

Neuromorphic X-ray Computed Tomography

Hongjian Wang^{1,3}, Goran Lovric¹, and Benjamín Béjar²

¹ Swiss Light Source, Paul Scherrer Institute, Villigen PSI, 5232, Switzerland

² Swiss Data Science Center, Paul Scherrer Institute, Villigen PSI, 5232, Switzerland

³ Institute for Biomedical Engineering, ETH Zurich, Zurich, 8092, Switzerland
{hongjian.wang, goran.lovric, benjamin.bejar}@psi.ch

Abstract. X-ray computed tomography (CT) reconstruction is inherently an ill-posed inverse problem, particularly with sparse-view projections. Unlike traditional frame-based sensors that acquire projections at fixed intervals, neuromorphic event cameras operate asynchronously at the pixel level and trigger events only when the log-intensity change exceeds a threshold. This asynchronous sensing naturally captures high-frequency angular variations during object rotation, providing complementary information between sparsely sampled projections. However, the application of event cameras to CT reconstruction remains largely unexplored. To bridge this gap, we introduce a Neuromorphic X-ray CT framework. Specifically, we propose an Event-enhanced Neural Attenuation Field (ENAF) that models the volumetric attenuation field by jointly leveraging sparse projections and event streams. To support and validate this framework, we construct synthetic neuromorphic CT datasets and acquire a real-world dataset. Experiments on both synthetic and real-world datasets demonstrate that ENAF consistently outperforms state-of-the-art frame-only methods while maintaining training efficiency. To the best of our knowledge, this work presents the first application of neuromorphic sensing to X-ray CT and establishes the feasibility of event-enhanced sparse-view CT reconstruction. The code and datasets are available at <https://wanghongjian98.github.io/projects/neuroxct/>.

Keywords: X-ray Computed Tomography · Event-based Vision · Implicit Neural Representations · Sparse-view Reconstruction

1 Introduction

X-ray computed tomography (CT) enables non-destructive volumetric imaging of internal structures and is widely used in materials science, biology, and industrial inspection [13, 34, 40]. High-quality CT reconstruction typically requires a large number of projection views acquired with dense angular sampling and sufficient exposure time. However, practical constraints on radiation dose, acquisition speed, and data throughput often limit the number of projections that can be acquired. For example, medical CT may acquire fewer X-ray projections to reduce radiation exposure [17, 20, 24], whereas high-speed synchrotron CT must

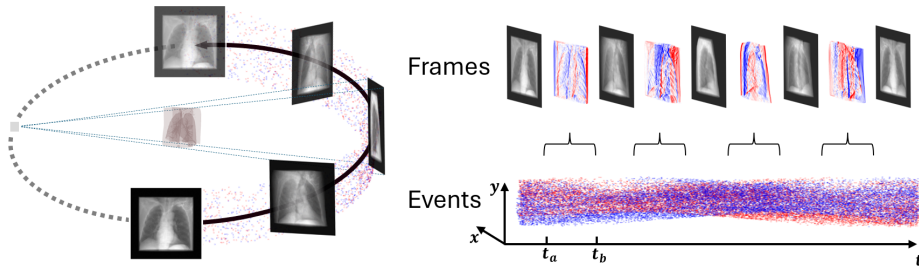


Fig. 1: Neuromorphic X-ray CT. Our system integrates two complementary modalities: sparse projections acquired at uniform angular intervals, and asynchronous event streams with microsecond temporal resolution. By accumulating events over arbitrary time windows $[t_a, t_b]$, we generate accumulated event maps that provide complementary angular-change information between sparse projections.

capture fast dynamic processes within very short scan times, restricting the number of available angular measurements [13]. In both cases, the reconstruction problem enters a sparse-view regime, where limited sparse angular information makes the inverse problem severely ill-posed. While various regularization- and learning-based methods have been developed to mitigate reconstruction artifacts [3, 5, 32, 38], their performance remains fundamentally bounded by the limited information available at the measurement level.

Unlike conventional frame-based cameras that capture images at fixed intervals, event cameras asynchronously record per-pixel brightness changes. An event is triggered whenever the change in log-intensity exceeds a preset contrast threshold. Each event records the pixel location, timestamp, and polarity, indicating the direction of the brightness change. This sensing paradigm offers microsecond-level temporal resolution, high dynamic range (> 120 dB), and low temporal redundancy [12]. Leveraging these advantages, event cameras have recently demonstrated strong potential in image restoration [22, 23, 37] and 3D reconstruction [16, 21, 31]. In the X-ray domain, event cameras have also been applied to temporal super-resolution in 2D radiography [33]. However, their application to 3D CT reconstruction remains largely unexplored.

In this work, we propose to combine conventional frame-based acquisition and neuromorphic sensing for CT reconstruction (Fig. 1). By integrating these complementary sensing modalities, the proposed paradigm aims to leverage absolute attenuation information from sparse projections and high-temporal-resolution angular-change information from events for high-fidelity volumetric reconstruction. However, these two modalities have fundamentally different characteristics: frame projections are spatially synchronous but temporally discrete, whereas event streams are asynchronously triggered at the pixel level and temporally adaptive. This mismatch makes it challenging to integrate events into conventional grid-based reconstruction pipelines. To address this, inspired by recent successes in applying implicit neural representations (INRs) to CT reconstruction [5, 38], we model the volumetric attenuation field as a continuous function.

Specifically, we propose an Event-enhanced Neural Attenuation Field (ENAF) that fuses sparse projections and asynchronous events to improve sparse-view CT reconstruction. In summary, our main contributions are threefold:

- **Neuromorphic X-ray CT framework and dataset.** To the best of our knowledge, we introduce the first CT framework that incorporates neuromorphic sensing. To support this new paradigm, we introduce *NeuroTomX*, a dataset comprising synthetic datasets with simulated events and a real-world dataset with synchronized frame and event measurements.
- **Event-enhanced Neural Attenuation Field (ENAF).** We propose a unified continuous representation that integrates sparse projections with asynchronous events. By exploiting the high-temporal-resolution information encoded in events, ENAF provides complementary constraints for reducing the angular ambiguity inherent in sparse-view CT reconstruction.
- **Improved sparse-view reconstruction performance.** Extensive experiments demonstrate that ENAF suppresses reconstruction artifacts and recovers fine-grained structural details, consistently outperforming existing frame-only reconstruction methods on both synthetic and real-world datasets.

2 Related Work

2.1 X-ray CT Reconstruction

Classical analytical methods, such as FDK [10] for cone-beam geometry or FBP [15] for parallel-beam geometry, are computationally efficient but often degrade severely in ill-posed settings. Model-based iterative reconstruction incorporates physical priors but is often computationally expensive and relies on handcrafted regularizers [2, 3, 32]. Data-driven deep learning methods improve reconstruction quality using learned priors, yet typically require large paired training datasets [1, 7]. Recently, implicit neural representations (INRs) have emerged as a promising alternative for self-supervised volumetric reconstruction, showing strong potential in sparse-view CT [5, 9, 18, 38, 39]. Unlike grid-based approaches, INRs such as Neural Attenuation Fields (NAF) represent the attenuation volume as a continuous coordinate-based function and enable reconstruction through differentiable projection consistency [38]. Building on this foundation, SAX-NeRF introduces Line-Segment Transformers to capture fine-grained structures [5]. Concurrently, 3D Gaussian Splatting has been adapted for tomography to leverage explicit spatial primitives for efficiency [35], while diffusion models have been employed as generative priors for ill-posed and limited-view reconstruction tasks [8, 19]. In this work, we adopt NAF as our backbone for sparse-view settings. Its continuous and efficient formulation naturally facilitates the integration of asynchronous event streams alongside sparse projections.

2.2 Event-based 3D Reconstruction

Event cameras have recently been explored for 3D reconstruction, owing to their microsecond temporal resolution and high dynamic range. Early geometric methods, such as event-based multi-view stereo [28] and SLAM [29], demonstrated

that 3D structures can be recovered even under rapid motion or extreme lighting conditions where traditional cameras fail. However, these methods remain tied to hand-crafted geometric assumptions and do not directly provide a flexible, learnable representation for complex 3D scenes. Building on the success of INRs, recent research has integrated event cameras into implicit 3D reconstruction frameworks [36]. Some of these approaches rely solely on event data for scene representation [14, 16, 30], while others adopt a hybrid fusion of events and RGB frames [6, 27]. For example, EventNeRF reconstructs a dense neural radiance field (NeRF) from a monocular event stream for 3D-consistent novel view synthesis [30], while E-NeRF estimates a NeRF from a fast-moving event camera and shows improved robustness under fast motion and high-dynamic-range conditions [16]. Hybrid methods like E2NeRF [27] leverage events to compensate for motion blur in conventional video. Despite this rapid progress, neuromorphic sensing has yet to be explored in X-ray CT. Unlike existing event-based neural rendering methods that primarily model optical radiance or scene appearance, CT reconstruction estimates a volumetric attenuation field from X-ray line-integral measurements. This distinction prevents a direct transfer of event-based NeRF formulations to CT. In this work, we address this gap by proposing an event-enhanced 3D reconstruction framework tailored to X-ray imaging physics.

3 Method

3.1 Problem Formulation

We consider a neuromorphic X-ray CT acquisition setup. Let $f(\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{R}^+$ denote the unknown 3D attenuation field. According to the Beer–Lambert law [15], for a given projection angle θ and detector coordinates (x, y) , the measured intensity $I_\theta(x, y)$ is:

$$I_\theta(x, y) = I_0 \exp(-A_\theta(x, y)), \quad A_\theta(x, y) \triangleq \int_{L_{\theta, x, y}} f(\mathbf{x}) d\mathbf{x}, \quad (1)$$

where I_0 is the incident intensity and $A_\theta(x, y)$ is the line-integral projection (or simply projection) along the X-ray path $L_{\theta, x, y}$.

In addition to conventional intensity frames, the system also records a synchronized event stream that senses per-pixel changes in the log-intensity given by:

$$\log I_\theta(x, y) = \log I_0 - A_\theta(x, y). \quad (2)$$

Let $\theta(t)$ denote the rotation angle at time t , and $\log I_{\theta(t)}(x, y)$ be the log-intensity at pixel (x, y) at that moment. Under the standard idealized contrast-threshold model, an event $e_k = (x_k, y_k, t_k, p_k)$ is triggered whenever the accumulated change reaches a predefined contrast threshold $C > 0$:

$$\log I_{\theta(t_k)}(x_k, y_k) - \log I_{\theta(t_{k-1})}(x_k, y_k) = p_k C, \quad p_k \in \{+1, -1\}, \quad (3)$$

where t_k and t_{k-1} are the timestamps of successive events at the same pixel. Due to the inverse relationship between $\log I_\theta$ and A_θ (Eq. (2)), a positive event ($p_k =$

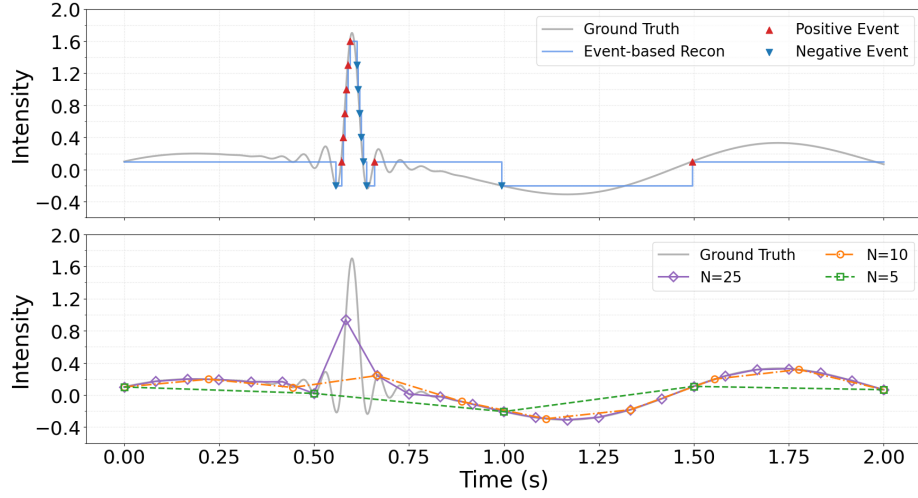


Fig. 2: Comparison of neuromorphic and uniform sampling. Neuromorphic event-based sampling ($C = 0.3$) adaptively captures the transient pulse at $t = 0.6$ s using only 16 events. In contrast, uniform sampling at varying densities ($N = 5, 10, 25$) fails to resolve the peak intensity and suffers from temporal aliasing. The 1D signal can be interpreted as the temporal log-intensity trace of a detector pixel during continuous rotation.

+1) indicates a decrease in the projected attenuation $A_{\theta(t_k)}(x_k, y_k)$ relative to its value at t_{k-1} , while a negative event ($p_k = -1$) indicates an increase.

Let $\mathcal{A} = \{A_{\theta_n}\}_{n=0}^{N-1}$ denote the set of acquired projections at regular angles (*i.e.*, $\theta_n = \pi n/N$), and $\mathcal{E} = \{e_k\}_{k=0}^{N'-1}$ the set of events. Our goal is to reconstruct the attenuation field $f(\mathbf{x})$ from $\mathcal{A}$ and $\mathcal{E}$.

3.2 Neuromorphic Sampling

Unlike conventional frame-based sensors that capture signals at fixed temporal intervals (*e.g.*, equi-spaced angles under constant rotation), event cameras adaptively trigger measurements in response to signal dynamics, as illustrated in Fig. 2. This fundamental difference leads to complementary sampling characteristics. Regular projection acquisition operates at a fixed sampling rate, which limits the recoverable temporal bandwidth according to the Nyquist criterion. Rapid temporal variations that exceed this limit cannot be faithfully resolved and may lead to aliasing. Meanwhile, slowly varying or static regions are repeatedly sampled, producing redundant measurements. In contrast, neuromorphic event-based sampling excels at adaptively capturing high-frequency angular variations, yet it does not directly capture slowly varying or constant components where no events are triggered.

This complementarity naturally motivates our hybrid acquisition strategy, which integrates asynchronous events with sparse projections to leverage the strengths of both modalities. By fusing the low-frequency structural information

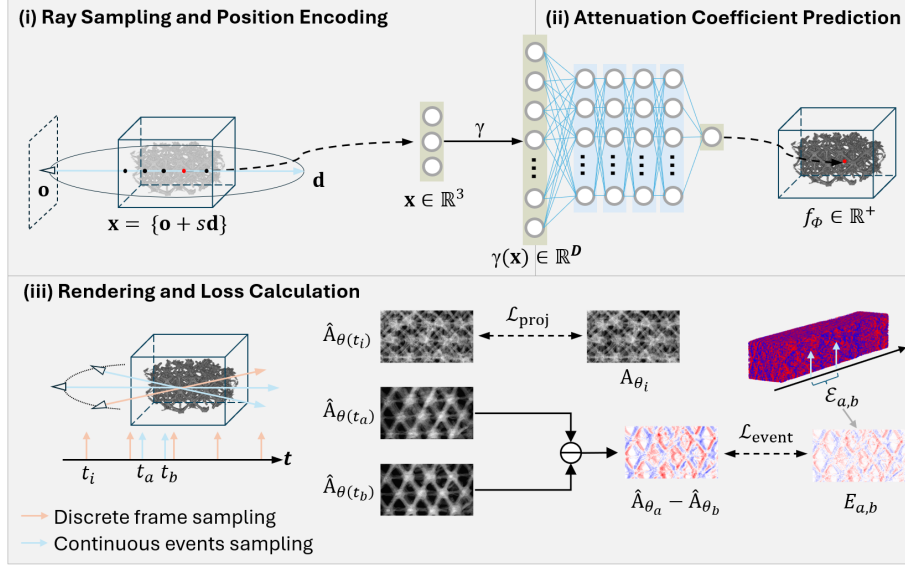


Fig. 3: Overview of the proposed Event-enhanced Neural Attenuation Field (ENAF). (i) Spatial coordinates $\mathbf{x}$ along a ray are sampled and encoded via multi-resolution hash encoding $\gamma(\cdot)$, then (ii) fed into an MLP $f_\Phi(\mathbf{x})$ parameterizing the 3D attenuation field. (iii) To supervise the network, projections are rendered at three distinct angles: a uniformly sampled angle θ_i is used to compute the projection loss $\mathcal{L}_{\text{proj}}$, while two randomly sampled angles θ_a and θ_b are used to calculate the predicted projection difference $\hat{A}_{\theta_a} - \hat{A}_{\theta_b}$. This predicted change is then supervised by the event loss $\mathcal{L}_{\text{event}}$, which compares the difference with the accumulated event map $E_{a,b}$ derived from the set of events $\mathcal{E}_{a,b}$ over the corresponding interval.

from projections with high-frequency angular dynamics from events, we aim to achieve high-fidelity volumetric reconstruction.

3.3 Event-enhanced Neural Attenuation Field (ENAF)

Following [38], we represent the 3D attenuation field as $f_\Phi(\mathbf{x}) = \text{MLP}(\gamma(\mathbf{x}); \Phi)$, where $\gamma(\cdot)$ is a multi-resolution hash encoding and Φ denotes the trainable parameters of the latent embeddings within the encoding volume and the weights of the MLP. ENAF is designed to bridge the information gap in sparse-view CT by leveraging high-temporal-resolution event streams. The overall architecture of ENAF is illustrated in Fig. 3.

In the projection forward model, each X-ray path $L_{\theta,x,y}$ corresponds to a detector coordinate (x, y) at projection angle θ . Points along the ray are parameterized as $\mathbf{x}(s) = \mathbf{o}_{\theta,x,y} + s \mathbf{d}_{\theta,x,y}$, where $\mathbf{o}_{\theta,x,y}$ and $\mathbf{d}_{\theta,x,y}$ denote the ray origin and unit direction vector, respectively. s denotes the propagation distance along the ray. The explicit (θ, x, y) subscripts are omitted for clarity hereafter. Then

the predicted projection $\hat{A}_\theta(x, y)$ is approximated via numerical integration:

$$\hat{A}_\theta(x, y) = \int_{L_{\theta, x, y}} f_\Phi(\mathbf{o} + s \mathbf{d}) ds \approx \sum_{k=1}^S f_\Phi(\mathbf{o} + s_k \mathbf{d}) \Delta s, \quad (4)$$

where $s_k = k\Delta s$ are uniformly spaced samples, Δs is the discretization step, and S is the total number of sampled points along the ray.

To bridge the angular gap between sparse-view projections, we consider a pair of angles θ_a and θ_b (corresponding to timestamps t_a and t_b) randomly sampled from the temporal axis. Within the interval $[t_a, t_b]$, we define $\mathcal{E}_{a,b}(x, y) = \{e_k \mid t_k \in [t_a, t_b], x_k = x, y_k = y\}$ as the set of events occurring at pixel (x, y) . The accumulated event map $E_{a,b}$ approximates the total change in log-intensity and is then computed by summing the polarities p_k of these pixel-specific events:

$$E_{a,b}(x, y) = C \sum_{e_k \in \mathcal{E}_{a,b}(x, y)} p_k, \quad (5)$$

where C denotes the preset contrast threshold as in Eq. (3).

From the estimated attenuation field, we render the predicted projection difference between these two angles:

$$\Delta \hat{A}_{a,b} = \hat{A}_{\theta_a}(x, y) - \hat{A}_{\theta_b}(x, y). \quad (6)$$

Since $\log I_{\theta_b} - \log I_{\theta_a} = A_{\theta_b} - A_{\theta_a}$ under a constant incident intensity, the predicted attenuation difference $\Delta \hat{A}_{a,b}$ can be compared with the accumulated event map $E_{a,b}$. By enforcing consistency between these two, ENAF can incorporate high-frequency angular dynamics to reduce structural ambiguities.

3.4 Optimization Objectives

The network parameters Φ are optimized using a hybrid objective that combines projection consistency with event-based supervision:

$$\mathcal{L}(\Phi) = \mathbb{E}_{\theta_n \sim \mathcal{P}_\Theta} [\mathcal{L}_{\text{proj}}(\theta_n)] + \lambda_e \mathbb{E}_{(\theta_a, \theta_b) \sim \mathcal{P}_E} [\mathcal{L}_{\text{event}}(\theta_a, \theta_b)], \quad (7)$$

where $\lambda_e \geq 0$ balances the contribution of the event-based supervision. Here $\mathcal{P}_\Theta$ denotes the uniform sampling over the regular projection angles, while $\mathcal{P}_E$ represents the random sampling of temporal event intervals. During training, both expectations are approximated using mini-batch sampling.

Projection-based Consistency. To preserve the absolute scale, we minimize the mean squared error (MSE) between the predicted and measured projections:

$$\mathcal{L}_{\text{proj}}(\theta_n) = \left\| \hat{A}_{\theta_n} - A_{\theta_n} \right\|_2^2, \quad (8)$$

Table 1: Details of the datasets used in the experiments. **Dim.:** volume resolution in voxels; **Proj.:** projection resolution in pixels; **EPP:** average number of events per pixel. Synthetic projection intensities are derived from CT volumes normalized to the $[0, 1]$ range. *Real-world dataset acquired at the TOMCAT beamline of the Swiss Light Source. [†]Estimated threshold based on the original physical projection scale.

Dataset	Dim. ($N^2 \times D$)	Proj. ($W \times H$)	Events	Thresh. (C)	EPP
Chest	$128^2 \times 128$	256×256	5.52M	0.0005	84.19
Jaw	$256^2 \times 256$	512×512	12.95M	0.001	49.40
Foot	$256^2 \times 256$	512×512	10.70M	0.001	40.81
Abdomen	$512^2 \times 463$	1024×1024	11.33M	0.0025	10.81
Lattice*	$1280^2 \times 720$	1280×720	51.26M	$0.2^\dagger$	55.62

Event-based Supervision. For event supervision, we randomly sample angular intervals (θ_a, θ_b) and enforce consistency between the predicted projection difference $\Delta\hat{A}_{a,b}$ and the corresponding accumulated event map $E_{a,b}$. In practice, the exact contrast scale of an event camera is not precisely known and may be affected by sensor noise, temperature drift, and manufacturing variability [12]. Directly matching the magnitudes of the two quantities would therefore require accurate threshold calibration. We instead adopt a normalized MSE loss [16] that compares their normalized spatial patterns:

$$\mathcal{L}_{\text{event}}(\theta_a, \theta_b) = \left\| \frac{\Delta\hat{A}_{a,b}}{\|\Delta\hat{A}_{a,b}\|_2 + \epsilon} - \frac{E_{a,b}}{\|E_{a,b}\|_2 + \epsilon} \right\|_2^2, \quad (9)$$

where ϵ is a small constant for numerical stability.

4 Experiments

4.1 Datasets

We evaluate ENAF on four synthetic CT datasets with simulated event streams and one real-world dataset. Details of these datasets are summarized in Tab. 1.

Synthetic datasets. Following the experimental settings of NAF [38], we use the *Chest*, *Jaw*, *Foot*, and *Abdomen* datasets and generate projections uniformly distributed over 180° using the TIGRE toolbox under a cone-beam geometry [4]. The original CT volume serves as the ground truth (GT). To simulate events, we first forward-project each volume at 500 densely sampled angles. The resulting per-pixel log-intensity signals are linearly interpolated in time, and events are triggered using the contrast-threshold crossing model in Eq. (3). The threshold C is chosen to produce approximately 5M–13M events per dataset. Since the original volumes are normalized, these threshold values should be interpreted as relative parameters.

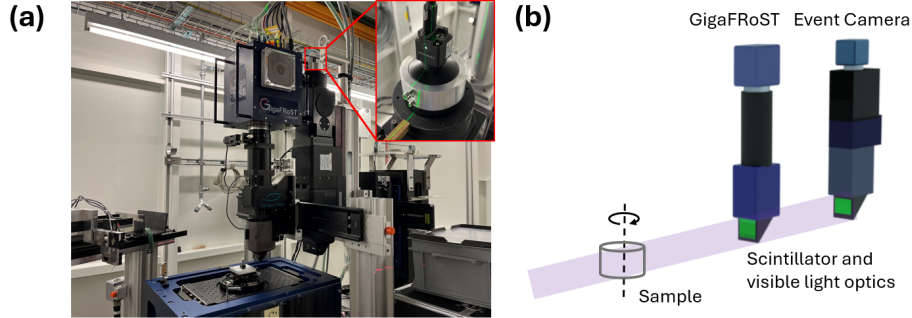


Fig. 4: Real-world acquisition setup at the TOMCAT beamline, Swiss Light Source [33]. **(a)** Photograph of the experimental setup, with the event camera highlighted by the red box. **(b)** Schematic illustration of the setup. The transmitted X-rays are converted into visible light by the scintillator, pass through optics and are captured by the GigaFRoST frame detector [26] and the Prophesee EVK4 event camera [11].

Real-world dataset. The real-world dataset was acquired at the TOMCAT beamline, Swiss Light Source, using parallel-beam geometry under continuous X-ray illumination, as used in [33]. A dual-camera system was used in our experiment (Fig. 4). It captures a nano-architected metallic lattice with fine geometric details [25]. The two cameras were temporally synchronized using a shared external trigger, and simultaneously recorded the sample during a 180° rotation over 4.5 s; the acquired 450 frames were cropped to a 1280×720 field of view ($1.41 \times 0.79 \text{ mm}^2$). Spatial alignment was estimated directly from the acquired data by registering frame-derived projection-difference maps with accumulated event maps. Further details are provided in the supplementary material. For reference, we estimate a contrast threshold of $C \approx 0.2$ (Tab. 1), by correlating the accumulated event maps with the corresponding projection-difference maps.

4.2 Implementation Details

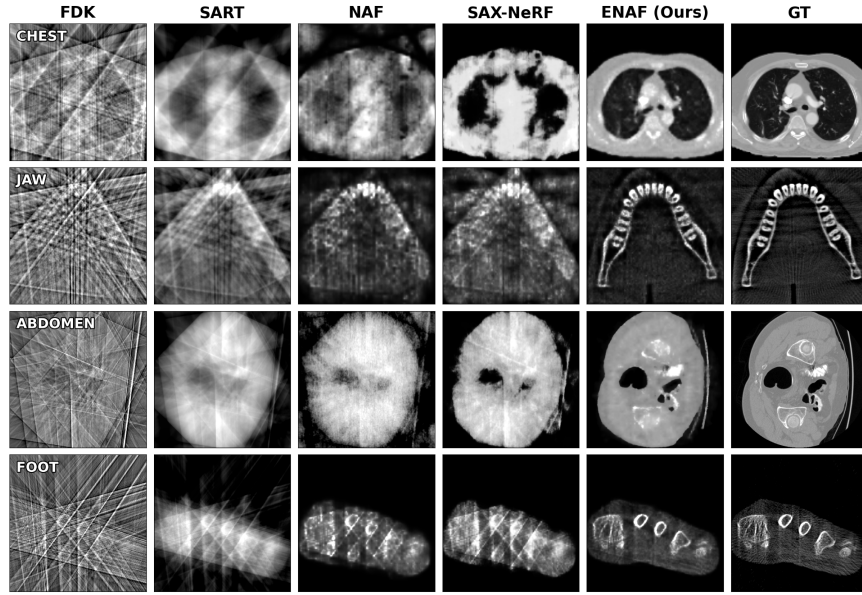
ENAF is implemented in PyTorch and optimized using Adam. The learning rate is initialized at 1×10^{-3} and decayed to 1×10^{-4} over the first 1,500 epochs. All models are optimized for 3,000 epochs. During training, we use a batch size of 1,024 rays. The number of samples per ray is adapted to the target volume resolution. Specifically, we use 192 sample points for *Chest*, 320 for *Jaw*, *Foot*, and 576 for *Abdomen*. For the lattice dataset, we use 1,280 sample points per ray. For event supervision, we randomly sample temporal windows $[t_a, t_b]$ to form event maps. For reproducibility, these settings of event-loss weight λ_e and the maximum event-window span are included in the released configuration files. For consistency with prior methods, we use the official NAF/SAX-NeRF evaluation protocol for computing 3D PSNR and SSIM across all compared methods. All experiments are conducted on a single NVIDIA H100 GPU.

Table 2: Quantitative comparison of reconstruction performance on four anatomical datasets (*Chest*, *Abdomen*, *Foot*, and *Jaw*). PSNR (dB) and SSIM are reported for 5, 10, 25, and 50 projection views. For consistency with the learning-based baselines, the metric computation follows the official evaluation implementation of NAF [38] and SAX-NeRF [5]. The learning-based baselines are retrained under the same experimental settings as ENAF. The **bold** and underlined values indicate the best and second-best results, respectively. All methods are evaluated using the same projection data and imaging geometry.

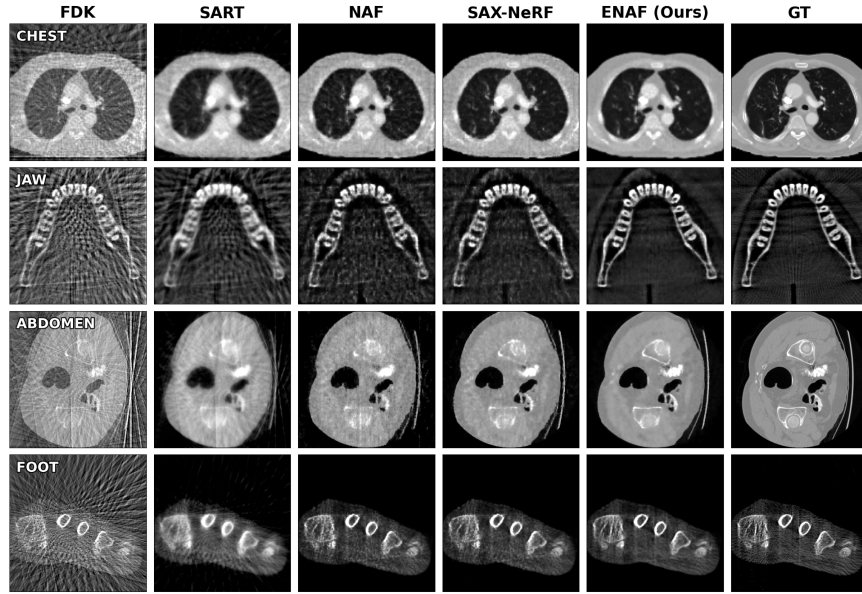
Dataset	Method	5 views		10 views		25 views		50 views	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Chest	FDK	12.59	0.2082	16.66	0.3651	21.82	0.6208	22.89	0.7861
	SART	19.83	0.5787	23.42	0.7347	28.15	0.8955	32.17	0.9594
	NAF	17.75	0.5157	24.75	0.7965	29.62	0.9168	33.50	0.9667
	SAX-NeRF	<u>20.52</u>	<u>0.6498</u>	<u>25.53</u>	<u>0.8240</u>	<u>30.46</u>	<u>0.9291</u>	<u>34.26</u>	<u>0.9701</u>
	ENAF (ours)	29.79	0.9330	31.10	0.9479	32.58	0.9622	34.57	0.9750
Abdomen	FDK	12.19	0.1674	15.62	0.2797	19.20	0.4623	22.63	0.6030
	SART	23.10	0.7246	25.12	0.7885	28.26	0.8598	31.40	0.9170
	NAF	21.42	0.6886	24.80	0.7690	29.53	0.8671	34.69	0.9544
	SAX-NeRF	<u>24.10</u>	<u>0.8005</u>	<u>26.62</u>	<u>0.8462</u>	<u>31.06</u>	<u>0.9192</u>	<u>34.95</u>	0.9589
	ENAF (ours)	29.14	0.8851	30.23	0.9038	32.76	0.9368	35.11	<u>0.9584</u>
Foot	FDK	11.79	0.1593	15.20	0.2371	20.22	0.4141	24.53	0.6000
	SART	22.67	0.7100	24.34	0.7930	27.26	0.8768	30.29	0.9296
	NAF	<u>24.29</u>	<u>0.8357</u>	26.59	0.8703	30.14	0.9135	31.82	0.9347
	SAX-NeRF	23.65	0.8326	<u>26.74</u>	<u>0.8737</u>	<u>30.32</u>	<u>0.9170</u>	<u>32.25</u>	<u>0.9404</u>
	ENAF (ours)	30.02	0.9105	30.78	0.9223	31.70	0.9330	32.45	0.9409
Jaw	FDK	15.49	0.1288	18.68	0.2496	23.71	0.5386	28.58	0.7816
	SART	24.80	<u>0.7091</u>	26.04	0.7428	29.07	0.8447	33.13	0.9301
	NAF	24.99	0.6869	26.87	0.7581	30.37	0.8629	34.37	0.9398
	SAX-NeRF	<u>25.12</u>	0.7017	<u>27.41</u>	<u>0.7764</u>	<u>31.39</u>	<u>0.8922</u>	<u>35.34</u>	<u>0.9510</u>
	ENAF (ours)	31.88	0.9026	33.16	0.9267	34.66	0.9467	35.54	0.9530

4.3 CT Reconstruction on Synthetic Dataset

Quantitative Results. Tab. 2 presents the quantitative results comparing the proposed ENAF method against traditional methods (FDK [10] and SART [2]) and recent state-of-the-art (SOTA) INR-based models (NAF [38] and SAX-NeRF [5]). We evaluate each method across four different settings corresponding to a limited number of 5, 10, 25, and 50 sparse projections/views. The results indicate that ENAF outperforms competing methods in almost all tested scenarios both in PSNR and SSIM, and it consistently outperforms NAF in all cases. The advantage of ENAF is most pronounced in extremely sparse settings, where a significant enhancement is observed by fusing the two modalities. For instance, in the 5-view *Chest* reconstruction, ENAF achieves a remarkable PSNR of 29.79 dB, which is a 12 dB improvement over the standard NAF and 9.27 dB higher than SAX-NeRF. This significant gain demonstrates that by leveraging event data, our model can recover structural information that is typically lost when projection data is severely undersampled.



(a) 5-view CT reconstruction.



(b) 25-view CT reconstruction.

Fig. 5: Qualitative comparison of sparse-view CT reconstructions. (a) Under extremely sparse 5-view sampling, our method preserves structural details better than FDK, SART, NAF, and SAX-NeRF. (b) With 25 views, integrating events further suppresses background artifacts and enhances high-frequency structures.

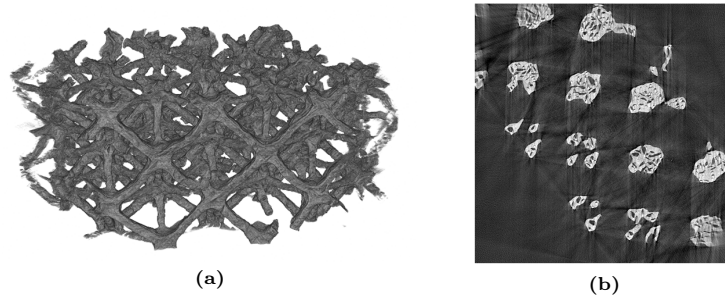


Fig. 6: Reference reconstruction of the real-world lattice dataset using FBP with all 450 recorded projections: (a) 3D rendering and (b) central slice.

As the number of views increases, the performance gap between the methods gradually narrows. In the 50-view setting, the additional intensity projections provide a sufficiently dense sampling in this setting, making the supplementary information from the event stream less critical. Nevertheless, ENAF still maintains a competitive performance, achieving the highest or second-highest scores across most anatomical datasets, and it consistently outperforms NAF. This confirms that ENAF is not only robust under extreme sparsity but also remains highly effective as more measurements become available.

Qualitative Results. Fig. 5a and Fig. 5b present the visual comparison of ENAF against traditional and SOTA baselines. Under the extreme sparsity of 5 views, analytical methods (FDK, SART) fail to reconstruct recognizable structures due to streak artifacts. While INR-based methods (NAF and SAX-NeRF) recover global shapes, they still suffer from significant structural loss and noise. In contrast, ENAF maintains sharp anatomical boundaries and structural integrity.

At 25 views, ENAF further demonstrates its superiority in suppressing background noise and recovering fine-grained details. In the *Jaw* dataset, existing approaches often fail to reconstruct intricate texture boundaries and introduce a noisy background. By leveraging events, ENAF effectively restores cleaner backgrounds and more complex structural patterns, enhancing high-frequency fidelity and producing reconstructions that are more consistent with the GT.

4.4 CT Reconstruction on Real-world Dataset

To further evaluate ENAF under experimental acquisition conditions, we conduct experiments on the real-world nano-architected lattice dataset. We use the reconstruction obtained by applying FBP to all 450 projections. Given the high projection resolution of 1280×720 , these 450 projections represent a relatively sparse-view sampling scenario. Fig. 6a shows its 3D rendering, and Fig. 6b shows the corresponding central slice. Although the reconstruction with FBP exhibits characteristic streak artifacts, it provides a qualitative reference for assessing the reconstruction fidelity of our proposed method.

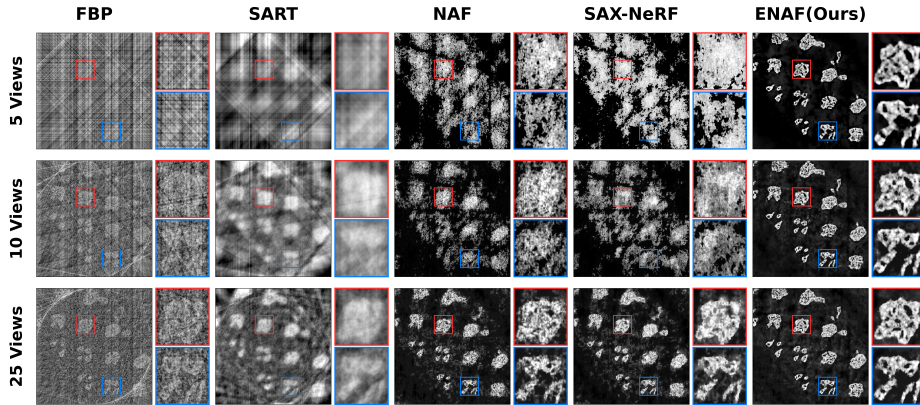


Fig. 7: Qualitative comparison on the real-world lattice dataset using 5, 10, and 25 frame projections. ENAF preserves sharper node structures and clearer lattice boundaries than the frame-only reconstruction methods.

For sparse-view evaluation, we uniformly subsample 5, 10, and 25 projections while using the event stream. As shown in Fig. 7, FBP and SART exhibit severe streak artifacts and structural blurring as the number of frame projections decreases. NAF improves the overall visual smoothness but still has blurred lattice boundaries and node structures. SAX-NeRF slightly improves local structures compared with NAF, but residual artifacts and boundary blurring remain. In comparison, ENAF restores sharper node structures and clearer lattice connectivity, particularly in the 5-view setting.

4.5 Sensitivity Analysis to Event Simulation Noise

We further analyze the robustness of ENAF to noise in the simulated event stream. Experiments are conducted on the *Chest* dataset in the 5-view setting. Starting from the clean simulated events, we independently corrupt the event stream with two types of perturbations: random temporal jitter in timestamps and spatially random spurious events.

As reported in Tab. 3, ENAF remains stable under both perturbations. Increasing the timestamp delay to 10 ms decreases PSNR by only 0.30 dB, while adding spurious events with a count equal to 50% of the original event stream reduces PSNR by 0.17 dB. This robustness is mainly due to the interval-based formulation of event supervision. Instead of enforcing consistency at the level of individual events, ENAF compares accumulated event maps over angular intervals. Consequently, small timestamp perturbations, even up to 10 ms over a 5 s acquisition, typically keep events within the same or neighboring accumulation windows. Similarly, isolated spurious events are spatially incoherent and do not form consistent structures aligned with object motion or projection changes; their contribution is therefore largely averaged out during event accumulation.

Table 3: ENAF robustness to event noise on the *Chest* dataset using 5 views. Timestamp jitter adds a random delay sampled uniformly from $[0, t_{\max}]$ to each event. Spurious noise uniformly adds isolated random events to the original stream, with the ratio denoting the added-event count relative to the original count.

Noise setting	Clean	Low	Medium	High
Jitter $t_{\max}$	0	1.0 ms	5.0 ms	10.0 ms
PSNR/SSIM	29.79/.9330	29.78/.9328	29.66/.9312	29.49/.9294
Spurious ratio	0	10%	20%	50%
PSNR/SSIM	29.79/.9330	29.77/.9322	29.73/.9316	29.62/.9292

4.6 Training Efficiency

We compare ENAF with the NAF and SAX-NeRF in terms of reconstruction quality and training time. As shown in Fig. 8, ENAF achieves a favorable quality–efficiency trade-off across the 5-, 10-, and 25-view settings. In the challenging 5-view setting, ENAF reaches approximately 30 dB PSNR within 5 minutes, substantially outperforming NAF (< 18 dB) and SAX-NeRF (~ 21 dB). Notably, SAX-NeRF requires more than 60 minutes and 25 views to achieve a reconstruction quality comparable to that of ENAF using only 5 views, highlighting the efficiency and data effectiveness of the proposed method.

4.7 Event-only Reconstruction

We further conduct an event-only reconstruction on the real-world dataset to analyze the structural information contained in the event stream alone. We optimize the model using only $\mathcal{L}_{\text{event}}$ and normalize the reconstructed volume for visualization. As shown in Fig. 9, event-only supervision recovers visible lattice structures in regions with sufficient event activity. However, regions with weak projection changes generate fewer events and are reconstructed less accurately. Adding sparse frame projections restores the missing absolute attenuation information and improves structural completeness, highlighting the complementary roles of frame projections and events.

4.8 Limitations

The current real-world evaluation is limited to a single lattice specimen acquired under a controlled synchrotron setup. Future work will evaluate broader sample types and imaging conditions, and design dedicated acquisition protocols with explicit dosimetry, low-SNR measurements, and high-speed operation.

5 Conclusion

This work establishes the feasibility of using neuromorphic event sensing as complementary angular-change supervision for sparse-view X-ray CT reconstruction. Experiments on synthetic and real-world data demonstrate improved reconstruction quality over frame-only methods.

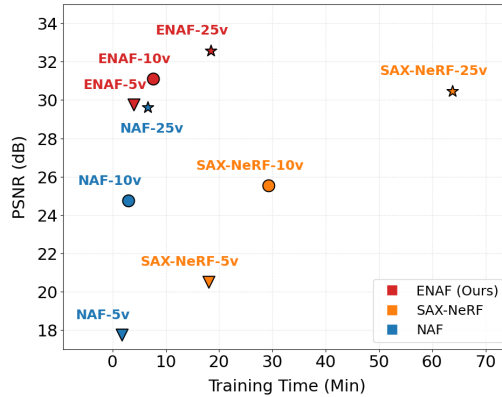


Fig. 8: Reconstruction quality versus training time on the *Chest* dataset. All timings are measured on a single NVIDIA H100 GPU over 3,000 training epochs. The suffix “-v” denotes the number of frame projections (5, 10, or 25).

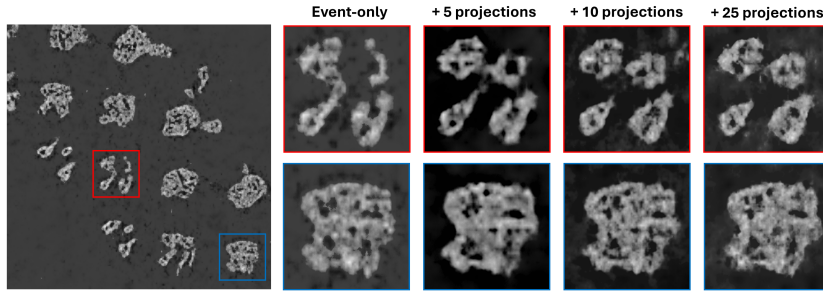


Fig. 9: Comparison of event-only and event-enhanced reconstructions on the real-world lattice dataset. The event-only reconstruction is normalized for visualization. The red boxes indicate a region with relatively sparse event activity, while the blue boxes highlight visible lattice structures recovered from event supervision.

Acknowledgements

We would like to thank M. Kagias (Lund University) for providing the sample and for participating in the beamline experiment. We acknowledge P. Zuppiger (PSI), A. Hadjiivanov (ESTEC) and E. Blazquez (ESTEC) for their support in setting up the dual-camera system. We further acknowledge X. Li (ETHZ), D. Vera Nieto (PSI) and M. Stampanoni (PSI) for fruitful discussions. H.W. acknowledges the China Scholarship Council. The authors acknowledge the computational resources provided by the Swiss National Supercomputing Centre (CSCS) under project ID lp91 on Alps. We acknowledge the Paul Scherrer Institute, Villigen, Switzerland for the provision of synchrotron radiation beamtime at the TOMCAT beamline X02DA of the SLS. The study was supported by the Swiss National Science Foundation (SNSF), grant no. 200021_219704.

References

1. Adler, J., Öktem, O.: Learned Primal-Dual Reconstruction. *IEEE Transactions on Medical Imaging* **37**(6), 1322–1332 (Jun 2018). <https://doi.org/10.1109/TMI.2018.2799231>
2. Andersen, A.H., Kak, A.C.: Simultaneous Algebraic Reconstruction Technique (SART): A Superior Implementation of the ART Algorithm. *Ultrasonic Imaging* **6**(1), 81–94 (Jan 1984). <https://doi.org/10.1177/016173468400600107>
3. Bian, J., Wang, J., Han, X., Sidky, E.Y., Shao, L., Pan, X.: Optimization-based Image Reconstruction from Sparse-view Data in Offset-Detector CBCT. *Physics in Medicine and Biology* **58**(2), 205–230 (Jan 2013). <https://doi.org/10.1088/0031-9155/58/2/205>
4. Biguri, A., Dosanjh, M., Hancock, S., Soleimani, M.: TIGRE: a MATLAB-GPU toolbox for CBCT image reconstruction. *Biomedical Physics and Engineering Express* **2**(5), 055010 (Sep 2016). <https://doi.org/10.1088/2057-1976/2/5/055010>
5. Cai, Y., Wang, J., Yuille, A., Zhou, Z., Wang, A.: Structure-Aware Sparse-View X-Ray 3D Reconstruction. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 11174–11183 (Jun 2024). <https://doi.org/10.1109/CVPR52733.2024.01062>
6. Cannici, M., Scaramuzza, D.: Mitigating Motion Blur in Neural Radiance Fields with Events and Frames. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 9286–9296 (Jun 2024). <https://doi.org/10.1109/CVPR52733.2024.00887>
7. Chen, H., Zhang, Y., Zhang, W., Liao, P., Li, K., Zhou, J., Wang, G.: Low-dose CT via convolutional neural network. *Biomedical Optics Express* **8**(2), 679–694 (Feb 2017). <https://doi.org/10.1364/BOE.8.000679>
8. Chung, H., Ryu, D., McCann, M.T., Klasky, M.L., Ye, J.C.: Solving 3D Inverse Problems Using Pre-Trained 2D Diffusion Models. In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 22542–22551 (Jun 2023). <https://doi.org/10.1109/CVPR52729.2023.02159>
9. Corona-Figueroa, A., Frawley, J., Taylor, S.B., Bethapudi, S., Shum, H.P.H., Willcocks, C.G.: MedNeRF: Medical Neural Radiance Fields for Reconstructing 3D-aware CT-Projections from a Single X-ray. In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. pp. 3843–3848 (Jul 2022). <https://doi.org/10.1109/EMBC48229.2022.9871757>
10. Feldkamp, L.A., Davis, L.C., Kress, J.W.: Practical cone-beam algorithm. *Journal of the Optical Society of America A* **1**(6), 612 (Jun 1984). <https://doi.org/10.1364/JOSA.1.000612>
11. Finateu, T., Niwa, A., Matolin, D., Tsuchimoto, K., Mascheroni, A., Reynaud, E., Mostafalu, P., Brady, F., Chotard, L., LeGoff, F., Takahashi, H., Wakabayashi, H., Oike, Y., Posch, C.: A 1280×720 Back-Illuminated Stacked Temporal Contrast Event-Based Vision Sensor with 4.86µm Pixels, 1.066GEPS Readout, Programmable Event-Rate Controller and Compressive Data-Formatting Pipeline. In: *2020 IEEE International Solid-State Circuits Conference - (ISSCC)*. pp. 112–114. IEEE (Feb 2020). <https://doi.org/10.1109/ISSCC19947.2020.9063149>
12. Gallego, G., Delbruck, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., Leutenegger, S., Davison, A., Conradt, J., Daniilidis, K., Scaramuzza, D.: Event-based vision: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(1), 154–180 (Jan 2022). <https://doi.org/10.1109/TPAMI.2020.3008413>

13. García-Moreno, F., Kamm, P.H., Neu, T.R., Bülk, F., Noack, M.A., Wegener, M., von der Eltz, N., Schlepütz, C.M., Stampanoni, M., Banhart, J.: Tomoscopy: Time-resolved tomography for dynamic processes in materials. *Advanced Materials* **33**(45), 2104659 (2021). <https://doi.org/10.1002/adma.202104659>
14. Hwang, I., Kim, J., Kim, Y.M.: Ev-NeRF: Event Based Neural Radiance Field. In: 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 837–847 (Jan 2023). <https://doi.org/10.1109/WACV56688.2023.00090>
15. Kak, A.C., Slaney, M.: Principles of computerized tomographic imaging. No. 33 in *Classics in applied mathematics*, Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), Philadelphia, Pa (2001). <https://doi.org/10.1137/1.9780898719277>
16. Klenk, S., Koestler, L., Scaramuzza, D., Cremers, D.: E-NeRF: Neural Radiance Fields From a Moving Event Camera. *IEEE Robotics and Automation Letters* **8**(3), 1587–1594 (Mar 2023). <https://doi.org/10.1109/LRA.2023.3240646>
17. Kudo, H., Suzuki, T., Rashed, E.A.: Image reconstruction for sparse-view CT and interior CT—introduction to compressed sensing and differentiated backprojection. *Quantitative Imaging in Medicine and Surgery* **3**(3), 147–161 (Jun 2013). <https://doi.org/10.3978/j.issn.2223-4292.2013.06.01>
18. Lai, S., Tian, X., Wu, Q., Du, C., Xu, X., Wei, H., Guan, X., Zhang, Y.: Reconstructing Knee CT Volumes from Biplanar X-Rays Via Self-Supervised Neural Field. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–5 (May 2024). <https://doi.org/10.1109/ISBI56570.2024.10635353>
19. Lee, S., Chung, H., Park, M., Park, J., Ryu, W.S., Ye, J.C.: Improving 3D Imaging with Pre-Trained Perpendicular 2D Diffusion Models. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 10676–10686 (Oct 2023). <https://doi.org/10.1109/ICCV51070.2023.00983>
20. Lell, M.M., Kachelrieß, M.: Recent and Upcoming Technological Developments in Computed Tomography: High Speed, Low Dose, Deep Learning, Multienergy. *Investigative Radiology* **55**(1), 8 (Jan 2020). <https://doi.org/10.1097/RLI.0000000000000601>
21. Li, W., Wan, P., Wang, P., Li, J., Zhou, Y., Liu, P.: BeNeRF: Neural Radiance Fields from a Single Blurry Image and Event Stream. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (eds.) *Computer Vision – ECCV 2024*. vol. 15089, pp. 416–434. Cham (2024). https://doi.org/10.1007/978-3-031-72751-1_24
22. Liu, M., Wang, H., Yoon, K.J., Wang, L.: Disentangled Cross-modal Fusion for Event-Guided Image Super-resolution. *IEEE Transactions on Artificial Intelligence* **5**(10), 5314–5324 (Oct 2024). <https://doi.org/10.1109/TAI.2024.3418376>
23. Lu, Y., Wang, Z., Liu, M., Wang, H., Wang, L.: Learning Spatial-Temporal Implicit Neural Representations for Event-Guided Video Super-Resolution. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 1557–1567 (Jun 2023). <https://doi.org/10.1109/CVPR52729.2023.00156>
24. McCollough, C.H., Primak, A.N., Braun, N., Kofler, J., Yu, L., Christner, J.: Strategies for Reducing Radiation Dose in CT. *Radiologic Clinics of North America* **47**(1), 27–40 (Jan 2009). <https://doi.org/10.1016/j.rc1.2008.10.006>
25. Miao, X., Hu, J., Xu, Y., Su, J., Jing, Y.: Review on mechanical properties of metal lattice structures. *Composite Structures* **342**, 118267 (Aug 2024). <https://doi.org/10.1016/j.compstruct.2024.118267>
26. Mokso, R., Schlepütz, C.M., Theidel, G., Billich, H., Schmid, E., Celcer, T., Mikuljan, G., Sala, L., Marone, F., Schlumpf, N., Stampanoni, M.: GigaFRoST: the gi-

- gabit fast readout system for tomography. *Journal of Synchrotron Radiation* **24**(6), 1250–1259 (Nov 2017). <https://doi.org/10.1107/S1600577517013522>
27. Qi, Y., Zhu, L., Zhang, Y., Li, J.: E² NeRF: Event Enhanced Neural Radiance Fields from Blurry Images. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 13208–13218. IEEE, Paris, France (Oct 2023). <https://doi.org/10.1109/ICCV51070.2023.01219>
 28. Rebecq, H., Gallego, G., Mueggler, E., Scaramuzza, D.: EMVS: Event-Based Multi-View Stereo—3D Reconstruction with an Event Camera in Real-Time. *International Journal of Computer Vision* **126**(12), 1394–1414 (Dec 2018). <https://doi.org/10.1007/s11263-017-1050-6>
 29. Rebecq, H., Horstschaef, T., Gallego, G., Scaramuzza, D.: EVO: A Geometric Approach to Event-Based 6-DOF Parallel Tracking and Mapping in Real Time. *IEEE Robotics and Automation Letters* **2**(2), 593–600 (Apr 2017). <https://doi.org/10.1109/LRA.2016.2645143>
 30. Rudnev, V., Elgharib, M., Theobalt, C., Golyanik, V.: EventNeRF: neural radiance fields from a single colour event camera. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 4992–5002 (Jun 2023). <https://doi.org/10.1109/CVPR52729.2023.00483>
 31. Rudnev, V., Fox, G., Elgharib, M., Theobalt, C., Golyanik, V.: Dynamic Event-NeRF: Reconstructing General Dynamic Scenes from Multi-view RGB and Event Streams. In: *Computer Vision and Pattern Recognition Conference Workshop*. pp. 4866–4876 (2025). <https://doi.org/10.1109/CVPRW67362.2025.00478>
 32. Sidky, E.Y., Pan, X.: Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine and Biology* **53**(17), 4777–4807 (Sep 2008). <https://doi.org/10.1088/0031-9155/53/17/021>
 33. Wang, H., Hadjiivanov, A., Blazquez, E., Schlepütz, C.M., Stampanoni, M., Lovric, G.: Event-guided temporally super-resolved synchrotron X-ray imaging. *Communications Physics* **8**(1), 1–11 (May 2025). <https://doi.org/10.1038/s42005-025-02142-w>
 34. Withers, P.J., Bouman, C., Carmignato, S., Cnudde, V., Grimaldi, D., Hagen, C.K., Maire, E., Manley, M., Du Plessis, A., Stock, S.R.: X-ray computed tomography. *Nature Reviews Methods Primers* **1**(1), 1–21 (Feb 2021). <https://doi.org/10.1038/s43586-021-00015-4>
 35. Wu, S., Lu, Y., Guo, Y., Ji, W., Huang, S., Yang, F., Sirejiding, S., He, Q., Tong, J., Ji, Y., Ding, Y., Lu, H.: Discretized Gaussian Representation for Tomographic Reconstruction. In: 2025 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 25073–25082 (Oct 2025). <https://doi.org/10.1109/ICCV51701.2025.02325>
 36. Xu, C., Zhou, H., Chen, L., Chen, H., Hu, Z.Z., Lu, Z., Zhou, Y., Chung, V., Qu, Q., Cai, W.: A Survey of 3D Reconstruction with Event Cameras (Jun 2026). <https://doi.org/10.48550/arXiv.2505.08438>, arXiv:2505.08438 [cs.CV]
 37. Yan, H., Lu, Z., Chen, Z., Ma, D., Tang, H., Zheng, Q., Pan, G.: EvSTVSR: Event Guided Space-Time Video Super-Resolution. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. pp. 9085–9093 (Apr 2025). <https://doi.org/10.1609/aaai.v39i9.32983>
 38. Zha, R., Zhang, Y., Li, H.: NAF: Neural Attenuation Fields for Sparse-View CBCT Reconstruction. *Medical Image Computing and Computer Assisted Intervention (MICCAI)* pp. 442–452 (Sep 2022). https://doi.org/10.1007/978-3-031-16446-0_42

39. Zhao, T., Ding, G., Liu, Z., Hu, P., Wei, H., Tan, M., Ding, J.: DE-NAF: decoupled neural attenuation fields for sparse-view CBCT reconstruction. *Pattern Analysis and Applications* **28**(1), 35 (Jan 2025). <https://doi.org/10.1007/s10044-025-01416-x>
40. Zwanenburg, E.A., Williams, M.A., Warnett, J.M.: Review of high-speed imaging with lab-based x-ray computed tomography. *Measurement Science and Technology* **33**(1), 012003 (Nov 2021). <https://doi.org/10.1088/1361-6501/ac354a>