STREAM: Stochastic Riemannian Flow Matching with Anisotropic Decoder for Digital Histopathology Image Generation

Won June Cho, Daeky Jeong, Hyeongyeol Lim, Hongjun Yoon
AI Research Lab, DEEPNOID Inc, Korea

This work was supported by the Technology Innovation Program (RS-2025-02221011, Development of Medical-Specialized Multimodal Hyperscale Generative AI Technology for Global Integration) funded by the Ministry of Trade Industry & Energy (MOTIE, South Korea).

NEURAL INFORMATION PROCESSING SYSTEMS

DEEP NOID

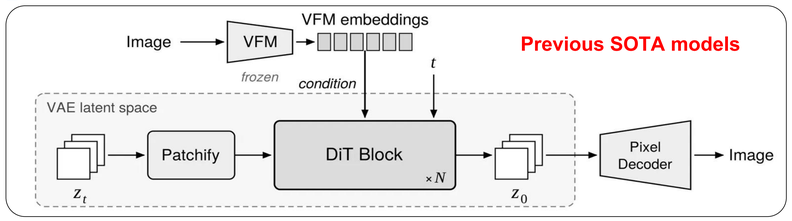
STREAM Project Page (NeurIPS 2026 Spotlight)

See also: SheafStain (NeurIPS 2026 Poster)

TL;DR

STREAM trains Riemannian flow matching in a pathology VFM's hyperspherical latent: a bridge perturbation guarantees per-token rectifiability and a decoder with Jacobian-shaped anisotropic noise adds superadditive gains for state-of-the-art generation on three different histopathology datasets.

What is the problem?



VFM vs VAE linear probing (benign vs malignant)

Model	d	AUROC	Acc_{dir}	Acc_{mag}
UNI	1024	0.995	0.993	0.583
UNI2-h	1536	**0.997**	**0.995**	**0.603**
DINOv3	1024	0.937	0.930	0.514
PixCell VAE	16	0.640	0.832	0.549
ZoomLDM VAE	3	0.624	0.609	0.543

Conditioning-dominated diversity

Model	H(X|C)	H(X)	ρ_{cond}
ZoomLDM	1.28	5.17	75%
PixCell	1.81	4.78	62%

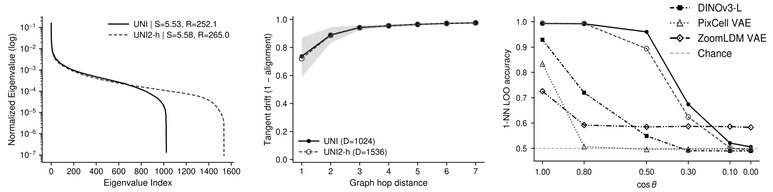
$$\rho_{\text{cond}} = \frac{I(X;C)}{H(X)} = 1 - \frac{H(X \mid C)}{H(X)},$$

🚨 **VAEs are semantically poor**

🚨 **VFM as conditioning is conditioning-dominant**

💡 **Instead of VAEs, use VFMs as representation autoencoders (RAE), directly on the latent space of the DiT!**

Why Riemannian Flow Matching?



✨ **Low effective rank** ✨ **Curved manifold** ✨ **Angular dominance**

SLERP
$$\mathrm{SLERP}(x_0, x_1, t) = \frac{\sin((1-t)\Omega)}{\sin\Omega}\, x_0 + \frac{\sin(t\Omega)}{\sin\Omega}\, x_1, \quad \Omega = d_g(x_0, x_1),$$

RFM loss
$$\mathcal{L}_{\mathrm{RFM}}(\theta) = \mathbb{E}_{t\sim U[0,1],\,(x_0,x_1)\sim\pi}\left[\left\| v_\theta(x_t,t) - u_t(x_t \mid x_1)\right\|_g^2\right],$$

Chord loss (STREAM)
$$\mathcal{L}_{\mathrm{chord}}(\theta) = \mathbb{E}_{t,\,(x_0,x_1)\sim\pi,\,\epsilon}\left[\frac{\|\hat{x}_1 - x_1\|^2}{(1-t)^2}\right].$$

💡 **Instead of vanilla v-pred/v-loss, use x-pred/x-loss (chord loss)**

✨ **Better training dynamics, better generation fidelity**

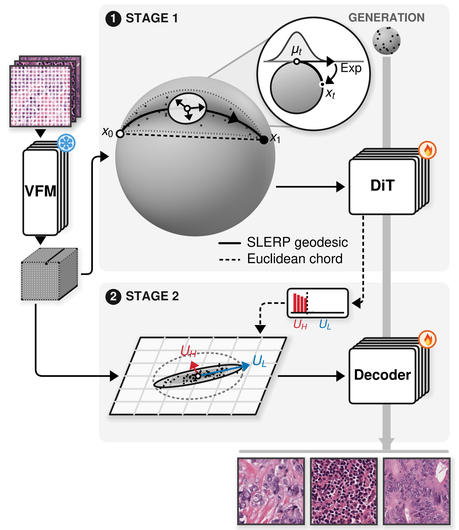
Ablation on loss/prediction configuration for RFM

Model	x-loss/x-pred	v-loss/v-pred	x-loss/v-pred	v-loss/x-pred
rFID↓	**2.23**	4.78	4.73	3.12
gFID↓	**6.16**	7.84	11.26	12.57

Ablation on geometry

Geometry	rFID ↓	gFID ↓
Euclidean	4.52	16.28
Riemannian	**2.23**	**6.16**

STREAM: Methods

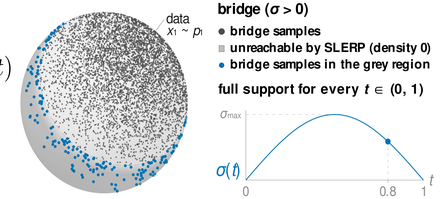


1. Diffusion transformer training with stochastic bridge perturbation

Training-time bridge perturbation
$$\mu_t = \mathrm{SLERP}(x_0, x_1, t),\; x_t = \mathrm{Exp}_{\mu_t}\big(\sigma(t)\,\epsilon\big),\;\; \epsilon \sim \mathcal{N}(0, \Pi_{\mu_t}),\;\; \sigma(t) = \sigma_{\max}\sin(\pi t)$$

Full support: Every point on sphere is reachable mid-flow
$$\mathrm{supp}\big(\mathrm{law}(X_t)\big) = \mathcal{S}^{d-1} \quad \forall\, t \in (0,1)$$

Rectifiability: Paths don't cross, unique solution for every initial condition
$$v_t \in C^\infty\big(\mathcal{S}^{d-1} \times [\epsilon_t, 1-\epsilon_t]\big) \quad \text{and} \quad \dot{x} = v_t(x) \text{ admits a unique solution}$$

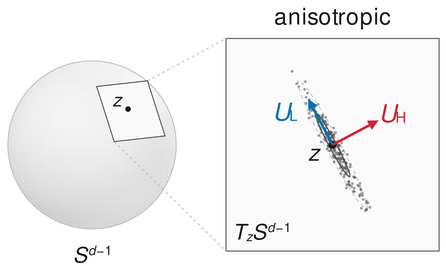


2. Decoder training with spectral-aligned anisotropy

SVD of the velocity-field Jacobian
$$J(z) = \nabla_z v_\theta(z,t) = USV^\top, \qquad U = \big[\,U_H \mid U_L\,\big]\;\; k^* = \min\left\{k : \frac{\sum_{i\le k} s_i^2}{\sum_i s_i^2} \ge \tau\right\}$$

Anisotropic noise injection
$$\Sigma_{\text{noise}}(z) = \sigma_H^2\, U_H U_H^\top + \sigma_L^2\, U_L U_L^\top, \quad \sigma_H \ll \sigma_L,$$
$$\tilde{z} = \frac{z+n}{\|z+n\|}, \quad n = \Pi_z \eta, \quad \eta \sim \mathcal{N}\big(0,\; \Sigma_{\text{noise}}(\bar{z}_c) + \sigma_L^2\, \bar{z}_c \bar{z}_c^\top\big)$$

Decoder training loss
$$\mathcal{L}_{\text{dec}} = \lambda_{\ell_1}\|\hat{I} - I\|_1 + \lambda_{\text{LPIPS}}\mathrm{LPIPS}(\hat{I}, I) + \lambda_{\cos}(1 - \langle z_{\text{rt}}, z\rangle) + \omega_G \lambda_{\text{adapt}} \mathcal{L}_{\text{adv}},$$



Results and Ablations

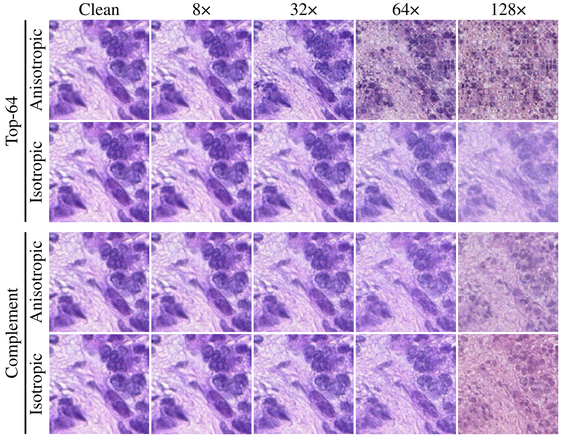
Main result on TCGA-BRCA

Task	Method	FID ↓	KID ↓	CMMD ↓	FvD ↓	KvD ↓	vMMD ↓
Reconstruction	ZoomLDM	4.88	0.0040	0.2685	37.26	0.038	4.12
	PixCell	2.88	0.0025	0.2590	18.11	0.018	2.04
	RAE	6.60	0.0058	1.0225	30.33	0.059	6.09
	SVG	**2.16**	**0.0017**	0.8534	16.68	0.020	2.21
	REPA-E	26.83	0.0211	1.4417	262.65	0.468	51.37
	STREAM	2.23	0.0019	**0.1509**	**9.91**	**0.008**	**0.98**
Generation	ZoomLDM	7.43	0.0058	0.2611	196.41	0.194	21.01
	PixCell	104.18	0.0984	0.6147	1298.00	1.410	169.38
	RAE	15.84	0.0126	1.0960	103.14	0.172	17.26
	SVG	17.37	0.0149	0.8895	123.04	0.185	19.90
	REPA-E	28.57	0.0224	1.4155	309.39	0.549	60.14
	STREAM	**6.16**	**0.0041**	**0.1282**	**78.04**	**0.048**	**5.16**

Main result on TCGA-COADREAD

Task	Method	FID ↓	KID ↓	CMMD ↓	FvD ↓	KvD ↓	vMMD ↓
Reconstruction	ZoomLDM	4.12	0.0033	0.1068	29.5	0.024	4.12
	PixCell	2.58	0.0021	0.0796	**15.87**	**0.012**	**2.58**
	RAE	6.31	0.005	0.0872	35.31	0.056	6.31
	SVG	**1.81**	**0.0012**	**0.0304**	21.37	0.028	1.81
	REPA-E	13.58	0.0096	0.2654	324.93	0.554	13.58
	STREAM	5.64	0.0043	0.0407	32.08	0.03	5.64
Generation	ZoomLDM	8.09	0.0053	0.0613	139.63	0.097	9.52
	PixCell	127.24	0.1132	0.957	1742.85	2.145	218.7
	RAE	19.64	0.0144	0.1605	**91.74**	0.129	11.51
	SVG	23.64	0.0189	0.1042	152.29	0.209	18.52
	REPA-E	13.41	0.009	0.2362	345.29	0.568	51.62
	STREAM	**7.68**	**0.0042**	**0.0293**	91.96	**0.055**	**5.1**

Visual effect of the anisotropic decoder



➕ **Superadditivity**

Bridge	Decoder	rFID ↓	gFID ↓
✗	Iso.	6.51	9.07
✗	Aniso.	5.02	8.27
✓	Iso.	6.51	9.37
✓	**Aniso.**	**3.52**	**6.86**

🧭 **Aniso. favors important directions**

R_{dec} ($L_{UH}^{dec}/L_{UL}^{dec}$)	Iso.	Aniso.
STREAM (k=64)	0.42	**3.41**
STREAM (k*=236)	1.12	**2.31**
No bridge (k=64)	0.62	**2.50**
No bridge (k*=232)	1.28	**2.01**

🎀 **Cutoff matters**

τ	rFID ↓	gFID ↓
0.9	**3.52**	**6.86**
0.8	5.81	8.80

🚨 **Doesn't work on OOD DINOv2 VFM**

Decoder	rFID ↓	gFID ↓
Iso.	7.67	11.05
Aniso.	7.71	11.22

