TopoCellGen: Generating Histopathology Cell Topology with a Diffusion Model

Meilong Xu[1]   Saumya Gupta[1]   Xiaoling Hu[2]   Chen Li[1]   Shahira Abousamra[3]
Dimitris Samaras[1]   Prateek Prasanna[1]   Chao Chen[1]

[1]Stony Brook University, NY, USA
[2]Athinoula A. Martinos Center for Biomedical Imaging,
Massachusetts General Hospital and Harvard Medical School, MA, USA
[3]Department of Biomedical Data Science, Stanford University, CA, USA

meixu@cs.stonybrook.edu, chao.chen.1@stonybrook.edu

Abstract

Accurately modeling multi-class cell topology is crucial in digital pathology, as it provides critical insights into tissue structure and pathology. The synthetic generation of cell topology enables realistic simulations of complex tissue environments, enhances downstream tasks by augmenting training data, aligns more closely with pathologists' domain knowledge, and offers new opportunities for controlling and generalizing the tumor microenvironment. In this paper, we propose a novel approach that integrates topological constraints into a diffusion model to improve the generation of realistic, contextually accurate cell topologies. Our method refines the simulation of cell distributions and interactions, increasing the precision and interpretability of results in downstream tasks such as cell detection and classification. To assess the topological fidelity of generated layouts, we introduce a new metric, Topological Fréchet Distance (TopoFD), which overcomes the limitations of traditional metrics like FID in evaluating topological structure. Experimental results demonstrate the effectiveness of our approach in generating multi-class cell layouts that capture intricate topological relationships. Code is available at https://github.com/Melon-Xu/TopoCellGen.

1. Introduction

Deep-learning methods have made substantial advances in foundational tasks for nuclei analysis, including instance segmentation [21, 29, 32, 40, 73], classification, and detection [1, 52]. These tasks provide the basis for downstream analyses, enabling detailed characterization of tissue architecture and cellular interactions, which are crucial for diagnostic and prognostic applications in pathology [12, 46]. However, accurately annotating multi-class cell arrangements in pathology images is challenging, as distinct cell types display unique spatial patterns, requiring significant domain expertise. Although annotated datasets for multiple cell types exist, they often lack the diversity needed for generalization across various tissues and organs.

To alleviate the burden of manual annotation and enhance the efficiency of analysis, there has been growing interest in utilizing generative models. Early works use Generative Adversarial Networks (GANs) [20] for automatic generation of pathology images [2, 10, 34]. In recent years, diffusion models [11, 31, 45, 50, 61, 69, 75, 76] have emerged as much more reliable alternatives, generating accurate, high-resolution histopathology images [4, 23, 48, 51, 72]. However, all these diffusion models are only trained to directly generate the histopathology images. Despite the impressive visual results, these powerful models deliver limited insight into the underlying biology. It is very hard to connect the learned distributions with human knowledge about tumor microenvionment. This makes it challenging to validate, generalize, or control these models.

We argue that a key issue is the lack of explicit generation of cells and their spatial arrangement. The spatial organization of cells and the interactions across different cell types are critical for understanding tumor microenvironments, disease progression, and tissue regeneration [59]. The density and spatial distribution of various cell types -such as lymphocytes, epithelial cells, and stromal cells- are essential for pathologists in making accurate diagnoses and prognoses. For instance, the detection and quantification of tumor-infiltrating lymphocytes (TILs), which are lymphocytes located within the boundaries of invasive tumors [56], have been strongly linked to improved clinical outcomes [57, 63]. The presence of isolated or small clusters of tumor cells at the invasive front, known as tumor budding, serves as a prognostic biomarker associated with a higher risk of lymph node metastasis in colorectal carcinoma and other solid malignancies [47].
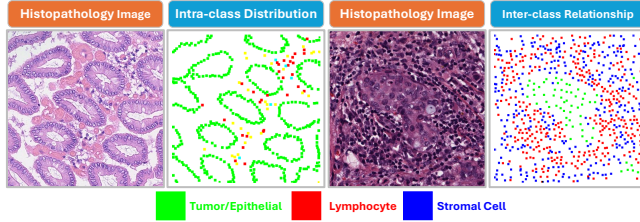
Figure 1. Illustrations of intra-class distribution and the inter-class relationship across various cell types. Here we only highlight the tumor/epithelial, lymphocytes, and stromal cells.

In this paper, we explore the problem of generating cell spatial layouts, with multiple benefits. First and foremost, directly modeling cell layouts brings us closer to aligning with pathologists' domain knowledge. This enables the direct verification of synthetic data through quantification and comparison with expert knowledge, facilitating greater accuracy and trust in the generated data. Additionally, it opens up the possibility of controlling the layout generation process, allowing the model to generalize to previously unseen scenarios. From a data augmentation perspective, generating cell layouts also enables the creation of histology images conditioned on these layouts. This capability allows for the production of synthetic images with cell annotations, which can significantly aid in the training of models for various downstream tasks, particularly cell detection and classification.

Cell topology is the key. To generate accurate cell layouts, we argue that the key is to model the topological relationships between cells, defined by their spatial organization, clustering, mixing, and connectivity. These topological patterns provide valuable insights into cellular communication, structural changes, and morphological abnormalities linked to pathological conditions such as cancer, inflammation, and fibrosis. Fig. 1 illustrates the significance of these topological relationships in understanding tissue organization and disease progression. These relationships reveal how cells are spatially arranged to form structural motifs that are crucial for tissue function and stability. For example, on the left, we show that in healthy glandular tissues, epithelial cells typically organize into circular or tubular structures, forming acinar units that are essential for secretion and nutrient transport [59]. In tissues with diverse cell types, topological relationships capture spatial configurations and connectivity patterns that reveal cellular interactions critical for tissue health or disease [74]. For example, immune cells clustering around tumor cells (Fig. 1 (right)) or specific fibroblast-epithelial arrangements highlight processes like immune surveillance, inflammation, and stromal support. By identifying these patterns, topological methods offer insights into multi-class cellular interactions, revealing potential biomarkers and enhancing our understanding of disease progression.

We propose the first diffusion model that generates cell topology for digital pathology. Our method guides the generation process with both 0- and 1-dimensional topological features, i.e., cell clusters and holes/gaps enclosed by cells. We implement this using the theory of persistent homology, which models cell topology in a multi-scale manner. As illustrated in Fig. 1, learning the topology ensures that the generated cell layouts not only preserve intra-class spatial/structural property but also accurately capture inter-class interactions.

The second contribution of this paper is a novel cell counting loss that ensures the diffusion model learns the correct distribution of cell numbers from the data. This addresses a key challenge in previous diffusion models for histopathology images [4, 23, 44], which is often biased toward unrealistically low or high cell counts.

By ensuring both accurate cell topology and cell count, our method generates layouts that more closely resemble real tissue microenvironments. This precision in modeling cell distributions and interactions enhances the utility of generated images for augmenting training in downstream tasks such as cell detection and classification, as demonstrated in our experiments. From a modeling perspective, our approach moves closer to generating biologically faithful representations of cellular environments, integrating human domain knowledge, and improving generalization to diverse scenarios.

Finally, we introduce the Topological Fréchet Distance (TopoFD), a novel metric designed to assess the spatial and topological accuracy of the generated layouts. TopoFD measures the similarity between the topological features of real and synthetic cell layouts, whereas traditional metrics such as FID focus solely on visual similarity.

In summary, our contribution is three-fold:

- We present the first diffusion model designed to generate cell topology for digital pathology, simulating realistic intra- and inter-class spatial distributions in the generated layouts.
- We introduce a novel cell counting loss that aligns the generated cell numbers with real data, addressing biases in prior diffusion models and ensuring realistic cell density in synthetic layouts.
- We introduce the Topological Fréchet Distance (TopoFD), a new metric designed to evaluate the topological similarity of generated cell layouts.

Extensive experiments demonstrate the effectiveness of our proposed method, showing that it not only enhances sample quality but also significantly improves performance in downstream tasks such as cell detection and classification.

2. Related Works

Diffusion Models for Digital Pathology. Diffusion models, such as Denoising Diffusion Probabilistic Models

(DDPMs) [11, 15, 31, 33, 50, 61, 62, 77] and Latent Diffusion Models (LDMs) [6, 53, 55], have significantly advanced image synthesis by modeling data distributions through iterative denoising processes on data sample or latent spaces. These models have been adapted to generate high-fidelity histopathology images in digital pathology [4, 23, 28, 72]. However, these methods often overlook the significance of multi-class cell layouts, which are crucial for accurately representing tissue structures.

Recent approaches to synthetic cell layout generation in digital pathology have explored nuclei labeling and cell arrangement. Abousamra et al. utilized spatial statistics and topological descriptors in GANs to model complex cell configurations [2]. DiffMix employs conditioned diffusion models for augmenting imbalanced nuclear pathology datasets, generating more realistic images than earlier methods [51]. Similarly, a inter diffusion framework proposed in [48] generates paired histopathology images and nuclei labels simultaneously, enhancing the context-awareness of synthetic data. Another strategy, a diffusion model focused on cell layout generation, uses density maps to incorporate spatial distributions, though it simplifies cell counts into five categorical conditions, which provides only limited guidance [44]. While these models have contributed to more realistic cell layout synthesis, they generally lack fine control over cell density and fail to explicitly preserve intra-class spatial distributions and inter-class spatial relationships across cell types. This limits their effectiveness in replicating the complex and varied structures of real tissue.

Topology-Driven Methods in Deep Learning. Incorporating algebraic topology [49] in deep learning frameworks is becoming increasingly significant. A key development in this front has been persistent homology [13, 14], a mathematical theory that analyzes how structural features (such as connected components, loops, and voids) persist across different scales in data. This approach has proven particularly valuable for understanding complex data structures while being resistant to noise. Its applications range from image segmentation [8, 36, 64, 71], feature extraction [19] and disease diagnosis [42, 43, 67, 70]. The field has further expanded through additional mathematical frameworks, including discrete Morse theory [17, 37, 38], homotopy warping [35], structural relationship studies [25, 27], and methods for analyzing shape features [60, 68].

For generative models, topological information has become increasingly essential, particularly in biomedical imaging, where capturing both visual realism and topological integrity is critical. One notable method is TopoGAN [66], which integrates topology into GANs by introducing a novel topological loss based on persistent homology. Recently, TopoDiffusionNet [26] has integrated topology with diffusion models, introducing a topology-based objective function that guides the model's denoising

process to generate images with exact object counts. However, these methods, designed for natural images, are unsuitable for our cell topology generation task. Their emphasis on overall structure numbers cannot capture the cell structural pattern and cell spatial interactions among multiple cell types as our method does.

3. Method

In this section, we introduce our methodology for generating realistic multi-class cell layouts with precise topological control. To accurately control the cell density in the generated layout, we employ a conditioning mechanism based on the cell count for each type. In the multi-class setting, assuming we have n classes of cells, we define the condition vector $c = [c_1, c_2, ..., c_n]$, where each element c_i represents the count of cells in the respective channel.

However, the condition vector c alone is insufficient to precisely control cell counts [26]. Additionally, it cannot preserve the intra-class spatial distributions as well as inter-class topological relationships in the generated layouts. To address this, we formulate a cell counting loss to strengthen the control on accurate cell counts and introduce two topology-aware objective functions based on persistent homology [13, 49] to guide the generation process, ensuring realistic and contextually accurate cell layouts that respect both spatial and topological constraints.

The remainder of this section is structured as follows: In Sec. 3.1, we present a brief introduction to diffusion models. Next, in Sec. 3.2, we provide a quick background on persistent homology. Following this, we introduce *TopoCellGen*, our topology-preserving cell layout generation method, in Sec. 3.3. We introduce our proposed Topological Fréchet Distance (TopoFD) in Sec. 3.4. Finally, in Sec. 3.5, we present the overall inference pipeline and how the generated cell layouts are transformed into H&E images for training augmentation for downstream tasks.

Preliminaries. In a multi-class cell layout containing n channels, each channel corresponds to a specific cell type (e.g., lymphocyte), with each cell represented as a square where the pixel value is set to 1, while the pixel value of the background is set to 0.

3.1. Diffusion Models

Our generative approach utilizes a denoising diffusion probabilistic model (DDPM) [31], which learns to reverse a forward process that incrementally adds Gaussian noise to transform a structured cell layout into a noise distribution. The reverse process then reconstructs the layout from noise via iterative denoising.

Let x_0 represent the target cell layout and x_T denote pure Gaussian noise. At each time step t, noise is added to the data based on a variance schedule β_t: $q(x_t|x_{t-1}) =$
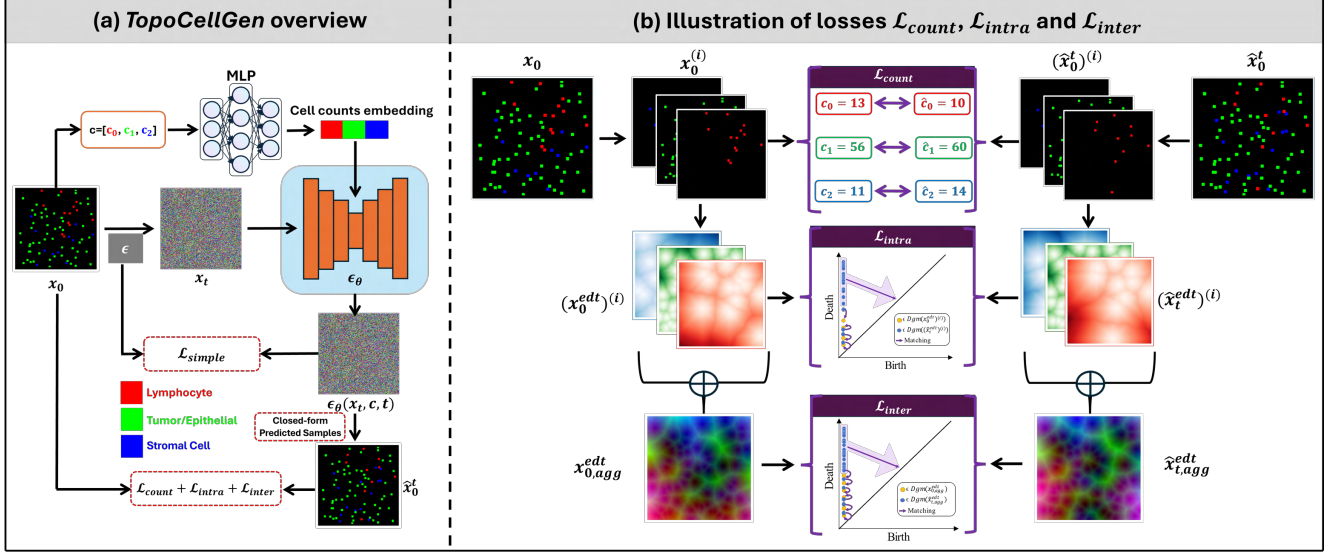
Figure 2. An overview of our method *TopoCellGen*. (**a**) denotes the overview workflow. (**b**) shows the details of $\mathcal{L}_{count}$, $\mathcal{L}_{intra}$ and $\mathcal{L}_{inter}$.

$\mathcal{N}(x_t; \sqrt{1 - \beta_t}x_{t-1}, \beta_t I)$. This forward process results in the progressively noisier version of the data, with x_T approximating an isotropic Gaussian distribution as t increases.

The reverse process, parameterized by a neural network $\epsilon_\theta(x_t, t)$ (typically UNet [54]), learns to iteratively denoise x_T back to x_0. Conditioning on the cell count vector c, the model is trained by minimizing a simplified variant of the variational lower bound, specifically focusing on predicting the noise added at each step:

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{t,x_0,\epsilon} \left[\|\epsilon - \epsilon_\theta(x_t, c, t)\|^2 \right] \tag{1}$$

where $\epsilon \sim \mathcal{N}(0, I)$ is the noise sampled during training. This objective enables the model to learn the reverse process effectively. Instead of the standard iterative denoising, we also approximate the noiseless layout $\widehat{x}_0^t$ deterministically for any noisy state x_t by marginalizing over the noise schedule:

$$\widehat{x}_0^t \approx \frac{1}{\sqrt{\bar{\alpha}_t}} \left(x_t - \sqrt{1 - \bar{\alpha}_t}\epsilon_\theta(x_t, c, t) \right) \tag{2}$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, which aggregates the effect of the variance schedule up to time t. This predicted noiseless layout, $\widehat{x}_0^t$, will be used to impose constraints in subsequent stages.

3.2. Background: Persistent Homology

In algebraic topology [49], homology classes provide a structured way to capture topological features across multiple dimensions. For instance, 0-, 1-, and 2-dimensional features represent connected components, loops, and voids, respectively. The d-dimensional Betti number, β_d, quantifies the number of d-dimensional features present, offering

insights into the underlying topological complexity. However, extending these concepts to real-world data, which is often continuous and noisy, introduces challenges for accurately capturing topological structures.

Persistent homology, developed in the early 2000s [13, 14], addresses this need by tracking the evolution of topological features across multiple scales, making it particularly effective for discrete datasets like cell point clouds. The process begins by constructing a filtration—a sequence of nested simplicial complexes built from the point cloud data by incrementally connecting points based on a scale parameter. As the parameter varies, topological features such as connected components and loops appear and eventually vanish, each represented by a point in a persistence diagram (Dgm). In a diagram, each point (b, d) marks the birth and death of a feature, capturing its persistence across scales and providing a compact, multi-scale summary of the underlying topology. More details are in the Supplementary.

3.3. Spatially Aligned Cell Layout Generation

The primary objective of our method is to generate multi-class cell layouts that accurately simulate both the topological and spatial properties of real-world biological cell distributions. To achieve this, we ensure accurate cell counts for each cell type through a cell counting loss, while also preserving spatial relationships within individual cell types via enforcing intra-class spatial consistency. Furthermore, we maintain structural coherence across all cell types by applying an inter-class structural regularization, leveraging 1-dimensional persistent homology to encapsulate both type-specific and collective spatial properties. The overall pipeline is shown in Fig. 2.

Cell Counting Loss. Given the target layout x_0, which

serves as the ground truth, for each time step t, we obtain the predicted noiseless layout $\widehat{x}_0^t$ using Eq. (2). To ensure precise control over the number of cells in the generated layout, we introduce a differentiable cell counting loss. The key challenge lies in making the counting operation differentiable for gradient-based optimization. We address this by employing the Straight-Through Estimator (STE) [5], which enables gradient flow through the discrete binarization operation. Specifically, after obtaining $\widehat{x}_0^t$, we apply a hard threshold to obtain binary values:

$$b(\widehat{x}_0^t) = float((\widehat{x}_0^t \geq \tau)) \tag{3}$$

where τ is the threshold parameter. Here we set it to the median value of $\widehat{x}_0^t$. During back-propagation, the STE treats the thresholding operation as an identity function, allowing gradients to flow through. The cell counting loss is then formulated as:

$$\mathcal{L}_{\text{count}} = \frac{1}{|n|} \sum_{i=1}^n \left| \frac{\sum b(\widehat{x}_0^t)^{(i)}}{\delta} - \frac{\sum x_0^{(i)}}{\delta} \right| \tag{4}$$

where $b(\widehat{x}_0^t)^{(i)}$ represents the binarized prediction for the i-th channel, and δ indicates the area (3×3) of a single cell in the layouts. This formulation provides a differentiable approximation to the discrete cell counting operation, enabling end-to-end training while maintaining precise control over the number of cells for each cell type.

Intra-Class Spatial Consistency. To enforce spatial consistency within each cell type, we first calculate the distance transform map [16] for each channel in both the target layout x_0 and the predicted noiseless layout $\widehat{x}_0^t$. The distance transform $D(x)$ is a function that assigns to each pixel the minimum Euclidean distance to the nearest cell (or non-zero pixel) in the channel. This can be formally written as:

$$D(x) = \min_{p \in \text{cells}} \|x - p\| \tag{5}$$

where p represents the positions of cells in the layout. After obtaining the distance transform maps of the target layout and the predicted noiseless layout, $\widehat{x}_t^{edt} = D(b(\widehat{x}_0^t))$ and $x_t^{edt} = D(x_0)$, we calculate the 1-dim persistence diagrams for both of them, $Dgm(\widehat{x}_t^{edt})$ and $Dgm(x_0^{edt})$ respectively. Similar to previous topological losses [36], we will use the classic Wasserstein distance between the two diagrams. Given two diagrams $Dgm(q)$ and $Dgm(s)$, the p-th Wasserstein distance is defined as follows:

$$W_p(Dgm(q), Dgm(s)) = \left(\inf_{\gamma \in \Gamma} \sum_{x \in Dgm(q)} \|x - \gamma(x)\|^p \right)^{\frac{1}{p}}$$

where Γ represents all bijections from $Dgm(q)$ to $Dgm(s)$.

The Wasserstein distance operates by identifying an optimal correspondence between points in two diagrams, assigning unmatched points to their projections on the diagonal. This distance metric is calculated by summing the distances between all paired points. The process of finding this optimal matching, as well as calculating the Wasserstein distance, can be accomplished using either the traditional Hungarian algorithm or more sophisticated methods [39, 41].

Next, we denote γ^*, the optimal matching between $Dgm(\widehat{x}_t^{edt})$ and $Dgm(x_0^{edt})$. Each persistence dot in $Dgm(\widehat{x}_t^{edt})$ is matched either to a target dot in $Dgm(x_0^{edt})$ or its projection on the diagonal. We can now formulate the spatial distribution consistency loss as the squared distance between every dot in $Dgm(\widehat{x}_t^{edt})$ and its match:

$$\mathcal{L}_{\text{spc}} = \sum_{q \in Dgm(\widehat{x}_t^{edt})} \|q - \gamma^*(q)\|^2 \tag{6}$$

For a multi-class cell layout containing n classes of cells, we formulate the intra-class spatial consistency loss as follows by averaging the Eq. (6) across multiple classes:

$$\mathcal{L}_{\text{intra}} = \frac{1}{|n|} \sum_{i=1}^n \mathcal{L}_{\text{spc}} \left(Dgm((\widehat{x}_t^{edt})^{(i)}), Dgm((x_0^{edt})^{(i)}) \right) \tag{7}$$

Inter-Class Structural Regularization. Beyond maintaining spatial distribution consistency within individual cell types, it is equally important to capture the relationships between different cell types. To achieve this, we construct a unified layout by combining all cell types into a single-channel representation, referred to as the aggregated layout: $x_0^{agg} = Agg(x_0)$ and $\widehat{x}_0^{t,agg} = Agg(\widehat{x}_0^t)$. We then compute the distance transform for the aggregated layouts, with $\widehat{x}_{t,agg}^{edt} = D(\widehat{x}_0^{t,agg})$ representing the distance transform of the predicted layout and $x_{0,agg}^{edt} = D(x_0^{agg})$ for the target layout. The inter-class structural loss $\mathcal{L}_{\text{inter}}$ is computed similarly to the intra-class loss:

$$\mathcal{L}_{\text{inter}} = \mathcal{L}_{\text{spc}} \left(Dgm(\widehat{x}_{t,agg}^{edt}), Dgm(x_{0,agg}^{edt}) \right) \tag{8}$$

Together, these class-specific and cross-class regularizations ensure that both individual cell distributions and their cumulative spatial interactions are enforced, preserving critical spatial dynamics within and between cell types in the generated layouts.

Final Objectives. The final training objective function of the model is the weighted sum of the three losses with $\mathcal{L}_{\text{simple}}$:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{simple}} + \lambda_c \mathcal{L}_{\text{count}} + \lambda_{\text{intra}} \mathcal{L}_{\text{intra}} + \lambda_{\text{inter}} \mathcal{L}_{\text{inter}} \tag{9}$$

where λ_c, λ_{intra} and λ_{inter} are hyper-parameters that control the relative contributions of the respective loss terms.

3.4. Topological Fréchet Distance (TopoFD)

Fréchet Inception Distance (FID) [30] is a key metric for evaluating the quality of generated image data by comparing the distributions of real and generated samples. FID operates by computing the mean and covariance of feature representations extracted from a pre-trained model, typically
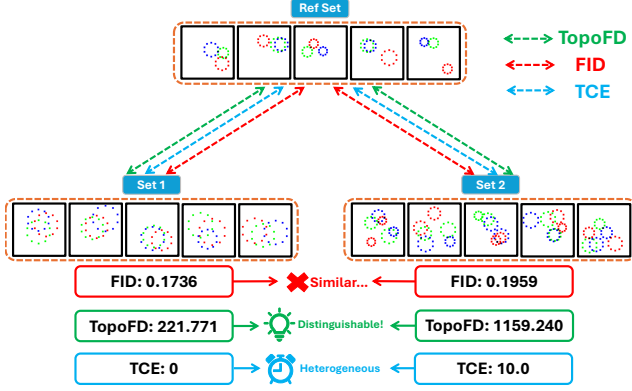
Figure 3. Intuition of our proposed Topological Fréchet Distance. TCE indicates the Total Count Error.

InceptionV3 [65]. The distance between these distributions is then measured using the Fréchet distance [18]. The formula for FID is as follows:

$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{\frac{1}{2}}) \quad (10)$$

where μ_r, Σ_r and μ_g, Σ_g are the mean and covariance of the real and generated data, respectively.

Traditional FID measures feature similarities but fail to capture complex spatial and topological cell interactions. For example, in Fig. 3, synthetic layouts of three cell types (lymphocyte, tumor/epithelial and stromal cells) demonstrate this limitation. Both Set 1 and Set 2 yield similar FID scores with respect to the Ref Set, yet differ significantly from each other in spatial configuration. Set 1 mirrors the reference with cohesive single cycles per cell type, while Set 2 presents multiple distinct cycles, deviating from the reference. The Total Count Error (TCE) reveals identical counts between Set 1 and the reference (TCE = 0) but substantial discrepancies for Set 2 (TCE = 10). Despite matching counts, Set 1's spatial arrangement still diverges from the reference. This suggests that a count metric alone does not ensure spatial and topological fidelity. We propose Topological Fréchet Distance (TopoFD) to capture higher-dimensional topological features, enabling greater sensitivity to spatial configurations essential in synthetic pathology data generation.

The pipeline of our Topological Fréchet Distance (TopoFD) is shown in Fig. 4, taking one type of cell as an example. For each layout in the reference and synthetic sets, we first obtain a point cloud, with the points corresponding to the cell center coordinates. Then, we calculate 1-dimensional persistence diagrams of each point cloud in each set:

$$Dgm_{ref} = \{Dgm_r^1, Dgm_r^2, ..., Dgm_r^n\}$$
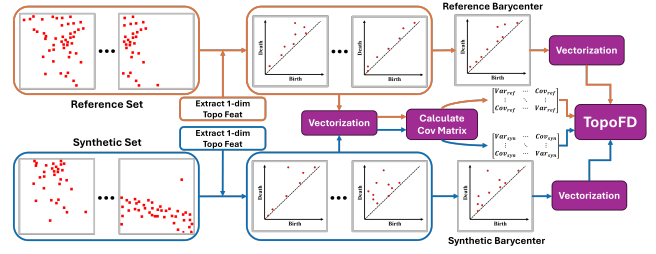$$Dgm_{syn} = \{Dgm_s^1, Dgm_s^2, ..., Dgm_s^n\}$$

Figure 4. The overall pipeline of calculating the Topological Fréchet Distance. Take the lymphocyte as an example.

where Dgm_r^i and Dgm_s^j are persistence diagrams in reference and synthetic sets, n is the number of the samples. We then compute each set's barycenter [7, 9], which minimizes the sum of Wasserstein distances [9] between individual diagrams and the barycenter. For the reference set, the barycenter $\overline{Dgm}_{real}$ is given by:

$$\overline{Dgm}_r = \arg \min_{\overline{Dgm}_r} \sum_{i=1}^{n} W_2^2(Dgm_r^i, \overline{Dgm}_r) \quad (11)$$

where W_2 is the 2-Wasserstein distance. A similar process is applied for $\overline{Dgm}_s$ from the synthetic set.

To enable statistical analysis, we transform the persistence diagrams and their barycenters into persistence landscapes, transforming each diagram into a sequence of continuous, piecewise linear functions that preserve topological information while enabling standard statistical operations [7]. The vectorized barycenters are regarded as mean vectors $\boldsymbol{\mu}$ and we compute the covariance matrices $\boldsymbol{\Sigma}$ from both sets of persistence diagrams.

For multi-class layouts, TopoFD is calculated in two steps. First, for each cell type, we compute the vectorized persistence diagrams for both reference and synthetic layouts containing that type. Then, the final TopoFD is the average Fréchet Distance across cell types:

$$\text{TopoFD} = \frac{1}{|n|} \sum_{i=1}^{n} \text{FD}(\boldsymbol{\mu}_i^r, \boldsymbol{\Sigma}_i^r, \boldsymbol{\mu}_i^s, \boldsymbol{\Sigma}_i^s) \quad (12)$$

where $\text{FD}(\cdot)$ denotes the Fréchet distance computation.

3.5. Layout-Guided Pathology Image Generation

In the final stage of our methodology, we focus on generating high-resolution pathology images based on the previously created cell layouts. These layouts serve as spatial and structural guides, ensuring that the pathology images adhere to the organization of cells within the layout. We employ a guided diffusion model [11] to synthesize realistic pathology images, conditioned on the underlying cell arrangement. This approach ensures that cells' spatial distribution and organization in the generated images align closely with the provided cell layouts. More details are provided in the Supplementary.
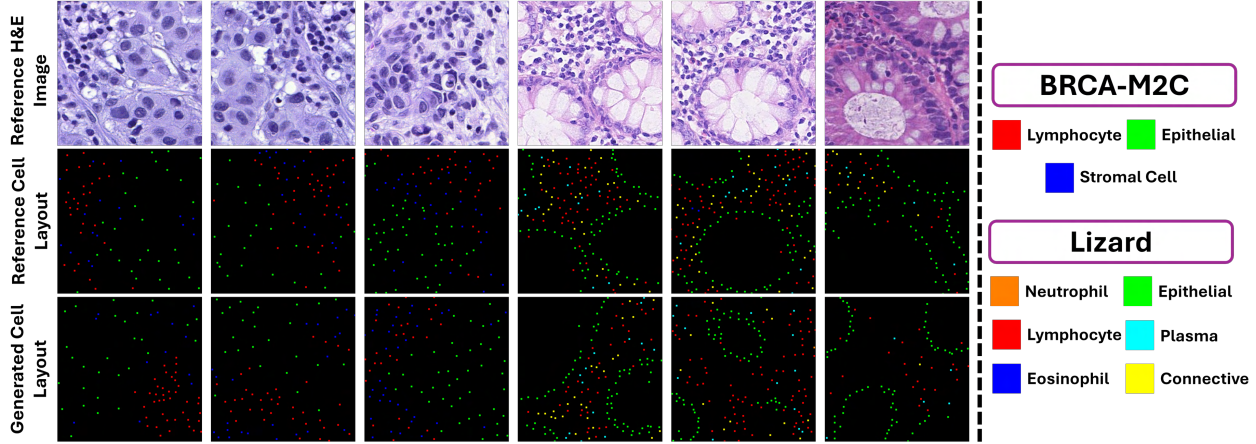
Figure 5. The qualitative results of our proposed *TopoCellGen*. Columns 1-3: BRCA-M2C dataset. Columns 4-6: Lizard dataset. The cell types and their corresponding colors are shown on the right side of the figure.

4. Experiments

We conduct extensive experiments on two public and widely used nuclei analysis datasets. We compare our method against SoTA layout generation methods in terms of sample quality and performance on downstream tasks.

Datasets. We evaluate our proposed method on *TCGA Breast Cancer Cell Classification Dataset (BRCA-M2C)* [1] and *Lizard dataset* [22]. More details about the datasets, the test set split, and the cell count distribution analysis are provided in the Supplementary.

Evaluation Metrics. We evaluate our proposed method on both sample quality and the performance of downstream tasks. We use **Fréchet Inception Distance (FID)** [30], the **cell count error** for each cell type, the **total count error (TCE)**, our proposed **TopoFD** and **maximum mean discrepancy (MMD)** [66] to evaluate how well the generated cell layouts align with the reference layouts. Note that for FID, feature extraction is tailored to each dataset with custom-trained models. On the other hand, we generate 2,000 image-layout pairs as augmented training data for cell detection and classification tasks, evaluating their performance with the **F1-score**. More details are provided in the Supplementary.

Implementation Details are in the Supplementary.

Biological Plausibility. To further validate the quality of our generated cell layouts, we conducted an evaluation involving a board-certified pathologist with 7+ years of experience. From a pathologist's perspective, the synthetic cell layouts are nearly indistinguishable from real ones and accurately capture the defining benign or malignant features of their real counterparts. More details are in the Supplementary.

Backpropagation from Distance Transform Map to Binary Mask. To backpropagate the gradient from the dis-

tance transform map to a binary mask, we leverage an efficient implementation of the geodesic distance transform provided by the FastGeodis library [3]. This method enables the computation of a differentiable distance map, which allows us to propagate gradients through the distance transform during training. The library implements a parallelized raster scan method that efficiently computes the Euclidean distance transforms on GPU hardware.

4.1. Experimental Results

Qualitative Results. Qualitative results are shown in Fig. 5. Our proposed *TopoCellGen* method demonstrates cell layouts that closely match the reference layouts in both density and spatial arrangement. The generated layouts exhibit a high degree of distribution consistency, preserving density patterns and spatial relationships among cell types while reflecting the structural characteristics of the reference layouts.

Quantitative Results on Sample Quality. Tab. 1 shows the results of sample quality and generation accuracy on the BRCA-M2C and Lizard datasets compared with the state-of-the-art cell layout generation baselines: ADM [11], TM-CCG [2] and Spatial Diffusion [44]. Across all multi-class datasets, our proposed *TopoCellGen* attains the lowest FID and TopoFD, indicating superior visual fidelity and topological accuracy. Moreover, *TopoCellGen* notably reduces multi-class count errors, resulting in more precise and topologically consistent synthetic cell layouts. These findings highlight *TopoCellGen*'s capacity to preserve realistic cell distributions, maintain topological relationships, and enable accurate control over cell density.

Performance on Downstream Tasks. In Tab. 2, we show the results of cell detection and classification tasks using our synthetic image-layout pairs as data augmentations. We present the results using two frameworks, UNet [54] and

	Method	FID ↓	Lym. ↓	Epi. ↓	Stro. ↓	Neu. ↓	Pla. ↓	Eos. ↓	Con. ↓	TCE ↓	TopoFD ↓	MMD ↓
BRCA-M2C	ADM [11]	1.150	13.757	40.230	15.491	–	–	–	–	22.465	133.012	0.732
	TMCCG [2]	0.634	11.503	34.032	12.907	–	–	–	–	19.687	89.252	0.635
	Spatial Diffusion [44]	0.263	10.852	35.954	13.496	–	–	–	–	20.806	97.584	0.589
	TopoCellGen	**0.005**	**2.090**	**3.824**	**2.468**	–	–	–	–	**5.192**	**69.354**	**0.421**
Lizard	ADM [11]	0.059	16.508	11.796	–	1.123	4.328	1.598	10.737	23.964	65.910	0.783
	TMCCG [2]	1.093	15.548	10.011	–	2.376	4.293	1.872	11.643	22.604	63.120	0.667
	Spatial Diffusion [44]	0.137	10.740	9.062	–	3.040	6.552	2.173	11.225	20.606	79.591	0.883
	TopoCellGen	**0.027**	**6.155**	**6.560**	–	**1.022**	**2.982**	**1.167**	**7.288**	**11.590**	**31.607**	**0.536**

Table 1. Results for BRCA-M2C and Lizard datasets on the quality of the generated samples.

Data	Method	F1-Score ↑				
		Lymphocytes	Epithelial	Stromal	Mean	Detection
Real.		0.569 ± 0.010	0.736 ± 0.012	0.507 ± 0.015	0.604 ± 0.011	0.857 ± 0.006
Real+Syn. (Rand)		0.549 ± 0.009	0.693 ± 0.014	0.472 ± 0.016	0.571 ± 0.013	0.848 ± 0.008
Real+Syn (TMCCG)	UNet	0.650 ± 0.007	0.768 ± 0.010	0.511 ± 0.012	0.643 ± 0.009	0.852 ± 0.005
Real+Syn (SpaDM)		0.647 ± 0.006	0.797 ± 0.003	0.554 ± 0.011	0.666 ± 0.007	0.853 ± 0.005
Real+Syn (*TopoCellGen*)		**0.656 ± 0.003**	**0.803 ± 0.005**	**0.574 ± 0.004**	**0.678 ± 0.004**	**0.860 ± 0.004**
Real.		0.615 ± 0.008	0.777 ± 0.010	0.540 ± 0.013	0.644 ± 0.009	0.855 ± 0.005
Real+Syn. (Rand)		0.578 ± 0.009	0.756 ± 0.012	0.502 ± 0.014	0.612 ± 0.010	0.851 ± 0.006
Real+Syn (TMCCG)	MCSpatNet	**0.678 ± 0.006**	0.800 ± 0.005	0.522 ± 0.014	0.667 ± 0.007	0.853 ± 0.004
Real+Syn (SpaDM)		0.639 ± 0.005	0.804 ± 0.007	0.563 ± 0.012	0.669 ± 0.006	0.855 ± 0.005
Real+Syn (*TopoCellGen*)		0.652 ± 0.004	**0.817 ± 0.006**	**0.582 ± 0.005**	**0.684 ± 0.004**	**0.862 ± 0.004**

Table 2. Results on cell detection and classification tasks on BRCA-M2C dataset. The best and statistically significant results are highlighted in **bold**.

MCSpatNet [1]. The results indicate that *TopoCellGen* achieves the highest F1 scores across various cell types, including inflammation and epithelial cells, resulting in superior mean F1 scores and detection metrics on the BRCA-M2C dataset. This demonstrates our *TopoCellGen*'s capability to capture both spatial fidelity and topological accuracy, ensuring that synthetic data closely resembles real biological structures. By accurately modeling complex spatial distributions and inter-class relationships, *TopoCellGen* provides biologically plausible synthetic samples that improve the generalizability of detection and classification models. Furthermore, its balanced representation of cellular compositions reduces class-wise biases, allowing the classifiers to better learn fine-grained distinctions.

4.2. Ablation Studies

Extensive experiments are conducted to elucidate the effectiveness and robustness of our loss components and hyperparameters. All experiments are performed on the BRCA-M2C dataset.

Ablation Study on Loss Components. We evaluate the contributions of three loss functions to the model's performance in generating synthetic cell layouts, measured by FID, Total Counting Error, and TopoFD. As shown in Tab. 3, the results demonstrate that $\mathcal{L}_{\text{count}}$ significantly reduces the counting errors, achieving better accuracy when combined with the other losses. The $\mathcal{L}_{\text{intra}}$ improves the fidelity and topological accuracy of the generated multi-class layouts, as indicated by lower FID and TopoFD. The $\mathcal{L}_{\text{inter}}$ has a smaller effect on the cell counting error but enhances structural consistency, as reflected in improved TopoFD. Combining all three losses yields superior results across all metrics, highlighting their complementary contributions to accurate and realistic multi-class cell layouts.

$\mathcal{L}_{\text{count}}$	$\mathcal{L}_{\text{intra}}$	$\mathcal{L}_{\text{inter}}$	BRCA-M2C		
			FID ↓	TCE ↓	TopoFD ↓
✗	✗	✗	1.150	22.465	133.012
✓	✗	✗	0.842	12.253	118.304
✗	✓	✗	0.621	16.315	98.798
✗	✗	✓	1.083	21.928	126.742
✓	✓	✗	0.232	6.854	85.672
✓	✗	✓	0.498	8.012	91.324
✗	✓	✓	0.327	17.573	73.612
✓	✓	✓	**0.005**	**5.192**	**69.354**

Table 3. Ablation study on loss components.

Ablation Study on Loss Weights. The ablation study on loss weights is presented in Tab. 4. The results indicate that when the loss weights are higher, such as 0.005 or 0.001, they impose overly strong regularization, leading to suboptimal outcomes. Conversely, when the loss weights are lowered to $1e-4$ or even $5e-5$, the regularization becomes too weak to achieve optimal performance. The configuration with moderate weights $5e-4$ achieves the best balance, yielding the lowest FID, total count error, and TopoFD.

λ_c	λ_{intra}	λ_{inter}	BRCA-M2C		
			FID ↓	TCE ↓	TopoFD ↓
0.005	0.005	0.005	0.289	14.920	92.718
0.001	0.001	0.001	0.153	12.471	85.642
1e-4	1e-4	1e-4	0.012	8.275	88.627
5e-5	5e-5	5e-5	0.129	11.378	83.629
5e-4	5e-4	5e-4	**0.005**	**5.192**	**69.354**

Table 4. Ablation study on loss weights.

5. Conclusion

In summary, *TopoCellGen* presents a robust framework for generating realistic cell topologies in digital pathology. It accurately preserves both intra- and inter-class spatial patterns, ensures cell count control, and achieves high structural fidelity. Experimental results confirm its close approximation of real tissue layouts, thereby enhancing downstream tasks such as cell detection and classification.

References

[1] Shahira Abousamra, David Belinsky, John Van Arnam, Felicia Allard, Eric Yee, Rajarsi Gupta, Tahsin Kurc, Dimitris Samaras, Joel Saltz, and Chao Chen. Multi-class cell detection using spatial context representation. In *ICCV*, 2021. 1, 7, 8, 12

[2] Shahira Abousamra, Rajarsi Gupta, Tahsin Kurc, Dimitris Samaras, Joel Saltz, and Chao Chen. Topology-guided multi-class cell context generation for digital pathology. In *CVPR*, 2023. 1, 3, 7, 8

[3] Muhammad Asad, Reuben Dorent, and Tom Vercauteren. Fastgeodis: Fast generalised geodesic distance transform. *Journal of Open Source Software*, 2022. 7

[4] Marco Aversa, Gabriel Nobis, Miriam Hägele, Kai Standvoss, Mihaela Chirica, Roderick Murray-Smith, Ahmed M Alaa, Lukas Ruff, Daniela Ivanova, Wojciech Samek, et al. Diffinfinite: Large mask-image synthesis via parallel random patch diffusion in histopathology. *NeurIPS*, 2024. 1, 2, 3

[5] Yoshua Bengio, Nicholas Léonard, and Aaron Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432*, 2013. 5

[6] Tim Brooks, Aleksander Holynski, and Alexei A Efros. Instructpix2pix: Learning to follow image editing instructions. In *CVPR*, 2023. 3

[7] Peter Bubenik et al. Statistical topological data analysis using persistence landscapes. *JMLR*, 2015. 6

[8] James R Clough, Nicholas Byrne, Ilkay Oksuz, Veronika A Zimmer, Julia A Schnabel, and Andrew P King. A topological loss function for deep-learning based image segmentation using persistent homology. *TPAMI*, 2020. 3

[9] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have l p-stable persistence. *Foundations of computational mathematics*, 2010. 6

[10] Srijay Deshpande, Muhammad Dawood, Fayyaz Minhas, and Nasir Rajpoot. Synclay: Interactive synthesis of histology images from bespoke cellular layouts. *Medical Image Analysis*, 2024. 1

[11] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *NeurIPS*, 2021. 1, 3, 6, 7, 8, 12, 13

[12] Ruiwen Ding, Prateek Prasanna, Germán Corredor, Cristian Barrera, Philipp Zens, Cheng Lu, Priya Velu, Patrick Leo, Niha Beig, Haojia Li, et al. Image analysis reveals molecularly distinct patterns of tils in nsclc associated with treatment outcome. *NPJ precision oncology*, 2022. 1

[13] Edelsbrunner, Letscher, and Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 2002. 3, 4, 12

[14] Herbert Edelsbrunner and John L Harer. *Computational topology: an introduction*. American Mathematical Society, 2022. 3, 4, 12

[15] Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. Generative diffusion prior for unified image restoration and enhancement. In *CVPR*, 2023. 3

[16] Pedro F Felzenszwalb and Daniel P Huttenlocher. Distance transforms of sampled functions. *Theory of computing*, 2012. 5

[17] Robin Forman. Morse theory for cell complexes. *Advances in mathematics*, 1998. 3

[18] Maurice Fréchet. Sur la distance de deux lois de probabilité. In *Annales de l'ISUP*, 1957. 6

[19] Rickard Brüel Gabrielsson, Bradley J Nelson, Anjan Dwaraknath, and Primoz Skraba. A topology layer for machine learning. In *AISTATS*, 2020. 3

[20] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 2020. 1

[21] Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *MedIA*, 2019. 1

[22] Simon Graham, Mostafa Jahanifar, Ayesha Azam, Mohammed Nimir, Yee-Wah Tsang, Katherine Dodd, Emily Hero, Harvir Sahota, Atisha Tank, Ksenija Benes, et al. Lizard: A large-scale dataset for colonic nuclear instance segmentation and classification. In *ICCV*, 2021. 7, 12

[23] Alexandros Graikos, Srikar Yellapragada, Minh-Quan Le, Saarthak Kapse, Prateek Prasanna, Joel Saltz, and Dimitris Samaras. Learned representation-guided diffusion models for large-image generation. In *CVPR*, 2024. 1, 2, 3

[24] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *JMLR*, 2012. 13

[25] Saumya Gupta, Xiaoling Hu, James Kaan, Michael Jin, Mutshipay Mpoy, Katherine Chung, Gagandeep Singh, Mary Saltz, Tahsin Kurc, Joel Saltz, et al. Learning topological interactions for multi-class medical image segmentation. In *ECCV*, 2022. 3

[26] Saumya Gupta, Dimitris Samaras, and Chao Chen. Topodiffusionnet: A topology-aware diffusion model. *arXiv preprint arXiv:2410.16646*, 2024. 3

[27] Saumya Gupta, Yikai Zhang, Xiaoling Hu, Prateek Prasanna, and Chao Chen. Topology-aware uncertainty for image segmentation. In *NeurIPS*, 2024. 3

[28] Robert Harb, Thomas Pock, and Heimo Müller. Diffusion-based generation of histopathological whole slide images at a gigapixel scale. In *WACV*, 2024. 3

[29] Hongliang He, Jun Wang, Pengxu Wei, Fan Xu, Xiangyang Ji, Chang Liu, and Jie Chen. Toposeg: Topology-aware nuclear instance segmentation. In *ICCV*, 2023. 1

[30] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *NeurIPS*, 2017. 5, 7, 13

[31] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020. 1, 3

[32] Fabian Hörst, Moritz Rempe, Lukas Heine, Constantin Seibold, Julius Keyl, Giulia Baldini, Selma Ugurel, Jens Siveke, Barbara Grünwald, Jan Egger, et al. Cellvit: Vision transformers for precise cell segmentation and classification. *MedIA*, 2024. 1

[33] Jinhui Hou, Zhiyu Zhu, Junhui Hou, Hui Liu, Huanqiang Zeng, and Hui Yuan. Global structure-aware diffusion process for low-light image enhancement. In *NeurIPS*, 2024. 3

[34] Le Hou, Ayush Agarwal, Dimitris Samaras, Tahsin M Kurc, Rajarsi R Gupta, and Joel H Saltz. Robust histopathology image analysis: To label or to synthesize? In *CVPR*, 2019. 1

[35] Xiaoling Hu. Structure-aware image segmentation with homotopy warping. In *NeurIPS*, 2022. 3

[36] Xiaoling Hu, Fuxin Li, Dimitris Samaras, and Chao Chen. Topology-preserving deep image segmentation. In *NeurIPS*, 2019. 3, 5

[37] Xiaoling Hu, Yusu Wang, Li Fuxin, Dimitris Samaras, and Chao Chen. Topology-aware segmentation using discrete morse theory. In *ICLR*, 2021. 3

[38] Xiaoling Hu, Dimitris Samaras, and Chao Chen. Learning probabilistic topological representations using discrete morse theory. In *ICLR*, 2023. 3

[39] Michael Kerber, Dmitriy Morozov, and Arnur Nigmetov. Geometry helps to compare persistence diagrams. In *2016 Proceedings of the Eighteenth Workshop on Algorithm Engineering and Experiments (ALENEX)*, 2016. 5

[40] Navid Alemi Koohbanani, Mostafa Jahanifar, Neda Zamani Tajadin, and Nasir Rajpoot. Nuclick: a deep learning framework for interactive segmentation of microscopic images. *MedIA*, 2020. 1

[41] Théo Lacombe, Marco Cuturi, and Steve Oudot. Large scale computation of means and clusters for persistence diagrams using optimal transport. In *NeurIPS*, 2018. 5

[42] D-S Lee, Juyong Park, KA Kay, Nicholas A Christakis, Zoltan N Oltvai, and A-L Barabási. The implications of human metabolic network topology for disease comorbidity. *PNAS*, 2008. 3

[43] Richard M Levenson, Yashbir Singh, Bastian Rieck, Quincy A Hathaway, Colleen Farrelly, Jennifer Rozenblit, Prateek Prasanna, Bradley Erickson, Ashok Choudhary, Gunnar Carlsson, et al. Advancing precision medicine: algebraic topology and differential geometry in radiology and computational pathology. *Laboratory Investigation*, 2024. 3

[44] Chen Li, Xiaoling Hu, Shahira Abousamra, Meilong Xu, and Chao Chen. Spatial diffusion for cell layout generation. In *MICCAI*, 2024. 2, 3, 7, 8, 13

[45] Ming Li, Taojiannan Yang, Huafeng Kuang, Jie Wu, Zhaoning Wang, Xuefeng Xiao, and Chen Chen. Controlnet++: Improving conditional controls with efficient consistency feedback. In *ECCV*, 2024. 1

[46] Cheng Lu, Can Koyuncu, German Corredor, Prateek Prasanna, Patrick Leo, XiangXue Wang, Andrew Janowczyk, Kaustav Bera, James Lewis Jr, Vamsidhar Velcheti, et al. Feature-driven local cell graph (flock): new computational pathology-based descriptors for prognosis of lung cancer and hpv status of oropharyngeal cancers. *MedIA*, 2021. 1

[47] Alessandro Lugli, Inti Zlobec, Martin D Berger, Richard Kirsch, and Iris D Nagtegaal. Tumour budding in solid cancers. *Nature Reviews Clinical Oncology*, 2021. 1

[48] Seonghui Min, Hyun-Jic Oh, and Won-Ki Jeong. Co-synthesis of histopathology nuclei image-label pairs using a context-conditioned joint diffusion model. In *ECCV*, 2024. 1, 3

[49] James R Munkres. Elements of algebraic topology, 1984. 3, 4, 12

[50] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *ICML*, 2021. 1, 3, 13

[51] Hyun-Jic Oh and Won-Ki Jeong. Diffmix: Diffusion model-based data synthesis for nuclei segmentation and classification in imbalanced pathology image datasets. In *MICCAI*, 2023. 1, 3

[52] Tim Prangemeier, Christoph Reich, and Heinz Koeppl. Attention-based transformers for instance segmentation of cells in microstructures. In *2020 IEEE international conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2020. 1

[53] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022. 3

[54] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015. 4, 7

[55] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. In *NeurIPS*, 2022. 3

[56] Roberto Salgado, Carsten Denkert, S Demaria, N Sirtaine, F Klauschen, Giancarlo Pruneri, S Wienert, Gert Van den Eynden, Frederick L Baehner, Frederique Pénault-Llorca, et al. The evaluation of tumor-infiltrating lymphocytes (tils) in breast cancer: recommendations by an international tils working group 2014. *Annals of oncology*, 2015. 1

[57] Joel Saltz, Rajarsi Gupta, Le Hou, Tahsin Kurc, Pankaj Singh, Vu Nguyen, Dimitris Samaras, Kenneth R Shroyer, Tianhao Zhao, Rebecca Batiste, et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell reports*, 2018. 1

[58] Hanan Samet and Markku Tamminen. Efficient component labeling of images of arbitrary dimension represented by linear bintrees. *TPAMI*, 1988. 13

[59] Denis Schapiro, Hartland W Jackson, Swetha Raghuraman, Jana R Fischer, Vito RT Zanotelli, Daniel Schulz, Charlotte Giesen, Raúl Catena, Zsuzsanna Varga, and Bernd Bodenmiller. histocat: analysis of cell phenotypes and interactions

in multiplex image cytometry data. *Nature methods*, 2017. 1, 2

[60] Suprosanna Shit, Johannes C Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien PW Pluim, Ulrich Bauer, and Bjoern H Menze. cldice-a novel topology-preserving loss function for tubular structure segmentation. In *CVPR*, 2021. 3

[61] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *ICML*, 2015. 1, 3

[62] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021. 3, 13

[63] Sasha E Stanton and Mary L Disis. Clinical significance of tumor-infiltrating lymphocytes in breast cancer. *Journal for immunotherapy of cancer*, 2016. 1

[64] Nico Stucki, Johannes C Paetzold, Suprosanna Shit, Bjoern Menze, and Ulrich Bauer. Topologically faithful image segmentation via induced matching of persistence barcodes. In *ICML*, 2023. 3

[65] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *CVPR*, 2016. 6

[66] Fan Wang, Huidong Liu, Dimitris Samaras, and Chao Chen. Topogan: A topology-aware generative adversarial network. In *ECCV*, 2020. 3, 7, 13

[67] Fan Wang, Zhilin Zou, Nicole Sakla, Luke Partyka, Nil Rawal, Gagandeep Singh, Wei Zhao, Haibin Ling, Chuan Huang, Prateek Prasanna, et al. Topotxr: A topology-guided deep convolutional network for breast parenchyma learning on dce-mris. *MedIA*, 2024. 3

[68] Haotian Wang, Min Xian, and Aleksandar Vakanski. Ta-net: Topology-aware network for gland segmentation. In *WACV*, 2022. 3

[69] Jueqi Wang, Jacob Levman, Walter Hugo Lopez Pinaya, Petru-Daniel Tudosiu, M Jorge Cardoso, and Razvan Marinescu. Inversesr: 3d brain mri super-resolution using a latent diffusion model. In *MICCAI*. Springer, 2023. 1

[70] Frederick H Xu, Michael Gao, Jiong Chen, Sumita Garai, Duy Anh Duong-Tran, Yize Zhao, and Li Shen. Topology-based clustering of functional brain networks in an alzheimer's disease cohort. *AMIA Summits on Translational Science Proceedings*, 2024. 3

[71] Meilong Xu, Xiaoling Hu, Saumya Gupta, Shahira Abousamra, and Chao Chen. Semi-supervised segmentation of histopathology images with noise-aware topological consistency. In *ECCV*, 2024. 3

[72] Srikar Yellapragada, Alexandros Graikos, Prateek Prasanna, Tahsin Kurc, Joel Saltz, and Dimitris Samaras. Pathldm: Text conditioned latent diffusion model for histopathology. In *WACV*, 2024. 1, 3

[73] Chenyu You, Weicheng Dai, Fenglin Liu, Yifei Min, Nicha C Dvornek, Xiaoxiao Li, David A Clifton, Lawrence Staib, and James S Duncan. Mine your own anatomy: Revisiting medical image segmentation with extremely limited labels. *TPAMI*, 2024. 1

[74] Yinyin Yuan. Spatial heterogeneity in the tumor microenvironment. *Cold Spring Harbor perspectives in medicine*, 2016. 2

[75] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *ICCV*, 2023. 1

[76] Shihao Zhao, Dongdong Chen, Yen-Chun Chen, Jianmin Bao, Shaozhe Hao, Lu Yuan, and Kwan-Yee K Wong. Uni-controlnet: All-in-one control to text-to-image diffusion models. In *NeurIPS*, 2024. 1

[77] Zixiang Zhao, Haowen Bai, Yuanzhi Zhu, Jiangshe Zhang, Shuang Xu, Yulun Zhang, Kai Zhang, Deyu Meng, Radu Timofte, and Luc Van Gool. Ddfm: denoising diffusion model for multi-modality image fusion. In *ICCV*, 2023. 3

TopoCellGen: Generating Histopathology Cell Topology with a Diffusion Model
— Supplementary Material —

In the supplementary material, we begin with notations for foreground and background in Sec. 6, followed by a description of the background knowledge about persistent homology in Sec. 7. Next, we provide detailed introduction to our layout-guided pathology image generation part in Sec. 8, followed by the comprehensive descriptions of the datasets in Sec. 9. The implementation details are provided in Sec. 10. In Sec. 11, we discuss the evaluation metrics in detail. To ensure the generation accuracy, we conduct the analysis on cell count distribution across training and test sets in Sec. 12. More ablation studies are given in Sec. 13. The biological plausibility analysis by the domain expert is provided in Sec. 14. Then, we provide the spatial point pattern analysis using multivariate Ripley's K-functions in Sec. 15, followed by the discussion on computational cost and scalability of our method in Sec. 16. Finally, we discuss the limitations of our method in Sec. 17.

6. Notes on Foreground and Background

Here, we provide some notations about foreground and background in our paper. In our experiments, the background of the layouts is black (the pixel value of 0) as can be seen in Fig. 2 and Fig. 5. For better visualization, we display the multi-class cell layouts with white as the background in Fig. 1, Fig. 3 and Fig. 4 of the main paper.

7. Background: Persistent Homology

In algebraic topology [49], homology classes capture topological features across different dimensions. For instance, 0-, 1-, and 2-dimensional structures represent connected components, loops (or holes), and voids, respectively. In binary images, the number of d-dimensional topological features is described by the d-dimensional Betti number, β_d.[1] While topological structures are well-defined in binary images, extending this theory to real-world data, which is often continuous and noisy, poses challenges.

In the case of analyzing cell point clouds, where data is inherently discrete, we require a robust framework to infer the underlying topological structures. Persistent homology, developed in the early 2000s [13, 14], addresses this need by tracking the evolution of topological features across multiple scales.

Given a point cloud in the 2D space $P \subseteq \mathbb{R}^2$, a filtration is built by considering a growing family of simplicial complexes constructed from the point cloud as a function of a parameter (e.g., radius). For each parameter value, we define a set of simplices connecting the points, starting from isolated vertices and gradually adding edges and higher-dimensional simplices as the parameter increases. This creates a series of nested simplicial complexes: $\varnothing \subseteq K_{r_1} \subseteq K_{r_2} \subseteq ... \subseteq K_{r_n}$. As the parameter grows, the topology of the complexes changes, with new connected components and loops emerging or vanishing.

Persistent homology captures these changes, tracking the birth and death of topological features over the filtration. The result is summarized in a persistence diagram (Dgm), which provides a multi-scale representation of the topological structures. A Dgm consists of points in a 2D plane, each representing a topological feature. The coordinates of each point, (b, d), correspond to the feature's birth and death filtration values, providing a concise description of its persistence across scales.

8. Layout To Image Generation

In this section, we introduce our layout-guided image generation framework in detail. The framework leverages a guided diffusion model (ADM) [11] to generate H&E images conditioned on multi-class cell layouts. The layouts serve as explicit conditional inputs to the diffusion model, which learns to reconstruct high-resolution pathology images from noisy counterparts during the reverse diffusion process. The conditioning mechanism is implemented using a cross-attention layer that seamlessly integrates cell layout information into the diffusion model. As shown in Fig. 6, the generated H&E images generated by the model accurately depict the relative densities and arrangements of different cell types, while preserving the fine-grained details characteristic of histopathology images, such as nuclear shapes and staining patterns. This helps greatly improve the performance of downstream tasks, such as cell detection and classification.

9. Details of the Datasets

BRCA-M2C dataset [1] is obtained from the TCGA dataset and contains 80, 10, and 30 pathology image patches for training, validating, and testing, respectively. This dataset provides dot annotations for multi-class classification in breast cancer images. All images are around 500×500 pixels. The cell classes are lymphocytes, tumor or epithelial, and stromal cells.

Lizard Dataset [22] is a large-scale resource for nuclear instance segmentation and classification, specifically target-

---

[1] Technically, β_d measures the dimension of the d-dimensional homology group. The number of distinct homology classes is exponential in β_d.
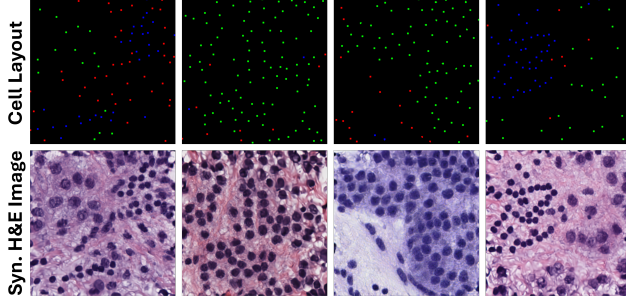
Figure 6. Qualitative results generated by our layout and image generation framework for downstream tasks.

ing colonic tissue in computational pathology. It includes nearly $495,000$ manually and semi-automatically annotated nuclei, categorized into six classes: epithelial cells, connective tissue cells, lymphocytes, plasma cells, neutrophils, and eosinophils. 238 images in the dataset are sourced from 6 publicly available datasets, ensuring diverse representations of normal, inflammatory, dysplastic, and cancerous colonic conditions.

10. Implementation Details

Our work is mainly based on guided-diffusion (ADM) [11]. The condition of our model is a list of cell counts. An embedding of the condition is obtained by using an encoding network. After that, we feed this embedding to all the residual blocks in the network by adding it to the timestep embedding [50]. For every dataset, the image resolution is 256×256. Our diffusion models use a cosine noise scheduler [50], with noising timesteps of 1000 for training. We first pre-train the diffusion model using only $\mathcal{L}_{simple}$ for 150K steps, then train with the three losses for 210K steps. During the inference, we use 100 steps of DDIM [62]. The learning rate is 2×10^{-5} and the batch size is 5. λ_c, λ_{intra} and λ_{inter} are all set to 0.0005.

For the layout-guided generation model, the learning rate is also 2×10^{-5} and we train the model only using $\mathcal{L}_{simple}$ for 360K steps. The batch size is 6. The image resolution is also 256×256. These experiments were conducted on 1 NVIDIA RTX A6000 GPU with 48GB RAM.

Our experiments designate specific test sets for each dataset to evaluate the synthetic cell layout generation process. For the BRCA-M2C dataset, we utilize 30 images in the test set, which were pre-defined in the dataset. To prepare these for testing, each image is segmented into patches using a sliding window approach with a stride of 32 pixels, resulting in patches of size 256×256. This process yields a total of $1,550$ patches for the BRCA-M2C dataset. We randomly select 20% of the cell layouts as the test set for the Lizard dataset, which lacks predefined training and test splits. The chosen images undergo the same patching pro-

cedure, generating 256×256 patches, resulting in $1,000$ patches for the test set of the Lizard dataset.

In generating synthetic layouts, we aim to match the channel-wise cell counts observed in the real layouts of the test set. For each real test layout, we calculate the counts of each cell type across the channels and use these as conditional inputs during inference. This ensures that the generated synthetic patches exhibit similar cell count distributions to those observed in the real test layouts.

11. Evaluation Metrics

To evaluate the quality of the generated cell layouts and pathology images, we employ a set of metrics focusing on different aspects, such as visual fidelity, topological similarity, and utility to downstream tasks.

First, the **Fréchet Inception Distance (FID)** [30] measures visual similarity by comparing the distributions of features extracted from a pre-trained Inception network between real and generated images. Lower FID scores indicate greater visual realism in the generated images. Feature extraction is tailored to each dataset with custom-trained models. Here, the FID we used is the **spatial-FID** proposed in Spatial Diffusion [44]. The spatial-FID replaces visual features with a spatial representation derived from an autoencoder's intermediate layer, and we trained the autoencoder in the same way. In addition, we extended it to the Lizard dataset by training another autoencoder in the same manner. We also evaluate the accuracy of the generation through cell count error, calculating discrepancies between real and generated cell counts per type and overall. In our experiments, we use the connected component labeling method [58] to count the cell numbers. Assume there are n types of cells. For each cell type i, the **cell count error (CCE)** across N test samples is defined as:

$$Cell\_Count\_Error^{(i)} = \frac{1}{N} \sum_{j=1}^{N} \left| c_{real,j}^{(i)} - c_{syn,j}^{(i)} \right|$$

with **total count error (TCE)** calculated as:

$$Total\_Count\_Error = \frac{1}{N} \sum_{j=1}^{N} \left| \sum_{i=1}^{n} c_{real,j}^{(i)} - \sum_{i=1}^{n} c_{syn,j}^{(i)} \right|$$

where c is the cell count. In addition, our proposed **TopoFD** metric is used to evaluate the topological similarity between real and generated cell layouts. Lower TopoFD scores indicate closer alignment in spatial structure.

We also use the metric proposed in [66], Maximum Mean Discrepancy (MMD) [24] to measure the topological difference between the real and synthetic distributions. The persistence diagrams from synthetic and real layouts are embedded into a reproducing kernel Hilbert space (RKHS).

The MMD computes the distance between the mean embeddings of these two distributions in the RKHS. Given two sets of persistence diagrams, $\mathcal{D}_{syn} = \{Dgm_i^{syn}\}_{i=1}^N$ from the synthetic data and $\mathcal{D}_{real} = \{Dgm_j^{real}\}_{j=1}^N$ from the real data, we can define the mean of each diagram set,

$$\Phi(\mathcal{D}_{syn}) := \frac{1}{N}\sum_{i=1}^N \Phi(Dgm_i^{syn})$$

$$\Phi(\mathcal{D}_{real}) := \frac{1}{N}\sum_{j=1}^N \Phi(Dgm_i^{real})$$

Then, the MMD is defined as:

$$\mathrm{MMD}(\mathcal{D}_{\mathrm{syn}}, \mathcal{D}_{\mathrm{real}}) := \|\Phi(\mathcal{D}_{\mathrm{syn}}) - \Phi(\mathcal{D}_{\mathrm{real}})\|_{\mathcal{H}}$$

In terms of the kernel for persistence diagrams, we use the Gaussian kernel based on the 1-Wasserstein distance between diagrams,

$$k_{W_1}(Dgm_i, Dgm_j) = \exp\left(-\frac{W_1(Dgm_i, Dgm_j)}{\sigma^2}\right)$$

Lastly, to enhance downstream utility, we used $2,000$ generated image-layout pairs as augmented training data for cell detection and classification tasks, evaluating their performance with the **F1-score**.

12. Cell Count Distribution Analysis

Also, to ensure the accurate generation of cell distributions, the training set encompasses a wide range of cell count values. As shown in Fig. 7, we randomly select $2,000$ patches during the training. We analyzed and observed each cell type's range of cell counts in the training patches to confirm coverage across typical values observed in test conditions. This observation is crucial for the diffusion model, as it needs exposure to the range of cell counts during training to accurately generate corresponding counts during the inference.

13. Additional Ablation Study

Ablation Study on learning rate. This ablation study examines the effect of different learning rates on model performance. The results indicate that a learning rate of 2×10^{-5} achieves the best overall performance across all metrics, with the lowest FID, Total Counting Error, and TopoFD values. Higher learning rates, such as 1×10^{-4}, result in a higher total counting error and TopoFD, suggesting that an overly large learning rate may hinder convergence. Conversely, lower learning rates, including 1×10^{-5} and 5×10^{-5}, show some improvements but do not reach the optimal balance across all metrics. The chosen learning rate of 2×10^{-5}, therefore, appears to provide the best trade-off, facilitating convergence that enhances both cell counting accuracy and fidelity in the synthetic cell layouts.
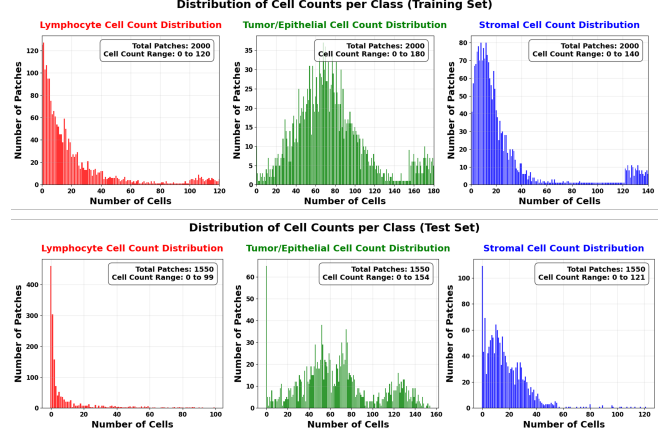

Figure 7. The statistical analysis of the cell count distributions on the BRCA-M2C training and test sets.

learning rate	BRCA-M2C		
	FID ↓	TCE ↓	TopoFD ↓
1e-4	0.021	12.357	75.667
1e-5	0.015	6.314	81.397
5e-5	0.066	12.367	85.949
2e-5	**0.005**	**5.192**	**69.354**

Table 5. Ablation study on learning rate.

14. Biological Plausibility

Specifically, we randomly selected 10 pairs of real and synthetic cell layouts as shown in Fig. 8. Without revealing their type (synthetic/real), we asked the expert to (1) identify which layout is synthetic; (2) characterize the tissue biology of these layouts. The expert achieves only a 60% accuracy in identifying the synthetic layout, confirming the realism of our synthetic layouts even to a domain expert. Regarding the characterization of tissue biology, as shown in Fig. 8, the pathologist concluded that for each pair of layouts, the synthetic layout preserved the defining biological characteristics of its corresponding real sample, consistently reflecting benign/low-grade or cancerous/high-grade properties. These experiments with a domain expert offer direct evidence, beyond quantitative measures and downstream analyses, that our generated layouts align well with actual biological structures.
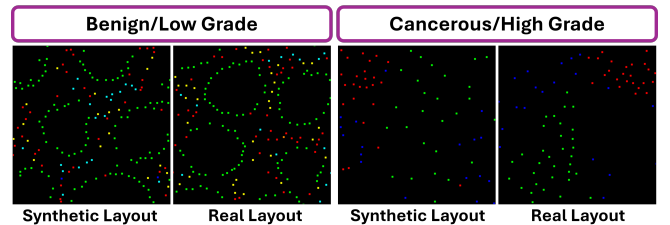

Figure 8. Biological plausibility validated by the domain expert.

15. Spatial Point Pattern Analysis

We also evaluate our synthetic layouts using one standard statistical method for spatial point pattern analysis. Specifically, we employ multivariate Ripley's K-functions to evaluate the synthetic layouts of the BRCA-M2C dataset, which comprises 3 cell types. For each test reference layout, we have a corresponding synthetic layout and extract cell centroids from both. We then compute 3 K-functions to capture intra-class clustering (one per cell type) and 6 cross-K functions to describe inter-class interactions. Next, we examine the difference between real and synthetic K-values over 6 radii: $[15, 30, 45, 60, 75, 90]$. For each radius and each cell-type pair, we perform a paired t-test to check if synthetic data deviates significantly from real layouts. This procedure yields 54 p-values (18 from intra-class and 36 from inter-class analyses). We then count the number of cases where these p-values exceed 0.05, indicating no statistically significant difference. Overall, as shown in Tab. 6 and Tab. 7, *TopoCellGen* achieves a greater number of radii with no significant difference is observed, compared to other methods. It most accurately produces both intra-class clustering and inter-class interactions, demonstrating the closest alignment with real data across the evaluated radii.

Method	BRCA-M2C			
	Lym. – Lym.	Epi. – Epi.	Stro. – Stro.	Total
ADM	0/6	0/6	2/6	2/18
TMCCG	2/6	1/6	2/6	5/18
Spatial Diffusion	1/6	3/6	2/6	6/18
TopoCellGen	3/6	5/6	4/6	12/18

Table 6. Number of radii with no statistically significant difference ($p > 0.05$) for intra-class spatial clustering.

Method	BRCA-M2C						
	Lym. – Epi.	Lym. – Stro.	Epi. – Lym.	Epi. – Stro.	Stro. – Lym.	Stro. – Epi.	Total
ADM	1/6	3/6	2/6	1/6	1/6	2/6	10/36
TMCCG	3/6	2/6	3/6	4/6	3/6	2/6	17/36
Spatial Diffusion	3/6	4/6	2/6	3/6	1/6	2/6	15/36
TopoCellGen	4/6	3/6	4/6	5/6	3/6	5/6	24/36

Table 7. Number of radii with no statistically significant difference ($p > 0.05$) for inter-class spatial interactions.

16. Computational Costs and Scalability

Currently, our model is trained on a single NVIDIA A6000 GPU with 48 GB of memory for approximately 360K steps, using a batch size of 5 at 256×256 resolution within 200 hours. The experiments can also be seamlessly scaled with data parallel training.

17. Limitations

Our proposed *TopoCellGen* will fail in some cases. First, the model is limited by its dependence on the cell types present in the training data, preventing it from generating layouts containing unseen cell types. Additionally, the model currently generates cell layouts in 256×256 patches, which constrains its application to small-scale regions.