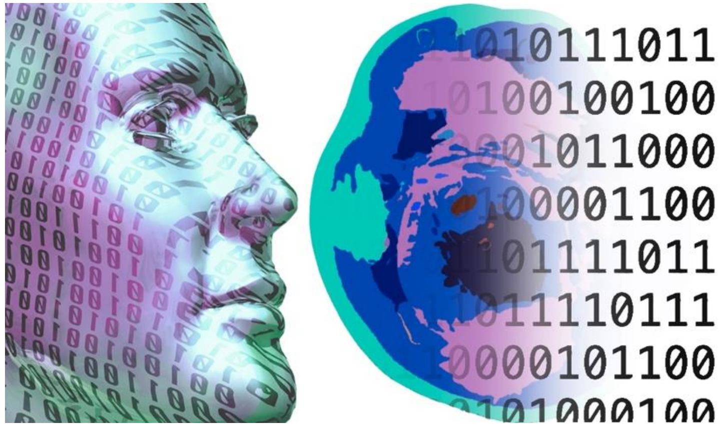
ImageFlowNet

Smita Krishnaswamy



ImageFlowNet: Forecasting Multiscale Image-Level Trajectories of Disease Progression with Irregularly-Sampled Longitudinal Medical Images

Chen Liu[1*] **Ke Xu**[1*] **Liangbo L. Shen**[2] **Guillaume Huguet**[3,4] **Zilong Wang**[3,5]
Alexander Tong[3,4§] **Danilo Bzdok**[3,5§] **Jay Stewart**[2§] **Jay C. Wang**[1,2,6§]
Lucian V. Del Priore[1§] **Smita Krishnaswamy**[1§]

[1]Yale University [2]University of California, San Francisco [3]Mila - Quebec AI Institute
[4]Université de Montréal [5]McGill University [6]Northern California Retina Vitreous Associates

* These authors are joint first authors: `{chen.liu.cl2482, k.xu}@yale.edu`. § Senior authors.
Please direct correspondence to: `smita.krishnaswamy@yale.edu` or `lucian.delpriore@yale.edu`.

ImageFlowNet

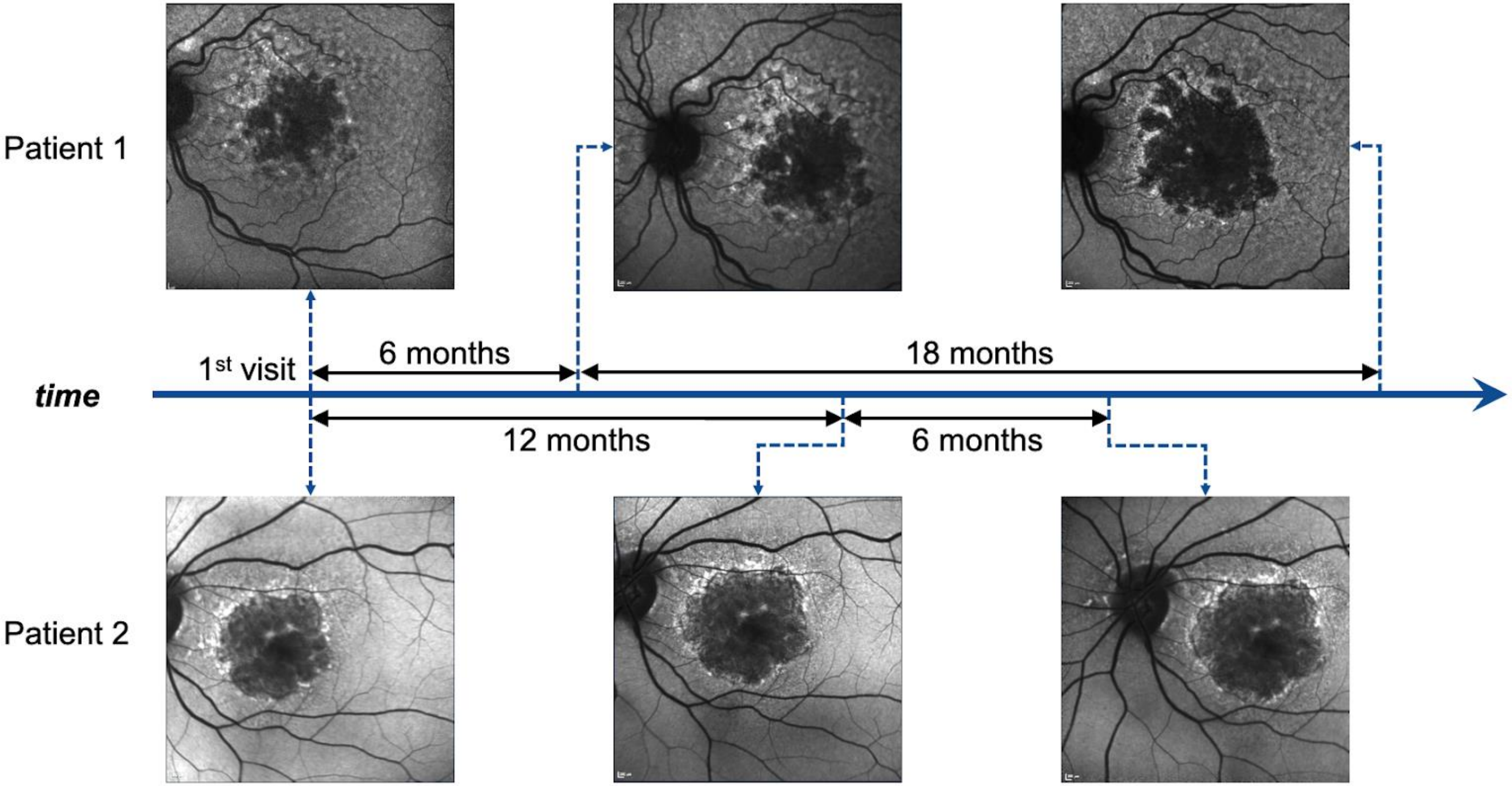
- Motivation
- Preliminaries
- Methods
  - ImageFlowNet
- Experiments & Results
  - Theoretical Results
  - Empirical Results

ImageFlowNet

- ❑ Motivation
- ❑ Preliminaries
- ❑ Methods
  - ❑ ImageFlowNet
- ❑ Experiments & Results
  - ❑ Theoretical Results
  - ❑ Empirical Results

ImageFlowNet

Longitudinal Medical Images: repeated scanning of the same patient over time



Temporal Sparsity

Unlike videos (many frames per second), these longitudinal images are separated by weeks, months or years.

Sampling Irregularity

Irregularly sampled over time for the same patient, and different sampling schedules among patients.

Spatial Misalignment

Almost never spatially aligned.

ImageFlowNet

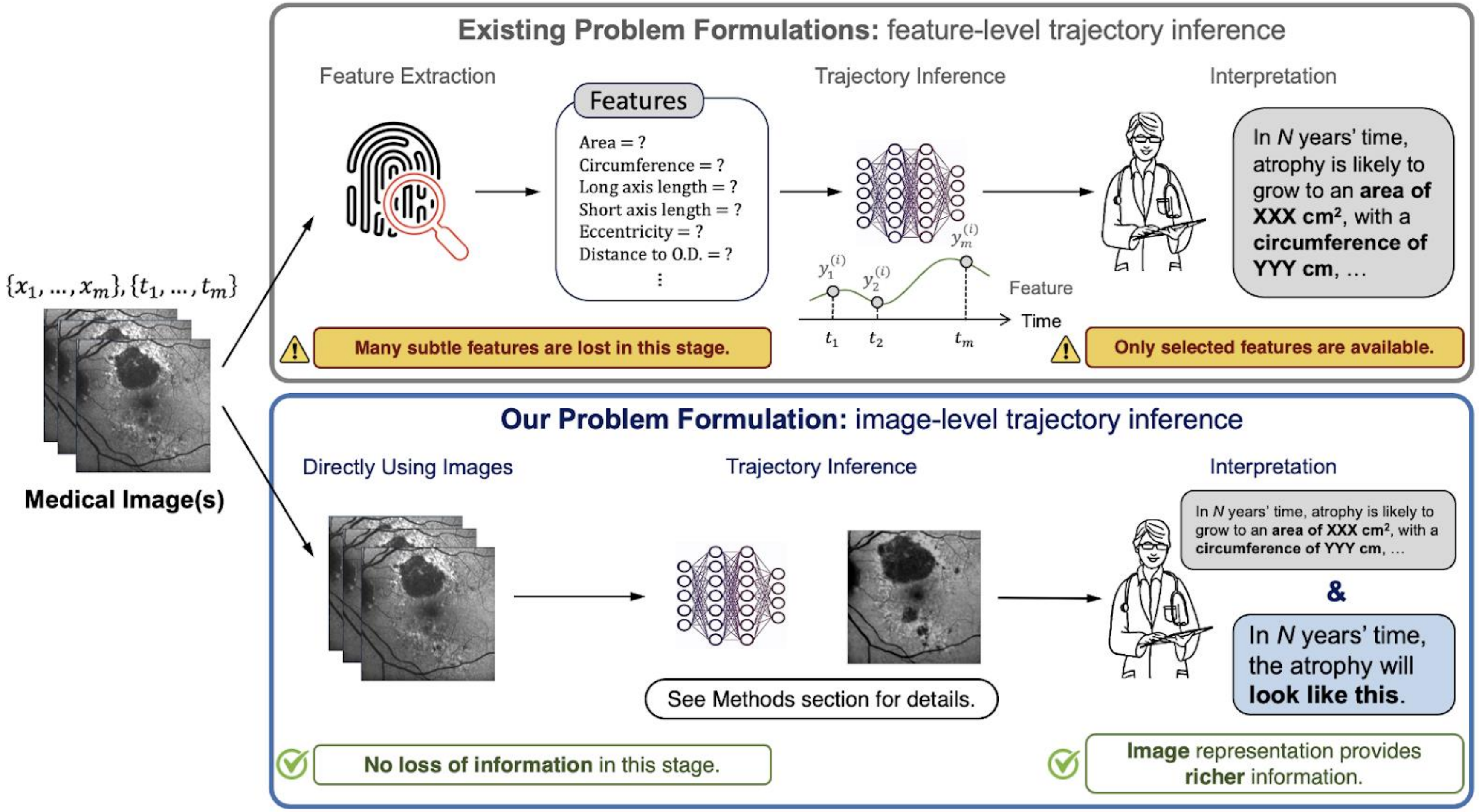


Fig. 1. Advantages of image-level trajectory inference.

ImageFlowNet

- Motivation
- Preliminaries
- Methods
  - ImageFlowNet
- Experiments & Results
  - Theoretical Results
  - Empirical Results

ImageFlowNet

Neural ODE

$$\frac{\mathrm{d}y(\tau)}{\mathrm{d}\tau} = f_\theta(y(\tau), \tau) \qquad (1a)$$

$$y(t_1) = y(t_0) + \int_{t_0}^{t_1} f_\theta(y(\tau), \tau)\mathrm{d}\tau \qquad (1b)$$

Parameterize the continuous dynamics of hidden units using an ordinary differential equation (ODE) specified by a neural network

$$y(t_1) = \mathrm{ODESolve}(f(y(t), t, \theta), y(t_0), t_0, t_1)$$

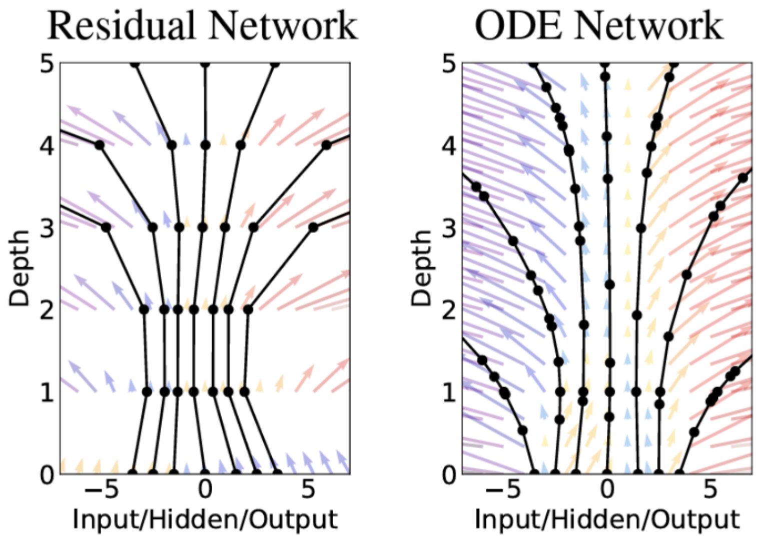


Figure 1: *Left:* A Residual network defines a discrete sequence of finite transformations. *Right:* A ODE network defines a vector field, which continuously transforms the state. *Both:* Circles represent evaluation locations.

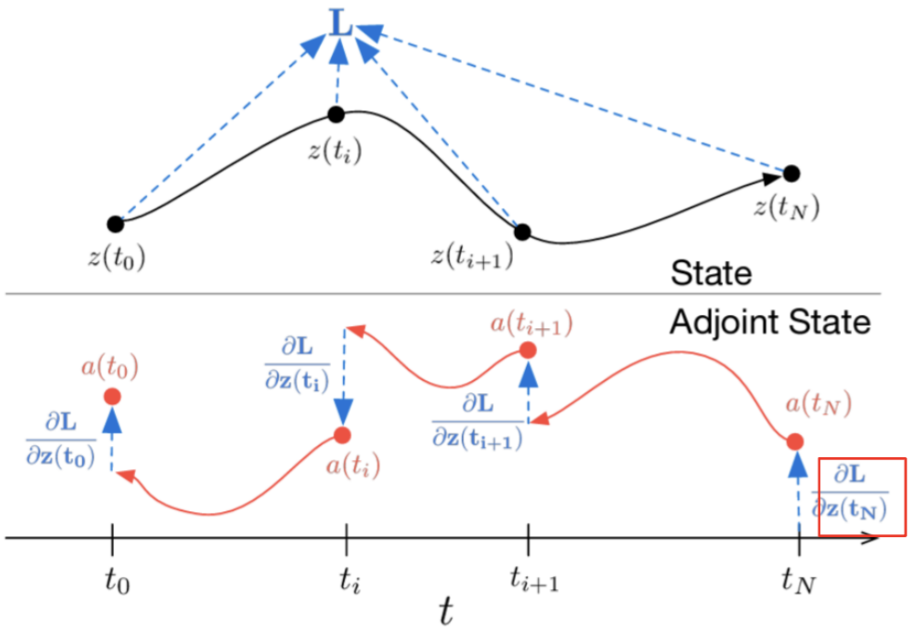
Chen, Ricky TQ, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud. "Neural ordinary differential equations." *Advances in neural information processing systems* 31 (2018).

ImageFlowNet

Neural ODE

$$L(\mathbf{z}(t_1)) = L\left(\mathbf{z}(t_0) + \int_{t_0}^{t_1} f(\mathbf{z}(t), t, \theta)dt\right) = L\left(\text{ODESolve}(\mathbf{z}(t_0), f, t_0, t_1, \theta)\right)$$

adjoint sensitivity method



$$\frac{d\mathbf{a}(t)}{dt} = -\mathbf{a}(t)^{\mathsf{T}} \frac{\partial f(\mathbf{z}(t), t, \theta)}{\partial \mathbf{z}}$$

$$\frac{dL}{d\theta} = -\int_{t_1}^{t_0} \mathbf{a}(t)^{\mathsf{T}} \frac{\partial f(\mathbf{z}(t), t, \theta)}{\partial \theta} dt$$

Chen, Ricky TQ, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud. "Neural ordinary differential equations." *Advances in neural information processing systems* 31 (2018).

ImageFlowNet

Neural SDE

$$\mathrm{d}X_t = f(t, X_t)\,\mathrm{d}t + g(t, X_t) \circ \mathrm{d}W_t,$$

$f(t, X_t)\,\mathrm{d}t$ – deterministic term

$g(t, X_t) \circ \mathrm{d}W_t$ – stochastic term

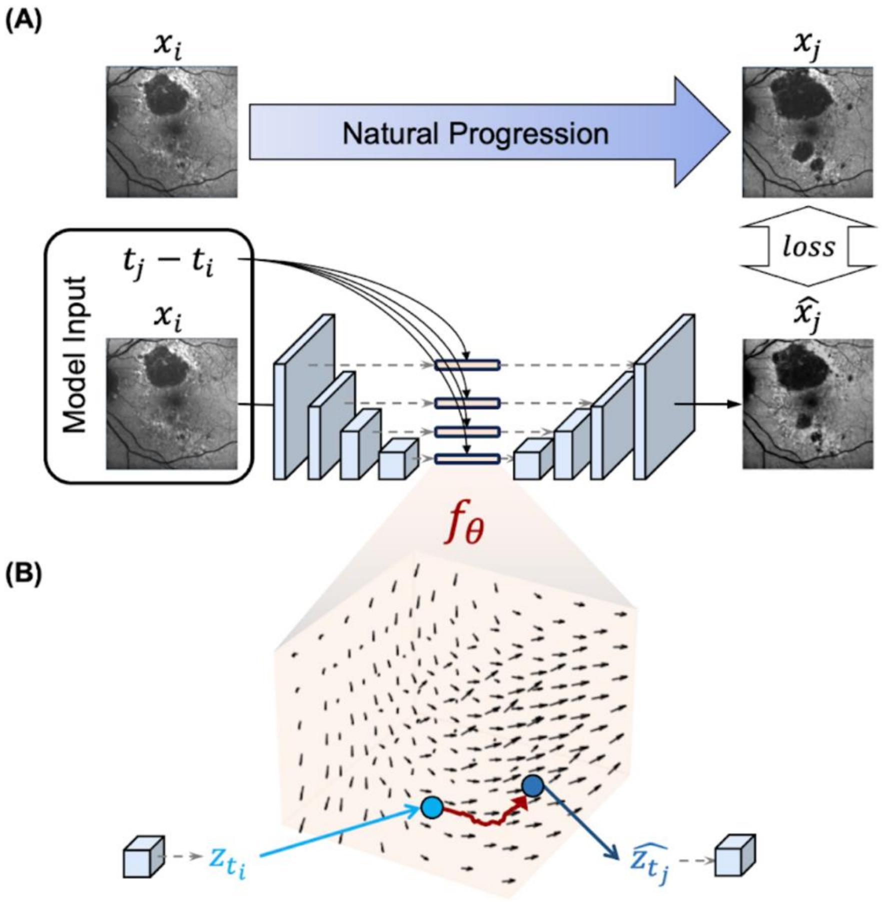
Kidger, Patrick, James Foster, Xuechen Li, and Terry J. Lyons. "Neural sdes as infinite-dimensional gans." In *International conference on machine learning*, pp. 5453-5463. PMLR, 2021.

ImageFlowNet

- Motivation
- Preliminaries
- Methods
  - ImageFlowNet
- Experiments & Results
  - Theoretical Results
  - Empirical Results

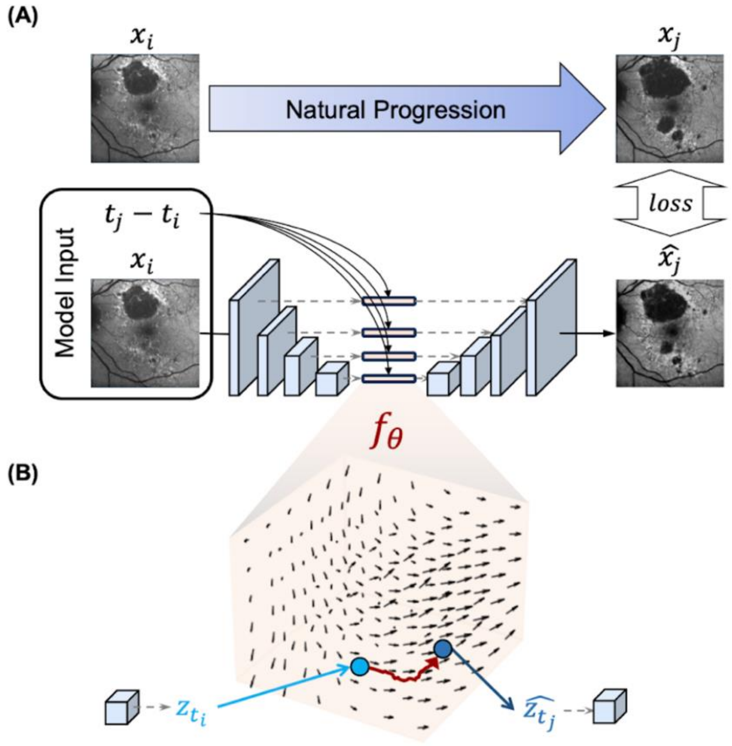
ImageFlowNet

Methods (1/3): ImageFlowNet predicts future image from earlier image and time gap, by learning a vector field of latent features.



ImageFlowNet

Methods (2/3): The latent features are flowed with ODE or SDE.

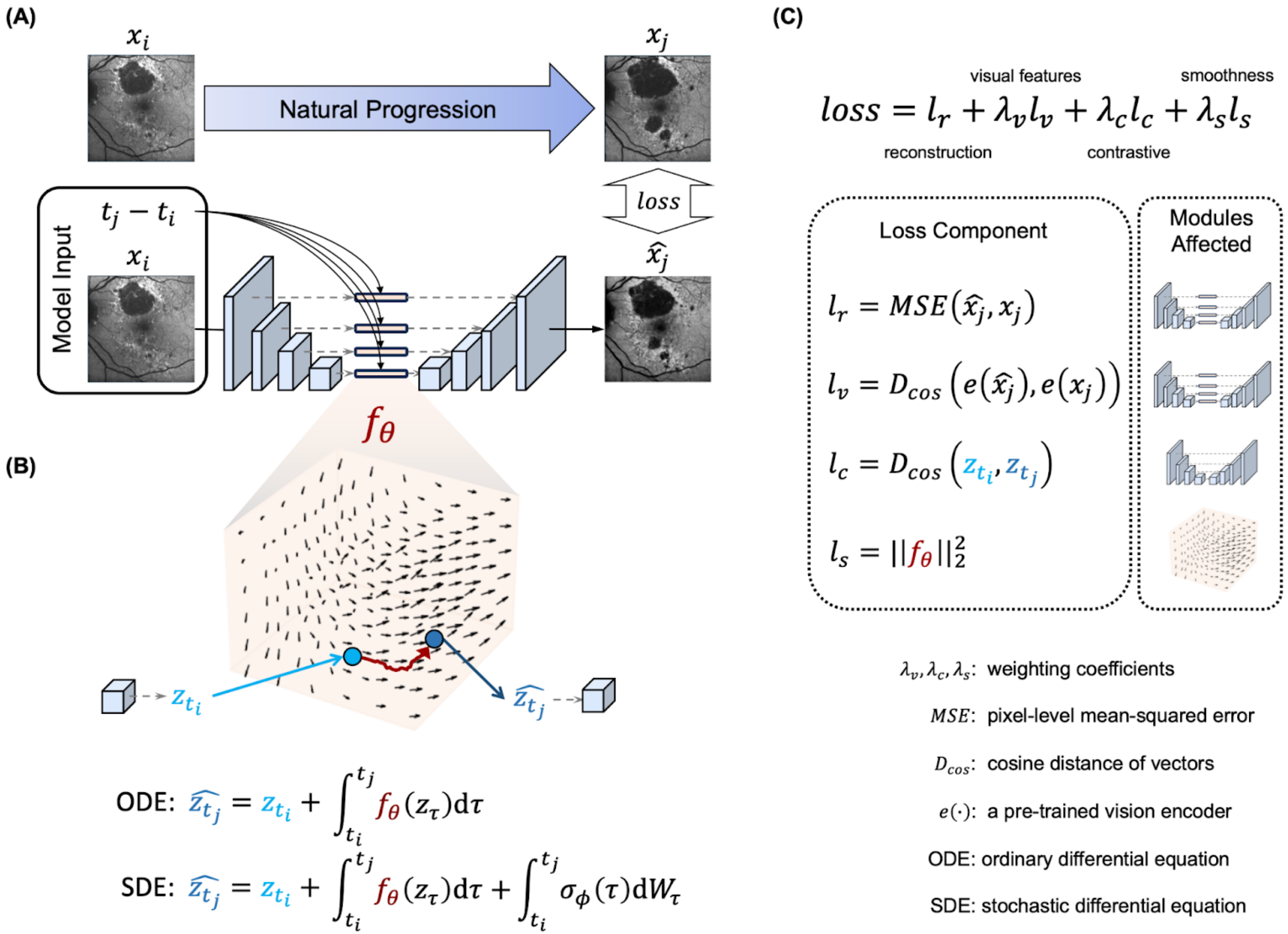


ODE: $\widehat{z_{t_j}} = z_{t_i} + \int_{t_i}^{t_j} f_\theta(z_\tau)\mathrm{d}\tau$

SDE: $\widehat{z_{t_j}} = z_{t_i} + \int_{t_i}^{t_j} f_\theta(z_\tau)\mathrm{d}\tau + \int_{t_i}^{t_j} \sigma_\phi(\tau)\mathrm{d}W_\tau$

ImageFlowNet

Methods (3/3): Loss components affect different modules.



ImageFlowNet

Loss components

$$loss = \overbrace{\frac{1}{HWC}\sum_{h\in H}\sum_{w\in W}\sum_{c\in C}||\widehat{x_j}[h,w,c]-x_j[h,w,c]||_2^2}^{\text{(1) } l_r\text{: reconstruction}} + \overbrace{\lambda_v\left(-\frac{e(\widehat{x_j})^\top e(x_j)}{||e(\widehat{x_j})||_2||e(x_j)||_2}\right)}^{\text{(2) } l_v\text{: visual feature}}$$

$$+ \overbrace{\lambda_c\left(-\frac{p_d(p_j(z_{t_i}))^\top p_j(z_{t_j})}{2||p_d(p_j(z_{t_i}))||_2||p_j(z_{t_j})||_2}-\frac{p_d(p_j(z_{t_j}))^\top p_j(z_{t_i})}{2||p_d(p_j(z_{t_j}))||_2||p_j(z_{t_i})||_2}\right)}^{\text{(3) } l_c\text{: contrastive learning (SimSiam)}} + \overbrace{\lambda_s||f_\theta||_2^2}^{\text{(4) } l_s\text{: trajectory smoothness}}$$

(1) *Image reconstruction objective* is achieved by a MSE loss, attending to low-level features on the pixel level.

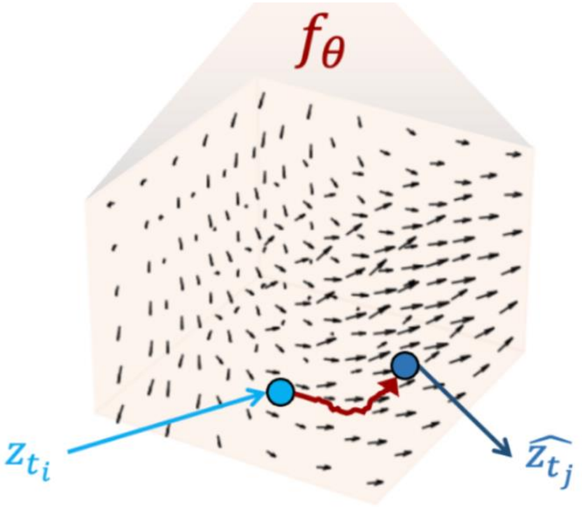
(2) *Visual feature regularization* guides the network to produce images that resemble the ground truth on high-level features judged by an encoder pretrained on ImageNet [15].

(3) *Contrastive learning regularization* organizes a well-structured ImageFlowNet latent space, by encouraging proximity of representations from images within the same longitudinal series, following the SimSiam formulation [16].

(4) *Trajectory smoothness regularization* leverages a theorem in convex optimization (Lemma 2.2 in [17]) to enforce smoothness of trajectories by regularizing the norm of the field. *Notably, this achieves Lipschitz continuity, satisfying a crucial assumption for our theoretical results.*

ImageFlowNet

Flowed latent features are collected hierarchically to form an image.



$$\frac{\mathrm{d}z_\tau^{(b)}}{\mathrm{d}\tau} = f_\theta^{(b)}(z_\tau^{(b)}) \quad \text{for } b \in [1, B] \quad (3a) \qquad z_{t_j}^{(b)} = z_{t_i}^{(b)} + \int_{t_i}^{t_j} f_\theta^{(b)}(z_\tau^{(b)})\mathrm{d}\tau \quad \text{for } b \in [1, B] \quad (3b)$$

$$\begin{aligned} \widehat{x_j} &= \text{ResBlock}(\text{Concat}(\tilde{z}_{t_j}^{(2)}, z_{t_j}^{(1)})), \text{ where} \\ \tilde{z}_{t_j}^{(b)} &= \text{Upsample}(\text{ResBlock}(\text{Concat}(\tilde{z}_{t_j}^{(b+1)}, z_{t_j}^{(b)}))) \text{ for } b \in [2, B-1], \text{ with } \tilde{z}_{t_j}^{(B)} = z_{t_j}^{(B)} \end{aligned} \quad (4)$$

ImageFlowNet

NOTE: We used a novel ODE formulation, which we call a *position-parameterized* ODE.

$$\frac{\mathrm{d}y(\tau)}{\mathrm{d}\tau} = f_\theta(y(\tau), \tau)$$

Standard ODE

$$\frac{\mathrm{d}z_\tau^{(b)}}{\mathrm{d}\tau} = f_\theta^{(b)}(z_\tau^{(b)})$$

Our ODE

ImageFlowNet

Change of variable to make the comparison more obvious.

$$\frac{\mathrm{d}z_{\tau}^{(b)}}{\mathrm{d}\tau} = f_{\theta}^{(b)}(z_{\tau}^{(b)}, \tau)$$

Standard ODE

$$\frac{\mathrm{d}z_{\tau}^{(b)}}{\mathrm{d}\tau} = f_{\theta}^{(b)}(z_{\tau}^{(b)})$$

Our ODE

ImageFlowNet

Why the position-parameterized ODE.

A bigger field to learn

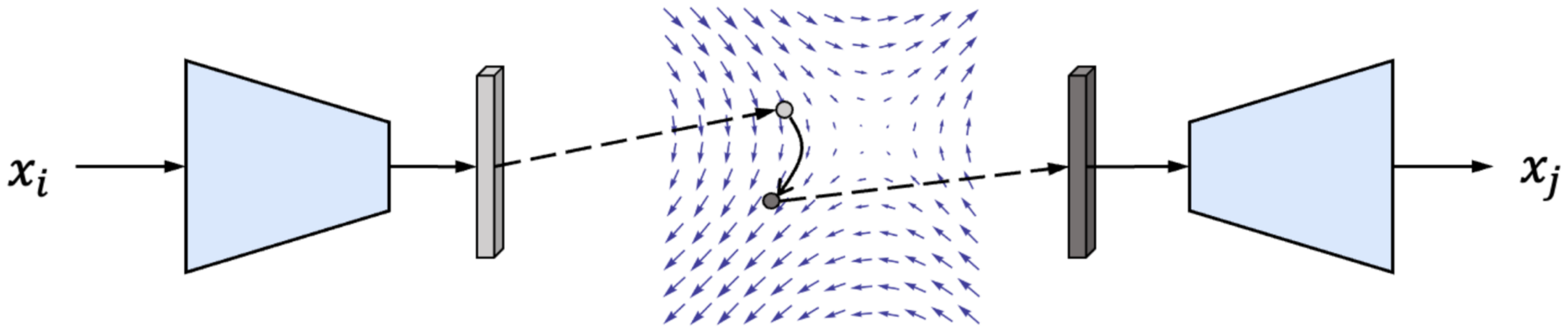
$$\frac{\mathrm{d}z_\tau^{(b)}}{\mathrm{d}\tau} = f_\theta^{(b)}(z_\tau^{(b)}, \tau)$$

Parameterization by “time”

A more compact field

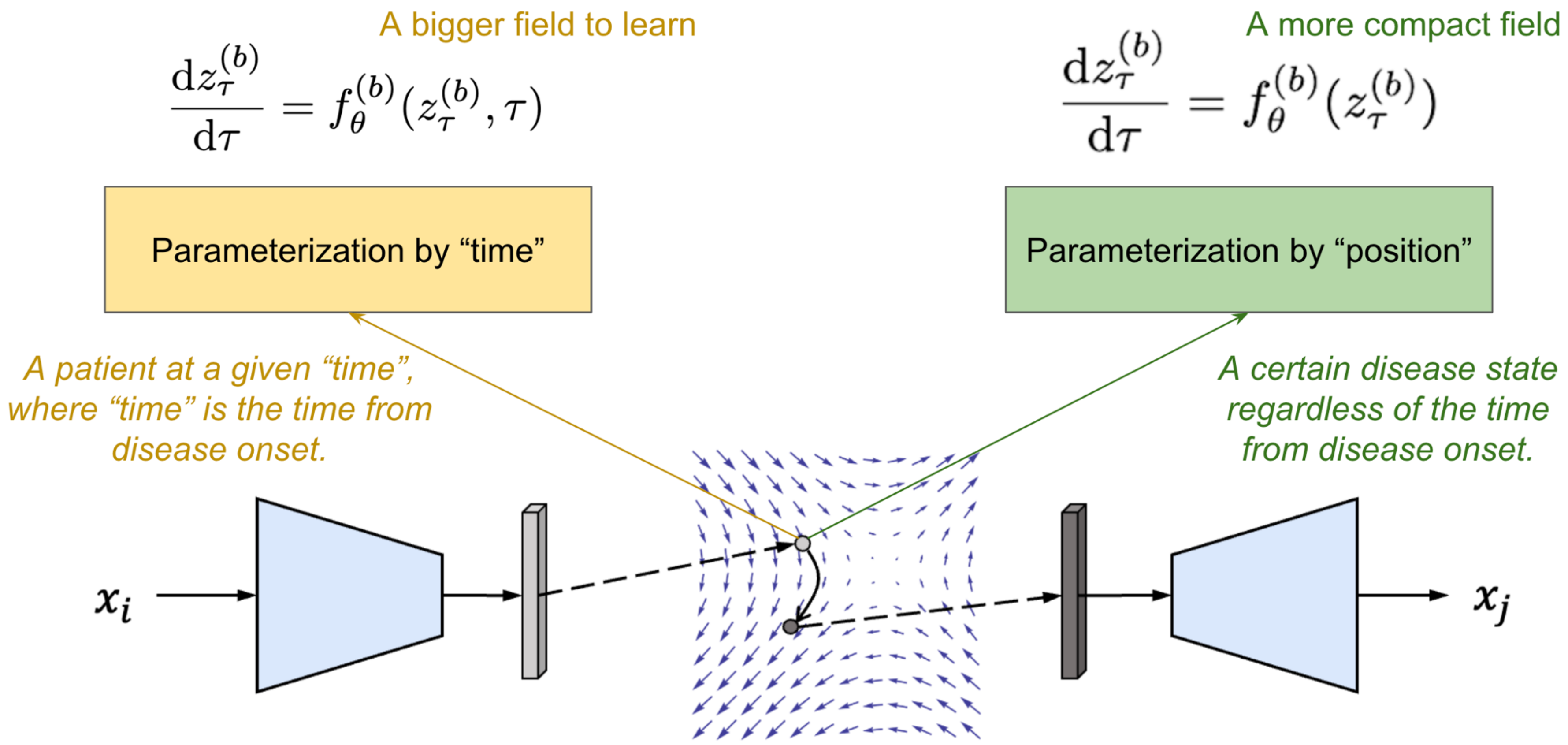
$$\frac{\mathrm{d}z_\tau^{(b)}}{\mathrm{d}\tau} = f_\theta^{(b)}(z_\tau^{(b)})$$

Parameterization by “position”



ImageFlowNet

Why the position-parameterized ODE.



ImageFlowNet

- ❑ Motivation
- ❑ Preliminaries
- ❑ Methods
  - ❑ ImageFlowNet
- ❑ Experiments & Results
  - ❑ Theoretical Results
  - ❑ Empirical Results

ImageFlowNet

Theoretical Results (1/2)

Equivalent Expressiveness of our ODE and standard ODE.

Proposition IV.1. *Let f_θ be a continuous function that satisfies the Lipschitz continuity and linear growth conditions. Also, let the initial state $y(t_0) = y_0$ satisfy the finite second moment requirement. Suppose $z(t_0)$ is the latent representation learned by ImageFlowNet in the initial state corresponding to t_0. Then, our neural ODEs are at least as expressive as the original neural ODEs, and their solutions capture the same dynamics.*

ImageFlowNet

Theoretical Results (2/2)

Connection between ImageFlowNet and dynamic optimal transport.

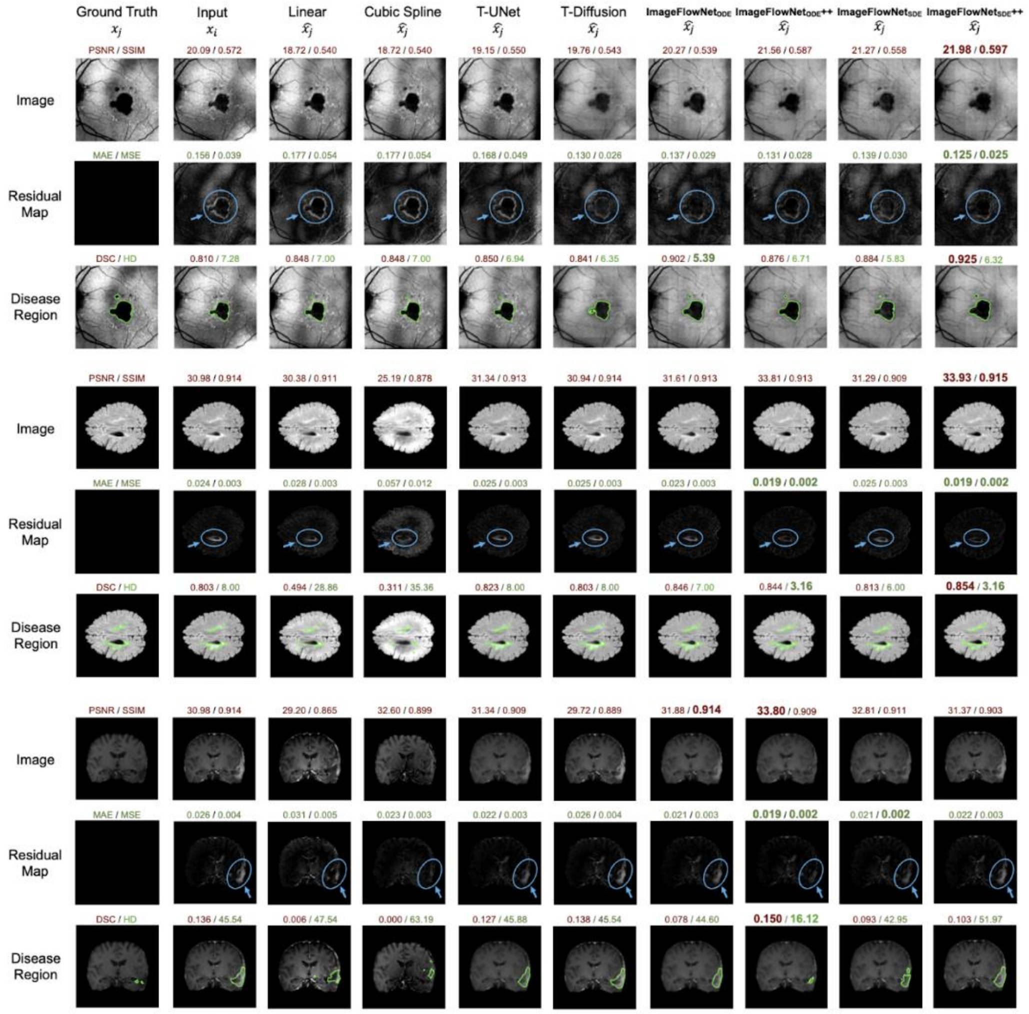
Proposition IV.2. *If we consider an image as a distribution over a 2D grid, ImageFlowNet is equivalently solving a dynamic optimal transport problem, as it meets 3 essential criteria: (1) matching the density, (2) smoothing the dynamics, and (3) minimizing the transport cost, where the ground distance is the Euclidean distance in the latent joint embedding space.*

ImageFlowNet

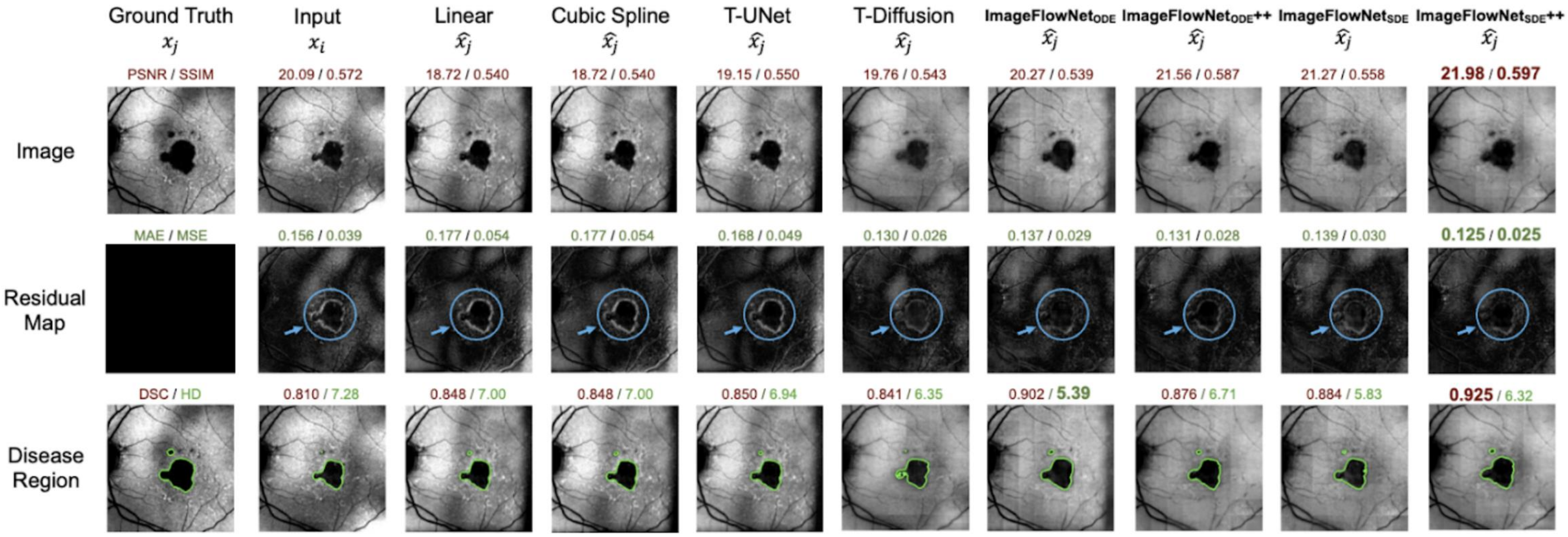
Empirical Results (1/3): Future Image Forecasting Performance

Datasets:

1. Retinal geographic atrophy
   a. 2-5 years
   b. <24 month gap
2. Brain multiple sclerosis
   a. ~5 years
   b. ~4.4 time points per person
3. Brain glioblastoma
   a. <5 years
   b. 2-18 time points per person



ImageFlowNet



ImageFlowNet

Table 1: Image forecasting performance: $\text{metric}(x_j, \widehat{x_j})$. $\widehat{x_j} = \mathcal{F}(x_i, t_i, t_j), \forall i < j$.
†Extrapolation methods use the entire history. "++" means using the 3 regularizations in Eqn (6).

Dataset	Metric	Linear† [24]	Cubic Spline† [25]	T-UNet [33]	T-Diffusion [28]	ImageFlowNet$_{\text{ODE}}$ (ours)	ImageFlowNet$_{\text{ODE}}$++ (ours)	ImageFlowNet$_{\text{SDE}}$ (ours)	ImageFlowNet$_{\text{SDE}}$++ (ours)
Retinal	PSNR ↑	20.22 ± 0.00	19.79 ± 0.00	22.06 ± 0.33	22.29 ± 0.33	22.63 ± 0.26	22.74 ± 0.25	22.32 ± 0.29	**22.89** ± 0.28
Images	SSIM ↑	0.535 ± 0.000	0.505 ± 0.000	0.635 ± 0.015	0.624 ± 0.016	0.646 ± 0.012	0.647 ± 0.013	**0.651** ± 0.015	**0.651** ± 0.012
all	MAE ↓	0.163 ± 0.000	0.177 ± 0.000	0.126 ± 0.005	0.122 ± 0.004	0.119 ± 0.004	0.118 ± 0.004	0.124 ± 0.005	**0.115** ± 0.004
cases	MSE ↓	0.050 ± 0.000	0.060 ± 0.000	0.029 ± 0.002	0.027 ± 0.002	0.024 ± 0.001	0.024 ± 0.001	0.027 ± 0.002	**0.023** ± 0.001
1	DSC ↑	0.833 ± 0.000	0.756 ± 0.000	0.872 ± 0.012	0.867 ± 0.014	0.874 ± 0.012	0.873 ± 0.011	**0.885** ± 0.011	0.883 ± 0.012
	HD ↓	51.64 ± 0.00	54.30 ± 0.00	44.59 ± 4.66	44.41 ± 4.74	**42.68** ± 4.82	47.10 ± 4.89	48.14 ± 4.87	45.14 ± 4.89
minor	PSNR ↑	21.36 ± 0.00	21.08 ± 0.00	22.56 ± 0.55	22.99 ± 0.55	23.23 ± 0.34	23.44 ± 0.33	23.28 ± 0.36	**23.63** ± 0.43
atrophy	SSIM ↑	0.599 ± 0.000	0.586 ± 0.000	0.662 ± 0.023	0.657 ± 0.024	0.682 ± 0.018	0.685 ± 0.018	**0.693** ± 0.018	0.687 ± 0.019
growth	MAE ↓	0.141 ± 0.000	0.147 ± 0.000	0.121 ± 0.007	0.114 ± 0.007	0.110 ± 0.005	0.108 ± 0.004	0.109 ± 0.005	**0.106** ± 0.005
2	MSE ↓	0.038 ± 0.000	0.042 ± 0.000	0.027 ± 0.003	0.024 ± 0.002	0.021 ± 0.002	**0.020** ± 0.002	0.021 ± 0.002	**0.020** ± 0.002
	DSC ↑	0.900 ± 0.000	0.874 ± 0.000	**0.949** ± 0.004	**0.949** ± 0.004	0.936 ± 0.009	0.939 ± 0.007	0.948 ± 0.005	0.948 ± 0.006
	HD ↓	38.15 ± 0.00	41.67 ± 0.00	35.74 ± 5.67	**29.40** ± 4.77	34.59 ± 6.20	39.86 ± 6.40	31.66 ± 5.21	36.98 ± 6.04
major	PSNR ↑	19.02 ± 0.00	18.41 ± 0.00	21.40 ± 0.33	21.68 ± 0.32	21.94 ± 0.34	22.01 ± 0.33	22.01 ± 0.30	**22.10** ± 0.31
atrophy	SSIM ↑	0.468 ± 0.000	0.420 ± 0.000	0.607 ± 0.017	0.588 ± 0.017	0.607 ± 0.014	0.606 ± 0.014	0.607 ± 0.014	**0.613** ± 0.013
growth	MAE ↓	0.186 ± 0.000	0.210 ± 0.000	0.135 ± 0.006	0.131 ± 0.006	0.129 ± 0.006	0.129 ± 0.006	0.128 ± 0.005	**0.126** ± 0.005
3	MSE ↓	0.063 ± 0.000	0.080 ± 0.000	0.032 ± 0.003	0.030 ± 0.002	0.028 ± 0.002	0.028 ± 0.002	**0.027** ± 0.002	**0.027** ± 0.002
	DSC ↑	0.762 ± 0.000	0.631 ± 0.000	0.784 ± 0.016	0.779 ± 0.019	0.807 ± 0.014	0.803 ± 0.012	**0.817** ± 0.016	0.814 ± 0.017
	HD ↓	65.97 ± 0.00	67.73 ± 0.00	61.43 ± 7.26	60.36 ± 7.37	**51.28** ± 7.13	54.79 ± 7.19	65.65 ± 7.17	53.81 ± 7.49
Brain	PSNR ↑	30.07 ± 0.00	29.56 ± 0.00	31.55 ± 0.20	31.57 ± 0.23	32.01 ± 0.19	32.34 ± 0.20	32.40 ± 0.20	**32.41** ± 0.20
MS	SSIM ↑	0.895 ± 0.000	0.888 ± 0.000	0.909 ± 0.003	0.907 ± 0.003	0.914 ± 0.002	**0.915** ± 0.002	0.913 ± 0.002	**0.915** ± 0.002
Images	MAE ↓	0.028 ± 0.000	0.030 ± 0.000	0.024 ± 0.000	0.024 ± 0.001	0.023 ± 0.000	**0.021** ± 0.000	**0.021** ± 0.000	**0.021** ± 0.000
4	MSE ↓	0.004 ± 0.000	0.005 ± 0.000	0.004 ± 0.000	0.004 ± 0.000	**0.003** ± 0.000	**0.003** ± 0.000	**0.003** ± 0.000	**0.003** ± 0.000
	DSC ↑	0.739 ± 0.000	0.682 ± 0.000	0.774 ± 0.007	0.771 ± 0.007	0.775 ± 0.007	**0.777** ± 0.007	**0.777** ± 0.007	0.774 ± 0.007
	HD ↓	22.73 ± 0.00	26.23 ± 0.00	22.00 ± 1.30	**20.91** ± 1.23	22.38 ± 1.28	21.72 ± 1.16	22.21 ± 1.27	21.28 ± 1.27
Brain	PSNR ↑	35.32 ± 0.00	33.60 ± 0.00	35.73 ± 0.13	35.49 ± 0.17	35.86 ± 0.12	**35.90** ± 0.14	35.77 ± 0.12	35.79 ± 0.15
GBM	SSIM ↑	0.929 ± 0.000	0.895 ± 0.000	0.935 ± 0.001	0.940 ± 0.001	0.940 ± 0.001	**0.943** ± 0.001	0.937 ± 0.001	0.939 ± 0.001
Images	MAE ↓	0.017 ± 0.000	0.024 ± 0.000	0.015 ± 0.000	**0.014** ± 0.000	**0.014** ± 0.000	**0.014** ± 0.000	0.015 ± 0.000	0.015 ± 0.000
5	MSE ↓	0.002 ± 0.000	0.005 ± 0.000	**0.001** ± 0.000	0.002 ± 0.000	**0.001** ± 0.000	**0.001** ± 0.000	**0.001** ± 0.000	**0.001** ± 0.000
	DSC ↑	0.300 ± 0.000	0.287 ± 0.000	0.258 ± 0.018	0.253 ± 0.017	**0.302** ± 0.019	0.266 ± 0.018	0.286 ± 0.019	0.287 ± 0.017
	HD ↓	170.44 ± 0.00	**165.62** ± 0.00	195.52 ± 7.69	189.61 ± 7.64	198.19 ± 7.78	185.14 ± 7.69	196.37 ± 7.74	181.66 ± 7.66
1, 4, 5	**Rank** ↓	6.3 ± 1.6	7.3 ± 2.0	4.9 ± 1.4	4.6 ± 1.9	2.9 ± 1.9	2.3 ± 1.6	3.4 ± 2.0	**2.1** ± 1.3
1, 2, 3, 4, 5	**Rank** ↓	6.5 ± 1.3	7.6 ± 1.5	4.9 ± 1.5	4.5 ± 1.8	3.1 ± 1.6	2.7 ± 1.7	3.0 ± 1.8	**2.0** ± 1.2

ImageFlowNet

Empirical Results (2/3): Latent Space Regularization

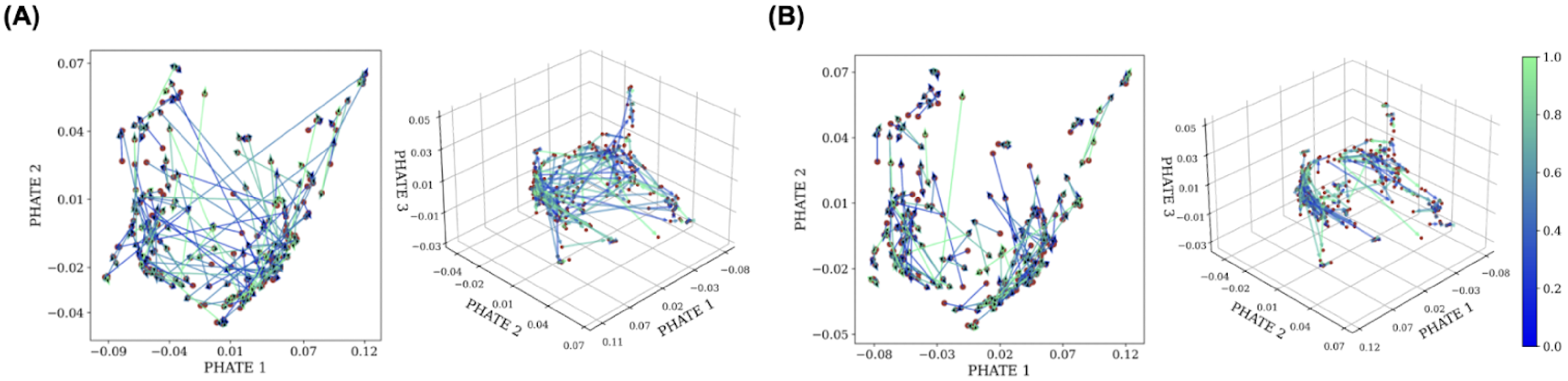


Figure 4: Joint representation space and the effect of contrastive learning regularization. Red dots are the observed disease states and arrows connect adjacent transitions. Normalized time is color coded. (A) Without regularization ($\lambda_c = 0$). (B) With contrastive learning regularization ($\lambda_c = 0.01$).

ImageFlowNet

Empirical Results (3/3): Test-Time Optimization

(Using the entire history to locally fine-tune the vector field)

Table 2: Effect of test-time optimization.

Iterations	Learning Rate	PSNR↑	SSIM↑	MAE↓	MSE↓	DSC↑	HD↓
N/A	N/A	22.31	0.643	0.123	0.027	0.827	51.07
1	10^{-4}	22.52	**0.646**	0.120	**0.025**	**0.829**	**48.97**
1	10^{-5}	22.36	0.643	0.122	0.027	0.827	51.02
1	10^{-6}	22.31	0.643	0.123	0.027	0.827	51.07
10	10^{-4}	20.63	0.605	0.157	0.042	0.749	64.79
10	10^{-5}	<u>22.59</u>	**0.646**	**0.119**	**0.025**	**0.829**	49.92
10	10^{-6}	22.36	0.644	0.122	0.027	0.827	51.01
100	10^{-4}	19.63	0.571	0.177	0.056	0.726	70.12
100	10^{-5}	20.92	0.614	0.152	0.040	0.759	58.76
100	10^{-6}	**22.61**	**0.646**	**0.119**	**0.025**	**0.829**	<u>49.74</u>

ImageFlowNet

Ablation

TABLE II
FLOW FIELD FORMULATION.

	PSNR↑	SSIM↑	MAE↓	MSE↓	DSC↑	HD↓
$f_\theta(z_t, t)$	22.42	0.643	0.123	0.027	0.872	48.38
$f_\theta(z_t)$	**22.63**	**0.646**	**0.119**	**0.024**	**0.874**	**42.68**

TABLE III
LATENT REPRESENTATION.

	PSNR↑	SSIM↑	MAE↓	MSE↓	DSC↑	HD↓
bottleneck only	22.33	0.639	0.122	0.026	0.850	48.13
all unique resolutions	22.49	0.643	0.122	0.025	0.859	43.39
all unique layers	**22.63**	**0.646**	**0.119**	**0.024**	**0.874**	**42.68**

TABLE IV
VISUAL FEATURE REGULARIZATION.

λ_v	PSNR↑	SSIM↑	MAE↓	MSE↓	DSC↑	HD↓
0	22.63	0.646	0.119	**0.024**	**0.874**	**42.68**
0.001	**22.65**	**0.658**	**0.118**	**0.024**	0.872	44.27
0.01	22.64	0.650	0.120	0.025	0.872	45.89
0.1	22.57	0.647	0.120	0.025	0.869	50.69
1	22.54	0.634	0.124	0.027	0.867	48.13

TABLE V
CONTRASTIVE REGULARIZATION.

λ_c	PSNR↑	SSIM↑	MAE↓	MSE↓	DSC↑	HD↓
0	22.63	0.646	0.119	**0.024**	0.874	42.68
0.001	22.63	0.646	0.119	0.025	0.872	46.23
0.01	**22.65**	**0.652**	**0.118**	**0.024**	**0.875**	**42.18**
0.1	22.38	0.651	0.121	0.025	0.871	45.30
1	22.25	0.644	0.121	0.025	0.868	46.85

TABLE VI
SMOOTHNESS REGULARIZATION.

λ_s	PSNR↑	SSIM↑	MAE↓	MSE↓	DSC↑	HD↓
0	22.63	0.646	0.119	**0.024**	0.874	**42.68**
0.001	22.38	0.649	0.123	0.027	0.870	46.91
0.01	22.65	0.648	0.119	0.024	0.870	45.71
0.1	**22.70**	**0.657**	**0.118**	**0.024**	**0.878**	47.44
1	22.69	0.655	**0.118**	**0.024**	0.875	45.16