

EMGauss: Continuous Slice-to-3D Reconstruction via Dynamic Gaussian Modeling in Volume Electron Microscopy

Yumeng He Zanwei Zhou Yekun Zheng Chen Liang Yunbo Wang* Xiaokang Yang
 MoE Key Lab of Artificial Intelligence, AI Institute, School of Computer Science
 Shanghai Jiao Tong University

{ymhe, sjtu19zzw, beta_cat, no8irch6en, yunbow, xkyang}@sjtu.edu.cn

Abstract

Volume electron microscopy (vEM) enables nanoscale 3D imaging of biological structures but remains constrained by acquisition trade-offs, leading to anisotropic volumes with limited axial resolution. Existing deep learning methods seek to restore isotropy by leveraging lateral priors; yet their assumptions break down for morphologically anisotropic structures. We present **EMGauss**, a general framework for 3D reconstruction from planar scanned 2D slices with applications in vEM, which circumvents the inherent limitations of isotropy-based approaches. Our key innovation is to reframe slice-to-3D reconstruction as a 3D dynamic scene rendering problem based on Gaussian splatting, where the progression of axial slices is modeled as the temporal evolution of 2D Gaussian point clouds. To enhance fidelity in data-sparse regimes, we incorporate a **Teacher–Student bootstrapping mechanism** that uses high-confidence predictions on unobserved slices as pseudo-supervisory signals. Compared with diffusion- and GAN-based reconstruction methods, EMGauss substantially improves interpolation quality, enables continuous slice synthesis, and eliminates the need for large-scale pretraining. Beyond vEM, it potentially provides a generalizable slice-to-3D solution across diverse imaging domains.

1. Introduction

Volume electron microscopy (vEM) has revolutionized the understanding of biological ultrastructures, enabling nanoscale 3D imaging of cells, tissues, and even entire organisms [5, 9, 20, 30, 32]. However, due to the inherent trade-off among resolution, field of view, and acquisition time—commonly referred to as the “impossible triangle”—direct acquisition of isotropic data remains costly and inefficient [14, 27]. This often results in anisotropic volumes,

with axial (z) resolution substantially lower than the in-plane (xy) resolution.

Deep learning has emerged as a powerful tool to address this anisotropy by computationally enhancing axial resolution [6, 11, 13, 14, 21]. These methods aim to restore high-resolution details from inexpensive anisotropic acquisitions, thereby boosting imaging throughput while preserving structural fidelity. Conventional learning-based approaches typically rely on supervised training using paired high- and low-resolution volumes [14], or on partially isotropic samples [21], leading to strong dependence on well-curated data and poor adaptability across different domains and imaging conditions. In practice, isotropic reference data are rarely available, as the need for isotropic reconstruction mostly arises from the absence of such acquisitions. Self-supervised alternatives have been explored to alleviate the data reliance on isotropic data, including GAN-based [6, 11] and diffusion-based [21, 25] methods. Their key insight is to leverage the high-resolution xy-planes as an internal reference. As shown in Figure 1, to reconstruct a volume from anisotropic slices, they either perform *video frame-interpolation* between consecutive xy-slices to generate intermediate frames [6, 13], or apply *image super-resolution* to orthogonal views (e.g., xz or yz) to improve resolution along the z-axis [21]. Both strategies rely on the critical assumption that local tissue structures are approximately isotropic across the x, y, and z dimensions, such that geometric patterns in the high-resolution xy-planes can be extrapolated along the axial direction.

In practice, however, the assumption of isotropy is often violated in real biological specimens, where morphological anisotropy is prevalent—e.g., elongated neuronal fibers or dendritic spines exhibit strong directional continuity [1, 37]. Consequently, these methods suffer from limitations in accurately segmenting or analyzing such structures. This motivates the need for frameworks that move beyond 2D image priors to directly reason about 3D space and structural continuity, capable of preserving the true morphology of biological specimens.

*Corresponding author.

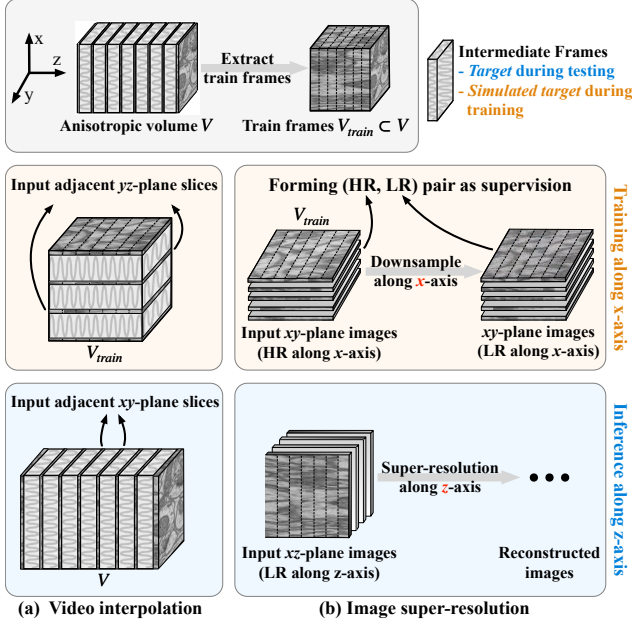


Figure 1. *Different paradigms of isotropic reconstruction from anisotropic datasets.* Existing approaches typically extract volume features from abundant xy -plane images. After stacking anisotropic slices along the z -axis (top), they: (a) **Video interpolation along the z -axis**: models are trained along the x or y direction, where adjacent frames are input to synthesize the intermediate frame; during inference, the trained model is repurposed to interpolate along the z -axis, generating intermediate slices corresponding to continuous temporal coordinates ($\tau \in [0, 1]$). (b) **Image super-resolution on xz/yz planes**: Since the xy -plane images are of much higher resolution, a low-resolution counterpart can be generated either by manual downsampling [21] or by learning a degradation model [6], forming paired HR–LR supervision for training; during inference, the resulting super-resolution model is then applied along the z -axis to enhance xz/yz slices. Both paradigms implicitly assume spatial isotropy. In contrast, our method directly performs inference in continuous 3D space, without relying on such isotropic assumptions.

In this paper, we introduce EMGauss, a novel framework that reformulates anisotropic volume reconstruction as a *dynamic 3D scene rendering* problem via dynamic Gaussian splatting. Instead of processing individual slices separately, we represent the entire axial slice sequence as the temporal evolution of a dynamic 2D Gaussian point cloud. By enriching each Gaussian with dynamic attributes that allow deformation, spatial shifting, and opacity variation across adjacent slices, our model implicitly recovers the underlying continuous 3D structure and synthesizes intermediate slices at arbitrary depths without relying on large models or external datasets. To further improve fidelity under sparse supervision, we employ a *teacher-student pseudo-labeling scheme* in which an EMA-based teacher provides stable pseudo targets for unseen slices, progressively guiding the student toward smoother and more consistent interpolation

along the axial dimension. This strategy effectively regularizes the network against overfitting to the limited observed slices and promotes smooth, geometry-aware interpolation across the volume.

In this way, our method fundamentally circumvents the domain mismatch issue inherent in prior isotropic-assumed models. Moreover, EMGauss offers two additional advantages. First, unlike methods such as EMDiffuse [21], which require additional isotropic sub-volumes of the same tissue for training, our approach operates in a fully self-contained optimization loop that uses only the anisotropic slices from the target volume. This makes it suitable for data-scarce or computationally constrained scenarios without the need for large-scale pretraining or auxiliary datasets. Second, by explicitly modeling geometric and photometric evolution across slices, our method enforces spatial continuity and physical coherence in the reconstructed volume, leading to stable training and reduced artifacts in the final reconstructed volume.

Extensive experiments on multiple vEM datasets demonstrate that EMGauss achieves lower test-time reconstruction error and delivers superior visual quality compared to existing methods, producing more realistic details with fewer artifacts. By providing a stable and highly generalizable solution, our framework not only advances the state of the art in vEM reconstruction but also establishes a scalable paradigm for slice-to-3D reconstruction across diverse imaging modalities.

2. Related Work

2.1. Isotropic Reconstruction of Volume EM

Volume Electron Microscopy (vEM) refers to a suite of techniques for generating 3D volumes of biological tissues, which are foundational for analyzing complex 3D cellular structures [9, 20, 32, 41]. A standard vEM workflow consists of sequentially sectioning and imaging a specimen, then computationally aligning and reconstructing them into a coherent 3D volume [5, 27]. However, achieving isotropic reconstruction for large volumes remains elusive, as high-throughput methods like ssTEM [2], ssSEM [12] and SBF-SEM [7] prioritize speed over Z -axis fidelity, while isotropic techniques like enhanced FIB-SEM [3] are limited to small volumes and require specialized instrumentation [27]. Thus, achieving isotropic reconstruction for biological specimens remains a persistent challenge.

Early approaches addressed anisotropic reconstruction through slice interpolation or optical-flow-based alignment [4, 8, 10, 16]. While straightforward to implement, these methods often introduce severe artifacts under high anisotropy factors. Deep learning approaches initially relied on fully supervised training. For instance, 3DSRUNet [14] learns interpolation from paired high- and low-resolution

volumes and achieves strong performance on synthetic data. However, acquiring such paired isotropic/anisotropic volumes from the same tissue requires specialized imaging protocols, such as dual-mode acquisition or repeated scans, which are rarely feasible for most vEM systems [27]. Moreover, synthetic degradations (e.g., row removal or Gaussian blur) poorly approximate real imaging processes, leading to domain gaps and limited generalization. These limitations motivate self-supervised approaches that avoid the need for isotropic training data.

Early self-supervised methods [19, 34] construct pseudo HR-LR pairs internally, for example by masking pixels or exploiting axis re-interpretation. However, they typically assume a known point spread function (PSF), whereas the true degradation in practice is often unknown and deviates from handcrafted models. Alternative strategies such as CycleGAN [6] attempt to learn the degradation process directly from target tissue data. However, adversarial training can introduce instability and artifacts, including texture aliasing and geometric distortion. More recently, diffusion-based models [21, 25] learn generative priors from abundant xy -plane images and leverage them to enhance axial resolution. Although these approaches improve fidelity and stability compared with GAN-based methods, they operate in a slice-wise 2D manner and implicitly assume spatial isotropy. As a result, degradations learned in the lateral (xy) plane may not transfer well to the axial (z) direction. In contrast, our method directly models continuous 3D geometry across slices, avoiding isotropy assumptions and enabling more faithful reconstruction of anisotropic structures.

2.2. 3D Gaussian Splatting and Dynamic Variants

Recent advances in neural rendering have been significantly driven by the emergence of 3D Gaussian Splatting (3DGS) and its extensions to dynamic scenes. 3DGS [17] represents a scene as a set of anisotropic 3D Gaussians parameterized by position, covariance, opacity, and spherical harmonic coefficients for view-dependent appearance. This explicit representation enables efficient optimization and high-quality real-time rendering, substantially improving both speed and fidelity compared with earlier explicit [24] and implicit [23] scene representations.

To handle dynamic scenes, subsequent works introduce temporal modeling through deformation fields that map canonical Gaussians to time-dependent observations. Deformable 3D Gaussians [36], for example, employ a lightweight multi-layer perceptron (MLP) to predict per-Gaussian deformations conditioned on timestamps, enabling reconstruction of non-rigid motions from monocular videos. Similarly, 4D Gaussian Splatting (4DGS) [35] models motion using a 4D neural voxel representation decomposed with a HexPlane-style encoding, followed by a compact MLP decoder for efficient deformation prediction.

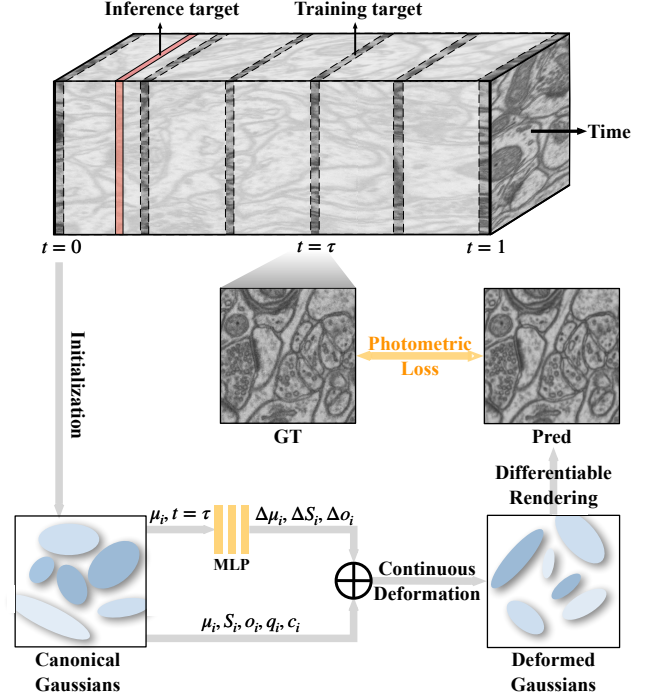


Figure 2. *Overview of the pipeline of EMGauss.* Given a volume EM dataset (top), the gray-shaded regions denote the missing intermediate slices to be interpolated, while the visible regions represent the available anisotropic training data. We interpret the low-resolution z -axis as a temporal dimension and employ a **deformable 2D Gaussian splatting** representation to model the geometric evolution of 2D structures along depth. During inference, the model synthesizes the target slice by directly conditioning on the relative temporal coordinate corresponding to the desired interpolation position.

These approaches extend the efficiency of Gaussian primitives to dynamic scene reconstruction while explicitly modeling temporal variation.

Beyond volumetric rendering, recent studies also explore Gaussian splatting in the 2D image domain for tasks such as image restoration, compression, and super-resolution [15, 39, 40, 43]. These works highlight the flexibility of Gaussian primitives and suggest their broader applicability to imaging problems beyond 3D scene reconstruction.

3. Method

3.1. Preliminaries

Our method builds upon Deformable 3D Gaussians [36], an extension of 3D Gaussian Splatting (3DGS) [17] for dynamic representations. 3DGS models a scene as a collection of explicit 3D Gaussian primitives, each parameterized by an opacity $o \in [0, 1]$, a center $\mu \in \mathbb{R}^{3 \times 1}$, and a covariance matrix $\Sigma \in \mathbb{R}^{3 \times 3}$:

$$G(\mathbf{X}) = \exp \left[-\frac{1}{2} (\mathbf{X} - \mu)^\top \Sigma^{-1} (\mathbf{X} - \mu) \right]. \quad (1)$$

The covariance Σ is typically decomposed as $\Sigma = RSS^\top R^\top$, where $S \in \mathbb{R}^{3 \times 1}$ denotes the scaling factors and $R \in \mathbb{R}^{3 \times 3}$ is a quaternion-parameterized rotation matrix, enabling efficient gradient-based optimization. For image rendering from a given viewpoint, Gaussians within the camera frustum are projected to the image plane via a splatting process. The projected 2D covariance Σ' is given by $\Sigma' = JW\Sigma W^\top J^\top$, where W and J denote the view transformation and the Jacobian of the local projective mapping, respectively. Each pixel color $\mathbf{p}$ is then obtained by alpha-blending the overlapping Gaussians $G_{i=1}^N$ in front-to-back order.

Building on this foundation, Deformable 3D Gaussians [36] extend 3DGS to dynamic scenes by learning a canonical Gaussian set $\mathcal{G}_C = (\mu_i, S_i, q_i, o_i, C_i)_{i=1}^M$ along with a lightweight MLP-based deformation network Φ_θ . Conditioned on a temporal coordinate $t \in [0, 1]$, Φ_θ predicts per-Gaussian offsets:

$$\Delta\mu_i, \Delta S_i, \Delta q_i, \Delta o_i = \Phi_\theta(\mu_i, t), \quad (2)$$

The deformed Gaussian at time t is then obtained by applying these deltas to each attribute correspondingly. This formulation enables smooth temporal interpolation while preserving fine-grained spatial details.

3.2. Slice-to-3D Dynamic Gaussians

Given a set of anisotropic slices I_t sparsely sampled along the z-axis, our goal is to reconstruct a continuous isotropic volume $\mathcal{V}$ by learning a temporally continuous deformation field that bridges adjacent slices. Unlike dynamic 3D reconstruction tasks, which deform Gaussians over time for 3D motion modeling, our deformation field is designed to parameterize spatial continuity along the axial (z) dimension. Specifically, we treat the slice index $t \in [0, 1]$ as a normalized spatial coordinate along the z-axis, and train a deformation MLP to learn the local geometric variations between neighboring slices.

As shown in Figure 2, starting from a canonical Gaussian set $\mathcal{G}_C = (\mu_i, S_i, R_i, o_i, C_i)_{i=1}^M$ initially optimized from the observed anisotropic frames, the deformation network Φ_θ predicts local offsets for each Gaussian as:

$$\Delta\mu_i, \Delta S_i, \Delta o_i = \Phi_\theta(\mu_i, t), \quad (3)$$

where $\Delta\mu_i = (\Delta x_i, \Delta y_i, 0)$ encodes lateral displacements, $\Delta S_i = (\Delta s_x, \Delta s_y, 0)$ controls in-plane scaling, and Δo_i modulates the opacity change along z .

To maintain training stability and prevent degenerate deformation along the depth axis, we fix both the absolute z-coordinate and the z-scale of all Gaussians to shared constants, while allowing rotation to be learned but not time-

Algorithm 1 EMA-based Teacher-Student Bootstrapping.

- 1: **Input:** Training set $\{\mathcal{D}_i\}_{1:N}$, where $\mathcal{D}_i = \{\text{image } I_i, \text{timestamp } t_i \in [0, 1]\}$
 - 2: **Hyperparameters:** Learning rates η , EMA decay α , Loss weight λ_{pseudo}
 - 3: **Parameters:** Gaussian attributes $\{\mathcal{G}_j\}_{1:M}$ where $G_j = (\mu_i, S_i, R_i, o_i, C_i)$, Pretrained deformation network $\Phi_{\theta_{\text{student}}}$
 - 4: Initialize teacher network $\Phi_{\theta_{\text{teacher}}} \leftarrow \Phi_{\theta_{\text{student}}}$
 - 5: **while** not converged **do**
 - 6: Sample an unseen timestamp $\tilde{t} \notin \{\mathcal{D}_i\}_{1:N}$
 - 7: $\mathcal{L}_{\text{pseudo}} = \lambda_{\text{pseudo}} \|\Phi_{\theta_{\text{student}}}(\{\mathcal{G}_j\}, \tilde{t}) - \Phi_{\theta_{\text{teacher}}}(\{\mathcal{G}_j\}, \tilde{t})\|$
 - 8: $\theta_{\text{student}} \leftarrow \theta_{\text{student}} - \eta \nabla_{\theta_{\text{student}}} \mathcal{L}_{\text{pseudo}}$
 - 9: $\theta_{\text{teacher}} \leftarrow \alpha \cdot \theta_{\text{teacher}} + (1 - \alpha) \cdot \theta_{\text{student}}$
 - 10: **end while**
-

dependent. Formally:

$$z_i = z_0, \quad (4)$$

$$s_{z,i} = s_{z,0}, \quad (5)$$

$$\Delta z_i = \Delta s_z = \Delta R_i = 0, \quad (6)$$

where z_0 and $s_{z,0}$ are global constants shared across all Gaussians. This ensures consistent axial alignment while enabling learnable in-plane deformation and per-slice appearance variation.

During inference, we query Φ_θ with intermediate values of t to generate deformed Gaussian sets $\mathcal{G}_t$, which are then rendered via Gaussian splatting to produce interpolated slices $\hat{I}_t$. This process enables continuous z-axis upsampling without requiring any additional training data beyond the given anisotropic volume.

3.3. Bootstrapping with Pseudo Gaussian Targets

To further enhance the fidelity of interpolated slices in data-sparse regimes, we introduce a *Teacher-Student pseudo-labeling framework* that leverages high-confidence predictions on unobserved intermediate slices as supervisory signals. It can be viewed as a bootstrapping method as we use slice predictions to generate new slice predictions. This approach is particularly effective in our setting, where only a small subset of axial slices (e.g., 10% – 20%) are provided during training, leaving the majority of the z-axis unobserved.

As shown in Algorithm 1, we maintain a stable teacher model as an exponential moving average (EMA) version of the student model (the actively trained model). The teacher parameters θ_{teacher} are updated at each optimization step (Line 9), where $\alpha = 0.995$ is the EMA decay rate. Concretely, the student is trained with a matching pseudo-label loss on selected unseen slices, which is detailed in Section 3.4. The teacher model is initialized when the training iteration reaches a predefined threshold, ensuring initial convergence before introducing pseudo-supervision. After a short

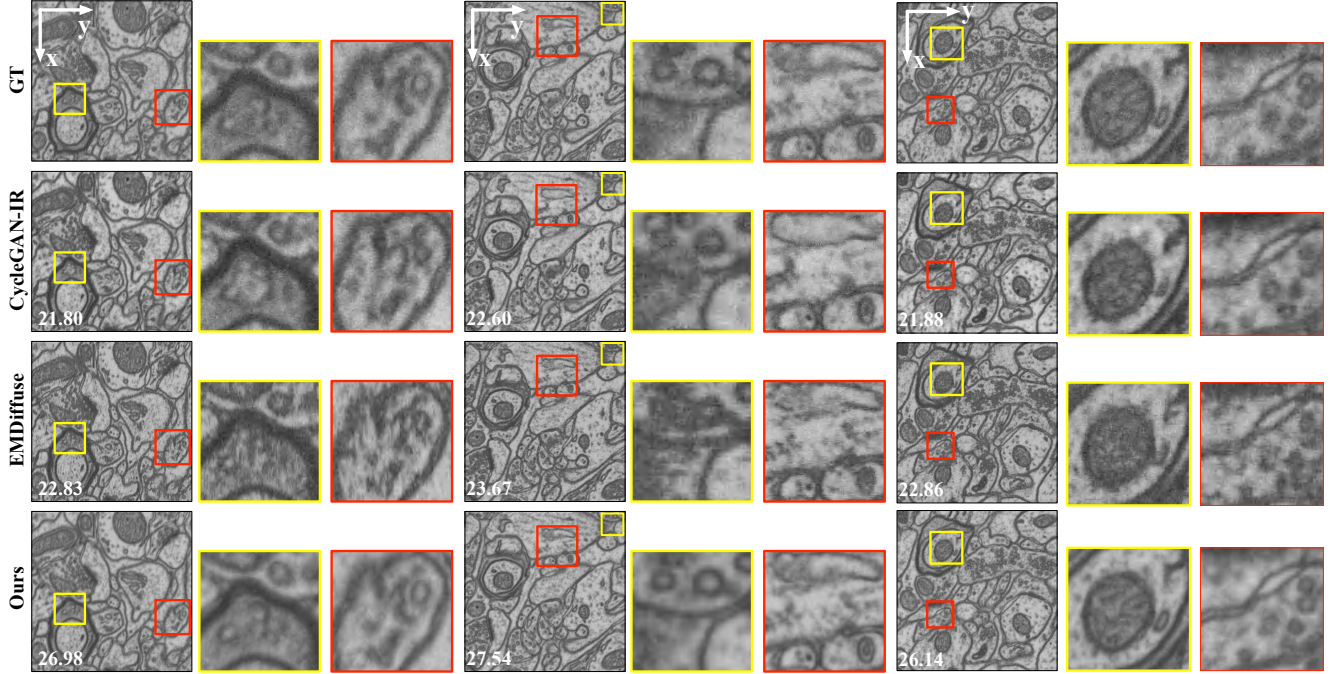


Figure 3. *Isotropic xy-slice reconstruction results on EPFL dataset [22].*

warm-up, we activate the EMA teacher to generate pseudo targets on a subset of unseen slices and gradually include more unseen indices as training progresses (from mid-slices to other interleaving positions). Note that the algorithm describes the pseudo-supervision iterations, where the student is optimized using the teacher’s predictions on non-training frames. In practice, these pseudo-supervised iterations are interleaved with standard iterations using ground-truth training data, ensuring balanced learning and stable convergence.

This Teacher-Student framework enables robust learning of the deformation field by providing stable pseudo-supervision for unobserved intermediate slices, effectively regularizing the model against overfitting to sparse training data while facilitating smooth interpolation along the axial dimension.

3.4. Training Pipeline

We adopt a three-stage training pipeline. First, we optimize the canonical Gaussian set $\mathcal{G}_c$ and the rendering parameters on the observed slices $\mathcal{T}$ while freezing the deformation MLP, establishing a stable radiance baseline. Next, we enable the deformation MLP Φ_θ and jointly train it with $\mathcal{G}_c$ on $\mathcal{T}$ so that in-plane position offsets ($\Delta x, \Delta y$), in-plane scales offsets ($\Delta s_x, \Delta s_y$), and opacity offsets capture the observed axial transitions. This second stage is deliberately brief: extending it tends to overfit and trap the model in suboptimal local minima, yet omitting it leads to an initially weak teacher that hinders convergence. Fi-

nally, we activate the EMA teacher to generate pseudo targets on a progressively enlarged subset of unseen timestamps. Training alternates between supervised iterations using ground-truth samples and pseudo-supervised iterations with teacher-generated labels, gradually increasing the proportion of pseudo-supervised steps as the model stabilizes.

Objective functions. Our model can effectively function with only RGB photometric supervision. Following 3DGS [17], we compute the deviation between the rendered image and the ground-truth RGB frame using an ℓ_1 loss combined with a D-SSIM regularization term. For pseudo-supervised iterations, we apply the same objective but scale the loss weight λ_{pseudo} by an exponentially increasing factor, gradually enhancing the impact of pseudo supervision as training progresses.

Implementation details. We implement EMGauss in PyTorch [26] and optimize with Adam [18] following the default hyperparameters of 3DGS [17] for Gaussian attributes and learning rates. Training proceeds in three phases: a warm-up phase of 2k iterations optimizing the canonical Gaussian set $\mathcal{G}_c$ with the deformation MLP frozen, a short joint-training phase of 1k iterations enabling the $\mathcal{G}_c$ and the deformation MLP, and a final phase of 15k iterations with the EMA teacher providing pseudo supervision on unseen timestamps until convergence. During the pseudo-supervised phase, the corresponding loss weight is linearly ramped from 0.1 to 1.0 between 3k and 10k iterations. The EMA teacher employs a decay factor of $\alpha = 0.995$. All experiments are conducted on an NVIDIA RTX 3090 GPU.

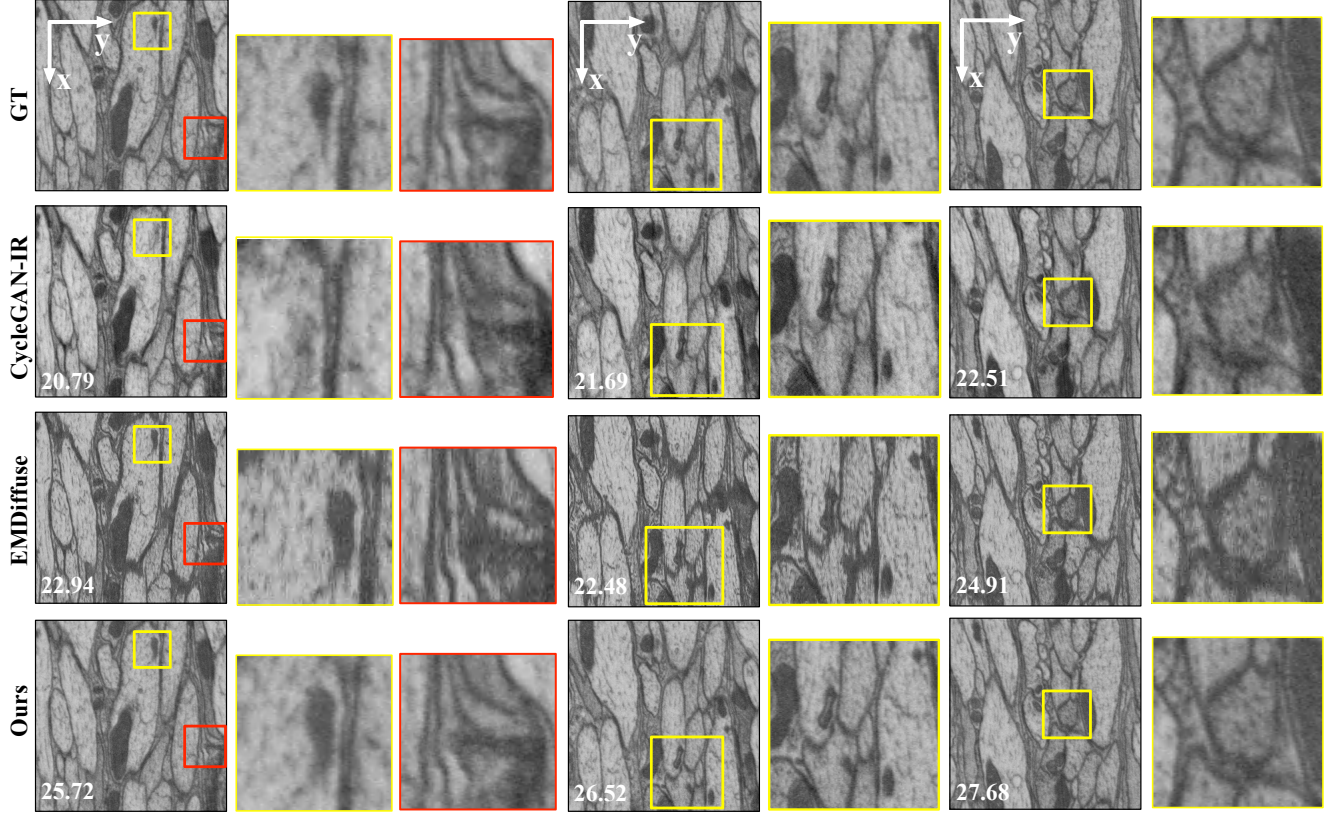


Figure 4. Isotropic xy-slice reconstruction results on FIB-25 dataset [31].

4. Results

4.1. Experimental Settings

Datasets. We conduct experiments on diverse large-scale electron microscopy (EM) datasets under both isotropic and anisotropic imaging conditions, as shown in Table 1. For isotropic data, we select the *EPFL mouse brain dataset* [22] acquired from the CA1 hippocampus region of the mouse brain with a uniform voxel resolution of $5 \times 5 \times 5$ nm, and the *FIB-25 drosophila brain dataset* [31] with a resolution of $8 \times 8 \times 8$ nm. Both datasets are collected by FIB-SEM. To simulate anisotropic acquisition, we subsample the z -axis of these isotropic volumes by a factor of $\times 6$, using the withheld intermediate slices as ground truth for evaluation. Following CycleGAN-IR [6], we use a $500 \times 500 \times 500$ sub-volume from each dataset. To further assess performance under real-world anisotropic imaging, we use the *FANC dataset* [28] acquired from the ventral nerve cord of an adult female *Drosophila* with an anisotropy ratio of 10 ($4 \times 4 \times 40$ nm), where a $500 \times 500 \times 50$ sub-volume is used for experiments. These datasets collectively span multiple organisms, tissue types, and imaging resolutions, providing a robust benchmark for both simulated and real isotropic reconstruction. We use PSNR, SSIM [33], and FSIM [38] for quantitative evaluation.

Baselines. We compare our method with representative isotropic reconstruction approaches spanning GAN-based, diffusion-based, and transformer-based paradigms, including vEMDiffuse-a from EMDiffuse [21], CycleGAN-IR [6], and IsoVEM [13]. vEMDiffuse-a is a diffusion-based model that learns anisotropic-to-isotropic mappings using additional training data from the same tissue type. CycleGAN-IR is an unsupervised GAN-based framework built on CycleGAN [42] that learns cross-domain mappings between high-resolution lateral and low-resolution axial slices. IsoVEM [13] is a transformer-based approach designed for anisotropic EM volume reconstruction. All baselines rely on large-scale datasets and extensive training, whereas our method generates intermediate slices using only a small sub-batch from the target volume, greatly reducing data and training cost.

4.2. Isotropic Reconstruction for Simulated Anisotropic vEM Data

Results on xy-slices. We evaluate all the methods on the xy-slice images, as shown in Figure 3, Figure 4, and Table 2. In Figure 3 and Figure 4, compared with CycleGAN-IR and EMDiffuse, our method can generate more high-quality images that are closer to the GT images. The generated images from CycleGAN-IR contain some structures that do

Table 1. *Dataset description.* We use both isotropic and anisotropic datasets for our experiments. For isotropic datasets, we simulate anisotropies via manual sub-sampling.

Datasets	Isotropy	Resolution (nm)	Anisotropy ratio	Issue
EPFL [22]	✓	$5 \times 5 \times 5$	$6\times$	brain
FIB-25 [31]	✓	$8 \times 8 \times 8$	$6\times$	brain
FANC [28]	×	$4 \times 4 \times 40$	$10\times$	nerve

Table 2. *Results of isotropic xy-slice reconstruction on synthetic anisotropic datasets.* [†]EMDiffuse uses additional data for training.

Method	EPFL			FIB-25		
	PSNR	SSIM	FSIM	PSNR	SSIM	FSIM
CycleGAN-IR[6]	22.05	0.491	0.856	22.39	0.554	0.856
EMDiffuse[21] [†]	23.34	0.519	0.899	24.10	0.514	0.878
IsoVEM [13]	23.91	0.597	0.8558	21.51	0.546	0.8456
Ours	26.59	0.6977	0.943	27.37	0.7275	0.920

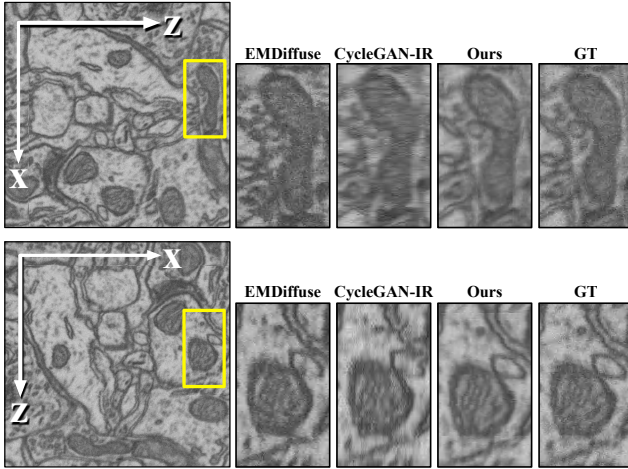


Figure 5. *Isotropic xz/yz-slice Reconstruction results on synthetic anisotropic datasets.*

not belong to the GT images, which not only leads to a loss in image metrics but also more easily affects the results of subsequent 3D reconstruction. The images generated by EMDiffuse exhibit relatively obvious artifacts, with details and structures significantly differing from the ground truth images. This is because the training and testing domains of this method are different: the training set used downsampled xz/yz -slice images, causing the structural details on the test set to resemble those of the downsampled images. This severely degrades the quality of the generated images and poses difficulties for subsequent 3D reconstruction. Our method avoids the use of xz/yz slices and instead performs continuous modeling directly on xy -slices, thereby avoiding the problem of differing training and testing domains. Compared with the other two baseline methods, the images generated by our method contain neither extraneous structures nor blurry details, thus achieving higher image metrics and facilitating the downstream 3D reconstruction process.

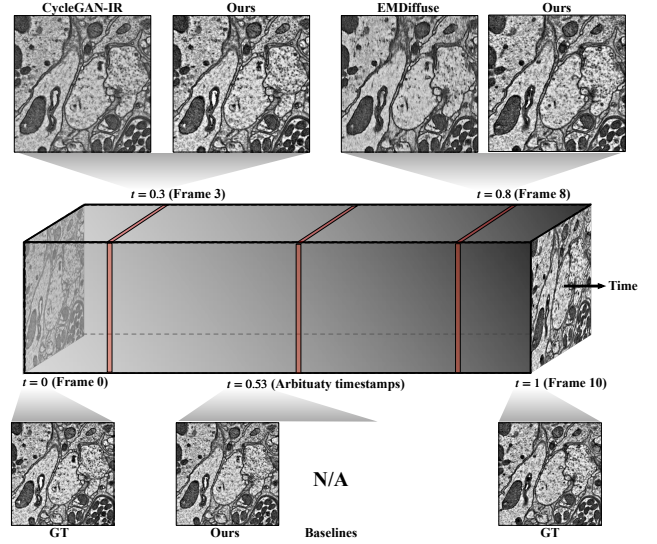


Figure 6. *Isotropic xy-slice Reconstruction results on real-anisotropic FANC datasets with anisotropic factor $10\times$.* Notably, beyond the predefined slice intervals, our model can freely interpolate frames at any arbitrary time step (e.g. $t = 0.53$).

Results on xz/yz -slices. We further evaluate all methods on reconstructed xz/yz slices, as shown in Figure 5. Although EMDiffuse [21] and CycleGAN-IR [6] are trained using xz/yz slices, the generated results still exhibit noticeable artifacts and structural inconsistencies. This suggests that the downsampled xz/yz slices differ substantially from the original slices, making it difficult to accurately model the underlying volumetric structure. In contrast, although our method is trained only on xy slices, the proposed 3D continuous deformation field enables the model to capture coherent volumetric deformation across the entire volume. As a result, our method produces smoother and more realistic reconstructions on the xz/yz views.

4.3. Isotropic Reconstruction for Real Anisotropic vEM Data

To demonstrate the generalizability of our method to real-world scenarios, we evaluate all methods on the real anisotropic Volume EM dataset FANC [28]. The results are shown in Figure 6. Compared with CycleGAN-IR [6] and EMDiffuse [21], our method produces more accurate and structurally detailed reconstructions, indicating that it remains effective on real anisotropic data. Moreover, only our method enables continuous generation along the z -axis, allowing images to be synthesized at an arbitrary timestamp t . As illustrated in Figure 6, when evaluated at an arbitrary timestamp $t = 0.53$, CycleGAN-IR [6] and EMDiffuse [21] cannot produce valid results because they rely on discrete xz/yz slice interpolation. In contrast, our method models the entire 3D volume and can therefore generate consistent slices at any timestamp.

Table 3. Video segmentation results on EPFL datasets with synthetic anisotropy ratio of $\times 6$. Segmentation is performed on the reconstructed isotropic slice sequences using SAM2 [29], and IoU is computed against the ground-truth intermediate slices.

method	CycleGAN-IR	EMDiffuse	Ours
IoU	0.9099	0.9555	0.9687

Table 4. Ablation studies on key components of our framework. The results are averaged over the two isotropic datasets.

Method	PSNR	SSIM	FSIM
w/o teacher-student module	25.19	0.6272	0.9035
w/o Warmup	25.76	0.6528	0.9076
w/o Joint	24.35	0.5773	0.8513
w/o Δo	25.44	0.6295	0.8936
w/ ΔR	25.07	0.6398	0.9055
Ours	26.98	0.7126	0.9315

4.4. Downstream Segmentation Results

Isotropic slice interpolation is essential for accurate 3D reconstruction in volume electron microscopy (vEM). To evaluate downstream utility, we follow [21] and perform video-based segmentation on the reconstructed volumes. Specifically, we adopt Segment Anything Model 2 (SAM2) [29] to segment cellular structures across continuous xy slices. With ground-truth intermediate slices available, we compute their segmentation masks and measure the Intersection-over-Union (IoU) between predictions and ground truth to assess segmentation accuracy. The results are summarized in Table 3. We further provide 3D reconstruction visualizations and comparisons in the *supplementary video*. These experiments collectively demonstrate that high-quality anisotropy correction is crucial for downstream 3D vEM analysis, and that our reconstruction pipeline yields more coherent structural continuity and higher morphological fidelity than existing methods.

4.5. Ablation Study

We conduct ablation studies to systematically evaluate the contribution of key components in our framework, including the teacher-student module, training pipeline design, and Gaussian configuration. As shown in Table 4, removing the teacher-student module leads to a noticeable degradation in generation quality and structural fidelity. Skipping the warm-up stage results in coarse initialization and compensatory Gaussian proliferation, while omitting joint training yields low-quality pseudo-labels and suboptimal convergence. For our Gaussian configuration, we adopt dynamic opacity with time-independent rotation. Static opacity introduces noise when modeling structural appearance and disappearance, whereas dynamic rotation causes temporal jitter. Further ablations on different anisotropy ratios ($\times 4$ and $\times 8$) and the number of training slices per bundle are provided in the *supplementary materials*.

5. Discussion

Our method provides a general and flexible framework for reconstructing 3D structures from planar-scanned 2D slices, with a particular focus on applications in vEM. Unlike prior approaches that rely on anisotropic priors or heuristic slice-to-slice matching, our formulation reconstructs the underlying 3D volume in a more principled manner by explicitly modeling a continuous deformation field directly from the high-resolution xy slices. This design not only yields smoother inter-slice transitions but also enables a more faithful recovery of morphological variations across depth. Despite its advantages, our framework still exhibits several limitations. A primary drawback is that the number of Gaussian splatting primitives can grow significantly in certain challenging regions—especially when the input slices contain substantial noise. The resulting increase in Gaussian density may lead to excessive memory consumption. A practical remedy is to incorporate a lightweight denoising module before our reconstruction pipeline, which can stabilize the optimization and prevent the uncontrolled proliferation of Gaussian primitives. In addition, future extensions may explore adaptive Gaussian pruning or joint learning with image-space regularizers to further improve robustness and computational efficiency.

6. Conclusion

We introduced EMGauss, a simple yet powerful framework for anisotropic volume reconstruction in volume electron microscopy. By recasting slice interpolation as dynamic 3D scene rendering with deformable Gaussian splatting, EMGauss removes the strong isotropy assumptions commonly embedded in previous approaches and provides a more principled, geometry-aware formulation of the reconstruction problem. Leveraging a self-supervised teacher-student bootstrapping scheme, our method learns directly from the target anisotropic slices without requiring external isotropic volumes or paired HR-LR data. This enables continuous, high-fidelity slice synthesis while suppressing artifacts that often arise in purely 2D formulations. Experiments demonstrate improved structural coherence across depth while preserving fine morphological details critical for vEM analysis. More broadly, EMGauss establishes a general and modality-agnostic paradigm for slice-to-volume reconstruction. The combination of deformable Gaussians, continuous depth modeling, and a self-supervised learning strategy offers a promising foundation for high-throughput 3D imaging in diverse scientific domains, enabling stable reconstruction without relying on large paired datasets or modality-specific supervision. We believe that the proposed framework opens up new possibilities for scalable, robust, and data-efficient 3D reconstruction pipelines across a wide range of planar acquisition systems.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant 62250062), the Smart Grid National Science and Technology Major Project (Grant 2024ZD0801200), the Shanghai Municipal Science and Technology Major Project (Grant 2021SHZDZX0102), and the Fundamental Research Funds for the Central Universities.

References

- [1] Christian Beaulieu. The basis of anisotropic water diffusion in the nervous system—a technical review. *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo*, 15(7-8):435–455, 2002. [1](#)
- [2] Kevin L Briggman and Winfried Denk. Towards neural circuit reconstruction with volume electron microscopy techniques. *Current opinion in neurobiology*, 16(5):562–570, 2006. [2](#)
- [3] Andrew J Bushby, Kenneth MY P’ng, Robert D Young, Christian Pinali, Carlo Knupp, and Andrew J Quantock. Imaging three-dimensional tissue architectures by focused ion beam scanning electron microscopy. *Nature protocols*, 6(6):845–858, 2011. [2](#)
- [4] Lucian Carata, Dan Shao, Markus Hadwiger, and Eduard Groeller. Improving the visualization of electron-microscopy data through optical flow interpolation. In *Proceedings of the 27th spring conference on computer graphics*, pages 103–110, 2011. [2](#)
- [5] Lucy M Collinson, Carles Bosch, Anwen Bullen, Jemima J Burden, Raffaella Carzaniga, Cheng Cheng, Michele C Darrow, Georgina Fletcher, Errin Johnson, Kedar Narayan, et al. Volume em: a quiet revolution takes shape. *Nature methods*, 20(6):777–782, 2023. [1](#), [2](#)
- [6] Shiyu Deng, Xueyang Fu, Zhiwei Xiong, Chang Chen, Dong Liu, Xuejin Chen, Qing Ling, and Feng Wu. Isotropic reconstruction of 3d em images with unsupervised degradation learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 163–173. Springer, 2020. [1](#), [2](#), [3](#), [6](#), [7](#)
- [7] Winfried Denk and Heinz Horstmann. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS biology*, 2(11):e329, 2004. [2](#)
- [8] Fisseha A Ferede, Ali Khalighifar, Jaison John, Krishnan Venkataraman, and Khaled Khairy. Z-upscaling: Optical flow guided frame interpolation for isotropic reconstruction of 3d em volumes. In *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2025. [2](#)
- [9] Jill R Glausier, Matthew Maier, Cedric Bouchet-Marquis, Ken Wu, Tabitha Banks-Tibbs, Darlene Melchitzky, Jiying Ning, David A Lewis, and Zachary Freyberg. Volume electron microscopy reveals 3d synaptic nanoarchitecture in postmortem human prefrontal cortex. *iScience*, 28(7), 2025. [1](#), [2](#)
- [10] Vicente González-Ruiz, Juan Pablo García-Ortiz, María Rosario Fernández-Fernández, and José-Jesús Fernández. Optical flow driven interpolation for isotropic fib-sem reconstructions. *Computer Methods and Programs in Biomedicine*, 221:106856, 2022. [2](#)
- [11] Katsumi Hagita, Takeshi Higuchi, and Hiroshi Jinnai. Super-resolution for asymmetric resolution of fib-sem 3d imaging using ai with deep learning. *Scientific reports*, 8(1):5877, 2018. [1](#)
- [12] Kenneth J Hayworth, Josh L Morgan, Richard Schalek, Daniel R Berger, David GC Hildebrand, and Jeff W Lichtman. Imaging atom ultrathin section libraries with wafermapper: a multi-scale approach to em reconstruction of neural circuits. *Frontiers in neural circuits*, 8:68, 2014. [2](#)
- [13] Jia He, Yan Zhang, Wenhao Sun, Ge Yang, and Fei Sun. IsoVem: isotropic reconstruction for volume electron microscopy based on transformer. *bioRxiv*, pages 2023–11, 2023. [1](#), [6](#), [7](#)
- [14] Larissa Heinrich, John A Bogovic, and Stephan Saalfeld. Deep learning for isotropic super-resolution from non-isotropic 3d electron microscopy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 135–143. Springer, 2017. [1](#), [2](#)
- [15] Jintong Hu, Bin Xia, Bin Chen, Wenming Yang, and Lei Zhang. Gaussiansr: High fidelity 2d gaussian splatting for arbitrary-scale image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3554–3562, 2025. [3](#)
- [16] Saurabh Joshi, André Forjaz, Kyu Sang Han, Yu Shen, Vasco Queiroga, Florin A Selaru, Marie Gérard, Daniel Xenos, Jordan Matelsky, Brock Wester, et al. Interpolai: deep learning-based optical flow interpolation and restoration of biomedical images for improved 3d tissue mapping. *Nature Methods*, pages 1–12, 2025. [2](#)
- [17] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. [3](#), [5](#)
- [18] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. [5](#)
- [19] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2129–2137, 2019. [3](#)
- [20] Xia Li, Feng Qiao, Jiansheng Guo, Ting Jiang, Huifang Lou, Huixia Li, Gangcai Xie, Hangjun Wu, Weizhen Wang, Ruoyu Pei, et al. In situ architecture of the intercellular organelle reservoir between epididymal epithelial cells by volume electron microscopy. *Nature Communications*, 16(1):1664, 2025. [1](#), [2](#)
- [21] Chixiang Lu, Kai Chen, Heng Qiu, Xiaojun Chen, Gu Chen, Xiaojuan Qi, and Haibo Jiang. Diffusion-based deep learning method for augmenting ultrastructural imaging and volume electron microscopy. *Nature Communications*, 15(1):4677, 2024. [1](#), [2](#), [3](#), [6](#), [7](#), [8](#)
- [22] Aurélien Lucchi, Yunpeng Li, and Pascal Fua. Learning for structured prediction using approximate subgradient descent

- with working sets. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1987–1994, 2013. [5](#), [6](#), [7](#)
- [23] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. [3](#)
- [24] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022. [3](#)
- [25] Mingjie Pan, Yulu Gan, Fangxu Zhou, Jiaming Liu, Ying Zhang, Aimin Wang, Shanghang Zhang, and Dawei Li. Diffuseir: diffusion models for isotropic reconstruction of 3d microscopic images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 323–332. Springer, 2023. [1](#), [3](#)
- [26] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *NeurIPS*, 32, 2019. [5](#)
- [27] Christopher J Peddie, Christel Genoud, Anna Kreshuk, Kimberly Meechan, Kristina D Micheva, Kedar Narayan, Constantin Pape, Robert G Parton, Nicole L Schieber, Yannick Schwab, et al. Volume electron microscopy. *Nature Reviews Methods Primers*, 2(1):51, 2022. [1](#), [2](#), [3](#)
- [28] Jasper S Phelps, David Grant Colburn Hildebrand, Brett J Graham, Aaron T Kuan, Logan A Thomas, Tri M Nguyen, Julia Buhmann, Anthony W Azevedo, Anne Sustar, Sweta Agrawal, et al. Reconstruction of motor control circuits in adult drosophila using automated transmission electron microscopy. *Cell*, 184(3):759–774, 2021. [6](#), [7](#)
- [29] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024. [8](#)
- [30] Florian Robert, Alexia Calovoulos, Laurent Facq, Fanny Decoeur, Etienne Gontier, Christophe F Grosset, and Baudouin Denis de Senneville. Improving cell instance segmentation in scanning electron microscopy via semantic image synthesis. In *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2025. [1](#)
- [31] Shin-ya Takemura, C Shan Xu, Zhiyuan Lu, Patricia K Rivlin, Toufiq Parag, Donald J Olbris, Stephen Plaza, Ting Zhao, William T Katz, Lowell Umayam, et al. Synaptic circuits and their variations within different columns in the visual system of drosophila. *Proceedings of the National Academy of Sciences*, 112(44):13711–13716, 2015. [6](#), [7](#)
- [32] Marta Turegano-Lopez, Felix de Las Pozas, Andrea Santuy, Jose-Rodrigo Rodriguez, Javier DeFelipe, and Angel Merchan-Perez. Tracing nerve fibers with volume electron microscopy to quantitatively analyze brain connectivity. *Communications Biology*, 7(1):796, 2024. [1](#), [2](#)
- [33] 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. [6](#)
- [34] Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broadbuss, Siân Culley, et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature methods*, 15(12):1090–1097, 2018. [3](#)
- [35] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20310–20320, 2024. [3](#)
- [36] Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20331–20341, 2024. [3](#), [4](#)
- [37] Hui Zhang, Torben Schneider, Claudia A Wheeler-Kingshott, and Daniel C Alexander. Noddi: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage*, 61(4):1000–1016, 2012. [1](#)
- [38] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 20(8):2378–2386, 2011. [6](#)
- [39] Pingping Zhang, Xiangrui Liu, Meng Wang, Shiqi Wang, Sam Kwong, et al. 2d gaussian splatting for image compression. *APSIPA Transactions on Signal and Information Processing*, 13(6), 2024. [3](#)
- [40] Xinjie Zhang, Xingtong Ge, Tongda Xu, Dailan He, Yan Wang, Hongwei Qin, Guo Lu, Jing Geng, and Jun Zhang. Gaussianimage: 1000 fps image representation and compression by 2d gaussian splatting. In *European Conference on Computer Vision*, pages 327–345. Springer, 2024. [3](#)
- [41] Jingjing Zhao, Xiaoping Yu, Xuping Shentu, and Danting Li. The application and development of electron microscopy for three-dimensional reconstruction in life science: a review. *Cell and Tissue Research*, 396(1):1–18, 2024. [2](#)
- [42] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. [6](#)
- [43] Lingting Zhu, Guying Lin, Jinnan Chen, Xinjie Zhang, Zhenchao Jin, Zhao Wang, and Lequan Yu. Large images are gaussians: High-quality large image representation with levels of 2d gaussian splatting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10977–10985, 2025. [3](#)