Stony Brook University

CVPR Nashville JUNE 11-15, 2025

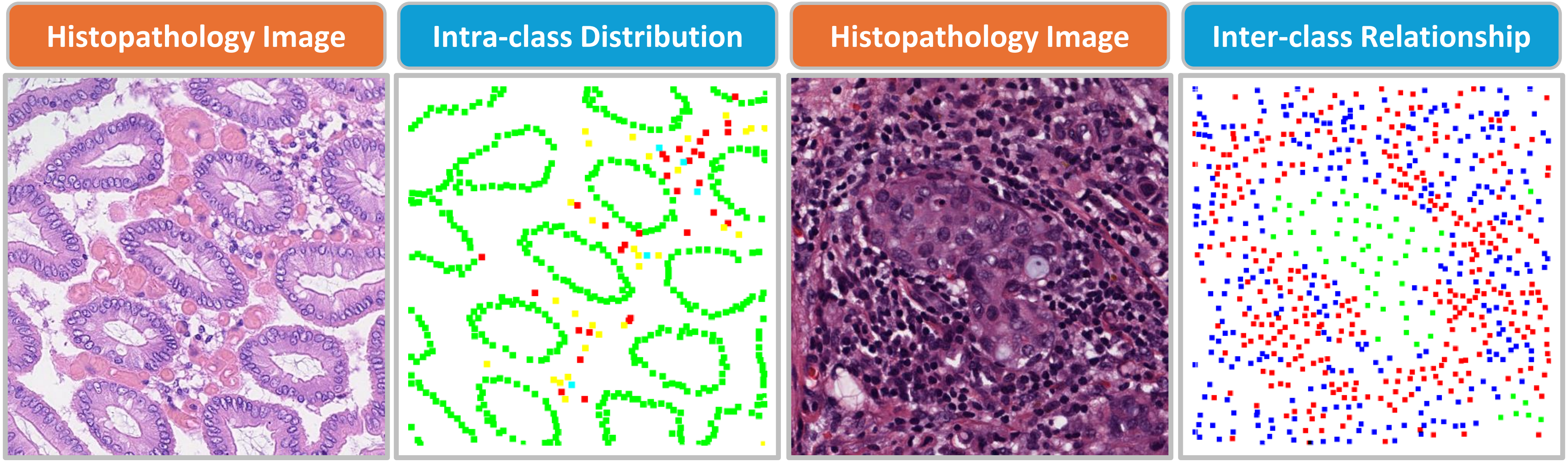
VE RI TAS

TopoCellGen
Generating Histopathology Cell Topology with a Diffusion Model

🏆 **Oral Presentation**

Meilong Xu, Saumya Gupta, Xiaoling Hu, Chen Li, Shahira Abousamra,
Dimitris Samaras, Prateek Prasanna, **Chao Chen**

What is Cell Topology?



Histopathology Image

Intra-class Distribution

Histopathology Image

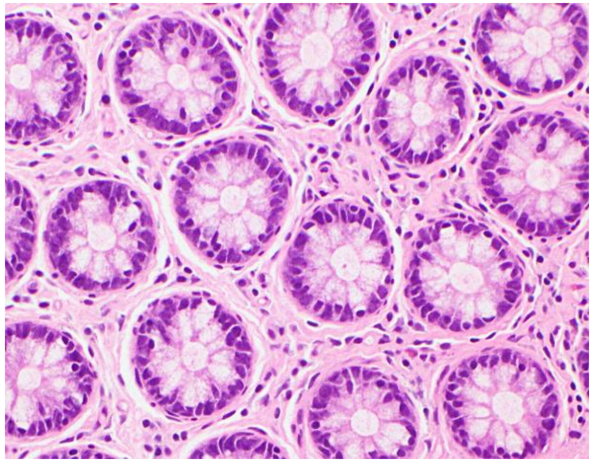
Inter-class Relationship

🟩 Tumor/Epithelial 🟥 Lymphocyte 🟦 Stromal Cell

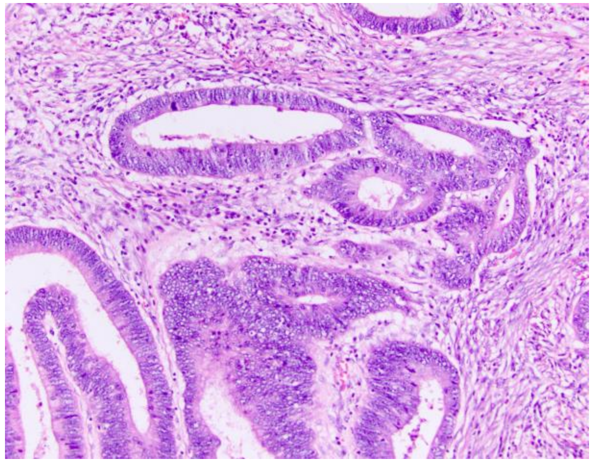
Why Cell Topology?

Diagnosis depends on who a cell's neighbors are.

Single-class cell distribution



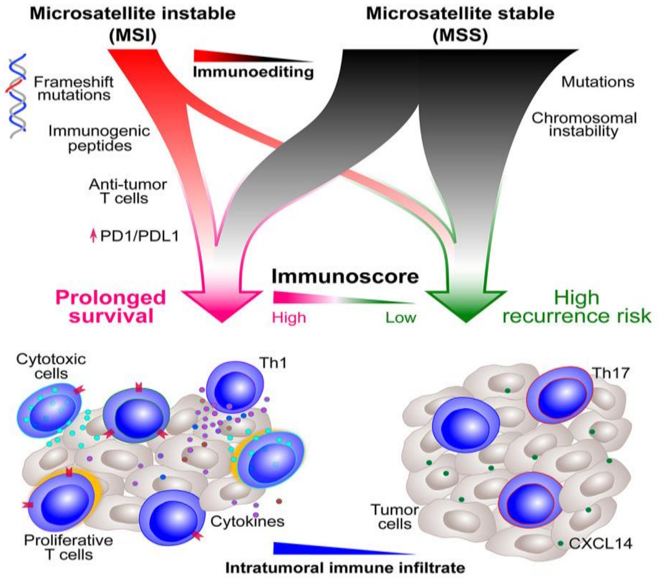
Normal glandular structure



Grade 4 of Gleason

Wikimedia Commons

Multi-class cell relationships



Mlecnik et al. *Immunity*, 2016[1]

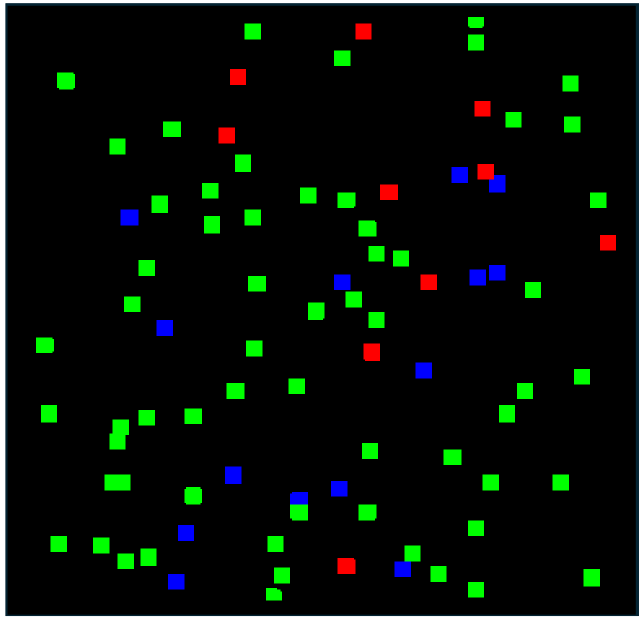
---

[1] Mlecnik, Bernhard, et al. *Immunity* 44.3 (2016)

Cell Topology Generation

$c=[c_0, c_1, c_2]$

Cell Count Per Class



🟥 Lymphocyte

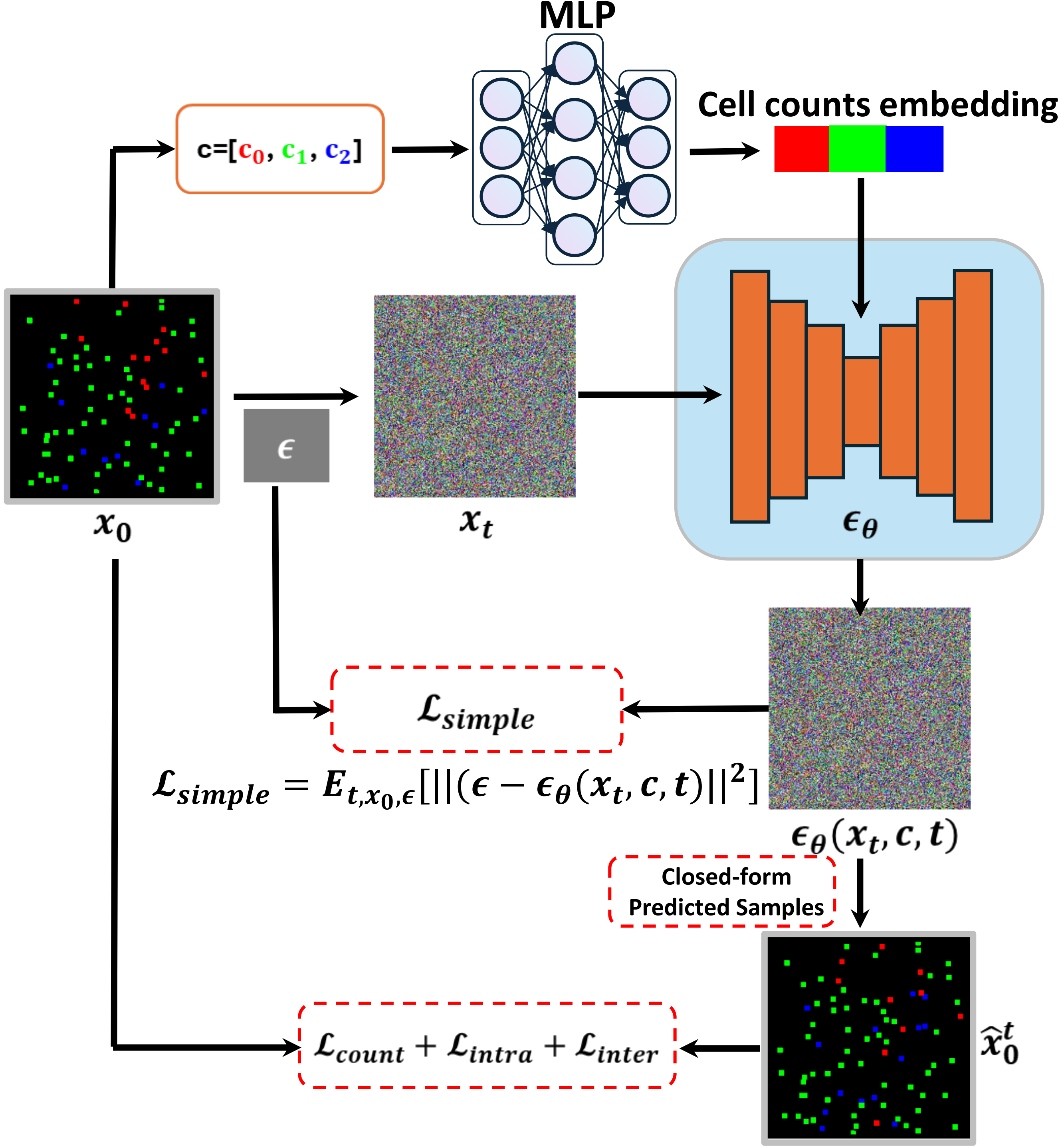
🟩 Tumor/Epithelial

🟦 Stromal Cell

- Cell count and density
- Intra-class cell distribution
- Inter-class cell interactions

Aiming to generate accurate cell topology!

Method: Overview



MLP

$c=[c_0, c_1, c_2]$

Cell counts embedding

x_0

ϵ

x_t

ϵ_θ

$\mathcal{L}_{simple}$

$$\mathcal{L}_{simple} = E_{t,x_0,\epsilon}[||(\epsilon - \epsilon_\theta(x_t, c, t)||^2]$$

$\epsilon_\theta(x_t, c, t)$

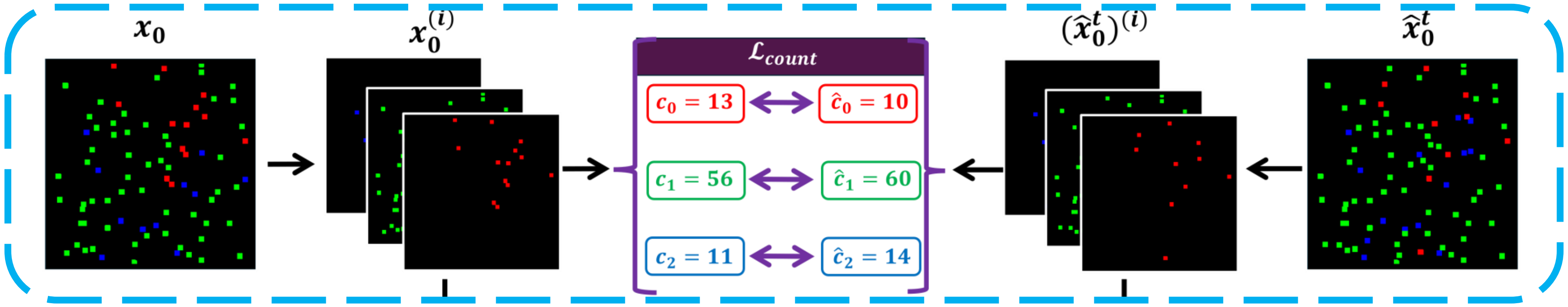
Closed-form Predicted Samples

$\hat{x}_0^t$

$\mathcal{L}_{count} + \mathcal{L}_{intra} + \mathcal{L}_{inter}$

Method: Ensure accurate count per class

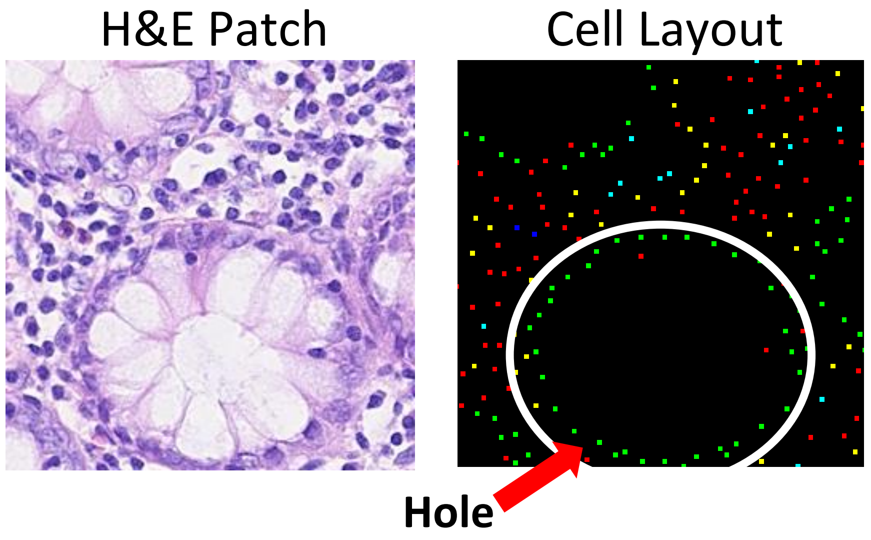
- Pixel-level enforcement
- Maintain the count of cells for each cell type



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

Topological Feature

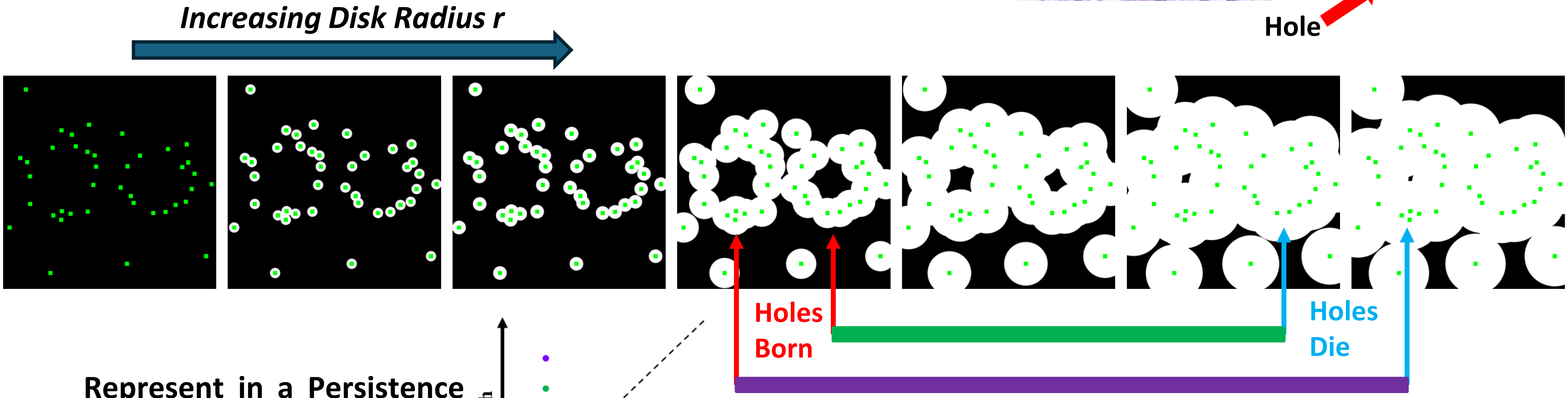
- Cell topology – Here mainly represents the holes
- Use Persistent Homology[2] to capture the holes.



H&E Patch

Cell Layout

Hole

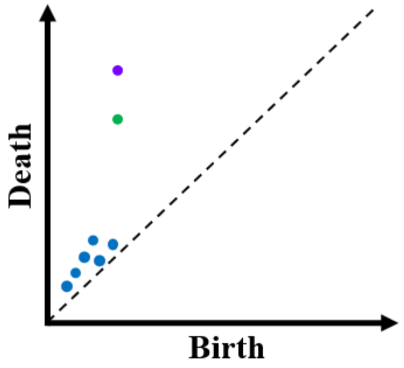


Increasing Disk Radius r

Holes Born

Holes Die

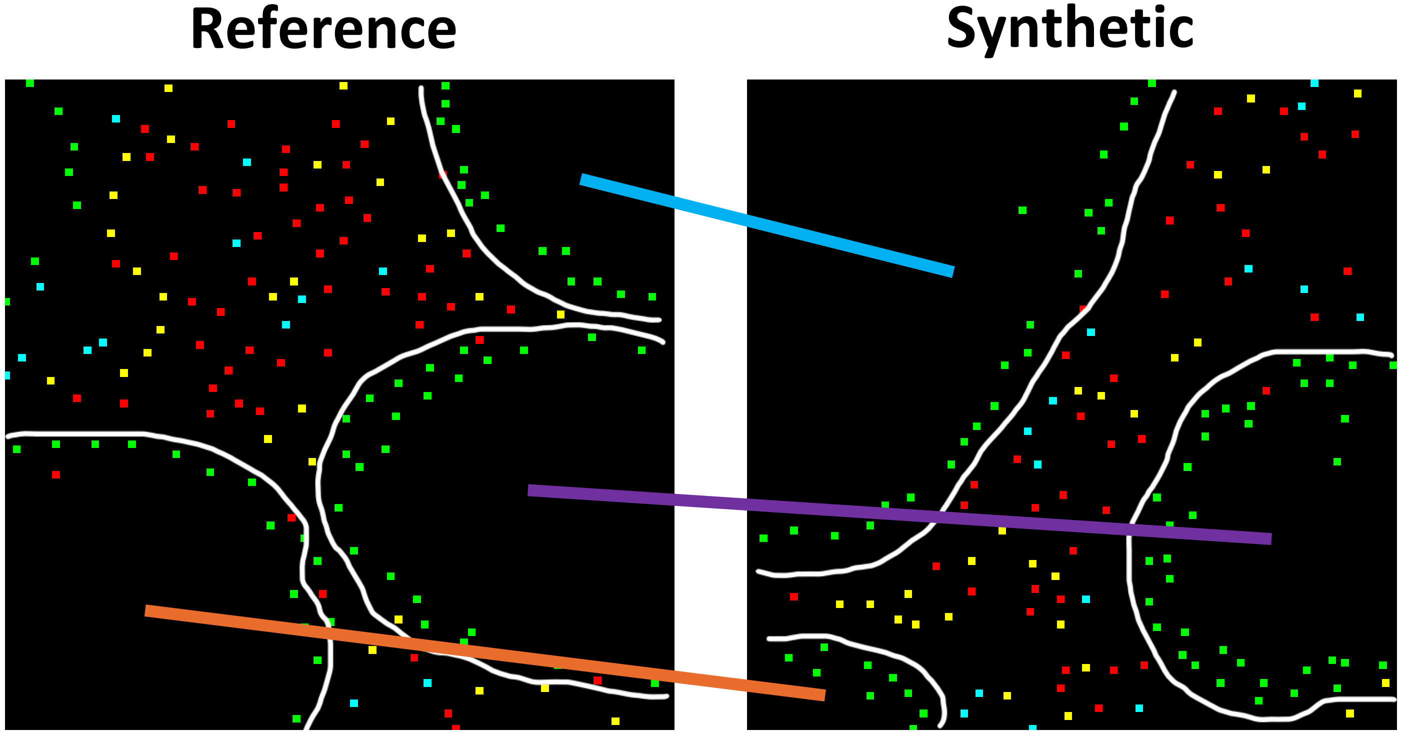
Represent in a Persistence Diagram (Dgm):



[2] Edelsbrunner, Herbert, and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

Matching Structures: Holes

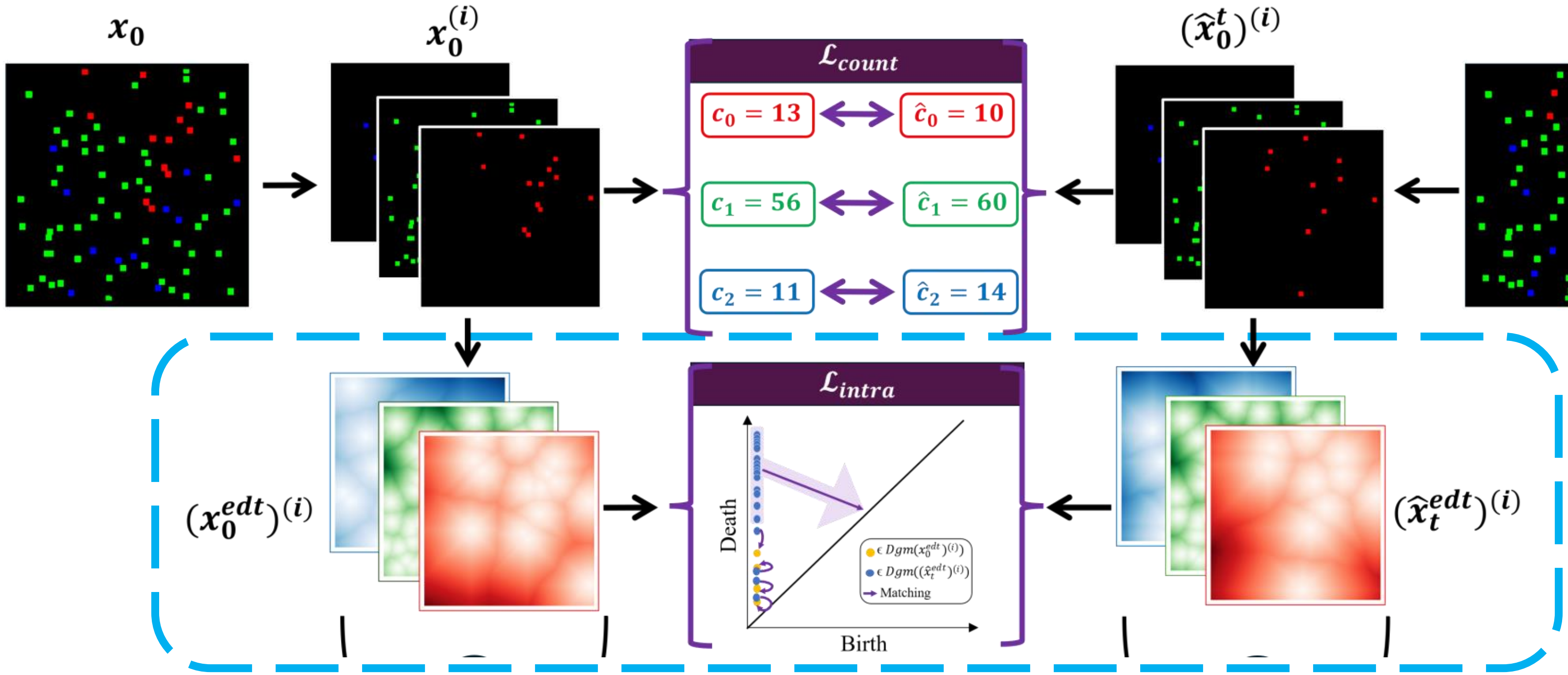
- Target: Ref. and Syn. have similar intra/inter-class topological patterns.
- Find the correspondence between the Ref. and Syn. structures.
- The matching is based on the Wasserstein distance[3].



[3] Villani, Cédric. *Optimal transport: old and new*. Vol. 338. Berlin: springer, 2008.

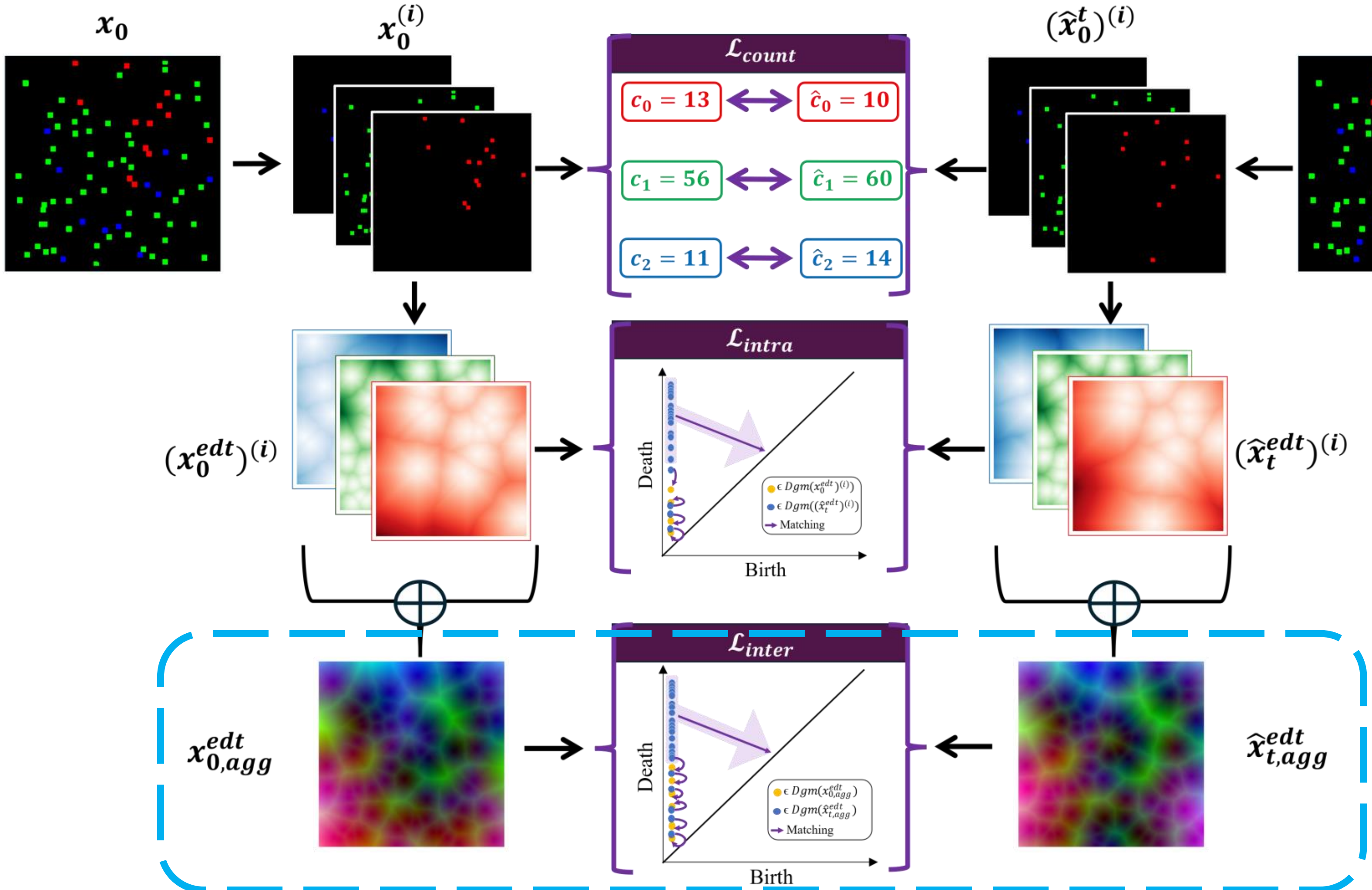
Method: $\mathcal{L}_{intra}$

- Matched structures should have similar spatial patterns of same type of cells.
- e.g. have similar values at the matched locations in class-wise distance transform map.



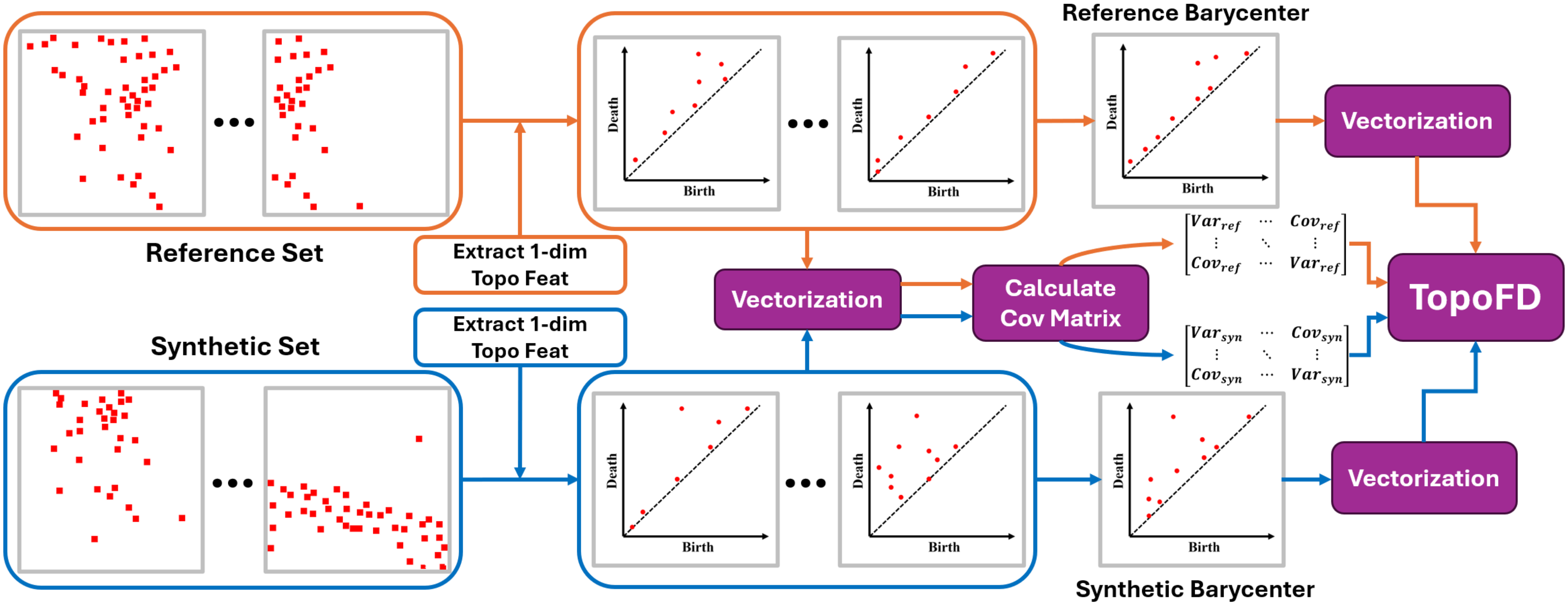
Method: $\mathcal{L}_{inter}$

- Matched structures should have similar spatial patterns of different types of cells.
- e.g. have similar values at the matched locations in a combined distance transform map.



New Metric: Topological Fréchet Distance

- Conventional FID measures feature similarity, but fails to capture complex spatial and topological cell distributions.



Reference Set

Extract 1-dim Topo Feat

Synthetic Set

Extract 1-dim Topo Feat

Reference Barycenter

Vectorization

Vectorization

Calculate Cov Matrix

$$\begin{bmatrix} Var_{ref} & \cdots & Cov_{ref} \\ \vdots & \ddots & \vdots \\ Cov_{ref} & \cdots & Var_{ref} \end{bmatrix}$$

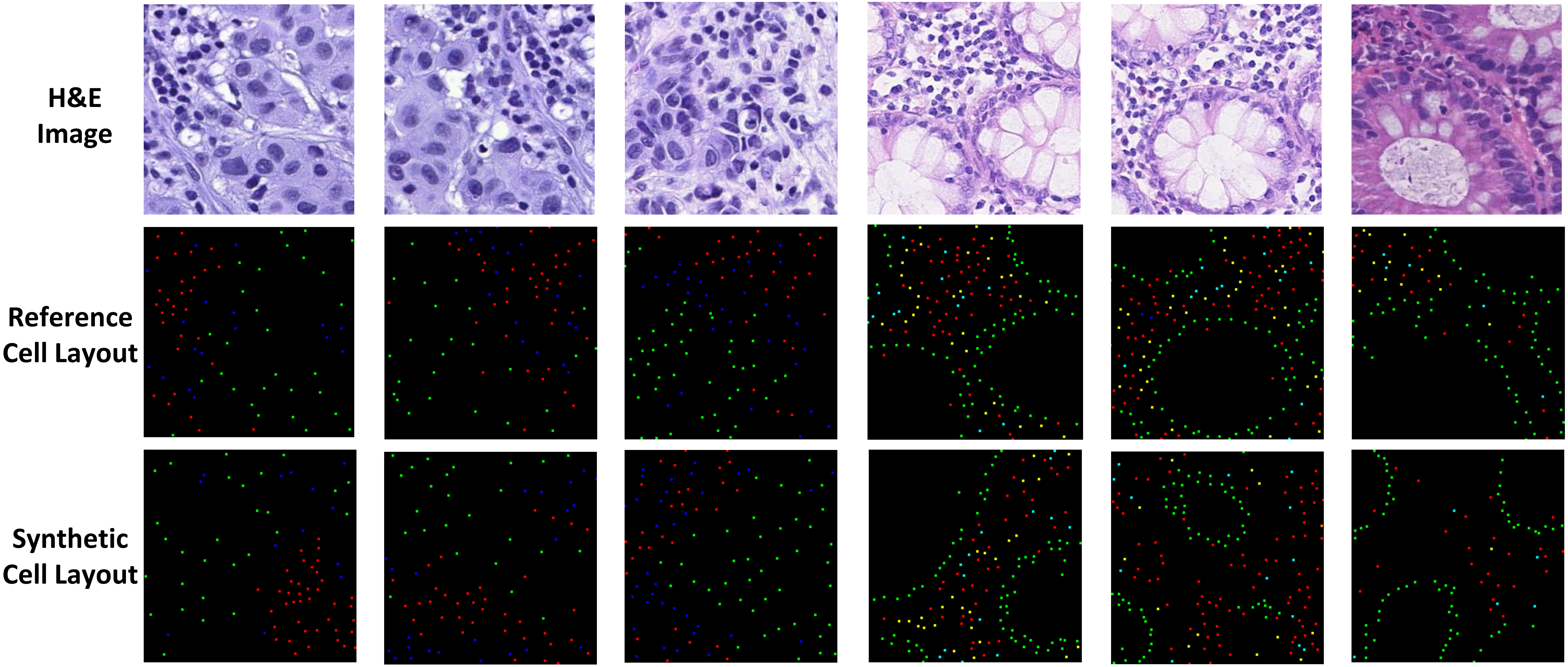
$$\begin{bmatrix} Var_{syn} & \cdots & Cov_{syn} \\ \vdots & \ddots & \vdots \\ Cov_{syn} & \cdots & Var_{syn} \end{bmatrix}$$

TopoFD

Vectorization

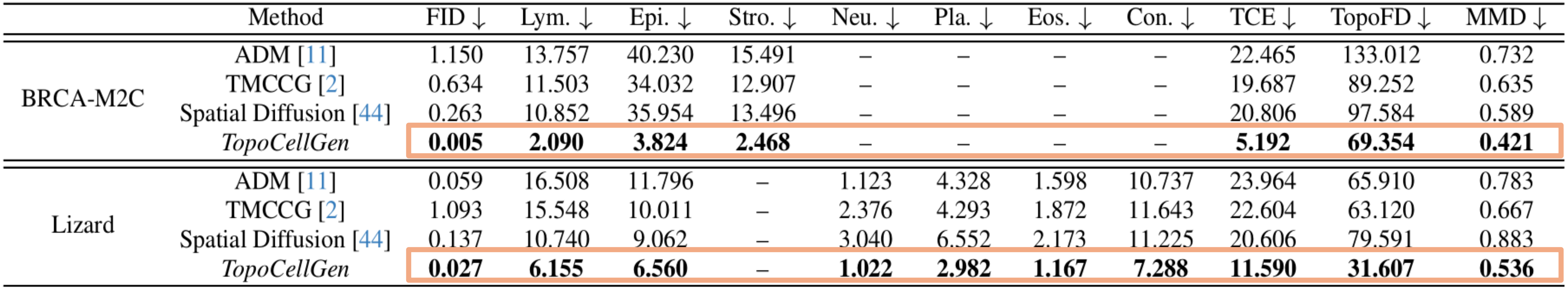
Synthetic Barycenter

Qualitative Results



Quantitative Evaluation: Generation Quality

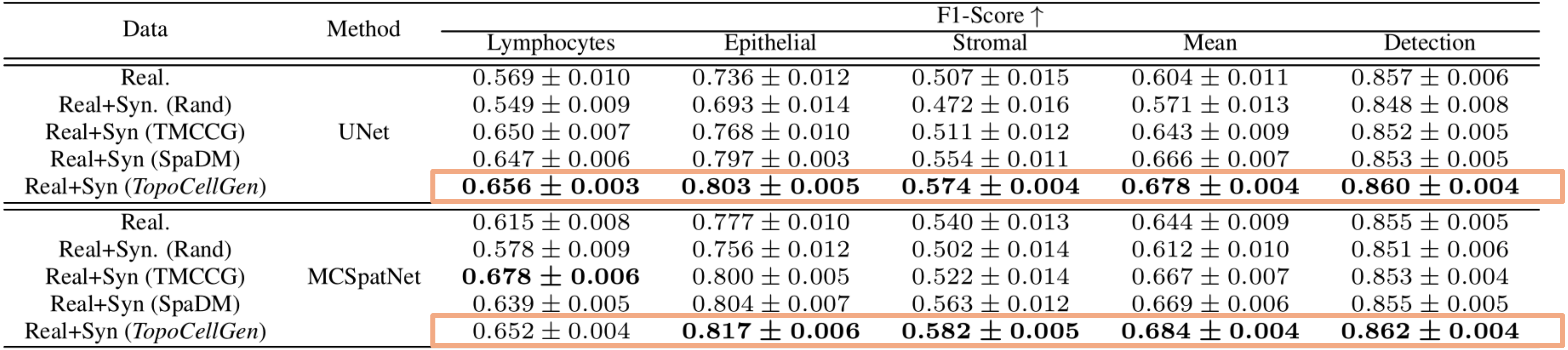
- Conduct experiments on BRCA-M2C and Lizard datasets.



	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**

Quantitative Evaluation: Downstream Performance

- Downstream Task: cell detection and classification



Data	Method	F1-Score ↑				
		Lymphocytes	Epithelial	Stromal	Mean	Detection
Real.	UNet	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)		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.	MCSpatNet	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)		**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**

Conclusion

- The first diffusion-based framework for cell topology generation.
- Introduce a novel topology-aware metric to quantify topological fidelity of synthetic layouts.
- Enhance downstream task performance greatly using augmented image-layout pairs.

Thank you!