Journal of Computer Science and Cybernetics, V.42, N.x (2026), 1–
DOI: 10.15625/1813-9663/24272

GENERATIVE MODELS FOR TIME-LAPSE CELL DATASETS A STRUCTURED REVIEW OF GAN AND DIFFUSION APPROACHES

THANH-HA DO*, MINH TOAN HOANG

Posts and Telecommunications Institute of Technology, Km10, Nguyen Trai Street, Ha Dong Ward, Ha Noi, Viet Nam

Crossref Similarity Check
Powered by iThenticate

Abstract. Time-lapse cell microscopy provides detailed information on cell dynamics, morphology, interactions, and development. However, both the techniques and datasets have limitations that hinder cell analysis. Recent advances in generative models, especially generative adversarial networks (GANs) and diffusion models, offer new tools to alleviate data availability limitations by generating samples that resemble the probability distribution of real data. This review summarizes applications of GANs and diffusion models in time-lapse microscopy of cells. For each group, we analyze representative studies with respect to data, architecture design, evaluation metrics, and biological plausibility. This review provides an insightful overview of the theoretical foundation and a concise summary of existing methods for time-lapse cell microscopy.

Keywords. Time-lapse cell, generative models, diffusion models, time-lapse microscopy.

1. INTRODUCTION

Time-lapse cell microscopy is a powerful tool for observing morphological changes, cell division, motility, and phenotypic transitions [1]. Despite its importance, the cell microscopy data remains challenging due to several limitations. First, acquiring high-quality time-lapse images is challenging due to microscopy techniques and lab environments that introduce low signal-to-noise ratios, variable conditions, and artifacts, often leading to segmentation, tracking, or phenotype classification failures [2, 3]. Second, annotating cell structures and events across long temporal sequences requires substantial expert effort, thereby limiting the scalability of supervised learning approaches [4, 5]. These challenges underscore the urgent need for methods to generate high-quality microscopy data.

Recent advances in generative adversarial networks (GANs) and diffusion models have opened new directions for addressing these limitations. GANs and diffusion models have been widely adopted for microscopy image synthesis, data augmentation, and super-resolution [6, 7, 8]. Their ability to generate realistic high-frequency textures makes them attractive for improving the quality and diversity of cellular images. Meanwhile, diffusion models characterized by stable training dynamics, high sample quality, and flexibility in conditioning, have rapidly become state-of-the-art for many synthesis and restoration tasks. Diffusion-based approaches have recently been adapted for denoising, biological structure reconstruction [7, 8, 9, 10], and even for predicting time-lapse images through adapted video diffusion frameworks [11, 12]. Together, GANs and diffusion models provide promising foundations

*Corresponding author.
E-mail addresses: dothanhha@ptit.edu.vn (T.H. Do); ToanHM.B21CN713@stu.ptit.edu.vn (M.T. Hoang).

for overcoming data scarcity, reducing noise, and modelling dynamic behaviors in biological systems.

However, despite the rapid growth of generative models, applications to time-lapse cell microscopy remain fragmented and relatively limited in scope. Existing works differ widely in their goals, ranging from single-image synthesis to temporal prediction, and employ heterogeneous datasets, architectures, and evaluation protocols [9, 10, 11, 12]. Furthermore, generative models for cell microscopy often prioritize visual realism over biological plausibility and fail to preserve relevant biological patterns.

To address these challenges, this paper offers a structured literature review of GAN- and diffusion-based techniques for time-lapse cell microscopy. The main contributions are as follows. First, methods are organized by model family (GAN-based and diffusion-based). Second, generative applications for data augmentation, synthesis, restoration, and denoising within time-lapse cell microscopy are examined, with these tasks being related to segmentation, tracking, and classification. Third, an illustrative experimental evaluation of four representative models is conducted on the DIC-C2DH-HeLa dataset, providing observational findings on their behavior and relative performance within this specific setting [3]. The representative models were selected based on their prominence in recent literature and their relevance to typical generative tasks in cell microscopy. Evaluation was carried out using quantitative metrics such as peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM), as well as qualitative assessment protocols to ensure a comprehensive comparison. Finally, the paper identifies key challenges, present constraints, and potential directions for future research to support further methodological development and applied use in the field.

The paper is organised as follows. Section 2 introduces the background on time-lapse cell microscopy data, as well as GANs and diffusion models. Section 3 presents a task-oriented taxonomy of GANs and diffusion methods for time-lapse cell microscopy. The experimental results and evaluation are presented in Section 4, and the conclusion is in Section 5.

2. BACKGROUND

2.1. Time-lapse cell microscopy data

Cell microscopy has a wide range of imaging types, including fluorescence, bright-field, phase-contrast, and confocal microscopy for both 3D and 2D images. These microscopies differ in their contrast mechanisms, recording conditions, and resolution, which pose challenges for observation and analysis. Moreover, time-lapse microscopy enables researchers to observe cellular dynamics, migration, apoptosis, and cell-cycle progression by acquiring image sequences over time.

Downstream analysis tasks include cell segmentation and instance detection, multi-object tracking, phenotype or state classification, and event detection. However, the quality and quantity of available data are often limited by acquisition constraints (phototoxicity, imaging speed, storage), variable imaging conditions (illumination changes, focus drift), and the high cost of manual annotation [2, 3, 13, 14]. These factors motivate the use of generative models to synthesize additional data, restore degraded images, and model complex temporal patterns.

In recent years, several public benchmarks have provided time-lapse microscopy sequences for segmentation, tracking, phenotype classification, and biological analysis. In 3D microscopy datasets, 3DeeCellTracker provides three datasets from 3 different cells for tracking

Table 1: Representative public benchmarks for live-cell and time-lapse microscopy. Modality: Fluo, BF, PhC, DIC. Tasks: Seg, Trk, Lin, Mit det, Pheno, Causal, Dyn.

Dataset name	Modality	Dim.	System / scale	Main tasks
3D / complex ecosystems				
3DeeCell Tracker [15]	Fluo	3D+t	Moving/deforming organs (*C. elegans* brain, zebrafish heart) and tumor spheroids	Seg, Trk
CausalXtract tumor-on-chip [16]	Live-cell microscopy	2D+t	Breast tumor-on-chip ecosystem; 2 min sampling up to 48 h (1440 frames)	Trk, Pheno, Causal
Subcellular nanostructures [17]	Confocal Fluo	2D+t	Endo-lysosomal organelles (early/late endosomes, lysosomes)	Det/Seg, Trk, Dyn
2D time-lapse benchmarks				
CTC [3]	Fluo, BF, PhC, DIC	2D/3D	Diverse cell types; standardized GT and ST for segmentation & tracking	Seg, Trk, Lin
DeepSea [18]	PhC	2D+t	Phase-contrast live in mouse embryonic stem cells	Seg, Trk
CTMC [19]	DIC	2D+t	14 different mammalian cells in brightfield/phase-contrast videos with dense mitosis annotations	Trk, Mit det
Ker *et al.* [20]	PhC	2D+t	48 long-term phase-contrast sequences of C2C12 myoblast cells (up to 3.5 days); manual & automated tracks	Trk, Seg
Genome-wide RNAi screen [21]	Fluo	2D+t	H2B-GFP HeLa time-lapse under systematic gene knock-down (48 h; 30 min interval)	Pheno, Mit det, Trk
ALFI datasets [22]	DIC	2D+t	HeLa, U2OS, and hTERT RPE-1 cells under different conditions (237.9 h)	Pheno, Mit det, Seg, Trk

and segmentation [15], while CausalXtract tumor-on-chip is used to infer dynamic causal relationships between cells and treatments [16], and Subcellular nanostructures is a benchmark dataset designed to quantify the structure and dynamics of nanometer-scale organelles [17].

In 2D microscopy datasets, there are the DeepSea dataset for single-cell segmentation and tracking in phase-contrast movies [18], the CTMC "Cell Tracking with Mitosis Detection" challenge [19], the phase-contrast time-lapse collections of Ker *et al.* with automated and manual tracking annotations [20], and the genome-wide time-lapse RNAi screen of Neumann *et al.*, which releases thousands of two-day H2B-GFP HeLa movies of dividing human cells under systematic gene knockdown [21] or ALFI datasets with various annotations that are useful for identifying interphase and mitotic events and reconstructing their lineage [22]. In particular, the Cell Tracking Challenge (CTC) datasets provide 2D and 3D data across various cell types and microscopy modalities [3]. A more detailed overview of representative datasets is given in Table 1.

2.2. Generative adversarial networks

Generative Adversarial Networks (GANs) have become a prominent class of deep generative models for image synthesis. A GAN consists of a generator G that produces synthetic samples that resemble the real data distribution and a discriminator D that learns to distinguish between real and generated samples [23]. The training procedure of GANs can be viewed as a competition between a generator that produces realistic samples to fool the discriminator and a discriminator that learns to identify these generated samples. The input

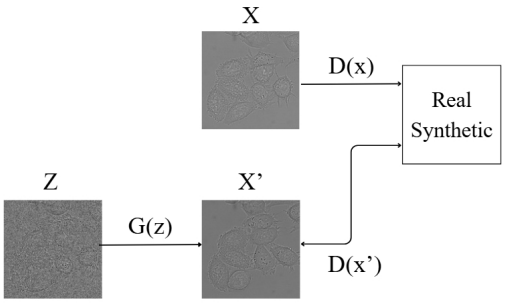

Figure 1: General GAN architecture

of the generator G is random noise z, and it is transformed into a synthetic sample $x' \in X'$ that is similar to the training data distribution. The discriminator takes input samples from both X and X' and attempts to determine whether the generated data is real or synthetic [23]. The objective G is to closely approximate the generated probability distribution to the training data probability distribution by utilizing the information retrieved from D. Consequently, D will have greater difficulty in locating the fake samples after G is adequately educated. The GAN architecture is illustrated in Figure 1.

In recent years, GANs have been used to generate realistic cell images from noise or label maps, augment training data, enhance resolution or contrast, and denoise low-SNR microscopy images [5, 7]. Different GAN-based microscopy methods employ distinct loss functions [6]. These losses can be broadly divided into two groups: adversarial and auxiliary losses. Adversarial losses include vanilla GAN loss, Wasserstein loss, least-squares loss, and relativistic loss, among which vanilla loss is widely used [6]. In the original vanilla GAN formulation, the objective is defined as in Equation 1.

$$\min_G \max_D \mathcal{L}(G, D) = \mathbb{E}_{x \sim p_{\text{data}}} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log (1 - D(G(z)))] , \tag{1}$$

here, x denotes a real microscopy image sampled from the data distribution p_{data}, while z is a latent noise vector sampled from a prior distribution p_z. The discriminator D is optimized to assign high probability to real images and low probability to synthetic images generated by G.

Auxiliary losses are commonly introduced to improve image fidelity and structural consistency. The most common auxiliary losses are pixel-wise reconstruction losses, such as the L_1 and L_2 norms, as indicated in Equation 2.

$$\mathcal{L}\text{pix} = \|x - \hat{x}\|_1 \quad \text{or} \quad \mathcal{L}\text{pix} = \|x - \hat{x}\|_2^2. \tag{2}$$

In this context, where either $\hat{x} = G(z)$ or $\hat{x} = G(y)$ represents the generated image, GANs often combine adversarial and auxiliary objectives as defined in Equation 3 in practical implementations.

$$\mathcal{L}G = \mathcal{L}G^{\text{adv}} + \lambda \mathcal{L}_{\text{aux}}, \tag{3}$$

here, λ controls the contribution of the auxiliary loss, balancing both objectives. This combination of losses is particularly important in microscopy applications, where specific requirements must be met. In these scenarios, generated images should not only appear visually

realistic but also preserve biologically meaningful structures, such as cell boundaries, morphology, and texture patterns. Achieving this requires incorporating auxiliary information, including segmentation masks, class labels, and imaging conditions [24]. This enables more guided and task-specific synthesis, as demonstrated by architectures such as Pix2Pix [25].

Moreover, CellCycleGAN, a conditional GAN, employs multiple statistical models to model morphological features [26]. However, this approach generates sequenced images via predefined 6 phases of the cell cycle under morphology conditions, rather than learning the full sequence dynamics. Despite their success, GANs are highly sensitive to hyperparameters and can suffer from mode collapse when the model fails to capture all modes of the data distribution, leading to samples with limited variability [27, 28]. Only a few recent works extend GANs to model time-lapse sequences or video-like cell dynamics directly; these approaches will be discussed in detail in Section 3.1.

2.3. Diffusion models

Diffusion models have recently achieved state-of-the-art performance in natural image synthesis [29, 30]. Their core mechanism has two stages. First, a forward diffusion process gradually corrupts data by adding Gaussian noise. Second, a learned reverse process, parameterized by neural networks, denoises the corrupted samples to reconstruct clean data [29, 31]. By learning reverse dynamics via a denoising objective, diffusion models generate high-quality, diverse samples by starting from randomness and progressively refining them [29, 32].

A widely used formulation is the Denoising Diffusion Probabilistic Model (DDPM), which uses variational inference to approximate the target data distribution [29, 33]. DDPM defines the forward diffusion process as a Markov chain in which Gaussian noise is incrementally added to the original data sample. Let $\mathbf{x}_0 \sim q(\mathbf{x}_0)$ denote a clean sample drawn from the data distribution. The forward noising process is defined as in Equation 4.

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}\right), \quad t = 1, \ldots, T, \tag{4}$$

where T denotes the total number of diffusion steps, $\beta_t \in [0, 1)$ is the variance schedule at time step t, $\mathbf{I}$ is the identity matrix, and $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

By defining $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, an arbitrary noisy sample $\mathbf{x}_t$ can be directly obtained from the clean sample $\mathbf{x}_0$ as

$$q(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}\right). \tag{5}$$

Equivalently, $\mathbf{x}_t$ can be sampled using the reparameterization form

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{6}$$

The reverse process aims to recover a clean sample from noise by learning the conditional distribution $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$. Starting from a Gaussian noise sample $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, DDPM generates data by iteratively applying the learned reverse transitions

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T) \prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t), \tag{7}$$

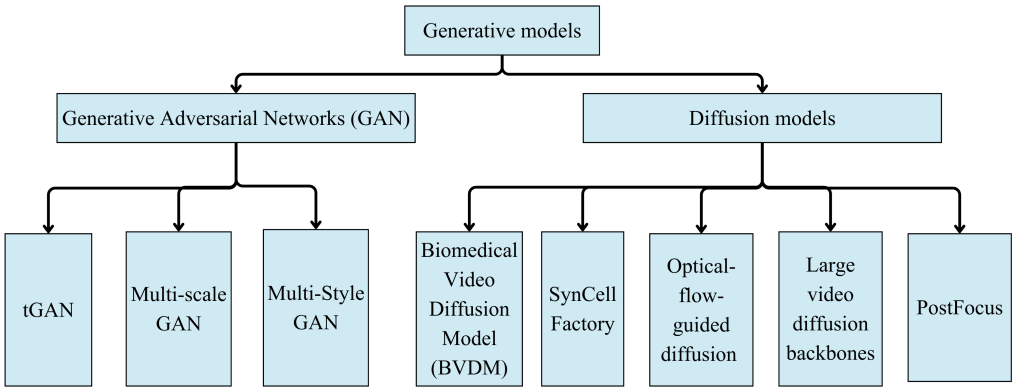

Figure 2: Task-Oriented taxonomy of GANs and diffusion methods for time-lapse cell microscopy.

where each reverse transition is commonly modeled as a Gaussian distribution

$$p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t) = \mathcal{N}\left(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)\right), \tag{8}$$

here, $p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$ denotes the learned denoising step that maps a noisy sample $\mathbf{x}_t$ to a cleaner sample $\mathbf{x}_{t-1}$.

For microscopy and bioimage analysis, diffusion models have been adapted for tasks such as denoising fluorescence images, super-resolution reconstruction, inpainting missing structures [8, 10, 34], and, more recently, video or time-lapse prediction via extensions of image diffusion models to the temporal domain [11, 12]. Compared to GANs, diffusion models typically exhibit more stable training dynamics and better mode coverage, at the cost of increased computational requirements due to the multi-step sampling process. Their flexibility in conditioning on auxiliary information (segmentation maps, time indices, text prompts, or physical parameters) makes them a promising tool for controllable biological image synthesis and restoration [34].

Most existing diffusion-based approaches for microscopy, however, still operate on individual frames or short 2D patches, and only a small number of methods explicitly learn to generate full time-lapse sequences of cells. In Section 3.2 we review these sequence-level diffusion models and analyze their potential for simulating and augmenting cell dynamics.

3. TASK-ORIENTED TAXONOMY OF GANS AND DIFFUSION METHODS

In this section, we present a task-oriented taxonomy of GANs and diffusion methods for time-lapse cell microscopy. Rather than grouping works solely by architecture, we categorize them according to their primary application: (i) GAN-based generative models and (ii) diffusion-based generative models. Within each category, we discuss representative studies that address datasets, architectures, evaluation metrics, and biologically plausible synthetic microscopy data. The categorization is illustrated in Figure 2.

Table 2: Summary of representative generative models for time-lapse cell microscopy.

Method	Family	Main Task	Core Idea	Key Limitation
tGAN, 2025 [35]	GAN	Cell tracking augmentation	Dual-resolution GANs for video-to-video and image-to-image.	Requires a specialized multi-stage pipeline.
Multi-StyleGAN, 2021 [32]	GAN	Time-lapse fluorescence microscopy simulation	Multi-domain StyleGAN sequence generator with U-Net discriminator.	Limited sequence length and benchmarking depth.
Multi-scale GAN, 2020 [36]	GAN	Temporal super-resolution	Multi-scale GAN based on a Laplacian pyramid.	Risk of hallucinated intermediate dynamics and limited generalization.
SynCellFactory, 2024 [37]	Diffusion	Cell tracking augmentation	Denoising diffusion U-Net with dual ControlNets.	Limited biological fidelity and long-sequence robustness.
BVDM, 2024 [38]	Diffusion	Biomedical video generation	DDPM-based texture modeling with statistical shape model and flow propagation.	Limited fine biological control and cross-domain generalization.
PostFocus, 2024 [39]	Diffusion	Focus restoration	CoAT-Net focus assessment followed by DDPM-based restoration.	Not designed for broader simulation or data augmentation.
Video Diffusion Adaptation, 2025	Diffusion	Conditioned microscopy video generation	Fine-tuned pretrained video diffusion model.	High computational cost and weaker interpretability.
Optical-Flow Diffusion, 2022 [40]	Diffusion	Neural cell video synthesis	Video diffusion model guided by optical flow.	Sensitive to optical-flow errors and pipeline complexity.

3.1. GAN-based generative models for time-lapse cell microscopy

GAN-based models are the earliest generative approaches applied to time-lapse microscopy. Their main capability is in effective image or video synthesis, especially when paired with explicit conditioning. As shown in Table 2, tGAN operates in two stages to produce annotated microscopy sequences for cell-tracking augmentation from binary mask sequences. First, a low-resolution video-to-video stage produces cell images depicting spatial and temporal patterns, with a Generator inspired by the 2D-Unet and a PatchGAN discriminator. Next, a high-resolution super-resolution module enhances these images by adding fine morphological details at the single-cell level, using a U-Net-based generator and a PatchGAN discriminator. They also add training for FlowNet to calculate optical flow losses, enhancing the realism and scientific utility of the synthetic images' sequences. Moreover, tGAN develops an augmentation process to capture both morphological characteristics and diverse generative [35].

Other GAN-based methods focus less on annotation generation and more on simulation or temporal enhancement. Multi-StyleGAN with a U-Net discriminator and Vanilla loss, with R1 regularization, extends StyleGAN-like generation to multi-domain fluorescence time-lapse microscopy (BF and GFP), showing that latent generative models can reproduce visual and temporal patterns of living cell populations [32]. Meanwhile, multi-scale GANs operating at various scales (16×16, 32×32, 64×64) based on a Laplacian pyramid have been used for temporal super-resolution in organ-on-chip microscopy by synthesizing intermediate frames between real observations [36]. These studies show the potential of GANs for efficient video synthesis; however, they may suffer from limited sequence length, instability during training, and the risk of hallucinating biologically implausible dynamics.

3.2. Diffusion-based generative models for time-lapse cell microscopy

Diffusion-based models have recently become more notable due to their ability to generate high-fidelity images through iterative denoising and to incorporate structural or motion priors. BVDM is a statistical shape model for capturing information, a DDPM model for learning texture characteristics, and an FPM (flow prediction model - VovelMorph) for image registration to generate temporally coherent biomedical videos with aligned pixel-level annotations [38]. Similarly, SynCellFactory uses ControlNet-style conditioning to guide diffusion models with cell layout and motion information, enabling controllable generation of annotated cell-tracking videos [37]. These methods are particularly suitable for data augmentation because they preserve an explicit relationship between generated images and annotation masks. Beyond synthetic data generation, diffusion models have also been applied to restoration and domain adaptation. PostFocus uses a CoAT-net for focus assessment and a DDPM for the focus-restoration pipeline, therefore improving the usability of degraded videos [39]. Moreover, optical-flow-guided diffusion models based on 3D U-Net and Optical flow separate the video processing stream and the motion information stream using the Horn-Schunck algorithm, showing that explicit motion priors can improve temporal flow in neural cell video synthesis [40]. More recently, the fine-tuned pretrained video diffusion model based on CogVideoX has been adapted to microscopy sequences under different conditions (text prompts, numeric phenotype scores, numeric embeddings, and image-conditioned generation), suggesting that generative priors can improve biological video generation and discover cell behaviors in various conditions [12].

4. EXPERIMENTS AND EVALUATION

4.1. Strengths and limitations of GANs and diffusion models

GANs and diffusion models exhibit complementary strengths. GANs are attractive for their fast sampling, relatively low inference cost, and long history of adaptation to medical and biological imaging. When well trained, they can produce high-frequency details and sharp textures, which are valuable for super-resolution and realistic data augmentation. Moreover, GANs have a long history of adaptation to various medical imaging modalities and enable semantic manipulation and exploration of generated images [6, 41, 42, 43]. However, adversarial training is notoriously unstable, especially in low-data regimes, and mode collapse can lead to under-representation of rare yet biologically important phenotypes [6, 27, 28].

Diffusion models, by contrast, offer more stable training objectives and improved mode coverage, leading to greater diversity in generated images [29, 32, 34]. The noise-to-signal process is well-suited for denoising, restoration, and super-resolution tasks, and its flexibility enables sophisticated conditioning, thereby enabling fine control in applications such as cell microscopy and time-lapse imaging [8, 10, 34]. Nevertheless, the multi-step sampling process is computationally expensive, resulting in slow inference and high hardware requirements, which can hinder large-scale or real-time applications [29, 32, 34]. In addition, the performance of diffusion models can degrade with limited or low-quality training data, making them less suitable for data-scarce scenarios and challenging in semantic manipulation [34].

Overall, the choice between GANs and diffusion models should be guided by application needs, data availability, computational resources, and the trade-off between sampling speed and diversity/stability.

4.2. Comparison on DIC-C2DH-Hela dataset

To provide a controlled reference point across representative methods, we conducted a small-scale comparison on the DIC-C2DH-HeLa dataset, which has 84 frames with time-step 10 minutes recording on glass plate at a spatial resolution of 0.19 μm $\times$ 0.19 μm [3]. Due to limited compute, we used a batch size of 1 for all runs and trained each method with the maximum epochs feasible within our hardware budget.

We report results from the final checkpoints of each run. PostFocus was trained for 300 epochs and converged to losses on the order of 10^{-2} near the end of training. For the tGAN lower-resolution and super-resolution modules, training stabilized in the last 300 epochs for each module. For BVDM, training was resumed from an existing checkpoint (epoch 1000), achieving a final training loss of 0.0189, whereas SynCellFactory used the training already provided by its own research.

In this experiment, the generated image is evaluated on five metrics: Fréchet inception distance (FID) [44], structural similarity index measure (SSIM), peak signal-to-noise ratio (PSNR) [45], learned perceptual image patch similarity (LPIPS) [46], and their temporal extensions. FID assesses the similarity between the distributions of generated and real images, while LPIPS quantifies perceptual dissimilarity [46, 44]. Moreover, SSIM measures structural similarity based on luminance, contrast, and spatial associations, whereas PSNR evaluates the maximum signal-to-noise ratio using mean squared error (MSE) between images [45]. Finally, temporal consistency was assessed using Multi-Scale SSIM (MS-SSIM) for multi-scale structural similarity and Temporal LPIPS (T-LPIPS) for inter-frame perceptual distance [46, 47].

Table 3: Per-frame image quality evaluation across generative models on DIC-C2DH-HeLa.

Method	SSIM ↑	PSNR ↑	LPIPS ↓	FID ↓
tGAN	0.3628 ± 0.0159	19.85	0.3698	344.54
BVDM	0.2342 ± 0.1389	16.40	0.4515	203.30
SynCellFactory	0.1542 ± 0.1463	12.14	0.7376	376.90
PostFocus	0.5482 ± 0.0539	19.67	0.2307	**114.23**

Overall, PostFocus achieves the best frame-wise fidelity (lowest FID/LPIPS and highest SSIM/PSNR). However, PostFocus is a restoration method on paired degraded/clean frames, rather than a conditional time-lapse generator. Therefore, its strong performance should be interpreted in terms of restoration quality, not as direct evidence of superior long-term sequence generation.

Among these models, BVDM obtains a lower FID than tGAN and SynCellFactory (203.30 vs. 344.54 and 376.9), indicating closer alignment to the real-image distribution under our evaluation setup. In contrast, tGAN achieves higher SSIM/PSNR than BVDM (0.3628/19.85 vs. 0.2342/16.40), indicating better pixel/structure-level similarity to the reference frames, but with a worse distributional distance, as reflected by FID. SynCellFactory yields the weakest frame-wise similarity metrics (SSIM=0.1542, PSNR=12.14) and the highest FID/LPIPS in this experiment, suggesting sensitivity to data scale and domain shift.

Table 4: Temporal consistency evaluation with MS-SSIM and Temporal LPIPS (T-LPIPS) between consecutive frames in real and generated sequences.

	MS-SSIM ↑			T-LPIPS ↓		
Method	Real	Gen.	Δ	Real	Gen.	Δ ↓
tGAN	0.2649	0.2689	+0.0040	0.2032	0.2110	+0.0078
BVDM	0.4211	0.4136	−0.0075	0.3162	0.4092	+0.0931
SynCellFactory	0.3884	0.6371	+0.2487	0.2269	0.2795	+0.0526
PostFocus	0.3903	0.2871	−0.1032	0.2074	0.2582	+0.0508

In Table 4, tGAN shows the most consistent temporal properties, with generated MS-SSIM very close to the real value (0.004) and the smallest ΔT-LPIPS (+0.0078). This indicates that its frame-to-frame structural and perceptual variations are closest to real cell sequences. Similarly, BVDM not only shows a small MS-SSIM gap, but also it has the largest ΔT-LPIPS (+0.0931), suggesting that structural similarity changes as closely as the real sequences, whereas perceptual detail fluctuates more strongly. Furthermore, SynCellFactory's MS-SSIM gap is much higher than the actual value (0.2487), indicating overly transitions and a lack of variation diversity. In contrast, PostFocus shows a lower generated MS-SSIM than real data, suggesting stronger structural variation between time-lapse frames.

Table 5: Comparison of Morphological Metrics: Area and Eccentricity. W_1 denotes the Wasserstein distance (lower is better).

Method	Area			Eccentricity		
	Real	Gen	$W_1 \downarrow$	Real	Gen	$W_1 \downarrow$
tGAN	3517 ± 1226	3346 ± 953	**0.0714**	0.72 ± 0.14	0.74 ± 0.12	**0.0302**
BVDM†	4997 ± 4488	38097 ± 19489	1.2653	0.80 ± 0.15	0.52 ± 0.25	0.4009
SynCellFactory	3440 ± 1564	2119 ± 1957	0.1556	0.74 ± 0.15	0.46 ± 0.28	0.3492
PostFocus	2881 ± 1093	2282 ± 816	0.1012	0.75 ± 0.15	0.71 ± 0.17	0.0435

Table 6: Comparison of Morphological Metrics: Shape Integrity (Solidity) and Movement Dynamics. W_1 denotes the Wasserstein distance (lower is better)

Method	Solidity			Movement ($W_1 \downarrow$)	
	Real	Gen	$W_1 \downarrow$	Disp.	Speed
tGAN	0.90 ± 0.06	0.95 ± 0.04	0.1514	0.1410	0.1025
BVDM†	0.96 ± 0.08	0.97 ± 0.03	**0.0209**	—	—
SynCellFactory	0.92 ± 0.05	0.93 ± 0.06	0.0737	**0.0528**	0.6691
PostFocus	0.95 ± 0.03	0.97 ± 0.01	0.0597	0.1688	**0.0697**

To assess biological plausibility, morphologically consistent, generated, and real images are evaluated on morphological and movement features by Cellpose segmentation [48]. For morphological features, we measure each cell's Area, Eccentricity, and Solidity. In contrast, movement features are tracked using the centroid positions of individual nuclei throughout

the sequence. To compare morphological accuracy and phenotype alignment, we calculate the Wasserstein Distance [12]. The results are shown in Tables 5 and 6. Notably, tGAN achieves the best scores for area ($W_1 = 0.0714$), eccentricity ($W_1 = 0.0302$), and solidity ($W_1 = 0.1514$), suggesting that it preserves cell shapes. Meanwhile, BVDM has a Wasserstein Distance of 1.2653 in area (W_1) and 0.0209 in solidity, likely due to a mismatch between cell-level and frame-level outputs. By comparison, SynCellFactory performs worst on shape metrics (eccentricity $W_1 = 0.3492$, area $W_1 = 0.1556$), but achieves the best displacement score ($W_1 = 0.0528$). In summary, PostFocus demonstrates the most balanced performance, with solid performance ($W_1 = 0.0597$), low eccentricity ($W_1 = 0.0435$), and high speed fidelity ($W_1 = 0.0697$). Overall, the disadvantages of these models suggest that neither adversarial nor conditional synthesis alone captures migrating cell behavior, therefore supporting gap-based future approaches.

4.3. Temporal consistency and long-term dynamics

Modeling long-term biological processes in cell microscopy requires generative models that achieve both spatial realism and temporal uniformity. Most current methods focus on static images or short sequences, but recent advances are beginning to address the challenges of generating temporally coherent, biologically plausible long time-lapse data. In this setting, synthetic sequences should preserve not only smooth frame-to-frame transitions but also plausible cell motion, growth, and replication events, lineage topology, and stable fluorescence or contrast statistics across time.

GAN models often exhibit flickering or inconsistent object shapes across frames, thereby limiting their ability to model smooth, biologically realistic cell dynamics throughout long sequences [6, 35]. Diffusion-based video models mitigate some of these issues but stay computationally demanding and have rarely been trained on large-scale microscopy datasets [12]. Moreover, existing evaluations of temporal generative models are largely qualitative or rely on generic video metrics that can be difficult to capture biologically relevant aspects.

For this reason, temporal evaluation should go beyond measuring similarity between adjacent frames. Instead, biologically grounded temporal readouts, such as migration speed, mitotic timing, cell-cycle duration, trajectory smoothness, and lineage consistency, are needed to determine whether generated sequences reflect realistic cellular dynamics. Furthermore, when synthetic videos are used for pre-training or testing tracking techniques, metrics from benchmarks such as the Cell Tracking Challenge (SEG, TRA) provide a more task-relevant view of utility [3]. However, a key open challenge remains: developing models and training strategies that explicitly encode temporal structure, such as integrating autoregressive, state-space, or biophysical constraints [49, 50]. Additionally, there is a need for temporal evaluation measures that go beyond frame-wise quality and assess the consistency and plausibility of entire trajectories [12, 35, 50].

4.4. Biological plausibility and evaluation protocols

Synthetic time-lapse images must conform to biological rules, such as morphology, division rates, and plausible motion. Nevertheless, much of the existing research still relies primarily on indices, without scientifically assessing whether the generated data preserve biological statistics and structures [32, 35]. At the same time, expert manual inspection

remains subjective and time-consuming. These disadvantages hinder accurate results and affect downstream activities, especially in augmentation, model pre-training, and biological interpretation.

Therefore, ensuring biological plausibility requires tighter integration between method development, domain expertise, and task evaluation protocols [6, 32, 34, 38, 37]. A common strategy is to evaluate generative models by measuring their effects on segmentation, tracking, or phenotype classification using synthetic data. For example, Baniukiewicz et al. showed that GAN-based augmentation can improve segmentation quality for fluorescence images, while geometric transformations can produce biologically implausible artifacts [51]. However, these observed gains require careful interpretation, since improvements in downstream performance do not necessarily reflect preserved biological variability.

A further direction is to use multiview and temporal representation techniques to understand how generative models are affected by biological factors and the type of input, helping provide deep insight into biological states and transition patterns [52].

4.5. Data, benchmarks, and reproducibility

The development of generative models for time-lapse microscopy is significantly limited by the lack of uniform datasets, benchmarks, and reproducibility protocols. Public cell imaging benchmarks often focus on segmentation and tracking, but there are few dedicated benchmarks for time-lapse synthesis or temporal prediction [6]. Existing studies also differ substantially in their choice of metrics, baselines, datasets, and downstream evaluation, making it difficult to compare methods fairly [12, 36, 37].

While image-level scores such as FID, SSIM, PSNR, and LPIPS provide useful reference points, they are not designed for interpreting or inspecting biological plausibility or insight-driven analysis [53]. Many generative models also lack open-source implementations and pre-trained weights, additionally hindering repeatability and fair benchmarking [6, 12, 38].

In order to advance the field, there is a clear need for:

- Public datasets with carefully selected training and test splits for generative evaluation, including static and time-lapse modalities;
- Standardized evaluation that combines image-level, sequence-level, and downstream-task metrics;
- Open-source implementations and pre-trained models to enable reproducibility and fair comparison.

Such resources would not only enable more rigorous benchmarking of GAN and diffusion methods but also accelerate their adoption in real biological workflows.

4.6. Future research directions

Based on current trends and gaps, several promising directions for future work can be identified:

- Scaling and diversifying time-lapse training data: Increasing the scale and diversity of microscopy datasets across cell lines, perturbations, and treatments has been demonstrated to significantly improve the robustness and transferability of generative models.

- Multiview representation input: Enabling models to grasp cell states across time, supporting understanding of how the generated results are affected by morphology and motion features. This helps identify insightful information to develop models that synthesize more plausible biology.

- Advanced conditioning, multi-modal control, and optimized models: Multimodal fusion of text, numeric features, and visual cues in hierarchical conditioning can help develop conditioning strategies, improving diversity and controllability. Moreover, optimizing models of downstream tasks such as segmentation, tracking, and classification would improve their accuracy and productivity.

- In-silico hypothesis testing: Time-lapse generators can support virtual experiments, helping discover enriched or rare behavior within various conditions and treatments. Furthermore, it could help biologists explore and observe different behaviors or treatments in time-lapse cells, without the need for repeated lab assays.

- Biologically grounded evaluation benchmarks: Developing evaluation benchmarks on frame level, time-lapse levels, temporal morphology statistics, lineage descriptors, and biological characteristics. This would help evaluate synthetic results that are not merely visually realistic but also biologically meaningful.

5. CONCLUSION

In this paper, we presented a structured literature review of GAN- and diffusion-based methods for time-lapse cell microscopy. We organized existing work primarily by model family, distinguishing between GAN-based and diffusion-based approaches. Within each family, we analyzed their roles in data augmentation and synthetic image or video generation, image restoration and denoising within time-lapse pipelines, and temporal prediction and simulation of cell dynamics. Furthermore, we examined overview design patterns, strengths, and limitations of current approaches and identified key gaps related to temporal consistency, biological plausibility, and reproducibility.

Despite rapid advances in generative modeling, the use of GANs and diffusion models for time-lapse data and dynamic cell-state analysis still faces challenges. Future research will depend not only on improved architectures and training strategies but also on the development of biological evaluation methods, standardized benchmarks, and integrative frameworks.

REFERENCES

[1] M. H. Qureshi, N. Ozlu, and H. Bayraktar, "Adaptive tracking algorithm for trajectory analysis of cells and layer-by-layer assessment of motility dynamics," *Computers in Biology and Medicine*, vol. 150, p. 106193, 2022. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2022.106193

[2] K. E. G. Magnusson, "Segmentation and tracking of cells and particles in time-lapse microscopy," Ph.D. dissertation, Royal Institute of Technology, Stockholm, Sweden, 2016. [Online]. Available: https://www.base-search.net/Record/e9cbbdba73c6c8b71b31237df195352665c13a6fe1bd225aef977a9c2d4dc79a

[3] V. Ulman, M. Maška, K. E. G. Magnusson, O. Ronneberger, C. Haubold, N. Harder, P. Matula, P. Matula, D. Svoboda, M. Radojević, I. Smal, K. Rohr, J. Jaldén, H. Blau, O. Dzyubachyk, B. Lelieveldt, P. Xiao, Y. Li, S.-Y. Cho, A. Dufour, J.-C. Olivo-Marin, C. Reyes-Aldasoro, J. Solís-Lemus, R. Bensch, T. Brox, J. Stegmaier, R. Mikut, S. Wolf, F. Hamprecht, T. Esteves, P. Quelhas, Ö. Demirel, L. Malmström, F. Jug, P. Tomančák, E. Meijering, A. Muñoz-Barrutia, M. Kozubek, and C. O. de Solórzano, "An objective comparison of cell tracking algorithms," *Nature Methods*, vol. 14, pp. 1141–1152, 2017. [Online]. Available: https://doi.org/10.1038/nmeth.4473

[4] D. N. Lesmes-Leon, M. Miranda, M. Caroprese, G. Lovell, A. Dengel, and S. Ahmed, "Synthesizing annotated cell microscopy images with generative adversarial networks," in *Proceedings of the 17th International Conference on Agents and Artificial Intelligence - Volume 3: ICAART*, 2025, pp. 592–599. [Online]. Available: https://doi.org/10.5220/0013163200003890

[5] Şerban Vădineanu, D. Pelt, O. Dzyubachyk, and K. Batenburg, "Reducing manual annotation costs for cell segmentation by upgrading low-quality annotations," *Journal of Imaging*, vol. 10, 2024. [Online]. Available: https://doi.org/10.3390/jimaging10070172

[6] D. N. Lesmes-Leon, A. Dengel, and S. Ahmed, "Systematic review of generative adversarial networks (gans) in cell microscopy: Trends, practices, and impact on image augmentation," *PLOS One*, vol. 20, 2025. [Online]. Available: https://doi.org/10.1371/journal.pone.0291217

[7] H. Zhang, C. Fang, X. Xie, Y. Yang, W. Mei, D. Jin, and P. Fei, "High-throughput, high-resolution deep learning microscopy based on registration-free generative adversarial network." *Biomedical Optics Express*, vol. 10, no. 3, pp. 1044–1063, 2019. [Online]. Available: https://doi.org/10.1364/BOE.10.001044

[8] A. Saguy, T. Nahimov, M. Lehrman, E. G. de Mariscal, I. Hidalgo-Cenalmor, O. Alalouf, A. Balakrishnan, M. Heilemann, R. Henriques, and Y. Shechtman, "This microtubule does not exist: Super-resolution microscopy image generation by a diffusion model," *Small Methods*, vol. 9, no. 3, p. 2400672, 2025. [Online]. Available: https://doi.org/10.1002/smtd.202400672

[9] S. Kaondal, A. Taassob, S. Jeon, S. H. Lee, H. L. Nuñez, B. A. Akindipe, H. Lee, S. Y. Joo, S. M. D. Oliveira, and O. Argüello-Miranda, "Generative frame interpolation enhances tracking of biological objects in time-lapse microscopy," *bioRxiv*, 2025. [Online]. Available: https://doi.org/10.1101/2025.03.23.644838

[10] P. Osuna-Vargas, M. H. Wehrheim, L. Zinz, J. V. Rahm, A. Balakrishnan, A. Kaminer, M. Heilemann, and M. Kaschube, "Denoising diffusion models for high-resolution microscopy image restoration," *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 4320–4330, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:272704328

[11] S. Deltadahl, J. D. Gilbey, C. V. Laer, N. Boeckx, M. P. G. Leers, T. Freeman, L. Aiken, T. Farren, M. L. Smith, M. Zeina, S. MacDonald, D. Gleghorn, M. F. S. S. M. Y. N. U. S. A. K. W. K. H. de Wit, M. Schut, F. W. Asselbergs, S. Kar, S. Williams, M. Koh, Y. Henskens, N. C. J. de Wit, U. D'alessandro, B. Bah, O. Secka, R. Gupta, S. Trompeter, C. V. Laer, G. A. Awandare, K. A. N. Sarpong, L. N. Amenga-Etego, W. H. Ouwehand, J. H. Rudd, J. H. Rudd, C. Piazzese, J. Taylor, N. S. Gleadall, C.-B. Schönlieb, S. Sivapalaratnam, M. Roberts, and P. Nachev, "Deep generative classification of blood cell morphology," *Nature Machine Intelligence*, vol. 7, no. 11, pp. 1791–1803, 2025. [Online]. Available: https://doi.org/10.1038/s42256-025-01122-7

[12] A. Holmberg, N. Mechtel, and O. Wei, "Adapting video diffusion models for time-lapse microscopy," *ArXiv*, vol. abs/2503.18583, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:277272340

[13] D. H. Vu, T. T. A. Phi, M. C. An, T. T. H. Nguyen, T. H. Nguyen, H. B. N. Le, and V. H. Nguyen, "Prediction of pregnancy outcomes of single vitrified-warmed blastocyst transfer using a combination of the Early Embryo Viability Assessment and morphological assessment," *Vietnam Journal of Biotechnology*, vol. 23, no. 3, pp. 275–295, 2025. [Online]. Available: https://doi.org/10.15625/vjbt-22794

[14] L. H. Son and O. Damilola.I, "Multimedia multimodal artificial intelligence (MMAI): Foundations, challenges, and future directions," *Journal of Computer Science and Cybernetics*, vol. 42, no. 1, pp. 1–24, 2026. [Online]. Available: https://doi.org/10.15625/1813-9663/23021

[15] C. Wen, T. Miura, V. Voleti, K. Yamaguchi, M. Tsutsumi, K. Yamamoto, K. Otomo, Y. Fujie, T. Teramoto, T. Ishihara, K. Aoki, T. Nemoto, E. M. C. Hillman, and K. D. Kimura, "3DeeCellTracker, a deep learning-based pipeline for segmenting and tracking cells in 3D time lapse images," *eLife*, vol. 10, p. e59187, 2021. [Online]. Available: https://elifesciences.org/articles/59187

[16] F. Simon, M. C. Comes, T. Tocci, L. Dupuis, V. Cabeli, N. Lagrange, A. Mencattini, M. C. Parrini, E. Martinelli, and H. Isambert, "CausalXtract, a flexible pipeline to extract causal effects from live-cell time-lapse imaging data," *eLife*, vol. 13, p. RP95485, 2025. [Online]. Available: https://doi.org/10.7554/eLife.59187

[17] G. Ferri, L. Digiacomo, F. D'Autilia, W. Durso, G. Caracciolo, and F. Cardarelli, "Time-lapse confocal imaging datasets to assess structural and dynamic properties of subcellular nanostructures," *Scientific Data*, vol. 5, no. 1, p. 180191, 2018. [Online]. Available: https://doi.org/10.1038/sdata.2018.191

[18] A. Zargari, G. A. Lodewijk, N. Mashhadi, N. Cook, C. W. Neudorf, K. Araghbidikashani, R. Hays, S. Kozuki, S. Rubio, E. Hrabeta-Robinson, A. N. Brooks, L. Hinck, and S. A. Shariati, "Deepsea is an efficient deep-learning model for single-cell segmentation and tracking in time-lapse microscopy," *Cell Reports Methods*, vol. 3, no. 6, p. 100500, 2023. [Online]. Available: https://doi.org/10.1016/j.crmeth.2023.100500

[19] S. Anjum and D. Gurari, "Ctmc: Cell tracking with mitosis detection dataset challenge," *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 4228–4237, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:220892151

[20] D. F. E. Ker, S. Eom, S. Sanami, R. Bise, C. Pascale, Z. Yin, S. Huh, E. Osuna-Highley, S. Junkers, C. Helfrich, P. Liang, J. Pan, S. Jeong, S. S. Kang, J. Liu, R. A. Nicholson, M. F. Sandbothe, P. Van, A. Liu, M. Chen, T. Kanade, L. E. Weiss, and P. G. Campbell, "Phase contrast time-lapse microscopy datasets with automated and manual cell tracking annotations," *Scientific Data*, vol. 5, no. 1, p. 180237, 2018. [Online]. Available: https://doi.org/10.1038/sdata.2018.237

[21] B. Neumann, T. Walter, J.-K. Hériché, J. Bulkescher, H. Erfle, C. Conrad, P. Rogers, I. Poser, M. Held, U. Liebel, C. Çetin, F. Sieckmann, G. Pau, R. Kabbe, A. Wünsche, V. P. Satagopam, M. H. A. Schmitz, C. Chapuis, D. W. Gerlich, R. Schneider, R. Eils, W. Huber, J.-M. Peters, A. A. Hyman, R. Durbin, R. Pepperkok, and J. Ellenberg, "Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes," *Nature*, vol. 464, pp. 721 – 727, 2010. [Online]. Available: https://doi.org/10.1038/nature08869

[22] L. Antonelli, F. Polverino, A. B. Albu, A. Hada, I. A. Asteriti, F. Degrassi, G. Guarguaglini, L. Maddalena, and M. R. Guarracino, "ALFI: Cell cycle phenotype annotations of label-free time-lapse imaging data from cultured human cells," *Scientific Data*, vol. 10, no. 1, p. 677, 2023. [Online]. Available: https://doi.org/10.1038/s41597-023-02540-1

[23] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, vol. 27, 2014, pp. 2672–2680. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2014/file/f033ed80deb0234979a61f95710dbe25-Paper.pdf

[24] L. B. Thu, H. T. Nam, P. D. Trung, and N. L. Minh, "A survey on generative adversarial networks for malware analysis," *Journal of Computer Science and Cybernetics*, vol. 41, pp. 135–161, 2025. [Online]. Available: https://doi.org/10.15625/1813-9663/22522

[25] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5967–5976. [Online]. Available: https://api.semanticscholar.org/CorpusID:6200260

[26] D. Bähr, D. Eschweiler, A. Bhattacharyya, D. Moreno-Andrés, W. Antonin, and J. Stegmaier, "Cellcyclegan: Spatiotemporal microscopy image synthesis of cell populations using statistical shape models and conditional gans," in *IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 2021. [Online]. Available: https://doi.org/10.1109/isbi48211.2021.9433896

[27] I. J. Goodfellow, "Nips 2016 tutorial: Generative adversarial networks," *ArXiv*, vol. abs/1701.00160, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:5216145

[28] A. Al-Yaari and Y. Deng, "Generative adversarial networks: A comprehensive survey," *Systems and Soft Computing*, vol. 8, p. 200460, 2026. [Online]. Available: https://doi.org/10.1016/j.sasc.2026.200460

[29] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 6840–6851. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf

[30] A. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 139, 2021, pp. 8162–8171. [Online]. Available: https://proceedings.mlr.press/v139/nichol21a.html

[31] A. Nguyen, B. Gong, B. T. D. Trinh, T. L. Hong, M. N. Tue, and H. D. Phuong, "Sdags: Smote+forest diffusion-based data augmentation and gbt-based stacking ensemble learning for holistic ai-powered diabetes mellitus prediction," *Journal of Computer Science and Cybernetics*, vol. 40, pp. 147–163, 2024. [Online]. Available: https://doi.org/10.15625/1813-9663/20361

[32] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *International Conference on Learning Representations*, 2021. [Online]. Available: https://openreview.net/forum?id=PxTIG12RRHS

[33] J. N. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," *ArXiv*, vol. abs/1503.03585, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:14888175

[34] A. Kazerouni, E. K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, and D. Merhof, "Diffusion models in medical imaging: A comprehensive survey," *Medical Image Analysis*, vol. 88, p. 102846, 2023. [Online]. Available: https://doi.org/10.1016/j.media.2023.102846

[35] A. Zargari, N. Mashhadi, and S. A. Shariati, "Enhanced cell tracking using a gan-based super-resolution video-to-video time-lapse microscopy generative model," *iScience*, vol. 28, 2025. [Online]. Available: https://doi.org/10.1101/2024.06.11.598572

[36] M. C. Comes, J. Filippi, A. Mencattini, P. Casti, G. Cerrato, A. Sauvat, E. Vacchelli, A. D. Ninno, D. D. Giuseppe, M. D'Orazio, F. Mattei, G. Schiavoni, L. Businaro, C. D. Natale, G. Kroemer, and E. Martinelli, "Multi-scale generative adversarial network for improved evaluation of cell–cell interactions observed in organ-on-chip experiments," *Neural Computing and Applications*, vol. 33, pp. 3671–3689, 2021. [Online]. Available: https://doi.org/10.1007/s00521-020-05226-6

[37] M. Sturm, L. Cerrone, and F. A. Hamprecht, " SynCellFactory: Generative Data Augmentation for Cell Tracking ," in *Proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI*, vol. LNCS 15012, 2024, pp. 304–313. [Online]. Available: https://papers.miccai.org/miccai-2024/752-Paper3680.html

[38] R. Yilmaz, D. Eschweiler, and J. Stegmaier, "Annotated biomedical video generation using denoising diffusion probabilistic models and flow fields," in *Simulation and Synthesis in Medical Imaging*, 2025, pp. 197–207. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-73281-2_19

[39] K.-L. Wu, M. J. Montalvo, P. S. Menon, B. Roysam, and N. Varadarajan, "Postfocus: automated selective post-acquisition high-throughput focus restoration using diffusion model for label-free time-lapse microscopy," *Bioinformatics*, vol. 40, 2024. [Online]. Available: https://doi.org/10.1093/bioinformatics/btae467

[40] M. Serna-Aguilera, K. Luu, N. Harris, and M. Zou, "Neural cell video synthesis via optical-flow diffusion," *ArXiv*, vol. abs/2212.03250, 2022. [Online]. Available: https://doi.org/10.48550/arxiv.2212.03250

[41] Y. Chen, X.-H. Yang, Z. Wei, A. A. Heidari, N. Zheng, Z. Li, H. Chen, H. Hu, Q. Zhou, and Q. Guan, "Generative adversarial networks in medical image augmentation: A review," *Computers in Biology and Medicine*, vol. 144, p. 105382, 2022. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2022.105382

[42] J. Hussain, M. Båth, and J. Ivarsson, "Generative adversarial networks in medical image reconstruction: A systematic literature review," *Computers in Biology and Medicine*, vol. 191, p. 110094, 2025. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2025.110094

[43] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Medical Image Analysis*, vol. 58, p. 101552, 2019. [Online]. Available: https://doi.org/10.1016/j.media.2019.101552

[44] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in Neural Information Processing Systems*, vol. 30, 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a1d694707eb0fefe65871369074926d-Paper.pdf

[45] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *20th International Conference on Pattern Recognition*, 2010, pp. 2366–2369. [Online]. Available: https://doi.org/10.1109/ICPR.2010.579

[46] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 586–595. [Online]. Available: https://api.semanticscholar.org/CorpusID:4766599

[47] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003, pp. 1398–1402. [Online]. Available: https://doi.org/10.1109/ACSSC.2003.1292216

[48] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu, "Cellpose: a generalist algorithm for cellular segmentation," *Nature Methods*, vol. 18, pp. 100–106, 2021. [Online]. Available: https://doi.org/10.1038/s41592-020-01018-x

[49] N. Pandhe, B. Rada, and S. Quinn, "Generative spatiotemporal modeling of neutrophil behavior," in *IEEE 15th International Symposium on Biomedical Imaging (ISBI)*, 2018, pp. 969–972. [Online]. Available: https://doi.org/10.1109/ISBI.2018.8363732

[50] T. Kaiser, M. Schier, and B. Rosenhahn, "Cell tracking according to biological needs – strong mitosis-aware multi-hypothesis tracker with aleatoric uncertainty," *IEEE Transactions on Medical Imaging*, 2025, early access. [Online]. Available: https://doi.org/10.1109/TMI.2025.3583148

[51] S. Tosi, J. J. Merelo, C. Ruiz, T. Bretschneider, P. Baniukiewicz, E.JosiahLutton, and S. Collier, "Generative adversarial networks for augmenting training data of microscopic cell images," *Frontiers in Computer Science*, vol. 1, 2019. [Online]. Available: https://doi.org/10.3389/fcomp.2019.00010

[52] Y. Ji, M. F. A. Cutiongco, B. S. Jensen, and K. Yuan, "Generating realistic single-cell images from cellprofiler representations," *Medical Image Analysis*, vol. 103, p. 103574, 2025. [Online]. Available: https://doi.org/10.1016/j.media.2025.103574

[53] Y. Deo, Y. Jia, T. Lassila, W. A. P. Smith, T. Lawton, S. Kang, A. F. Frangi, and I. Habli, "Metrics that matter: Evaluating image quality metrics for medical image generation," *ArXiv*, vol. abs/2505.07175, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:278501413

Received on March 02, 2026
Accepted on June 09, 2026