SUGAR: STRUCTURED UNSUPERVISED GEOMETRY-AWARE CT ATTENUATION ALIGNMENT AND PET ACTIVITY RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

A misaligned computed tomography (CT)-derived attenuation map can bias time-of-flight positron emission tomography (TOF-PET) activity reconstruction. This bias can persist after subsequent image registration, motivating joint estimation of activity and CT geometry from emission data. We present **SUGAR**, a **S**tructured **U**nsupervised **G**eometry-aware framework for CT attenuation **A**lignment and TOF-PET activity **R**econstruction. SUGAR fits a differentiable acquisition model to the measured data. A coordinate network is fitted to the CT attenuation map and frozen; a global transformation and neural residual warp control its sampling coordinates. A separate positive neural field represents the PET activity. To reduce redundancy between the two geometry components, SUGAR suppresses sampled infinitesimal rigid modes in the residual deformation. It further factorizes the positive activity coefficients into a scalar magnitude and a simplex-valued composition. The activity field initially uses coarse hash-grid features, with finer features introduced on a preset schedule. In evaluations on digital brain and torso phantoms, SUGAR improves activity reconstruction over the evaluated comparison methods, while controlled tests assess attenuation-prior alignment. The results support structured geometry, activity coefficients, and resolution scheduling for coupled activity–attenuation inverse problems.

1 INTRODUCTION

In time-of-flight positron emission tomography (TOF-PET) (Surti, 2015), measurement discrepancies reflect tracer activity or attenuation along detection paths. Computed tomography (CT) constrains attenuation (Kinahan et al., 1998), but respiration and patient motion can misalign it with PET. Reconstruction may then absorb attenuation errors into activity; subsequent registration does not revisit this attribution. Joint reconstruction and alignment address this coupling (Rezaei et al., 2016), as do motion-aware gated and self-gated models (Bousse et al., 2016; Elhamiasl et al., 2025).

Maximum likelihood reconstruction of attenuation and activity (MLAA) estimates both fields, with attenuation sinograms identifiable up to an additive constant (Rezaei et al., 2012). Maximum-likelihood reconstruction and registration (MLRR) instead deforms a CT attenuation prior while reconstructing activity (Rezaei et al., 2016). This retains CT attenuation values while estimating spatial correspondence. Global pose and local deformation must both be represented to capture spatially varying mismatch. Continuous neural fields model images and displacements (Sitzmann et al., 2020; Wolterink et al., 2022); differentiable projection guides both fields using emission data.

Learned PET–CT registration (Schaefferkoetter et al., 2023) and short-frame motion correction (Zhou et al., 2023) provide additional priors. Our setting instead uses one static TOF acquisition and its CT prior, motivating structure within per-scan optimization. Both fields must explain the same counts despite different PET and CT intensities, making residual allocation central.

Yet reducing the emission loss does not by itself establish correct alignment. With flexible neural fields, a residual warp can duplicate global motion, so different parameter updates can describe a similar coordinate map. Activity can also absorb errors caused by a misaligned CT prior: an expressive activity field may improve the measurement fit before the geometry has been corrected. A

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

further interaction arises within activity itself, where positive basis coefficients control both overall magnitude and relative composition. These interactions motivate structuring the joint optimization.

To address these interactions, **SUGAR** jointly reconstructs activity and aligns a frozen CT field from one TOF acquisition. A coordinate network is fitted to the CT attenuation map and frozen; optimization updates its sampling geometry and a separate positive activity field. No external pretraining, paired motion supervision, or PET–CT similarity loss is required.

SUGAR structures this optimization through three complementary components. Gauge correction suppresses sampled infinitesimal rigid modes in the residual warp, reducing overlap with the global pose branch. Polar coefficients use a magnitude–direction parameterization (Salimans & Kingma, 2016) to separate activity scale from normalized basis composition. Following coarse-to-fine neural optimization (Lin et al., 2021; Wu et al., 2025), multiresolution continuation gradually introduces fine activity features, limiting early compensation while preserving later spatial detail.

Experiments on BrainWeb and extended cardiac-torso (XCAT) phantoms evaluate activity reconstruction under rigid and respiratory mismatch at normal and low count levels. Alongside volumetric image metrics, tumor contrast recovery assesses the reconstruction of small uptake targets. An eight-configuration XCAT ablation measures the combined benefit of the three components. Controlled tests link the components to geometry recovery, convergence speed, and early activity compensation.

2 Preliminaries and Related Work

2.1 TOF-PET Measurement Model with Alignment

TOF localizes emission along a line of response (LOR), with uncertainty encoded by a kernel (Moses, 2003; Surti, 2015). Let $\Omega \subset \mathbb{R}^3$ be the reconstruction box, λ the activity, and μ_ϕ the aligned CT attenuation (Kinahan et al., 1998). LOR i follows $\mathbf{x}_i(t) = \mathbf{s}_i + t\mathbf{d}_i$, with unit direction $\mathbf{d}_i$ and segment $\mathcal{T}_i$ inside Ω. For $y_{ik} \sim \mathrm{Poisson}(\bar{y}_{ik})$ (Shepp & Vardi, 1982),

$$\bar{y}_{ik} = q_i \underbrace{\exp\Big[-\int_{\mathcal{T}_i} \mu_\phi(\mathbf{x}_i(t))\,\mathrm{d}t\Big]}_{\text{attenuation}} \cdot \underbrace{\int_{\mathcal{T}_i} \lambda(\mathbf{x}_i(t))\, h_i(t - t_{ik})\,\mathrm{d}t}_{\text{activity}} + b_{ik}. \tag{1}$$

Here q_i is known sensitivity, b_{ik} is random/scatter background, and h_i is the TOF kernel centered at t_{ik}. Denote the factors by $a_i(\mu_\phi)$ and $p_{ik}(\lambda)$. Attenuation is shared across each LOR's TOF bins; unrestricted joint estimation retains a scale ambiguity (Rezaei et al., 2012). Activity updates change the TOF-weighted projection; geometry updates change the shared attenuation factor by resampling the CT prior. Appendix A.1 details the acquisition physics.

2.2 Joint PET Reconstruction and Alignment

Demons estimates image correspondence (Thirion, 1998); MLRR couples CT registration with activity reconstruction through emission data (Rezaei et al., 2016). Gated and self-gated models combine motion, warped attenuation, and anatomical priors (Bousse et al., 2016; Tsai & Liu, 2023; Elhamiasl et al., 2025). Learned methods address gated reconstruction (Li et al., 2022), PET–CT alignment (Schaefferkoetter et al., 2023), or short-frame motion correction and count restoration (Zhou et al., 2023), requiring validation across acquisition conditions (Arabi et al., 2021). Appendix A.4 details adaptations.

2.3 Neural Representations and Continuation

Implicit neural representations (INRs) model continuous signals and deformations (Sitzmann et al., 2020; Wolterink et al., 2022; Byra et al., 2023). In PET, PD-INR uses an ordered-subsets expectation-maximization (OSEM) image prior (Long et al., 2025); IMREPET shares a network across dynamic frames (Fan et al., 2025). Neither jointly estimates activity and CT alignment. Spatial encodings control neural-field detail (Tancik et al., 2020; Müller et al., 2022); BARF schedules positional frequencies for scene/pose estimation (Lin et al., 2021), and Moner uses coarse-to-fine hash encoding for motion-aware magnetic resonance (MR) imaging (Wu et al., 2025). Nonnegative

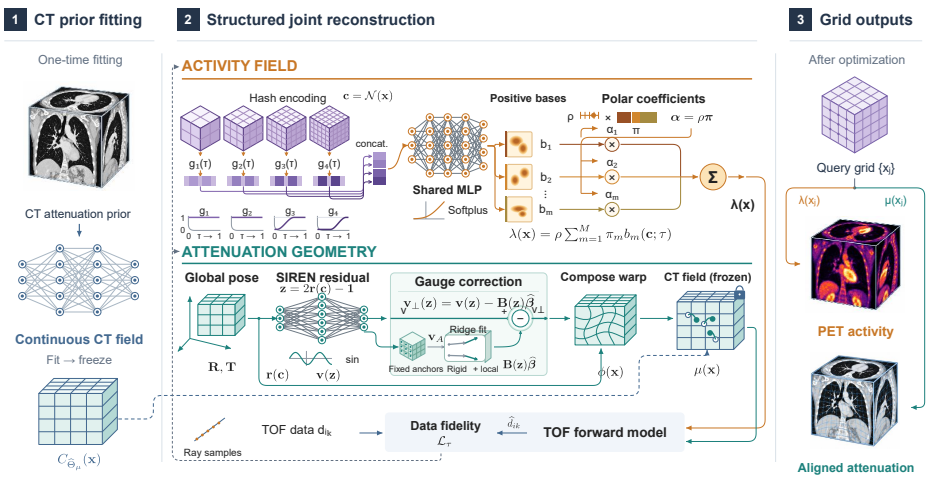


Figure 1: Overview of SUGAR. A global–residual warp queries the frozen CT field; gauge correction removes the residual's fitted rigid modes. An MLP produces positive activity bases with polar coefficients and progressively activated hash levels. Both fields enter the TOF emission objective. Blue dashed links reuse frozen CT parameters; gray dashed links denote alternating activity and geometry updates. Continuation advances every epoch (Algorithm 2).

factorization provides additive bases (Lee & Seung, 2000); weight normalization separates magnitude and direction (Salimans & Kingma, 2016). SUGAR couples these representations with control of rigid-mode overlap between global pose and residual deformation.

3 Proposed Method

In this section, we first formulate the emission-domain objective in Section 3.1. We then introduce gauge-fixed CT alignment to reduce rigid-mode redundancy (Section 3.2), polar activity coefficients to separate magnitude and composition (Section 3.3), and multiresolution continuation to control spatial detail during optimization (Section 3.4). Finally, we describe the alternating updates of activity and geometry. Figure 1 summarizes the SUGAR architecture.

3.1 Problem Formulation

We represent the activity and aligned attenuation fields as

$$\lambda_{\Theta_\lambda}(\mathbf{x}) = F_{\Theta_\lambda}(\mathbf{x}), \tag{2}$$

$$\mu_{\Theta_\phi}(\mathbf{x}) = C_{\widehat{\Theta}_\mu}\left(\phi_{\Theta_\phi}(\mathbf{x})\right), \tag{3}$$

where F_{Θ_λ} is positive. Let $\mathcal{N}(\mathbf{x}) = \mathbf{D}^{-1}(\mathbf{x} - \mathbf{a})$, with box origin $\mathbf{a}$ and diagonal side-length matrix $\mathbf{D}$, map the physical domain to $[0,1]^3$. We use $C_{\Theta_\mu}(\mathbf{x}) = \mathcal{C}_{\Theta_\mu}(\mathcal{N}(\mathbf{x}))$ for the physical-coordinate wrapper of the normalized CT network $\mathcal{C}$. The CT decoder uses rectified linear unit (ReLU) activations and a nonnegative ReLU output. After fitting to the scan's attenuation prior, it is frozen. The backward map ϕ_{Θ_ϕ} sends PET coordinates to physical CT queries. Only its global and residual parameters are optimized, changing the sampled values of the fixed CT field.

Substituting Eq. 2 and Eq. 3 into Eq. 1 yields the expected measurements $\bar{y}_{ik}(\Theta_\lambda, \Theta_\phi)$. Let d_{ik} and $\widehat{d}_{ik}$ denote observations and predictions in the same numerical units, over the training index set $\mathcal{M}$. We minimize the stabilized data fidelity

$$\mathcal{L}_\tau(\Theta_\lambda, \Theta_\phi) = \frac{1}{|\mathcal{M}|} \sum_{(i,k)\in\mathcal{M}} \left[\widehat{d}_{ik} - d_{ik}\log(\widehat{d}_{ik} + \epsilon_{\mathrm{d}})\right], \tag{4}$$

where $\epsilon_{\mathrm{d}} > 0$ stabilizes the logarithm. For raw counts, $d = y$ and $\widehat{d} = \bar{y}$; at $\epsilon_{\mathrm{d}} = 0$ this is the mean Poisson negative log-likelihood up to parameter-independent terms. The continuation stage τ is fixed within each epoch and suppressed in the field notation below.

As in emission-based joint reconstruction and registration (Rezaei et al., 2016), differentiable ray integration lets the same objective guide activity and geometry without a PET–CT similarity loss. We structure their updates through the three components below.

3.2 Gauge-Fixed CT Alignment

We model PET–CT misalignment as the composition of a global pose change and a local residual deformation. The global component captures normalized-coordinate rotation and translation, whereas a sinusoidal representation network (SIREN) represents spatially varying displacement (Sitzmann et al., 2020). This follows continuous registration models (Wolterink et al., 2022; Byra et al., 2023).

The residual field can duplicate global translation or infinitesimal rotation. Subtracting its fitted rigid component reduces this redundancy, assigning pose to the global branch and local deformation to the residual.

Let $\mathbf{c} = \mathcal{N}(\mathbf{x})$, $\mathbf{c}_0 = \frac{1}{2}\mathbf{1}$, and let Π clip coordinates componentwise to $[0,1]^3$. The global query and complete physical-coordinate warp are

$$\begin{aligned} \mathbf{r}(\mathbf{c}) &= \Pi(\mathbf{R}(\mathbf{c}-\mathbf{c}_0)+\mathbf{c}_0+\mathbf{T}), \\ \phi_{\Theta_\phi}(\mathbf{x}) &= \mathcal{N}^{-1}\left[\Pi\big(\mathbf{r}(\mathbf{c})+\tfrac{1}{2}\mathbf{v}_\perp(2\mathbf{r}(\mathbf{c})-\mathbf{1})\big)\right]. \end{aligned} \tag{5}$$

Here $\mathbf{R} \in SO(3)$ is parameterized by an axis-angle vector and $\mathbf{T} \in \mathbb{R}^3$ is a normalized translation. The factor $1/2$ converts residual displacements from $[-1,1]^3$ to unit-box units. Before clipping, the physical linear part is $\mathbf{D}\mathbf{R}\mathbf{D}^{-1}$, generally nonorthogonal for unequal box side lengths. Thus "rigid" refers to the normalized global branch, not an arbitrary physical Euclidean motion. Clipping implements boundary queries and can give zero coordinate gradients outside the box.

Let $\mathbf{v}_{\Theta_{\mathrm{nr}}} : [-1,1]^3 \to \mathbb{R}^3$ be the raw SIREN displacement. At $\mathbf{z} \in [-1,1]^3$, define the normalized rigid modes by $\mathbf{B}(\mathbf{z}) = [\mathbf{I}_3, -[\mathbf{z}]_\times]$, where $[\mathbf{z}]_\times \mathbf{u} = \mathbf{z} \times \mathbf{u}$:

$$\mathbf{B}(\mathbf{z})\boldsymbol{\beta} = \mathbf{t} + (\boldsymbol{\omega} \times \mathbf{z}), \qquad \boldsymbol{\beta} = \begin{bmatrix} \mathbf{t} \\ \boldsymbol{\omega} \end{bmatrix} \in \mathbb{R}^6, \tag{6}$$

where $\mathbf{t}$ and $\boldsymbol{\omega}$ describe normalized translation and infinitesimal rotation.

At fixed anchors $\{\mathbf{z}_j\}_{j=1}^J$, we vertically stack the sampled displacements into $\mathbf{v}_A \in \mathbb{R}^{3J}$ and the corresponding rigid-mode matrices into $\mathbf{B}_A \in \mathbb{R}^{3J\times 6}$. The fitted rigid coefficients and corrected residual field are

$$\begin{aligned} \left(\mathbf{B}_A^\top \mathbf{B}_A + \epsilon_{\mathrm{g}} \mathbf{I}_6\right) \widehat{\boldsymbol{\beta}} &= \mathbf{B}_A^\top \mathbf{v}_A, \\ \mathbf{v}_\perp(\mathbf{z}) &= \mathbf{v}_{\Theta_{\mathrm{nr}}}(\mathbf{z}) - \mathbf{B}(\mathbf{z})\widehat{\boldsymbol{\beta}}. \end{aligned} \tag{7}$$

Eq. 7 is a six-dimensional ridge least-squares correction (Hoerl & Kennard, 1970). Geometry gradients propagate through this linear solve.

At the anchors, $\mathbf{B}_A^\top \mathbf{v}_{\perp,A} = \epsilon_{\mathrm{g}} \widehat{\boldsymbol{\beta}}$. Thus $\perp$ denotes approximate orthogonality when $\epsilon_{\mathrm{g}} > 0$; exact removal requires an unregularized full-rank solve. This sampled, first-order correction reduces overlap but does not guarantee global identifiability or invertibility of the clipped warp.

3.3 Polar Activity Representation

We use a nonnegative basis expansion (Lee & Seung, 2000), separating coefficient scale and composition via a magnitude–direction decomposition (Salimans & Kingma, 2016):

$$\begin{aligned} \lambda_{\Theta_\lambda}(\mathbf{x}) &= \rho \sum_{m=1}^{M} \pi_m b_m(\mathbf{c};\tau), \\ \rho = \exp(s), \qquad \pi_m &= \frac{\exp(\eta_m)}{\sum_{j=1}^{M} \exp(\eta_j)}. \end{aligned} \tag{8}$$

Here, the M positive basis functions b_m are the Softplus outputs of a shared decoder. The spatially shared scalar s determines the coefficient magnitude $\rho > 0$, while the logits $\{\eta_m\}_{m=1}^M$ determine the

relative basis weights, with $\pi_m > 0$ and $\sum_m \pi_m = 1$. The decoder receives the gated coordinate encoding described in Sec. 3.4, where $\tau \in [0, 1]$ denotes continuation progress.

We call this representation *polar* because the positive coefficient vector $\boldsymbol{\alpha} = \rho\boldsymbol{\pi}$ is decomposed into an L_1 magnitude, $\rho = \sum_m \alpha_m$, and a direction $\boldsymbol{\pi}$ on the probability simplex. For fixed bases, changing s uniformly rescales the activity, whereas changing the logits modifies its basis composition. Because the bases are also learned, this parameterization does not eliminate basis–coefficient scale ambiguity, and ρ should not be interpreted as total tracer activity. We clamp s to $[-4, 4]$ after coefficient updates; this bounds coefficient magnitude, not the learned bases.

3.4 Multiresolution Continuation

Before the attenuation prior is correctly aligned, an expressive activity field may explain part of the resulting measurement discrepancy through fine spatial details. We therefore restrict the activity representation during early optimization and progressively introduce finer hash-grid levels. This continuation strategy follows coarse-to-fine neural optimization (Lin et al., 2021; Wu et al., 2025).

For an L-level multiresolution hash encoder (Müller et al., 2022), let $\mathbf{h}_\ell(\mathbf{c})$ denote the features at level ℓ, ordered from coarse to fine. The decoder receives

$$\mathbf{h}_\tau(\mathbf{c}) = \mathrm{concat}_{\ell=1}^{L} \left[g_\ell(\tau)\mathbf{h}_\ell(\mathbf{c}) \right]. \tag{9}$$

The level gates follow the raised-cosine schedule

$$g_\ell(\tau) = \frac{1}{2}\left\{1 - \cos\Big[\pi\ \mathrm{clip}_{[0,1]}(L_{\min} + \tau(L - L_{\min}) - \ell + 1)\Big]\right\}. \tag{10}$$

Here, $L_{\min} \in \{1, \ldots, L\}$ is the number of initially active levels. At $\tau = 0$, only the first $L_{\min}$ levels contribute; as τ increases, the remaining levels are activated smoothly from coarse to fine, and all gates equal one at $\tau = 1$. The schedule controls when different spatial resolutions become available to the activity field; it does not impose a strict Fourier-frequency cutoff.

Optimization. We first fit the CT field and freeze its parameters (Algorithm 1). Joint optimization alternates activity and geometry epochs, starting with activity: activity epochs hold geometry fixed, while geometry epochs hold all activity parameters fixed. Within activity epochs, coefficients are updated every mini-batch; scheduled basis updates recompute the loss using the updated coefficients. Continuation advances at every epoch, including geometry epochs, with gates fixed within each epoch. Thus, frozen activity parameters do not imply an unchanged activity field across epoch boundaries during continuation. Algorithm 2 gives the complete procedure.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate 3D activity and attenuation reconstruction on BrainWeb (Cocosco et al., 1997) and XCAT (Segars et al., 2010). BrainWeb tissue labels define paired activity and CT-equivalent attenuation volumes; XCAT provides thoracic anatomies at different respiratory states.

Simulation and Preprocessing. For the main comparisons, activity and attenuation volumes are represented on $128 \times 128 \times 96$ grids with an isotropic voxel size of 1.5 mm. TOF sinograms are generated from spatially aligned activity and attenuation references using PyTomography (Polson et al., 2025) and the Joseph3D TOF forward projector implemented in parallelproj (Schramm & Thielemans, 2024). We use 13 TOF bins and 200 ps timing resolution (full width at half maximum).

To construct PET–CT mismatch, we perturb only the attenuation prior supplied to reconstruction. For BrainWeb, we apply synthetic 3D rigid transformations to the attenuation prior, with translations of up to ± 30 mm along each axis and rotations of up to $\pm 30^\circ$. For XCAT, we pair the end-inhale and end-exhale anatomies generated by its four-dimensional respiratory motion model, using the attenuation prior from a different respiratory phase than the aligned emission reference. This introduces nonrigid mismatch in the lungs, liver, and diaphragm. We simulate two relative dose levels: normal dose (ND) and low dose (LD), with LD using 10% of the ND expected counts. For each

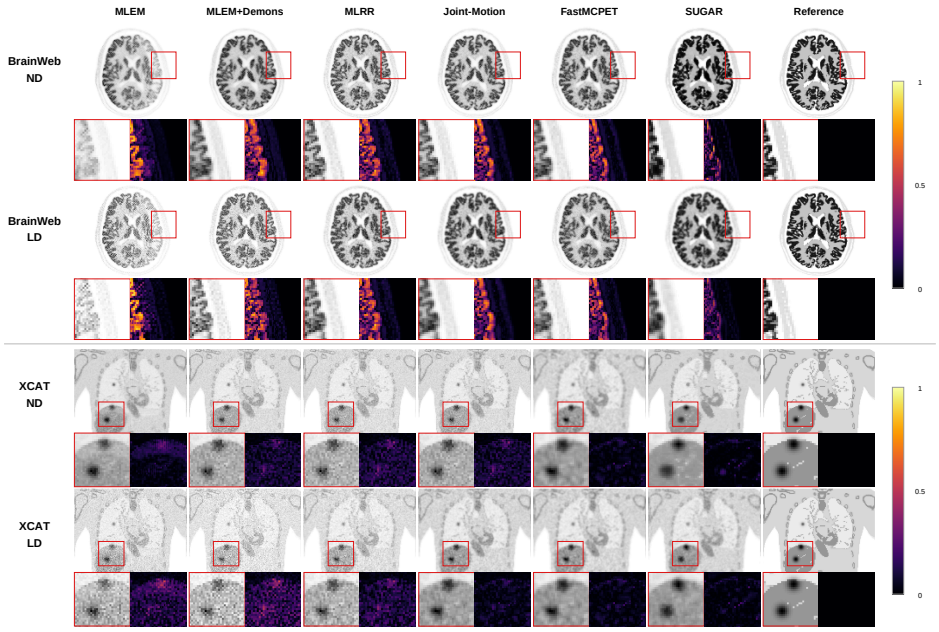


Figure 2: PET reconstruction comparisons on BrainWeb (top) and XCAT (bottom), each under ND and LD. Reference images appear at right. Red boxes mark ROIs; insets show enlarged ROIs (left) and absolute display-intensity differences from the reference (right). Scales are shared across methods and doses within each dataset.

Table 1: PSNR and SSIM comparison of PET reconstruction on BrainWeb and XCAT under ND and LD conditions. Results summarize 10 three-dimensional reconstructions as mean $\pm$ standard deviation. Paired t-tests against SUGAR are marked by ** ($p < 0.01$), * ($p < 0.05$), and $^{\blacktriangledown}$ (not significant, $p \geq 0.05$). Best and second-best values are bold and underlined.

Dataset	Counts	Optim.				Sup.	Unsup.
		TOF-MLEM	MLEM+Demons	TOF-MLRR	Joint-Motion	FastMCPET	SUGAR (Ours)
(a) PSNR (dB) ↑							
BrainWeb	ND	$23.48 \pm 1.76^{**}$	$24.42 \pm 2.13^{**}$	$24.08 \pm 1.59^{**}$	$25.13 \pm 2.41^{*}$	$\underline{25.57 \pm 1.88}^{\blacktriangledown}$	$\mathbf{26.38 \pm 1.67}$
	LD	$21.33 \pm 2.64^{**}$	$22.17 \pm 1.92^{**}$	$22.43 \pm 2.37^{**}$	$23.28 \pm 1.73^{**}$	$\underline{24.02 \pm 2.58}^{**}$	$\mathbf{26.13 \pm 2.46}$
XCAT	ND	$31.72 \pm 1.43^{**}$	$32.08 \pm 2.09^{**}$	$30.93 \pm 1.81^{**}$	$32.37 \pm 1.56^{**}$	$\underline{32.82 \pm 2.32}^{*}$	$\mathbf{34.23 \pm 2.17}$
	LD	$29.97 \pm 2.72^{**}$	$30.43 \pm 2.18^{**}$	$28.98 \pm 2.53^{**}$	$31.12 \pm 1.87^{**}$	$\underline{31.77 \pm 2.39}^{**}$	$\mathbf{33.42 \pm 2.04}$
(b) SSIM ↑							
BrainWeb	ND	$0.941 \pm 0.010^{**}$	$0.957 \pm 0.009^{**}$	$0.952 \pm 0.009^{**}$	$0.960 \pm 0.008^{**}$	$\underline{0.962 \pm 0.008}^{*}$	$\mathbf{0.967 \pm 0.008}$
	LD	$0.872 \pm 0.014^{**}$	$0.915 \pm 0.011^{**}$	$0.910 \pm 0.012^{**}$	$0.932 \pm 0.010^{**}$	$\underline{0.940 \pm 0.009}^{**}$	$\mathbf{0.960 \pm 0.008}$
XCAT	ND	$0.935 \pm 0.010^{**}$	$0.938 \pm 0.011^{**}$	$0.923 \pm 0.010^{**}$	$0.936 \pm 0.009^{**}$	$\underline{0.939 \pm 0.009}^{**}$	$\mathbf{0.964 \pm 0.006}$
	LD	$0.871 \pm 0.012^{**}$	$0.882 \pm 0.012^{**}$	$0.869 \pm 0.012^{**}$	$0.915 \pm 0.011^{**}$	$\underline{0.925 \pm 0.010}^{**}$	$\mathbf{0.946 \pm 0.009}$

dose level, the noise-free expected TOF sinogram is scaled by the corresponding count factor, after which Poisson noise is sampled independently for each LOR–TOF-bin pair. All methods use the same noisy measurements for a given case and dose level.

Baselines and Metrics. We compare fixed-attenuation TOF maximum-likelihood expectation-maximization (TOF-MLEM), MLEM+Demons, TOF-MLRR (Rezaei et al., 2016), and static Joint-Motion (Tsai & Liu, 2023) and FastMCPET (Zhou et al., 2023) adaptations.

PET quality is evaluated on reconstructed 3D volumes using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) (Wang et al., 2004). PSNR uses volumetric mean squared error (MSE), not averaged slice-wise PSNR. Main-table summaries use 10 three-dimensional reconstruc-

tions; paired t-tests compare each baseline with SUGAR on the same cases. Component analyses and appendix diagnostics assess alignment and optimization; slices serve only as illustrations.

Optimization. SUGAR alternates activity and geometry updates for approximately 40 epochs per scan without external pretraining. All four activity hash levels are available after four epochs. Appendix A.2.1 specifies architectures, sampling, optimization, and hardware.

4.2 Main Results

Reconstruction quality. SUGAR achieves the highest mean PET PSNR and SSIM in all four settings (Table 1). On BrainWeb, SUGAR obtains mean PSNR values of 26.38 and 26.13 dB under normal-dose and low-dose conditions, respectively, exceeding the strongest competing baseline by 0.81 and 2.11 dB. The corresponding SSIM values reach 0.967 and 0.960. On XCAT, SUGAR achieves 34.23 and 33.42 dB, improving upon the strongest competing baseline by 1.41 and 1.65 dB, respectively. Both datasets show a larger PSNR advantage under low counts.

Effect of reduced counts. Reducing expected counts to 10% of ND lowers SUGAR's mean PSNR by 0.25 dB on BrainWeb and 0.81 dB on XCAT, compared with 1.55 and 1.05 dB for FastMCPET. SUGAR retains the highest mean LD SSIM, 0.960 and 0.946, respectively, under reduced counts.

Table 2: CRC (%) for three XCAT tumors. Ideal: 100%.

Method	ND	LD
MLEM	86.7	74.8
MLEM+Demons	82.0	71.4
MLRR	79.1	77.6
Joint-Motion	82.2	79.1
FastMCPET	89.0	87.8
SUGAR	**91.7**	**88.5**

Comparison with alignment baselines. TOF-MLEM holds attenuation fixed; the other baselines include sequential registration, joint optimization, and learning-assisted adaptation. Under LD, SUGAR exceeds Joint-Motion by 2.85 dB on BrainWeb and 2.30 dB on XCAT. These datasets test global pose and nonrigid respiratory mismatch, respectively. The BrainWeb ND PSNR margin over FastMCPET is not significant in the paired test; LD PSNR margins over every baseline are significant on both datasets. This supports global–residual geometry.

Visual comparison and tumor contrast. Figure 2 shows clearer BrainWeb tissue boundaries with SUGAR; enlarged ROIs highlight local differences. For XCAT tumors, the contrast recovery coefficient (CRC) measures recovered tumor-to-background contrast relative to ground truth. SUGAR achieves CRCs of 91.7% (ND) and 88.5% (LD), closest to 100% among the compared methods (Table 2), exceeding FastMCPET by 2.7 and 0.7 percentage points. CRC complements volumetric PSNR by assessing uptake contrast in the selected lesions.

4.3 Ablation and Component Analyses

Volumetric reconstruction. We evaluate all eight combinations of gauge correction (G), polar coefficients (P), and continuation (C) on XCAT ND. Table 3 uses a different XCAT phantom configuration from the main comparison. Base retains residual geometry without gauge correction, uses direct-positive coefficients, and activates all hash levels initially.

Adding G, P, or C alone improves PET PSNR over Base by 0.36, 0.23, or 0.27 dB. G+C and P+C reach 33.43 and 33.55 dB; adding C to G or P therefore gives 0.85 or 1.10 dB, exceeding its isolated gain. This pattern suggests that continuation complements both parameterizations. Full SUGAR reaches 33.97 dB, gaining 1.75 dB over Base and 0.42 dB over the best pair. Removing G, P, or C from the full model costs 0.42, 0.54, or 1.23 dB. All three components contribute in the complete configuration. The controlled tests below examine their optimization behavior.

Table 3: XCAT ND PET PSNR (dB). G: gauge correction; P: polar coefficients; C: continuation.

G	P	C	PSNR ↑
–	–	–	32.22
✓	–	–	32.58
–	✓	–	32.45
–	–	✓	32.49
✓	✓	–	32.74
✓	–	✓	33.43
–	✓	✓	33.55
✓	✓	✓	**33.97**

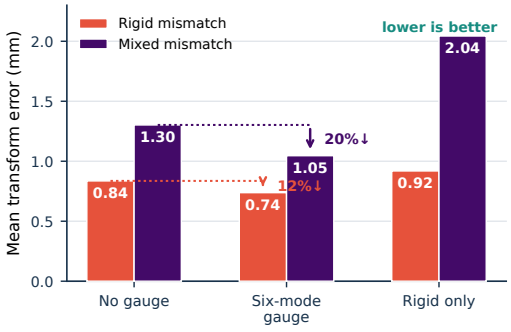


Figure 3: Gauge correction reduces mean map error under rigid and mixed perturbations on BrainWeb.

Gauge correction: geometry recovery. We compare no gauge, six-mode gauge, and rigid-only alignment on BrainWeb with a known 5° rotation, alone or combined with a smooth 3-mm warp. Activity representation, initialization, and LOR streams are shared. Mean map error averages Euclidean distances between estimated and known coordinate maps. Figure 3 shows that gauge correction reduces mean map error from 0.836 to 0.738 mm under rigid mismatch and from 1.302 to 1.046 mm under mixed mismatch. Rigid-only alignment reaches 2.044 mm in the mixed case. PET PSNR differs by at most 0.016 dB within either perturbation, isolating a coordinate-recovery benefit. Appendix A.3.1 gives the protocol and additional diagnostics.

The nearly identical PET scores show why image quality alone is insufficient to assess geometry: coordinate errors reveal differences that reconstruction metrics can obscure.

Polar coefficients: optimization speed. We freeze eight positive basis projections and compare direct-positive and polar coefficients across four target magnitudes, with 30 runs per magnitude (three mixtures and ten Poisson seeds). Both start from the same coefficients and use Adam with shared settings for 1,200 iterations. We define t_{99} as the first iteration attaining 99% of each run's total loss decrease within that budget. Figure 4 shows 34–57% fewer iterations on average with polar coefficients. This supports faster relative objective reduction; composition accuracy remains scale dependent (Appendix A.3.2).

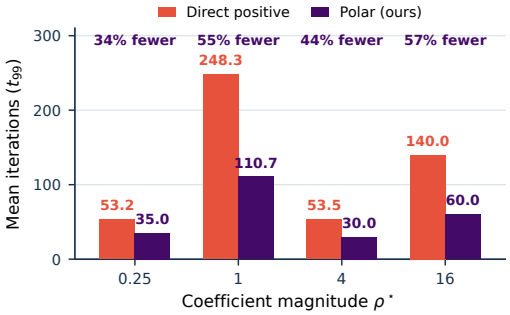


Figure 4: Fixed-basis convergence (30 runs per scale). Polar coefficients reduce mean t_{99} by 34–57%.

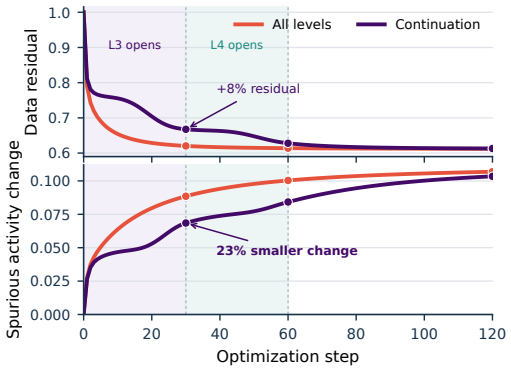


Figure 5: Continuation in a local TOF proxy. Staging limits spurious activity change at step 30 despite a higher data residual; both methods reach similar residuals with all levels active.

Continuation: early activity compensation. We isolate activity compensation using a two-dimensional linearized TOF proxy in which only attenuation geometry is perturbed. Four nested spatial spaces model coarse-to-fine activity capacity. With geometry held fixed, activity updates can reduce the resulting measurement residual by changing an otherwise correct activity distribution. Spurious activity change is the correction's ℓ_2 norm divided by the reference activity norm.

Figure 5 compares staged activation with all levels active initially. At step 30, continuation reduces spurious activity change from 0.089 to 0.068, with a data residual of 0.668 versus 0.620 (8% higher). This gap closes after all levels open: continuation limits early compensation while retaining later fitting capacity (Appendix A.3.3).

These tests isolate geometry recovery (G), relative convergence speed (P), and early compensation control (C). Table 3 measures their combined benefit: Full SUGAR has the highest PSNR. The diagnostics use distinct targets and protocols; their gains are not directly comparable. The factorial ablation tests all components together and links their separate roles to reconstruction quality.

4.4 Evolution of CT Attenuation Errors

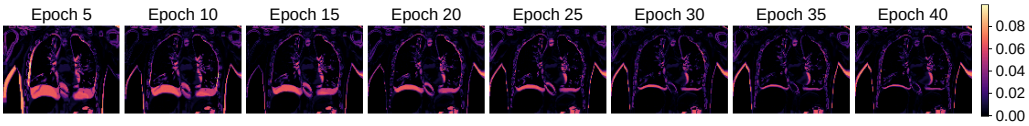


Figure 6: CT attenuation-error maps at epochs 5–40. All snapshots use one shared color scale, shown at right, to compare the spatial error patterns over training.

Figure 6 shows attenuation-error bands near the lung bases and diaphragm becoming narrower and less intense toward epochs 30–40, while localized peripheral residuals remain. The fixed color scale allows changes to be followed at the same locations across epochs. These maps complement the controlled geometry test by showing where attenuation errors persist during joint reconstruction. The spatial views distinguish concentrated boundary errors from changes spread across the volume, adding anatomical context to the aggregate reconstruction metrics.

4.5 Exploratory Extension to MR-Derived Priors

We also test whether emission-driven geometry refinement can improve an MR-derived attenuation prior. A pretrained MR-to-synthetic-CT (sCT) generator supplies a frozen prior; reconstruction then updates only activity and geometry. In a 14-volume validation study with noiseless simulated TOF measurements, geometry refinement raises PET PSNR from 23.53 to 24.12 dB without imposed mismatch. Under a known 5° rotation and $(+6, -4, +4)$-mm translation, it raises PSNR from 20.29 to 23.31 dB and reduces mean geometric endpoint error from 11.72 to 7.91 mm. These results support the feasibility of adapting an MR-derived prior without updating its generator. The larger improvement under imposed mismatch is accompanied by reduced endpoint error, linking the activity gain to improved spatial alignment in this setting. The extension changes the source of the prior while retaining the same emission-driven reconstruction principle. Appendix A.6 reports the protocol, paired results, and limitations of this exploratory validation study.

5 Conclusion

SUGAR jointly reconstructs PET activity and aligns a frozen CT attenuation field through a differentiable TOF model. Its structure gives global pose, residual deformation, activity magnitude, and spatial detail distinct roles within one emission objective. BrainWeb and XCAT comparisons show improved PET reconstruction at both count levels, accompanied by better XCAT tumor contrast recovery. Controlled analyses separately support rigid-mode suppression, faster coefficient optimization, and reduced early activity compensation. The XCAT ND ablation gains 1.75 dB over Base with all three components, connecting these mechanisms to reconstruction quality. External-template and PET/MR studies further explore alternative attenuation priors while retaining the same joint reconstruction principle.

The combined evidence connects end-to-end reconstruction quality with the behavior of individual optimization components. The factorial ablation tests their contribution within the complete model, while controlled diagnostics distinguish coordinate recovery, coefficient convergence, and early activity compensation. This separation is useful because activity quality and geometric accuracy need not change together. SUGAR therefore provides both a reconstruction framework and a way to examine how activity and geometry account for the measured signal, supporting explicit structure in coupled activity–geometry optimization.

Limitations. SUGAR requires per-scan optimization and depends on initialization and prior quality; geometric alignment cannot correct missing anatomy or attenuation-value errors. Sampled gauge correction does not guarantee global identifiability or invertible deformations. Numerically scaled projections and omitted random/scatter terms limit the raw-count likelihood interpretation. Evaluation uses phantoms and simulated PET/MR measurements, so performance on acquired clinical emission data remains to be established. Future work will examine activity-basis capacity and initialization sensitivity, incorporate background models and deformation constraints, and assess reconstruction time and robustness across acquisition settings.

AI USE STATEMENT

AI tools assisted with literature search, experiment planning, interpretation of results, manuscript editing, figure preparation, and consistency checks. The authors take responsibility for the final text, code, figures, and scientific claims.

ETHICS STATEMENT

The main experiments use BrainWeb and XCAT digital phantoms with simulated PET measurements. The exploratory PET/MR extension uses existing paired MR/CT volumes to construct attenuation priors and evaluate reconstruction with simulated emission data.

REPRODUCIBILITY STATEMENT

Section 3 defines the representation and objective; Algorithms 1 and 2 specify CT-field fitting and alternating optimization. Section 4.1 reports simulation and evaluation settings; Appendix A.2.1 gives network and optimizer configurations. Appendix A.3 describes controlled module checks, Appendix A.4 explains baseline adaptations, and Appendices A.5 and A.6 detail the exploratory extension protocols.

REFERENCES

Hossein Arabi, Azadeh AkhavanAllaf, Amirhossein Sanaat, Isaac Shiri, and Habib Zaidi. The promise of artificial intelligence and deep learning in PET and SPECT imaging. *Physica Medica*, 83:122–137, 2021.

A. Bousse, O. Bertolli, D. Atkinson, S. Arridge, S. Ourselin, B. F. Hutton, and K. Thielemans. Maximum-likelihood joint image reconstruction/motion estimation in attenuation-corrected respiratory gated PET/CT using a single attenuation map. *IEEE Transactions on Medical Imaging*, 35(1):217–228, 2016.

Michal Byra, Charissa Poon, Muhammad Febrian Rachmadi, Matthias Schlachter, and Henrik Skibbe. Exploring the performance of implicit neural representations for brain image registration. *Scientific Reports*, 13:17334, 2023.

Chris A. Cocosco, Vasken Kollokian, Remi K.-S. Kwan, and Alan C. Evans. BrainWeb: Online interface to a 3D MRI simulated brain database. *NeuroImage*, 5(4):S425, 1997.

Masoud Elhamiasl, Frederic Jolivet, Ahmadreza Rezaei, Michael Fieseler, Klaus Schäfers, Johan Nuyts, Georg Schramm, and Fernando Boada. Joint estimation of activity, attenuation and motion in respiratory-self-gated time-of-flight PET. *Physics in Medicine & Biology*, 70(7):075003, 2025.

Kailong Fan, Yubo Ye, Huafeng Liu, and Linwei Wang. IMREPET: Implicit neural representation for unsupervised dynamic PET reconstruction. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, volume 15961, pp. 271–281. Springer Nature Switzerland, 2025.

Xiao Han. MR-based synthetic CT generation using a deep convolutional neural network method. *Medical Physics*, 44(4):1408–1419, 2017.

Arthur E. Hoerl and Robert W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

P. E. Kinahan, D. W. Townsend, T. Beyer, and D. Sashin. Attenuation correction for a combined 3D PET/CT scanner. *Medical Physics*, 25(10):2046–2053, 1998.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, volume 13, 2000.

Tiantian Li, Mengxi Zhang, Wenyuan Qi, Evren Asma, and Jinyi Qi. Deep learning based joint PET image reconstruction and motion estimation. *IEEE Transactions on Medical Imaging*, 41 (5):1230–1241, 2022.

Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. BARF: Bundle-adjusting neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5741–5751, 2021.

Yuxuan Long, Yulin Zhang, Hong Wang, Xiaodong Kuang, Hailiang Huang, Fan Rao, Huafeng Liu, Yefeng Zheng, and Wentao Zhu. PD-INR: Prior-driven implicit neural representations for TOF-PET reconstruction. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, volume 15962, pp. 459–469. Springer Nature Switzerland, 2025.

William W. Moses. Time of flight in PET revisited. *IEEE Transactions on Nuclear Science*, 50(5): 1325–1330, 2003.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4):102:1–102:15, 2022.

Lucas A. Polson, Roberto Fedrigo, Chenguang Li, Maziar Sabouri, Obed Dzikunu, Shadab Ahamed, Nicolas Karakatsanis, Sara Kurkowska, Peyman Sheikhzadeh, Pedro Esquinas, Arman Rahmim, and Carlos Uribe. PyTomography: A Python library for medical image reconstruction. *SoftwareX*, 29:102020, 2025.

Sashank J. Reddi, Satyen Kale, and Sanjiv Kumar. On the convergence of Adam and beyond. In *International Conference on Learning Representations*, 2018.

Ahmadreza Rezaei, Michel Defrise, Girish Bal, Christian Michel, Maurizio Conti, Charles Watson, and Johan Nuyts. Simultaneous reconstruction of activity and attenuation in time-of-flight PET. *IEEE Transactions on Medical Imaging*, 31(12):2224–2233, 2012.

Ahmadreza Rezaei, Christian Michel, Michael E. Casey, and Johan Nuyts. Simultaneous reconstruction of the activity image and registration of the CT image in TOF-PET. *Physics in Medicine & Biology*, 61(4):1852–1874, 2016.

Tim Salimans and Diederik P. Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Joshua Schaefferkoetter, Vijay Shah, Charles Hayden, John O. Prior, and Sven Zuehlsdorff. Deep learning for improving PET/CT attenuation correction by elastic registration of anatomical data. *European Journal of Nuclear Medicine and Molecular Imaging*, 50(8):2292–2304, 2023.

Georg Schramm and Kris Thielemans. PARALLELPROJ—an open-source framework for fast calculation of projections in tomography. *Frontiers in Nuclear Medicine*, 3:1324562, 2024.

W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui. 4D XCAT phantom for multimodality imaging research. *Medical Physics*, 37(9):4902–4915, 2010.

L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1(2):113–122, 1982.

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems*, volume 33, pp. 7462–7473, 2020.

Suleman Surti. Update on time-of-flight PET imaging. *Journal of Nuclear Medicine*, 56(1):98–105, 2015.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*, volume 33, pp. 7537–7547, 2020.

J.-P. Thirion. Image matching as a diffusion process: An analogy with Maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998.

Yu-Jung Tsai and Chi Liu. Joint motion estimation and penalized image reconstruction algorithm with anatomical priors for gated TOF-PET/CT. *Physics in Medicine & Biology*, 68(2):025020, 2023.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004.

Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

Jelmer M. Wolterink, Jesse C. Zwienenberg, and Christoph Brune. Implicit neural representations for deformable image registration. In *Proceedings of the 5th International Conference on Medical Imaging with Deep Learning*, volume 172 of *Proceedings of Machine Learning Research*, pp. 1349–1359. PMLR, 2022.

Qing Wu, Chenhe Du, Xuanyu Tian, Jingyi Yu, Yuyao Zhang, and Hongjiang Wei. Moner: Motion correction in undersampled radial MRI with unsupervised neural representation. In *The Thirteenth International Conference on Learning Representations*, 2025.

Bo Zhou, Yu-Jung Tsai, Jiazhen Zhang, Xueqi Guo, Huidong Xie, Xiongchao Chen, Tianshun Miao, Yihuan Lu, James S. Duncan, and Chi Liu. Fast-MC-PET: A novel deep learning-aided motion correction and reconstruction framework for accelerated PET. In *Information Processing in Medical Imaging*, volume 13939 of *Lecture Notes in Computer Science*, pp. 523–535. Springer, 2023.

A Appendix

A.1 Principles of TOF-PET Imaging

PET visualizes the spatial distribution of a radiotracer by detecting pairs of photons produced by positron–electron annihilation. The two photons, each with an energy of approximately 511 keV, travel in nearly opposite directions. Their coincident detection defines an LOR between the two detectors, constraining the annihilation location to that line.

As illustrated in Fig. 7, TOF-PET additionally uses the difference in photon arrival times to localize the event along the LOR. Let $\Delta\tau_{ik}$ denote the measured arrival-time difference associated with TOF bin k on LOR i. Its corresponding spatial position relative to the LOR midpoint is

$$s_{ik} = \frac{c_{\text{light}}\Delta\tau_{ik}}{2}, \tag{11}$$

where c_{light} is the speed of light. The factor $1/2$ appears because moving the annihilation point by a distance s shortens one photon path and lengthens the other by the same amount.

Timing uncertainty prevents exact event localization. The TOF response along an LOR is commonly approximated by a Gaussian kernel (Moses, 2003; Surti, 2015):

$$h_i(s - s_{ik}) = \frac{1}{\sqrt{2\pi}\sigma_{s,i}} \exp\left[-\frac{(s - s_{ik})^2}{2\sigma_{s,i}^2}\right], \qquad \sigma_{s,i} = \frac{c_{\text{light}}\sigma_{\tau,i}}{2}, \tag{12}$$

where $\sigma_{\tau,i}$ and $\sigma_{s,i}$ denote the timing and spatial standard deviations, respectively. Equivalently, a coincidence timing resolution $\Delta\tau_{\text{FWHM}}$ corresponds to the spatial FWHM

$$\Delta s_{\text{FWHM}} = \frac{c_{\text{light}}\Delta\tau_{\text{FWHM}}}{2}.$$

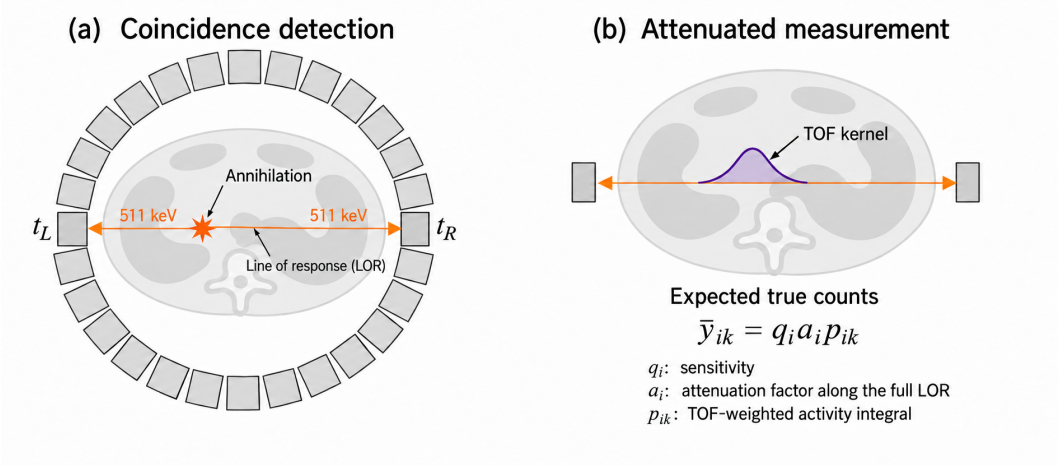


Figure 7: Illustration of TOF-PET imaging. (a) Coincident detection of two approximately 511 keV photons defines a line of response. (b) The TOF kernel describes localization uncertainty along the LOR, while tissue attenuation modifies the probability of detecting an unscattered photon pair. The displayed expected true-count component is $q_i a_i(\mu) p_{ik}(\lambda)$; additive random and scattered coincidences are omitted from the schematic for clarity.

In a binned acquisition, the probability assigned to a TOF bin is obtained by integrating the response over the corresponding bin interval. Improved timing resolution produces a narrower kernel and therefore more precise localization along the LOR.

Measurement model with alignment. Let $\Omega \subset \mathbb{R}^3$ be the reconstruction box within the scanner field of view, with physical coordinates $\mathbf{x} \in \Omega$ and activity $\lambda(\mathbf{x}) \geq 0$. Let $\widetilde{\mu}_{\mathrm{CT}}$ denote the CT-derived attenuation field at the PET photon energy. A transformation ϕ maps reconstruction coordinates to CT coordinates, giving the aligned field $\mu_\phi(\mathbf{x}) = \widetilde{\mu}_{\mathrm{CT}}(\phi(\mathbf{x}))$. CT-derived attenuation correction accounts for photon loss along coincidence detection paths (Kinahan et al., 1998).

For LOR i, let $\mathbf{x}_i(t) = \mathbf{s}_i + t\mathbf{d}_i$, where $\mathbf{s}_i$ is a reference point, $\mathbf{d}_i$ is a unit direction, and $\mathcal{T}_i$ is the segment inside the reconstruction domain. The expected counts follow Eq. 1, whose two marked factors are attenuation and TOF-weighted activity projection. Here, $q_i > 0$ is a known LOR-dependent sensitivity factor, b_{ik} denotes the expected additive contribution from random and scattered coincidences, h_i is the TOF response kernel that weights emission locations along LOR i, and t_{ik} denotes the center of TOF bin k in the corresponding TOF coordinate.

TOF information localizes emission along each LOR (Surti, 2015), whereas attenuation is shared across its TOF bins. CT misalignment therefore changes the multiplicative factor applied to the activity projection and can bias the reconstructed PET image. Although TOF improves joint activity–attenuation estimation, unconstrained attenuation retains a scale ambiguity (Rezaei et al., 2012). A CT prior constrains attenuation, while estimating its alignment within the forward model allows emission data to guide both reconstruction and registration (Rezaei et al., 2016).

PET counts are modeled as independent Poisson variables (Shepp & Vardi, 1982). Equation 1 additionally includes sensitivity and additive background. For a Poisson variable, variance equals its mean. Lower counts increase relative noise and complicate reconstruction.

The Poisson interpretation in Eq. 1 applies to raw count data. When observed and predicted projections are numerically scaled, the corresponding objective is treated as a Poisson-form data-fidelity term rather than an exact raw-count likelihood.

A.2 Optimization and CT-Field Validation

We report the main reconstruction settings and optimization procedure, then validate the fitted CT field used as the frozen attenuation prior.

A.2.1 IMPLEMENTATION OF THE MAIN RECONSTRUCTION EXPERIMENTS

The activity field uses a four-level multiresolution hash encoder (Müller et al., 2022), with two feature channels per level, a base resolution of 16, and at most 2^{16} hash-table entries per level. A four-layer MLP with hidden width 32 maps the encoded coordinates to eight Softplus outputs, forming the positive neural bases. The CT-field decoder uses ReLU activations; its fitting and activation comparison are detailed in Appendix A.2.3. The residual displacement field uses a SIREN (Sitzmann et al., 2020) with one hidden layer of width 16. Gauge correction uses a fixed $6 \times 6 \times 6$ anchor grid. Continuation starts with two active hash levels and activates the rest over four epochs.

We use 256 integration samples per LOR and mini-batches of 32,768 LORs. Optimization uses Adam (Kingma & Ba, 2015) with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and AMSGrad (Reddi et al., 2018) enabled. Learning rates are 2×10^{-3} for the activity encoder and basis decoder, 5×10^{-3} for the log-amplitude, 1×10^{-2} for the mixture logits, and 2×10^{-3} for the global transformation and residual displacement parameters. During activity epochs, coefficients are updated for every mini-batch, while the encoder and basis decoder are updated every two global training steps. SUGAR is typically optimized for approximately 40 epochs with constant learning rates and evaluation after each epoch. Experiments use NVIDIA Tesla V100 accelerators with 32 GB of memory.

A.2.2 OPTIMIZATION PROCEDURE

SUGAR first fits the CT field (Algorithm 1), then alternates emission-domain updates (Algorithm 2).

Algorithm 1 CT-field fitting

Require: CT samples $\{(\mathbf{x}_n, \widetilde{\mu}_n)\}_{n=1}^{N_\mu}$; iterations T_μ
Ensure: Frozen field $C_{\widehat{\Theta}_\mu}$

```
Initialize Θ_μ
for t = 1, ..., T_μ do
    B_μ ← Sample({1, ..., N_μ})
    L_CT ← 1/|B_μ| Σ_{n∈B_μ} |C_{Θ_μ}(x_n) − μ̃_n|^2
    Θ_μ ← Adam(Θ_μ, ∇_{Θ_μ} L_CT)
end for
Θ̂_μ ← stopgrad(Θ_μ)
```

Algorithm 1 uses no emission data. Here $C(\mathbf{x}_n) = \mathcal{C}(\mathcal{N}(\mathbf{x}_n))$, so physical samples are normalized exactly once. In Algorithm 2, $C_{\widehat{\Theta}_\mu}$ has frozen weights; geometry gradients pass through its query coordinates and the gauge solve. The counter q counts mini-batches across both update phases. Geometry epochs evaluate the frozen activity network with the current epoch's gates.

A.2.3 CT-FIELD FITTING AND TRANSFORMATION CONSISTENCY

We evaluate CT-field fitting on a BrainWeb attenuation volume ($128 \times 128 \times 96$) before freezing the field for joint reconstruction. The canonical field has six hash-grid levels, two features per level, base resolution 16, a maximum hash-table size of 2^{19}, and a four-layer MLP of width 32 with ReLU hidden activations and a ReLU output. Here $\mathrm{ReLU}(u) = \max(0, u)$ ensures nonnegative attenuation at the output. All variants retain this output activation and use Adam with learning rate 10^{-3} for 10,000 fitting steps. Each step samples every voxel once with independent coordinate jitter of at most half a voxel; complete-volume PSNR and SSIM are evaluated on the fixed grid every 100 steps. We vary one factor at a time: hidden activation, features per level, MLP depth, hidden width, and number of hash levels. Unless the random seed is being tested, all runs use seed 2024.

Figure 8 shows a rapid initial improvement followed by slower gains. The canonical field reaches 39.50 dB and 0.9924 SSIM at step 10,000. Its three-seed endpoint PSNR is 38.84 ± 0.64 dB. Despite high endpoint fidelity, higher-capacity variants show occasional sharp drops.

The hidden-activation comparison supports our use of ReLU over sigmoid linear unit (SiLU) and Softplus activations. Separately, four features per level, a deeper MLP, width 16, and additional hash

Algorithm 2 Alternating joint reconstruction and CT alignment

Require: Scaled measurements $\{d_{ik}\}_{(i,k)\in\mathcal{M}}$; $C_{\widehat{\Theta}_\mu}$; epochs E; warm-up $E_{\mathrm{w}} \geq 2$; basis-update interval K

Ensure: $\widehat{\lambda}$ and $\widehat{\mu}$

$\Theta_\lambda \leftarrow (\Theta_{\mathrm{enc}}, \Theta_{\mathrm{dec}}, s, \boldsymbol{\eta})$; $s \leftarrow 0$, $\boldsymbol{\eta} \leftarrow \mathbf{0}$, $q \leftarrow 0$
$\Theta_\phi \leftarrow (\Theta_{\mathrm{rig}}, \Theta_{\mathrm{nr}})$; $\Theta_{\mathrm{rig}} \leftarrow \mathrm{Identity}$, $\Theta_{\mathrm{nr}} \leftarrow \mathrm{InitNearZero}$
for $e = 1, \ldots, E$ **do**
  $\tau_e \leftarrow \min\left(1, \dfrac{e-1}{E_{\mathrm{w}}-1}\right)$; set $\{g_\ell(\tau_e)\}_{\ell=1}^{L}$ for this epoch
  if e is odd **then** ▷ activity update
    for mini-batch $\mathcal{B} \subset \mathcal{M}$ **do**
      $\lambda_{\Theta_\lambda}(\mathbf{x}) \leftarrow e^{s} \sum_{m=1}^{M} \mathrm{softmax}(\boldsymbol{\eta})_m b_m(\mathbf{c}; \tau_e)$, $\mathbf{c} = \mathcal{N}(\mathbf{x})$
      $\mu_{\Theta_\phi}(\mathbf{x}) \leftarrow C_{\widehat{\Theta}_\mu}\left(\phi_{\Theta_\phi}(\mathbf{x})\right)$
      $\widehat{d}_{ik} \leftarrow \mathrm{TOFForward}_{ik}\left(\lambda_{\Theta_\lambda}, \mu_{\Theta_\phi}\right)$, $(i,k) \in \mathcal{B}$
      $\mathcal{L}_{\mathcal{B}} \leftarrow \dfrac{1}{|\mathcal{B}|} \sum_{(i,k)\in\mathcal{B}} \left[\widehat{d}_{ik} - d_{ik} \log\left(\widehat{d}_{ik} + \epsilon_{\mathrm{d}}\right)\right]$
      $(s, \boldsymbol{\eta}) \leftarrow \mathrm{Adam}_c\left(s, \boldsymbol{\eta}; \nabla_{s,\boldsymbol{\eta}} \mathcal{L}_{\mathcal{B}}\right)$
      $s \leftarrow \mathrm{clip}(s, -4, 4)$
      if $q \bmod K = 0$ **then**
        Recompute $\widehat{d}_{ik}$ and $\mathcal{L}_{\mathcal{B}}$ with updated $(s, \boldsymbol{\eta})$ held fixed
        $(\Theta_{\mathrm{enc}}, \Theta_{\mathrm{dec}}) \leftarrow \mathrm{Adam}_b\left(\Theta_{\mathrm{enc}}, \Theta_{\mathrm{dec}}; \nabla_{\Theta_{\mathrm{enc}}, \Theta_{\mathrm{dec}}} \mathcal{L}_{\mathcal{B}}\right)$
      end if
      $q \leftarrow q + 1$
    end for
  else ▷ geometry update
    for mini-batch $\mathcal{B} \subset \mathcal{M}$ **do**
      $\widehat{\boldsymbol{\beta}} \leftarrow \arg\min_{\boldsymbol{\beta}} \sum_{j=1}^{J} \|\mathbf{v}_{\Theta_{\mathrm{nr}}}(\mathbf{z}_j) - \mathbf{B}(\mathbf{z}_j)\boldsymbol{\beta}\|_2^2 + \epsilon_{\mathrm{g}} \|\boldsymbol{\beta}\|_2^2$
      $\mathbf{v}_\perp(\mathbf{z}) \leftarrow \mathbf{v}_{\Theta_{\mathrm{nr}}}(\mathbf{z}) - \mathbf{B}(\mathbf{z})\widehat{\boldsymbol{\beta}}$
      $\mu_{\Theta_\phi}(\mathbf{x}) \leftarrow C_{\widehat{\Theta}_\mu}\left(\phi_{\Theta_\phi}(\mathbf{x}; \mathbf{v}_\perp)\right)$
      Re-evaluate λ_{Θ_λ} at τ_e with Θ_λ fixed; compute $\widehat{d}_{ik}$ and $\mathcal{L}_{\mathcal{B}}$ as above
      $\Theta_\phi \leftarrow \mathrm{Adam}_\phi\left(\Theta_\phi, \nabla_{\Theta_\phi} \mathcal{L}_{\mathcal{B}}\right)$
      $q \leftarrow q + 1$
    end for
  end if
end for
$\widehat{\lambda} \leftarrow \lambda_{\Theta_\lambda}$; $\widehat{\mu} \leftarrow C_{\widehat{\Theta}_\mu} \circ \phi_{\Theta_\phi}$

levels provide favorable fidelity within the tested ranges. These one-factor comparisons identify practical choices for this volume, not a jointly optimal configuration.

Prescribed transformation test. We compare transformed frozen-field queries with the transformed discrete CT. To limit CT-fitting error, this diagnostic uses the best checkpoint from a separate 10,000-step repeat of the four-features-per-level field (step 9,900; 41.85 dB). For a global map $T(\mathbf{x}) = \mathbf{x}_0 + s\mathbf{R}(\mathbf{x} - \mathbf{x}_0) + \mathbf{t}$, we query the transformed field by the pullback $C_{\widehat{\Theta}_\mu}(T^{-1}(\mathbf{x}))$. We test an eight-voxel translation, a 10° in-plane rotation, and isotropic scaling by 0.90. We additionally use the smooth nonlinear source-coordinate offset

$$\boldsymbol{\delta}(\mathbf{c}) = \begin{bmatrix} \frac{6}{n_x - 1} \sin(2\pi c_y) \sin(\pi c_z) \\ \frac{4}{n_y - 1} \sin(2\pi c_x) \sin(\pi c_z) \\ \frac{2}{n_z - 1} \sin(\pi c_x) \sin(\pi c_y) \end{bmatrix}, \qquad \mathbf{c}_{\mathrm{src}} = \mathbf{c} + \boldsymbol{\delta}(\mathbf{c}), \tag{13}$$

whose maximum component amplitudes are six, four, and two voxels. The discrete reference uses identical source coordinates and trilinear interpolation. Metrics are evaluated only where the source query remains inside the original field of view.

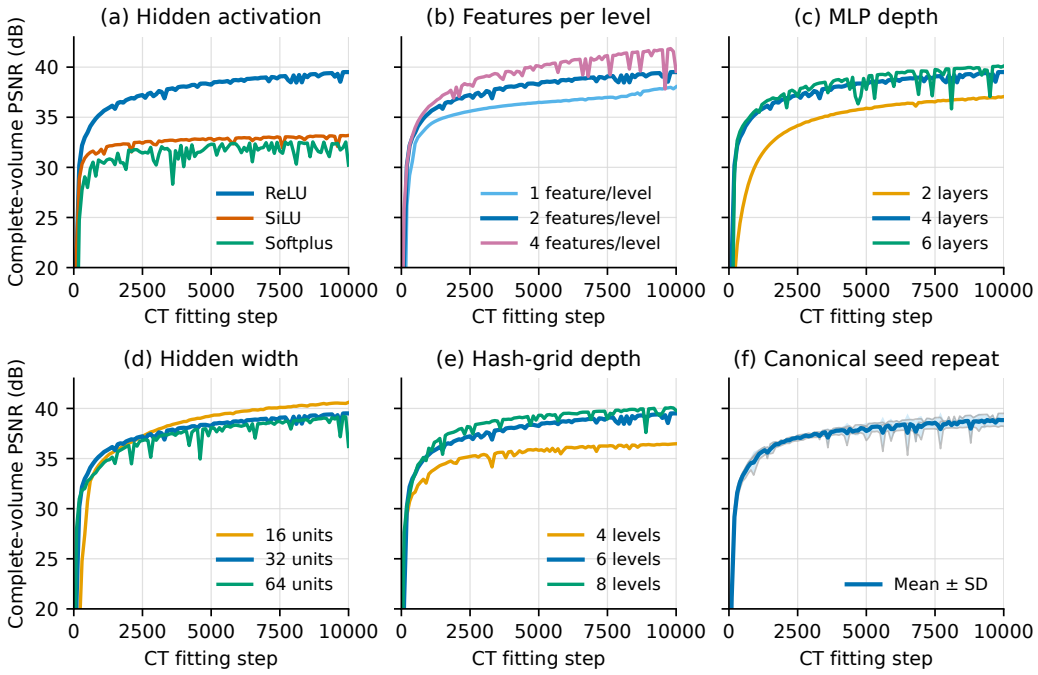


Figure 8: Complete-volume PSNR over 10,000 CT-field fitting steps. We vary (a) hidden activation, (b) hash features per level, (c) MLP depth, (d) hidden width, and (e) number of hash levels. Panel (f) shows three canonical runs and their mean ± sample standard deviation. Curves below 20 dB during the earliest iterations are clipped from the displayed range.

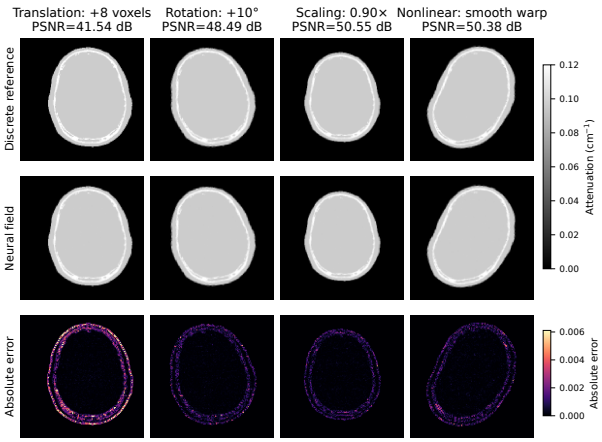


Figure 9: Frozen CT-field queries under prescribed transforms. Rows: transformed CT, neural query, and absolute error. PSNR uses common valid support.

For translation, rotation, scaling, and nonlinear deformation, neural queries reach 41.54, 48.49, 50.55, and 50.38 dB, respectively. Valid-support normalized root-mean-square error (NRMSE) is 0.62%–1.70%.

Boundary residuals reflect field approximation and differences from trilinear interpolation. Scaling and nonlinear warping test queries, without adding global motion parameters.

This validates query direction, normalization, and transform consistency; emission-based deformation recovery is not tested.

A.3 Controlled Analysis of Structured Optimization

A.3.1 Anchor-Space Analysis of the Gauge Correction

We analyze rigid-mode removal at fixed anchors and evaluate known-transform reconstruction.

Proposition (ridge residualization). Let $\mathbf{B}_A \in \mathbb{R}^{3J\times 6}$ have full column rank and let $\mathbf{v}_A \in \mathbb{R}^{3J}$ stack the raw residual displacements. Each of the J anchors contributes three displacement components; the six columns represent three translations and three infinitesimal rotations. Define

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_\epsilon &= (\mathbf{B}_A^\top \mathbf{B}_A + \epsilon \mathbf{I}_6)^{-1}\mathbf{B}_A^\top \mathbf{v}_A,\\ \mathbf{P}_\epsilon &= \mathbf{I} - \mathbf{B}_A(\mathbf{B}_A^\top \mathbf{B}_A + \epsilon \mathbf{I}_6)^{-1}\mathbf{B}_A^\top.\end{aligned} \tag{14}$$

Then $\mathbf{v}_{\perp,A} = \mathbf{P}_\epsilon \mathbf{v}_A$ satisfies

$$\mathbf{B}_A^\top \mathbf{v}_{\perp,A} = \epsilon \widehat{\boldsymbol{\beta}}_\epsilon. \tag{15}$$

For $\epsilon = 0$, $\mathbf{P}_0$ is the orthogonal projector onto $\ker(\mathbf{B}_A^\top)$, and hence $\mathbf{B}_A^\top \mathbf{v}_{\perp,A} = \mathbf{0}$. For $\epsilon > 0$, a rigid singular direction with singular value σ_i is retained with gain $\epsilon/(\sigma_i^2 + \epsilon)$.

Proof. Substitution of Eq. 14 gives

$$\begin{aligned}\mathbf{B}_A^\top \mathbf{P}_\epsilon \mathbf{v}_A &= \left[\mathbf{I}_6 - (\mathbf{B}_A^\top \mathbf{B}_A)(\mathbf{B}_A^\top \mathbf{B}_A + \epsilon \mathbf{I}_6)^{-1}\right]\mathbf{B}_A^\top \mathbf{v}_A\\ &= \epsilon(\mathbf{B}_A^\top \mathbf{B}_A + \epsilon \mathbf{I}_6)^{-1}\mathbf{B}_A^\top \mathbf{v}_A = \epsilon \widehat{\boldsymbol{\beta}}_\epsilon.\end{aligned}$$

When $\epsilon = 0$, direct multiplication shows that $\mathbf{P}_0$ is symmetric and idempotent, while $\mathbf{B}_A^\top \mathbf{P}_0 = \mathbf{0}$. The singular-direction gain follows by inserting the singular value decomposition of $\mathbf{B}_A$. □

Numerical operator check. We reproduced the implemented $6 \times 6 \times 6$ anchor grid ($J = 216$) and rigid-mode construction in double precision. The resulting 648×6 matrix has rank six and Gram eigenvalues between 201.6 and 216.0. At the implemented $\epsilon_{\mathrm{g}} = 10^{-6}$, Eq. 15 holds with maximum absolute residual 1.17×10^{-14}. The maximum rigid-mode gain predicted above is 4.96×10^{-9}; for the tested smooth-plus-rigid displacement, the remaining fitted rigid-component ratio is 4.52×10^{-8}.

Linearized recovery test. We next constructed a known displacement containing translation, infinitesimal rotation, and a smooth residual made exactly orthogonal to the six sampled rigid modes. The raw residual design contains a duplicate of the rigid basis, which models the local redundancy of an expressive neural field. We fit the known displacement by minimum-norm least squares using no correction, translation-only correction, the complete six-mode correction, or no residual branch. Define the anchor-space principal cosine

$$\gamma^A_{\mathrm{rigid,nr}} = \sigma_{\max}(\mathbf{Q}_B^\top \mathbf{Q}_R), \tag{16}$$

where $\mathbf{Q}_B$ and $\mathbf{Q}_R$ are orthonormal bases of the sampled rigid and residual design spaces. This is an unweighted anchor-space measure.

Table 4: Linearized known-field recovery. RMSE denotes root-mean-square error. Coefficient and warp errors are in normalized coordinates. e_t and e_ω denote translation and rotation coefficient errors; Leakage measures rigid content in the residual branch. The uncorrected models reproduce the complete field while assigning part of the rigid motion to the residual branch. Bold marks the smallest reported value in each column.

Variant	$\gamma^A_{\mathrm{rigid,nr}}$	Leakage	e_t	e_ω	Warp RMSE
No gauge	1.000	0.970	6.16×10^{-2}	5.59×10^{-2}	6.40×10^{-17}
Translation only	1.000	0.935	2.52×10^{-16}	5.59×10^{-2}	1.45×10^{-16}
Six-mode gauge	$\mathbf{4.24\times10^{-17}}$	$\mathbf{1.26\times10^{-16}}$	$\mathbf{1.70\times10^{-17}}$	$\mathbf{9.81\times10^{-18}}$	$\mathbf{1.61\times10^{-17}}$
Rigid only	–	–	3.40×10^{-17}	4.60×10^{-17}	1.19×10^{-2}

The principal cosine decreases from one to numerical zero only for the exact six-mode residualization. Translation-only correction removes the translational ambiguity but leaves the rotational subspace duplicated. The rigid-only model recovers the global coefficients but cannot reproduce the orthogonal nonrigid part. With a nonzero ridge, arbitrarily small retained directions affect the principal cosine, making gain and leakage more informative than numerical rank.

Identity control. We set the target displacement to zero and initialized all parameters away from zero. For a linear least-squares objective, gradient flow preserves the component of the initialization in the design null space. Over 100 random initializations, all variants had a mean output root-mean-square (RMS) displacement below 9×10^{-17}. Without the gauge, however, the global and residual rigid branches retained equal and opposite RMS displacements of 0.047 ± 0.016 in normalized units. Translation-only correction retained 0.033 ± 0.014 through the duplicated rotational modes. The six-mode gauge reduced both the global drift and the residual rigid component to numerical precision (below 3×10^{-17} mean RMS). An identity output can thus hide drift in duplicated rigid modes.

Geometry-recovery protocol and additional metrics. Figure 3 in the main text reports mean map errors. The BrainWeb reconstruction experiment applies either a known 5° rotation or that rotation composed with a smooth 3-mm deformation. All compared variants share the activity representation, initialization, and LOR streams. Under the mixed perturbation, mean/95th-percentile map error is $1.046/1.992$ mm with the six-mode gauge, $1.302/2.699$ mm without it, and $2.044/4.052$ mm with rigid-only alignment. Under the rigid perturbation, the corresponding mean errors are 0.738, 0.836, and 0.919 mm. PET PSNR differs by at most 0.016 dB within either perturbation. The no-gauge variant obtains higher global attenuation PSNR in both cases. This supports reduced parameter leakage and better coordinate recovery, but not uniformly improved image metrics.

Translation-only correction was tested only in the linearized experiment; the trained reconstruction comparison contains one phantom and one seed. Anchor-space orthogonality does not establish measurement-space decoupling, global identifiability, or warp invertibility.

A.3.2 Analysis of the Polar Activity Coefficients

First-order decomposition. For fixed positive bases, write the coefficient vector as $\boldsymbol{\alpha} = \rho\boldsymbol{\pi}$, where $\rho = \exp(s)$ and $\boldsymbol{\pi} = \mathrm{softmax}(\boldsymbol{\eta})$. Let

$$\mathbf{J}_\pi = \mathrm{Diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^\top. \tag{17}$$

The coefficient differential is

$$\mathrm{d}\boldsymbol{\alpha} = \boldsymbol{\alpha}\,\mathrm{d}s + \rho\mathbf{J}_\pi\,\mathrm{d}\boldsymbol{\eta}. \tag{18}$$

Because $\mathbf{1}^\top\mathbf{J}_\pi = \mathbf{0}^\top$ and $\mathbf{1}^\top\boldsymbol{\alpha} = \rho$, Eq. 18 implies

$$\mathbf{1}^\top\mathrm{d}\boldsymbol{\alpha} = \rho\,\mathrm{d}s. \tag{19}$$

Thus s changes the L_1 magnitude of the positive coefficient vector, whereas an infinitesimal logit update preserves its coefficient sum. This separates scale and composition without implying tangent orthogonality in coefficient or measurement space.

For a differentiable objective $\mathcal{L}$ and $\mathbf{d} = \partial\mathcal{L}/\partial\boldsymbol{\alpha}$, the corresponding gradients are

$$\frac{\partial\mathcal{L}}{\partial s} = \boldsymbol{\alpha}^\top\mathbf{d}, \qquad \nabla_{\boldsymbol{\eta}}\mathcal{L} = \boldsymbol{\alpha} \odot \left[\mathbf{d} - (\boldsymbol{\pi}^\top\mathbf{d})\mathbf{1}\right]. \tag{20}$$

We verified Eqs. 18–20 over 100 random configurations using central finite differences. The maximum relative errors of the coefficient differential and logit gradient were 1.77×10^{-6} and 4.77×10^{-9}, respectively; the maximum absolute error in Eq. 19 was 7.11×10^{-15}.

Adding a constant to every logit does not change $\boldsymbol{\pi}$, since $\mathbf{J}_\pi\mathbf{1} = \mathbf{0}$. Finite logits give strictly positive coefficients, with zero attained only as a limit. The clamp on s projects updates at its boundaries. With jointly learned bases, scaling remains interchangeable with $\boldsymbol{\alpha}$, so ρ denotes coefficient magnitude rather than integrated tracer activity.

Fixed-basis recovery experiment. We isolated the coefficient parameterization by freezing eight positive, correlated basis projections in a matrix $\mathbf{A} \in \mathbb{R}_+^{256\times 8}$. Its columns were normalized to unit mean. For each target $\boldsymbol{\alpha}^\star = \rho^\star\boldsymbol{\pi}^\star$, measurements were sampled as

$$\mathbf{y} \sim \mathrm{Poisson}\left(100\mathbf{A}\boldsymbol{\alpha}^\star + \mathbf{1}\right). \tag{21}$$

We used four magnitudes $\rho^\star \in \{0.25, 1, 4, 16\}$, three mixture profiles, and ten Poisson seeds. Direct-positive and polar coefficients were initialized at the same vector $\boldsymbol{\alpha}_0 = \mathbf{1}/8$ and optimized

for 1,200 iterations using Adam with learning rate 0.05. We report the iterate with the lowest Poisson objective. To quantify convergence speed, t_{99} is the first iteration that achieves 99% of the total loss reduction observed within the 1,200-step budget:

$$t_{99} = \min\left\{t : \mathcal{L}_t \le \mathcal{L}_{\min} + 0.01(\mathcal{L}_0 - \mathcal{L}_{\min})\right\}, \tag{22}$$

where $\mathcal{L}_{\min}$ is the lowest value attained within the fixed budget. It measures speed relative to each run's achieved decrease, not a global optimum.

Figure 4 compares direct-positive and polar coefficients with the basis projections, initialization, and optimizer settings held fixed. Polar required 34–57% fewer iterations on average to reach 99% of each run's achieved loss reduction across the four magnitudes. Both parameterizations recovered the global scale with mean relative errors below 1.2% at every tested magnitude. At $\rho^\star = 0.25$, polar had a larger mixture L_1 error than direct-positive coefficients (0.570 versus 0.279), whereas at $\rho^\star = 16$ it had a smaller error (0.093 versus 0.222). Polar thus speeds relative objective reduction under shared optimizer settings, with composition accuracy depending on scale.

Conditioning and gradient diagnostic. Let $\mathbf{g}_s = 100\mathbf{A}\boldsymbol{\alpha}$ and $\mathbf{G}_\eta = 100\mathbf{A}\rho\mathbf{J}_\pi$ be the measurement-space tangents for scale and composition. We report

$$c_{s,\eta} = \max_m \frac{|\mathbf{g}_s^\top \mathbf{G}_{\eta,:,m}|}{\|\mathbf{g}_s\|_2 \, \|\mathbf{G}_{\eta,:,m}\|_2}. \tag{23}$$

Across the three mixtures, $c_{s,\eta}$ ranged from 0.143 to 0.314. Hence the coefficient-sum identity does not make the two update directions orthogonal after projection. The polar Jacobian contains one exact null direction from softmax shift invariance; after removing zero singular values, its measurement-space condition numbers were 176.6, 381.0, and 274.9 for the uniform, skewed, and multimodal mixtures. Thus, the polar form supplies optimization coordinates without guaranteeing identifiability or favorable conditioning.

A.3.3 Local Analysis of Multiresolution Continuation

We analyze how the available activity directions can absorb a residual caused by attenuation misalignment. Let $\bar{\mathbf{y}}(\Theta_\lambda, \Theta_\phi)$ collect the expected TOF measurements, and linearize it at a fixed activity and attenuation geometry. After Poisson whitening with $\mathbf{W} = \mathrm{diag}(1/\max\{\bar{y}_m, \delta\})$, define

$$\mathbf{J}_\lambda(\tau) = \mathbf{W}^{1/2}\frac{\partial \bar{\mathbf{y}}}{\partial \Theta_\lambda}(\tau), \qquad \mathbf{r}_\phi = \mathbf{W}^{1/2}\mathbf{J}_\phi \Delta\Theta_\phi. \tag{24}$$

Here $\mathbf{J}_\phi = \partial\bar{\mathbf{y}}/\partial\Theta_\phi$, and $\mathbf{r}_\phi$ is the first-order measurement perturbation generated by attenuation misalignment. Its fraction explainable by an activity update is

$$R_{\mathrm{absorb}}(\tau) = \frac{\|\mathbf{P}_{\mathcal{S}_\lambda(\tau)}\mathbf{r}_\phi\|_2^2}{\|\mathbf{r}_\phi\|_2^2}, \qquad \mathcal{S}_\lambda(\tau) = \mathrm{col}(\mathbf{J}_\lambda(\tau)), \tag{25}$$

where $\mathbf{P}_{\mathcal{S}}$ is the orthogonal projector onto $\mathcal{S}$. By the least-squares projection theorem, $0 \le R_{\mathrm{absorb}} \le 1$, and it equals the maximum fractional reduction of $\|\mathbf{r}_\phi - \mathbf{J}_\lambda\Delta\Theta_\lambda\|_2^2$ under the local linearization. This guarantees neither nonlinear identifiability nor attainment of the projection.

We also measure local activity–geometry overlap. Let $\mathbf{Q}_\lambda$ and $\mathbf{Q}_\phi$ be orthonormal bases for the two whitened Jacobian column spaces. Their largest canonical correlation is

$$\gamma_{\lambda,\phi} = \sigma_{\max}(\mathbf{Q}_\lambda^\top \mathbf{Q}_\phi) \in [0, 1]. \tag{26}$$

A value near one indicates local directions with similar measurement effects, not global ambiguity. In the linear proxy below, nonzero gates rescale fixed tangent columns without changing their span. The subspace diagnostics change only when a level enters that span, while gate amplitudes affect finite-step optimization. This argument need not hold for a nonlinear hash-MLP decoder.

Controlled two-dimensional proxy. We construct a 24×24 activity–attenuation phantom and simulate 24 projection angles, 28 detector bins, and seven TOF bins. The forward model retains the relevant PET structure: one attenuation factor is shared by all TOF bins of an LOR, while the activity integral is TOF resolved. A 3° rotation and $(0.85, -1.15)$-pixel translation are applied only to the

attenuation prior. Four nested masked cosine spaces, containing 4, 12, 48, and 80 new directions, approximate the four coarse-to-fine hash levels and permit explicit Jacobians. This local proxy is distinct from the three-dimensional reconstruction experiment.

For the activity-only trajectories below, let $\Delta\boldsymbol{\lambda}_t$ denote the activity correction at step t. Since only attenuation geometry is perturbed, we measure spurious activity compensation as

$$B_\lambda(t) = \frac{\|\Delta\boldsymbol{\lambda}_t\|_2}{\|\boldsymbol{\lambda}^\star\|_2}. \tag{27}$$

This measures erroneous activity change, not population-level statistical bias.

The relative weighted residual is $\|\mathbf{r}_\phi - \mathbf{J}_\lambda \Delta\Theta_\lambda\|_2 / \|\mathbf{r}_\phi\|_2$; the final change in the table is B_λ.

Table 5: Local subspace diagnostics and matched-step activity compensation. "Final" denotes gradient step 120 in the linearized proxy; arrows report the initial and fully open continuation states. Bold marks the lowest final residual or activity change; subspace diagnostics are not ranked.

Configuration	R_{absorb}	$\gamma_{\lambda,\phi}$	Final residual	Final change
Coarse only	0.425	0.722	0.758	**0.049**
All levels	0.626	0.809	**0.612**	0.107
Continuation	0.425→0.626	0.722→0.809	0.613	0.103

Interpretation of the main-text trajectories. Figure 5 compares the matched-step trajectories. Opening all levels increases R_{absorb} from 0.425 to 0.626 and $\gamma_{\lambda,\phi}$ from 0.722 to 0.809. Thus, the richer activity tangent space can explain more of the error caused by the fixed, misaligned attenuation prior. The condition number of the column-normalized joint Jacobian correspondingly increases from 2.941 to 4.100, indicating less favorable local conditioning. At step 30, continuation has a relative weighted residual of 0.668 and spurious activity change of 0.068, compared with 0.620 and 0.089 when all levels are active from initialization. After all levels open, the residual gap closes, with a slightly smaller endpoint activity change under continuation. The proxy shows an early compensation–fit tradeoff, without establishing an endpoint advantage for nonlinear reconstruction.

A.4 Baseline Methods and Task-Specific Adaptations

All methods receive the same measurements and misaligned CT attenuation prior for each case and dose condition. Joint-Motion and FastMCPET are adapted from gated or short-frame PET to a single static acquisition; the main-table names refer to these adaptations.

TOF-MLEM. Multiplicative TOF-MLEM updates estimate nonnegative activity with attenuation held fixed (Shepp & Vardi, 1982), serving as the fixed-attenuation control.

MLEM+Demons. This sequential control reconstructs PET with the supplied attenuation map, registers the CT prior to that reconstruction using Demons (Thirion, 1998), and reconstructs PET again with the transformed map. Registration operates across modalities in image space.

TOF-MLRR. MLRR alternates activity estimation and CT-prior registration (Rezaei et al., 2016). Emission-driven attenuation updates provide the registration target for this classical joint reconstruction–alignment baseline.

Joint-Motion. The gated source method combines joint reconstruction and motion estimation with a parallel level sets (PLS) anatomical prior (Tsai & Liu, 2023). Our static adaptation replaces intergate motion with one CT-to-emission alignment transform and retains PLS regularization for consistent boundary orientations.

FastMCPET. Our Fast-MC-PET-inspired adaptation (Zhou et al., 2023) uses learned registration between the initial PET reconstruction and CT prior, reconstructs with the transformed attenuation map, and applies count restoration under LD. It adapts temporal PET-to-PET motion correction to cross-modal PET-to-CT alignment.

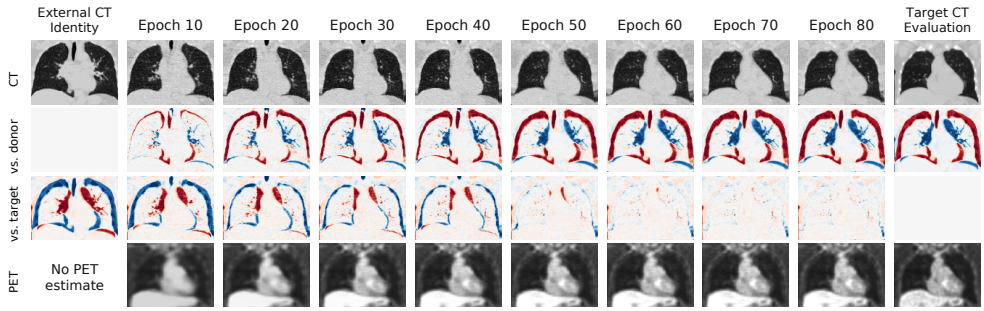


Figure 10: Controlled external-template reconstruction on a fixed coronal plane. Columns show the external donor CT, intermediate epochs, and the withheld target CT. Rows show the displayed attenuation image, its signed difference from the external template, its signed difference from the target CT, and the reconstructed PET activity. Blue and red denote negative and positive attenuation differences under a shared zero-centered window. Epochs 10–40 are direct SUGAR iterates. The displayed attenuation panels at epochs 50–80 include a target-CT-assisted two-dimensional Demons refinement, shown only as the oracle diagnostic below, not as patient-CT-free SUGAR results.

A.5 External-Template Reconstruction without Patient-Specific CT

Question and scope. The main experiments use a spatially misaligned CT prior from the PET subject. Here we explore *external-template reconstruction*: the target provides TOF measurements, while a different subject supplies the CT prior for the same anatomical region. SUGAR adapts the donor geometry while reconstructing target activity. The frozen donor field preserves attenuation values and cannot create anatomy absent from the template.

External-template formulation. For target subject t, let $\mathbf{y}^{(t)}$ denote the TOF measurements and let $\widetilde{\mu}_{\text{ext}}^{(d)}$ be a CT-derived attenuation template from donor subject $d \neq t$. We fit the continuous field $C_{\widehat{\Theta}_\mu}^{(d)}$ to the donor template and then freeze its parameters. The attenuation map used for target reconstruction is

$$\mu_{\text{ext}}^{(t,d)}(\mathbf{x}) = C_{\widehat{\Theta}_\mu}^{(d)}\big(\phi_{\Theta_\phi}(\mathbf{x})\big)\,, \tag{28}$$

where ϕ_{Θ_ϕ} is the gauge-fixed rigid–nonrigid sampling map defined in Sec. 3.2. Activity and geometry are jointly optimized from $\mathbf{y}^{(t)}$ with the same emission-domain objective as in Eq. 4; no voxel-wise similarity term involving the target attenuation map is included.

Controlled feasibility protocol. The current pilot uses an external thoracic CT template and a distinct target thorax. The donor CT, target attenuation map, and target activity are resampled to a common reconstruction grid before forward projection. We isolate adaptation from count noise using noiseless target TOF measurements and 80 SUGAR epochs initialized at identity geometry.

The common field of view is established with the target anatomy available during data construction and remains fixed across epochs. Primary optimization uses the target TOF measurements and donor CT, with a fixed endpoint. Target CT and reference activity are excluded from optimization and checkpoint selection; the target-CT-assisted refinement below is a separate oracle diagnostic.

Evolution of the reconstructed fields. Figure 10 shows the donor template, intermediate attenuation estimates, and the corresponding PET reconstructions on a fixed coronal plane. The direct SUGAR iterates at epochs 10–40 alter the thoracic outline and cardiac region while the activity estimate becomes more structured. These panels illustrate emission-driven geometric adaptation to an external prior. The target-CT-assisted panels at epochs 50–80 are discussed separately.

Reference-assisted diagnostic. Direct SUGAR retains disagreement around the heart and body boundary. Registering selected estimates to the target CT with two-dimensional Demons reduces the cardiac-boundary residuals in Fig. 10. This oracle diagnostic indicates remaining geometric headroom; it does not demonstrate recovery from emission data alone.

Interpretation and limitations. The pilot provides qualitative evidence for external CT templates as approximate attenuation priors. One target–donor pair, noiseless measurements, and a reference-informed field of view do not establish population-level or low-count robustness. Without unique ground-truth cross-subject motion, displacement magnitude is not registration error.

A.6 Exploratory Extension to PET/MR

From CT-based SUGAR to an MR-derived prior. We explore reconstruction from MR and emission data without CT at reconstruction time. Building on MR-to-sCT synthesis (Han, 2017), we use a pretrained 3D network to construct the 511-keV attenuation prior. This simulation retains SUGAR's polar activity field, gauge-fixed geometry, and multiresolution optimization.

Let m be the normalized MR volume and $G_{\widehat{\Psi}}$ the pretrained 3D generator. Its normalized output s_0 gives an sCT volume c_0 in Hounsfield units (HU) and attenuation prior μ_0:

$$s_0 = G_{\widehat{\Psi}}(m), \qquad c_0 = 3000 s_0 - 1000 \text{ HU}, \qquad \mu_0 = \mathcal{H}(c_0). \tag{29}$$

For this simulation, the conversion in cm^{-1} is

$$\mathcal{H}(c) = \max\left\{0, \begin{cases} 0.096(c+1000)/1000, & c \le 0, \\ 0.096 + 0.024c/1000, & c > 0. \end{cases}\right\}. \tag{30}$$

The main PET/CT method first fits and freezes a continuous CT field $C_{\widehat{\Theta}_\mu}$. Here we instead freeze the synthesized attenuation volume and sample it directly with trilinear interpolation $\mathcal{S}$:

$$\underbrace{C_{\widehat{\Theta}_\mu}\big(\phi_{\Theta_\phi}(\mathbf{x})\big)}_{\text{PET/CT: fitted CT field}} \quad \longrightarrow \quad \underbrace{\mu_{\Theta_\phi}(\mathbf{x}) = \mathcal{S}\big(\mu_0, \phi_{\Theta_\phi}(\mathbf{x})\big)}_{\text{PET/MR: synthesized volume}}. \tag{31}$$

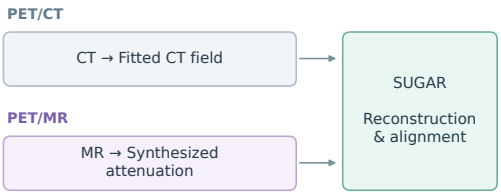


Figure 11: MR-derived prior with unchanged joint activity–geometry optimization.

The same gauge-fixed mapping in Eq. 5 acts after synthesis. Activity and geometry are then optimized through the TOF emission data term, keeping both $G_{\widehat{\Psi}}$ and μ_0 fixed. Thus, paired MR/CT data are required for offline prior construction, while reconstruction uses MR and emission measurements without the subject's CT. Figure 11 summarizes the change.

Data and preprocessing. The paired collection contains 98 MR/CT volume sets. We constructed an engineering split with 68 training, 14 validation, and 16 test entries while grouping candidate duplicates by identifier and identical normalized MR arrays. Because subject provenance is incomplete, this extension is treated as a feasibility study rather than an independent-patient evaluation. MR is normalized by its positive-voxel 0.5th and 99.5th percentiles, CT is clipped to $[-1000, 2000]$ HU, and paired volumes are resampled to an endpoint-preserving 128^3 grid.

Offline generator and sCT initialization. The 3D U-Net uses channel widths 24, 48, 96, and 192. Its epoch-95 checkpoint was selected by validation body-HU mean absolute error (MAE). Across the 14 validation cases, full-volume CT MAE was 65.8 ± 7.2 HU and PSNR/SSIM were 25.08 ± 0.67 dB/0.776 ± 0.032. Table 6 reports the resulting attenuation prior as the fixed-sCT baseline.

Controlled TOF-PET protocol. The 14-case study uses eight positive activity bases with four hash levels (resolutions 8–64, two features per level, table bound 2^{16}) and a four-layer width-32 decoder. Measurements are noiseless TOF expectations with 2×10^6 total expected counts and no scatter or randoms. We use 16,384 training and 2,048 held-out random 3D box chords, 128 samples per chord, and 21 TOF bins. Optimization runs for 30 epochs with batch size 512 and Adam learning rates 2×10^{-3} for activity and 10^{-4} for geometry; generator adaptation uses 10^{-6}. The larger single-case experiment below uses a separate staged schedule. For the MR-prior variants, reference PET and CT are used only to synthesize measurements and compute evaluation metrics; neither enters the reconstruction objective. All variants share rays, initialization, and endpoints.

Table 6: Common-endpoint validation results over 14 cases (mean±standard deviation). PSNR is in dB and attenuation MAE is in cm^{-1}. The adapted-generator row is a diagnostic, and the reference-CT row is a known-attenuation reference control. Bold marks the best mean among the three estimated-prior variants, including generator adaptation; the reference-CT control is excluded.

	PET		Attenuation		
Configuration	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	MAE ↓
Fixed sCT	23.53±0.75	0.8855±0.0250	21.35±0.81	0.6918±0.0454	0.00471±0.00063
Frozen G + geometry (SUGAR)	24.12±0.92	0.8908±0.0273	22.05±0.80	0.7018±0.0449	0.00446±0.00064
Adapted G + geometry (diagnostic)	**24.58±0.93**	**0.8963±0.0255**	**22.37±0.80**	**0.7068±0.0449**	**0.00435±0.00063**
Reference CT	27.84±1.61	0.9334±0.0227	Known	Known	Known

Table 7: Recovery from the known synthetic MR–PET mismatch over 14 cases (mean±standard deviation). Fold/OOB denote nonpositive-Jacobian and pre-clamping out-of-box percentages. Bold marks the best mean among the three estimated-prior variants, including generator adaptation; the reference-CT control and geometric plausibility diagnostics are excluded.

	PET		Attenuation		Geometry	Diagnostics
Configuration	PSNR	SSIM	PSNR	SSIM	EPE (mm)	Fold/OOB (%)
Fixed sCT	20.29±1.21	0.8366±0.0449	17.46±0.92	0.5747±0.0329	11.72±0.63	0/0
Frozen G + geometry (SUGAR)	23.31±1.14	0.8793±0.0333	20.87±0.75	0.6603±0.0328	7.91±1.03	0.233/10.66
Adapted G + geometry (diagnostic)	**23.64±1.17**	**0.8833±0.0327**	**21.09±0.78**	**0.6634±0.0326**	**7.86±0.97**	0.192/10.79
Reference CT	27.97±1.66	0.9345±0.0225	Known	Known	–	–

Metrics and statistical summaries. Image fidelity is evaluated on 3D volumes using PSNR, SSIM (Wang et al., 2004), and MAE. For a reference volume u and an estimate $\widehat{u}$ on N evaluated voxels, $\mathrm{MSE} = N^{-1}\sum_n(\widehat{u}_n - u_n)^2$, $\mathrm{MAE} = N^{-1}\sum_n|\widehat{u}_n - u_n|$, and $\mathrm{PSNR} = 10\log_{10}(R^2/\mathrm{MSE})$, where R is the preset peak: $R = 1$ for PET and $R = 0.144\,\mathrm{cm}^{-1}$ for attenuation. SSIM uses the same respective data ranges. Geometric endpoint error (EPE) is the mean Euclidean distance in millimetres between the estimated and known coordinate maps at valid points of a fixed 25^3 evaluation grid. Fold and out-of-box (OOB) rates are the percentages of sampled points with nonpositive Jacobian determinants and coordinates outside the domain before clamping, respectively. Tables report case-level mean±standard deviation; each case contributes one endpoint observation, and paired comparisons use the two-sided Wilcoxon signed-rank test (Wilcoxon, 1945). The mismatch comparisons report unadjusted per-metric p-values as exploratory summaries.

Reconstruction with an imperfect sCT prior. Table 6 reports the common-endpoint 14-case results. Relative to fixing the generated sCT, emission-driven geometry improves PET PSNR from 23.53 to 24.12 dB and SSIM from 0.8855 to 0.8908. Attenuation PSNR increases by 0.70 dB, while full-volume attenuation MAE decreases from 0.00471 to $0.00446\,\mathrm{cm}^{-1}$. These gains isolate the contribution of adapting the MR-derived prior geometry without changing the pretrained generator.

Updating the 3.29-million-parameter generator produces a further 0.46 dB PET gain over frozen-generator geometry, but increases mean runtime from 28.4 ± 3.2 s to 312.5 ± 35.2 s. All 14 cases improve in PET PSNR relative to fixed sCT. This accuracy–runtime tradeoff motivates keeping the generator frozen in the primary extension.

Numerical plausibility of the learned geometry. In the unperturbed 14-case experiment, no nonpositive Jacobian determinants are observed on a fixed 25^3 full-box grid. However, $13.68 \pm 2.35\%$ of points lie outside the field of view before clamping, mainly near the boundary. These are plausibility checks; spatial recovery is assessed using the known-transform EPE in Table 7.

Recovery under a known MR–PET mismatch. We next introduce known geometry before sCT generation while leaving the PET activity and CT used for measurement synthesis unchanged. Each of the 14 MR volumes is translated by $(+6, -4, +4)$ mm and rotated 5° about the axial axis.

Table 7 shows that frozen-generator geometry raises PET PSNR/SSIM over the fixed mismatched prior by 3.02 dB/0.0427 and attenuation PSNR/SSIM by 3.41 dB/0.0856. All 14 cases improve in all four metrics (two-sided paired Wilcoxon $p = 1.22 \times 10^{-4}$). Mean EPE decreases from 11.72 to 7.91 mm, corresponding to a $32.8\% \pm 5.4\%$ reduction across cases. Generator adaptation changes EPE only marginally, from 7.91 to 7.86 mm. Geometry drives most spatial recovery.

The larger perturbation in Fig. 12 is visualized below and excluded from the 14-case statistics.

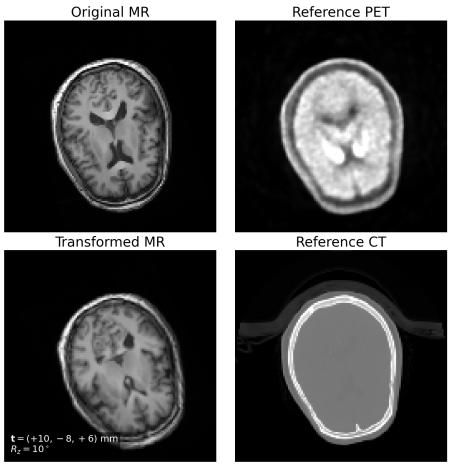


Figure 12: Larger single-case stress test used for the optimization visualizations: $\mathbf{t} = (+10, -8, +6)$ mm and $R_z = 10^\circ$. Reference PET and CT remain aligned.

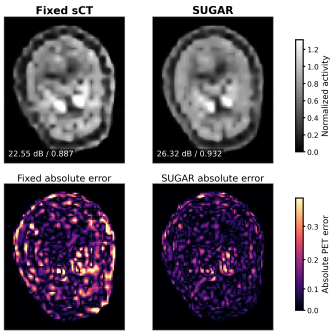


Figure 13: Fixed-sCT versus SUGAR at epoch 90.

Fixed-prior control. To isolate emission-driven alignment from activity reconstruction, the fixed-sCT control uses the same frozen MR-derived attenuation volume, polar activity representation, ray stream, and 90-epoch budget, but fixes the PET-to-prior transformation to the identity. Figure 13 uses a common error-map range and volumetric PET metrics. The fixed prior retains structured PET errors and reaches 22.55 dB/0.887 PSNR/SSIM. Allowing SUGAR to estimate the attenuation geometry raises the result to 26.32 dB/0.932 and suppresses much of the anatomically structured error, gaining 3.77 dB and 0.045 with the additional geometry updates. The 100-epoch trajectory below uses a separate run.

Optimization dynamics. The separate 100-epoch PET/MR run first optimizes coarse activity for 10 epochs. Rigid alignment then begins while only the polar activity coefficients are updated; activity-basis parameter updates resume after epoch 30. Separately, the four hash levels are progressively released during the first 30 epochs, with a maximum resolution of 32. Weak gauge-fixed nonrigid refinement and learning-rate decay begin at epoch 75. Activity and rigid-motion learning rates are 10^{-3} and 2×10^{-4}; geometry updates use 4,096-ray batches.

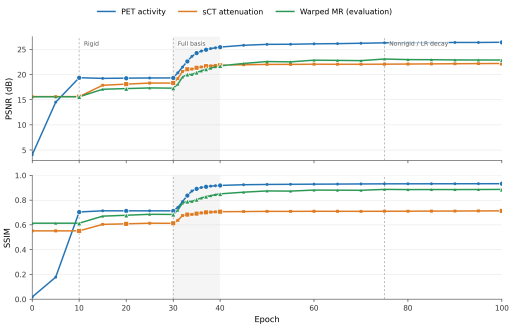


Figure 14: PSNR and SSIM over 100 epochs. Dashed lines mark optimization stages; shading marks epochs 30–40.

Figure 14 shows the complete 100-epoch trajectory. At epoch 100, PET PSNR/SSIM reach 26.39 dB/0.9323 and attenuation PSNR/SSIM reach 22.19 dB/0.7131. The evaluation-only warped-MR PSNR increases from 15.56 to 22.87 dB, while MR MAE decreases from 0.0687 to 0.0247 (about a 64% reduction). After epoch 75, PET and attenuation fidelity improve slightly, but MR-space alignment does not improve further. Thus, later image refinement does not imply better geometric recovery.

Detail at epochs 30–40 comes from an identical-setting rerun; shared checkpoints agree within 0.001 dB PSNR and 0.0001 SSIM. Figure 15 shows matched spatial estimates.

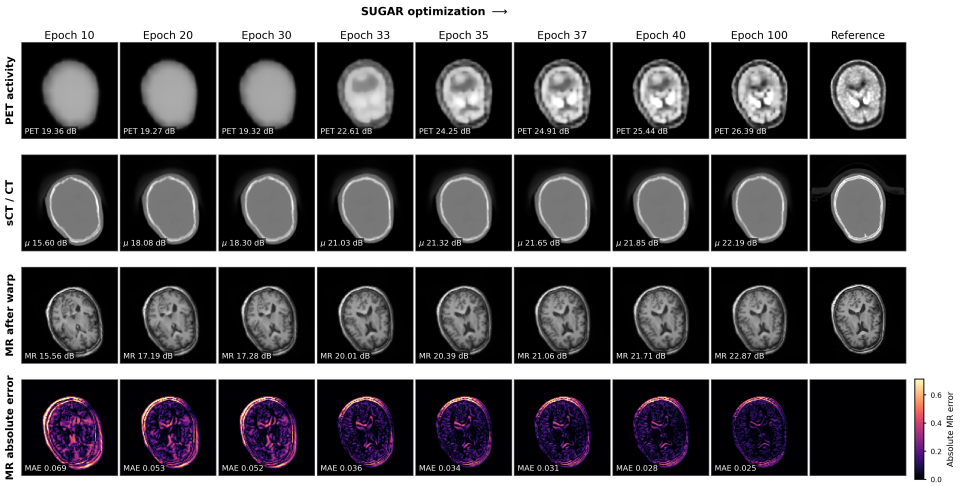


Figure 15: PET, transformed sCT, evaluation-only warped MR, and absolute MR error at matched epochs. Image windows and error scales are fixed; warped MR is not a reconstruction.

Regional attenuation behavior. In the 14-case study without imposed mismatch, the average full-volume attenuation MAE of frozen-generator geometry is 5.4% lower than that of fixed sCT. The reduction is concentrated in air/background (7.4%) and low-density regions (6.2%). Soft-tissue MAE instead increases by 8.3%, bone MAE changes by less than 0.4%, and the aggregate body MAE increases by 2.2%. These aggregate gains do not extend to every tissue.

Interpretation and limitations. These exploratory gains do not establish unique motion recovery. The brain simulations use random chords, a matched projector, simplified attenuation conversion, and no count noise, scatter, or randoms. The generator uses paired training data. The 14-case cohort informed development and checkpoint selection; it is not an independent test set.