Beyond Stochastic Diffusion: Trajectory-Consistent Deterministic Flow Matching for Fluorescence Molecular Tomography

Qianqian Xue[1], Xingyu Liu[1], Hongming Shan[2], Peng Zhang[1](✉), and Wenjian Wang[1](✉)

[1] School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China
{pengzhang,wjwang}@sxu.edu.cn
[2] Science and Technology for Brain-inspired Intelligence, Fudan University, Shanghai 200433, China

Abstract. Fluorescence molecular tomography (FMT) is a non-invasive optical imaging modality for three-dimensional molecular visualization, yet its reconstruction is fundamentally limited by the severe ill-posedness of the inverse problem, resulting in poor spatial resolution and localization instability. While diffusion models improve detail recovery, their dependence on small-step Gaussian approximations necessitates long stochastic denoising processes, leading to high computational cost and reconstruction instability under few-step sampling. We present TCDFlow-FMT, a fast deterministic flow-matching framework that reformulates FMT reconstruction as continuous-time transport between noise and fluorescent sources distribution. By replacing stochastic denoising with ODE-based deterministic flow, our approach achieves reproducible and efficient reconstruction in only a few steps. We introduce a spatially guided flow Transformer that integrates multi-view fluorescence measurements into velocity field prediction via structure alignment and spatial modulation, resolving the global–local mismatch between diffuse measurements and sparse sources. To accelerate inference, we propose a Gibbs-based geometric consistency reweighting strategy that implicitly approximates optimal transport and straightens noisy coupling paths into stable deterministic trajectories. Phantom and *in vivo* experiments demonstrate that TCDFlow-FMT achieves efficient few-step reconstruction with improved spatial localization and quantitative accuracy. Code will be available at: https://github.com/Qianqian-Xue/TCDFlow-FMT.

Keywords: Fluorescence molecular tomography · Flow matching · Image reconstruction.

1 Introduction

Fluorescence molecular tomography (FMT) is a non-invasive optical technique widely used in early tumor detection [7, 12], drug analysis [13, 18], and dynamic

molecular activity monitoring [4,21]. However, FMT reconstruction is a highly ill-posed nonlinear inverse problem governed by the radiative transfer equation [9, 21]. Due to strong photon scattering and absorption [1] in biological tissues, depth information is degraded, resulting in non-unique and unstable solutions.

Traditional iterative reconstruction methods [3, 5, 10, 22] rely on explicit regularization to constrain the solution space but suffer from prohibitive computational cost and over-smoothing caused by handcrafted priors. Deterministic deep learning models [6, 16, 26, 29] (e.g., convolutional neural networks and Transformers) significantly accelerate reconstruction by learning nonlinear mappings. However, pixel-wise regression losses induce a mean-averaging effect, leading to blurred edges and degraded localization accuracy for sparse fluorescent sources. Diffusion models [8, 15, 28] alleviate mean regression by modeling the full data distribution, yet their Gaussian transition assumption enforces thousands of denoising steps, resulting in inefficient inference. Although stochastic sampling can characterize aleatoric uncertainty, compressing the denoising process into only a few steps may lead to structural instability and positional drift.

To address inference inefficiency and few-step reconstruction instability, we propose **TCDFlow-FMT**, a trajectory-consistent deterministic flow-matching framework for FMT reconstruction. Instead of modeling the inverse problem through stochastic diffusion, we reformulate reconstruction as a continuous-time deterministic transport process between noise and fluorescent sources distribution. This formulation enables efficient ordinary differential equation (ODE)-based sampling while preserving geometric fidelity and strict reproducibility under observational constraints.

Our contributions are threefold. (1) We introduce TCDFlow-FMT, the first deterministic flow-matching framework for fluorescence molecular tomography reconstruction, which reformulates the ill-posed inverse problem as a continuous-time velocity field evolution process, enabling few-step and reproducible reconstruction. (2) We design a spatially guided flow Transformer (SGFT) that explicitly bridges the feature-space gap between multi-view fluorescence measurements and fluorescent sources through structure-aligned encoding and spatially modulated decoding, improving geometric fidelity and localization precision. (3) We propose a geometric consistency reweighting (GCR) strategy grounded in optimal transport theory to enforce near-straight deterministic transport trajectories, effectively mitigating trajectory instability under few-step sampling and maintaining stable high-quality reconstruction.

2 Method

2.1 TCDFlow-FMT Overview

FMT reconstruction aims to recover three-dimensional (3D) fluorescent sources $x \in \mathbb{R}^N$ from fluorescence measurements $\Phi \in \mathbb{R}^M$ based on the linear system $\Phi = Wx$ [27]. To improve inference efficiency and reconstruction stability under few-step sampling, we propose TCDFlow-FMT, as shown in Fig. 1, a deterministic framework governed by continuous-time flow matching theory [14]. Instead of

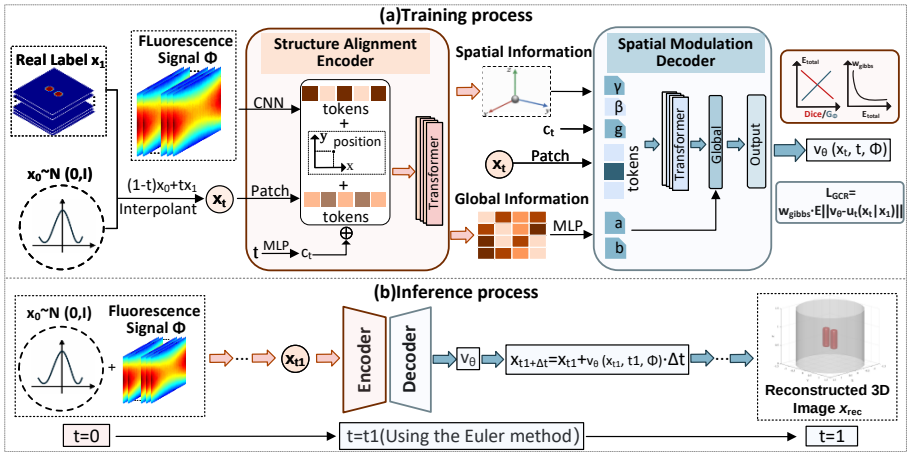

Fig. 1. Overview of the TCDFlow-FMT. (a) Training process; (b) Inference process.

modeling a stochastic reverse stochastic differential equation (SDE), TCDFlow-FMT defines a probability flow ODE that deterministically transports a simple noise distribution $p_0(x)$ to the conditional fluorescent source distribution $p_{data}(x|\Phi)$:

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = v_\theta(x_t, t, \Phi), \quad x_0 \sim \mathcal{N}(0, I), \quad x_1 \sim p_{data}(x|\Phi), \tag{1}$$

where the conditional velocity field v_θ is parameterized by the spatially guided flow Transformer (SGFT). During training, we construct a linear interpolation path, $x_t = (1-t)x_0 + tx_1$, with a target velocity field $u_t(x_t|x_1) = x_1 - x_0$. To accelerate inference, we introduce a geometric consistency reweighting (GCR) strategy. By assigning Gibbs-energy-based weights w_{gibbs} to training pairs, GCR guides the model to learn stable and shorter trajectories. The final weighted regression objective is defined as:

$$\mathcal{L}_{\text{GCR}} = \mathbb{E}_{t,x_0,x_1,\Phi}\left[w_{\text{gibbs}} \cdot ||v_\theta(x_t, t, \Phi) - (x_1 - x_0)||_2^2\right]. \tag{2}$$

Given a fixed initialization, our inference follows a deterministic integration process. The high-fidelity reconstruction x_{rec} is uniquely determined by solving the ODE starting from a fixed x_0: $x_{rec} = x_0 + \int_0^1 v_\theta(x_t, t, \Phi)dt$.

2.2 Spatially Guided Flow Transformer (SGFT)

In TCDFlow-FMT, reconstruction quality relies on accurate velocity field estimation. However, a severe feature space mismatch exists: fluorescence measurements are a globally diffusive field with strong scattering, whereas fluorescent sources are sparse and localized. Standard decoupled diffusion Transformer [19], often designed for low-dimensional conditioning, compresses inputs into a single vector. This leads to the loss of high-dimensional spatial cues essential for

the inverse problem, causing localization errors. To address this, we propose the spatially guided flow Transformer (SGFT), modeling the velocity field as:

$$v_\theta\left(x_t, t, \Phi\right) = \mathcal{D}_\theta\left(x_t, t \mid \mathcal{E}_\phi\left(x_t, t, \Phi\right)\right), \tag{3}$$

where the structure alignment encoder $\mathcal{E}_\phi$ extracts and structures priors from multi-channel fluorescence measurements, and the spatial modulation decoder $\mathcal{D}_\theta$ then generates precise velocity field estimates under the guidance of priors.

The structure alignment encoder $\mathcal{E}_\phi$ extracts spatially aligned priors from the noisy image x_t, time step t, and fluorescence measurements Φ via an early embedding strategy. Specifically, we first map the time step to $c_t = \mathrm{MLP}(t)$ and process the raw signal $\Phi \in \mathbb{R}^{24\times64\times64}$ into patch tokens $\tilde{\Phi} = \mathrm{Patch}(\mathcal{F}_{\mathrm{CNN}}(\Phi)) \in \mathbb{R}^{L\times D}$, in which $\mathcal{F}_{\mathrm{CNN}}$ is a lightweight two-layer CNN capturing local correlations, and the $\mathrm{Patch}(\cdot)$ operation projects features to dimension D, ensuring exact alignment with x_t. To establish geometric correspondence, we inject $\tilde{\Phi}$ directly into the noisy patches via a residual connection $h^{(0)} = \mathrm{Patch}(x_t) + P_{abs} + \tilde{\Phi}$, where P_{abs} denotes the absolute positional encoding. This pixel-wise injection anchors the fluorescence measurements to the coordinate system of the noisy image, enabling precise feature decoupling. Finally, the multi-layer Transformer module outputs sequence features z_{seq} and vector features z_{vec} under c_t:

$$z_{\mathrm{seq}} = \mathrm{SiLU}\big(\mathrm{Linear}(\mathrm{LN}(\tilde{h}^{(L)}))\big), z_{\mathrm{vec}} = \mathrm{Linear}\big(\mathrm{SiLU}(\mathrm{Linear}(\mathrm{GAP}(\tilde{h}^{(L)})))\big), \tag{4}$$

where $\tilde{h}^{(L)} = \mathrm{RMSNorm}(h^{(L)})$ denotes the normalized output of the last transformer block. The LayerNorm-based z_{seq} retains fine-grained spatial information for local distribution modeling, while the MLP-based z_{vec} aggregates global context to regulate overall generative attributes.

The spatial modulation decoder $\mathcal{D}_\theta$ utilizes the extracted priors to generate precise velocity estimates. We first initialize the state $h^{(0)}_{\mathrm{dec}} = \mathrm{Patch}(x_t) + P_{abs}$ to maintain spatial consistency. To address the local sparsity of fluorescence sources, we extend adaptive layer normalization (AdaLN) [17] to a spatial AdaLN, which injects z_{seq} to generate spatially varying parameters $\gamma = \gamma_t(c_t) + \gamma_s(z_{\mathrm{seq}})$, $\beta = \beta_t(c_t) + \beta_s(z_{\mathrm{seq}})$ and $\mathrm{g} = \mathrm{gate}_t(c_t) + \mathrm{gate}_s(z_{\mathrm{seq}})$. These parameters are embedded into both attention and MLP sub-layers via $\mathrm{modulate}(x, \gamma, \beta) = x \cdot (1 + \gamma) + \beta$, followed by a gated residual connection with g. This mechanism allows each patch to receive independent modulation weights, enabling the network to enhance signals in fluorescence regions and suppress background noise based on local structural information extracted by the encoder. Finally, to ensure that the overall intensity distribution of the velocity field remains consistent with the physical measurements, we employ a calibration module using z_{vec}. Specifically, we compute calibration factors $a, b = \mathrm{MLP}_{\mathrm{calib}}(z_{\mathrm{vec}})$ to correct the linear velocity projection v_{proj}, yielding the final estimate $v_\theta = v_{\mathrm{proj}} \odot (1 + a) + b$, where $\odot$ denotes channel-wise multiplication.

2.3 Geometric Consistency Reweighting (GCR)

Independent coupling in flow matching randomly pairs noise with targets, ignoring spatial structural correlations and leading to intersecting evolution paths

and instability. This issue is further exacerbated by the extreme sparsity of fluorescence distributions in FMT, where effective gradients are often drowned by numerous background paths. To straighten the transport trajectories, we propose geometric consistency reweighting (GCR). Instead of costly mini-batch optimal transport (OT), GCR implicitly approximates the Entropic Optimal Transport (EOT) plan by reweighting training pairs through a Gibbs kernel.

To ensure the transport cost is physically well-defined, we first map the sparse target x_1 into a continuous potential field $G_{\Phi} = \omega(x_1 * \mathcal{K}_{\text{core}}) + (1-\omega)(x_1 * \mathcal{K}_{\text{guide}})$, where $*$ denotes convolution. This dual-field design utilizes a small convolution kernel $\mathcal{K}_{\text{core}}$ to lock precise boundaries and a large convolution kernel $\mathcal{K}_{\text{guide}}$ to provide long-range guidance. By creating a wide attraction domain, this multi-scale representation ensures that even evolution paths spatially distant from the target can be effectively guided toward the correct direction. To quantify the transport cost, we must compute the geometric similarity between the noise x_0 and the target field G_{Φ}. However, calculating direct similarity between unstructured Gaussian noise and the sparse potential field is mathematically ill-posed, as the disjoint support of their respective distributions leads to uninformative gradients. Therefore, motivated by noise preconditioning [11], we project x_0 into a feature space aligned with the potential field via $\tilde{x}_0 = \sigma(k(t) \cdot x_0)$, where $k(t) = k_{\text{base}} + \lambda t$ increases linearly over time. This dynamic scaling enables the model to transition from capturing coarse positions via a smoother projection to resolving precise boundaries as the projection sharpens.

To address the ineffectiveness of MSE in distinguishing sparse sources from zero-value backgrounds [2], we quantify the transport cost using a composite energy function that combines structural potential with sparsity regularization:

$$E_{\text{total}} = \underbrace{(1 - \text{Dice}(\tilde{x}_0, G_{\Phi}))}_{\text{Geometric alignment}} + \lambda \cdot \underbrace{(||G_{\Phi}||_1)}_{\text{Sparsity}}. \tag{5}$$

The first term leverages the continuous Dice to enforce geometric alignment, while the second term adjusts energy baseline based on target sparsity. Finally, to prioritize stable trajectories, we transform this energy into a Gibbs training weight [2], defined as $w_{\text{gibbs}} = \frac{1}{Z} \exp(-E_{\text{total}}/\tau)$, where τ is a temperature hyperparameter and Z is a normalization constant. The Gibbs weight w_{gibbs} is inversely proportional to the transport cost E_{total}: samples with better geometric alignment (higher Dice) and enhanced sparsity (lower L_1 norm) receive lower energy and thus higher training weights. In this manner, GCR effectively prioritizes physically plausible transport paths.

3 Experiments and Results

3.1 Dataset

Due to the scarcity of in vivo data, we generated 5,000 training samples via numerical simulations governed by the diffusion approximation and Green function on a finite element mesh [25]. The dataset features randomized cylindrical fluorophores (diameters: 0.13 – 0.17 cm, height: 0.5 cm, center coordinates:

Table 1. Quantitative results of the physical phantom experiment.

Method	EED (1 mm / 2 mm / 3 mm)			
	CNR	LE1 (cm)	LE2 (cm)	Dice
ART	3.06/2.17/2.99	0.03/0.03/0.03	0.02/0.03/0.03	0.37/0.21/0.29
StOMP	2.81/2.27/2.57	0.03/0.02/0.02	0.02/0.02/0.01	0.24/0.20/0.24
3D-CNN	18.41/17.08/18.39	0.02/0.02/0.02	0.02/0.02/0.02	0.68/0.71/0.68
TransUnet	20.49/23.56/15.25	0.02/0.02/0.02	0.02/0.01/0.02	0.70/0.76/0.63
SegMamba	9.82/15.02/13.45	0.03/0.02/0.03	0.02/0.01/0.01	0.40/0.58/0.60
MDiff-FMT	24.49/19.18/16.88	0.03/0.01/0.01	0.02/0.02/0.01	0.63/0.75/0.68
TCDFlow-FMT	**28.29/30.18/21.30**	**0.02/0.01/0.02**	**0.01/0.01/0.01**	**0.70/0.81/0.73**

$[-1.2, 1.2]$ cm, edge-to-edge distances (EEDs): 0.03 – 0.5 cm) with physiological optical attributes. Inputs consist of 24-view fluorescence measurements (64×64) covering 360°, while the ground truth comprises 16 spatially stacked 64×64 transverse slices representing the fluorescent sources distribution.

3.2 Implementation Details

We compared TCDFlow-FMT against ART [10], StOMP [5], 3D-CNN [6], MDiff-FMT [28], TransUnet [26], and SegMamba (adapted for FMT reconstruction) [20] on phantom and *in vivo* datasets, evaluated by Dice [24], CNR [23], and LE [6]. Computations used an NVIDIA RTX 4090 and an A800 for traditional and deep learning models, respectively. For TCDFlow-FMT, we set $D = 384$, $P = 2$, $heads = 6$, $L_{enc} = 12$, $L_{dec} = 6$, and GCR parameters $\omega = 0.7$, $\tau = 0.1$, $k(t) = 4 + 6t$. During inference, we employed the Euler method with 25 function evaluations (NFE=25) for both phantom and *in vivo* experiments. For MDiff-FMT, the sampling process used 1000 denoising steps. ART used 10,000 iterations with a regularization parameter of 0.001, and StOMP used 20 iterations with a threshold of 0.8. All deep learning models were trained to convergence, with the best-performing validation checkpoints selected for evaluation.

3.3 Experimental Results

Physical Phantom Experiments and Results. To evaluate the robustness of TCDFlow-FMT, we conducted physical phantom experiments using a cylindrical phantom (diameter: 3 cm, height: 6 cm, 1% Intralipid, $\mu_s = 10\text{cm}^{-1}$, $\mu_a = 0.02\text{cm}^{-1}$). Two glass tubes (0.15 cm diameter, 20 μL of 1.3 μM indocyanine green (ICG)) served as targets with EEDs of 1, 2, and 3 mm. As shown in Fig. 2, the results from ART and StOMP are overly smoothed, with ART failing to distinguish between the two fluorescent targets. SegMamba exhibits background artifacts and reconstructs one of the fluorescent targets with an irregular shape. While 3D-CNN, TransUnet, and MDiff-FMT produce clear edges, the positions of the reconstructed fluorescent targets are inaccurate. In comparison, TCDFlow-FMT reconstructs the two fluorescent targets with more precise localization and regular shapes. The quantitative results in Table 1 show that TCDFlow-FMT achieved the highest CNR and Dice, demonstrating improved shape recovery and noise suppression.

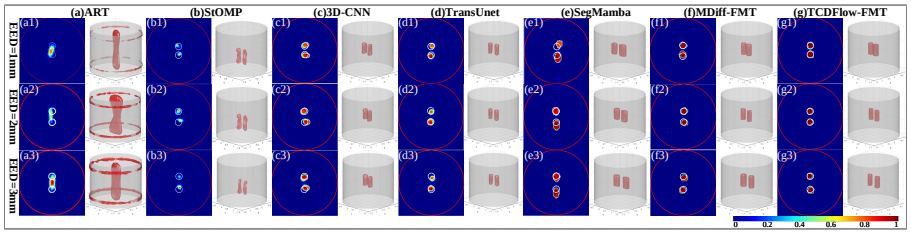

Fig. 2. The physical phantom reconstruction results of seven methods.

Table 2. Quantitative results of the *in vivo* experiment.

Metric	ART	StOMP	3D-CNN	TransUnet	SegMamba	MDiff-FMT	**TCDFlow-FMT**
CNR	1.93	1.43	16.86	14.22	22.14	16.29	**26.13**
LE (cm)	0.02	0.03	0.02	0.03	0.01	0.02	**0.01**
Dice	0.31	0.27	0.52	0.42	0.67	0.48	**0.71**

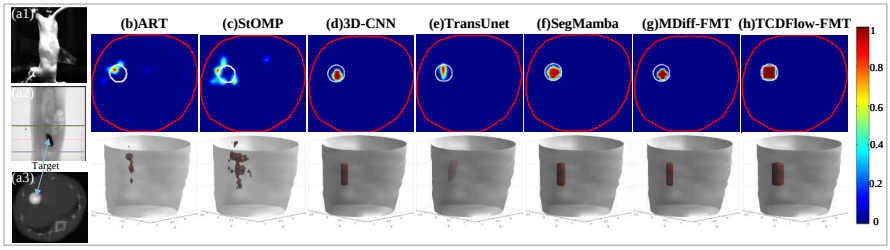

Fig. 3. *in vivo* reconstruction results of seven methods.

in vivo Experiments and Results. To evaluate the *in vivo* feasibility of TCDFlow-FMT, an ICG-DMSO (dimethyl sulfoxide) glass tube was implanted into the abdominal cavity of an 8-week-old female nude mouse (Fig. 3(a1-a2)), using X-ray computed tomography as the anatomical ground truth (Fig. 3(a3)). As shown in Fig. 3(b-h), StOMP and ART fail to achieve accurate morphological reconstruction. Although 3D-CNN, TransUnet, and SegMamba recover the targets, they exhibit irregular boundaries. MDiff-FMT achieves accurate localization but tends to underestimate the target size. In contrast, TCDFlow-FMT yields the most precise localization and regular morphology. Table 2 further confirms that TCDFlow-FMT outperforms other methods in terms of CNR, Dice, and LE, indicating improved performance in signal enhancement, morphological recovery, and localization accuracy.

Ablation Study. We conducted an ablation study on the simulation data with an EED of 3 mm (target centers at (0, 0, 0.75) and (-0.6, 0, 0.75) cm). As shown in Fig. 4(a-b) and Table 3, without GCR, reducing the number of function evaluations (NFE) to 25 significantly intensifies background noise and causes the CNR to drop to 46.71. In contrast, incorporating GCR maintains a CNR of 55.27

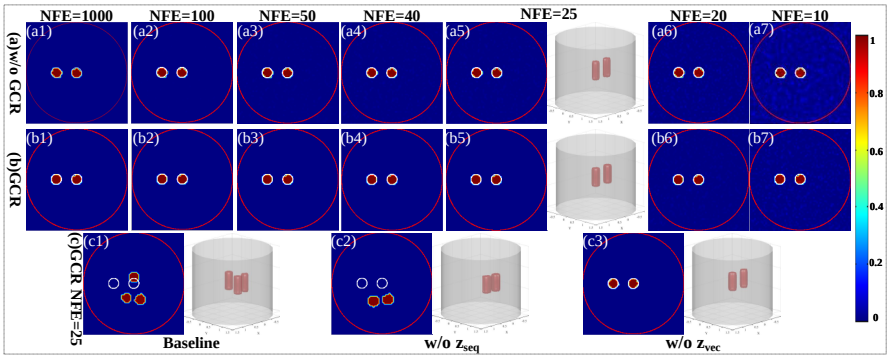

Fig. 4. Reconstruction results of ablation study.

Table 3. Quantitative results of the ablation study.

Method	NFE	time(s)	CNR	LE1(cm)	LE2(cm)	Dice
GCR/ w/o GCR	10	5.12/4.85	44.12/28.06	0.02/0.02	0.02/0.02	0.92/0.91
	20	10.64/11.58	52.10/43.06	0.02/0.02	0.02/0.02	0.92/0.92
	25	13.90/12.90	**55.27**/46.71	**0.02**/0.02	**0.02**/0.02	**0.92**/0.92
	40	15.35/16.33	55.27/51.22	0.02/0.02	0.02/0.02	0.92/0.92
	50	17.38/17.85	55.28/53.01	0.02/0.02	0.02/0.02	0.92/0.92
	100	26.94/25.31	55.28/54.68	0.02/0.01	0.02/0.02	0.92/0.92
	1000	326.66/332.68	55.28/55.04	0.02/0.01	0.02/0.02	0.92/0.92
Baseline	25	12.98	1.58	0.05	0.04	0.22
w/o z_{seq}		11.35	0.07	0.07	0.06	0.20
w/o z_{vec}		13.30	49.66	0.02	0.02	0.91

and a Dice of 0.92 even at NFE=25, effectively suppressing few-step sampling noise. This demonstrates its ability to effectively suppress few-step sampling noise and enhance efficiency by reweighting geometrically consistent trajectories. Furthermore, removing spatial information z_{seq} leads to localization failure (LE $>$ 0.05 cm), while removing intensity calibration z_{vec} degrades signal fidelity (CNR: 49.66). The full model (Fig. 4(b5)) integrates priors, achieving optimal spatial localization and intensity quantification (CNR: 55.27, Dice: 0.92).

4 Conclusion

We introduced TCDFlow-FMT, the first deterministic flow-matching framework for fast and high-fidelity FMT reconstruction. By replacing stochastic diffusion with an ODE-based deterministic flow, our method alleviates the computational inefficiency of diffusion-based models and improves reconstruction stability under few-step sampling. To address the global-local feature mismatch between diffuse measurements and sparse sources, we designed a spatially guided flow Transformer. Coupled with a geometric consistency reweighting strategy that encourages optimal transport, TCDFlow-FMT promotes stable and precise reconstructions with limited inference steps. Phantom and *in vivo* evaluations

demonstrate its effectiveness in spatial localization and quantitative accuracy, while future work will further investigate uncertainty and validate the method under more diverse and realistic fluorophore configurations.

Acknowledgments. This work was partially supported by the National Natural Science Foundation of China (No. 62506220, 62476157, 62471148, and U21A20513), Fundamental Research Program of Shanxi Province (202403021222024), and the Special Project for Patent Transformation in Shanxi Province (No. 20250016).

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

References

1. Bai, J., Xu, Z.: Fluorescence Molecular Tomography, pp. 185–216. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
2. Calvo-Ordonez, S., Meunier, M., Cartea, A., Reisinger, C., Gal, Y., Hernandez-Lobato, J.M.: Weighted conditional flow matching. arXiv preprint arXiv:2507.22270 (2025)
3. Cao, X., He, Y., Zhou, X., Du, J., Chen, Y., Liu, Y., Gao, C., Yi, H.: Adagrad-optimized variational bayesian reconstruction with sparsity-adaptive normal-generalized inverse gaussian prior for fluorescence molecular tomography. Journal of Applied Physics **138**(24), 243103 (2025)
4. Fujita, K., Urano, Y.: Activity-based fluorescence diagnostics for cancer. Chemical Reviews **124** (2024)
5. Gu, Y., Jin, J., Mei, S.: l_0 norm constraint lms algorithm for sparse system identification. IEEE Signal Processing Letters **16**(9), 774–777 (2009)
6. Guo, L., Liu, F., Cai, C., Liu, J., Zhang, G.: 3D deep encoder–decoder network for fluorescence molecular tomography. Opt. Lett. **44**(8), 1892–1895 (2019)
7. Harada, Y., Murayama, Y., Takamatsu, T., Otsuji, E., Tanaka, H.: 5-aminolevulinic acid-induced protoporphyrin ix fluorescence imaging for tumor detection: Recent advances and challenges. International Journal of Molecular Sciences **23**(12), 6478 (2022)
8. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS). Curran Associates Inc., Red Hook, NY, USA (2020)
9. Jiang, H.: Fluorescence Molecular Tomography: Principles and Applications. Springer, Cham (2022)
10. Kak, A.C., Slaney, M.: Principles of computerized tomographic imaging. IEEE Press, Piscataway, NJ, USA (2001)
11. Karras, T., Aittala, M., Aila, T., Laine, S.: Elucidating the design space of diffusion-based generative models. In: Advances in Neural Information Processing Systems. vol. 35, pp. 26565–26577. Curran Associates, Inc. (2022)
12. Kravchenko, Y., Sikora, K., Wireko, A.A., Lyndin, M.: Fluorescence visualization for cancer detection: Experience and perspectives. Heliyon **10**(2), e24390 (2024)
13. Li, X., Wang, W., Gao, Q., Lai, S., Liu, Y., Zhou, S., Yan, Y., Zhang, J., Wang, H., Wang, J., Feng, Y., Yang, R., Su, J., Li, B., Liao, Y.: Intelligent bacteria-targeting zif-8 composite for fluorescence imaging-guided photodynamic therapy of drug-resistant superbug infections and burn wound healing. Exploration **4**(6), 20230113 (2024)

14. Lipman, Y., Chen, R.T.Q., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. In: The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda. OpenReview.net (2023)
15. Liu, J., Anirudh, R., Thiagarajan, J.J., He, S., Mohan, K.A., Kamilov, U.S., Kim, H.: Dolce: A model-based probabilistic diffusion framework for limited-angle ct reconstruction. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 10464–10474 (2023)
16. Meng, H., Gao, Y., Yang, X., Wang, K., Tian, J.: K-nearest neighbor based locally connected network for fast morphological reconstruction in fluorescence molecular tomography. IEEE Transactions on Medical Imaging **39**(10), 3019–3028 (2020)
17. Peebles, W., Xie, S.: Scalable diffusion models with transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4195–4205 (2023)
18. Ronzetti, M., Simeonov, A.: A comprehensive update on the application of high-throughput fluorescence imaging for novel drug discovery. Expert Opinion on Drug Discovery **20**(6), 785–797 (2025)
19. Wang, S., Tian, Z., Huang, W., Wang, L.: DDT: Decoupled diffusion transformer. arXiv preprint arXiv:2504.05741 (2025)
20. Xing, Z., Ye, T., Yang, Y., Liu, G., Zhu, L.: Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. In: Linguraru, M.G., Dou, Q., Feragen, A., Giannarou, S., Glocker, B., Lekadir, K., Schnabel, J.A. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. pp. 578–588. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-72111-3_54
21. Yang, L., Zhang, Z., Zhao, C., Huo, Y., Zhang, E., He, S., Jia, C., Guo, X.: Monitoring molecular dynamics with single-molecule electronic devices and fluorescence techniques. Chinese Journal of Chemistry **41**, 2889–2907 (2023)
22. Yi, H., Ma, S., Yang, R., Zhong, S., Guo, H., He, X., He, X., Hou, Y.: Neighbor-based adaptive sparsity orthogonal least square for fluorescence molecular tomography. Journal of Biomedical Optics **28**, 066005 (2023)
23. Zhang, G., Pu, H., He, W., Liu, F., Luo, J., Bai, L.: Bayesian framework based direct reconstruction of fluorescence parametric images. IEEE Transactions on Medical Imaging **34**, 1378–1391 (2015)
24. Zhang, G., Tzoumas, S., Cheng, K., Liu, F., Liu, J., Luo, J., Bai, J., Xing, L.: Generalized adaptive gaussian markov random field for x-ray luminescence computed tomography. IEEE Transactions on Biomedical Engineering **65**(9), 2130–2133 (2018)
25. Zhang, P., Fan, G., Xing, T., Song, F., Zhang, G.: UHR-DeepFMT: Ultra-high spatial resolution reconstruction of fluorescence molecular tomography based on 3-d fusion dual-sampling deep neural network. IEEE Transactions on Medical Imaging **40**(11), 3217–3228 (2021)
26. Zhang, P., Liu, X., Xue, Q., Shang, Y., Liu, C., Chen, R., Gao, H., Liang, J., Wang, W., Zhang, G.: PAH2T-Former: Paired-attention hybrid hierarchical transformer for synergistically enhanced fmt reconstruction quality and efficiency. IEEE Transactions on Computational Imaging **11**, 536–545 (2025)
27. Zhang, P., Ma, C., Song, F., Fan, G., Sun, Y., Feng, Y., Ma, X., Liu, F., Zhang, G.: A review of advances in imaging methodology in fluorescence molecular tomography. Physics in Medicine Biology **67**(10), 10TR01 (2022)
28. Zhang, P., Xue, Q., Liu, X., Zhang, G., Wang, W., Liang, J.: MDiff-FMT: Morphology-aware diffusion model for fluorescence molecular tomography with small-scale datasets. arXiv preprint arXiv:2410.06757 (2024)

29. Zhao, X., Zhang, L., Qiu, C., Zhang, H., He, X., Leng, X., He, X., Guo, H.: MS-DRN: Multi-scale deep residual network for fluorescence molecular tomography. Computer Methods and Programs in Biomedicine **276**, 109217 (2026)