

Special Section on EG VCBM 2025

Deep learning for brain electron microscopy segmentation: Advances, challenges, and future directions in connectomics and ultrastructure analysis[☆]

Uzair Shah^a, Mahmood Alzubaidi^{a,*,1}, Marco Agus^{a,*}, Corrado Calí^b, Pierre J. Magistretti^c, Mowafa Househ^{a,*}

^a College of Science and Engineering, Hamad Bin Khalifa University, Education City, Doha, Qatar

^b Neuroscience Institute Cavalieri Ottolenghi, University of Turin, Italy

^c Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

ARTICLE INFO

Keywords:

Deep learning
Electron microscopy
Brain segmentation
Connectomics
Neural circuits
Convolutional neural networks
Transformers
Meta-analysis

ABSTRACT

This systematic review and meta-analysis comprehensively analyzes deep learning approaches for brain electron microscopy (EM) segmentation, addressing the critical challenge of extracting neuroanatomical information at nanometer resolution. Following PRISMA guidelines, we identified 60 studies through structured database searches, with quantitative meta-analysis of 27 studies (46 experiments) across 10 datasets providing the first unified benchmark comparison in this domain. Our analysis reveals a field transitioning from traditional CNN approaches toward foundation models and hybrid architectures. The meta-analysis demonstrates that foundation models outperform traditional CNNs by 13%–35% across key metrics, with the 3D Transformer + U-Net achieving the highest composite score (0.954) across five datasets. Meta-analysis confirms significant advantages for foundation models in instance-based metrics (Cohen's $d = -6.44$), while only 26% of experiments validate across multiple datasets. Four key evolutionary trends emerge: (1) transition from 2D to 3D architectures optimized for ultrastructural complexity; (2) development of topology-preserving loss functions and evaluation metrics (cDice, ERL) that prioritize neural connectivity over pixel-wise accuracy; (3) emergence of self-supervised and foundation model adaptation techniques reducing annotation dependency; and (4) evolution toward specialized architectures capturing long-range dependencies critical for neural structures. Performance analysis reveals that mitochondria segmentation achieves highest accuracy (Jaccard scores 87.2–90.5%), while computational requirements vary from single-GPU implementations to distributed systems with 48 GPUs for teravoxel-scale volumes. Despite progress, reproducibility challenges persist with only 54% of studies providing public code repositories. These advances drive innovation in 3D computer vision, establish new benchmarks for volumetric instance segmentation, and address fundamental challenges in processing massive biological datasets. Our unified benchmarks and comprehensive analysis provide a foundation for systematic progress tracking and evidence-based method selection, positioning brain EM segmentation to enable large-scale connectomics studies and detailed neuroanatomical mapping across scales.

1. Introduction

Understanding the brain's complex architecture—its synaptic connections, organelles, and nanoscale neuronal processes—is fundamental for decoding the biological basis of cognition, behavior, and neurological disease. Electron microscopy (EM) has emerged as a crucial modality for this task, enabling structural imaging at sub-nanometer resolution—far surpassing the capabilities of light microscopy [1,2]. Multiple specialized EM and complementary techniques have been

developed to capture the complicated details of neural tissue, each with distinct advantages for different brain mapping applications.

Serial section Transmission EM (ssTEM) involves imaging tissue sections that are computationally aligned, providing resolution for neural circuit reconstruction [3], while Transmission EM (TEM) enables examination of structures with electrons passing through specimens to image components [4]. Scanning EM (SEM) detects secondary electrons from specimen surfaces to create topography images [5], with

[☆] This article is part of a Special issue entitled: 'EG VCBM 2025' published in Computers & Graphics.

* Corresponding authors.

E-mail addresses: MAgus@hbku.edu.qa (M. Agus), mhouseh@hbku.edu.qa (M. Househ).

<https://doi.org/10.1016/j.cag.2025.104391>

Received 9 May 2025; Received in revised form 31 July 2025; Accepted 15 August 2025

Available online 1 September 2025

0097-8493/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

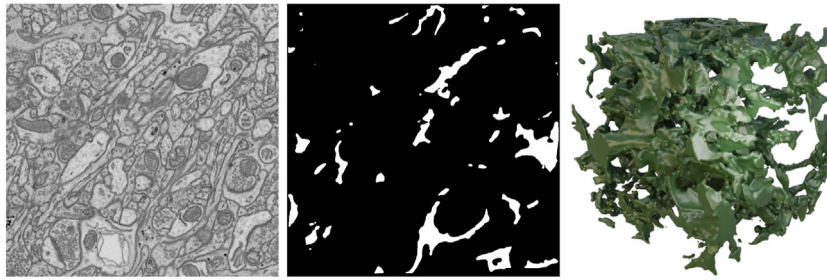


Fig. 1. Current SOTA segmentation methods for 3D EM are effective for well-defined cellular structures like mitochondria, while they struggle when segmenting complex processes with fractal structures like glia cells Shah et al. [21].

variants including Focused Ion Beam-SEM (FIB-SEM) that mills tissue layers between imaging for z-axis resolution, Serial Block-Face SEM (SBF-SEM) using ultra-microtome for section acquisition, Automated Tape-collecting Ultra-microtome SEM (ATUM-SEM) collecting sections on tape for imaging, and Multi-beam SEM employing multiple electron beams to increase imaging throughput. Together with EM, optical microscopy techniques including Fluorescence Micro-Optical Sectioning Tomography (fMOST) [6], light-sheet microscopy, confocal microscopy, and X-ray microscopy are used to image samples at different resolutions. Through all these specialized imaging techniques, researchers can capture ultrastructural details—subcellular structures visible at nanometer resolution—such as synaptic vesicles, mitochondrial cristae, and myelinated axons [7,8], thereby supporting both the mapping of complete neural circuits (*connectomics*) and the characterization of subcellular variations in neuro-degenerative disorders [9, 10].

Despite EM's imaging power, the extraction of biological meaning remains hampered by the segmentation process—labeling each pixel or voxel according to cellular structures such as neurons, mitochondria, synapses, or nuclei. Manual and semi-automated approaches are prohibitively labor-intensive: reconstructing just one cubic millimeter of mouse cortex can require annotating more than 1000 terabytes of image data [11,12], making manual workflows unsustainable for modern-scale projects. Even widely-used interactive tools like *ilastik* or Fiji's Trainable Weka Segmentation struggle to scale efficiently and are sensitive to heterogeneities in staining quality, anisotropic resolution, and complex tissue architecture [13,14]. Algorithmic segmentation methods—including watershed and region-growing techniques—are computationally demanding and error-prone, especially in noisy or ambiguous boundaries [15,16]. These computational constraints are intensified by inter-annotator variability, which can exceed 20% in complex regions like dendritic spines and postsynaptic densities [17, 18], undermining the reproducibility of quantitative analyses.

Deep learning (DL) offers a transformative alternative by automating both feature extraction and multiscale contextual understanding, which are critical for accurate EM segmentation. Convolutional Neural Networks (CNNs), such as U-Net and Mask R-CNN, are currently used for segmenting complex biological structures in EM datasets [2,19], while hybrid, transformer-based, and foundation models are rapidly improving state-of-the-art metrics across multiple benchmarks [18,20, 21] (as shown in Fig. 1). DL scales where manual methods cannot: distributed training pipelines can process large volumes efficiently [9], while lightweight architectures achieve real-time inference capabilities [22].

Methodological innovations—including topology-preserving loss functions like centerline-Dice (cDice), which measures structural connectivity preservation [9], contrastive learning modules [23,24], and joint semantic-instance training paradigms [25]—further refine segmentation accuracy and generalization.

Despite these advances, brain-specific EM segmentation faces unique challenges that necessitate specialized approaches. Neural tissues exhibit intricate ultrastructures including synaptic clefts [9],

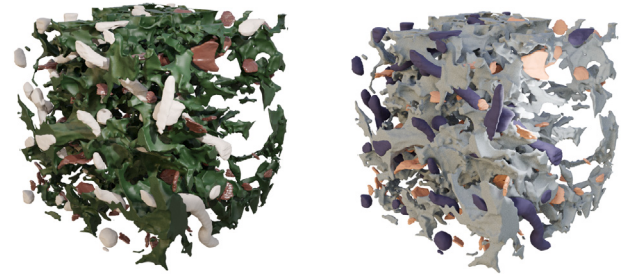


Fig. 2. Mouse rendering: Full 3D reconstruction of the data in the stack rendered using Blender software. Left: glia (green), mitochondria (white), synapses (pink). Right: glia (light gray), mitochondria (purple), and post-synaptic densities (orange).

myelinated axons [10], and complex layered mitochondria [26], which require sub-nanometer precision and topology-aware models to accurately capture delicate morphological relationships. Fig. 2 illustrates this complexity by showcasing a 3D reconstruction of mouse neural tissue with distinctly labeled glial processes (green/light gray), mitochondria (white/purple), and synaptic elements (pink/orange). Scalability remains critical—whole-brain segmentation requires substantial computational resources [7,9]—while reproducibility challenges persist, with only about half of published studies providing accessible code and pretrained models [2,25,27] versus others lacking implementation disclosure [1].

1.1. Comparison with existing reviews and objectives

Previous reviews have examined deep learning in microscopy or EM image analysis, but none have specifically synthesized DL-based segmentation for brain EM datasets (see Table 1 for detailed comparison). A recent survey on general EM segmentation [28] provides a broad overview of DL advancements in semantic and instance segmentation (2017–2022), but covers various cell types without focusing on brain-specific challenges. Our review exclusively addresses brain EM segmentation, emphasizing regions like cortex, hippocampus, and peripheral nervous system. A broader microscopy review [29] surveys DL applications across microscopy tasks including classification and tracking, highlighting CNN architectures and dataset challenges, but remains general and overlooks segmentation challenges specific to neural ultrastructure such as membrane continuity and synaptic connections. A third survey [30] focuses on nuclei and cell detection in light microscopy images from cancer and Alzheimer's tissues, offering insights into stereological analysis but remaining confined to stained light microscopy data without addressing high-resolution, densely packed brain EM environments necessary for neuroanatomical mapping. A fourth review [31] explores broader machine learning applications in EM across physical and life sciences, including denoising and inpainting, while identifying common neural network architectures and EM-specific adaptations, but lacks systematic analysis of

Table 1

Comparison of this survey with five existing reviews. A checkmark (✓) indicates explicit coverage; a dash (–) means the aspect is outside the stated scope.

Dimension	Our	Aswath et al. [28]	Liu et al. [29]	Alahmari et al. [30]	Treder et al. [31]	Huo et al. [32]
Primary modality = EM	✓	✓	–	–	✓	–
Brain tissue only	✓	–	–	–	–	–
Segmentation focus	✓	✓	partial ^a	✓ ^b	–	partial ^c
Spatial scales	✓	✓	cell–tissue	nuclei	partial	partial
Architectures	✓	CNN	CNN	CNN	mixed	CNN
Brain-specific metrics (CREMI, cDice)	✓	–	–	–	–	–
Scalability (teravoxel, multi-GPU)	✓	partial	–	–	partial	–
Reproducibility/code audit	✓	–	–	–	–	–
Meta-Analysis	✓	–	–	–	–	–
Years covered	19–25	16–22	17–21	16–19	18–22	20–24
Unique contribution	brain EM seg.	generic EM	broad micro.	nuclei seg.	ML for EM	cell img.

^a Segmentation addressed among multiple tasks.

^b Light-microscopy nuclei, not EM.

^c Focuses on enhancement/recognition; segmentation only tangential.

segmentation models and anatomical specificity critical for brain EM tasks. Finally, a recent survey [32] primarily addresses low-level image processing tasks like denoising and single-cell behavior modeling, without considering neural tissue segmentation, extreme data scales, or fine-grained annotation requirements essential for brain ultrastructure reconstruction.

As detailed in Table 1, our review fills this critical gap by focusing exclusively on DL-based segmentation methods applied to brain-specific EM datasets across model organisms, including mouse, human, *Drosophila*, zebrafish, and *C. elegans*. These datasets span multiple spatial scales—from subcellular organelles to whole neurons and neural circuits—including neuronal morphology reconstruction, organelle segmentation, and fine-scale subcellular analysis. Our objectives are fourfold: (1) to **synthesize and compare key DL methodologies**, including CNN variants, transformer-based models, and foundation models, along with specialized loss functions and training paradigms; (2) to **evaluate segmentation performance** through a detailed meta-analysis of 27 studies (46 experiments) across 10 datasets, providing unified benchmark tables, normalized comparisons, and statistical analysis of method categories and cross-dataset trends; (3) to **identify existing challenges**, including limited reproducibility and task-specific weaknesses; and (4) to **propose future directions**, advocating for topology-aware approaches and open-source frameworks.

This review is guided by key questions designed to identify trends, benchmark performance, and expose limitations across seven major dimensions. We thoroughly examine which brain EM datasets dominate the literature and how data processing pipelines affect segmentation quality, analyze the evolution and comparative performance of deep learning architectures from CNNs to transformers, categorize segmentation tasks by complexity and structural targets, evaluate performance metrics and scalability considerations, address practical implementation challenges including computational requirements and reproducibility, and conduct an extensive meta-analysis of segmentation methods across datasets and metrics. To ensure rigor, we include only DL-based segmentation studies of brain EM, excluding non-neural or non-deep learning approaches, spanning semantic segmentation, instance segmentation, and panoptic modeling across conventional CNNs, vision transformers, and hybrid models.

In contrast to broader or differently focused reviews, our work offers several distinct contributions structured across our seven findings sections. Through a focused investigation of **Datasets in Brain EM Segmentation**, we identify dominant datasets and their characteristics; **Data Processing Pipeline** examination reveals critical preprocessing steps affecting performance; **Deep Learning Architectures** synthesis compares architectural evolution from U-Net variants to transformer-based models; **Segmentation Tasks** categorization addresses task-specific challenges and solutions; **Evaluation and Performance** assessment provides robust benchmarking; **Practical Considerations** analysis addresses real-world implementation challenges; and our novel

Meta-Analysis of Segmentation Methods provides the first statistical synthesis of cross-dataset performance trends through unified benchmark tables and normalized comparisons—directly addressing the concerns about rational performance presentation. Through this seven-dimensional analysis, our review serves as both a definitive reference and practical roadmap for future developments in neural tissue analysis, uniquely combining technical depth with statistical rigor in the brain EM segmentation domain.

2. Methods

This review was conducted following established guidelines for systematic reviews of computational studies, with particular attention to the unique considerations required for deep learning and medical imaging research. Our methodology encompassed extensive literature searching, rigorous study selection, data extraction, quality assessment, and structured synthesis approaches designed to capture the multidimensional nature of brain EM segmentation research while maintaining scientific rigor and reproducibility.

2.1. Search strategy

A broad literature search was conducted across five major databases: PubMed, IEEE Xplore, Web of Science (WoS), Google Scholar, and arXiv. The search aimed to identify studies that applied deep learning (DL) methods to segmentation tasks in brain electron microscopy (EM) data. Our search strategy incorporated a detailed set of terms organized into three conceptual categories: (1) deep learning techniques, (2) connectomics and brain mapping applications, and (3) volumetric and microscopy data characteristics.

The following search string was consistently applied across all databases, with appropriate syntax modifications for each platform:

```
(((((('deep learning' OR 'artificial intelligence' OR 'neural network*' OR 'machine learning' OR 'AI' OR 'deep neural network*' OR 'transformer*' OR 'convolutional neural network*')) AND (('connectom*' OR 'brain mapping' OR 'neural circuit*' OR 'brain reconstruction' OR 'neuroanatomy' OR 'brain imaging' OR 'neural reconstruction')))) AND (('3D' OR 'three-dimensional' OR 'multiscale' OR 'multi-scale' [ OR 'volume*' OR 'large-scale' OR 'microscopy'])))
```

The search encompassed publications from January 2019 to April 2025, including both peer-reviewed journal articles and conference proceedings as well as preprints. The initial search yielded 2,330 records distributed across databases as follows: PubMed (n = 621), Web of Science (n = 859), IEEE Xplore (n = 496), arXiv (n = 154), and Google Scholar (n = 200). After removing 684 duplicates through automated and manual screening processes, 1546 unique records remained for further evaluation.

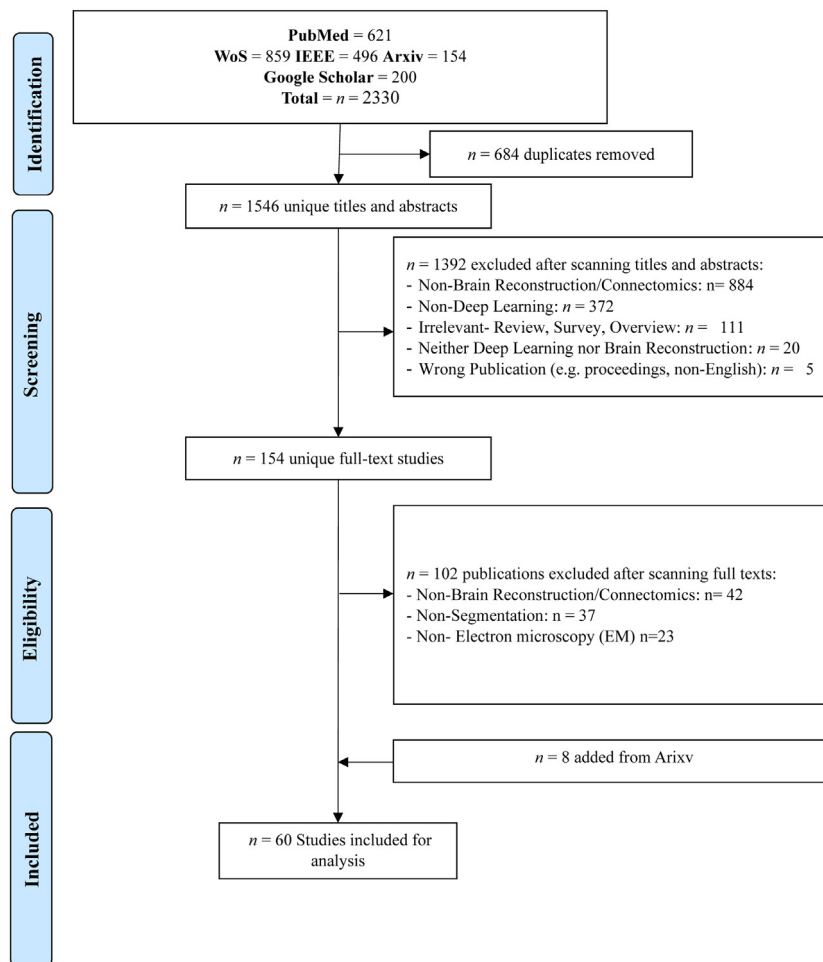


Fig. 3. PRISMA flowchart illustrating study selection.

2.2. Study selection

The study selection process was implemented in two sequential phases: initial title/abstract screening followed by full-text review. Predetermined inclusion and exclusion criteria were established to ensure consistent evaluation across all authors and maintain focus on the research objectives.

During the title and abstract screening phase, 1392 studies were excluded based on predetermined criteria. The majority of exclusions ($n = 884$) comprised studies not focused on brain connectomics or neural reconstruction. Additional exclusions included research not employing deep learning methodologies ($n = 372$), reviews or surveys without primary research content ($n = 111$), publications addressing neither deep learning nor brain reconstruction/connectomics ($n = 20$), and inappropriate publication types or non-English language publications ($n = 5$).

Subsequent full-text screening was conducted on the remaining 154 articles, resulting in 102 further exclusions. These comprised studies focusing on non-brain connectomics applications ($n = 42$), research primarily addressing tasks other than segmentation ($n = 37$), and studies using imaging modalities other than EM or light microscopy ($n = 23$).

During the review process, an additional 8 relevant studies were identified through supplementary searches. Consequently, a final corpus of **60 studies** was included in the review. The PRISMA flow diagram in Fig. 3 provides a visual representation of the complete selection process.

2.3. Data extraction

A standardized data extraction form was developed to consistently capture relevant information from each included study. The extraction framework was designed to capture both technical methodological details and practical implementation considerations essential for a thorough analysis of deep learning approaches in brain EM segmentation.

The extracted data were organized into four primary categories. Study details included authors, journal or conference name, and publication year. Dataset characteristics encompassed dataset identification (e.g., CREMI, MitoEM), acquisition modality (FIB-SEM, ssTEM), brain region examined (e.g., cortex, hippocampus), voxel resolution, volume dimensions, and annotated structure types (e.g., synapses, mitochondria, axons).

Methodological information comprised segmentation approach (semantic or instance), model architecture (e.g., U-Net, feature pyramid network), training paradigm (supervised, semi-supervised, non-supervised), preprocessing and postprocessing techniques, data augmentation strategies, loss functions, evaluation metrics (e.g., Dice, Variation of Information (VOI), ARAND, cDice, ERL), and validation methodology. Performance and reproducibility data included quantitative results, computational scalability, and code availability assessment.

2.4. Quality assessment

Methodological quality and risk of bias were evaluated for each included study using a customized assessment checklist specifically designed for computational imaging research. The assessment framework

addressed the unique challenges of evaluating deep learning studies where traditional clinical trial quality measures may not directly apply.

The assessment criteria examined three core dimensions: (1) **experimental rigor**, including adequacy of train/test splits, validation methodology, and statistical analysis; (2) **reporting completeness**, encompassing documentation of hyperparameters, training procedures, and implementation details; and (3) **reproducibility factors**, evaluating code availability, data accessibility, and sufficient methodological detail for replication.

Each study received independent assessment by two authors using a standardized scoring rubric, with discrepancies resolved through consensus discussion. Quality scores informed confidence levels for individual study findings and weighted meta-analysis contributions where appropriate.

2.5. Data synthesis and meta-analysis

The synthesis strategy combined qualitative narrative analysis with quantitative meta-analysis to maximize information extraction while accommodating methodological heterogeneity across studies. The approach was designed to identify field-wide patterns while enabling statistical comparison where data permitted.

Narrative Synthesis: Qualitative synthesis examined architectural trends, dataset utilization patterns, evaluation metric evolution, and computational considerations across the research landscape. Studies were categorized methodically by method type, target structures, and experimental approaches to identify convergent practices and emerging directions. **Quantitative Meta-Analysis:** For studies reporting comparable performance metrics on shared datasets, quantitative synthesis was conducted to establish unified benchmark comparisons. Inclusion criteria required studies to report: (i) clearly documented test sample sizes, (ii) evaluation on datasets used by ≥ 2 studies, and (iii) at least one standardized metric (Dice coefficient, Jaccard index, Average Precision, F1-score, Precision, Recall, VOI, or ARAND).

Of the 60 studies encompassing 28 datasets, 27 studies (46 experiments) across 10 datasets met stringent selection criteria for quantitative analysis. The synthesis employed descriptive statistical analysis, one-way ANOVA testing for between-group comparisons, effect size analysis using Cohen's d , and correlation analysis for metric relationships. Performance comparisons were standardized across method categories (CNN-based, Transformer-based, Hybrid, Foundation models), datasets (CREMI, MitoEM, SNEMI3D, Lucchi, etc.), and segmentation targets (neurons, mitochondria, synapses).

A composite scoring system was developed to rank methods across multiple metrics, enabling unified benchmark table generation for direct cross-dataset performance comparison. Statistical significance testing was conducted using one-way ANOVA ($\alpha = 0.05$), with effect sizes calculated to quantify practical significance of performance differences between method categories.

3. Findings

This in-depth review of brain EM segmentation presents our findings across seven critical dimensions of the field. By examining 60 studies spanning 28 datasets, we characterize the current landscape from data foundations through methodological innovations to practical deployment considerations. Our analysis reveals both convergent best practices and emerging technological frontiers that collectively define the state-of-the-art in neural ultrastructure analysis.

3.1. Datasets in brain EM segmentation

The foundation of effective brain EM segmentation lies in the quality, diversity, and accessibility of training and evaluation datasets. Our analysis of 28 distinct datasets reveals how dataset characteristics—from anatomical targets to acquisition modalities—fundamentally shape methodological development and performance benchmarking across the field.

3.1.1. Overview

This review identified 28 distinct datasets across the selected studies designed for brain EM segmentation, each contributing uniquely to the field through their anatomical focus, imaging modality, and annotated structures (Supplementary A Table 1). The mouse somatosensory cortex emerged as the most frequently studied brain region, represented in high-resolution datasets including SNEMI3D [19,33–37], Kasthuri++ [22,38–41], and Mouse ATUM-SEM [17,19,42]. These datasets have proven instrumental in advancing neuronal circuit mapping at ultrastructural resolution. Beyond the somatosensory cortex, researchers have developed specialized collections targeting other critical brain regions, including the hippocampus presented by EPFL Hippocampus/Lucchi++ [15,21,25,43–45] and DendriteSAM [20,35], cerebellum evident in Cerebellar EGL ssSEM [46], and visual cortex notably in AxonEM [10,47] and h01/MICrONS [48,49]. The anatomical diversity extends through cross-species collections such as EM-Neuron [49], which incorporates samples from multiple species, and BigNeuron [2,13,50–55], which spans mouse, human, and zebrafish nervous systems.

The subcellular targets varied significantly across datasets but consistently focused on key neuroanatomical components essential for understanding neural connectivity and function. Neuronal membrane and whole-cell segmentation formed the central focus in datasets like CREMI [9,23,25,38,40,56–59], SNEMI3D [19,33–37], and FIB-25 [16,26,58], enabling complete reconstruction of neuronal morphologies and connectomes. Synaptic structures were specifically annotated in several datasets including ISBI 2012 [14,42,60], Kasthuri++ [22,38–41], and WASPSYN [61], with attention to both synaptic clefts and their pre- and post-synaptic partners. Mitochondrial annotations featured prominently in specialized datasets such as MitoEM [15,25,37,38,43–45] and EPFL Hippocampus [15,21,25,43–45], reflecting the critical importance of these organelles in neural energetics. Nuclei segmentation was emphasized in datasets like NucMM [25,38,43,62] and Zebrafish EM dataset [60], providing essential context for cellular organization.

3.1.2. Data acquisition and properties

Fig. 4 illustrates acquisition methodologies across datasets. 3D imaging dominates, with ssEM/ssTEM most prevalent across 26 studies [9,19,21,23,25,57,60], reflecting its reliability for neural tissue reconstruction; (see Supplementary Material A, Table 2). FIB-SEM appears in 11 studies [25,43,56,58,61], offering superior resolution consistency. ATUM-SEM was used in 6 studies [11,17,33,42] and SBEM/SBF-SEM in 5 studies [11,20,58,63], demonstrating automated sectioning approaches.

In 2D acquisition, Light/Optical microscopy appeared in 8 studies [2,13,50,51,53], matching Multi-beam SEM with 8 studies [15,38,40,43,44]. Other SEM variants were used in 9 studies [1,34,38,39,57], general EM in 4 studies [12,26,64], and TEM in 2 studies [27,48]. Specialized techniques included fMOST in 4 studies [2,18,52], Light-sheet Microscopy in 2 studies [7,65], Confocal Microscopy [66], and 3D X-ray Microscopy [62].

The acquisition methodology directly determines resolution characteristics and anisotropy patterns observed across datasets (Supplementary A Table 1). Serial section transmission EM (ssTEM) datasets typically exhibit moderate anisotropy due to sectioning limitations: CREMI provides $4 \times 4 \times 40$ nm resolution, ISBI 2012 offers $4 \times 4 \times 50$ nm, while AC3/AC4 datasets present $12 \times 12 \times 29$ nm voxels. This anisotropy stems from fundamental technical constraints—xy-plane resolution is determined by electron beam focus and detector capabilities (readily achieving sub-10 nm precision), while z-axis resolution depends on physical sectioning thickness, typically constrained to 30–50 nm to avoid section folding, charging artifacts, and signal-to-noise degradation.

FIB-SEM demonstrates superior isotropy capabilities, with EPFL Hippocampus datasets maintaining $5 \times 5 \times 5$ nm resolution and FIB-25 providing $8 \times 8 \times 8$ nm voxels. The controlled ion beam milling

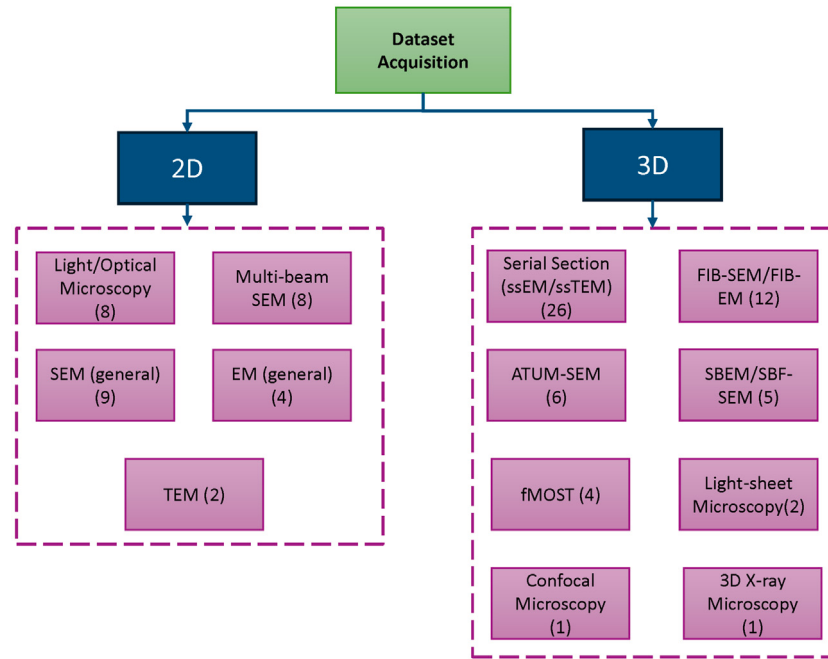


Fig. 4. Overview of Dataset Acquisition Modalities. The figure categorizes dataset acquisition techniques used in brain EM segmentation studies, classified by imaging dimensionality (2D vs. 3D) and modality type. Two-dimensional methods include light microscopy, SEM variants, and TEM, while 3D techniques are dominated by serial section EM (ssEM/ssTEM), FIB-SEM, and ATUM-SEM. The number in parentheses indicates the number of studies utilizing each method, as detailed in (Supplementary A Table 1).

enables finer z-resolution but requires balanced consideration of acquisition time and potential sample damage. Multi-beam and serial block-face SEM show intermediate characteristics: MitoEM datasets offer highly anisotropic $30 \times 8 \times 8$ nm voxels optimized for large-volume acquisition, while SNEMI3D provides moderate $6 \times 6 \times 30$ nm resolution.

Volume sizes demonstrate the scale diversity driving methodological development, ranging from small subvolumes $512 \times 512 \times 30$ voxels in ISBI 2012 [14,42,60] to massive reconstructions EMNeuron's 22 billion voxels [49] or whole-brain imaging FlyWire/FAFB with 213 tera-voxels [9,24]. Contemporary datasets increasingly favor near-isotropic configurations—newer acquisitions like Lucchi $5 \times 5 \times 5$ nm, FIB-25 $8 \times 8 \times 8$ nm, and WASPYN $8 \times 8 \times 8$ nm demonstrate the field's progression toward isotropic imaging. However, large-scale connectomics projects often accept moderate anisotropy to balance resolution with acquisition throughput and coverage requirements.

These resolution characteristics fundamentally influence segmentation methodology across our analyzed studies. Methods addressing highly anisotropic datasets (e.g., MitoEM's $30 \times 8 \times 8$ nm) employ specialized architectures with anisotropic convolution kernels, resolution-adaptive loss functions, or multi-scale processing pipelines that explicitly model differential information content across spatial dimensions. Conversely, near-isotropic datasets enable more straightforward 3D convolution approaches and simplified topology preservation strategies.

3.1.3. Labeling and availability

Labeling strategies across these datasets aligned closely with the segmentation task types they were designed to address. Semantic segmentation datasets primarily annotated class boundaries such as neuronal membranes—evident in ISBI 2012's membrane annotations [14,42,60]—or broad cellular compartments. Instance segmentation datasets targeted the delineation of individual structures—complete neurons in 6 SNEMI3D studies [19,33–35] and 9 CREMI studies [9,23,25,56,57], mitochondria in 7 MitoEM studies [15,25,38,43,44] and 6 EPFL Hippocampus studies [15,25,43,44], axons in AxonEM Wei et al.

[10], Liu et al. [47], dendrites in DendriteSAM [20,35], or synapses in WASPSYN [61]. This distinction in annotation philosophy reflects different downstream applications. More recently, panoptic tasks have emerged, integrating both semantic and instance components, as presented by Dorkenwald et al. [48], Zhang et al. [49], on which simultaneously segmented both neurons and glial cells; see Supplementary Material A, Table 1 for complete dataset details.

Dataset availability varied considerably across the field; see Supplementary Material A, Table 1 for comprehensive details. Many key benchmarks were openly accessible via GitHub—e.g., ISBI 2012 [14,42,60], 8 BigNeuron studies [2,13,50,51,53], DendriteSAM [20,35]—specialized challenge platforms such as Codalab WASPSYN [61], or institutional repositories such as Zenodo Cerebellar EGL ssSEM [46]. CREMI with 9 studies [9,23,25,56,57], SNEMI3D with 6 studies [19,33–35], and MitoEM with 7 studies [15,25,38,43,44] exemplify this open-science approach with well-documented public distributions. Some datasets provided only partial annotations rather than exhaustive ground truth, reflecting the immense human effort required for complete manual annotation of large EM volumes.

3.1.4. Methodological impact

The impact of these datasets extended far beyond simple benchmarking, fundamentally shaping methodological innovation in the field. CREMI with 9 studies [9,23,25,56,57] established a precedent for standardized evaluation of 3D neuron segmentation, creating a common framework that facilitated innovations in graph-based agglomeration techniques and hybrid CNN-Transformer architectures; see Supplementary Material A, Table 1 for complete study list. The dataset's carefully controlled evaluation metrics drove consistent progress in addressing challenging boundary detection problems in densely packed neuropil.

SNEMI3D with 6 studies [19,33–37] similarly helped formalize synaptic segmentation challenges and was instrumental in developing focal losses and instance-aware networks. The dataset's emphasis on anisotropic resolution ($6 \times 6 \times 30$ nm) spurred innovations specifically addressing z-direction discontinuities. Its standardized format

for synaptic cleft segmentation provided a consistent benchmark for precision in synapse detection algorithms.

FIB-25 [16,26,58], with its isotropic voxel sampling ($8 \times 8 \times 8$ nm), enabled fine-scale modeling of *Drosophila* medulla spanning $52 \times 53 \times 65 \mu\text{m}^3$ and catalyzed advances in adversarial boundary detection strategies. The dataset's synaptic-resolution data proved critical for improving detection accuracy in small neuronal structures typical of invertebrate systems.

MitoEM [15,25,37,38,43–45] offered one of the most extensive volumes specifically targeting mitochondria segmentation ($1000 \times 4096 \times 4096$ voxels), pushing the field toward sophisticated U-Net variants with attention mechanisms. The dataset's cross-species approach, encompassing both rat and human samples, encouraged the development of architectures robust to biological variability.

EMNeuron [49], with its massive 22-billion-voxel scale and multi-modal, multi-species architecture, supported multi-resolution frameworks suitable for cross-species generalization, establishing new paradigms for handling extreme-scale volumes.

The technological evolution evident in (Supplementary A Tables 1 and 2) demonstrates a clear progression from traditional EM approaches toward more sophisticated volume imaging techniques. This shift parallels algorithmic developments, with newer acquisition methods enabling and demanding more sophisticated computational approaches. Automated sectioning technologies like ATUM-SEM featured in 4 studies: [11,17,19,42] have proven particularly valuable for large-scale 3D reconstructions. The emerging prominence of light microscopy techniques 8 studies: [2,13,50–53,55,66] points toward future integration of live-tissue imaging with ultrastructural analysis. Specialized techniques such as multi-beam SEM 3 studies: [15,38,40] and FIB-SEM 6 studies: [16,25,45,56,58,61] further illustrate the field's methodological diversification.

Collectively, these datasets highlight critical challenges in the field: the scale of data demands increasingly sophisticated computational infrastructure; high-resolution imaging substantially increases annotation complexity; and dense annotations of specific structures require significant expert effort. Despite these challenges, the continuous development of specialized datasets has accelerated methodological innovation, enabling increasingly accurate and detailed neural circuit mapping at ultrastructural scales.

3.2. Data processing pipeline

Effective transformation of raw EM data into analysis-ready formats requires sophisticated preprocessing, data handling, and postprocessing strategies. This section examines how studies address the unique challenges of brain EM data, from intensity normalization and resolution management to specialized postprocessing techniques that preserve neuronal topology and connectivity. (see Supplementary A Table 3: Data Processing Methods)

3.2.1. Preprocessing

Effective preprocessing constitutes a critical step in enhancing the quality and consistency of EM data prior to deep learning analysis. Brain-specific preprocessing methodologies demonstrate significant variation across studies, reflecting the diverse challenges inherent to different imaging modalities and segmentation targets.

Normalization strategies represent a fundamental preprocessing component implemented in numerous studies (Supplementary A Table 3). Xin et al. [56] implemented mean/standard deviation normalization, while Heinrich et al. [9] and Shamsi et al. [66] applied value normalization to standardize intensity distributions. Klinghoffer et al. [53] utilized intensity normalization for axon tracing in SHIELD PVGPe datasets, and Huang et al. [52] performed intensity normalization before geometric transformations. More specialized approaches included, Knowles-Barley et al. [39] implemented Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance local contrast.

Resolution adjustments were consistently applied to manage volume size and computational requirements. Multiple studies incorporated downsampling techniques. For example, Lin et al. [62] implemented resolution downsampling for 3D neuronal nuclei segmentation, while Guo et al. [17] employed patch division combined with downsampling for synapse reconstruction. Liu et al. [47] utilized $2\times$ in-plane downsampling to process large volumes, and Turner et al. [41] specifically applied $12\times 12\times 30$ nm downsampling for membrane segmentation in Kasthuri++. For anisotropic correction, trilinear interpolation was used by Schmidt et al. [11], whereas the ROTO protocol was implemented by Cordero Cervantes et al. [46] to normalize resolution across dimensions.

Structural enhancement techniques were customized to specific anatomical targets. Gornet et al. [7] implemented topology inflation for neuronal arbor preservation, Januszewski et al. [63] utilized distance transforms, and Shi et al. [38] applied Line Segment Detector (LSD) generation to enhance linear structures. Wang et al. [50] implemented scale-space transformation for neural structure preservation in BigNeuron datasets.

Several studies employed specialized preprocessing for data preparation: Troidl et al. [24] utilized point cloud sampling for point-based neuron segmentation, Gornet et al. [7] applied topology inflation designed to preserve neuronal connectivity, and Potocek et al. [1] employed sparse reconstruction techniques to process large datasets efficiently.

Contrast enhancement was implemented through various methods, with Plebani et al. [27] applying histogram equalization for unmyelinated axon segmentation. Both Urakubo et al. [33] and Knowles-Barley et al. [39] utilized CLAHE for multi-structure segmentation, while Wei et al. [18] employed maximum intensity projection alongside block division for 3D soma detection.

It is noteworthy that several studies [15,16,19,23,43] reported no specific preprocessing steps, suggesting either reliance on high-quality original data or the integration of normalization directly within their neural network architectures.

3.2.2. Data handling techniques

Dataset distribution. Dataset partitioning strategies varied considerably across studies, reflecting the diverse nature of EM datasets and segmentation tasks. Standard train/validation/test splits were common but implemented with different ratios: Shamsi et al. [66] employed a 75/25 split, Cheng et al. [34], Xu et al. [8] used 50/25/25 divisions, while Wei et al. [65] implemented a 6:1:1 ratio. Volume-based partitioning proved particularly suitable for 3D EM data, with Potocek et al. [1] utilizing 50 locations for training with 46 validation/test locations, and Liu et al. [19] employing 80–100 training slices with 98–256 test slices.

Cross-dataset validation emerged as a robust approach for testing generalization capabilities. Luo et al. [15] applied standard splits across multiple datasets including MitoEM and EMNeuron, while Sun et al. [23] utilized both CREMI and AC4 standard splits. Zhuo et al. [20] trained on 100 slices and tested across different datasets, and Cheng et al. [37] followed standard splits across SNEMI3D and MitoEM. Zhang et al. [49] demonstrated extensive cross-dataset validation by training on 13 datasets and testing on 3 others.

Several studies employed specialized partitioning strategies: Schmidt et al. [11] used mouse data for training and human data for testing, demonstrating cross-species generalization. Januszewski et al. [63] trained on 29 small volumes plus 4 medium volumes to evaluate scalability. Lin et al. [62] implemented a 5%/5%/90% split to emphasize testing performance. Matejek et al. [36] utilized 4 training, 3 validation, and 4 testing volumes across multiple datasets, while Liu et al. [47] divided volumes of approximately $3300 \times 2800 \times 64$ voxels for training and $1900^2 \times 512$ voxels for testing.

Studies employed two distinct approaches to increase training data diversity: *data expansion* involves pre-generating and permanently storing transformed versions of original samples to enlarge the training

dataset, while *data augmentation* applies transformations dynamically during training without storing additional samples. This distinction is crucial for understanding computational requirements and storage needs across different methodological approaches. Some studies employed data expansion techniques to increase dataset size for training. Xiao et al. [42] pre-generated transformed versions using elastic distortion, rotation, and flipping to expand training sets to 9000, 7000, and 8000 samples respectively across different datasets, while Xin et al. [56] generated 1 million patch pairs for feature registration training. Cordero Cervantes et al. [46] implemented an 81/19 train/test split after applying the ROTO protocol, while Troidl et al. [24] utilized 15,000 training neurons and 2,000 testing neurons after point cloud sampling. These expansion strategies aimed to provide sufficient training data for complex deep learning models.

Data augmentation. To enhance generalizability and mitigate overfitting, data augmentation was widely implemented across studies; see Supplementary Material A, Table 3. Geometric transformations formed the core of most augmentation strategies: rotation was employed in 7 studies [19,50,51,56,64], while flipping along various axes was similarly common across 9 studies [19,27,46,50,61,67]. More complex geometric transformations were also reported, with Schmidt et al. [11] introducing flight axis rotation, Troidl et al. [24] applying multiple geometric transformations tailored for point clouds, and Xin et al. [56] utilizing affine transformations.

Intensity-based augmentations complemented geometric approaches, with many studies implementing contrast adjustments, brightness variations, and noise addition. Liu et al. [19] applied intensity scaling alongside geometric transformations, while several studies, including Li et al. [61], Xiao et al. [2], Shamsi et al. [66], and Wei et al. [18], employed various intensity transformations. Additionally, Plebani et al. [27] implemented jittering alongside flipping, and Lee et al. [67] combined multiple transformations for neuron instance segmentation.

Specialized augmentation techniques addressed specific segmentation challenges: Gornet et al. [7] implemented simulated artifacts to enhance robustness, Chen et al. [13] generated synthetic defects to improve generalization, and Cheng et al. [34] applied Gaussian blur as a form of augmentation for axon centerline detection. Cordero Cervantes et al. [46] employed multiple transformations specifically tailored for granule cell segmentation, while Guo et al. [17] combined various transformations with patch-based approaches for synapse reconstruction.

Diverse augmentation strategies were employed across several studies. Dorkenwald et al. [48] applied random transformations for neuronal membrane segmentation, Zhang et al. [49] implemented multiple transformations for dense neuron and glia segmentation, and Lin et al. [25] employed various transformations for multi-class organelle segmentation.

A substantial number of studies—18 in total—reported no data augmentation [1,15,38,43,55,63], suggesting either reliance on sufficient training data or architectural innovations that reduced dependency on data augmentation; see Supplementary Material A, Table 3 for complete study list.

3.2.3. Postprocessing

Postprocessing methods were critical for translating raw segmentation outputs into biologically meaningful structures, with approaches tailored to specific segmentation tasks and neural structures (Supplementary A Table 3).

Watershed-based algorithms dominated the postprocessing landscape, appearing in various specialized forms across studies as presented in Supplementary Material A, Table 3. Eleven studies employed watershed techniques to separate touching objects and refine instance boundaries [12,15,25,33,37,58,62]. Specialized variants were also developed, with Lee et al. [67] introducing Mutex Watershed for dense

voxel embeddings, Santurkar et al. [12] utilizing watershed with agglomeration, and Cheng et al. [37] combining watershed variants with DBSCAN clustering.

Connected component analysis and clustering approaches were frequently implemented to identify distinct structures. For instance, Luo et al. [15] combined connected components with watershed, while Troidl et al. [24] and Guo et al. [17] employed hierarchical clustering. Wei et al. [18] integrated connected components with DBSCAN clustering, and Chen et al. [40] utilized 3C clustering for neural structure segmentation. Additionally, Plebani et al. [27] applied size filtering alongside dilation techniques.

Skeletonization methods played a crucial role in neuron morphology reconstruction. Wang et al. [50] implemented a backtracking algorithm, while Cordero Cervantes et al. [46] applied skeleton extraction for granule cell reconstruction. Zhao et al. [51] used skeleton aggregation for 3D neuron segmentation, and both Shamsi et al. [66] and Xu et al. [8] employed 3D skeletonization techniques. Specialized approaches included the ST-LVF tracing method introduced by Huang et al. [52] and the SmartTracing algorithm developed by Cheng et al. [55].

Threshold-based refinement appeared in various forms. For example, Liu et al. [57] and Heinrich et al. [9] applied simple thresholding, while Berman et al. [59] implemented threshold-based refinement for point-cloud neuron segmentation. Graph-based refinement was utilized in several studies, with Chen et al. [13] employing BFS graph reconstruction, Dorkenwald et al. [48] using embedding aggregation, and Zhang et al. [49] applying graph-based segmentation.

Specialized postprocessing techniques addressed unique challenges in neuronal reconstruction. For instance, Schmidt et al. [11] employed curvature integration for axon and spine neck reconstruction, and Gornet et al. [7] utilized the UltraTracer tool. Both Xiao et al. [42] and Matejek et al. [36] applied lifted multicut for global consistency, while Huang et al. [54] implemented tree pruning for point-cloud neuron reconstruction. Additionally, Liu et al. [19] utilized block-wise inference, and Zhuo et al. [20] implemented interactive refinement for multi-task segmentation.

Advanced pipelines combined multiple techniques. Li et al. [61] applied local maxima detection, Mai et al. [44] implemented sliding window inference, and Urakubo et al. [33] combined binarization with watershed. Further, Xin et al. [56] employed feature regularization, and Wu [14] utilized probability map refinement. Consensus agglomeration for tissue-aware neuron segmentation was implemented by Liu et al. [47], while Turner et al. [41] applied dilated segment matching for membrane and nucleus segmentation.

Notably, several studies [1,16,21,35,44,53,64,68] reported no post-processing, suggesting end-to-end approaches in which the model output was considered sufficient without additional refinement.

This analysis reveals both methodological convergence around certain techniques and significant innovation in addressing specialized challenges posed by different brain structures, imaging modalities, and reconstruction tasks.

3.3. Deep learning architectures

The evolution of deep learning architectures for brain EM segmentation reflects the unique computational challenges posed by neural ultrastructure analysis. From traditional CNNs to emerging transformer and foundation models, architectural choices must balance the demands of multi-scale feature extraction, volumetric context integration, and topology preservation across diverse segmentation tasks. (see Supplementary A Tables 4–6)

3.3.1. Core architectures

The analysis of brain (EM data has witnessed a consistent evolution of deep learning architectures, with a clear distinction emerging between 2D, 3D, and hybrid approaches. CNNs remain dominant, with U-Net variants widely adopted for both semantic and instance segmentation tasks.

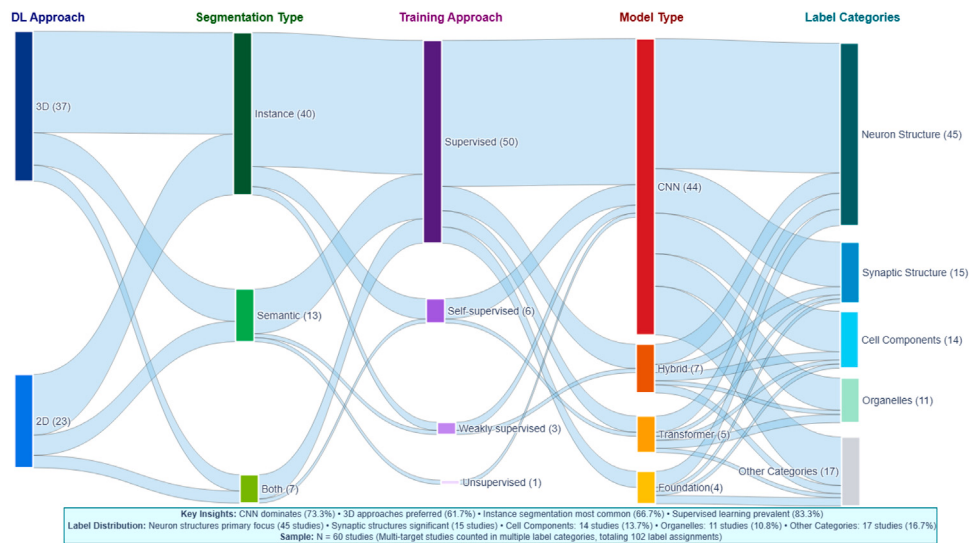


Fig. 5. Sankey diagram illustrating the distribution of 60 deep learning studies on brain EM segmentation. The flow begins with the dimensionality of the deep learning approach (2D or 3D), followed by the segmentation task type (semantic, instance, or both), the training strategy (supervised, self-supervised, weakly-supervised, or unsupervised), the architecture type (CNN, Transformer, Hybrid, or Foundation Model), and finally ends with the targeted anatomical label categories. The width of each connection reflects the number of studies using a given combination. Node labels include the number of studies at each level, showing that 3D approaches (37 studies) are more prevalent than 2D (23 studies), with CNN-based architectures (44 studies) and supervised learning (50 studies) being the most dominant choices. “Neuronal Structures” (45 studies) remains the most common segmentation target across all approaches.

2D vs. 3D approaches. 2D architectures remain foundational for brain EM analysis, particularly for datasets with anisotropic resolution or when computational resources are limited; see Supplementary Material A, Tables 4–6 for complete details. Our review revealed that CNN-based models, especially U-Net variants, dominate this category. Potocek et al. [1] implemented FusionNet with residual U-Net and multi-resolution blocks for cell membrane segmentation, emphasizing the preservation of fine details in neuronal boundaries. Similarly, Turner et al. [41] employed FusionNet with mean absolute error (MAE) loss for membrane and nucleus segmentation on Kasthuri++, demonstrating the value of fully residual networks for preserving edge continuity. Of the 60 studies examined, 23 studies utilized 2D approaches [1,12,20,27,42,46,56], highlighting their continued relevance despite advances in 3D methodologies as shown in Fig. 5; see Supplementary Material A, Tables 4–6 for complete study list.

3D architectures have gained significant traction due to their ability to leverage volumetric context crucial for neuronal continuity. Among the 60 studies analyzed, 37 studies—representing 62.7%—employed 3D approaches [9,11,15,19,23,25,43], reflecting the growing trend toward volumetric analysis in neuronal reconstruction tasks; see Supplementary Material A, Tables 4–6 for complete study list.

Among these, 3D U-Net implementations stand out as particularly effective for handling z-axis discontinuities in anisotropic data. Heinrich et al. [9] utilized 3D U-Net for synaptic cleft segmentation on CREMI, employing signed distance transform regression to precisely delineate synaptic junctions. Xu et al. [8] applied 3D U-Net with cDice and L2 losses for axon tracing in SHIELD PVGP, focusing on centerline reconstruction to maintain structural continuity.

Most notably, recurrent 3D CNNs such as Flood-Filling Networks (FFNs) have emerged as powerful tools for dense neuronal reconstruction. Januszewski et al. [63] implemented FFNs with recursive field-of-view movement and object mask channels, enabling voxelwise log-loss optimization for dense neurite skeletonization. Similarly, Dong et al. [16] utilized FFNs with FOV/POM feedback for neuronal structure segmentation in FIB-25, demonstrating the effectiveness of this approach for tracing complex neuronal processes across large volumes.

Cnn, transformers, hybrid models and foundation models. CNNs remain the dominant architecture class in brain EM segmentation, with 44 of

the 60 studies utilizing CNN implementations, as shown in Fig. 5. U-Net variants were especially common, appearing in 22 studies [7–9,25,27,61,62], highlighting their effectiveness for capturing multi-scale features necessary for neuronal structure delineation; see Supplementary Material A, Tables 4–6 for complete study list. Among CNN variants, FusionNet was used for membrane segmentation by Potocek et al. [1] and Quan et al. [60], while Flood-Filling Networks (FFNs) were employed for recursive object tracking by 5 studies [10,16,33,63].

Graph-based approaches have emerged for capturing structural connectivity. Zhao et al. [51] implemented a Dynamic Graph Convolutional Neural Network (DGCNN) for 3D neuron segmentation, while Matejek et al. [36] employed a graph-based optimization with CNN components for partitioning. Wei et al. [65] utilized a hybrid approach combining 2D Mask R-CNN with a GNN-based contextual graph for inter-slice merging, modeling local features and global neuronal topology.

Transformer-based architectures represent the frontier in brain EM segmentation for multi-modal and large-scale datasets. Five studies [15,21,24,34,35,37,44,65] implemented transformer approaches. Cheng et al. [55] presented a 3D Vision Transformer (ViT) approach for neuron segmentation in BigNeuron, utilizing deformable tubular transfer in a self-supervised manner. This approach captures long-range dependencies for maintaining neuronal continuity across volumes.

Foundation models have emerged as a promising direction for brain EM segmentation, with four studies utilizing finetuned variants of the Segment Anything Model (SAM). Three studies, including those by Shi et al. [38], Zhuo et al. [20], and Archit et al. [69], employed adaptations based on the original SAM architecture. Notably, Shi et al. [38] introduced an innovative hybrid foundation model that integrates 3D Transformer components with U-Net and Mamba blocks to effectively capture both spatial and sequential dependencies in volumetric data. The μ SAM approach, presented by Archit et al. [69], extends SAM specifically for microscopy data, demonstrating improved performance across different microscopy modalities, including EM, through specialized fine-tuning and a napari plugin for interactive segmentation. Meanwhile, Shah et al. [21] leveraged the more recent SAM2 foundation model with its advanced memory encoder and attention system, significantly improving segmentation consistency across EM slices, particularly for complex structures like astrocytic processes. These

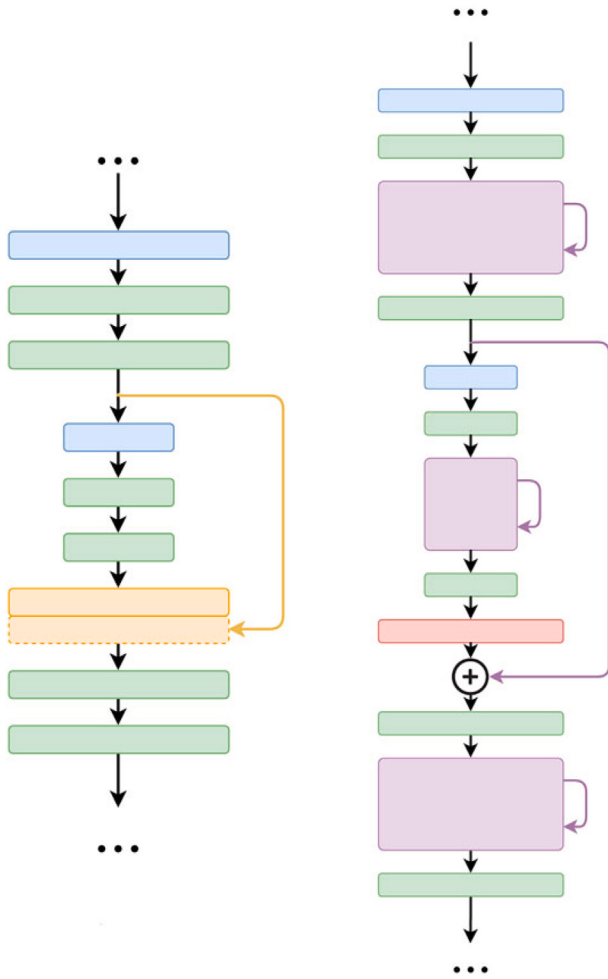


Fig. 6. Difference between the core connections of U-net Ronneberger et al. [70] (left) and FusionNet (right) Potocek et al. [1]. Note that FusionNet is a fully residual network due to the summation-based skip connections and is a much deeper network.

foundation model adaptations effectively address the data efficiency challenges in brain EM segmentation, establishing a new paradigm for applying large-scale pre-trained vision models to specialized neuroanatomical tasks.

Hybrid models combining CNNs and transformers have emerged for brain EM segmentation, with 7 studies [23,34,35,43,45,57,65] employing such approaches. Luo et al. [15] introduced FragViT (Fragment Vision Transformer), which fused fragment encoders with hierarchical aggregation for mitochondria and neuron instance segmentation across EMNeuron and MitoEM datasets. The Adaptive Template Transformer proposed by Pan et al. [43] combined CNN and transformer layers with transport-based optimization for mitochondria tracking in MitoEM, modeling structural priors while maintaining flexibility. Shi et al. [38] presented a hybrid foundation model integrating a 3D Transformer with U-Net and Mamba blocks, demonstrating the potential of combining attention mechanisms with state space models for capturing spatial and sequential dependencies in volumetric data.

These architectural trends reflect the evolution toward models capable of capturing local features and global context, with hybrid approaches bridging the gap between CNN efficiency and transformer expressivity. The emergence of foundation models suggests a shift toward transfer learning from pre-trained models, addressing the data efficiency challenges in brain EM segmentation.

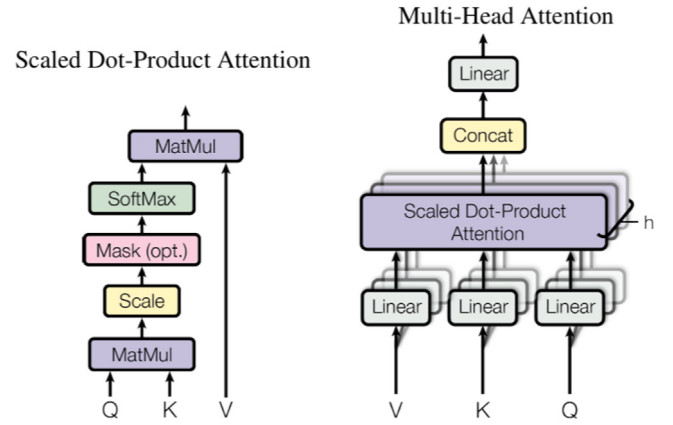


Fig. 7. Attention mechanism: (left) Scaled Dot-Product Attention. (right) Multi-Head Attention consists of several attention layers running in parallel. Source: Adapted from Vaswani et al. [71].

3.3.2. Specialized components

Skip connections and connectivity-preserving designs. Beyond core architectural choices, specialized components tailored for brain-specific challenges have proven crucial for successful EM segmentation. Our analysis identified several key adaptations designed to address the unique characteristics of brain ultrastructure (Supplementary A Tables 4–6).

Skip connections have evolved beyond their traditional U-Net implementation to specifically address neuronal continuity challenges, particularly in anisotropic datasets. Potocek et al. [1] proposed FusionNet by implemented enhanced skip connections to mitigate feature loss in cell membrane segmentation as show in Fig. 6. Gornet et al. [7] incorporated connectivity-preserving regularization into their U-Net design for neuronal segmentation, enforcing topological constraints through a specialized loss function targeting non-simple points. These adaptations enable the preservation of thin neuronal processes often lost in standard approaches, maintaining topological fidelity in complex segmentations.

Attention modules. Attention mechanisms have proven essential in focusing learning capacity on ultrastructurally relevant targets. These mechanisms, including self-attention (Fig. 7 left) and multi-head attention (Fig. 7 right), enable models to selectively focus on relevant features. Mai et al. [44] implemented a Reliable Pixel Aggregation module (RPiA) and cross-attention in their Dual-branch Net for mitochondria segmentation across multiple datasets, while Chen et al. [40] utilized SE-ResNext-50 with self-attention for neural structure segmentation in various datasets. Additionally, Guo et al. [17] implemented a region-focused Mask R-CNN with an SE-ResNet101 backbone for synapse and post-synaptic density reconstruction in Mouse ATUM-SEM, leveraging hierarchical clustering to identify synaptic regions within the broader neuronal context.

Similarly, Thawakar et al. [45] incorporated attention gates into their U-Net architecture for organelle instance segmentation, demonstrating improved performance for fine subcellular structures. The inclusion of deformable attention by Cheng et al. [55] and appearance prompt modulation by Sun et al. [23] demonstrated significant promise for generalizing across morphologically diverse datasets, providing adaptability to varying neuronal morphologies.

Loss functions and optimization. Specialized loss functions have emerged as critical components for addressing the unique challenges of brain EM segmentation (Supplementary A Tables 4 and 5). Weighted binary cross-entropy was widely adopted to address the severe class imbalance inherent in organelle-rich datasets such as MitoEM and in axon

segmentation tasks, as demonstrated by Casser et al. [22] and Plebani et al. [27].

Boundary-aware losses have proven particularly effective for the precise delineation of neuronal structures. For example, Heinrich et al. [9] implemented signed distance transform regression on CREMI, while Guo et al. [17] optimized synaptic cleft and membrane segmentation using a combined focal and CIoU loss. Lin et al. [62] employed a hybrid representation learning approach, combining binary cross-entropy with mean squared error (MSE) for 3D neuronal nuclei segmentation in NucMM-Z, enabling more precise localization of cell boundaries.

Self-supervised and contrastive losses have emerged as important innovations for reducing annotation dependency. For instance, Cheng et al. [55] utilized deformable tubular transfer, while Zhang et al. [49] employed NT-Xent and decorrelation losses for contrastive learning in dense neuron and glia segmentation on EMNeuron. Additionally, Cheng et al. [37] adapted MAE pre-training for neural structure segmentation, and Sun et al. [23] employed centroid contrastive and diversity losses for neural connectivity reconstruction.

Advanced metrics like cDice, as used by Cheng et al. [34], Wei et al. [65], and Xu et al. [8], have proven particularly effective at mitigating topological errors in neuron reconstructions by focusing optimization on preserving connectivity rather than simple pixel-wise accuracy. These specialized metrics prioritize topological preservation, which is especially critical in neuron tracing applications.

3.4. Segmentation tasks

Brain EM segmentation encompasses a spectrum of analytical tasks, each with distinct computational requirements and biological significance. This section examines how different segmentation paradigms—from semantic boundary detection to complex multi-task instance segmentation—address the diverse needs of neural circuit analysis and subcellular structure characterization. (see Supplementary A Tables 4–6).

3.4.1. Semantic segmentation

Semantic segmentation forms the foundation of brain EM analysis, enabling the delineation of key cellular structures without distinguishing individual instances (Supplementary A Table 4). Out of 60 studies analyzed in this review, 13 studies employed semantic segmentation approaches, as shown in Fig. 5. This approach has been applied to three primary targets: cell membranes, mitochondria, and neural structures.

Cell membrane delineation serves as the basis for establishing cellular boundaries in neural tissue. Potocek et al. [1] and Turner et al. [41] employed FusionNet architectures with different loss functions (L2-norm and MAE, respectively), demonstrating effective preservation of membrane continuity across datasets. Mitochondria segmentation, critical for metabolic analysis, has been addressed through transformer-enhanced networks by Mai et al. [44], who utilized cross-attention mechanisms, while Casser et al. [22] implemented a lightweight U-Net with BCE loss to handle class imbalance characteristic of organelle segmentation.

Neural structure segmentation approaches varied based on morphological complexity. For example, Jiang et al. [64] employed ResNet-50 with ASPP and focal loss for cell body delineation, while Wang et al. [50] utilized 3D CNNs with specialized boundary components for dendritic processes. The diverse architectural choices reflect the scale-dependent nature of semantic segmentation tasks in EM data, with membrane segmentation requiring fine boundary precision and neural structure segmentation demanding multi-scale context integration.

Cross-architectural performance analysis reveals that iterative refinement approaches, such as the I-CNN proposed by Wu [14], and foundation model adaptations, like the modified SAM introduced by Shah et al. [21], demonstrate particular promise for generalizing across datasets. This suggests that semantic segmentation benefits from progressive feature refinement rather than single-pass inference. The integration of specialized morphological priors, as demonstrated in the

VoxResNet approach by Huang et al. [52], which combines region growing with skeletonization, has proven especially effective for maintaining structural continuity in sparsely annotated volumes.

3.4.2. Instance segmentation

Instance segmentation in brain EM analysis requires the precise delineation of individual cellular components within densely packed neural tissue (Supplementary A Table 5). As shown in Fig. 5, this approach dominates the methodological landscape, with 40 out of the 60 studies analyzed in this review employing instance segmentation techniques.

For synapse instance segmentation on SNEMI3D, Liu et al. [19] implemented a 3D network based on ResNet50 with Feature Pyramid Network (FPN), incorporating a region proposal network and instance segmentation branch. For neuron reconstruction in AC3/AC4, Lee et al. [67] employed Residual Symmetric U-Net with embedding loss and Mutex Watershed to separate neurons.

Flood-Filling Networks (FFN) have proven effective for neuron segmentation. Januszewski et al. [63] utilized recursive mask extension for neurite skeletonization, while Dong et al. [16] extended this approach with 3D fields-of-view and memory channels for object instance predictions in FIB-25. For axon instance segmentation in AxonEM, Wei et al. [10] combined U-Net with FFN, employing affinity prediction to trace axonal morphologies.

Point-based and graph-based methods address topological structure preservation. Berman et al. [59] employed CurveNet for neuron segmentation in CREMI, utilizing curve-feature extraction for structural continuity. Several studies implemented graph network architectures for neuronal reconstruction, including Zhao et al. [51], Santurkar et al. [12], and Huang et al. [54]; notably, the Contextual Graph Model proposed by Santurkar et al. [12] used message passing for 2D-to-3D correspondence in nuclei segmentation.

For domain transfer, the Point Affinity Transformer for neuron segmentation in FlyWire, MANC, and Hemibrain, introduced by Troidl et al. [24], combined BCE with contrastive loss for point affinity prediction. Additionally, the modified SAM-Med3D for DendriteSAM developed by Zhuo et al. [20] incorporated Fact fine-tuning and a 3D LSD Encoder to adapt foundation models to EM data.

3.4.3. Subcellular structure segmentation

Synapses and vesicles. Subcellular structure segmentation represents a challenging aspect of brain EM analysis, requiring precise delineation at nanometer scale. For synaptic cleft segmentation on CREMI and FAFB, Heinrich et al. [9] employed 3D U-Net with signed distance transform regression, enabling geometric characterization of synaptic junctions. Li et al. [61] utilized a two-step detection network based on 3D U-Net for pre/post-synapse point annotations in WASPSYN to identify synaptic partners across synaptic clefts. Guo et al. [17] implemented region-focused Mask R-CNN for synapse and PSD reconstruction in Mouse ATUM-SEM with hierarchical clustering to identify synaptic regions. Matejek et al. [26] implemented compositional ConvNets with rules-based composition of marginal features for synapse detection and connectivity analysis across brain regions, providing a structured approach to synapse connectivity mapping.

Dendritic spines. Dendritic spine analysis presents challenges due to complex morphology. Schmidt et al. [11] presented a specialized approach for axon and spine neck reconstruction on SBEM/multiSEM datasets, implementing Bishop frame projection with membrane-avoiding flight mechanisms to trace thin connections of dendritic spines. Gonda et al. [35] utilized a vision foundation model with transformer architecture for dendrite and spine segmentation on DendriteSAM, implementing a ViT encoder with mask decoder to capture distinctive morphological features of dendritic spines. This approach outperformed traditional 2D methods for spine segmentation in rat hippocampus, demonstrating the value of transformer architectures for complex morphological features.

Mitochondria tracking. Mitochondrial tracking across image volumes has emerged as an important subcellular segmentation task. FragViT by Luo et al. [15] and the Adaptive Template Transformer by Pan et al. [43] employed hierarchical aggregation and structural template learning approaches to track mitochondria across MitoEM datasets. Their hierarchical attention strategies and template learning approaches effectively captured diverse mitochondrial morphologies across brain regions, enabling consistent mapping across large-scale datasets.

3.4.4. Multi-task segmentation

The simultaneous segmentation of multiple cellular and subcellular structures has gained traction as datasets and models have matured. As highlighted in Supplementary A Table 6, several approaches have demonstrated success in this challenging domain:

FragViT, proposed by Luo et al. [15], simultaneously segmented mitochondria and neurons in the MitoEM and EMNeuron datasets by utilizing fragment encoders with affinity heads and hierarchical fragment aggregation. This approach maintained distinct instance boundaries while sharing feature extraction layers. The Adaptive Template Transformer introduced by Pan et al. [43] combined structural template learning with optimal transport regularization for 3D mitochondria segmentation across MitoEM, Lucchi, and NucMM datasets. It effectively modeled shape priors while adapting to biological variability. RC-SPCNet, developed by Xiao et al. [42] for neuronal boundary segmentation in the ISBI EM Challenge, integrated deep residual networks with subpixel convolution and lifted multicut post-processing to achieve accurate boundary delineation while preserving neuronal connectivity. The Multi-CNN architecture presented by Urakubo et al. [33] fused 2D convolutional neural networks with 3D flood-filling networks (FFNs) for the segmentation of neurons, mitochondria, and synapses in the SNEMI3D dataset, demonstrating the effectiveness of specialized components for different cellular structures within a unified framework. U-Net variants proposed by Quan et al. [60] for multi-structure segmentation in the ISBI and Zebrafish EM datasets employed hybrid-representation learning using mixed 2D/3D convolutions. This design enabled effective handling of anisotropic data while segmenting multiple subcellular structures simultaneously. The hybrid encoder-decoder architecture introduced by Thawakar et al. [45] for 3D mitochondria segmentation in MitoEM and Lucchi incorporated split spatio-temporal attention (SST) and deformable convolutions to capture the complex morphological features of mitochondria across different datasets.

These multi-task approaches demonstrate the increasing sophistication of deep learning models for brain EM segmentation, moving beyond single-structure analysis toward a more holistic volumetric understanding of neural ultrastructure.

3.4.5. Training paradigms

The effectiveness of deep learning models for brain EM segmentation is highly dependent on the training approaches employed (Supplementary A Tables 4 and 5). Our analysis of 60 studies revealed four distinct training paradigms with varying degrees of representation: supervised learning (50 studies), self-supervised learning (6 studies), weakly supervised learning (3 studies), and unsupervised learning (1 study), as illustrated in Fig. 5.

Supervised learning methods, which require expert annotations, dominated the methodological landscape, comprising 83% of the reviewed studies [9,20,25,43,60,61,67]. This prevalence is exemplified by Heinrich et al. [9] and Quan et al. [60] in their work on synaptic cleft and multiple structure segmentation, respectively. These approaches consistently demonstrate high accuracy but are inherently constrained by the limited availability of labeled data—a significant challenge in EM analysis due to the labor-intensive nature of expert annotation; see Supplementary Material A, Tables 4–6 for complete study list.

Self-supervised methods, accounting for approximately 10% of studies [8,37,48,49,53,66], have emerged as promising alternatives to reduce annotation dependency. For example, Xu et al. [8] employed a

self-supervised 3D U-Net architecture for axon tracing in the SHIELD PVGPe dataset, while Cheng et al. [37] utilized MAE pre-training techniques for neural structure segmentation in SNEMI3D and MitoEM datasets. Notably, the contrastive learning approach introduced by Zhang et al. [49] for neuron and glia segmentation in EMNeuron employed NT-Xent and decorrelation losses to facilitate effective learning with minimal labeled data.

Weakly supervised approaches, representing 5% of the reviewed studies [8,37,48,49,53,66], effectively bridge the methodological gap between fully supervised and self-supervised methods. For instance, Huang et al. [52] utilized region growing with skeleton methods for 3D neuron reconstruction in fMOST and BigNeuron datasets, while Klinghoffer et al. [53] implemented similar techniques for 3D axon segmentation in SHIELD PVGPe. These approaches strategically leverage sparse annotations, which is particularly valuable in large-scale datasets where dense labeling is prohibitively expensive or time-consuming.

Semi-supervised learning, though less prevalent, demonstrates considerable promise as evidenced by the dual-branch network for mitochondria segmentation proposed by Mai et al. [44]. This approach incorporated prototype consistency alongside conventional cross-entropy loss mechanisms to efficiently utilize both labeled and unlabeled data resources. Unsupervised learning remains the least explored paradigm, with only Xin et al. [56] employing fully unsupervised techniques for EM segmentation.

The distribution of training paradigms reflects the inherent trade-off between segmentation accuracy and annotation efficiency in brain EM analysis. While supervised approaches currently dominate due to their superior performance metrics, the growing representation of alternative training paradigms suggests an evolving methodological landscape increasingly oriented toward annotation efficiency and generalizability.

3.4.6. Performance trends

Analysis across studies reveals key patterns (Supplementary A Table 7). 3D architectures consistently outperform two-dimensional approaches in anisotropic volumes, as demonstrated by the FFN implementation of Dong et al. [16] in FIB-25, which achieved superior continuity in neurite tracking compared to slice-based approaches. Transformer models enhance multi-task generalization capabilities. FragViT by Luo et al. [15] and the Point Affinity Transformer by Troidl et al. [24] both demonstrated superior performance in joint segmentation tasks and cross-dataset generalization. These models leverage their capacity for modeling long-range dependencies to capture structural relationships. Specialized loss functions also improve segmentation quality. Topological losses such as cDice, as used by Cheng et al. [34], Wei et al. [65], and Xu et al. [8], enhance neuron topology preservation compared to standard pixel-wise metrics, focusing optimization on connectivity patterns that are crucial for circuit reconstruction. Hybrid architectures that combine multiple components—such as the integration of SAM with 3D watershed algorithms by Thawakar et al. [45] and the modified SAM-Med3D by Zhuo et al. [20]—represent a promising direction for leveraging pre-trained foundation models while adapting to the challenges of brain EM data.

3.5. Evaluation and performance

Rigorous evaluation of brain EM segmentation methods requires metrics and validation strategies that capture both computational accuracy and biological relevance. Our analysis of performance trends across studies reveals how evaluation approaches have evolved to address the unique challenges of neural ultrastructure analysis, from traditional overlap metrics to connectivity-preserving measures. (see Supplementary A Table 7: Performance Evaluation and Reproducibility)

3.5.1. Validation strategies

Validation approaches for brain EM segmentation studies reveal methodological considerations specific to neural ultrastructure analysis (Supplementary A Table 7). Test set evaluation dominates, with studies employing held-out volumes or slices for performance assessment.

Rigorous studies implement cross-validation alongside test set evaluation to ensure generalizability. Liu et al. [57] employed cross-validation for boundary prediction in SEM and CREMI datasets, while Knowles-Barley et al. [39] utilized it for validating membrane segmentation across diverse brain regions. Hold-out validation was common in studies with limited data availability, as demonstrated in [53,64,65].

Expert validation through manual gold standards represents another critical approach. Huang et al. [52] validated neuron reconstruction against manual annotations, achieving average precision/recall above 0.99 in BigNeuron datasets. Similarly, Liu et al. [47] employed manual proofreading to validate axonal connectivity, reporting 86.61% correct edge identification with 98.10% path length coverage. Zhuo et al. [20] utilized expert validation to assess mask quality in adapting foundation models to brain EM data.

Out-of-distribution testing emerged as a rigorous validation approach for assessing generalization capabilities. Zhang et al. [49] evaluated contrastive learning methods across different datasets (vEM4, Basil, Harris), demonstrating robust performance (VOI=0.772–1.022) despite substantial domain shifts. This approach identified models capable of generalizing across brain regions and imaging conditions—a critical consideration for practical deployment.

3.5.2. Performance metrics and results

Performance evaluation in brain EM segmentation employs task-specific metrics designed to capture different aspects of segmentation quality (Supplementary A Table 7). Semantic segmentation relies primarily on overlap metrics (Dice/IoU/Jaccard), instance segmentation emphasizes topological correctness through VOI and ARAND, while specialized tasks employ connectivity-preserving measures such as Expected Run Length (ERL) and skeleton-focused metrics.

Membrane and Cell Segmentation achieved the highest performance levels. Xiao et al. [42] reported Rand scores up to 0.9878 on ISBI challenge data, while Jiang et al. [64] achieved IoU values of 0.9163 for neural cell bodies and 0.9883 for nuclei. These high scores reflect the maturity of membrane segmentation methods and the distinct signal characteristics of cell boundaries.

Mitochondria Segmentation emerged as another high-performing domain. Luo et al. [15], Pan et al. [43], and Mai et al. [44] achieved Jaccard scores between 85.0–90.5% and DSC between 91.9–95.0% on Lucchi datasets. Performance remained strong but decreased on more complex MitoEM datasets, with Luo et al. [15] and Pan et al. [43] reporting AP75 values of 0.679–0.958 for instance segmentation and AP scores of 76.9–78.2 (rat) and 67.5–68.2 (human), highlighting dataset complexity impact.

Neuron Instance Segmentation presented greater challenges due to complex branching structures. On CREMI datasets, composite scoring combining VOI and ARAND [9,25,58] revealed VOI ranging from 0.031 to 1.118 and ARAND from 0.008 to 0.117 across studies [23,40]. Troidl et al. [24] demonstrated strong performance on diverse datasets (Fly-Wire: VOI=0.17, ARE=0.04; MANC: VOI=0.20, ARE=0.05), while Matejek et al. [36] achieved VOI improvements of 20.9–28.7% on PNI datasets through recurrent network architectures.

Synapse Detection faced challenges due to small size and sparse distribution. Li et al. [61] reported F1 scores between 0.408 (baseline) and 0.616 (best performer) in the WASPSYN challenge, employing specialized synapse matching criteria with biologically relevant distance thresholds (88/52 nm). Shi et al. [38] achieved Dice=0.834 and mAP=0.865 on CREMI synaptic data, though these lower scores reflect the inherent difficulty in identifying small synaptic structures within complex neuropil.

Axon Tracing performance was evaluated through connectivity-preserving metrics that assess biological utility. Wei et al. [10] reported ERL values of 9.6–18.5 μm for AxonEM-H and 38.5 μm for AxonEM-M, measuring distance along neuronal processes before encountering segmentation errors. Skeleton-focused measures showed varied performance, with Zhao et al. [51] and Huang et al. [52] reporting ESA values of 1.43 ± 0.29 and 1.28 ± 0.30 , respectively. Additionally, topology-preserving metrics such as cDice and ρ -Dice, employed by Shamsi et al. [66] and Xu et al. [8], prioritized connectivity preservation over boundary precision.

Denosing and Reconstruction tasks utilized image quality metrics, with Potocek et al. [1] and Cheng et al. [34] reporting PSNR and SSIM values in denoising contexts, though these appeared less frequently in the literature.

Clear performance patterns emerged: subcellular organelles like mitochondria achieved higher accuracy than branching structures like axons; sparse structures like synapses presented persistent challenges; and performance consistently decreased with dataset complexity. Direct comparison between studies remains challenging due to differences in datasets, metrics, and evaluation protocols, underscoring the need for standardized benchmarking approaches in brain EM segmentation.

3.6. Practical considerations

The transition from research prototypes to practical deployment in neuroscience workflows involves critical considerations of computational scalability, methodological reproducibility, and cross-domain generalization. This section examines how studies address these real-world constraints and the strategies employed to make brain EM segmentation methods accessible and reliable for broader scientific application. (see Supplementary Table 7: Performance Evaluation and Reproducibility)

3.6.1. Scalability

Processing large EM volumes presents computational challenges, with studies employing diverse hardware configurations and algorithmic optimizations for scalability (Supplementary A Table 7). Hardware requirements scale with dataset complexity and model sophistication. High-end GPUs dominate, with RTX 3090, V100, and A100 models commonly employed. Shi et al. [38] utilized 8 $\times$ A800 GPUs for multi-task segmentation, while Zhuo et al. [20] leveraged A100 (40 GB) for foundation model adaptation. For large volumes, distributed systems become necessary: Januszewski et al. [63] employed 32 $\times$ K40 GPUs, Heinrich et al. [9] utilized 48 GPUs for CREMI challenge performance, and Wei et al. [10] required 32 $\times$ A100 GPUs for axon reconstruction. The most extreme case was Dong et al. [16], utilizing 2048 KNL nodes for FIB-25 segmentation.

Processing efficiency varies across approaches. Potocek et al. [1] reported inference times of 4–5s per image on Quadro P5000, while Gornet et al. [7] achieved throughput of 348.8 ± 1.9 gigavoxels/hour on GTX 1080 Ti. The approach proposed by Funke et al. [58] required 2.6 s per megavoxel, while Matejek et al. [26] reported a processing rate of 1 million voxels per second. These metrics provide useful information for estimating computational requirements.

Several studies reported computational optimizations. Schmidt et al. [11] achieved a 5–80 $\times$ speedup through membrane-avoiding flight techniques. Xin et al. [56] reduced feature extraction time from 7.04 to 1.52 ms while simultaneously improving performance. Block-based processing emerged as a common strategy, with Potocek et al. [1], Liu et al. [19], and Liu et al. [47] implementing sliding window approaches to address memory constraints.

3.6.2. Reproducibility

Code availability varied across studies (Supplementary A Tables 7 and 8). Approximately 40% (24 out of 60) provided public GitHub repositories, including studies such as Januszewski et al. [63], Plebani et al. [27], Cordero Cervantes et al. [46], Xiao et al. [42], Urakubo et al. [33], Lee et al. [67], Funke et al. [58], Zhuo et al. [20], Dorkenwald et al. [48], Zhang et al. [49], and Lin et al. [25]. Several studies also provided access through specialized platforms, for example, Lin et al. [62] via a custom website, Wei et al. [10] through a project website, and Knowles-Barley et al. [39] via a dedicated portal.

To provide a more granular analysis of the state of reproducibility, we conducted a structured audit of each available code repository, evaluating it based on three key criteria: the provision of pretrained model weights, the quality of documentation and instructions, and the overall completeness of the package for reproducing the paper's results. Our findings reveal a wide spectrum of reproducibility, from 'Full Reproduction Packages' that include pretrained models and clear instructions e.g., Plebani et al. [27], Chen et al. [13], Heinrich et al. [9], Dorkenwald et al. [48], and Archit et al. [69] to repositories that provide only source code with minimal guidance e.g., Liu et al. [19], Guo et al. [17], and Zhuo et al. [20]. A detailed breakdown of this audit for each study is presented in Supplementary A Table 8, which categorizes each repository to provide a clear overview of its utility for other researchers.

The depth of implementation details varied significantly. Most repositories included model definitions and training scripts, but fewer provided pretrained weights or complete pipelines. Comprehensive examples include the implementation of lifted multicut by Xiao et al. [42], the end-to-end pipeline for neuron instance segmentation by Funke et al. [58], and the adaptation of foundation models by Zhuo et al. [20].

Domain-specific reproducibility challenges emerged due to the use of specialized EM datasets. Many studies employed proprietary datasets, complicating replication efforts, while others relied on pre-processing or specialized data formats that were not fully documented. The most reproducible approaches, presented by Cordero Cervantes et al. [46], Xiao et al. [42], and Heinrich et al. [9], provided thorough documentation of preprocessing steps and model implementation.

Several studies highlighted the impact of training methodology on reproducibility. For example, Luo et al. [15] reported performance metrics for 160k training iterations using 2×RTX 3090 GPUs. Huang et al. [52] detailed the importance of hyperparameter selection, and Urakubo et al. [33] specified hardware configurations alongside public code.

3.6.3. Domain adaptation

Domain adaptation emerges as a critical consideration for generalizing segmentation approaches across datasets, brain regions, and species—a persistent challenge in brain EM given the variability in tissue preparation, imaging protocols, and neuroanatomy (Supplementary A Table 7).

Cross-species adaptation has demonstrated promising results. For example, Schmidt et al. [11] trained on mouse SBEM data and tested on human multiSEM datasets, reducing split errors from 65 to 28 mm. Chen et al. [13] employed synthetic defect generation for cross-species neuron morphology reconstruction. In another example, Zhang et al. [49] trained on 13 datasets and transferred to three out-of-distribution test sets, achieving VOI values of 0.772–1.022 and ARE of 0.102–0.153 through contrastive learning.

Foundation model adaptation offers an approach for handling domain variation. Zhuo et al. [20] adapted the SAM model to brain EM data, reporting improvements of approximately 0.1 and 0.3 in quality over standard SAM and Micro SAM, respectively. Cheng et al. [55] employed deformable tubular transfer in self-supervised learning contexts, while Shah et al. [21] utilized bi-directional self-prompting to achieve robust performance across mouse and human datasets.

Table 2

Summary statistics for key performance metrics (N = 46 experiments)

Metric	N	Mean	SD	Min	Max	Q50
Dice	15	0.896	0.132	0.549	0.974	0.948
Jaccard	11	0.903	0.033	0.846	0.947	0.905
AP	8	0.820	0.160	0.536	0.973	0.845
F1	7	0.880	0.083	0.717	0.973	0.880
Precision	10	0.945	0.068	0.790	0.998	0.978
Recall	10	0.919	0.108	0.632	0.996	0.951

Transfer learning strategies also appear in several studies. For instance, Liu et al. [19] leveraged detection models pretrained on natural images for synaptic segmentation, achieving AP mask values of 53.6–55.4% across datasets. Li et al. [61] applied transfer learning for pre- and post-synapse identification in the WASPSYN challenge, and Huang et al. [52] utilized transfer learning for neuron reconstruction across BigNeuron datasets.

Self-supervised and unsupervised domain adaptation techniques further reduce annotation requirements. Xin et al. [56] employed an unsupervised triplet margin loss for patch registration, improving CREMI Dice scores from 0.896 to 0.901 while reducing computational requirements. Additionally, Chen et al. [40] utilized self-attention mechanisms for domain adaptation across CREMI, Kasthuri, and Octopus datasets.

Collectively, these approaches address a fundamental challenge in brain EM analysis: variability across imaging conditions, brain regions, and species. Successful methods combine architectural innovations with specialized training regimes to achieve generalization, a capability essential for practical deployment across diverse neuroscience applications.

3.7. Meta-analysis of segmentation methods

To address the synthesis of cross-dataset performance trends, we conducted a comprehensive meta-analysis of segmentation methods. While our survey encompassed 60 studies across 28 datasets, stringent selection criteria yielded 27 studies (46 experiments) across 10 datasets for quantitative analysis. Selection criteria included: (i) clearly reported test sample sizes, (ii) evaluation on datasets used by ≥ 2 studies, and (iii) reporting of at least one standardized metric (Dice, Jaccard, AP, F1, Precision, Recall, VOI, ARAND, or Rand). (see Supplementary B Tables 1–6 and Figure 1–10).

Normalized performance overview. Table 2 presents normalized summary statistics across all evaluated metrics. The analysis reveals consistently high performance for overlap-based metrics (Dice: 0.896 ± 0.132 , Jaccard: 0.903 ± 0.033) with greater variability in instance-based metrics (AP: 0.820 ± 0.160).

Unified benchmark of top methods. Table 3 presents the unified benchmark of top-performing methods ranked by composite score. Foundation models and hybrid architectures dominate the top positions, with the 3D Transformer + U-Net achieving the highest composite score (0.954) with consistent performance across 5 datasets.

Method category performance analysis. Fig. 8 illustrates performance differences across method categories. Statistical analysis reveals significant differences only for AP scores ($F=19.03, p=0.008$), while Dice and Jaccard show no significant differences between categories ($p > 0.4$). Effect size analysis indicates large improvements when comparing CNN to foundation models (Cohen's $d=-1.28$ for Dice, -6.44 for AP).

Cross-dataset generalization. Critical to assessing method robustness, only 26% (12/46) of experiments evaluated methods on multiple datasets. The 3D Transformer + U-Net demonstrated the best generalization with coefficient of variation (CV) < 0.033 across 5 datasets. Methods with cross-dataset validation showed superior consistency (mean CV: 0.032) compared to single-dataset evaluations.

Table 3

Top 10 brain EM segmentation methods ranked by composite performance score in descending order, showing performance metrics and code availability.

Method	Cat.	Dice	Jacc.	AP	Comp.	DS	Code
3D Trans.+U-Net [38]	Fnd.	0.955	–	0.951	0.954	5	No
DDeep3M+ [2]	CNN	–	0.947	–	0.947	1	Yes
Hyb. enc-dec [45]	Hyb.	0.962	0.913	–	0.938	2	Yes
U-Net+FFN [33]	CNN	0.950	0.910	–	0.930	1	Yes
Dual-br. attn Mai et al. [44]	Trn.	0.938	0.884	–	0.911	2	No
UTR [56]	CNN	0.901	–	–	0.901	1	No
Mod. U-Net [22]	CNN	–	0.893	–	0.893	2	Yes
SAM2 [21]	Fnd.	0.924	0.861	–	0.893	1	Yes
ATFormer [43]	Hyb.	0.948	0.902	0.782	0.854	3	No
FragViT [15]	Trn.	0.950	0.905	0.769	0.848	3	No

Methods ranked by composite score in descending order. Fnd.: Foundation, Hyb.: Hybrid, Trn.: Transformer, Jacc.: Jaccard, Comp.: Composite Score, DS: Datasets, enc-dec: encoder-decoder, attn: attention, br.: branch, Mod.: Modified.

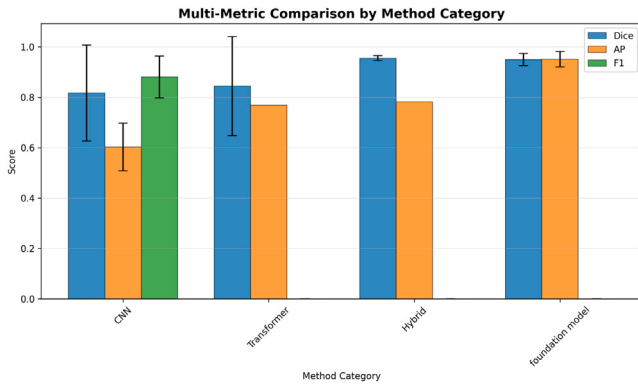


Fig. 8. Multi-metric performance comparison across method categories. Foundation models and hybrid approaches show superior performance across most metrics. See Supplementary Material B for detailed statistical analysis.

Key findings. Our meta-analysis reveals: (1) Foundation models outperform traditional CNNs by 13%–35% on average metrics; (2) Code availability remains limited (54% overall), with CNNs showing highest availability (67%); (3) Dataset complexity significantly impacts performance, with specialized datasets (e.g., BigNeuron) showing 40% lower scores; (4) Strong metric correlations exist for overlap measures (Dice-Jaccard: $r=0.89$) but weak correlations with instance metrics ($r<0.4$).

For detailed analyses including effect sizes, dataset-specific performance, and additional visualizations, see Supplementary Material B.

4. Discussion

Our review of 60 studies, combined with quantitative meta-analysis of 27 studies (46 experiments) across 10 datasets, reveals a field in rapid transition from traditional CNN-based approaches toward sophisticated hybrid and foundation model architectures. The evolution of brain EM segmentation reflects both advancing deep learning capabilities and growing understanding of the unique challenges posed by neuroanatomical data at nanometer resolution.

4.1. Performance landscape and architectural evolution

Our meta-analysis provides the first extensive benchmark comparison across brain EM segmentation methods, revealing significant performance disparities between architectural approaches. Foundation models and hybrid architectures demonstrate clear superiority, with the 3D Transformer + U-Net achieving the highest composite score (0.954) across five datasets. Statistical analysis confirms that while

overlap-based metrics (Dice, Jaccard) show no significant differences between method categories, instance-based metrics reveal substantial advantages for foundation models, with effect sizes reaching Cohen's $d = -6.44$ for Average Precision when compared to traditional CNNs.

The dominance of U-Net variants (appearing in 42% of studies) reflects their architectural advantages for maintaining spatial precision while capturing multi-scale features. However, the most successful implementations have moved far beyond the original architecture, incorporating residual connections, attention mechanisms, and specialized 3D optimizations. The emergence of transformer-based approaches represents the most significant recent innovation, with these methods demonstrating superior ability to capture long-range dependencies critical for preserving neural connectivity patterns.

Our analysis reveals that only 26% of experiments evaluated methods on multiple datasets, highlighting a critical gap in cross-dataset validation. Methods with cross-dataset evaluation showed superior consistency (mean coefficient of variation: 0.032) compared to single-dataset approaches, underscoring the importance of robust generalization testing. The 3D Transformer + U-Net demonstrated exceptional generalization capability with coefficient of variation < 0.033 across five datasets.

Flood-filling networks have emerged as particularly effective for neurite segmentation through their iterative field-of-view approach, while topology-preserving loss functions have shown substantial improvements in preserving thin axonal structures. Multi-representation approaches that transition between voxel, skeleton, and graph representations effectively capture the inherent network properties of neural structures. Foundation model adaptation techniques, as demonstrated in recent studies, reduce annotation requirements while improving cross-dataset generalization—a critical advancement given the annotation burden inherent in EM data.

4.2. Dataset characteristics and evaluation evolution

The dataset landscape encompasses 28 distinct datasets across multiple species and brain regions, creating both opportunities and challenges for method development. Our meta-analysis reveals substantial performance variations across datasets, with specialized datasets like BigNeuron showing 40% lower scores compared to standardized benchmarks. Publicly available datasets (MitoEM, CREMI, SNEMI3D, Lucchi) have enabled more consistent benchmarking, particularly for mitochondria segmentation where current state-of-the-art achieves Jaccard scores of 87.2–90.5%.

A critical finding from our analysis is the evolution of evaluation metrics beyond pixel-wise accuracy toward topology-aware measures. Traditional metrics (Dice, Jaccard) show strong correlation ($r=0.89$) but weak correlation with instance-based metrics ($r<0.4$), highlighting their limitation in capturing biological significance. The emergence of topology-preserving metrics (cDice, ERL) represents a paradigm shift toward evaluation frameworks that prioritize structural continuity and connectivity preservation over pixel-wise accuracy.

Dataset fragmentation reflects the diversity of research questions in connectomics but complicates direct method comparison. Most EM datasets exhibit substantial anisotropy with lower z -axis resolution, and while several specialized architectures have addressed this challenge, no definitive solution has emerged. The median training dataset contains only a few hundred annotated slices, insufficient for fully capturing neuroanatomical complexity, driving the recent focus on self-supervised and foundation model approaches.

4.3. Visual platform capabilities and segmentation examples

Our review reveals that the advancement of brain EM segmentation has been significantly enabled by sophisticated visualization platforms that integrate computer graphics techniques with neuroscience workflows. Modern platforms such as WEBKNOSSOS Boergens et al.

[72] demonstrate the critical intersection of high-performance computing, 3D visualization, and collaborative data analysis in contemporary connectomics research.

Fig. 9 illustrates the advanced capabilities of modern brain EM segmentation platforms. The specialized streaming and rendering technology (Fig. 9a) enables real-time browsing of multi-terabyte EM datasets through standard web browsers, eliminating computational barriers for collaborative research. GPU-accelerated 3D mesh visualization (Fig. 9b) allows researchers to explore segmented neural structures in three dimensions, with direct export capabilities to professional rendering software.

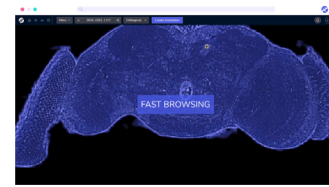
Optimized interfaces for Volume EM data navigation (Fig. 9c) demonstrate how raw EM datasets can be efficiently visualized and analyzed. The platform's segmentation capabilities encompass both automated AI-assisted volume annotations and manual refinement tools (Fig. 9d), enabling researchers to generate training data for machine learning models while maintaining annotation quality control. Skeleton-based segmentation approaches (Fig. 9e) complement volume segmentation by enabling high-speed neural tracing, with trained annotators achieving speeds of 1.5 ± 0.6 mm/h for axons and 2.1 ± 0.9 mm/h for dendrites in 3D EM data Boergens et al. [72].

These platform capabilities directly support the methodological diversity observed in our review, where 83% of studies employed supervised learning approaches with varying degrees of manual annotation refinement. To demonstrate practical applications of these platform capabilities, Fig. 10 presents real-world segmentation examples from our brain EM datasets. The multi-view interface (Fig. 10a) showcases interactive segmentation and 3D reconstruction capabilities, with orthogonal slice views (XY, YZ, XZ) displaying segmented dendritic structures and myelinated axons overlaid on high-resolution grayscale EM data. Multi-planar navigation (Fig. 10b) demonstrates synchronous cross-sectional analysis enabling verification of segmentation accuracy across spatial dimensions. Fine-scale synaptic structure annotation (Fig. 10c,d) illustrates precise mapping of dendritic spines, presynaptic boutons, and glial processes, revealing the complex ultrastructural relationships essential for understanding neural microcircuits. The 3D volumetric reconstructions (Fig. 10e) provide high-fidelity mesh representations that enable quantitative morphological analysis and spatial relationship assessment at nanoscale resolution.

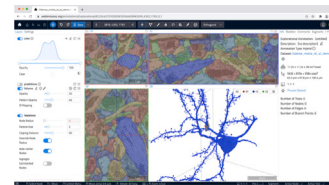
These visual examples demonstrate both the remarkable capabilities and persistent limitations identified in our in-depth analysis of current segmentation approaches across diverse neural structures. While platforms successfully segment clearly defined neuronal components and synaptic structures, significant challenges remain in regions with poor membrane contrast, imaging artifacts, or complex ultrastructural arrangements—consistent with our findings that CNN-based methods achieve superior performance on well-defined structures but struggle with complex morphologies. Our meta-analysis confirms that the diversity of available tools, from AI-assisted volume annotations to high-speed skeleton tracing, reflects the methodological fragmentation observed across the 46 experiments reviewed. The visual workflows presented illustrate how computer graphics advances enable sophisticated analysis, yet our evaluation reveals substantial gaps in standardized benchmarking and reproducibility standards. Moving forward, the integration of these advanced visualization platforms with the emerging transformer-based and foundation models identified in our review presents significant opportunities for improving both segmentation accuracy and collaborative research workflows in connectomics.

4.4. Challenges, limitations, and reproducibility

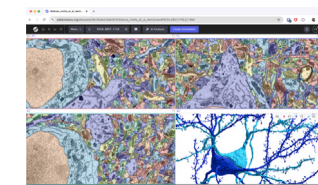
Despite significant progress, several interconnected challenges persist across both the field and our review methodology. The annotation burden remains prohibitive, requiring specialized expertise and extensive time investment that limits dataset scale and diversity. This challenge is compounded by architectural tradeoffs where 3D approaches



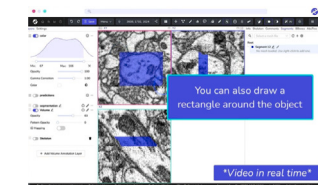
(a) Fast browsing capabilities with specialized storage, streaming, and rendering technology enabling real-time access to multi-terabyte EM datasets through web browsers without additional installations.



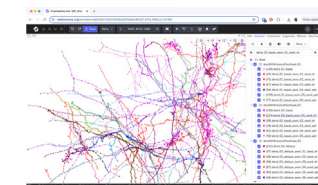
(b) 3D mesh visualization of segmented neural structures with automatic mesh generation from volume annotations and export capabilities to professional 3D rendering software.



(c) Volume EM interface optimized for large raw EM data navigation with real-time viewport rendering and 3D mesh computation capabilities.



(d) AI-assisted volume annotation tools featuring quick-select functionality, interpolation features, and manual brush/trace refinement for machine learning training data generation.



(e) Skeleton annotation interface for high-speed neural tracing with Flight Mode capabilities, achieving professional annotation speeds of 1.5 ± 0.6 mm/h for axons and 2.1 ± 0.9 mm/h for dendrites, with hierarchical organization and annotation features.

Fig. 9. Overview of modern brain EM segmentation platform capabilities demonstrating the integration of computer graphics techniques with neuroscience workflows. These visual examples illustrate both successful segmentation of diverse neural structures (neurons, synapses, organelles) and the technological infrastructure supporting collaborative connectomics research. Images courtesy of WEBKNOSSOS platform (<https://webknossos.org/>).

better capture structural continuity but impose substantially higher computational demands, forcing many studies toward patch-based processing or downsampling that can compromise fine structural details.

Computational scalability presents a significant barrier to practical deployment, with processing teravoxel-scale volumes requiring distributed computing resources unavailable to many research groups. The most computationally intensive approaches required dozens of high-end GPUs, achieving processing throughput of approximately 3 mega-voxels per second per GPU but limiting widespread adoption. This computational burden is particularly challenging given that only 54% of studies provide publicly available code, with CNNs showing the highest availability (67%) compared to newer architectural approaches.

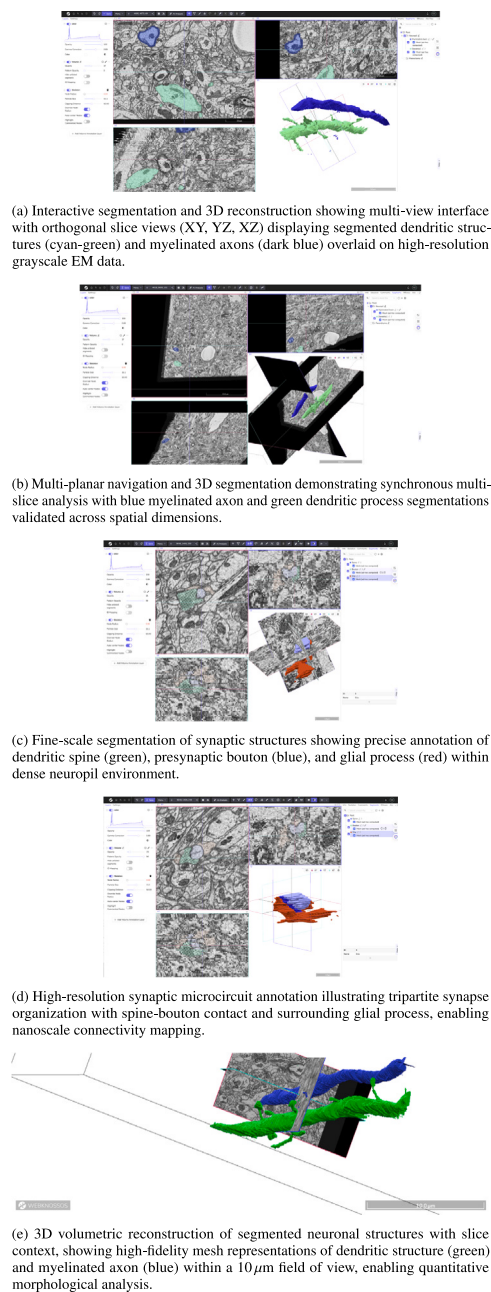


Fig. 10. Brain EM segmentation examples illustrating platform performance on neural tissue datasets. Diverse ultrastructural components—neurons, synapses, glia—are segmented at nanoscale resolution, demonstrating both precision and challenges in complex neuropil. Data acquired with Zeiss Volutome on Gemini 460 SEM at Neuroscience Institute Cavalieri Ottolenghi (Orbassano, Italy).

Domain adaptation across different brain regions, species, or imaging protocols remains challenging, with only 23% of studies explicitly addressing generalization through transfer learning or augmentation techniques. Our meta-analysis confirms this limitation, showing that methods evaluated on single datasets often fail to maintain performance when applied more broadly.

Our review itself faces limitations including potential publication bias, particularly for conference proceedings not indexed in major databases, and heterogeneous reporting that limited quantitative synthesis capabilities. The rapid evolution of deep learning methods means some findings may be superseded by the time of publication, while our

reproducibility evaluation was based on code availability rather than verification of implementation completeness or functionality.

4.5. Implications and future directions

The findings from our meta-analysis and systematic review have significant implications for both connectomics research and the broader visual computing community. The superior performance of foundation models (13%–35% improvement over traditional CNNs) suggests that adapting large pretrained vision models represents the most promising near-term direction for reducing annotation requirements while improving generalization. Self-supervised learning techniques show particular promise for large-scale EM datasets where complete manual annotation is infeasible.

For researchers in connectomics, methodological choices should be guided by specific research questions and computational constraints. Our analysis confirms that specialized algorithms consistently outperform general-purpose approaches for targeted studies of specific structures, while approaches prioritizing topological correctness over pixel-wise accuracy yield more biologically meaningful results for large-scale connectivity mapping. The strong performance of topology-aware metrics suggests that evaluation strategies should shift away from traditional pixel-based measures toward metrics that better reflect biological significance.

The challenges addressed in brain EM segmentation extend far beyond neuroscience, offering significant contributions to visual computing through innovations in processing volumetric data at scale, topology-preserving evaluation metrics, and multi-representation approaches. The volume sizes encountered (exceeding 50 teravoxels) have driven developments in distributed computing frameworks and memory-efficient inference techniques that benefit other data-intensive domains. Topology-preserving metrics ensure structural continuity in complex 3D networks—a challenge shared across medical imaging, scientific visualization, and computer graphics.

Future research priorities should focus on establishing standardized benchmark datasets and evaluation protocols to enable fair method comparison and track progress over time. Multi-task learning frameworks that simultaneously segment multiple structures and predict their relationships could improve both biological consistency and computational efficiency. Human-in-the-loop systems that efficiently incorporate expert feedback present opportunities for minimizing manual effort while ensuring biological plausibility.

The field's evolution toward integration across scales—connecting ultrastructural EM data with other imaging modalities—represents perhaps the most significant frontier for visual computing research in connectomics. This integration, combined with advancing foundation model capabilities and topology-aware evaluation frameworks, positions brain EM segmentation to enable high-resolution neural circuit mapping at scales previously considered intractable, potentially bridging the gap between ultrastructural organization and functional properties of neural networks.

Success in addressing these challenges requires stronger reproducible research practices, including standardized code sharing, model distribution, and evaluation protocols. The field must move beyond proof-of-concept demonstrations toward practical tools that accelerate neuroscience discovery through reliable, scalable, and biologically meaningful segmentation of brain ultrastructure.

5. Conclusion

This review and meta-analysis of 60 studies, with quantitative synthesis of 27 studies (46 experiments), provides the first extensive benchmark comparison for brain EM segmentation methods. Our analysis reveals a field transitioning from traditional CNN approaches toward foundation models and hybrid architectures that demonstrate clear performance superiority. Our meta-analysis establishes that foundation

models outperform traditional CNNs by 13%–35% across key metrics, with the 3D Transformer + U-Net achieving the highest composite score (0.954) across five datasets. Statistical analysis confirms significant advantages for foundation models in instance-based metrics (Cohen's $d = -6.44$), while topology-aware evaluation metrics (cDice, ERL) better capture biological significance than traditional pixel-wise measures. Critical challenges persist: only 26% of experiments validate across multiple datasets, 54% of studies lack public code repositories, and computational scalability limits practical deployment. However, methods with cross-dataset validation show superior consistency (coefficient of variation: 0.032), indicating that robust evaluation practices yield more reliable approaches. The field's evolution toward foundation model adaptation and topology-preserving evaluation frameworks positions brain EM segmentation to enable large-scale connectomics studies previously considered intractable. Our unified benchmarks provide a foundation for reliable progress tracking, while the superior performance of specialized, domain-specific architectures over computationally intensive general approaches offers practical guidance for resource-constrained laboratories.

Success in advancing this field requires strengthened reproducible research practices and standardized evaluation protocols. As brain EM segmentation matures, integration with multimodal imaging promises to bridge nanometer-resolution ultrastructure with whole-brain connectivity, fundamentally advancing our understanding of neural organization and enabling neuroscientific discoveries impossible through manual analysis alone.

CRediT authorship contribution statement

Uzair Shah: Writing – original draft, Visualization, Methodology, Investigation, Conceptualization. **Mahmood Alzubaidi:** Writing – original draft, Visualization, Methodology, Investigation, Conceptualization. **Marco Agus:** Writing – original draft, Validation, Methodology, Investigation, Conceptualization. **Corrado Calì:** Writing – review & editing. **Pierre J. Magistretti:** Writing – review & editing. **Mowafa Househ:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used Claude (sonnet 3.7) to assist with English editing and to enhance sentence construction. After using these tools, the author(s) thoroughly reviewed and edited the content as needed and take full responsibility for the final content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This publication was funded by the PPM-7th Cycle grant (PPM 07-0409-240041, AMAL-For-Qatar) from the Qatar National Research Fund, a member of the Qatar Foundation. The findings herein reflect the work and are solely the responsibility, of the authors. We thank the PICO imaging facility at the NICO institute, Roberta Parolisi for the imaging support, and Christel Genoud at UNIL for the help in sample preparation. Open Access funding provided by the Qatar National Library.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.cag.2025.104391>.

Data availability

No data was used for the research described in the article.

References

- [1] Potocek P, Trampert P, Peemen M, Schoenmakers R, Dahmen T. Sparse scanning electron microscopy data acquisition and deep neural networks for automated segmentation in connectomics. *Microsc Microanal* 2020;26(3):403–12. <http://dx.doi.org/10.1017/S1431927620001361>.
- [2] Xiao R, Zhu L, Liao J, Wu X, Gong H, Huang J, et al. DDeep3M+: Adaptive enhancement powered weakly supervised learning for neuron segmentation. *Neurophotonics* 2023;10(3). <http://dx.doi.org/10.1117/1.NPh.10.3.035003>, 035003–035003.
- [3] Zheng Z, Own CS, Wanner AA, Koene RA, Hammerschmith EW, Silversmith WM, et al. Fast imaging of millimeter-scale areas with beam deflection transmission electron microscopy. *Nat Commun* 2024;15(1):6860. <http://dx.doi.org/10.1038/s41467-024-50846-4>.
- [4] Zhang J, Reyes BA, Ross JA, Trovillion V, Van Bockstaele EJ. Advances in neuroscience using transmission electron microscopy: A historical perspective. *Transm Electron Microsc Methods Underst Brain* 2016;1–20. http://dx.doi.org/10.1007/7657_2016_101.
- [5] Wanner A, Kirschmann M, Genoud C. Challenges of microtome-based serial block-face scanning electron microscopy in neuroscience. *J Microsc* 2015;259(2):137–42. <http://dx.doi.org/10.1111/jmi.12244>.
- [6] Qi X, Yang T, Li L, Wang J, Zeng S, Lv X. Fluorescence micro-optical sectioning tomography using acousto-optical deflector-based confocal scheme. *Neurophotonics* 2015;2(4). <http://dx.doi.org/10.1117/1.NPh.2.4.041406>, 041406–041406.
- [7] Gornet J, Venkataraju KU, Narasimhan A, Turner N, Lee K, Seung HS, et al. Reconstructing neuronal anatomy from whole-brain images. In: 16th international symposium on biomedical imaging. IEEE; 2019, p. 218–22. <http://dx.doi.org/10.1109/ISBI.2019.8759197>.
- [8] Xu AS, Shamsi NI, Gjestebj LA, Brattain LJ. Self-supervised edge detection reconstruction for topology-informed 3D axon segmentation and centerline detection. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2024, p. 7831–9. <http://dx.doi.org/10.1109/WACV57701.2024.00765>.
- [9] Heinrich L, Funke J, Pape C, Nunez-Iglesias J, Saalfeld S. Synaptic cleft segmentation in non-isotropic volume electron microscopy of the complete drosophila brain. In: Medical image computing and computer assisted intervention–MICCAI. Springer; 2018, p. 317–25. http://dx.doi.org/10.1007/978-3-030-00934-2_36.
- [10] Wei D, Lee K, Li H, Lu R, Bae JA, Liu Z, et al. AxonEM dataset: 3D axon instance segmentation of Brain Cortical Regions. In: Medical image computing and computer assisted intervention–MICCAI. Springer; 2021, p. 175–85. http://dx.doi.org/10.1007/978-3-030-87193-2_17.
- [11] Schmidt M, Motta A, Sievers M, Helmstaedter M. RoboEM: Automated 3D flight tracing for synaptic-resolution connectomics. *Nature Methods* 2024;21(5):908–13. <http://dx.doi.org/10.1038/s41592-024-02226-5>.
- [12] Santurkar S, Budden D, Matveev A, Berlin H, Saribekyan H, Meirovitch Y, et al. Toward streaming synapse detection with compositional ConvNets. 2017. <http://dx.doi.org/10.48550/arXiv.1702.07386>, arXiv preprint [arXiv:1702.07386](https://arxiv.org/abs/1702.07386).
- [13] Chen W, Liu M, Du H, Radojević M, Wang Y, Meijering E. Deep-learning-based automated neuron reconstruction from 3D microscopy images using synthetic training images. *IEEE Trans Med Imaging* 2021;41(5):1031–42. <http://dx.doi.org/10.1109/TMI.2021.3130934>.
- [14] Wu X. An iterative convolutional neural network algorithm improves electron microscopy image segmentation. 2015. <http://dx.doi.org/10.48550/arXiv.1506.05849>, arXiv preprint [arXiv:1506.05849](https://arxiv.org/abs/1506.05849).
- [15] Luo N, Sun R, Pan Y, Zhang T, Wu F. Electron microscopy images as set of fragments for mitochondrial segmentation. In: Proceedings of the AAAI conference on artificial intelligence, vol. 38, 2024, p. 3981–9. <http://dx.doi.org/10.1609/aaai.v38i4.28191>.
- [16] Dong W, Kececi M, Vescevi R, Li H, Adams C, Jennings E, et al. Scaling distributed training of flood-filling networks on HPC infrastructure for brain mapping. In: Third workshop on deep learning on supercomputers. IEEE; 2019, p. 52–61. <http://dx.doi.org/10.1109/DLS49591.2019.00012>.
- [17] Guo J, Liu J, Hong B, Zhang Y, Xu Y, Han H. Planar to spatial: A synapse reconstruction method to rebuild voxel connections for anisotropic serial EM images. In: 20th international symposium on biomedical imaging. IEEE; 2023, p. 1–5. <http://dx.doi.org/10.1109/ISBI53787.2023.10230767>.
- [18] Wei X, Liu M, Wang Y, Meijering E. 3D soma detection in large-scale whole brain images via a two-stage neural network. *IEEE Trans Med Imaging* 2022;42(1):148–57. <http://dx.doi.org/10.1109/TMI.2022.3206605>.
- [19] Liu J, Hong B, Xiao C, Zhai H, Shen L, Xie Q, et al. A novel 3D instance segmentation network for synapse reconstruction from serial electron microscopy images. *Expert Syst Appl* 2024;255:124562. <http://dx.doi.org/10.1016/j.eswa.2024.124562>.

- [20] Zhuo Z, Belevich I, Leinonen V, Jokitalo E, Malm T, Sierra A, et al. Segment anything for dendrites from electron microscopy. In: IEEE 6th international conference on image processing, applications and systems. 2025, p. 1–6. <http://dx.doi.org/10.1109/IPAS63548.2025.10924514>.
- [21] Shah U, Agus M, Boges D, Chiappini V, Alzubaidi M, Schneider J, et al. SAM4EM: Efficient memory-based two stage prompt-free segment anything model adapter for complex 3D neuroscience electron microscopy stacks. In: Proceedings of the computer vision and pattern recognition conference (CVPR) workshops. 2025, p. 4678–87.
- [22] Casser V, Kang K, Pfister H, Haehn D. Fast mitochondria detection for connectomics. In: Medical imaging with deep learning. PMLR; 2020, p. 111–20.
- [23] Sun R, Luo N, Pan Y, Mai H, Zhang T, Xiong Z, et al. Appearance prompt vision transformer for connectome reconstruction. In: IJCAI. 2023, p. 1423–31. <http://dx.doi.org/10.24963/ijcai.2023/158>.
- [24] Trold J, Knittel J, Li W, Zhan F, Pfister H, Turaga S. Global neuron shape reasoning with point affinity transformers. 2024, <http://dx.doi.org/10.1101/2024.11.24.625067>, BioRxiv.
- [25] Lin Z, Wei D, Lichtman J, Pfister H. Pytorch connectomics: A scalable and flexible segmentation framework for emconnectomics. 2021, <http://dx.doi.org/10.48550/arXiv.2112.05754>, arXiv preprint arXiv:2112.05754.
- [26] Matejek B, Franzmeyer T, Wei D, Wang X, Zhao J, Palágyi K, et al. Scalable biologically-aware skeleton generation for connectomic volumes. IEEE Trans Med Imaging 2022;41(9):2360–70. <http://dx.doi.org/10.1109/TMI.2022.3164343>.
- [27] Plebani E, Biscola NP, Havton LA, Rajwa B, Shemonti AS, Jaffey D, et al. High-throughput segmentation of unmyelinated axons by deep learning. Sci Rep 2022;12(1):1198. <http://dx.doi.org/10.1038/s41598-022-04854-3>.
- [28] Aswath A, Alsahaf A, Giepmans BN, Azzopardi G. Segmentation in large-scale cellular electron microscopy with deep learning: A literature survey. Med Image Anal 2023;89:102920. <http://dx.doi.org/10.1016/j.media.2023.102920>.
- [29] Liu Z, Jin L, Chen J, Fang Q, Ablameyko S, Yin Z, et al. A survey on applications of deep learning in microscopy image analysis. Comput Biol Med 2021;134:104523. <http://dx.doi.org/10.1016/j.combiomed.2021.104523>.
- [30] Alahmari SS, Goldhof D, Hall LO, Mouton PR. A review of nuclei detection and segmentation on microscopy images using deep learning with applications to unbiased stereology counting. IEEE Trans Neural Netw. Learn Syst 2022;35(6):7458–77. <http://dx.doi.org/10.1109/TNNLS.2022.3213407>.
- [31] Treder KP, Huang C, Kim JS, Kirkland AI. Applications of deep learning in electron microscopy. Microscopy 2022;71(Supplement_1):i100–15. <http://dx.doi.org/10.1093/jmicro/dfab043>.
- [32] Huo Y, Lu Z, Deng Z, Zhang F, Xiong J, Zhang P, et al. A survey of deep learning-based microscopic cell image understanding. Displays 2025;102968. <http://dx.doi.org/10.1016/j.displa.2025.102968>.
- [33] Urakubo H, Bullmann T, Kubota Y, Oba S, Ishii S. UNI-EM: An environment for deep neural network-based automated segmentation of neuronal electron microscopic images. Sci Rep 2019;9(1):19413. <http://dx.doi.org/10.1038/s41598-019-55431-0>.
- [34] Cheng A, Ma G, Zheng L, Chen Y, Wang L, Zhang R. Enhancing neural segmentation on low-quality SEM images using self-attention based fusion module with auxiliary deblurring network. In: 2023 11th international conference on bioinformatics and computational biology. IEEE; 2023, p. 189–95. <http://dx.doi.org/10.1109/ICBCB57893.2023.10246655>.
- [35] Gonda F, Wei D, Pfister H. Consistent recurrent neural networks for 3D neuron segmentation. In: 18th international symposium on biomedical imaging. IEEE; 2021, p. 1012–6. <http://dx.doi.org/10.1109/ISBI48211.2021.9434092>.
- [36] Matejek B, Haehn D, Zhu H, Wei D, Parag T, Pfister H. Biologically-constrained graphs for global connectomics reconstruction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019, p. 2089–98. <http://dx.doi.org/10.1109/CVPR.2019.00219>.
- [37] Cheng A, Shi J, Wang L, Zhang R. Learning the heterogeneous representation of brain's structure from serial SEM images using a masked autoencoder. Front Neuroinformatics 2023;17:1118419. <http://dx.doi.org/10.3389/fninf.2023.1118419>.
- [38] Shi R, Pang Q, Ma L, Duan L, Huang T, Jiang T. ShapeMamba-EM: Fine-tuning foundation model with local shape descriptors and mamba blocks for 3D EM image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer; 2024, p. 731–41. http://dx.doi.org/10.1007/978-3-031-72390-2_68.
- [39] Knowles-Barley S, Kaynig V, Jones TR, Wilson A, Morgan J, Lee D, et al. RhonaNet pipeline: Dense automatic neural annotation. 2016, <http://dx.doi.org/10.48550/arXiv.1611.06973>, arXiv preprint arXiv:1611.06973.
- [40] Chen M, Renuka MB, Mi L, Lichtman J, Shavit N, Meirovitch Y. Learning to correct sloppy annotations in electron microscopy volumes. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023, p. 4273–84. <http://dx.doi.org/10.1109/CVPRW59228.2023.00450>.
- [41] Turner NL, Lee K, Lu R, Wu J, Ih D, Seung HS. Synaptic partner assignment using attentional voxel association networks. In: 17th international symposium on biomedical imaging. IEEE; 2020, p. 1–5. <http://dx.doi.org/10.1109/ISBI45749.2020.9098489>.
- [42] Xiao C, Hong B, Liu J, Tang Y, Xie Q, Han H. Deep residual contextual and subpixel convolution network for automated neuronal structure segmentation in micro-connectomics. Comput Methods Programs Biomed 2022;219:106759. <http://dx.doi.org/10.1016/j.cmpb.2022.106759>.
- [43] Pan Y, Luo N, Sun R, Meng M, Zhang T, Xiong Z, et al. Adaptive template transformer for mitochondria segmentation in electron microscopy images. In: Proceedings of the IEEE/CVF international conference on computer vision. 2023, p. 21474–84. <http://dx.doi.org/10.1109/ICCV51070.2023.01963>.
- [44] Mai H, Sun R, Zhang T, Xiong Z, Wu F. DualRel: Semi-supervised mitochondria segmentation from a prototype perspective. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023, p. 19617–26. <http://dx.doi.org/10.1109/CVPR52729.2023.01879>.
- [45] Thawakar O, Anwer RM, Laaksonen J, Reiner O, Shah M, Khan FS. 3D mitochondria instance segmentation with spatio-temporal transformers. In: International conference on medical image computing and computer-assisted intervention. Springer; 2023, p. 613–23. http://dx.doi.org/10.1007/978-3-031-43993-3_59.
- [46] Cordero Cervantes D, Khare H, Wilson AM, Mendoza ND, Coulon-Mahdi O, Lichtman JW, et al. 3D reconstruction of the cerebellar germinal layer reveals tunneling connections between developing granule cells. Sci Adv 2023;9(14):eadf3471. <http://dx.doi.org/10.1126/sciadv.adf3471>.
- [47] Liu Z, Zhang Y, Yu Y, Wang Y, He Y, Zhao Q, et al. High-speed automated reconstruction of drosophila larval brain from volumetric EM data. In: 45th annual international conference of the IEEE Engineering in Medicine & Biology Society. IEEE; 2023, p. 1–4. <http://dx.doi.org/10.1109/EMBC40787.2023.10340599>.
- [48] Dorkenwald S, Li PH, Januszewski M, Berger DR, Maitin-Shepard J, Bodor AL, et al. Multi-layered maps of neuropil with segmentation-guided contrastive learning. Nature Methods 2023;20(12):2011–20. <http://dx.doi.org/10.1038/s41592-023-02059-8>.
- [49] Zhang Y, Guo J, Zhai H, Liu J, Han H. SegNeuron: 3D neuron instance segmentation in any EM volume with a generalist model. In: International conference on medical image computing and computer-assisted intervention. Springer; 2024, p. 589–600. http://dx.doi.org/10.1007/978-3-031-72111-3_55.
- [50] Wang H, Zhang D, Song Y, Liu S, Huang H, Chen M, et al. Multiscale Kernels for Enhanced U-Shaped Network to Improve 3D Neuron Tracing. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2019, <http://dx.doi.org/10.1109/CVPRW.2019.00144>.
- [51] Zhao R, Wang H, Zhang C, Cai W. PointNeuron: 3D neuron reconstruction via geometry and topology learning of point clouds. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2023, p. 5787–97. <http://dx.doi.org/10.1109/WACV56688.2023.00574>.
- [52] Huang Q, Chen Y, Liu S, Xu C, Cao T, Xu Y, et al. Weakly supervised learning of 3D deep network for neuron reconstruction. Front Neuroanat 2020;14:38. <http://dx.doi.org/10.3389/fnana.2020.00038>.
- [53] Klinghoffer T, Morales P, Park YG, Evans N, Chung K, Brattain LJ. Self-supervised feature extraction for 3D axon segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020, p. 978–9. <http://dx.doi.org/10.1109/CVPRW50498.2020.00497>.
- [54] Huang Q, Cao T, Chen Y, Li A, Zeng S, Quan T. Automated neuron tracing using content-aware adaptive voxel scooping on CNN predicted probability map. Front Neuroanat 2021;15:712842. <http://dx.doi.org/10.3389/fnana.2021.712842>.
- [55] Cheng YS, Zhao R, Wang H, Peng H, Lo Y, Chen Y, et al. DiNeuro: Distilling knowledge from 2D natural images via deformable tubular transferring strategy for 3D neuron reconstruction. In: IEEE 22nd international symposium on biomedical imaging. 2025, p. 1–5. <http://dx.doi.org/10.1109/ISBI60581.2025.10981082>.
- [56] Xin T, Chen B, Chen X, Han H. UTR: Unsupervised learning of thickness-insensitive representations for electron microscope image. In: International conference on image processing. IEEE; 2021, p. 155–9. <http://dx.doi.org/10.1109/ICIP42928.2021.9506597>.
- [57] Liu J, Hong B, Chen X, Xie Q, Tang Y, Han H. An Effective AI Integrated System for Neuron Tracing on Anisotropic Electron Microscopy Volume. Biomed Signal Process Control 2021;69:102829. <http://dx.doi.org/10.1016/j.bspc.2021.102829>.
- [58] Funke J, Tschopp F, Grisaitis W, Sheridan A, Singh C, Saalfeld S, et al. Large scale image segmentation with structured loss based deep learning for connectome reconstruction. IEEE Trans Pattern Anal Mach Intell 2018;41(7):1669–80. <http://dx.doi.org/10.1109/TPAMI.2018.2835450>.
- [59] Berman J, Chklovskii DB, Wu J. Bridging the gap: Point clouds for merging neurons in connectomics. In: International conference on medical imaging with deep learning. PMLR; 2022, p. 150–9.
- [60] Quan TM, Hildebrand DGC, Jeong W-K. FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics. Front Comput Sci 2021;3:613981. <http://dx.doi.org/10.3389/fcomp.2021.613981>.
- [61] Li Y, Li W, Chen Q, Huang W, Zou Y, Xiao X, et al. WASPSYN: A challenge for domain adaptive synapse detection in microwasp brain connectomes. IEEE Trans Med Imaging 2024. <http://dx.doi.org/10.1109/TMI.2024.3400276>.
- [62] Lin Z, Wei D, Petkova MD, Wu Y, Ahmed Z, K KS, et al. NucMM dataset: 3D neuronal nuclei instance segmentation at sub-cubic millimeter scale. In: Medical image computing and computer assisted intervention–MICCAI. Springer; 2021, p. 164–74. http://dx.doi.org/10.1007/978-3-030-87193-2_16.

- [63] Januszewski M, Kornfeld J, Li PH, Pope A, Blakely T, Lindsey L, et al. High-precision automated reconstruction of neurons with flood-filling networks. *Nature Methods* 2018;15(8):605–10. <http://dx.doi.org/10.1038/s41592-018-0049-4>.
- [64] Jiang Y, Xiao C, Li L, Chen X, Shen L, Han H. An effective encoder-decoder network for neural cell bodies and cell nucleus segmentation of EM images. In: 41st annual international conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2019, p. 6302–5. <http://dx.doi.org/10.1109/EMBC.2019.8857887>.
- [65] Wei Z, Dan T, Ding J, McCormick C, Kyere FA, Kim M, et al. High throughput deep model of 3D nucleus instance segmentation by stereo stitching contextual gaps. In: 20th international symposium on biomedical imaging. IEEE; 2023, p. 1–5. <http://dx.doi.org/10.1109/ISBI53787.2023.10230745>.
- [66] Shamsi NI, Gjestebj LA, Chavez D, Snyder M, Eastwood BS, Fay MG, et al. Self-supervised learning to improve topology-optimized axon segmentation and centerline detection. In: 20th international symposium on biomedical imaging. IEEE; 2023, p. 1–4. <http://dx.doi.org/10.1109/ISBI53787.2023.10230780>.
- [67] Lee K, Lu R, Luther K, Seung HS. Learning and segmenting dense voxel embeddings for 3D neuron reconstruction. *IEEE Trans Med Imaging* 2021;40(12):3801–11. <http://dx.doi.org/10.1109/TMI.2021.3097826>.
- [68] Bhattacharjee AD, Majumdar D, Dey S, Sarkar S. Advancing connectomics: High-precision neural structure segmentation using U-net-based model. In: 14th annual ubiquitous computing, electronics & mobile communication conference. IEEE; 2023, p. 0420–6. <http://dx.doi.org/10.1109/UEMCON59035.2023.10316141>.
- [69] Archit A, Freckmann L, Nair S, Khalid N, Hilt P, Rajashekar V, et al. Segment anything for microscopy. *Nature Methods* 2025;22(3):579–91. <http://dx.doi.org/10.1038/s41592-024-02580-4>.
- [70] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *Medical image computing and computer-assisted intervention—MICCAI*. Springer; 2015, p. 234–41. http://dx.doi.org/10.1007/978-3-319-24574-4_28.
- [71] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. In: *Advances in neural information processing systems*, vol. 30, Curran Associates, Inc.; 2017, URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
- [72] Boergens KM, Berning M, Bocklisch T, Bräunlein D, Drawitsch F, Frohnhofen J, et al. webKnossos: efficient online 3D data annotation for connectomics. *Nature Methods* 2017;14(7):691–4. <http://dx.doi.org/10.1038/nmeth.4331>.