Spatiotemporal Modeling of Tumor Dynamics via Optimal Transport for Enhanced Longitudinal Analysis

Jiannan Yang[1,2(✉)], Jue Jiang[2], Tengfei Ma[1], and Harini Veeraraghavan[2]

[1] Department of Computer Science, Stony Brook University, Stony Brook, NY, USA
[2] Memorial Sloan Kettering Cancer Center, New York, NY, USA
jiannan.yang@stonybrook.edu

Abstract. Data-driven longitudinal monitoring of tumor treatment response from computed tomography (CT) is challenged by temporally sparse and irregular follow-ups between patients. Naïve interpolation or unconstrained generative augmentation to fill missing scans often produces anatomically and temporally unrealistic tumor evolution, reducing realism and utility for tasks that reason about response trajectories. We propose *OxFlow*, a framework combining geometry conditioning via optimal transport (OT) mask interpolation and flow matching to synthesize smoothly varying tumor morphology. Our approach models tumor shape evolution as Wasserstein geodesic interpolation in mask space, yielding smooth and geometrically consistent intermediate shapes. These masks are then used to condition a flow matching model to produce anatomically coherent CT images. We evaluated the proposed framework on a longitudinal CT cohort of patients with lung cancer. Quantitative evaluations confirmed that our approach produced geometrically smooth tumor trajectories and highly consistent anatomical images. Furthermore, we observed that utilizing these synthesized sequences of temporal scans enhanced model accuracy for predicting immunotherapy response compared to standard augmentation techniques. Overall, this work highlights the value of incorporating geometric constraints into generative modeling for longitudinal medical imaging and suggests a general strategy for geometry-aware medical image synthesis.

Keywords: Longitudinal CT · Optimal Transport · Flow Matching

1 Introduction

Longitudinal CT imaging is routinely used in oncology to monitor tumor progression and treatment response. Discriminative models that leverage serial follow-up scans have demonstrated the prognostic value of multi-timepoint data, achieving strong performance in predicting overall survival [25], durable immunotherapy benefit [21], and treatment response through joint segmentation-prediction frameworks [9]. Recent works on longitudinal tumor evolution further underscore the clinical importance of modeling temporal tumor dynamics [22, 19].

However, the full potential of these data-driven approaches is often limited because both the number of follow-up scans and the intervals between scans vary widely across patients, thereby resulting in sparse and irregularly sampled serial scans, leaving large temporal gaps in observed tumor trajectories. Consequently, training time-series model with longitudinal medical images is challenging. A natural strategy to address this scarcity is generative data augmentation. Diffusion-based synthesis [6] has shown promise for enriching medical training sets, with methods such as DiffBoost [28] and mask-conditioned generation [3, 5], diffusion-based augmentation for downstream classifiers [10] producing realistic images with tumors. However, tumors are synthesized without considering the temporal evolution of tumor morphology from prior time points, subsequently diminishing the utility of the augmented scans for longitudinal analysis.

Several recent studies introduce temporal awareness into generative modeling of disease progression. These include sequence-aware and temporally-conditioned diffusion models [26, 13], treatment-informed synthesis [15], image-level trajectory forecasting [14], and latent diffusion or flow-matching models of patient-specific dynamics [2, 8]. Nevertheless, these methods operate directly at the image or latent level: they learn to predict pixel intensities conditioned on time or metadata, without explicitly constraining the *geometric trajectory* along which tumor shape deforms. Because they do not explicitly disentangle geometry from surrounding appearance, they often produce unconstrained and temporally inconsistent geometric evolution.

We posit that realistic longitudinal synthesis requires decoupling shape evolution from appearance generation. For shape, optimal transport (OT) provides a rigorous solution. Unlike heuristic morphing or linear interpolation, OT defines a Wasserstein geodesic path that yields mass-preserving, non-crossing correspondences [17]. While OT has been widely used for non-rigid registration and shape analysis [18, 23], its potential as a hard geometric constraint for generative conditioning remains underexplored [29].

In this work, we propose **OxFlow**, a framework for spatiotemporal modeling of tumor dynamics, combining OT-based mask interpolation with the x-prediction flow matching [11]. We utilize OT to produce geometrically consistent and temporally plausible intermediate masks, characterized by gradual, reduced oscillatory changes in tumor volume and boundaries. These masks then spatially condition the generative model. Motivated by the theoretical links between FM and OT [12, 16, 24], our framework applies OT at two levels: (1) explicitly for mask-space geometric conditioning and (2) implicitly for image synthesis. By unifying the underlying principle while explicitly decoupling geometry from appearance, we enable trajectory quality to benefit downstream utility.

Our main contributions include: (i) A novel framework for longitudinal tumor synthesis that combines OT geodesic mask interpolation with flow-matching image generation, (ii) controlled ablations demonstrating that the downstream gains arise from trajectory quality rather than per-image fidelity, and (iii) validation on a real-world longitudinal CT cohort, showing improved durable clinical benefit prediction for immunotherapy patients.

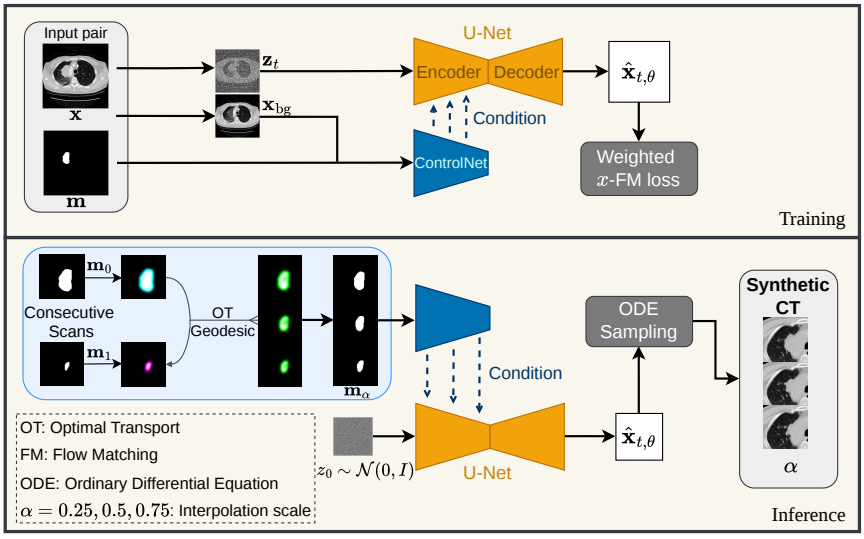

Fig. 1. Overview of OxFlow. Training conditions an x-FM model on tumor masks and background. At inference, OT geodesics interpolate consecutive masks $\mathbf{m}_0, \mathbf{m}_1$ to obtain $\mathbf{m}_\alpha$, which guide ODE sampling to synthesize intermediate CT images.

2 Methodology

2.1 Problem formulation

For each patient, we observe a temporally co-registered longitudinal CT sequence with tumor annotations $\{(\mathbf{X}_i, \mathbf{M}_i)\}_{i=1}^n$, where $\mathbf{X}_i \in \mathbb{R}^{H\times W\times D}$ denotes their i-th CT scan and $\mathbf{M}_i \in \{0,1\}^{H\times W\times D}$ is the corresponding tumor mask. Our goal is to synthesize anatomically plausible intermediate images conditioned on a target mask, enabling realistic modeling of tumor evolution. Synthesized samples should model realistic CT imaging appearance for training a classifier to predict treatment response or track tumor velocity.

2.5D preprocessing. Since the cohort is limited in size, we adopt a 2.5D setting by slicing each 3D volume $\mathbf{X}_i$ along the axial direction into 2D slices $\{\mathbf{x}_j^i\}_{j=1}^{N_i} \subset \mathbb{R}^{H\times W}$. Each slice is paired with its corresponding 2D tumor mask $\mathbf{m}_j^i \in \{0,1\}^{H\times W}$.

2.2 Geometry-aware mask interpolation

OxFlow models tumor evolution in mask space by implementing $\mathcal{F} : \mathbf{m} \mapsto \mathcal{P}$, which converts the binary masks into foreground point clouds using voxel coordinates. Given two point clouds of equal size $\mathcal{P}_0$, $\mathcal{P}_1$ from two adjacent

timepoints, we establish spatial correspondence between them by computing a discrete mass-preserving transport map $T : \mathcal{P}_0 \rightarrow \mathcal{P}_1$ that minimizes the quadratic transport cost:

$$\min_T \sum_{p \in \mathcal{P}_0} \|p - T(p)\|^2. \tag{1}$$

Using the optimized map T, we generate intermediate shapes via displacement interpolation, $\mathcal{P}_\alpha = \{(1-\alpha)p + \alpha T(p) \mid p \in \mathcal{P}_0\}$, $\alpha \in (0,1)$. Under the quadratic cost, this formulation yields a Wasserstein geodesic interpolation with globally consistent, non-crossing trajectories [17]. Inverse mapping $\mathcal{F}^{-1}$ rasterizes and morphologically refines the points $\mathcal{P}_\alpha$ back into intermediate binary masks $\mathbf{m}_\alpha$.

2.3 Geometry-conditioned generative model

We adopt a ControlNet-augmented U-Net [20, 27] to explicitly disentangle geometry from appearance. We construct a masked background to prevent appearance leakage, $\mathbf{x}_{\text{bg}} = \mathbf{x} \odot (1 - \mathbf{m}) + \bar{\mathbf{x}}\mathbf{m}$, where $\bar{\mathbf{x}}$ is the global mean intensity of $\mathbf{x}$. The joint spatial condition $\mathbf{c} = [\mathbf{m}; \mathbf{x}_{\text{bg}}]$ forces the network to synthesize tumor content driven solely by $\mathbf{m}$. The main U-Net processes the noisy state $\mathbf{z}_t$, while the ControlNet branch encodes $\mathbf{c}$ and injects the features via zero-initialized 1×1 convolutions (Fig. 1). Crucially, this preserves the pretrained generative prior while progressively enforcing the geometric constraints of the OT trajectory.

2.4 x-FM: flow matching via x-prediction

To synthesize images from geometric conditions, we deploy x-FM, a flow matching algorithm parameterized via x-prediction [11]. This choice aligns our framework conceptually at both levels: both methods share the principle of finding efficient, straight-line paths. Hence, while OT computes a Wasserstein geodesic in mask space, FM learns a continuous linear probability flow in image space.

Unlike standard velocity regression, the network $\hat{\mathbf{x}}_\theta(\mathbf{z}_t, t, \mathbf{c})$ directly estimates the clean image $\mathbf{x}$ from the noisy state $\mathbf{z}_t = t\mathbf{x} + (1-t)\varepsilon$. This formulation optimizes a mathematically equivalent velocity-matching objective and offers improved empirical stability and generation quality [11]:

$$\mathcal{L} = \mathbb{E}_{t,\epsilon,\mathbf{x}} \left[\frac{1}{(1-t)^2} \left\| \sqrt{w(\mathbf{m})} \odot (\mathbf{x} - \hat{\mathbf{x}}_\theta) \right\|_2^2 \right], \tag{2}$$

where $w(\mathbf{m}) = 1 + \mathbf{m}$ is a spatial re-weighting term that prioritizes synthesis fidelity within the actual tumor regions during training. During inference, conditioned on the OT-interpolated mask $\mathbf{m}_\alpha$ with $\mathbf{x}_{\text{bg}}$ from the smaller-tumor endpoint, we sample initial noise $\mathbf{z}_0 \sim \mathcal{N}(0, I)$, convert the network output to the velocity field $\hat{v}_t = (\hat{\mathbf{x}}_\theta - \mathbf{z}_t)/(1-t)$, and solve the probability-flow ODE. This continuous-time ODE formulation facilitates fast efficient synthesis of intermediate tracking data across the entire patient cohort.

2.5 Synthetic Augmentation and Downstream Task

Synthetic augmentation. To enrich longitudinal training data, we synthesize intermediate tumor states between consecutive timepoints at interpolation factors $\alpha \in (0, 1)$ using OxFlow.
Downstream task setup. We adopt a multiple instance learning (MIL) formulation for durable response prediction. Each patient is a temporal sequence of axial slices, and a ResNet-18 backbone extracts slice-level features aggregated via attention pooling into a patient-level prediction. Patient-level fold splits precede any synthetic augmentation: synthesized intermediate slices are appended only to training patients' sequences, while validation and test sets contain only real slices from held-out patients, precluding cross-validation leakage.

3 Experiments

3.1 Dataset and evaluation protocol

We evaluate a retrospective, IRB-approved longitudinal CT cohort of 229 lung cancer patients with at least two CT scans. All scans were acquired as part of routine clinical care. Scans include expert tumor segmentations and binary durable response labels (defined as progression-free survival > 6 months). After 2.5D axial slicing, the dataset yields 8,491 tumor-containing slices.

Synthesis quality evaluation. To evaluate conditioning consistency, we employ an independent nnU-Net [7] as a fixed anatomical evaluator. For each split, we train it on real tumor-containing axial slices from the training set and select the checkpoint on the validation set. We report the Dice Similarity Coefficient (DSC) between the evaluator-predicted tumor mask and the conditioning mask for synthesized images. For held-out real images, DSC is computed between the evaluator prediction and the corresponding ground-truth tumor mask.

Classification evaluation. Having established synthesis quality, we assess downstream performance using patient-level 10-fold stratified cross-validation. ROC curves are plotted from out-of-fold predictions, with shaded regions indicating 95% confidence intervals. Results are reported as mean AUC over three seeds.

3.2 Implementation details

ControlNet. Images are 256×256 single-channel axial slices. The U-Net backbone uses channel sizes $(64, 128, 256)$ with two residual blocks per level, GroupNorm, and SiLU activations. ControlNet training follows Sec. 2.3.
Flow-matching Training and Inference. We sample $t \sim \sigma(\mathcal{N}(-0.8, 0.8))$ and train for 100 epochs using AdamW ($\text{lr} = 10^{-4}$, batch size 16). An exponential moving average with decay 0.9999 is used for inference. The probability-flow ODE is solved using a second-order Heun solver with 50 steps.
Synthesis and Classification. Intermediate masks are generated at $\alpha = 0.25$,

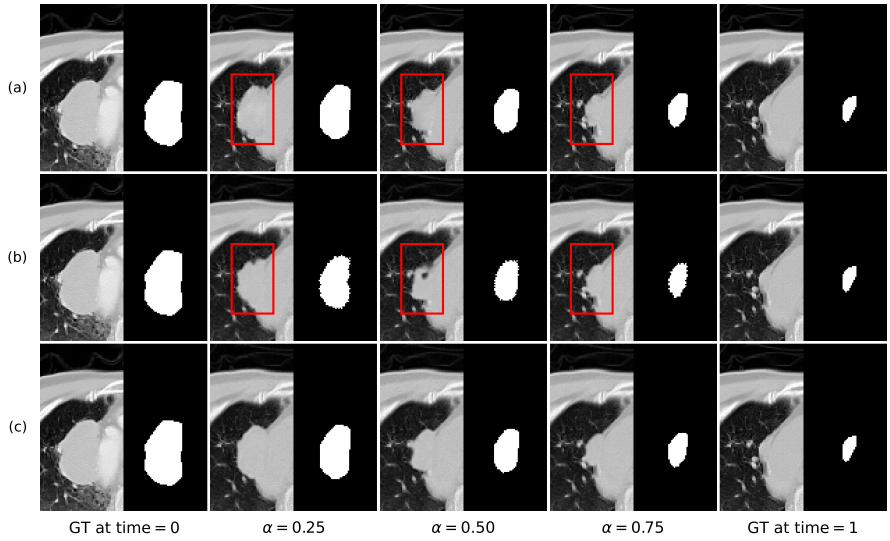

Fig. 2. Qualitative comparison of intermediate tumor synthesis. The first and last columns are consecutive ground truth scans; other columns are synthesized. Rows show: (a) OT mask + DDPM, (b) random matching + x-FM, and (c) OxFlow (ours).

0.5, 0.75, and the MIL classifier is trained for 50 epochs (averaged over 3 seeds). **Baselines.** To facilitate evaluation, we define: (1) *Masking variants:* We compare OT against *linear mask* interpolation (thresholded mask blending) and *random matching* (random point matching between $\mathcal{P}_0$ and $\mathcal{P}_1$). (2) *Generative algorithms:* We contrast x-FM with *DDPM* [6] using identical ControlNet conditioning. (3) *Augmentation schemes:* Downstream prediction is compared across *real-data-only*, *standard augmentation* (affine/elastic), and *pixel interpolation*.

3.3 Main results

Table 1. Trajectory smoothness metrics for mask interpolation methods. $N = 5078$ consecutive timepoint pairs. Median with [IQR] reported, best in **bold**.

	Non-monotonic vol. (%)↓	Perimeter TV ↓	CoM curvature (rad) ↓
Linear mask	92.5	0.47 [0.26, 0.93]	0.59 [0.21, 1.31]
Random matching	77.2	0.15 [0.09, 0.27]	0.28 [0.09, 0.74]
OT Mask	**51.7**	**0.13** [0.07, 0.23]	**0.21** [0.07, 0.51]

Trajectory smoothness. For each pair of consecutive slices, masks are interpolated using OT, linear, and random matching to form 5-point sequences. Motivated

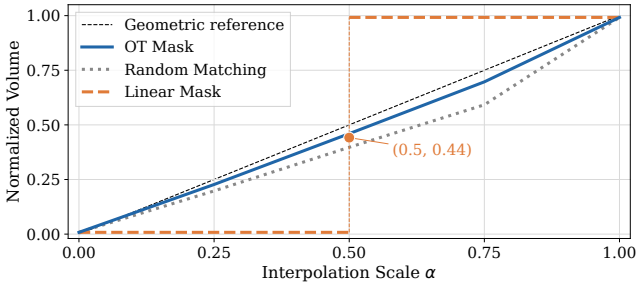

Fig. 3. Average normalized tumor volume progression across interpolation steps α.

by prior works [1, 4] on interpolation stability and trajectory smoothness, we quantify geometric stability using *non-monotonic volume rate*, *perimeter total variation (TV)*, and *center-of-mass (CoM) curvature*. As shown in Table 1, OT interpolation substantially reduces oscillatory behavior compared to linear and random baselines, while also achieving the lowest boundary variation and centroid curvature. This stability is visually corroborated by the volume progression (Fig. 3): OT interpolation yields a smooth, monotonic change in tumor volume between the two observed endpoints, in contrast to the abrupt step-transition of linear masking and the non-monotonic deviation of random matching. Overall, OT mask interpolation yields the most stable intermediate geometries.

Synthesis quality. Table 2 shows OxFlow and linear masking achieve similar DSC accuracies, as static overlap primarily captures spatial alignment rather than temporal coherence. OxFlow, however, produces the lowest failure rate, reflecting the most reliable adherence to the conditioning mask. Noticeable drop with random matching suggests a strong generative backbone alone is insufficient without geometric priors, often producing fragmented tumor boundaries despite realistic textures (Fig. 2). Moreover, replacing FM with DDPM yields the lowest DSC and exhibits boundary discontinuities that compromise anatom-

Table 2. Dice overlap (DSC) between the conditioning mask and nnU-Net segmentation on synthesized images. The failure rate is the ratio of images with DSC < 0.5.

Method	Mean ± SD ↑	Failure Rate (%) ↓
Real (Evaluator Reference)	0.912 ± 0.149	2.60
OT mask + DDPM	0.678 ± 0.294	20.4
Random matching + x-FM	0.711 ± 0.258	14.8
Linear mask + x-FM	0.722 ± 0.273	16.1
OxFlow (Ours)	**0.729 ± 0.264**	**14.3**

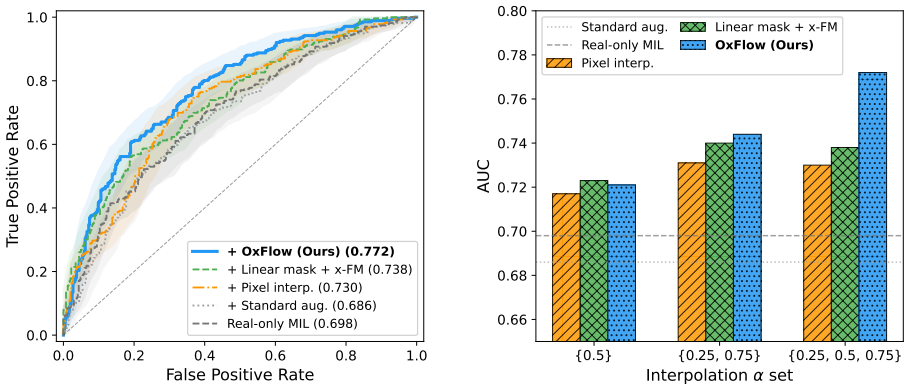

Fig. 4. Left: ROC curves for durable response prediction. Right: AUC as a function of interpolation density.

ical adherence. Collectively, these results support OxFlow's integration of OT trajectories with x-FM.

Classification performance. OxFlow outperforms all baselines in durable response prediction (Fig. 4). In particular, it maintains a consistently balanced and superior performance across a broader range of decision thresholds. This robust ROC profile suggests that the geometrically consistent tumor trajectories generated by OT provide more generalizable regularization for the classifier and may reduce overfitting to specific sensitivity-specificity trade-offs. Given that generative fidelity is comparable between OxFlow and linear mask + x-FM (Table 2), this improvement supports our claim that temporally coherent geometric trajectories are beneficial for downstream clinical utility in such scenarios.

Impact of interpolation density. To further validate the geometric stability of the OT trajectory, we evaluate downstream AUC as a function of interpolation density (Fig. 4, right). Increasing the number of intermediate states (α) consistently improves performance for OxFlow. In contrast, baseline methods plateau or degrade, indicating that denser temporal sampling only provides informative augmentation when the underlying geometric trajectory is intrinsically smooth.

4 Conclusion

We proposed OxFlow, a framework for spatiotemporal modeling of tumor dynamics, coupling optimal-transport mask interpolation with flow-matching generation. Experiments demonstrate our approach produces geometrically smooth, anatomically plausible trajectories that achieve superior downstream durable response prediction. This supports the premise that sequential geometric fidelity enhances clinical utility. Future work will extend this 2.5D, single-center study to multi-center 3D longitudinal cohorts.

Acknowledgments. This research was partially supported by NCI grant R01CA258821, NIBIB grant R01EB032825, and the Memorial Sloan Kettering Cancer Center Support Grant/Core Grant NCI P30CA008748.

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

References

1. Baust, M., Demaret, L., Storath, M., Navab, N., Weinmann, A.: Total variation regularization of shape signals. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2075–2083 (2015). https://doi.org/10.1109/CVPR.2015.7298819
2. Chen, H., Yin, R., Chen, Y., Chen, Q., Li, C.: Learning patient-specific disease dynamics with latent flow matching for longitudinal imaging generation. In: International Conference on Learning Representations (ICLR) (2026)
3. Dorjsembe, Z., Pao, H.K., Odonchimed, S., Xiao, F.: Conditional diffusion models for semantic 3D brain MRI synthesis. IEEE Journal of Biomedical and Health Informatics **28**(7), 4084–4093 (2024). https://doi.org/10.1109/JBHI.2024.3385504
4. Guan, D., Huang, J., Xiao, A., Lu, S.: Domain adaptive video segmentation via temporal consistency regularization. In: IEEE/CVF International Conference on Computer Vision (ICCV). pp. 8033–8044 (2021). https://doi.org/10.1109/ICCV48922.2021.00795
5. Guo, P., Zhao, C., Yang, D., et al.: MAISI: Medical AI for synthetic imaging. In: IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 4430–4441 (2025). https://doi.org/10.1109/WACV61041.2025.00435
6. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems. vol. 33, pp. 6840–6851 (2020)
7. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021)
8. Islam, M.M., Kuipers, T.P., Vadgama, S., de Vente, C., Khan, A., Sánchez, C.I., Bekkers, E.J.: Longitudinal flow matching for trajectory modeling. In: International Conference on Artificial Intelligence and Statistics (AISTATS) (2026)
9. Jin, C., Yu, H., Ke, J., et al.: Predicting treatment response from longitudinal images using multi-task deep learning. Nature Communications **12**(1), 1851 (2021). https://doi.org/10.1038/s41467-021-22188-y
10. Ktena, I., Wiles, O., Albuquerque, I., et al.: Generative models improve fairness of medical classifiers under distribution shifts. Nature Medicine **30**(4), 1166–1173 (2024). https://doi.org/10.1038/s41591-024-02838-6
11. Li, T., He, K.: Back to basics: Let denoising generative models denoise. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 36115–36125 (2026)
12. Lipman, Y., Chen, R.T.Q., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. In: ICLR (2023)
13. Litrico, M., Guarnera, F., Giuffrida, M.V., Ravì, D., Battiato, S.: TADM: Temporally-aware diffusion model for neurodegenerative progression on brain MRI. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. LNCS, vol. 15002, pp. 444–453 (2024). https://doi.org/10.1007/978-3-031-72069-7_42

14. Liu, C., Xu, K., Shen, L.L., et al.: ImageFlowNet: Forecasting multiscale image-level trajectories of disease progression with irregularly-sampled longitudinal medical images. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5 (2025). https://doi.org/10.1109/ICASSP49660.2025.10890535
15. Liu, Q., Fuster-García, E., Hovden, I.T., et al.: Treatment-aware diffusion probabilistic model for longitudinal MRI generation and diffuse glioma growth prediction. IEEE Transactions on Medical Imaging **44**(6), 2449–2462 (2025). https://doi.org/10.1109/TMI.2025.3533038
16. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. In: ICLR (2023)
17. McCann, R.J.: A convexity principle for interacting gases. Advances in Mathematics **128**(1), 153–179 (1997). https://doi.org/10.1006/aima.1997.1634
18. ur Rehman, T., Haber, E., Pryor, G., Melonakos, J., Tannenbaum, A.: 3D nonrigid registration via optimal mass transport on the GPU. Medical Image Analysis **13**(6), 931–940 (2009). https://doi.org/10.1016/j.media.2008.10.008
19. Rokuss, M., et al.: LesionLocator: Zero-shot universal tumor segmentation and tracking in 3D whole-body imaging. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 30872–30885 (2025)
20. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2015. LNCS, vol. 9351, pp. 234–241 (2015). https://doi.org/10.1007/978-3-319-24574-4_28
21. Sako, C., Kurland, B.F., Schmidt, T.G., et al.: Deep-learning serial CT prediction of survival in immunotherapy-treated non–small cell lung cancer. JAMA Network Open **9**(1), e2555759 (2026). https://doi.org/10.1001/jamanetworkopen.2025.55759
22. Sauty, B., Durrleman, S.: Progression models for imaging data with longitudinal variational auto encoders. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. LNCS, vol. 13431, pp. 3–13 (2022). https://doi.org/10.1007/978-3-031-16431-6_1
23. Su, Z., Wang, Y., Shi, R., Zeng, W., Sun, J., Luo, F., Gu, X.: Optimal mass transport for shape matching and comparison. IEEE Transactions on Pattern Analysis and Machine Intelligence **37**(11), 2246–2259 (2015). https://doi.org/10.1109/TPAMI.2015.2408346
24. Tong, A., Fatras, K., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Wolf, G., Bengio, Y.: Improving and generalizing flow-based generative models with minibatch optimal transport. Transactions on Machine Learning Research (2024)
25. Xu, Y., Hosny, A., Zeleznik, R., Parmar, C., Coroller, T., Franco, I., Mak, R.H., Aerts, H.J.W.L.: Deep learning predicts lung cancer treatment response from serial medical imaging. Clinical Cancer Research **25**(11), 3266–3275 (2019). https://doi.org/10.1158/1078-0432.CCR-18-2495
26. Yoon, J.S., Zhang, C., Suk, H.I., Guo, J., Li, X.: SADM: Sequence-aware diffusion model for longitudinal medical image generation. In: Information Processing in Medical Imaging – IPMI 2023. LNCS, vol. 13939, pp. 388–400 (2023). https://doi.org/10.1007/978-3-031-34048-2_30
27. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: IEEE/CVF International Conference on Computer Vision (ICCV). pp. 3836–3847 (2023)

28. Zhang, Z., Yao, L., Wang, B., Jha, D., Durak, G., Keles, E., Medetalibeyoglu, A., Bagci, U.: DiffBoost: Enhancing medical image segmentation via text-guided diffusion model. IEEE Transactions on Medical Imaging **44**(9), 3670–3682 (2025). https://doi.org/10.1109/TMI.2024.3519307
29. Zhu, J., Qiu, J., Guha, A., Yang, Z., Nguyen, X., Li, B., Zhao, D.: Interpolation for robust learning: Data augmentation on Wasserstein geodesics. In: International Conference on Machine Learning (ICML). PMLR, vol. 202, pp. 43129–43157 (2023)