 143–20 153.
- [242] L. H. Chiang, E. L. Russell, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*, 2000.
- [243] J. Van Impe and G. Gins, "An extensive reference dataset for fault detection and identification in batch processes," *Chemometr. Intell. Lab. Syst.*, vol. 148, pp. 20–31, 2015.
- [244] S. Han, X. Hu, H. Huang, M. Jiang, and Y. Zhao, "Adbench: Anomaly detection benchmark," *NeurIPS*, vol. 35, pp. 32 142–32 159, 2022.
- [245] C. Cao, Y. Lu, P. Wang, and Y. Zhang, "A new comprehensive benchmark for semi-supervised video anomaly detection and anticipation," in *CVPR*, 2023, pp. 20 392–20 401.
- [246] C. Cao, H. Zhang, Y. Lu, P. Wang, and Y. Zhang, "Scene-dependent prediction in latent space for video anomaly detection and anticipation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 1, pp. 224–239, 2025.
- [247] Y.-H. Chen, R. Sarokin, J. Lee, J. Tang, C.-L. Chang, A. Kulik, and M. Grundmann, "Speed is all you need: On-device acceleration of large diffusion models via gpu-aware optimizations," in *CVPR*, 2023, pp. 4651–4655.
- [248] K. Farid, S. Schrödi, M. Argus, and T. Brox, "Latent diffusion counterfactual explanations," 2023.
- [249] Z. Hu and M. Rostami, "Cognitively inspired cross-modal data generation using diffusion models," 2023.
- [250] B. J. Lengerich, S. Bordt, H. Nori, M. E. Nunnally, Y. Aphinyanaphongs, M. Kellis, and R. Caruana, "Llms understand glass-box models, discover surprises, and suggest repairs," 2023.
- [251] M. Chen, L. Wei, H. Cao, W. Zhou, and S. Hu, "Explore the potential of llms in misinformation detection: An empirical study," 2024.
- [252] M. Azarafza, M. Nayyeri, C. Steinmetz, S. Staab, and A. Rettinger, "Hybrid reasoning based on large language models for autonomous car driving," in *ICCM*, 2024, pp. 14–22.
- [253] Z. Zhang, H. Amiri, Z. Liu, L. Zhao, and A. Zuefle, "Large language models for spatial trajectory patterns mining," in *ACM SIGSPATIALW*, 2024, pp. 52–55.
- [254] D. S. G. Henrique, A. Kucharavy, and R. Guerraoui, "Stochastic parrots looking for stochastic parrots: Llms are easy to fine-tune and hard to detect with other llms," 2023.
- [255] B. Qi, K. Zhang, H. Li, K. Tian, S. Zeng, Z.-R. Chen, and B. Zhou, "Large language models are zero shot hypothesis proposers," 2023.
- [256] H. Shao, J. Huang, S. Zheng, and K. C.-C. Chang, "Quantifying association capabilities of large language models and its implications on privacy leakage," 2024.
- [257] S. Barua, "Exploring autonomous agents through the lens of large language models: A review," 2024.



Yang Liu Yang Liu received the B.S. degree as an honored undergraduate in automation from Nanjing University, Nanjing, China, in 2020, and the Ph.D. degree in computer science from Fudan University, Shanghai, China, in 2025. During 2023-2025, he was a visiting scholar in the Department of Computer Science at the University of Toronto, Toronto, Canada, as part of a joint Ph.D. program. His research interests focus on anomaly detection, with applications in industry and intelligent transportation systems.

Dr. Liu has published multiple first-author papers in prestigious journals including ACM CSUR, IEEE TIP, IEEE TII, and IEEE TIE. He has served as Area Chair for BMVC 2025 and IEEE ICIP 2025, Workshop/Special Session Co-Chair for IEEE ICIP, IEEE DSP, and IEEE WF-IoT, and TPC member for top-tier conferences including CVPR, ICCV, ICLR, NeurIPS, AAAI, and ACM MM. He is an active reviewer for ACM CSUR, PR, INFU, and multiple IEEE Transactions including TIP, TII, TNNLS, TCSV, TMM, and TCYB. His work has received over 1,300 citations with an h-index of 21.



Wenchao Li Wenchao Li received the B.S. degree in Software Engineering from Beihang University, Beijing, China, in 2023. From 2023 to 2024, he worked as a full-time algorithm researcher at SenseTime Research. He is currently pursuing the Ph.D. degree in the School of Computer Science at the University of Sydney, Australia.

His current research interests include generative models and embodied intelligence, on which he has published multiple first-authored papers at top-tier conferences (like CoreA* level).



Liang Cao (Graduate Student Member, IEEE) is currently a Postdoc at Massachusetts Institute of Technology. He received his Ph.D. degree in Chemical and Biological Engineering from the University of British Columbia in 2024, M.S. degree in Control Engineering and B.S. degree in Automation from Beijing University of Chemical Technology in 2018 and 2015, respectively.

His research interests include process control and modeling, causal machine learning, and fault diagnosis with applications in industrial systems.



Jing Liu (Member, IEEE) received the B.S. and M.S. degrees in Computer Science and Technology and Software Engineering from the University of Shanghai for Science and Technology, Shanghai, China, in 2014 and 2017, respectively, and the Ph.D. degree in Electronics and Information from Fudan University, Shanghai, China, in 2023.

He is currently a Postdoctoral Research Fellow at the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada. His research interests include end-cloud collaboration, anomaly detection, domain generalization, and knowledge transfer.



Jin Wang Prof. Jin Wang is currently a professor in the Department of Computer Science at Soochow University. He received his BSc degree in Mathematic from Ocean University of China, in 2006, and his Ph.D. degree in Computer Science from University of Science & Technology of China (USTC) and City University of Hong Kong (CityU) under the collaboration scheme of USTC and CityU supervised by Prof. Naijie Gu and Prof. Jianping Wang, at Suzhou Institute for Advanced Study. Prior to joining Soochow University, he worked at City University of Hong Kong as research assistant, research associate and research fellow.

Prof. Wang's research interests include Edge Intelligence, Intelligence IoT, Coded Computing and Secure AI. He has been actively publishing more than eighty papers in the high quality international journals (e.g., IEEE TPDS, IEEE TCC, IEEE TVT, IEEE TMM, IEEE TNSE, IEEE CL, etc.) and conferences (e.g., INFOCOM, ICDCS, IFIP Networking, ICC, GLOBECOM, etc.). His research has been supported by National Natural Science Foundation of China, China Postdoctoral Science Foundation and Application Foundation Research of Suzhou. He is currently a reviewer of IEEE JSAC, IEEE TPDS, IEEE Communications Letters, etc. He now is also serving as a Technical program committee (TPC) Member of many high quality international conferences.



Chengfang Li This is Chengfang Li.



Rui Xi This is Rui Xi.

Laurence T. Yang (Fellow, IEEE) received the B.E. degree in computer science and technology and the B.Sc. degree in applied physics from Tsinghua University, Beijing, China, in 1992, and the Ph.D. degree in computer science from the University of Victoria, Victoria, BC, Canada, in 2006. He is a Professor with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China, also with the School of Computer Science and Technology, Hainan University, Haikou, China, and also with the Department of Computer Science, St. Francis Xavier University, Antigonish, NS, Canada.

His research interests include parallel and distributed computing, embedded and ubiquitous computing, and big data. Dr. Yang is a Fellow of the Canadian Academy Engineering and the Engineering Institute of Canada.



Junsong Yuan (Fellow, IEEE) received the B.Eng. degree from Huazhong University of Science Technology (HUST) in 2002, the M.Eng. degree from the National University of Singapore in 2005, and the Ph.D. degree from Northwestern University in 2009 (advised by Prof. Ying Wu). He is currently a Professor and the Director of the Visual Computing Laboratory, Department of Computer Science and Engineering (CSE), State University of New York at Buffalo, USA. Before joining SUNY Buffalo, he was an Associate Professor (2015–2018) and a Nanyang Assistant Professor (2009–2015) with Nanyang Technological University (NTU), Singapore. He received Chancellor's Award for Excellence in Scholarship and Creative Activities from SUNY, Nanyang Assistant Professorship from NTU, Outstanding EECS Ph.D. Thesis Award from Northwestern University, and Best Paper Award from IEEE TRANSACTIONS ON MULTIMEDIA. He was elected as a Faculty Senator at Both SUNY Buffalo and NTU.

He is a fellow of IAPR. He serves as a Senior Area Editor for Journal of Visual Communication and Image Representation (JVCI); and an Associate Editor for IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (T-PAMI), IEEE TRANSACTIONS ON IMAGE PROCESSING (T-IP), IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (T-CSVT), and Machine Vision and Applications (MVA). He also serves as the General/Program Co-Chair for ICME and the Area Chair for CVPR, ICCV, ECCV, and ACM MM.



Wei Zhou (Senior Member, IEEE) is an Assistant Professor at Cardiff University, United Kingdom. Dr Zhou was a Postdoctoral Fellow at University of Waterloo, Canada. Wei received the Ph.D. degree from the University of Science and Technology of China in 2021, joint with the University of Waterloo from 2019 to 2021. Wei was a visiting professor at Dalian University of Technology, visiting scholar at National Institute of Informatics, Japan, a research assistant with Intel, and a research intern at Microsoft Research and Alibaba Cloud. Weis research interests span multimedia computing, perceptual image processing, and computational vision.