STAR-Net: Physics Inspired Spectral Topology Aware Reconstruction Network for Single-View Fluorescence Molecular Tomography

Xiangzheng Li[1,2], Jian Zhang[1], Mengxiang Chu[1], Xiaoli Luo[2], Hongbo Guo[1], Xiaowei He[1]*

[1] Key Lab of Radiomics and Intelligent Perception,School of Electronic Information (School of Artificial Intelligence), Northwest University, Xi'an, China

[2]School of Mathematics and Computer Science, Ningxia Normal University,Ningxia, China

82021011@nxnu.edu.cn, jian.zhang@stumail.nwu.edu.cn, chumengxiang@stumail.nwu.edu.cn, lxl@nxnu.edu.cn, guohb@nwu.edu.cn, hexw@nwu.edu.cn

Abstract

Fluorescence molecular tomography (FMT) serves as a pivotal modality for preclinical tumor screening. While single-view FMT offers distinct advantages in data acquisition efficiency and cost-effectiveness, the scarcity of projection views severely exacerbates photon scattering-induced depth ambiguity, rendering 3D volumetric recovery a highly ill-posed inverse problem. To address these challenges, we propose a physics-inspired spectral topology aware reconstruction network (STAR-Net). Specifically, STAR-Net establishes a synergistic framework: initially, a frequency domain decoupling strategy is introduced to simulate the physical characteristics of diffuse light fields; building on this, a differentiable inverse spectral gating (DISG) mechanism is utilized to explicitly impose low-pass spectral regularization for precise depth recovery; and further, a dual-domain synergistic module is integrated to dynamically fuse spatial and frequency features, achieving high-fidelity detail preservation. Extensive experiments on the Digimouse benchmark demonstrate that the proposed STAR-Net achieves the highest dice coefficient under single-view conditions, validating that explicit spectral topology modeling is a powerful paradigm for mitigating depth ambiguity.

1 Introduction

Fluorescence molecular tomography (FMT) plays a pivotal role in preclinical tumor screening and drug monitoring. Although the emergence of second near-infrared window (NIR-II) technology has enhanced tissue penetration, FMT fundamentally remains a severely ill-posed inverse problem. Recently, single-view FMT [Li *et al.*, 2025; Zhang *et al.*, 2023] has emerged as a research hotspot, owing to its distinct advantages in high throughput and cost-effectiveness. However, the task of reconstructing a complete 3D volume relying solely on a single 2D planar projection exposes single-view FMT to a challenge of depth ambiguity that is far more formidable than that encountered in multi-view reconstruction.

*Corresponding authors: Xiaowei He.

To mitigate the inherent ill-posedness of this inverse problem, early research predominantly relied on physical models and regularization constraints. For instance, variational bayesian methods incorporating normal-generalized inverse gaussian (NGIG) priors [Cao *et al.*, 2025] were proposed to address the uncertainty induced by tissue scattering. Concurrently, reconstruction models based on the truncated $L_{2,p}$ norm [Yuan *et al.*, 2025] were designed to suppress the impact of noise and outliers. However, such iterative algorithms are typically computationally intensive and highly sensitive to the selection of regularization parameters. Although hybrid approaches combining learning and iteration [Guo *et al.*, 2025] attempt to strike a balance between efficiency and accuracy, their performance remains constrained by the approximation errors inherent in the physical models.

In recent years, deep learning has progressively assumed a dominant role, leveraged by its potent non-linear mapping capabilities. Early convolutional neural networks (CNNs) and their variants significantly enhanced reconstruction speed and target recognition accuracy by introducing multi-scale feature extraction [Zhao *et al.*, 2026], time-frequency integration [Zhao *et al.*, 2025], and graph topological constraints [Wang *et al.*, 2025]. With the advent of Vision Transformers and generative models, the performance of FMT reconstruction has been propelled to new heights. To surmount the inherent limitation of restricted receptive fields in CNNs, Zhang et al. [Zhang *et al.*, 2025] significantly improved the network's ability to capture long-range dependencies via a spatial channel pairwise attention module (SC-PAM). Meanwhile, score-based generative models [He *et al.*, 2025] and diffusion models [Huang *et al.*, 2025] have demonstrated immense potential in morphological recovery and high-fidelity sampling. Furthermore, multiscale structural spectrum regularization (MSSR) [Pei *et al.*, 2026] has been introduced for NIR-II FMT reconstruction, jointly modeling structural correlations and Hessian-based curvature priors to improve localization accuracy and morphological fidelity.

Despite these advancements, existing deep learning paradigms still face a critical bottleneck in the challenging single-view scenario. Whether based on spatial convolutions, Transformers, or emerging generative models, current approaches often struggle to recover accurate depth topology while maintaining quantitative fidelity. We argue that this limitation stems from a fundamental oversight: the geometric

continuity of biological tissues, though manifested as complex spatial textures, inherently corresponds to spectral compactness in the frequency domain. However, prevailing end-to-end networks rarely model this spectral characteristic explicitly, leading to intrinsic ambiguity in depth decoding.

To address this challenge, we propose a physics-inspired spectral topology aware reconstruction network (STAR-Net). Our core insight is that leveraging spectral compactness to constrain geometric continuity offers a viable solution for eliminating non-physical artifacts and mitigating the ill-posedness of single-view depth recovery. Specifically, the network first introduces a frequency domain decoupling strategy inspired by radiative transfer theory. By mimicking the physical distinction between the isotropic diffuse background and source-specific boundaries, features are explicitly decoupled into low-frequency and high-frequency components, thereby aligning the network's inductive bias with light transport laws. Building on this, we reformulate the depth recovery task as a spectral prediction problem, utilizing a differentiable inverse spectral gating (DISG) mechanism to learn low-frequency fourier coefficients. This mechanism employs explicit low-pass spectral regularization to reconstruct a global depth topology with physical consistency. Finally, to circumvent the inherent limitations of 3D convolutional receptive fields, our dual-domain synergistic module (DDSM) establishes a collaborative pathway between the spatial and frequency domains. Through an adaptive gating mechanism, it dynamically fuses local spatial features with global frequency context, effectively capturing long-range dependencies while suppressing transmission noise.

The main contributions of this paper are summarized as follows:

- We propose the frequency domain decoupling strategy (FDDS), which significantly enhances the completeness and physical interpretability of feature representation by physically separating global structures from edge textures within the feature space.
- We design a differentiable inverse spectral gating (DISG) mechanism that innovatively leverages the spectral compactness prior. By employing differentiable spectral prediction and regularization, it effectively alleviates the ill-posed problem of depth ambiguity in single-view imaging.
- We construct a dual-domain synergistic Module (DDSM) to achieve effective dynamic and complementary fusion of spatial and frequency features, thereby reinforcing the network's explicit spectral topology modeling while improving high-fidelity 3D volumetric reconstruction.

2 Method

Unlike traditional reconstruction methods [Klose and Hielscher, 2008; Wang *et al.*, 2009], deep neural network (DNN)-based FMT reconstruction obviates the need for computationally intensive iterative solvers or precise priors regarding optical parameters. Instead, it leverages a data-driven, end-to-end DNN mapping model to reconstruct the fluorescent source distribution directly from surface measurements. Within the STAR-Net framework, the non-linear mapping between surface fluorescence projection signals and the internal 3D fluorescent source distribution is formulated as:

$$\hat{V} = \mathcal{F}_\theta(I). \tag{1}$$

Where $\mathcal{F}_\theta(\cdot)$ denotes the STAR-Net architecture parameterized by θ, which performs rapid forward inference for source reconstruction. I represents the input single-view 2D surface fluorescence projection measurements, and $\hat{V}$ denotes the predicted 3D fluorescent source density distribution. Consequently, the inverse problem of FMT is reformulated as a supervised learning optimization task:

$$\theta^* = \arg\min_\theta \mathcal{L}_{total}(\mathcal{F}_\theta(I), V_{gt}) \tag{2}$$

Where $\mathcal{F}_\theta(I)$ is the network prediction given the current parameters θ, and V_{gt} corresponds to the 3D Ground Truth. During training, the weight vector θ is iteratively updated by minimizing a hybrid objective function $\mathcal{L}_{total}$.

2.1 STAR-Net Architecture

STAR-Net adopts a physics-inspired 3D U-Net architecture. As illustrated in Figure 1, the core pipeline consists of three stages: (a) feature lifting via frequency domain decoupling and differentiable inverse spectral gating, (b) geometric prior embedding, and (c) dual-domain synergistic multi-scale reconstruction.

Initially, FDDS processes the 2D projection features X obtained through random cropping and patch sampling. Using the fast Fourier transform (FFT) and a radial low-pass mask M, FDDS decomposes X into a low-frequency component X_{lf}, encoding global topology, and a high-frequency component X_{hf}, preserving edge textures. To address depth ambiguity during dimensional lifting, DISG employs a lightweight network to predict depth spectral coefficients. A continuous depth gating signal $G_{(z)}$ is then generated through zero-padding and one-dimensional inverse FFT (IFFT), guiding the asymmetric weighted injection of X_{lf} and X_{hf} to synthesize the initial 3D feature volume V_{init}.

To overcome the limited spatial perception of 3D convolutions, Fourier positional encoding maps normalized 3D coordinates into high-dimensional positional embeddings $\mathcal{P}_{emb}$. These embeddings are concatenated with V_{init}, the depth-expanded high-frequency features X_{hf}, and the mask $P(M)$ along the channel dimension, forming hybrid features enriched with physical and geometric priors.

The encoder uses stacked 3D residual blocks to extract deep semantic features, with DDSM embedded at the bottleneck layer. DDSM performs local convolution in the spatial domain and global amplitude modulation in the frequency domain in parallel, and dynamically fuses the dual-path features through adaptive gating. This design suppresses transmission noise while preserving global contextual consistency. Finally, the decoder integrates multi-scale features via skip connections and progressively restores spatial resolution to output a high-fidelity 3D fluorescence distribution.

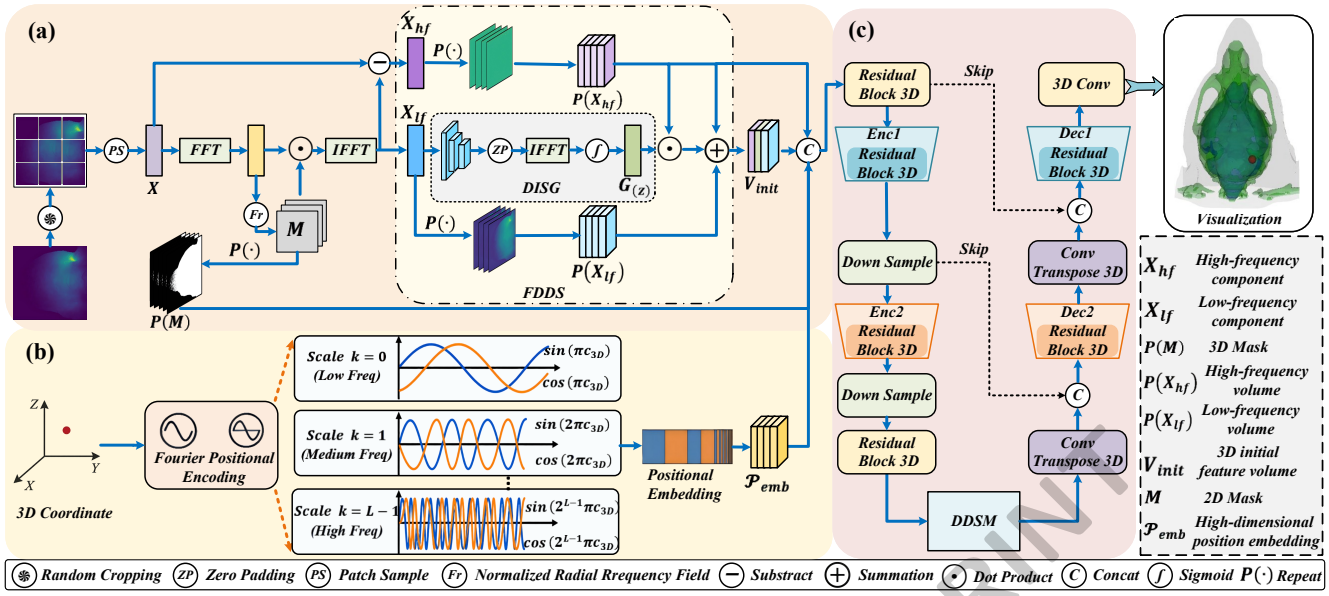

Figure 1: The architecture of the proposed STAR-Net. The pipeline begins with **(a)** the frequency-domain decoupling strategy (FDDS), which separates global structures (X_{lf}) from edge textures (X_{hf}). The differentiable inverse spectral gating (DISG) module then learns low-frequency Fourier coefficients to construct a depth-gating signal $G(z)$, effectively lifting 2D features into a 3D initial volume V_{init}. Simultaneously, **(b)** fourier positional encoding generates geometric embeddings $\mathcal{P}_{emb}$. Finally, **(c)** a dual-domain synergistic (DDSM) 3D U-Net, enhanced by the DDSM, progressively recovers the high-fidelity fluorescence distribution from the fused features.

2.2 Frequency Domain Decoupling Strategy

In optical transport modeling of biological tissue, the solution of the diffusion equation exhibits a pronounced separation in the frequency domain. In particular, low-frequency components predominantly capture the isotropic diffusive background, whereas high-frequency components encode boundary singularities and fine-grained texture details associated with fluorescent sources. To explicitly emulate this physical prior at the early feature-extraction stage, we propose a FDDS.

Instead of processing the full-resolution image directly, our STAR-Net first randomly crops the complete 2D fluorescence distribution map I into multiple patches. For each sampled patch, we obtain a feature tensor $X \in \mathbb{R}^{C\times H\times W}$. We then apply the 2D fast fourier transform $\mathcal{F}_{\text{FFT}}$ to map the features into the spectral domain, which also helps reduce redundant computation:

$$\hat{X}(u,v) = \mathcal{F}_{\text{FFT}}(X) \in \mathbb{C}^{C\times H\times(\frac{W}{2}+1)} \tag{3}$$

Where u and v denote the frequency indices along the height and width dimensions, respectively, and C, H, and W represent the number of channels, height, and width.

To enforce isotropic frequency separation, we compute a normalized radial frequency field $F_r(u,v)$ as:

$$F_r(u,v) = \sqrt{\left(\frac{u}{H}\right)^2 + \left(\frac{v}{W}\right)^2} \tag{4}$$

Given a predefined cutoff frequency threshold η, we construct a binary low-pass mask M to isolate low-frequency components:

$$M(u,v) = \begin{cases} 1, & \text{if } F_r(u,v) \le \eta \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

This design explicitly disentangles the smooth diffusive background from high-frequency, source-specific structures, thereby introducing a physics-aligned inductive bias for subsequent representation learning. In our experiments, we set the cutoff frequency threshold η to 0.25. By applying the spectral mask followed by the inverse fast fourier transform (IFFT), we orthogonally decompose the feature representation into a low-frequency structural component $\mathbf{X}_{lf}$ and a high-frequency detail component $\mathbf{X}_{hf}$:

$$\mathbf{X}_{lf} = \mathcal{F}_{\text{FFT}}^{-1}\left(\hat{\mathbf{X}} \odot \mathbf{M}\right), \qquad \mathbf{X}_{hf} = \mathbf{X} - \mathbf{X}_{lf} \tag{6}$$

Where $\odot$ denotes the hadamard product. Intuitively, low-frequency signals are more sensitive to variations in depth, whereas high-frequency signals are relatively translation-invariant with respect to depth. Consequently, this decoupling strategy encourages $\mathbf{X}_{lf}$ to encode global geometric topology, while preserving local boundaries and fine-grained details in $\mathbf{X}_{hf}$. This yields a physically interpretable feature basis for the subsequent asymmetric depth-lifting module.

2.3 Differentiable Invertible Spectral Gating Mechanism

Recovering 3D depth information from a single 2D fluorescence distribution map is a severely ill-posed inverse problem. Direct voxel-wise depth regression is often under-constrained, leading to non-physical discontinuities and spu-

rious artifacts along the Z-axis. Since biological tissues exhibit intrinsic physical continuity, the depth profile is expected to be spectrally sparse, with most energy concentrated in the low-frequency band.

Motivated by this observation, we propose a differentiable DISG mechanism. It exploits a spectral compactness prior by predicting a small set of frequency-domain coefficients and reconstructing a continuous depth-gating signal. Specifically, we design a lightweight mapping network Φ_{spec}, consisting of a spectral prediction head, a global average pooling layer $G_{\text{AP}}(\cdot)$, and two 1×1 convolutional layers. Given the low-frequency feature $\mathbf{X}_{lf}$, Φ_{spec} predicts the first K complex Fourier coefficients:

$$\boldsymbol{\psi} = \Phi_{\text{spec}}\big(G_{\text{AP}}(\mathbf{X}_{lf})\big) \in \mathbb{C}^{B\times K} \tag{7}$$

where B denotes the batch size. In practice, K is set much smaller than the Nyquist frequency limit along the depth dimension D, i.e., $N_{\text{freq}} = \left(\frac{D}{2} + 1\right)$. This design suppresses high-frequency depth components and imposes low-pass regularization, encouraging a smooth global topology in the reconstructed depth gate.

To convert the predicted spectral coefficients into a spatial depth gate, we construct a differentiable reconstruction operator. We first zero-pad $\boldsymbol{\psi}$ to the full length N_{freq}, obtaining an expanded spectrum $\hat{\boldsymbol{\psi}}$, and then apply inverse FFT:

$$\mathbf{g}_{\text{logits}} = \mathcal{F}^{-1}_{\text{FFT}}\Big(\hat{\boldsymbol{\psi}},\, \text{dim} = D\Big) \in \mathbb{R}^{B\times D} \tag{8}$$

where D denotes the depth dimension of the target volume. The depth-gating signal is obtained through a sigmoid activation: $\mathbf{G}_{(z)} = \delta(\mathbf{g}_{\text{logits}}) \in [0,1]^D$. This mechanism embeds a differentiable low-pass filter into the network, suppressing non-physical high-frequency depth jumps and promoting topological continuity.

Furthermore, inspired by physical considerations, we adopt an asymmetric injection strategy that preserves a smooth basis in high-probability depth regions while injecting high-frequency details near transitions. Using the generated depth gate $\mathbf{G}_{(z)}$, the 2D features are lifted into 3D space as:

$$\mathbf{V}_{\text{init}} = P(\mathbf{X}_{lf}) \odot \mathbf{G}_{(z)} + \Big(P(\mathbf{X}_{lf}) + \gamma \cdot P(\mathbf{X}_{hf})\Big) \odot \Big(1 - \mathbf{G}_{(z)}\Big) \tag{9}$$

where $P(\cdot)$ denotes replication along the depth axis, and γ is a learnable scalar controlling the injection strength of high-frequency textures. This design suppresses noise in core regions while enhancing boundary details.

2.4 Dual-Domain Synergistic Module

To recover the 3D geometric distribution of fluorophores while modeling long-range dependencies and suppressing transport-induced noise, we introduce DDSM. It performs cooperative learning through parallel spatial- and frequency-domain paths, enhancing the network's ability to capture globally consistent contextual information.

Specifically, we first normalize voxel coordinates as $c_{3D} = (x, y, z) \in [-1, 1]^3$, and map them to a high-dimensional Fourier positional embedding $\mathcal{P}_{emb}$ using multi-scale sinusoidal basis functions:

$$\mathcal{P}_{emb} = \bigcup_{k=0}^{L-1} \left[\sin\left(2^k \pi c_{3D}\right),\, \cos\left(2^k \pi c_{3D}\right)\right] \tag{10}$$

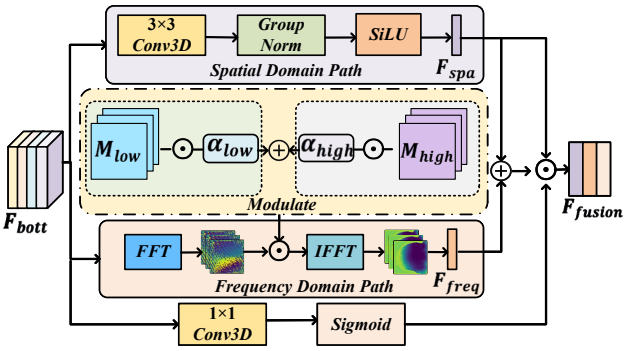

Figure 2: The dual-domain synergistic module (DDSM). It captures long-range dependencies by fusing local features from the spatial path (F_{spa}) and global context from the frequency path (F_{freq}). The latter employs learnable spectral modulation ($\alpha \cdot M$) to adaptively refine global structures before gating fusion.

where L denotes the number of encoding levels, and 2^k enables the model to capture spatial cues from coarse to fine scales. We concatenate $\mathcal{P}_{emb}$ with the initial 3D feature volume $\mathbf{V}_{init}$, the replicated mask $P(\mathbf{M})$, and the replicated high-frequency volume $P(\mathbf{X}_{hf})$, forming the input to the spectral topology-aware network: $\mathbf{X}^{3D}_{in} = \text{Concat}\,(\mathbf{V}_{init},\, P(\mathbf{M}),\, P(\mathbf{X}_{hf}),\, \mathcal{P}_{emb})$. This fused representation is processed by stacked 3D residual blocks, which abstract low-level textures into high-level semantics and produce a compact bottleneck feature $\mathbf{F}_{bott}$.

To alleviate the limited receptive field of 3D convolutions, we design a dual-domain collaborative pathway at the bottleneck layer, where global spectral cues calibrate local spatial features. As shown in Figure 2, the spatial branch uses 3D convolutions to extract local voxel features $\mathbf{F}_{spa}$, preserving sensitivity to small tumors. In parallel, the frequency branch performs slice-wise FFT to capture global correlations within each cross-sectional plane while maintaining depth-wise independence. Two learnable parameters, α_{low} and α_{high}, are introduced to modulate spectral amplitudes:

$$\mathbf{F}_{freq} = \mathcal{F}^{-1}_{\text{FFT}}\Big(\mathcal{F}_{\text{FFT}}(\mathbf{F}_{bott}) \odot (\alpha_{low}\mathbf{M}_{low} + \alpha_{high}\mathbf{M}_{high})\Big) \tag{11}$$

where $\mathbf{M}_{low}$ and $\mathbf{M}_{high}$ denote low- and high-frequency binary masks, respectively. This design adaptively enhances low-frequency global structures while suppressing high-frequency artifacts. Finally, an adaptive gating function $G_{agm}(\cdot)$ dynamically fuses spatial and spectral features: $\mathbf{F}_{fusion} = \mathbf{F}_{bott} + \mathbf{F}_{spa} \odot \mathbf{F}_{freq} + \mathbf{F}_{spa}$. The context-enriched representation is then fed into the decoder and convolutional refinement layers to predict the fluorescence intensity distribution.

2.5 Objective Function

Due to the intrinsic sparsity and ill-posedness of single-view FMT data, optimizing with a single loss term often results in unstable training and poor convergence. To address this issue, we propose a hybrid objective $\mathcal{L}_{total}$, which integrates a dynamically re-weighted binary cross-entropy loss

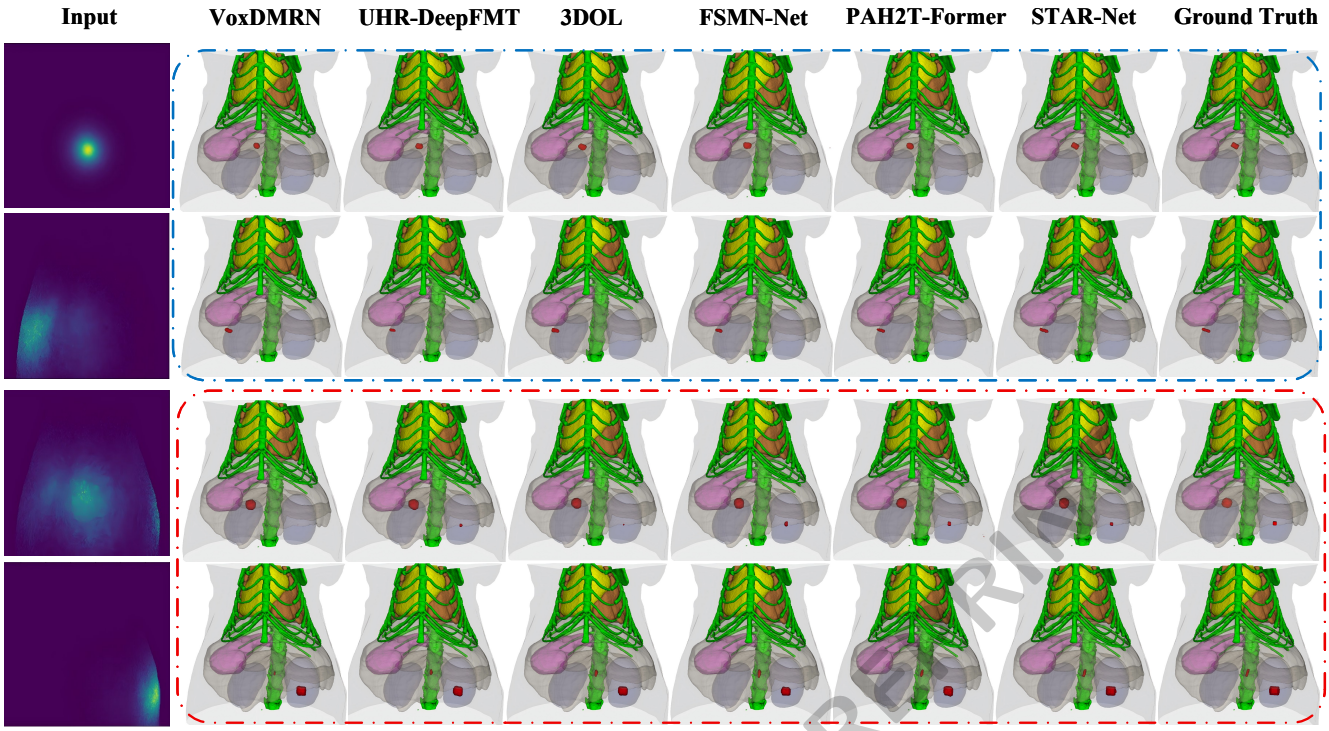

Figure 3: Visual comparison on the abdomen dataset. Blue dashed box: single-source results. Red dashed box: dual-source results. STAR-Net yields sharper boundaries and effectively suppresses artifactual merging in dual-source reconstruction compared to state-of-the-art methods.

$\mathcal{L}_{bce}$[Neyestanak *et al.*, 2025], a soft dice loss $\mathcal{L}_{dice}$[Wang *et al.*, 2025], a focal tversky loss $\mathcal{L}_{ft}$[Abraham and Khan, 2019], and a 3D total variation regularization term $\mathcal{L}_{tv}$[Niu *et al.*, 2014]. This composite objective jointly enforces geometric fidelity and topological continuity. Finally, the overall hybrid objective is formulated as:

$$\mathcal{L}_{total} = \lambda_{bce}\mathcal{L}_{bce} + \lambda_{dice}\mathcal{L}_{dice} + \lambda_{ft}\mathcal{L}_{ft} + \lambda_{tv}\mathcal{L}_{tv} \quad (12)$$

In our experiments, we set $\{\lambda_{bce}, \lambda_{dice}, \lambda_{ft}, \lambda_{tv}\} = \{1.0,\ 1.0,\ 0.6,\ 0.05\}$.

3 Experiments

To comprehensively validate the efficacy of STAR-Net in depth recovery and topological preservation for single-view FMT, we established a high-fidelity digital simulation dataset. This section details the dataset construction, experimental setup, and the quantitative and qualitative analysis results across different source configurations.

3.1 Dataset Construction

To overcome the limitations of simple geometric phantoms, which fail to capture realistic photon transport in biological tissues, we constructed the Digimouse-FMT benchmark dataset. The dataset is built upon the Digimouse digital atlas [Dogdas *et al.*, 2007] and generated using the Monte Carlo eXtreme (MCX) photon simulation software [Fang and Boas, 2009; Yu *et al.*, 2018; Yan and Fang, 2020]. Instead of assuming a homogeneous medium, we segmented the phantom into 11 anatomical tissue types, such as skin, bone, brain, liver, and kidney, and assigned tissue-specific optical properties [Jacques, 2013].

Specifically, we generated data from the brain region, characterized by low absorption, and the abdomen region, characterized by high absorption and strong scattering, to evaluate robustness under different signal-to-noise ratio (SNR) conditions. For physical simulation, 10^6 photon packets were propagated for each sample, and CCD detector shot noise was modeled using a Poisson distribution. Fluorescent sources were randomly sampled within target regions, with depths ranging from 2mm to 15mm. The brain and abdomen datasets each contain 20,000 sample pairs, where each pair consists of single-view projection data and the corresponding 3D ground-truth fluorescence distribution. The entire dataset was randomly split into training, validation, and testing subsets with an $8:1:1$ ratio.

3.2 Implementation Details and Evaluation Metrics

STAR-Net was implemented in PyTorch, with all experiments conducted on a single NVIDIA A100 GPU. The architecture is built upon a residual 3D U-Net backbone. For FDDS, the cutoff frequency was set to $\eta = 0.25$. In DISG, the number of low-frequency coefficients was set to $K = 16$ to balance spectral compactness and representation capacity, and the depth-gating signal was reconstructed via IFFT. Considering 3D memory constraints, we adopted patch-based training, where input projections were randomly cropped to 160×160, corresponding to a 3D volume of $96 \times 160 \times 160$,

Methods	Dice↑		ACS↑		ASSD(mm)↓		LE(mm)↓	
	Brain	Abdomen	Brain	Abdomen	Brain	Abdomen	Brain	Abdomen
VoxDMRN	0.485	0.603	0.652	0.794	0.185	0.136	0.325	0.274
UHR-DeepFMT	0.532	0.655	0.695	0.824	0.142	0.097	0.288	0.252
3DOL	0.548	0.661	0.712	0.823	0.183	0.158	0.412	0.394
FSMN-Net	0.615	0.737	0.795	0.895	0.082	0.058	0.195	0.167
PAH2T-Former	0.638	0.759	0.805	0.890	0.075	0.053	0.152	0.128
STAR-Net	**0.766**	**0.865**	**0.893**	**0.948**	**0.034**	**0.022**	**0.084**	**0.082**

Table 1: Quantitative comparison on brain and abdomen datasets for single-source reconstruction. Metrics include Dice, ACS, ASSD, and LE (↑: higher is better; ↓: lower is better). The best performance among all competing methods is shown in bold.

Methods	Dice↑		SSMP↑		ACS↑		ASSD(mm)↓		LE(mm)↓	
	Brain	Abdomen	Brain	Abdomen	Brain	Abdomen	Brain	Abdomen	Brain	Abdomen
VoxDMRN	0.352	0.455	0.045	0.065	0.553	0.649	0.848	0.819	0.654	0.522
UHR-DeepFMT	0.415	0.552	0.062	0.091	0.621	0.723	0.757	0.798	0.578	0.448
3DOL	0.478	0.607	0.075	0.096	0.645	0.712	0.677	0.792	0.546	0.436
FSMN-Net	0.535	0.648	0.082	0.112	0.705	0.768	0.591	0.726	0.475	0.441
PAH2T-Former	0.562	0.672	0.085	0.088	0.724	0.772	0.549	0.728	0.421	0.442
STAR-Net	**0.644**	**0.752**	**0.103**	**0.122**	**0.807**	**0.827**	**0.416**	**0.433**	**0.201**	**0.355**

Table 2: Quantitative results for dual-source reconstruction. The SSMP metric is introduced to evaluate the structural separability of multiple sources. Note that LE is reported in cm. Best results are marked in bold.

with a positive sampling ratio of 0.7. The model was optimized using AdamW [Zhou *et al.*, 2024] with an initial learning rate of 1×10^{-3}, weight decay of 1×10^{-4}, and a learning rate scheduler.

To comprehensively evaluate reconstruction performance, we employ the dice similarity coefficient (Dice) and localization error (LE) to quantify volumetric overlap and centroid localization accuracy. To address depth ambiguity and topological adhesion in single-view FMT, we further use the average symmetric surface distance (ASSD) [Shuai *et al.*, 2025] to assess morphological fidelity. In addition, we introduce the axial consistency score (ACS), which measures the similarity of normalized intensity distributions along the depth (z) axis and evaluates global depth topology recovery:

$$\text{ACS} = 1 - \frac{1}{2} \sum_{z} \left| \frac{\sum_{x,y} P(x,y,z)}{\sum P} - \frac{\sum_{x,y} Y(x,y,z)}{\sum Y} \right| \tag{13}$$

A higher ACS indicates stronger suppression of depth elongation artifacts. For dual-source scenarios, we propose the source separation and merging penalty (SSMP), which combines matched intersection over union (IoU) with connected component counts to penalize erroneous merging or splitting:

$$\text{SSMP} = \text{Sep} - 0.5 \cdot \mathbb{I}(N_{pred} < 2) - 0.25 \cdot \frac{\max(0, N_{pred} - 2)}{N_{pred}} \tag{14}$$

where Sep denotes the mean IoU of matched components, and N_{pred} is the number of predicted connected components. This metric effectively evaluates whether the model achieves physical separation of dual sources.

3.3 Single-Source Reconstruction Analysis

To demonstrate the superiority of STAR-Net in resolving depth ambiguity, we benchmarked it against a diverse set of five leading approaches, ranging from advanced CNN-based models (VoxDMRN [Li *et al.*, 2022], UHR-DeepFMT [Zhang *et al.*, 2021], FSMN-Net [Li *et al.*, 2024]) to Transformer-based architectures (PAH2T-Former [Zhang *et al.*, 2025]) and multi-task frameworks (3DOL [Li *et al.*, 2023]). All experiments were conducted under a consistent and controlled single-view configuration. Quantitative results presented in Table 1 demonstrate that STAR-Net achieves superior performance across all evaluated metrics. Notably, on the abdomen dataset characterized by strong scattering STAR-Net attains a Dice coefficient of 0.865, significantly outperforming the runner-up method, PAH2T-Former (0.759). Crucially, leveraging the explicit constraints imposed by the DISG mechanism on the physical depth distribution, our method reduces the localization error (LE) to a submillimeter level of 0.08 mm. This represents a substantial improvement over VoxDMRN (0.274 mm) and effectively mitigates the depth ambiguity inherent in single-view imaging. Furthermore, the frequency domain decoupling strategy ensures precise morphological delineation, resulting in an average symmetric surface distance (ASSD) of 0.022 mm. Regarding qualitative visualization, as highlighted by the dashed blue boxes in Figure 3 and Figure 4, STAR-Net effectively suppresses background noise and reconstructs 3D geometric topologies that are highly congruent with the ground truth. This stands in contrast to the diffuse contours generated by CNN-based methods and the depth elongation artifacts observed in Transformer-based approaches.

Model Variant	FDDS	DISG	DDSM	Dice	ASSD	LE
Baseline	×	×	×	0.715	0.088	0.215
Baseline+FDDS	✓	×	×	0.788	0.052	0.155
Baseline+FDDS+DISG	✓	✓	×	0.835	0.036	0.112
STAR-Net	✓	✓	✓	**0.865**	**0.022**	**0.082**

Table 3: Ablation study quantifying the impact of each component.

3.4 Dual-Source Reconstruction Analysis

Dual-source reconstruction serves as a critical benchmark for assessing a model's capacity to resolve topological structures in multi-target scenarios. Table 2 presents the quantitative results for this task. Due to photon scattering interference between multiple targets, the LE for all methods escalated to the centimeter order. However, STAR-Net excels in the SSMP metric, which reflects structural separability achieving scores of 0.103 and 0.122 in the brain and abdomen regions, respectively. These results significantly outperform PAH2T-Former, demonstrating our method's efficacy in mitigating artifactual merging. Regarding the ACS metric, STAR-Net achieved a score of 0.827 on the abdomen dataset, indicating that the DDSM successfully recovers global intensity trends via frequency-domain amplitude modulation.

In terms of qualitative visualization, as highlighted by the red dashed boxes in Figure 3 and Figure 4, distinct performance differences are observable when two sources are in close proximity. Baseline methods, such as PAH2T-Former and 3DOL, tend to conflate the two targets into a single connected mass, leading to topological errors. Conversely, STAR-Net successfully disentangles the two independent sources while preserving clear boundaries. This superior performance is attributed to the DDSM's capability to preserve high-frequency spatial information, thereby rendering the network highly sensitive to fine-grained edge details.

3.5 Ablation Study

To validate the contribution of each component, we performed a progressive ablation study on the Abdomen dataset (see Table 3), where strong scattering makes depth recovery particularly challenging.

Effectiveness of FDDS: The baseline U-Net yielded a Dice of 0.715, ASSD of 0.088 mm, and LE of 0.215 mm. Integrating FDDS improved the Dice to 0.788 and reduced LE to 0.155 mm. This confirms that decoupling high-frequency edge details from low-frequency global structures enhances feature completeness by mimicking the frequency separation property of diffuse light transport. In particular, the low-frequency component provides a more stable representation of global topology, while the high-frequency component preserves source-related boundary textures.

Efficacy of DISG: The addition of DISG further improved Dice to 0.835 and significantly reduced LE to 0.112 mm. By predicting compact low-frequency Fourier coefficients to generate a continuous depth gate, DISG imposes a spectral compactness prior on the lifting process. This explicitly constrains geometric continuity along the depth direction, suppresses non-physical axial fluctuations, and effectively mitigates the depth ambiguity inherent in single-view FMT.

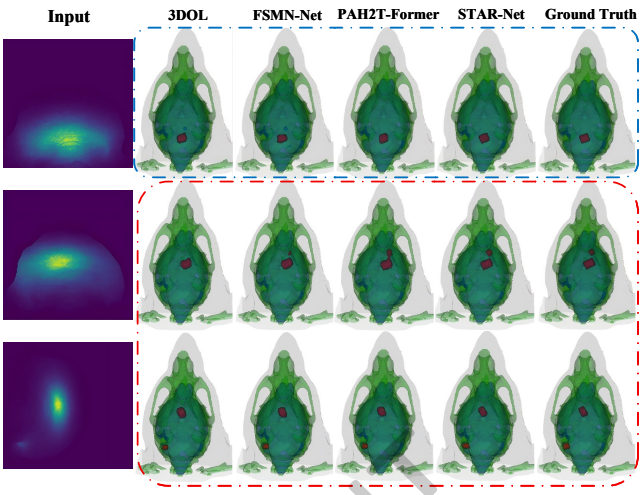

Figure 4: Evaluation on the brain dataset. Blue and red boxes indicate single and dual-source results. Consistent with the abdomen cases, STAR-Net achieves superior robustness and artifact-free separation.

Impact of DDSM: Finally, STAR-Net with DDSM achieved the best performance, with a Dice of 0.865 and a sub-millimeter LE of 0.082 mm. The reduction in ASSD to 0.022 mm demonstrates that synergistic learning between spatial and frequency domains effectively refines boundary sharpness and captures long-range dependencies. These results indicate that DDSM complements FDDS and DISG by enhancing global-local feature interaction while suppressing transport-induced noise.

4 Conclusion

In this paper, we proposed STAR-Net, a physics-inspired spectral topology-aware network, to mitigate the severe ill-posedness and depth ambiguity in single-view FMT reconstruction. The superior performance of STAR-Net stems from explicitly encoding photon transport priors into the deep architecture. Specifically, the frequency domain decoupling strategy (FDDS) leverages the spectral separation characteristics of diffuse light fields to disentangle global structures from edge textures. Combined with the differentiable inverse spectral gating (DISG) mechanism, the model enforces topological continuity through spectral compactness regularization inspired by biological tissues, thereby effectively alleviating depth ambiguity and achieving sub-millimeter localization precision (0.082 mm). Furthermore, the dual-domain synergistic module (DDSM) enhances physical separability in multi-source scenarios by capturing long-range dependencies via frequency-domain modulation, as reflected by superior SSMP scores. Extensive experiments on the scattering-intensive abdomen dataset demonstrate that STAR-Net significantly outperforms existing methods, achieving a Dice coefficient of 0.865. Future work will focus on bridging the Sim-to-Real gap using unsupervised domain adaptation to address complex heterogeneous optical properties in clinical data.

Acknowledgements

This research was supported in part by the National Natural Science Foundation of China under Grants 12271434, 62271394, and 62201459; the Science and Technology Projects of Xi'an under Grant 201805060ZD11CG44; the Science and Technology New Star Program of Shaanxi Province under Grant 2023KJXX-035; the Natural Science Basic Research Program of Shaanxi Province under Grant 2023-JC-JQ-57; the General Project of the Key Research and Development Plan of Shaanxi Province under Grant 2025SF-YBXM-240; the Ningxia Natural Science Foundation of China under Grant 2025AAC030591; and the Artificial Intelligence Empowering the "Scientists + Engineers" Team in Education under Grant 2024JH-KGDW-0084.

References

[Abraham and Khan, 2019] Nabila Abraham and Naimul Mefraz Khan. A novel focal tversky loss function with improved attention u-net for lesion segmentation. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, pages 683–687. IEEE, 2019.

[Cao *et al.*, 2025] Xin Cao, Yiting He, Xin Zhou, Jiaxin Du, Yi Chen, Yangyang Liu, Chengyi Gao, and Huangjian Yi. Adagrad-optimized variational bayesian reconstruction with sparsity-adaptive normal-generalized inverse gaussian prior for fluorescence molecular tomography. *Journal of Applied Physics*, 138(24), 2025.

[Dogdas *et al.*, 2007] Belma Dogdas, David Stout, Arion F Chatziioannou, and Richard M Leahy. Digimouse: a 3d whole body mouse atlas from ct and cryosection data. *Physics in medicine & biology*, 52(3):577, 2007.

[Fang and Boas, 2009] Qianqian Fang and David A Boas. Monte carlo simulation of photon migration in 3d turbid media accelerated by graphics processing units. *Optics express*, 17(22):20178–20190, 2009.

[Guo *et al.*, 2025] Ruchi Guo, Jiahua Jiang, Bangti Jin, Wuwei Ren, and Jianru Zhang. A warm-basis method for bridging learning and iteration: a case study in fluorescence molecular tomography. *arXiv preprint arXiv:2510.05926*, 2025.

[He *et al.*, 2025] Peng He, Jiayuan Lin, Yin Zhu, Qiao Wan, Chengzhong Wu, Wenbo Wan, and Qiegen Liu. Fluorescence molecular tomography via score-based generative model. *Optics and Lasers in Engineering*, 187:108863, 2025.

[Huang *et al.*, 2025] Qiushi Huang, Chunzhao Li, Anqi Xiao, Jie Tian, and Zhenhua Hu. Dang: Data augmentation based on nir-ii guided diffusion model for fluorescence molecular tomography. *IEEE Transactions on Computational Imaging*, 12:128–141, 2025.

[Jacques, 2013] Steven L Jacques. Optical properties of biological tissues: a review. *Physics in Medicine & Biology*, 58(11):R37, 2013.

[Klose and Hielscher, 2008] Alexander D Klose and Andreas H Hielscher. Optical tomography with the equation of radiative transfer. *International Journal of Numerical Methods for Heat & Fluid Flow*, 18(3/4):443–464, 2008.

[Li *et al.*, 2022] Shuangchen Li, Jingjing Yu, Xuelei He, Hongbo Guo, and Xiaowei He. Voxdmrn: a voxelwise deep max-pooling residual network for bioluminescence tomography reconstruction. *Optics Letters*, 47(7):1729–1732, 2022.

[Li *et al.*, 2023] Shuangchen Li, Beilei Wang, Jingjing Yu, Dizhen Kang, Xuelei He, Hongbo Guo, and Xiaowei He. 3d-deep optical learning: a multimodal and multitask reconstruction framework for optical molecular tomography. *Optics Express*, 31(15):23768–23789, 2023.

[Li *et al.*, 2024] Shuangchen Li, Beilei Wang, Jingjing Yu, Xuelei He, Hongbo Guo, and Xiaowei He. Fsmn-net: a free space matching network based on manifold convolution for optical molecular tomography. *Optics Letters*, 49(5):1161–1164, 2024.

[Li *et al.*, 2025] Yunfei Li, Qian Liu, and Fuhong Cai. Single-view fluorescence molecular tomography based on hyperspectral nir-ii imaging. *IEEE Transactions on Computational Imaging*, 2025.

[Neyestanak *et al.*, 2025] Mahdi Shafiei Neyestanak, Hamid Jahani, Mohsen Khodarahmi, Javad Zahiri, Mostafa Hosseini, Amirali Fatoorchi, and Mir Saeed Yekaninejad. A quantitative comparison between focal loss and binary cross-entropy loss in brain tumor auto-segmentation using u-net. *Journal of Biostatistics and Epidemiology*, 2025.

[Niu *et al.*, 2014] Shanzhou Niu, Yang Gao, Zhaoying Bian, Jing Huang, Wufan Chen, Gaohang Yu, Zhengrong Liang, and Jianhua Ma. Sparse-view x-ray ct reconstruction via total generalized variation regularization. *Physics in Medicine & Biology*, 59(12):2997, 2014.

[Pei *et al.*, 2026] Ziyu Pei, Yifei Du, Chunzhao Li, Jie Bao, Anqi Xiao, Yaqi Tian, Yuxi Deng, Hongbo Guo, Jie Tian, and Zhenhua Hu. Multiscale structural spectrum regularization for nir-ii fluorescence molecular tomography. *Optics & Laser Technology*, 200:115159, 2026.

[Shuai *et al.*, 2025] GENG Shuai, LI Yonghui, AO Yu, SHI Weili, MIAO Yu, WANG Shuhan, et al. Brain midline segmentation method based on prior knowledge and path optimization. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi= Journal of Biomedical Engineering*, 42(4):766, 2025.

[Wang *et al.*, 2009] Daifa Wang, Xin Liu, Yanping Chen, and Jing Bai. A novel finite-element-based algorithm for fluorescence molecular tomography of heterogeneous media. *IEEE Transactions on Information Technology in Biomedicine*, 13(5):766–773, 2009.

[Wang *et al.*, 2025] Beilei Wang, Shuangchen Li, Heng Zhang, Jia Li, Lizhi Zhang, Jingjing Yu, Xiaowei He, and Hongbo Guo. Deep system prior based graph convolution network for nir-ii fluorescence molecular tomography. *Computer Methods and Programs in Biomedicine*, page 108948, 2025.

[Yan and Fang, 2020] Shijie Yan and Qianqian Fang. Hybrid mesh and voxel based monte carlo algorithm for accurate and efficient photon transport modeling in complex bio-tissues. *Biomedical Optics Express*, 11(11):6262–6270, 2020.

[Yu *et al.*, 2018] Leiming Yu, Fanny Nina-Paravecino, David Kaeli, and Qianqian Fang. Scalable and massively parallel monte carlo photon transport simulations for heterogeneous computing platforms. *Journal of biomedical optics*, 23(1):010504–010504, 2018.

[Yuan *et al.*, 2025] Yating Yuan, Hongbo Guo, Jingjing Yu, Huangjian Yi, Xuelei He, and Xiaowei He. Robust reconstruction of fluorescence molecular tomography by minimizing the capped l2, p norm. *Biomedical Signal Processing and Control*, 109:108021, 2025.

[Zhang *et al.*, 2021] Peng Zhang, Guangda Fan, Tongtong Xing, Fan Song, and Guanglei Zhang. 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.

[Zhang *et al.*, 2023] Xuanxuan Zhang, Yunfei Jia, Jiapei Cui, Jiulou Zhang, Xu Cao, Lin Zhang, and Guanglei Zhang. Two-stage deep learning method for sparse-view fluorescence molecular tomography reconstruction. *Journal of the Optical Society of America A*, 40(7):1359–1371, 2023.

[Zhang *et al.*, 2025] Peng Zhang, Xingyu Liu, Qianqian Xue, Yu Shang, Chen Liu, Ruhao Chen, Honglei Gao, Jiye Liang, Wenjian Wang, and Guanglei Zhang. Pah2t-former: Paired-attention hybrid hierarchical transformer for synergistically enhanced fmt reconstruction quality and efficiency. *IEEE Transactions on Computational Imaging*, 2025.

[Zhao *et al.*, 2025] Yizhe Zhao, De Wei, Heng Zhang, Shuangchen Li, Lizhi Zhang, Jintao Li, Hongbo Guo, and Xiaowei He. Setdn: Signal-extraction and target-detection network for dynamic fluorescence molecular tomography. *Expert Systems with Applications*, page 129441, 2025.

[Zhao *et al.*, 2026] Xin Zhao, Liuyuan Zhang, Chunyu Qiu, Heng Zhang, Xiaowei He, Xin Leng, Xuelei He, and Hongbo Guo. Msdrn: Multi-scale deep residual network for fluorescence molecular tomography. *Computer Methods and Programs in Biomedicine*, 276:109217, 2026.

[Zhou *et al.*, 2024] Pan Zhou, Xingyu Xie, Zhouchen Lin, and Shuicheng Yan. Towards understanding convergence and generalization of adamw. *IEEE transactions on pattern analysis and machine intelligence*, 46(9):6486–6493, 2024.