Advances in Computer, Signals and Systems (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2026.100216
ISSN 2371-8838 Vol. 10 Num. 2

Advances and Future Prospects of Diffusion Models in Cellular Perturbation Modeling

Hanze Bao

School of Software, Taiyuan University of Technology, Taiyuan, 030000, China

Keywords: Diffusion models; Cellular perturbations; Single-cell transcriptomics; Generative models; Virtual cells

Abstract: Due to the fast development of single-cell transcriptomics technologies, society is provided unprecedented resolution for analyzing cellular heterogeneity and researching gene regulatory networks, and individuals believe that effective solutions to adequately reproduce the cellular perturbations should be urgently identified. Nonetheless, conventional computational algorithms can not efficiently describe non-linear behavior, transitions, or extrapolate across perturbations, so it becomes difficult to satisfy the practical requirements of precision medicine research. Recently, it has been shown that deep generative models (especially diffusion models) have tremendous benefits in cell perturbation prediction because of the excellent distribution learning and strong generation quality of these models. The provided paper is a systematic literature review of the underlying theoretical principles and technological advancements of diffusion models. It fully details all major contributions of representative models like Squidiff in predictions of responses to genetic perturbation and drug simulation. Moreover, it discusses the challenges of the field and its research opportunities depending on computational capabilities, biological insight, and the incorporation of multi-omics.

1. Introduction

The extensive use of the single-cell RNA sequencing (scRNA-seq) technology has enabled characterizing cellular heterogeneity and gene regulatory networks in an organized manner on the cellular level, which is an important source of data that may be utilized to define drug targets, pathologies of disease, and precision treatment programs. A complex and tightly regulated dynamic change will take place on the transcriptomes of a cell in response to a broad spectrum of perturbation types that include drug administration and gene editing and developmental differentiation, and radiation. Such molecular alterations are closely associated with the onset and development of the disease, as well as the mode of action of drugs, and thus form the molecular basis of precision medicine.

However, conventional experimentation techniques are slow and expensive to implement, which makes it challenging to perform systematic large-scale perturbation screens. At the same time, initial computational approaches like linear interpolation also have very low ability to describe nonlinear effects, and deep learning models like variational autoencoders (VAEs) have problems of non-interpretability and lack of extrapolation to perturbations[1], and therefore are unable to address

the scientific requirement of a practical simulation of multivariate perturbation responses at single-cell precision. Hence, highly accurate and useful computational modeling tools to predict and simulate perturbation responses, reduce experiment costs and speed up the drug discovery process have become one of the leading topics of research within the area of computational biology.

Deep generative models, such as diffusion models, have made remarkable advances in areas such as computer vision, natural language processing, and protein structure prediction, and are finding more applications in single-cell transcriptome modeling. Such diffusion models are actual distributions of the data throughout the bidirectional process as the slow introduction of noise and repeated efforts to remove it take place[2]. It is an exceptional tool that can be used to precisely model the complex, non-linear cellular state changes and is especially useful in predicting cellular responses in the presence of multiple disturbances and cellular types.

The provided paper is a systematic review of the recent advances in the field of diffusion model research and the associated deep generative models (e.g., VAE and graph neural network, GNNs) in relation to cellular perturbation predictive modeling. The study will entail looking at the technical innovations and efficacy of the early models, including Squidiff, and will provide some directions into the future and future direction of the research on how to build virtual cells.

2. The Biological Basis of Cellular Perturbation and Its Data Characteristics

The physiological definition of cellular perturbation refers to the disruption of the state of the cell due to the influence of external factors or endogenous signals. The classes might be divided into four broad groups based on the source of perturbation and how they act: all the range of experimental perturbation and natural development, including physical perturbation, chemical perturbation, genetic perturbation, and even natural perturbation. Physical perturbations that have a common use are ionizing radiation and heat stress, i.e. Neutrons can induce double-strand breaks in DNA and can affect the development process of vascular organs systematically; drug treatment and cytokine activation may be regarded as chemical perturbations, e.g. Panobinostat (histone deacetylase inhibitor) and etoposide (topoisomerase inhibitor) have their biological effects by blocking certain signaling pathways, whereas cytokines like granulocyte colony-stimulating factor (G-CSF) may help to repair the damage done by radiation. The term gene perturbation refers to modifications in the expression of specific genes by manipulating them using gene knockout, overexpression, or, in other cases, more precise CRISPR-Cas9 editing. One frequent case is the deletion of both transcription factors, ZBTB25 and protein tyrosine phosphatase, PTPN12, and this leads to a cellular reaction that is extremely non-additive. Natural perturbations are cell state transitions triggered by endogenous signals, including cell differentiation and transdifferentiation, i.e., an intended differentiation of an induced pluripotent stem cell (iPSC) to three-germ-layer cells[3].

Single-cell-based perturbation modeling has four main technical issues. To begin with, dynamic transitional states are hard to capture properly; cell state changes induced by perturbation persist as a continuous process in time (temporal continuum), although classical models are prone to approximate discrete states, missing important information about intermediate states. Second, the cumulative action of different perturbations exhibits nonlinear behaviour, which is notoriously challenging to model accurately: gene-drug interactions are inherently nonlinear because the cumulative effect is not the linear sum of the perturbation effects of individual genes. A standard example is the non-additive phenotype in dual-gene knockout experiments. Third, it is challenging to predict new perturbations under small sample conditions: In some cases, the size of the sample population is so tiny that it would be extremely expensive and time-consuming to conduct perturbation experiments, and, additionally, conventional supervised learning models cannot

extrapolate, they cannot predict new perturbation sets that were not included in the training set. Seventh, the biological interpretation of model predictions is weak: Most of the existing deep learning models do not provide any good correspondence between their predictions and particular mechanisms of gene regulation, which constrains their application in experimental design[4].

The existing public databases remain among the most useful information resources applied in perturbation modeling research. Common-purpose databases have GEO (Gene Expression Omnibus), with a vast amount of single-cell perturbation data, namely, GSE133344 gene perturbation data and GSE148842 drug perturbation data[5], and the Zenodo platform can be used to access datasets on natural perturbations, including iPSC directed differentiation[6]. In certain databases, the Sci-Plex database contains high-throughput data from single-cell sequencing in response to chemical perturbation[7], and the figshare platform has raw data from perturbation experiments on vascular organoids[8]. Data sources include numerous states of biochemistry, such as gene perturbation, drug treatment, and cell differentiation, and may serve as a great source of data to develop, train, and systematically test perturbation modeling algorithms.

3. Core Principles, Technological Evolution, and Adaptation to Biological Data of Diffusion Models

3.1 Core Principles of Diffusion Models

The diffusion model is a deep generative framework that is built upon a probabilistic generative framework. The main idea is to let the model learn the real distribution of the data by simulating a sequence of noisy addition-denoise reverse processes, eventually enabling the model to sample noise randomly and produce samples that are similar in distribution to the original data. This model can be reduced to two steps, which include the process of noise addition and the process of denoising with the aim of optimization being the loss of noise prediction.

The noise adding procedure has the characteristics of Markov chains, i.e., the states are changing at each discrete time step T based solely on the state of the dataset at the previous step. With the original single-cell transcriptomic data x_0, at any given number of time steps T, the model follows a certain noise schedule β_t to add Gaussian noise that is standard normal to the existing data x_{t-1} at every step, leading to With the progression of the noise addition process, the information entropy of the data will increase over time, and finally, T steps later, will be equal to the standard Gaussian distribution $q(x_T) \approx N(0,I)$, completely obscuring the structural characteristics of the pure data. One can view this as a gradual movement of biologically relevant patterns of cellular state expressions into a noise space in the modeling context of single-cell transcriptomics, which can be viewed as an extension of the evolutionary character of the data, and therefore has a good parallel to the mathematical description of cellular dynamics changes due to perturbations.

The opposite of denoising is the reverse of the forward process. Its goal is to teach a denoising network ε_θ with parameters θ (typically, it is a neural network) such that the model may begin with a randomly noisy vector x_T, predict the noise present in each step, and finally recover a sample conditioned upon the initial data distribution $p(x_0)$. The denoising network takes the current time step t and the associated data x_t as inputs to implement one step inference, i.e., to generate the estimate of the noise ε, $\varepsilon_\theta(x_t,t)$, which is used to estimate the denoised estimate of the previous step, f(x)=f(t-1) through the application of Bayesian inversion. After T iterations, the model finally presents a transcriptomic sample that corresponds to the desired cell state distribution.

The basic noise prediction loss is a core loss term in the diffusion model that seeks to reduce the prediction error of the denoising network 0theta by predicting real superimposed noise epsilon. The loss is very simple in its architecture and extremely stable in training, and has no common problems

in VAE and GAN training, including mode collapse and gradient instability, so the model can be trained to learn more complicated probability distributions of cellular transcriptomic data.

3.2 Technological Evolution of Diffusion Models and Adaptation to Single-Cell Data

The development of generative models in single-cell perturbation modeling took on a specific course of iterative method development based on the improvement of variational autoencoders (VAE) and generative adversarial networks (GAN) methods to diffusion models. Three of them can be used to show the intrinsic differences among them because they do not have any technical restrictions and can be adjusted to the specificity of single-cell data.

One of the oldest deep generative methods employed in popular applications is VAE, which uses data compression and mapping between high-dimensional data of single-cell transcriptomic data into a low-dimensional latent space via an encoder and subsequently regenerates and generates the low-dimensional data once more with a decoder, giving the first probability model of the cell state and dimensionality reduction analysis[9]. Nevertheless, VAE tends to produce infinite generated samples when there is an independent Gaussian prior to the latent space, which restricts its capability to simulate the continuous dynamics imposed by cellular perturbation and the delicate boundaries between cellular states. It is inherently flawed in respect to the representation of complex nonlinear relationships.

The adversarial training of GANs, where the generator and the discriminator work against each other, could be very beneficial in improving realism and diversity of the generated examples and might also offer some benefits in modeling cell state issues. Moreover, GAN training is seen as a 2-player zero-sum game whose Nash equilibrium does not converge. Its weaknesses include the usual ones, e.g., failure to converge and mode collapse. Moreover, because of its non-Markovian nature, it cannot appropriately represent continuous transitional processes that take place in cell perturbation, thus limiting its use in cases when the accuracy of predicting the behavior of the system under perturbation is of paramount importance[10].

The diffusion model is seen as an innovative advancement in comparison to the two approaches discussed earlier. Combining the capability of neural networks to represent themselves and the repetitive action of adding noise to the front and removing it backward, diffusion models can both precisely model the continuum dynamic behavior that results from perturbing one cell and produce high-quality examples of cell states at a steady pace and be highly resilient to modeling in the presence of complex, multi-perturbation, and non-linear responses. The training process is fixed, and the variety of examples generated by them has been growing steadily, and finally, became the highest-ranking core generation system of computational modeling of cellular perturbations. To sum up, this technology is the implementation of the recent enhancement of the flexibility of deep generative models to high-dimensional, sparse, continuous, and non-linear single-cell data.

4. Generation of Cellular Perturbations and Prediction of Effects Based on Diffusion Models

Consider Squidiff as one of the current uses of the systematic use of diffusion models to forecast how a specific cell would react to a perturbation using the structure of the diffusion model. The main objective of its architecture is to achieve two major goals, which are precisely to capture the dynamics of transcriptomic change caused by perturbation of single cells and enhance the accuracy of predictions of perturbation responses. Due to the inherent characteristics of single-cell data, Squidiff has been optimized in various ways depending on the traditional diffusion model paradigm and has proven to be more predictive when compared to other competitive algorithms in traditional contexts such as predicting the response to gene perturbation and drug response simulations.

Another major technological novelty of the Squidiff model is the use of the double-path

conditional variable mechanism, which uses two terms: (1) Semantic embedding of cell type and (2) Semantic embedding of perturbation type. The scheme also includes incorporating previous biological knowledge, namely cell-type identifiers and perturbation types, into low-dimensional embedding vectors and including them as conditional signals in the training and inference steps of the diffusion model to produce a distinctive transcriptomic response under particular perturbation conditions. The preprocessing phase of the model will provide standardized and normalized raw single-cell transcriptomic data and choose highly variable genes (HVGs) based on their variance, such that it retains most of the core gene properties that are highly correlated with the perturbation response and minimizes the influence of data noise. The information on cell types and perturbation types is transformed into dense embedding vectors in the condition-guided step, which are then included in the latent representations of the noisy-processed transcriptomic data and introduced to the denoising network. Thus, the model could be trained in a dynamic way on the patterns of responses of different cell types in various perturbation conditions.

The Transformer is used in denoising network of Squidiff as the primary model, and this specific Transformer makes use of the full potential of the multi-head self-attention mechanism to implicitly discover difficult regulatory interactions within the high-dimensional gene space, and therefore, it is capable of learning long-range interdependencies among various gene programs. At the same time, there are some residual connections to overcome the problem of vanishing gradients when training deep networks, optimize the model parameters, and eventually predict accuracy. Also, the noise scheduling scheme, which is unique to Squidiff because it was designed to be used in the context of diffusion, relies on the statistical properties of individual-cell transcriptomic data. It also uses a cosine noise schedule, which indicates that the signal decay function as it adds noise resembles the biodynamic behavior of cell disruption even more, enhancing the model's capacity to read and recreate the cellular transitional states.

Squidiff deals with two major prediction issues in its operations. Concerning the response to gene perturbation, the model may, based on the information about the knockout or overexpression of a particular target gene, predict the systemic remodeling of the downstream transcriptome, by accurately predicting the rearrangement of the downstream regulatory networks due to the perturbed gene followed by the reorganization of the entire gene expression patterns; To predict drug response, the model uses molecular data such as drug type and concentration to predict cellular phenotypic effects like changes in cell viability, apoptotic signal activation and other gene expression signatures, and hence offers valuable computer assistance to in vitro screening and evaluation of drug efficacy of drug candidates. The reference findings indicate that Squidiff has improved accuracy indicators in terms of the prediction of perturbation response compared to traditional modeling techniques (such as VAE and GNN). Also, it is useful in recovering the dynamical transition states under cell perturbation, which acts as an important methodological premise on which diffusion models could be applied to the context of perturbation of cells modeling.

5. Current Technical Challenges and Biological Limitations in the Field

Even though diffusion models (or deep generative methods) have achieved much in the area of cellular perturbation modeling, existing studies still have multiple systemic barriers that hinder the development of more clinical translation and application to larger populations.

The computational demand of the inference algorithm in diffusion models requires thousands or tens of thousands of denoising iterations. It is 10 times more expensive than one-pass forward inference models such as VAEs and GANs. This feature has the most apparent manifestation in the case of large amounts of single-cell data, where it is required to process hundreds of thousands of cells, and in practice, this requires much time and computational costs that restrict the possibility of

scaling the model to high-throughput experimental conditions.

Deep generative models have a black-box nature with respect to biological interpretability, and therefore it is extremely difficult to derive clear causal relationships between the models and particular biological processes. Many of the existing diffusion models are not equipped with efficient strategies to integrate previous biological information, including gene regulatory networks and pathway analysis, and therefore they cannot make predictions that can be used by experimental biologists in an effective way, and this constrains the effective translation of model predictions down to the computational level to the experimental validation level.

Regarding the accuracy of prediction and generalization abilities, the available models are susceptible to failure at different levels of predictive power when provided with an entirely new perturbation that was not observed in training or when used on data generated by cross-laboratory or cross-batch data.

6. Future Research Directions and Conclusions

6.1 Future Research Directions

The other possible future studies on this topic could also be informed by some more recent developments in speeding up the inference in the diffusion model, including successful sampling techniques like the Denoisy Diffusion Implicit Model (DDIM)[11] and the Consistency Model[12]. These effective sampling methods can also greatly reduce the number of iterations needed to generate one inference and it is extremely valuable since it reduces the quantity of computation resources consumed and the prediction quality and provides the technical foundation to use models on high-throughput single-cell data. In terms of model architecture, the bottleneck of limited generalization abilities of existing models will be resolved by investigating and formulating a universal foundation model paradigm that can be applied to the single-cell perturbation problem by training models on large-scale and multi-scenario environments with large amounts of pre-train data and then optimizing the parameters effectively to do downstream tasks.

When discussing the combination of multi-omics information, there are no other perturbation modeling approaches currently available that rely purely on data inputs in the transcriptomic field. Future studies ought to aim at investigating how to incorporate multi-modals of omics data such as proteomics, epigenomics, and spatial transcriptomics in a systematic way, and develop multi-modal cell perturbation response models that can simultaneously explain gene regulation, epigenetic alterations, and spatial microenvironmental components. Such models will be able to describe the function of perturbations on a more profound level, and increase both the biological validity and predictive value.

In order to improve the biological interpretability, one of the main methods that could be used in this regard is the amalgamation of the diffusion models with the mechanisms of background knowledge, including the gene regulatory networks and the cellular pathway maps. The application of causal inference models to create perturbation-phenotype causal graphs might correspond to the predictive power of models to regulatory events that have obvious biological relevance; that is, it might offer extra falsifiable computational hypotheses in one of the contexts of precision medicine research.

The digital twin system of virtual cells[13], which is the basis of the diffusion model to serve as the core generation engine with the integration of multi-omics data, cellular morphology, and intercellular interaction networks, is going to be one of the emerging frontiers of computational biology and precision medicine in the future and it will offer a new technology platform to the simulation of the complex biological processes in vitro and hasten the process of drug target discovery and preclinical experiments.

6.2 Conclusions

The current paper is a systematic review of the recent research on diffusion models in the context of cellular perturbation modeling. Theoretically, diffusion models may serve as a powerful probabilistic modeling system capable of modeling high-dimensional, non-linear, and continuous single-cell transcriptomic data through a bidirectional iterative process of noise addition and denoising that essentially eliminates the inherent constraints of VAEs and GANs with regard to the problem. Novel diffusion models such as Squidiff have made important contributions to the accuracy of gene perturbation and drug response prediction problems in terms of methodological contribution by implementing conditional semantic embeddings, denoising network optimization, and noise scheduling. In the future, new developments in other fields including, but not limited to, optimization of computational efficiency or integration of multi-omics data with enhanced bio-interpretability, will help find other potential applications of diffusion models in the domain of cell perturbation modeling and potentially in virtual cellular systems. It is therefore anticipated that they will make an enormous contribution toward the computational progress of precision medicine and drug discovery.

References

[1] Kingma, D. P. & Welling, M. (2014). Auto-encoding variational Bayes. In 2nd International Conference on Learning Representations (ICLR 2014) (eds Bengio, Y. & LeCun, Y.).

[2] Fuest, Michael, et al. (2026). "Diffusion models and representation learning: A survey." IEEE Transactions on Pattern Analysis and Machine Intelligence.

[3] He, S., Zhu, Y., Tavakol, D.N. et al. (2026). Squidiff: predicting cellular development and responses to perturbations using a diffusion model. Nat Methods 23, 65–77.

[4] Lei Sun, Jingwen Wang, Peng Hu, Xiuqing Mao, Cuiyun Hu, Zhihong Wang. (2026). IG-3D: Integrated-Gradients 3D Optimization for Private Transformer Inference[J]. Computers, Materials & Continua, 87 (5): 1158-1176.

[5] Clough, Emily, and Tanya Barrett. (2016). "The gene expression omnibus database." Statistical genomics: methods and protocols. New York, NY: Springer New York, 93-110.

[6] Sicilia, M. A., García-Barriocanal, E., & Sánchez-Alonso, S. (2017). Community curation in open dataset repositories: insights from Zenodo. Procedia Computer Science, 106, 54-60.

[7] Srivatsan, S. R., McFaline-Figueroa, J. L., Ramani, V., Saunders, L., Cao, J., Packer, J., ... & Trapnell, C. (2020). Massively multiplex chemical transcriptomics at single-cell resolution. Science, 367 (6473), 45-51.

[8] Singh J. (2011). Journal of pharmacology & pharmacotherapeutics, [J]. FigShare, 2(2): 138.

[9] Dai B, Wipf D. Diagnosing and enhancing VAE models[J]. arXiv preprint arXiv:1903.05789, 2019.

[10] Wang, Kunfeng, et al. (2017). "Generative adversarial networks: introduction and outlook." IEEE/CAA Journal of Automatica Sinica 4.4 : 588-598.

[11] Zhang, Qinsheng, Molei Tao, and Yongxin Chen. (2022). "gddim: Generalized denoising diffusion implicit models." arXiv preprint arXiv:2206.05564.

[12] Song, Y., & Dhariwal, P. (2024). Improved techniques for training consistency models. In International Conference on Learning Representations (Vol. 2024, pp. 15078-15097).

[13] Bunne, C., Roohani, Y., Rosen, Y., Gupta, A., Zhang, X., Roed, M., ... & Quake, S. R. (2024). How to build the virtual cell with artificial intelligence: Priorities and opportunities. Cell, 187(25), 7045-7063.