

Spatial Diffusion for Cell Layout Generation

Chen Li^{*1}, Xiaoling Hu², Shahira Abousamra¹, Meilong Xu¹, and Chao Chen¹

¹ Stony Brook University, Stony Brook, NY, USA

² Harvard Medical School, Boston, MA, USA

Abstract. Generative models, such as GANs and diffusion models, have been used to augment training sets and boost performances in different tasks. We focus on generative models for cell detection instead, i.e., locating and classifying cells in given pathology images. One important information that has been largely overlooked is the spatial patterns of the cells. In this paper, we propose a spatial-pattern-guided generative model for cell layout generation. Specifically, a novel diffusion model guided by spatial features and generates realistic cell layouts has been proposed. We explore different density models as spatial features for the diffusion model. In downstream tasks, we show that the generated cell layouts can be used to guide the generation of high-quality pathology images. Augmenting with these images can significantly boost the performance of SOTA cell detection methods. The code is available at <https://github.com/superlc1995/Diffusion-cell>.

Keywords: Diffusion Model · Cell Layout · Pathology Images.

1 Introduction

Cell detection focuses on identifying and locating multiple types of cells in images or videos. Although deep models have achieved great success in cell detection tasks [1,17,14,37,16,34], their deployment in the real world is largely constrained due to the demand for large amounts of labeled training data. Collecting and annotating cell detection datasets is high-cost.

Despite the rich literature on the generative model, its application in cell detection is still limited. Although GAN or diffusion models can generate high-quality images with a single or a small number of objects, these models cannot generate images with hundreds or thousands of cells. The primary issue is the lack of an explicit modeling of cell spatial distributions. Numerous factors, including cell-cell interactions, morphogenesis, and cellular functions, make the cells follow a specific spatial pattern [36]. If a generative model cannot learn these spatial distributions, it will not be able to generate realistic images for data augmentation purposes. This issue, however, has been largely overlooked by existing generative models. Most existing methods either generate object layouts randomly [15] or use existing layouts directly [10].

^{*} Email: li.chen.8@stonybrook.edu.

Another constraint for layout generation is the backbone methods. Generative adversarial network (GAN) [5,38,6,28,26,2] cannot handle the large amount of densely packed cells. Recent years have witnessed the rise of the diffusion model [8]. Diffusion models outperform GANs in image generation [9], and have shown great potential in many other tasks. The diffusion model’s ability to generate realistic images with fine details makes it ideal for layout generation.

We propose a spatial-distribution-guided diffusion model for cell layout generation. Our diffusion model learns to generate binary masks, in which each cell is represented by a square marker. To incorporate the spatial distribution into the diffusion model, we propose two major ideas. First, to handle the large variation of sparse/dense layouts, we propose to condition the diffusion model on the number of cells. Second, we incorporate spatial distributions into the model. Due to the heterogeneity of spatial distribution, it is unrealistic to simply condition the generation on summary statistics of the spatial density. Instead, we design the model to jointly generate both the layout map and the spatial density maps simultaneously. This way, the model will gradually learn the density distribution through the diffusion process.

As another contribution, we explore and analyze different density distribution models for layout generation: (1) Kernel density estimation (KDE); (2) Gaussian mixture model (GMM); and (3) Gaussian Mixture Copula Model (GMCM) [35,22]. While KDE is more flexible, GMM is more constrained and is less likely to overfit. GMCM is a compromise between the two. Our experiments provide a comprehensive analysis of the strengths and weaknesses of the three density models. Finally, we introduce metric *spatial-FID*, to evaluate the quality of generated object layouts. The metric maps the generated layouts into a spatial representation space and compares them with the spatial representation of real layouts. This provides an appropriate evaluation metric for cell layouts.

In experiments, we verify the effectiveness of our method through spatial-FID. Furthermore, we show that the generated layouts can guide diffusion models to generate high quality pathology images. These synthetic images can be utilized as an augmentation to boost the performance of supervised cell detection methods.

2 Related work

Diffusion Model. Diffusion models generate samples from random noise by learning to eliminate the noise in a multi-stage process [13,33]. Diffusion models are gaining attention because of their superior generation performance compared to GAN models [9,5,21]. There are extensive applications of diffusion models: semantic segmentation [11,4], point cloud generation [25], and video generation [12]. Saharia et al. [31] propose the super-resolution method SR3 by conditioning the diffusion model on low-resolution images. Dhariwal et al. [9] boost the quality of conditional generation for the diffusion model by using classifier guidance. Zhu et al. [39] apply the diffusion model to align the crowd distribu-

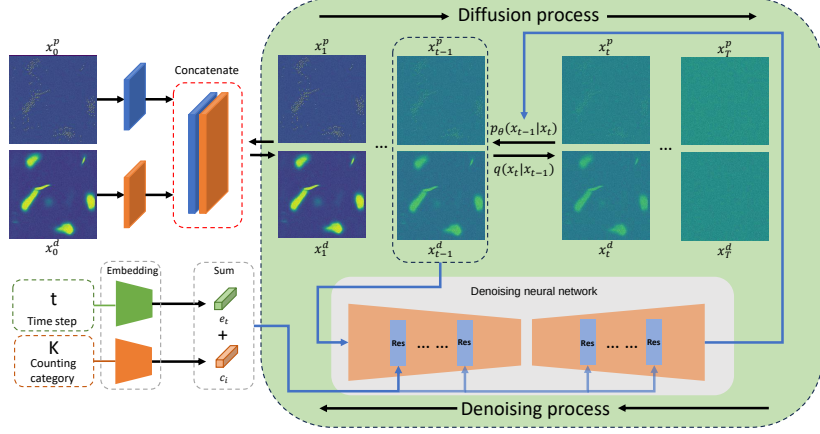


Fig. 1: Overview of proposed layout generation framework. In order to reduce the difficulty of learning spatial distributions behind object layouts, we propose a counting-category conditioned object layout generation framework. Meanwhile, the spatial density maps are incorporated into the training and generation process to help diffusion model learn and generate the spatial distribution of object layouts. For clearance, we only show one cell type’s cell layout and spatial density map. x_t^p and x_t^d are the cell layout and spatial density map at time step t .

tion for various domains. Unlike previous works, we focus on the cell layout and pathology image generation for cell detection.

Layout Generation. Layout generation is the task of synthesizing the arrangement of elements. Generative model based methods [24,20,3,19,23,7,18] are widely used in this field. LayoutGAN [24] and LayoutVAE [20] represent pioneering efforts in employing GAN and VAE to generate graphic and scene layouts. LayoutDM [7,18] explores the diffusion based model’s capability in layout generation. The works above need access to geometric parameters (location and size) of objects in training layouts. However, only the center coordinates of cells are available for cell detection layouts. TMCCG [2] attempts to generate cell layouts with GAN. Considering the superiority of the diffusion model in various generation tasks, we construct a diffusion model based cell layout generation framework.

3 Method

It is essential to generate realistic pathology images for cell detection tasks, and one challenge is to properly model the spatial distribution of these cell layouts. This paper explores multiple ways to summarize the spatial context in cell layouts. By incorporating spatial information in the training process, diffusion models can learn the underlying distribution of cells and generate realistic

layouts. We represent the cells in pathology images as a series of 3×3 square markers in layout maps. The markers are in the center of cells. Using narrow markers can prevent layout maps of highly dense cells from collapsing into a big mass. The generated layouts can guide the generation of realistic pathology images.

3.1 Layout and Image Generation

In this part, we start by explaining our approach to training spatial-aware diffusion models to generate realistic cell layouts. Next, we introduce the diffusion framework for generating synthetic pathology images under the guidance of generated layouts.

Diffusion Model. Given data $\mathbf{x}_0 \sim q(\mathbf{x}_0)$, the forward diffusion process of diffusion models corrupts original data $\mathbf{x}_0$ into $\mathbf{x}_T$ by introducing Gaussian noise with variance schedule $\beta_1, \dots, \beta_T$. This process is depicted as a Markov chain formulated as follows:

$$q(\mathbf{x}_1, \dots, \mathbf{x}_T | \mathbf{x}_{t-1}) := \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1}) \quad (1)$$

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}) \quad (2)$$

If T is large enough, the corrupted data $\mathbf{x}_T$ will nearly follow an isotropic normal distribution. The diffusion models sample data from $q(\mathbf{x}_0)$ by reverse of the diffusion process. The generative process $q(\mathbf{x}_{t-1} | \mathbf{x}_t)$ is approximated by a neural network:

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \sigma_t^2 \mathbf{I}) \quad (3)$$

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) \quad (4)$$

where $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$ is inferred by a denoising neural network. Both $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$ and σ_t are learned by optimizing a hybrid learning objective [8,27].

Due to the sparsity of cell layout maps, it is challenging for diffusion models to learn the underlying distribution of cells from the standard training process. To address this problem, we introduce spatial density maps. By modeling the spatial information in a dense way, the spatial density maps teach the model the spatial distribution behind cell layouts. The bond between layout maps and density maps assures the generated cell layouts follow the spatial distribution of the training set. To incorporate the spatial density map into the training process efficiently, we construct $\mathbf{x}_0$ by concatenating cell layout map $\mathbf{x}_0^p$ and spatial density maps $\mathbf{x}_0^d$ together: $\mathbf{x}_0 = \text{concat}(\mathbf{x}_0^p, \mathbf{x}_0^d)$. We construct $\mathbf{x}_0^p$ and $\mathbf{x}_0^d$ by combining the cell layouts and spatial density maps of different cell types. We generate spatial density maps for each cell type independently. As shown

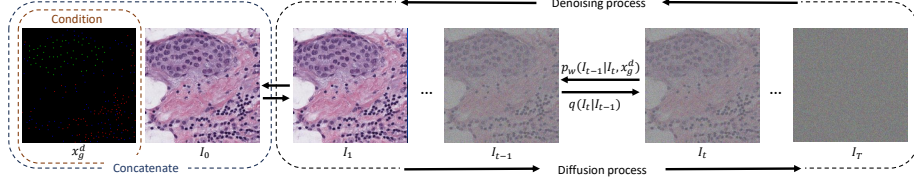


Fig. 2: Illustration of cell layout conditioned diffusion model for generating pathology images with ground truth.

in Fig. 1, the generated cell layouts are distributed in the corresponding spatial density generation pattern. More samples are available in the supplementary.

The spatial distribution of cells in pathology images is influenced by density. To make the generation model realize this correlation, we incorporate cell counting in the image patch as prior information for layout generation. However, due to the lack of samples for each quantity, using cell counting as a condition directly will lead to poor-quality layout generation. Here, we split the training image patches by the number of cells in each patch. We define C as the cell count distribution of the training set, and C_p is the p -th quantile of this distribution. We can divide the cell counting space into $K \in \mathbb{N}^*$ parts: $[C_{i/K}, C_{(i+1)/K}]$, $i \in \{0 \dots K-1\}$. During training, we treat the layouts belonging to the part i as a counting category e_i and condition the diffusion model on these counting categories. In the inference stage, we sample the same number of layouts from each counting category so that the generated layouts share the same counting distribution as the original data. An overview of layout generation and learning process is in Fig. 1.

We generate pathology images I using a diffusion model conditioned on the generated layouts. The layout maps $\mathbf{x}_g^p$ are fed into the denoising neural network as a condition to guide the cell distribution of the generated pathology image. Our layout conditional diffusion model $p(I|\mathbf{x}_g^p)$ generates pathology images with the layout map $\mathbf{x}_g^p$ through the neural network approximated denoising process $p_w(I_{t-1}|I_t, \mathbf{x}_g^p)$, where w is the parameter of neural network for pathology image generation. An illustration is shown in Fig. 2. Given the generated high-quality pathology images and their cell layouts, we can augment the existing cell detection methods for better performance.

3.2 Spatial Feature Extractors

An excellent spatial feature extractor should be able to capture the cluster patterns in cell layouts, including the position, density, and area of cell clusters, which is a challenging task. Here, we present three different ways to extract spatial features: Kernel density estimation (KDE), Gaussian Mixture Model (GMM) [29], and Gaussian Mixture Copula Model (GMCM) [35].

Kernel density estimation is a non-parametric probability estimation framework, which is widely used for point pattern analysis. Intuitively, KDE treats

training data points as density sources, and the combination of effects from training data points will create a smoothed estimate of probability distribution. For each estimating location, the closer data points have higher density contributions. The bandwidth is a crucial parameter for KDE. A lower bandwidth results in an estimation with more details but potentially introduces extra noise. On the other hand, a larger bandwidth produces a smoother result with the risk of oversmoothing. Here, to reach a good balance, we use Scott’s Rule to select bandwidth [32].

Unlike KDE, the Gaussian mixture is a parametric model, modeling the distributions of object cluster patterns as Gaussian distributions. Therefore, GMM has better interpretability with limited parameters. Here, we introduce another parametric model – the Gaussian mixture copula model (GMCM). The copula function is used to capture the dependency between marginal densities. GMCM uses GMM to estimate the marginal distribution of point cloud and Gaussian mixture copula function, derived from a mixture of Gaussians, to capture the marginal dependency. Due to the non-identifiability nature of the components in GMCM and GMM, we adopt an EM algorithm [35] to optimize the parameters of GMCM and GMM. Bayesian Information Criterion (BIC) decides the optimal number of components for both GMM and GMCM.

4 Experiment

We evaluate the effectiveness of cell layout and pathology image generation frameworks. We propose spatial-FID as a Fréchet Inception Distance (FID) modification to measure the quality of generated cell layouts. We also evaluate the effectiveness of our layout conditional pathology image generation framework on cell detection tasks.

Datasets. We evaluate our methods on cell detection dataset BRCA-M2C [1]. BRCA-M2C has 80, 10, and 30 pathology image patches for training, validating, and testing, respectively. This dataset has three cell types: tumor, lymphocyte, and stromal.

Implementation details. We separate the training set into five counting categories. To ensure the image patches for training have abundant spatial structures to learn, we crop 464×464 patches from whole training images for training diffusion models on BRCA-M2C. The diffusion model of layout generation is a U-Net architecture [9] with multiple attention heads at resolution: 32×32 , 16×16 , and 8×8 . For the image generation framework, we initialize the model with weights from pre-trained super-resolution diffusion model [9]. For cell detection tasks, we use the state-of-the-art methods MCSpatNet [1] and U-Net [30] as the cell detection frameworks.

Evaluation metric. To evaluate the quality of generated layout maps, we propose spatial-FID ($\downarrow$)³. The FID is used to measure the performance of the generative model on natural image datasets, e.g., ImageNet and CIFAR-10/100. It

³ The lower the better.

Table 1: Quantitative results evaluated by spatial-FID.

Method	w/o density	GMM	KDE	GMCM
Spatial-FID	1.078	0.039	0.357	0.953

Table 2: Results (F-scores) on the augmentation of cell detection. Columns “Infl.”, “Epi.”, “Stro.”, and “Det.” are F-scores on inflammatory, epithelial, stromal, and all cells.

Method	Infl.	Epi.	Stro.	Mean	Det
U-Net	0.498	0.744	0.476	0.572	0.838
U-Net + Rand.	0.625	0.735	0.472	0.611	-
U-Net + TMCCG	0.650	0.768	0.511	0.644	-
U-Net + No den	0.641	0.784	0.537	0.654	0.856
U-Net + KDE	0.642	0.800	0.545	0.662	0.853
U-Net + GMM	0.647	0.797	0.554	0.666	0.853
U-Net + GMCM	0.658	0.786	0.536	0.660	0.852
MCSpatNet	0.635	0.785	0.553	0.658	0.849
MCSpatNet + Rand.	0.652	0.772	0.506	0.644	-
MCSpatNet + TMCCG	0.678	0.800	0.522	0.667	-
MCSpatNet + No den	0.647	0.788	0.543	0.659	0.850
MCSpatNet + KDE	0.652	0.793	0.556	0.667	0.845
MCSpatNet + GMM	0.639	0.804	0.563	0.669	0.855
MCSpatNet + GMCM	0.615	0.806	0.555	0.659	0.850

thus relies on the visual features from a pre-trained inception model. However, there is a significant difference between natural images and layout maps. Therefore, we train an autoencoder (Fully Convolutional Networks) on the layout maps from the training set to capture the spatial information in the layouts. We use the spatial feature from the middle of the encoder ($f_s(\cdot)$) to replace the visual features. Moreover, introduce spatial-FID formed as follows:

$$s^2((\mu_T, \Sigma_T), (\mu_D, \Sigma_D)) = \|\mu_T - \mu_D\|_2^2 + Tr(\Sigma_T + \Sigma_D - 2(\Sigma_T \Sigma_D)^{(1/2)}) \quad (5)$$

where μ_T (μ_D) and Σ_T (Σ_D) are the mean and covariance of the spatial features extracted by $f_s(\cdot)$ from training layouts (generated layouts), respectively. We use F-scores to evaluate the cell detection performance.

4.1 Results

In this part, we train the model on the training set, generate 200 layouts with corresponding pathology images for each counting category, and evaluate our method from layout generation and cell detection augmentation.

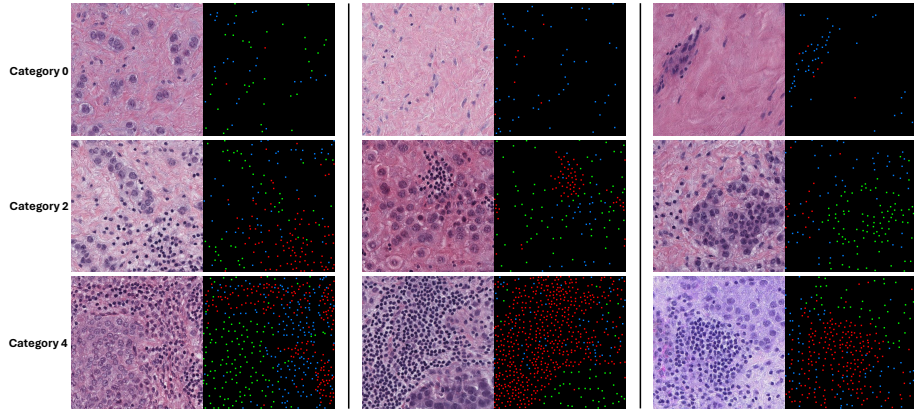


Fig. 3: Qualitative results generated by our layout and image generation framework for cell detection. Rows 1st, 2nd, and 3rd are from the counting categories 0, 2, and 4, respectively. Tumor, lymphocyte, and stromal cells are marked by green, red, and blue marks.

As shown in Tab. 1, the spatial density map generated by GMM achieved the best performance due to the distribution of cells in BRCA-M2C conforming to the mixture of Gaussian best. The cell layout distribution of BRCA-M2C data has distinct subgroups or clusters, and GMM can effectively capture these patterns. The flexibility of KDE leads to noisy density estimation results in our case, preventing our framework from getting better layout generation. GMCM models dependencies between variables using copulas, which can introduce additional complexity. The mismatches between GMCM’s assumption and data distribution lead to inferior layout generations. According to Tab. 2, the pathology image generations directed by GMM boost the cell detection performance best, reflecting that better cell layouts lead to higher-quality pathology image generations. As shown in Fig. 3, our pathology image generations are deceptively realistic. This is mainly attributed to high-quality cell layout generations and the excellent performance of diffusion models in pathology image generation.

4.2 Ablation study

We conduct ablation studies to show each component’s effectiveness and the effects of hyper-parameters on the generation framework. The supplementary includes more ablation studies.

Counting categories. We conduct ablation studies by discarding this condition or setting different numbers of counting categories (10 or 20) to show the effectiveness of counting categories in cell layout generation. According to Tab. 3, counting categories are important for generating realistic cell layouts, and our method is robust to the change of counting categories.

Table 3: Ablation study results on the number of counting categories

Classes number	0	5	10	20
Spatial-FID	0.122	0.039	0.231	0.558

5 Conclusion

In this paper, we propose a spatial-distribution-guided diffusion framework for generating high-quality cell layouts and pathology images. To represent the spatial distribution of cell layouts properly, we explore three alternative tools for spatial feature extraction: KDE, GMM, and GMCM. They all can significantly boost the generative quality of cell layouts. Due to the underlying cell layouts of BRCA-M2C complying with the mixture of Gaussian best, GMM achieves the best performance. We treat generated cell layouts as conditions for pathology image generation. These high-quality generated pathology images can improve the performance of SOTA cell detection methods.

Acknowledgements: This research was partially supported by the National Science Foundation (NSF) grant CCF-2144901, the National Institute of General Medical Sciences (NIGMS) grant R01GM148970, and the Stony Brook Trustees Faculty Award.

Disclosure of Interests: The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abousamra, S., Belinsky, D., Van Arnam, J., Allard, F., Yee, E., Gupta, R., Kurc, T., Samaras, D., Saltz, J., Chen, C.: Multi-class cell detection using spatial context representation. In: ICCV (2021)
2. Abousamra, S., Gupta, R., Kurc, T., Samaras, D., Saltz, J., Chen, C.: Topology-guided multi-class cell context generation for digital pathology. In: CVPR (2023)
3. Arroyo, D.M., Postels, J., Tombari, F.: Variational transformer networks for layout generation. In: CVPR (2021)
4. Baranchuk, D., Voynov, A., Rubachev, I., Khrulkov, V., Babenko, A.: Label-efficient semantic segmentation with diffusion models. In: ICLR (2021)
5. Brock, A., Donahue, J., Simonyan, K.: Large scale gan training for high fidelity natural image synthesis. In: ICLR (2018)
6. Casanova, A., Careil, M., Verbeek, J., Drozdal, M., Romero Soriano, A.: Instance-conditioned gan. In: NeurIPS (2021)
7. Chai, S., Zhuang, L., Yan, F.: Layoutdm: Transformer-based diffusion model for layout generation. In: CVPR (2023)
8. Choi, J., Lee, J., Shin, C., Kim, S., Kim, H., Yoon, S.: Perception prioritized training of diffusion models. 2022 ieee. In: CVPR (2022)
9. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. In: NeurIPS (2021)

10. Gong, X., Chen, S., Zhang, B., Doermann, D.: Style consistent image generation for nuclei instance segmentation. In: WACV (2021)
11. Graikos, A., Malkin, N., Jojic, N., Samaras, D.: Diffusion models as plug-and-play priors. In: NeurIPS (2022)
12. Harvey, W., Naderiparizi, S., Masrani, V., Weilbach, C., Wood, F.: Flexible diffusion modeling of long videos. In: NeurIPS (2022)
13. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: NeurIPS (2020)
14. Höfener, H., Homeyer, A., Weiss, N., Molin, J., Lundström, C.F., Hahn, H.K.: Deep learning nuclei detection: A simple approach can deliver state-of-the-art results. *Computerized Medical Imaging and Graphics* (2018)
15. Hou, L., Agarwal, A., Samaras, D., Kurc, T.M., Gupta, R.R., Saltz, J.H.: Robust histopathology image analysis: To label or to synthesize? In: CVPR (2019)
16. Huang, J., Li, H., Wan, X., Li, G.: Affine-consistent transformer for multi-class cell nuclei detection. In: ICCV (2023)
17. Hung, J., Goodman, A., Ravel, D., Lopes, S.C., Rangel, G.W., Nery, O.A., Malleret, B., Nosten, F., Lacerda, M.V., Ferreira, M.U., et al.: Keras r-cnn: library for cell detection in biological images using deep neural networks. *BMC bioinformatics* (2020)
18. Inoue, N., Kikuchi, K., Simo-Serra, E., Otani, M., Yamaguchi, K.: Layoutdm: Discrete diffusion model for controllable layout generation. In: CVPR (2023)
19. Jiang, Z., Sun, S., Zhu, J., Lou, J.G., Zhang, D.: Coarse-to-fine generative modeling for graphic layouts. In: AAAI (2022)
20. Jyothi, A.A., Durand, T., He, J., Sigal, L., Mori, G.: Layoutvae: Stochastic scene layout generation from a label set. In: ICCV (2019)
21. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: CVPR (2019)
22. Kasa, S.R., Rajan, V.: Improved inference of gaussian mixture copula model for clustering and reproducibility analysis using automatic differentiation. *Econometrics and Statistics* (2022)
23. Kikuchi, K., Simo-Serra, E., Otani, M., Yamaguchi, K.: Constrained graphic layout generation via latent optimization. In: ACM MM (2021)
24. Li, J., Yang, J., Hertzmann, A., Zhang, J., Xu, T.: Layoutgan: Synthesizing graphic layouts with vector-wireframe adversarial networks. *PAMI* (2020)
25. Luo, S., Hu, W.: Diffusion probabilistic models for 3d point cloud generation. In: CVPR (2021)
26. Miyato, T., Koyama, M.: cgans with projection discriminator. In: ICLR (2018)
27. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: ICML (2021)
28. Odena, A., Olah, C., Shlens, J.: Conditional image synthesis with auxiliary classifier gans. In: ICML (2017)
29. Reynolds, D.A., et al.: Gaussian mixture models. *Encyclopedia of biometrics* (2009)
30. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
31. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. *PAMI* (2022)
32. Scott, D.W.: *Multivariate Density Estimation. Theory, Practice, and Visualization*. Wiley (1992)
33. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: ICML (2015)

34. Sugimoto, T., Ito, H., Teramoto, Y., Yoshizawa, A., Bise, R.: Multi-class cell detection using modified self-attention. In: CVPR (2022)
35. Tewari, A.: On the estimation of gaussian mixture copula models. In: ICML (2023)
36. Tsai, T.Y.C., Sikora, M., Xia, P., Colak-Champollion, T., Knaut, H., Heisenberg, C.P., Megason, S.G.: An adhesion code ensures robust pattern formation during tissue morphogenesis. *Science* (2020)
37. Yousefi, S., Nie, Y.: Transfer learning from nucleus detection to classification in histopathology images. In: ISBI (2019)
38. Zhang, H., Goodfellow, I., Metaxas, D., Odena, A.: Self-attention generative adversarial networks. In: ICML (2019)
39. Zhu, H., Yuan, J., Yang, Z., Zhong, X., Wang, Z.: Fine-grained fragment diffusion for cross domain crowd counting. In: ACMMM (2022)