
Towards Generalized Synapse Detection Across Invertebrate Species

Samia Mohinta^{1,2,*†}
sm2667@cam.ac.uk

Daniel Franco-Barranco^{1,2,3,*}
dfranco@mrc-lmb.cam.ac.uk

Shi Yan Lee^{1,2}
syl58@cam.ac.uk

Albert Cardona^{1,2}
ac2040@cam.ac.uk

¹MRC Laboratory of Molecular Biology, University of Cambridge, UK

²Department of Physiology, Development and Neuroscience, University of Cambridge, UK

³Donostia International Physics Center (DIPC), San Sebastian, Spain

*Equal contribution †Corresponding author

Abstract

Behavioural differences across organisms—whether healthy or pathological—are closely tied to the structure of their neural circuits. Yet, the fine-scale synaptic changes that give rise to these variations remain poorly understood, in part due to persistent challenges in detecting synapses reliably and at scale. Volume electron microscopy (EM) offers the resolution required to capture synaptic architecture, but automated detection remains difficult due to sparse annotations, morphological variability, and cross-dataset domain shifts. To address this, we make three key contributions. First, we curate a diverse EM benchmark spanning four datasets across two invertebrate species: adult and larval *Drosophila melanogaster*, and *Megaphragma viggiani* (micro-WASP). Second, we propose SIMPSYN, a single-stage Residual U-Net trained to predict dual-channel spherical masks around pre- and post-synaptic sites, designed to prioritize training and inference speeds and annotation efficiency over architectural complexity. Third, we benchmark SIMPSYN against Buhmann et al.’s Synful [1], a state-of-the-art multi-task model that jointly infers synaptic pairs. Despite its simplicity, SIMPSYN consistently outperforms Synful in F1-score across all volumes for synaptic site detection. While generalization across datasets remains limited, SIMPSYN achieves competitive performance when trained on the combined cohort. Finally, ablations reveal that simple post-processing strategies—such as local peak detection and distance-based filtering—yield strong performance without complex test-time heuristics. Taken together, our results suggest that lightweight models, when aligned with task structure, offer a practical and scalable solution for synapse detection in large-scale connectomic pipelines.

1 Introduction

The brain’s capacity to perceive, learn, and adapt is fundamentally shaped by the organization of its synaptic connections. Synapses—the junctions where neurons communicate—undergo dynamic structural and functional changes in response to development, experience, and disease [2–4]. Capturing these fine-grained synaptic patterns is essential for understanding both normal brain function and neurological disorders.

Advances in volume electron microscopy (EM) now allow us to image entire brain regions at nanometer resolution. Yet, despite the anatomical richness in these datasets, reliable synapse detection at scale remains a major bottleneck. Synapses are sparse, morphologically diverse, and context-dependent [5], varying across species, brain regions, and developmental stages. Although current models can predict millions of synaptic sites, they often yield a high rate of false positives. Conversely, applying conservative filters—such as removing synapses with low connection counts [6] or based on cleft detection [1]—can lead to significant false negatives. These trade-offs underscore a deeper issue: the lack of a unified, generalizable detection framework that can robustly operate across biological and imaging variability without dataset-specific tuning.

Most existing synapse detection methods are developed and validated within a single dataset — trained and tested on anatomically similar volumes. These models often rely on densely annotated voxel-level labels such as synaptic clefts [7] or segmented neurites [8, 9], which are tedious to generate and tied to specific imaging conditions. As a result, their performance degrades significantly when applied to data from different species, brain regions, or acquisition modalities. Recent methods like Synful [1] represent meaningful progress by predicting synaptic pairs from sparse point annotations; however, they still require dataset-specific tuning and come with non-trivial computational overhead, which can limit their generalization to large or heterogeneous datasets.

To address this, we propose SIMPSYN: a simple yet effective deep learning model with the promise for generalized synapse detection. SIMPSYN forgoes complex multi-task learning or test-time heuristics, instead relies on a single-stage Residual 3D U-Net that predicts spherical masks around pre- and post-synaptic sites. It is trained solely on point annotations and paired with lightweight post-processing to infer synaptic connectivity. This task-aligned design significantly reduces annotation burden and computational cost, while still delivering competitive performance across diverse datasets.

We summarize our main contributions as follows:

- **A cross-species synapse detection benchmark**, spanning 16 volumes (in 4 datasets; Figure 1a) across adult and larval *Drosophila melanogaster* and *Megaphragma viggiani* (micro-wasp), including a new manually annotated larval fly dataset (Octo).
- **SIMPSYN**, a lightweight, task-aligned model for synapse detection that uses only point annotations and requires no manual or per-dataset test-time tuning.
- **Empirical evidence** that SIMPSYN outperforms or matches Synful in F1-score across all datasets, both in-distribution and out-of-distribution, despite being $15\times$ smaller.
- **Ablation studies** demonstrating that simple post-processing techniques (e.g., local peak detection, distance-based filtering) are sufficient to yield strong performance—highlighting that model simplicity need not come at the cost of accuracy.

2 Related work

To situate our work within the broader landscape of synapse detection approaches in EM, we first examine the morphological and imaging challenges that underlie synapse detection across species, before reviewing recent methodological advances and limitations in automated detection.

2.1 Synapse morphology across species

Vertebrate synapses are reliably identified in transmission EM by distinct ultrastructural features—synaptic clefts, vesicle clusters, and active zones [10]. These are well-resolved in anisotropic datasets such as MICrONS [11–13], which trade z-resolution for high x–y coverage (≤ 4 nm/pixel). Combined with predominantly monadic (1-to-1) connectivity, this enables straightforward partner assignment and supports supervised detection pipelines.

Invertebrate systems present greater challenges. Data are typically acquired via FIB-SEM (Focused Ion Beam Scanning Electron Microscope), which provides isotropic resolution (e.g., $8\times 8\times 8$ nm) but reduces x–y clarity, obscuring key features like synaptic clefts and PSDs. These structures are often faint or noisy, complicating both manual labelling and model training.

Insect synapses are structurally more complex, with polyadic connections where a single T-bar contacts multiple postsynaptic partners. As shown in Figure 1, this complexity varies across FIB-

SEM datasets (Hemibrain, Octo, WASP, MANC), developmental stages, and brain regions. WASP and MANC exhibit more frequent 1-to-5+ connections and greater EM variability (Figures 1b–e), complicating consistent detection. Structural similarity analyses (Figure 1e) reveal larger intra- and inter-dataset shifts. While SSIM captures these morphological changes; cosine similarity is more sensitive to intensity, reflecting contrast-driven resemblance. Both metrics are computed from synaptic patches (32^3), underscoring anatomical variability that can hinder model generalization.

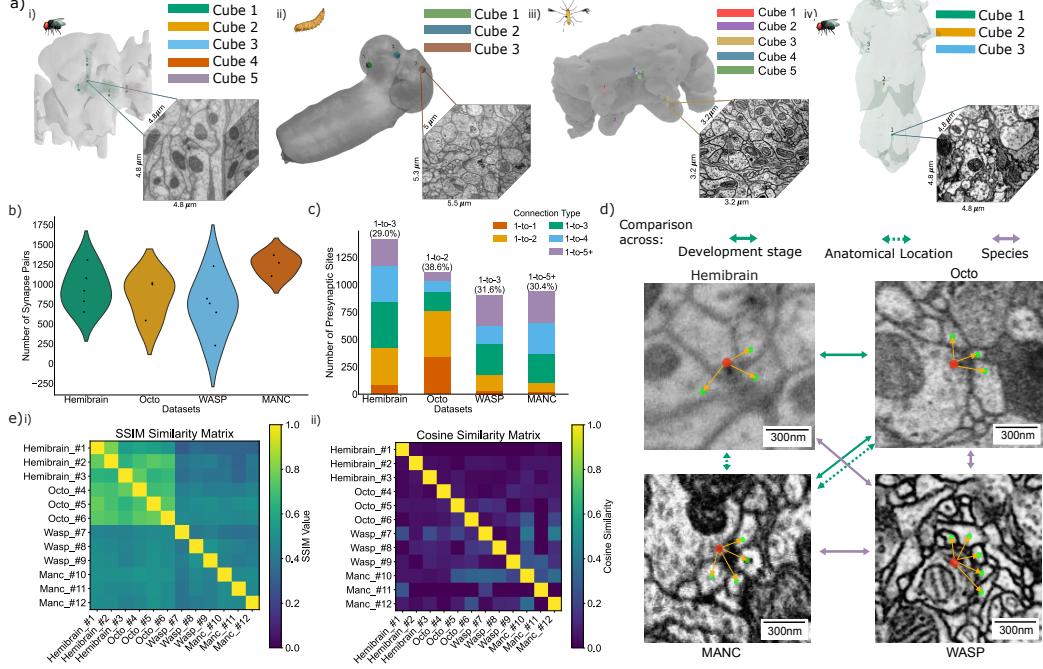


Figure 1: Diversity in EM structure and synaptic organization across species. a) Spatial distribution of annotated sub-volumes in Hemibrain, Octo, WASP, and MANC with example EM patches showing ultrastructural variability. b) Synaptic pair counts per sub-volume. c) Presynaptic connection types (1-to-1 to 1-to-5), highlighting Octo’s preference for low-degree connections versus polyadic dominance in others. d) Representative polyadic synapses across datasets, organized by developmental stage, region, and species. e) Structural similarity analysis. Higher values (yellow) denote greater morphological similarity: i) SSIM reveals higher intra (in-distribution)-/inter(out-of-distribution)-dataset consistency in Hemibrain/Octo and ii) cosine similarity suggests moderate overlaps in WASP and MANC. Scale bars: 300 nm.

2.2 Synapse labeling and recent approaches

Automated synapse detection in volume EM has historically been tied to dense, voxel-level annotations—including neurite skeletons [8], full segmentations [9, 14, 15], clefths [7, 11, 16, 17, 15], vesicles [15], and finely traced membranes [18]. For example, SynEM [9] requires full neurite segmentation before classifying interfaces as synaptic, while Heinrich et al. [7] regress signed-distance maps to cleft masks using a 3D U-Net. These pipelines demonstrate strong performance within their target domains but are tightly coupled to imaging resolution and annotation schemas. As a result, they require extensive re-annotation or tuning to adapt across datasets, limiting their scalability and generalization in comparative or cross-species connectomics.

More recent work has reduced the annotation burden by learning from sparse synaptic labels. In particular, Buhmann et al.[19], formulated synaptic partner prediction as an edge-classification task: a 3D U-Net is trained on long-range ‘edges’ linking pre- and post-synaptic neurons, using only point annotations at presynaptic and postsynaptic sites. Overall, the field is increasingly shifting towards the use of this type of annotation for synapse detection, as it is significantly more cost-effective [19, 6, 20, 21].

Buhmann et al.’s Synful approach [1] is widely recognized as the state-of-the-art method for synapse detection in large-scale invertebrate EM datasets. In recent years, several landmark projects have relied on Synful for synapse identification, including the reconstruction motor control circuits [22], and the ventral nerve cord [23] of *Drosophila melanogaster*, as well as the complete connectome of the adult fruit fly [6, 24, 25]. Additionally, Synful has been adapted to map the feedforward connectivity within the mouse cerebellar cortex [26].

2.3 Generalization challenge

Synapse detection is inherently a challenging problem in computational neuroscience, as noted by Buhmann et al.[19]. As discussed in 2.1, interspecies variability adds another layer of complexity to this problem. Many approaches that report high performance on synapse detection typically evaluate their models on data drawn from the same domain as the training set. Even when applied to different brain regions, these methods often assume consistency in staining protocols and tissue characteristics [9, 8, 14, 18, 27]. Existing methods frequently rely on the manual tuning of multiple threshold parameters [18], or report optimal performances selected from precision-recall curves—both of which presuppose the availability of ground truth annotations. This reliance poses a significant barrier to the automated deployment of these models across diverse datasets, especially in scenarios where such annotations are unavailable. There is a general lack of guidance within the field regarding how existing synapse detection approaches can be effectively adapted to novel datasets.

In response to these challenges, the WASPSYN23 competition [20] was recently introduced to foster the development of deep learning methodologies capable of generalizing across domains, with a particular emphasis on advancing synapse segmentation under varying biological and imaging conditions. Taking inspiration, our work avoids reliance on manually optimized thresholds and extreme hyperparameter tuning, and is tested on four different datasets across 16 non-overlapping brain regions. We introduce automated strategies for threshold selection, enabling the model to generalize to previously unseen data without the need for dataset-specific adjustments.

3 Materials and Methods

3.1 Dataset preparation

To rigorously evaluate our synapse detection methods, we use four heterogeneous datasets spanning two species: three from *Drosophila melanogaster* (adult: Hemibrain and MANC; larval: Octo in this study) and one from *Megaphragma viggiani* (WASP). Except for Octo, all datasets are publicly available. Curated sub-volumes from public datasets contain high-confidence annotations, primarily derived from machine predictions. In contrast, all synapse annotations in Octo were manually labelled.

Hemibrain The Hemibrain dataset [28] is a FIB-SEM volume of the adult *Drosophila melanogaster* central brain at $8 \times 8 \times 8$ nm resolution, featuring $\sim 25,000$ neurons and 20M+ synapses. We extract five 600^3 voxel subvolumes ($4.8 \mu\text{m}$ per side); four are used for training, one for testing.

Octo Octo is a FIB-SEM volume of the L1 larval *Drosophila melanogaster* central nervous system (brain + VNC) at $8 \times 8 \times 8$ nm resolution. Not yet public, it includes $\sim 2,500$ manually annotated synapses across three non-overlapping $25 \mu\text{m}^3$ sub-volumes from distinct brain regions, where two are used for training and one for testing.

MANC The MANC dataset [29] captures the ventral nerve cord of a 5-day-old adult male *Drosophila melanogaster* at $8 \times 8 \times 8$ nm resolution, with $\sim 10\text{M}$ presynaptic (T-bars) and 74M postsynaptic sites annotated across $\sim 23,000$ neurons. We use three 600^3 voxel sub-volumes with two for training and one for testing.

WASP The dataset of the WASPSYN23 challenge [20] consists of 14 FIB-SEM volumes (each 416^3 voxels, $8 \times 8 \times 8$ nm) from *Megaphragma viggiani*, covering diverse brain regions across three specimens. We select five volumes from one brain (specimen 3), and use four for training and one for testing.

3.2 Methods

3.2.1 SIMPSYN

SIMPSYN is a 3D Residual U-Net [30] for synapse detection that predicts two output channels: one corresponding to pre-synaptic regions and the other to post-synaptic regions as done in [20, 21]. To achieve this, a spherical 3D region is generated and centered at the coordinates of the pre-synaptic and post-synaptic points, following the work by Zhou et. al. [31] and in detection workflow in Franco-Barranco et al. [32]. These output masks are subsequently processed using connected component labelling to isolate individual synaptic structures. To establish correspondence between pre-synaptic and post-synaptic sites, each post-synaptic component is paired with its nearest pre-synaptic counterpart based on the nearest neighbour criterion. Figure 2 illustrates the complete pipeline of the SIMPSYN model.

3.2.2 Synful

Synful [1] is a 3D convolutional neural network designed for synapse detection from sparsely annotated pre- and postsynaptic sites. Synapses are represented as directed voxel pairs $(s, t) \in \Omega^2$, with s and t denoting presynaptic and postsynaptic voxels, respectively. This formulation supports polyadic synapses by modelling each connection independently. We use the MT1 architecture, comprising a single 3D U-Net with shared encoder-decoder pathways and dual output heads for jointly predicting postsynaptic masks and 3D direction vector fields, as illustrated in the Figure S2. At inference time, postsynaptic candidates are identified by thresholding the predicted mask and grouping connected components. Final postsynaptic locations \hat{t} are selected using the peak of a Euclidean distance transform over each component, and corresponding presynaptic sites are computed as $\hat{s} = \hat{t} + \hat{d}(\hat{t})$, forming the predicted pair (\hat{s}, \hat{t}) .

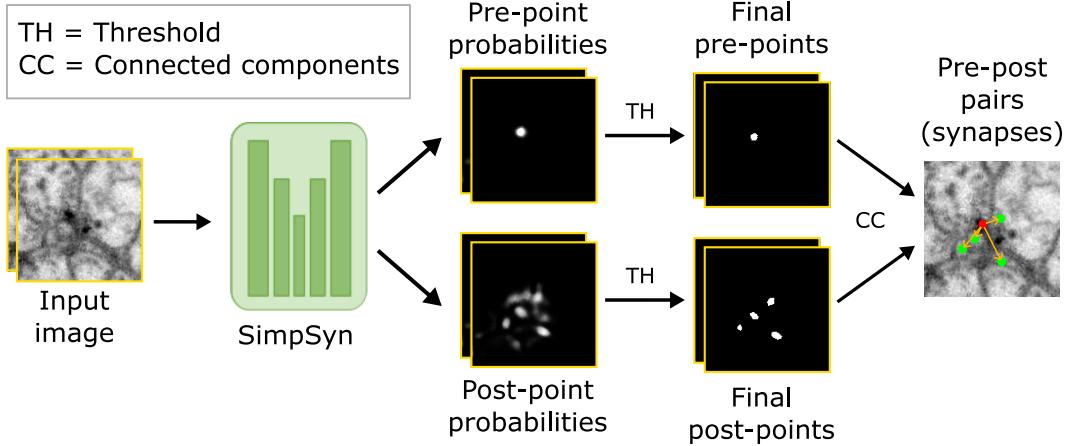


Figure 2: **General overview of SIMPSYN (our Simple Synapse Detection model).** SIMPSYN is a 3D Residual U-Net trained to predict 3D spherical masks of the pre and post synaptic sites. The segmented pre-post masks are paired to form pre-post synapses by using connected components and nearest neighbour pairing.

4 Experimental setup

4.1 Implementation details

4.1.1 SIMPSYN

Our SIMPSYN approach is fully integrated within an established deep-learning platform for bio-image analysis, Biapy [32], and is implemented using Pytorch version 2.4. During training, we employ standard data augmentation techniques for EM data [33, 34], including flips, random rotations, brightness and contrast adjustments, and elastic deformations. We set aside a randomly selected 10% of the training synaptic sites for validation, and this split is held constant across all random seeds for a

given model configuration. SIMPSYN is trained using cross-entropy loss with foreground-background re-weighting based on pixel counts, a cosine-decay learning rate scheduler with warm-up [35], and the AdamW optimizer. Training is performed for 1500 epochs on eight RTX 3090 24GB GPUs for ~ 15 hours, with early stopping using a patience of 250 epochs. Inference takes ~ 5 minutes. See Table S3 for a detailed overview of the hyperparameter search.

4.1.2 Synful

Our Synful baseline is based on the publicly available source code, which we refactored and encapsulated within a Dockerized environment for reproducibility [36]. Synful natively relies on TensorFlow version 1.x and trained on 128³ voxel cubes for 300,000 iterations. Validation is performed as a post-hoc process to select the checkpoint yielding the best performance, in order to reduce computational cost and time of training. The loss function combines binary cross-entropy (for postsynaptic masks) and mean squared error (for presynaptic direction vectors), and is optimized using Adam. We apply data augmentation strategies similar to those in Buhmann et al. [1], including random rotations, intensity scaling, and elastic deformations. Training and inference are conducted on a single 12GB Titan XP GPU, with training taking approximately 36 hours and inference requiring ~ 12 minutes per $5\mu\text{m} \times 5\mu\text{m} \times 5\mu\text{m}$ volume.

4.2 Evaluation metrics

We adopt the evaluation metrics used in the WASPSYN23 challenge [20], where synapse detection accuracy is assessed via bipartite matching to minimize Euclidean distance between predicted and ground-truth synapses [37]. Let D and G denote the sets of detected and ground-truth synapses, respectively. A matching $f : D \rightarrow G$ is computed using the Hungarian algorithm to minimize:

$$\sum_{d \in D} C(d, f(d)), \quad (1)$$

where $C(\cdot)$ is the Euclidean distance between matched pairs. A detection is considered a true positive if the distance is below a 120-voxel threshold (corresponding to $8 \times 8 \times 8$ nm resolution). The F1-score is then computed as:

$$F_1 = \frac{2 \cdot TP}{2TP + FP + FN}, \quad (2)$$

where TP , FP , and FN refer to the number of true positives, false positives, and false negatives, respectively. The F1-score thus captures the harmonic mean of precision and recall, offering a balanced measure of detection accuracy. Separate scores are reported for pre- and post-synaptic sites in Table 1.

4.3 Results and Discussion

We evaluate SIMPSYN against the Synful baseline across various training and testing configurations. Table 1 reports F1-scores for pre- and post-synaptic site detection across four EM datasets. Each cell shows the performance of SIMPSYN (top) and Synful (bottom); diagonal entries (shaded) represent in-distribution (ID) performance, while off-diagonal entries reflect out-of-distribution (OOD) generalization. For each test domain, the best and second-best performing configurations are highlighted. See Supplementary Table S1 for a detailed synaptic pairwise comparison using F1 scores across all model configurations.

In-distribution performance. SIMPSYN consistently outperforms Synful in all ID scenarios for both pre- and post-synaptic detection. On Hemibrain, SIMPSYN achieves 0.783 (pre) and 0.606 (post) F1-scores—substantially higher than Synful’s 0.262 and 0.437, respectively. Similarly, Octo and MANC models achieve strong ID performance, particularly in presynaptic detection (e.g., 0.538 on Octo vs. Synful’s 0.150). These results suggest that SIMPSYN handles intra-dataset variability and polyadic structures robustly, despite its architectural simplicity.

Out-of-distribution generalization. Generalization to unseen domains remains a challenging setting. Dataset-specific models show varying degrees of degradation under OOD shift. For example, a Hemibrain-trained model performs poorly on WASP (pre: 0.000, post: 0.000), likely due to

Table 1: **Pre- and post-synaptic site detection results** measured using F1 metric (higher is better). The upper row denotes the test datasets, while the first column indicates the training dataset. Consequently, diagonal entries (in gray) correspond to in-distribution (ID) evaluations, whereas off-diagonal entries represent out-of-distribution (OOD) scenarios. Within each cell, the upper value refers to performance achieved by SIMPSYN, and the lower value to that of Synful. For each column, the highest score is highlighted in **bold**, and the second-highest is underlined. SIMPSYN consistently achieves the highest F1-scores across all volumes. In nearly all cases, models trained on combined datasets outperform those trained solely on dataset-specific ground truth, highlighting the benefits of cross-domain learning.

		Pre synaptic sites					Post synaptic sites				
Train	Test	Hemibrain	Octo	WASP	MANC	All	Hemibrain	Octo	WASP	MANC	All
Hemibrain	0.783	0.021	0.000	0.007	0.203		0.606	0.030	0.000	0.029	0.166
	0.262	0.042	0.000	0.009	0.078		0.437	0.025	0.000	0.013	0.119
Octo	0.523	0.538	0.036	0.156	0.313		0.139	0.425	0.005	0.149	0.179
	0.017	0.150	0.012	0.021	0.046		0.004	0.171	0.002	0.022	0.050
WASP	0.057	0.400	0.781	0.194	0.358		0.000	0.182	0.655	0.300	0.284
	0.000	0.000	0.292	0.148	0.110		0.000	0.000	0.612	0.321	0.233
MANC	0.647	0.290	0.000	0.645	0.395		0.028	0.300	0.000	0.478	0.202
	0.009	0.005	0.000	0.201	0.054		0.018	0.002	0.000	0.364	0.096
All	0.846	0.654	0.810	0.736	0.762		0.635	0.506	0.664	<u>0.475</u>	0.570
	0.348	0.320	0.391	0.349	0.352		0.248	0.343	0.582	0.418	0.398

differences in tissue contrast and synapse morphology. However, models trained on Octo or MANC overall transfer better to other domains, likely due to their structural diversity. Notably, the ‘All’ model—trained jointly on all available training volumes achieves the highest average generalization performance across test domains, with strong pre-synaptic F1-scores (0.846, 0.654, 0.810, 0.736) and competitive post-synaptic results (0.635, 0.506, 0.664, 0.475). These results underscore the importance of anatomical and morphological diversity in training data.

Asymmetries in detection. Across datasets, we observe a general asymmetry between pre- and post-synaptic site performance. Presynaptic sites tend to be more reliably detected, while postsynaptic sites—especially in polyadic arrangements—show greater variability. This is visible both quantitatively in Table 1 and qualitatively in Figure 3. In regions with complex connectivity (e.g., WASP or MANC), SIMPSYN predictions occasionally include spurious post sites or miss diffuse ones. This reflects both intrinsic biological ambiguity and architectural limitations in point-pair association.

Qualitative evaluation. Figure 3 shows visual comparisons of SIMPSYN predictions across datasets. True positives (green), false negatives (red), and false positives (blue) reveal several patterns. Hemibrain samples show relatively clean detections, while Octo and WASP images highlight the challenge of distinguishing faint synaptic sites from surrounding structures. Despite this, SIMPSYN captures a large proportion of true sites without over-saturating predictions. Overall, these results confirm that SIMPSYN, despite being significantly simpler and smaller than Synful, offers competitive or superior accuracy across diverse domains. Its robustness across both ID and OOD conditions (when trained with representative data), with minimal tuning and reduced computational overhead, makes it a practical candidate for large-scale connectomics pipelines.

4.3.1 Ablation Study of Post-processing Strategies

We conducted a set of ablations to assess how different post-processing strategies affect final detection performance. Table 2 summarizes results under variations in point creation, thresholding, and filtering.

Peak detection strategy: As shown in Table 2a, we compared *peak_local_max* and *blob_log* from `scikit-image` [38]. The former, which uses a straightforward local maximum criterion, outperforms *blob_log* in both precision and recall, while also requiring fewer hyperparameters and less compute. This method aligns well with our model’s spherical output and minimizes false mergers.

Thresholding methods: We evaluated manual, automatic, and relative thresholding strategies (Table 2b). Relative thresholding consistently yielded the best results (pre: 0.762, post: 0.570),

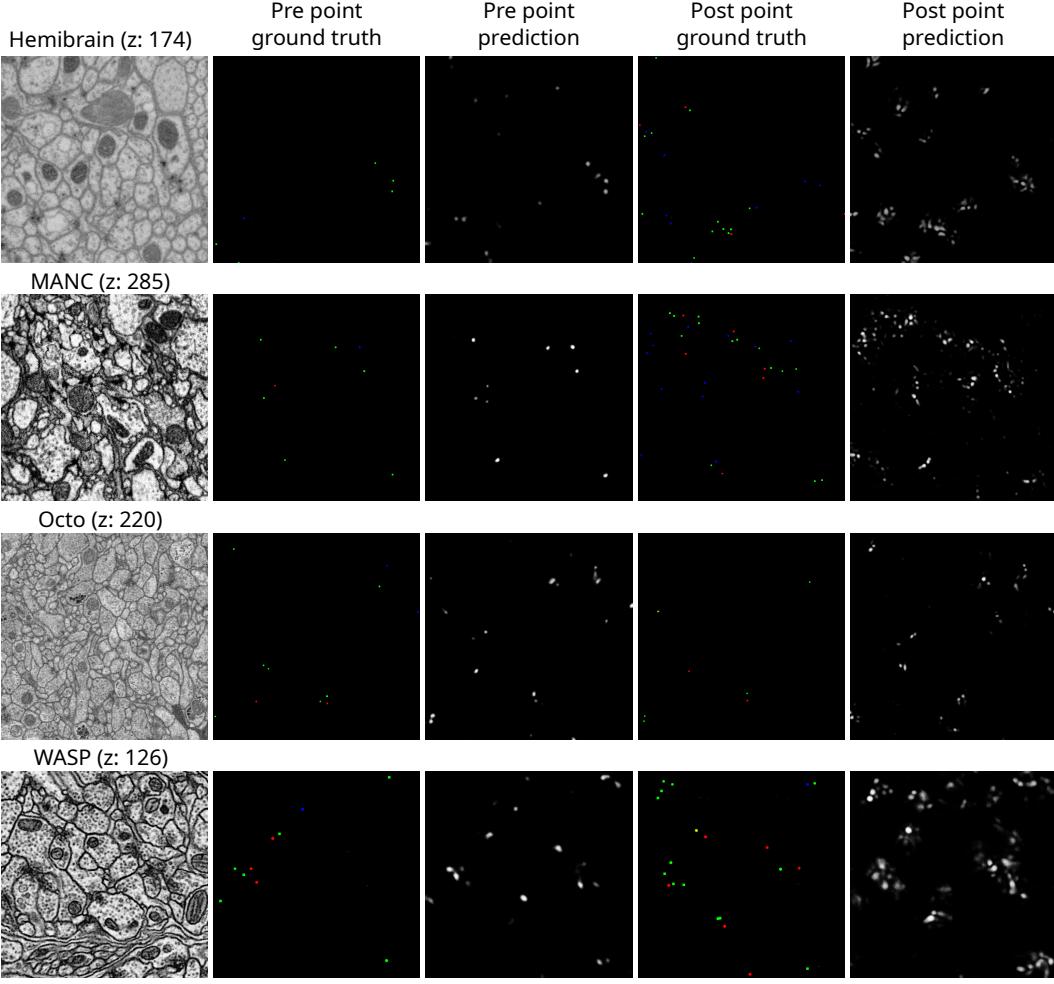


Figure 3: **Qualitative results SIMPSYN on test samples.** Correctly predicted point (true positives) are marked in green, un-detected points (false negatives) in red and spurious points (false positives) are in blue.

suggesting that global or per-batch intensity heuristics adapt well across volumes without requiring dataset-specific tuning.

Distance-based filtering: Synaptic sites predicted in close spatial proximity are often biologically implausible. Applying distance-based filtering improved precision, particularly for presynaptic sites (Table 2c). Further constraining predictions using both distance and predicted masks reduced false positives, but also suppressed some valid detections—highlighting the trade-off between specificity and sensitivity.

These ablations reinforce a key design principle of SIMPSYN: model simplicity can be complemented by targeted post-processing to achieve strong performance without relying on complex test-time inference pipelines. Moreover, when trained on anatomically diverse datasets, SIMPSYN enables the discovery of generalizable post-processing heuristics that remain effective across invertebrate model organisms (see Supplementary Figure S1).

5 Limitations

While SIMPSYN demonstrates strong in-distribution generalization and computational efficiency, several important limitations remain.

Table 2: **SimpSyn ablation experiments.** We conduct evaluations across all datasets and report the F1-scores for both pre- and post-synaptic site detection. Highest scores, and the default settings, are highlighted in **bold**.

(a) **Point creation method.** We explore two different approaches available in the scikit-image library [38].

Method	Pre ↑	Post ↑
<i>peak_local_max</i>	0.762	0.570
<i>blob_log</i>	0.675	0.482

(b) **Thresholding method.** An automatic way of thresholding can improve overall F1.

Thresholding	Pre ↑	Post ↑
manual	0.664	0.464
auto	0.672	0.432
relative	0.762	0.570
relative (batch)	0.717	0.533

(c) **Point filtering.** Excluding detected pre synaptic sites that are in close proximity is justified, as such configurations are biologically implausible.

Point filtering	Pre ↑	Post ↑
No-filter	0.682	0.570
By distance	0.762	0.558
By distance + mask	0.717	0.555

First, its out-of-distribution (OOD) generalization is sensitive to large domain shifts, particularly when models are trained on a single dataset. Variability in anatomical structure, polyadicity, and EM contrast continues to challenge detection robustness. Nonetheless, the model trained on the full cohort of datasets ('All') generalizes considerably better across held-out volumes, particularly for presynaptic sites. The lower performance observed for postsynaptic site detection can be partially attributed to the variability in synaptic partner distributions and configurations across different sub-volumes.

Second, although we report mean F1-scores across datasets, we do not currently include statistical confidence intervals or measures of variance. All SIMPSYN models were trained with multiple random seeds, and we observed minimal performance fluctuations and is therefore not shown. However, a more rigorous treatment of variance and significance testing would strengthen future benchmarking, especially when evaluating small performance margins.

Third, we use nearest-neighbour strategy to associate a pre-synaptic site with multiple nearby post-synaptic partners in a polyadic synapse setting. This can lead to spurious post-site assignments—particularly in densely innervated regions, where spatial proximity alone is an unreliable indicator of true connectivity. As explicitly modelling polyadicity remains difficult [1], future work could explore contrastive objectives or uncertainty-aware assignment strategies to disambiguate true partners from nearby distractors.

Fourth, while SIMPSYN is trained solely on sparse point annotations, offering clear efficiency advantages, it still depends on high-quality, manually curated ground truth. However, even expert annotators can introduce labelling inconsistencies, which can skew both model training and evaluation (see supplementary Fig S3). Addressing this issue may require human-in-the-loop proofreading, uncertainty-guided annotation, or robust objectives that account for label noise and incompleteness, especially in under-represented brain regions or species.

Finally, like other state-of-the-art approaches, SIMPSYN is susceptible to spurious predictions. Given the minute scale of synaptic junctions and their visual similarity to nearby organelles, distinguishing true synapses from noise remains a non-trivial task. Incorporating contrastive training objectives or hard negative mining strategies may help improve specificity by explicitly modelling context and suppressing false positives.

6 Conclusions and Future Work

Accurate and generalizable synapse detection is a critical component of large-scale connectomic mapping, key for advancing our understanding of how structural connectivity underlies neural function in health and in disease. Here, we introduced SIMPSYN, a lightweight, point-supervised

model for synapse detection, and evaluated it across 16 connectomic volumes spanning multiple species and developmental stages. SIMPSYN matches or surpasses a state-of-the-art baseline in both in-distribution and out-of-distribution scenarios. Ablation studies show that simple, biologically informed post-processing strategies—particularly relative thresholding and distance-based filtering—can substantially boost detection performance. Nevertheless, challenges remain in achieving robust out-of-distribution performance, particularly for postsynaptic site detection, due to polyadic variability and the lack of consistently prominent morphological features on the postsynaptic side. Future work will focus on enhancing pairing mechanisms and investigating label-denoising (uncertainty-aware) and contrastive learning strategies to further improve generalization within SIMPSYN’s framework. These directions, combined with SIMPSYN’s low computational overhead and reliance on sparse annotations, position SIMPSYN as a promising tool for scalable and accessible connectome mapping pipelines.

7 Acknowledgments

We thank Nicolo Ceffa (MRC Laboratory of Molecular Biology) for assistance with Blender rendering of mesh figures. We are grateful to Ana Correia, Amina Dulac, Michael Clayton, and Samuel Harris for proofreading synapses in the Octo dataset.

8 Funding

This project was funded by a Wellcome Trust Investigator award to AC (Ref: 205038/Z/16/Z) and the MRC LMB core funding.

9 Author contributions

The author contributions are as follows:

Conceptualization: SM and DFB; Methodology: SM and DFB; Investigation: SM and DFB; Visualization: SM, DFB, SYL and AC; Supervision: AC; Writing—original draft: SM and DFB; Writing—review and editing: SM, DFB, SYL, and AC

10 Competing Interests

The authors declare no competing interests.

11 Data and Materials Availability

All data needed to evaluate the conclusions are provided in the paper and/or the Supplementary Materials. Curated EM datasets will be released on Zenodo (DOIs to be added upon publication). Source code is available at Catena and BiaPy.

References

- [1] Julia Buhmann, Arlo Sheridan, Caroline Malin-Mayor, Philipp Schlegel, Stephan Gerhard, Tom Kazimiers, Renate Krause, Tri M. Nguyen, Larissa Heinrich, Wei-Chung Allen Lee, Rachel Wilson, Stephan Saalfeld, Gregory S. X. E. Jefferis, Davi D. Bock, Srinivas C. Turaga, Matthew Cook, and Jan Funke. Automatic detection of synaptic partners in a whole-brain *Drosophila* electron microscopy data set. *Nature Methods*, 18(7):771–774, 7 2021.
- [2] Michael Winding, Benjamin D. Pedigo, Christopher L. Barnes, Heather G. Patsolic, Youngser Park, Tom Kazimiers, Akira Fushiki, Ingrid V. Andrade, Avinash Khandelwal, Javier Valdes-Aleman, Feng Li, Nadine Randel, Elizabeth Barsotti, Ana Correia, Richard D. Fetter, Volker Hartenstein, Carey E. Priebe, Joshua T. Vogelstein, Albert Cardona, and Marta Zlatic. The connectome of an insect brain. *Science*, 379(6636), 3 2023.

- [3] J.R. Wolff and M. Missler. Synaptic reorganization in developing and adult nervous systems. *Annals of Anatomy - Anatomischer Anzeiger*, 174(5):393–403, 10 1992.
- [4] Katarzyna Lepeta, Mychael V. Lourenco, Barbara C. Schweitzer, Pamela V. Martino Adami, Priyanjalee Banerjee, Silvina Catuara-Solarz, Mario de La Fuente Revenga, Alain Marc Guillem, Mouna Haidar, Omamuyowwi M. Ijomone, Bettina Nadorp, Lin Qi, Nirma D. Perera, Louise K. Refsgaard, Kimberley M. Reid, Mariam Sabbar, Arghyadip Sahoo, Natascha Schaefer, Rebecca K. Sheean, Anna Suska, Rajkumar Verma, Cinzia Vicidomini, Dean Wright, Xing-Ding Zhang, and Constanze Seidenbecher. Synaptopathies: synaptic dysfunction in neurological disorders – A review from students to students. *Journal of Neurochemistry*, 138(6):785–805, 9 2016.
- [5] Fang-Shuai Shen, Cui Liu, Hui-Zhe Sun, Xin-Yi Chen, Yan Xue, and Lei Chen. Emerging evidence of context-dependent synapse elimination by phagocytes in the CNS. *Journal of Leukocyte Biology*, 116(3):511–522, 9 2024.
- [6] Sven Dorkenwald, Arie Matsliah, Amy R Sterling, Philipp Schlegel, Szi-Chieh Yu, Claire E McKellar, Albert Lin, Marta Costa, Katharina Eichler, Yijie Yin, et al. Neuronal wiring diagram of an adult brain. *Nature*, 634(8032):124–138, 2024.
- [7] Larissa Heinrich, Jan Funke, Constantin Pape, Juan Nunez-Iglesias, and Stephan Saalfeld. Synaptic Cleft Segmentation in Non-Isotropic Volume Electron Microscopy of the Complete Drosophila Brain. 5 2018.
- [8] Sven Dorkenwald, Philipp J Schubert, Marius F Killinger, Gregor Urban, Shawn Mikula, Fabian Svara, and Joergen Kornfeld. Automated synaptic connectivity inference for volume electron microscopy. *Nature methods*, 14(4):435–442, 2017.
- [9] Benedikt Staffler, Manuel Berning, Kevin M Boergens, Anjali Gour, Patrick van der Smagt, and Moritz Helmstaedter. SynEM, automated synapse detection for connectomics. *eLife*, 6, 7 2017.
- [10] Vignesh Jagadeesh, James Anderson, Bryan Jones, Robert Marc, Steven Fisher, and B.S. Manjunath. Synapse classification and localization in Electron Micrographs. *Pattern Recognition Letters*, 43:17–24, 7 2014.
- [11] J. Alexander Bae, Mahaly Baptiste, Caitlyn A. Bishop, Agnes L. Bodor, Derrick Brittain, JoAnn Buchanan, Daniel J. Bumbarger, Manuel A. Castro, Brendan Celii, Erick Cobos, Forrest Collman, Nuno Maçarico da Costa, Sven Dorkenwald, Leila Elabbady, Paul G. Fahey, Tim Fliss, Emmanouil Froudarakis, Jay Gager, Clare Gamlin, William Gray-Roncal, Akhilesh Halageri, James Hebditch, Zhen Jia, Emily Joyce, Justin Joyce, Chris Jordan, Daniel Kapner, Nico Kemnitz, Sam Kinn, Lindsey M. Kitchell, Selden Koolman, Kai Kuehner, Kisuk Lee, Kai Li, Ran Lu, Thomas Macrina, Gayathri Mahalingam, Jordan Matelsky, Sarah McReynolds, Elanine Miranda, Eric Mitchell, Shanka Subhra Mondal, Merlin Moore, Shang Mu, Taliah Muhammad, Barak Nehoran, Oluwaseun Ogedengbe, Christos Papadopoulos, Stelios Papadopoulos, Saumil Patel, Xaq Pitkow, Sergiy Popovych, Anthony Ramos, R. Clay Reid, Jacob Reimer, Patricia K. Rivlin, Victoria Rose, Casey M. Schneider-Mizell, H. Sebastian Seung, Ben Silverman, William Silversmith, Amy Sterling, Fabian H. Sinz, Cameron L. Smith, Shelby Suckow, Marc Takeno, Zheng H. Tan, Andreas S. Tolias, Russel Torres, Nicholas L. Turner, Edgar Y. Walker, Tianyu Wang, Adrian Wanner, Brock A. Wester, Grace Williams, Sarah Williams, Kyle Willie, Ryan Willie, William Wong, Jingpeng Wu, Chris Xu, Runzhe Yang, Dimitri Yatsenko, Fei Ye, Wenjing Yin, Rob Young, Szi-chieh Yu, Daniel Xenes, and Chi Zhang. Functional connectomics spanning multiple areas of mouse visual cortex, 7 2021.
- [12] Narayanan Kasthuri, Kenneth Jeffrey Hayworth, Daniel Raimund Berger, Richard Lee Schalek, José Angel Conchello, Seymour Knowles-Barley, Dongil Lee, Amelio Vázquez-Reina, Verena Kaynig, Thouis Raymond Jones, Mike Roberts, Josh Lyskowski Morgan, Juan Carlos Tapia, H. Sebastian Seung, William Gray Roncal, Joshua Tzvi Vogelstein, Randal Burns, Daniel Lewis Sussman, Carey Eldin Priebe, Hanspeter Pfister, and Jeff William Lichtman. Saturated Reconstruction of a Volume of Neocortex. *Cell*, 162(3):648–661, 7 2015.
- [13] Alexander Shapson-Coe, Michał Januszewski, Daniel R. Berger, Art Pope, Yuelong Wu, Tim Blakely, Richard L. Schalek, Peter H. Li, Shuhong Wang, Jeremy Maitin-Shepard, Neha

- Karlupia, Sven Dorkenwald, Evelina Sjostedt, Laramie Leavitt, Dongil Lee, Luke Bailey, Angerica Fitzmaurice, Rohin Kar, Benjamin Field, Hank Wu, Julian Wagner-Carena, David Aley, Joanna Lau, Zudi Lin, Donglai Wei, Hanspeter Pfister, Adi Peleg, Viren Jain, and Jeff W. Lichtman. A connectomic study of a petascale fragment of human cerebral cortex, 5 2021.
- [14] Changjoo Park, Jawon Gim, Sungjin Lee, Kea Joo Lee, and Jinseop S Kim. Automated synapse detection method for cerebellar connectomics. *Frontiers in Neuroanatomy*, 16:760279, 2022.
 - [15] Fabian Svara, Dominique Förster, Fumi Kubo, Michał Januszewski, Marco Dal Maschio, Philipp J Schubert, Jörgen Kornfeld, Adrian A Wanner, Eva Laurell, Winfried Denk, et al. Automated synapse-level reconstruction of neural circuits in the larval zebrafish brain. *Nature Methods*, 19(11):1357–1366, 2022.
 - [16] Nicholas L Turner, Kisuk Lee, Ran Lu, Jingpeng Wu, Dodam Ih, and H Sebastian Seung. Synaptic partner assignment using attentional voxel association networks. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2020.
 - [17] Casey M Schneider-Mizell, Agnes L Bodor, Derrick Brittain, JoAnn Buchanan, Daniel J Bumbarger, Leila Elabbady, Clare Gamlin, Daniel Kapner, Sam Kinn, Gayathri Mahalingam, et al. Inhibitory specificity from a connectomic census of mouse visual cortex. *Nature*, 640(8058):448–458, 2025.
 - [18] Chi Xiao, Weifu Li, Hao Deng, Xi Chen, Yang Yang, Qiwei Xie, and Hua Han. Effective automated pipeline for 3D reconstruction of synapses based on deep learning. *BMC bioinformatics*, 19:1–18, 2018.
 - [19] Julia Buhmann, Renate Krause, Rodrigo Ceballos Lentini, Nils Eckstein, Matthew Cook, Srinivas Turaga, and Jan Funke. Synaptic partner prediction from point annotations in insect brains. 6 2018.
 - [20] Yicong Li, Wanhu Li, Qi Chen, Wei Huang, Yuda Zou, Xin Xiao, Kazunori Shinomiya, Pat Gunn, Nishika Gupta, Alexey Polilov, et al. WASPSYN: A Challenge for Domain Adaptive Synapse Detection in Microwasp Brain Connectomes. *IEEE Transactions on Medical Imaging*, 2024.
 - [21] Qi Chen, Wei Huang, Yueyi Zhang, and Zhiwei Xiong. Domain adaptive synapse detection with weak point annotations. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2024.
 - [22] Jasper S Phelps, David Grant Colburn Hildebrand, Brett J Graham, Aaron T Kuan, Logan A Thomas, Tri M Nguyen, Julia Buhmann, Anthony W Azevedo, Anne Sustar, Sweta Agrawal, et al. Reconstruction of motor control circuits in adult drosophila using automated transmission electron microscopy. *Cell*, 184(3):759–774, 2021.
 - [23] Anthony Azevedo, Ellen Lesser, Jasper S Phelps, Brandon Mark, Leila Elabbady, Sumiya Kuroda, Anne Sustar, Anthony Moussa, Avinash Khandelwal, Chris J Dallmann, et al. Connectomic reconstruction of a female Drosophila ventral nerve cord. *Nature*, 631(8020):360–368, 2024.
 - [24] Philipp Schlegel, Yijie Yin, Alexander S Bates, Sven Dorkenwald, Katharina Eichler, Paul Brooks, Daniel S Han, Marina Gkantia, Marcia Dos Santos, Eva J Munnelly, et al. Whole-brain annotation and multi-connectome cell typing of Drosophila. *Nature*, 634(8032):139–152, 2024.
 - [25] Dustin Garner, Emil Kind, Jennifer Yuet Ha Lai, Aljoscha Nern, Arthur Zhao, Lucy Houghton, Gizem Sancer, Tanya Wolff, Gerald M Rubin, Mathias F Wernet, et al. Connectomic reconstruction predicts visual features used for navigation. *Nature*, 634(8032):181–190, 2024.
 - [26] Tri M Nguyen, Logan A Thomas, Jeff L Rhoades, Ilaria Ricchi, Xintong Cindy Yuan, Arlo Sheridan, David GC Hildebrand, Jan Funke, Wade G Regehr, and Wei-Chung Allen Lee. Structured cerebellar connectivity supports resilient pattern separation. *Nature*, 613(7944):543–549, 2023.
 - [27] Gary B Huang, Louis K Scheffer, and Stephen M Plaza. Fully-automatic synapse prediction and validation on a large data set. *Frontiers in neural circuits*, 12:87, 2018.

- [28] Louis K Scheffer, C Shan Xu, Michal Januszewski, Zhiyuan Lu, Shin-ya Takemura, Kenneth J Hayworth, Gary B Huang, Kazunori Shinomiya, Jeremy Maitlin-Shepard, Stuart Berg, et al. A connectome and analysis of the adult drosophila central brain. *elife*, 9:e57443, 2020.
- [29] Shin-ya Takemura, Kenneth J Hayworth, Gary B Huang, Michal Januszewski, Zhiyuan Lu, Elizabeth C Marin, Stephan Preibisch, C Shan Xu, John Bogovic, Andrew S Champion, Han SJ Cheong, Marta Costa, Katharina Eichler, William Katz, Christopher Knecht, Feng Li, Billy J Morris, Christopher Ordish, Patricia K Rivlin, Philipp Schlegel, Kazunori Shinomiya, Tomke Stürner, Ting Zhao, Griffin Badalamente, Dennis Bailey, Paul Brooks, Brandon S Canino, Jody Clements, Michael Cook, Octave Duclos, Christopher R Dunne, Kelli Fairbanks, Siqi Fang, Samantha Finley-May, Audrey Francis, Reed George, Marina Gkantia, Kyle Harrington, Gary Patrick Hopkins, Joseph Hsu, Philip M Hubbard, Alexandre Javier, Dagmar Kainmueller, Wyatt Korff, Julie Kovalyak, Dominik Krzeminski, Shirley A Lauchie, Alanna Lohff, Charli Maldonado, Emily A Manley, Caroline Mooney, Erika Neace, Matthew Nichols, Omotara Ogundeyi, Nneoma Okeoma, Tyler Paterson, Elliott Phillips, Emily M Phillips, Caitlin Ribeiro, Sean M Ryan, Jon Thomson Rymer, Anne K Scott, Ashley L Scott, David Shepherd, Aya Shinomiya, Claire Smith, Natalie Smith, Alia Suleiman, Satoko Takemura, Iris Talebi, Imaan FM Tamimi, Eric T Trautman, Lowell Umayam, John J Walsh, Tansy Yang, Gerald M Rubin, Louis K Scheffer, Jan Funke, Stephan Saalfeld, Harald F Hess, Stephen M Plaza, Gwyneth M Card, Gregory SXE Jefferis, and Stuart Berg. A Connectome of the Male Drosophila Ventral Nerve Cord, 5 2024.
- [30] Daniel Franco-Barranco, Arrate Muñoz-Barrutia, and Ignacio Arganda-Carreras. Stable deep neural network architectures for mitochondria segmentation on electron microscopy volumes. *Neuroinformatics*, 20(2):437–450, 2022.
- [31] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. Objects as points. *arXiv preprint arXiv:1904.07850*, 2019.
- [32] Daniel Franco-Barranco, Andrés-San Román, A Jesús, Ivan Hidalgo-Cenalmor, Lenka Backová, Aitor González-Marfil, Clément Caporal, Anatole Chessel, Pedro Gómez-Gálvez, Luis M Escudero, et al. BiaPy: Accessible deep learning on bioimages. *Nature Methods*, pages 1–3, 2025.
- [33] Dan Cire, san Cire, san Alessandro Giusti, Luca Gambardella, and Jürgen Schmidhuber. Neural Networks for Segmenting Neuronal Structures in EM Stacks.
- [34] Kisuk Lee, Jonathan Zung, Peter Li, Viren Jain, and H Sebastian Seung. Superhuman accuracy on the snemi3d connectomics challenge. *arXiv preprint arXiv:1706.00120*, 2017.
- [35] I. Loshchilov and F. Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.
- [36] Samia Mohinta. Catena: A comprehensive platform for automated large-scale connectomics., 2022.
- [37] David F Crouse. On implementing 2d rectangular assignment algorithms. *IEEE Transactions on Aerospace and Electronic Systems*, 52(4):1679–1696, 2016.
- [38] Stefan Van der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014.

Towards Generalized Synapse Detection Across Invertebrate Species

Supplementary Material

Table S1: **Synaptic pairwise comparison** measured using F1 scores. The upper row denotes the test datasets, while the first column indicates the training dataset. Consequently, diagonal entries (in gray) correspond to in-distribution (ID) evaluations, whereas off-diagonal entries represent out-of-distribution (OOD) scenarios. Within each cell, the upper value refers to performance achieved by SIMPSYN, and the lower value to that of Synful. For each column, the highest score is highlighted in **bold**, and the second-highest is underlined.

\diagdown	Test		Hemibrain	Octo	WASP	MANC	All
Train \		Hemibrain	Octo	WASP	MANC	All	
Hemibrain	0.832	0.014	0.0	0.021	0.217		
	0.401	0.155	0.0	0.142	0.175		
Octo	0.194	0.618	0.003	0.215	0.258		
	0.157	0.320	0.196	0.176	0.212		
WASP	0.006	0.318	<u>0.901</u>	0.574	0.450		
	0.0	0.032	<u>0.766</u>	0.651	0.362		
MANC	0.032	0.522	0.0	<u>0.680</u>	0.310		
	0.007	0.0	0.109	0.109	0.056		
All	<u>0.687</u>	0.624	0.909	<u>0.719</u>	0.734		
	0.385	<u>0.620</u>	0.875	<u>0.732</u>	<u>0.653</u>		

Table S2: **Synful MT1 Model Hyperparameters**

Parameter	Value
Input Size	$128 \times 128 \times 128$
Downsample Factors	$[2, 2, 2] \rightarrow [2, 2, 2] \rightarrow [2, 2, 2]$
Feature Maps (fmap_num)	12
Feature Map Increment (fmap_inc_factor)	5
Losses	BCE (mask) and MSE (vector)
Learning Rate	5×10^{-5}
Loss Combination Type	sum
Main Loss Scale (m_loss_scale)	1.0
Direction Loss Scale (d_loss_scale)	1.0
Reject Probability	0.95
Blob Radius	20
Max Iterations	300,000
Blob Mode	ball
Direction Scale (d_scale)	1
Direction Blob Radius (d_blob_radius)	150
Clip Range	$[7 \times 10^{-4}, 0.9993]$
Voxel Size	$8 \times 8 \times 8$

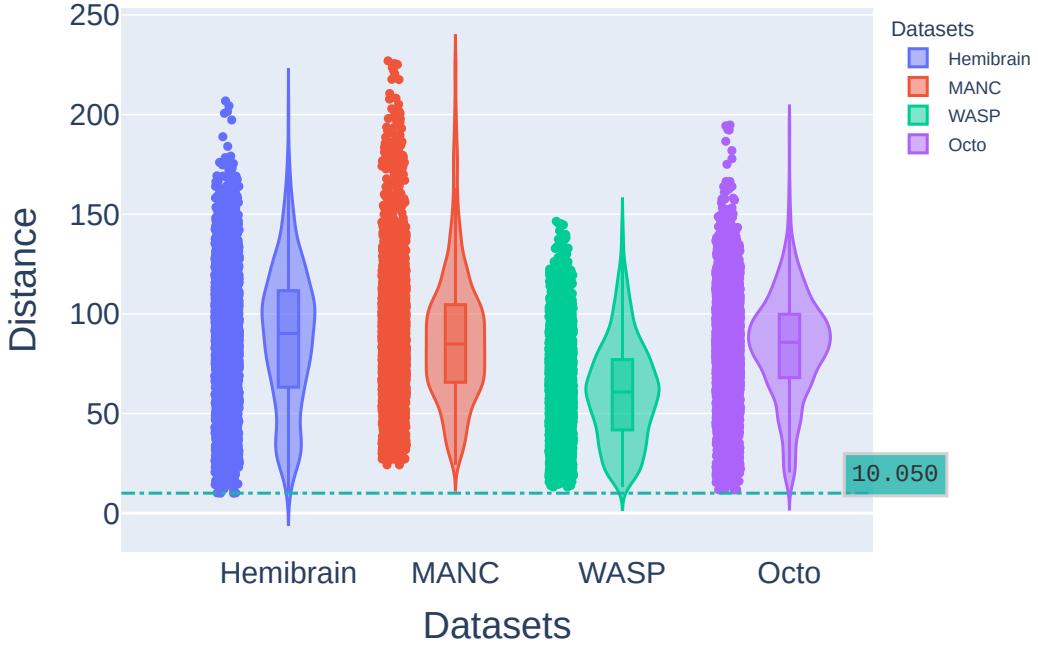


Figure S1: Pre point distribution in the ground truth data across all datasets. The green line and its corresponding value in the annotation 10.050 represent the minimum distance (in pixels) found across all datasets. This translates to 80nm taking into account that the data resolution is $8 \times 8 \times 8$ for all datasets.

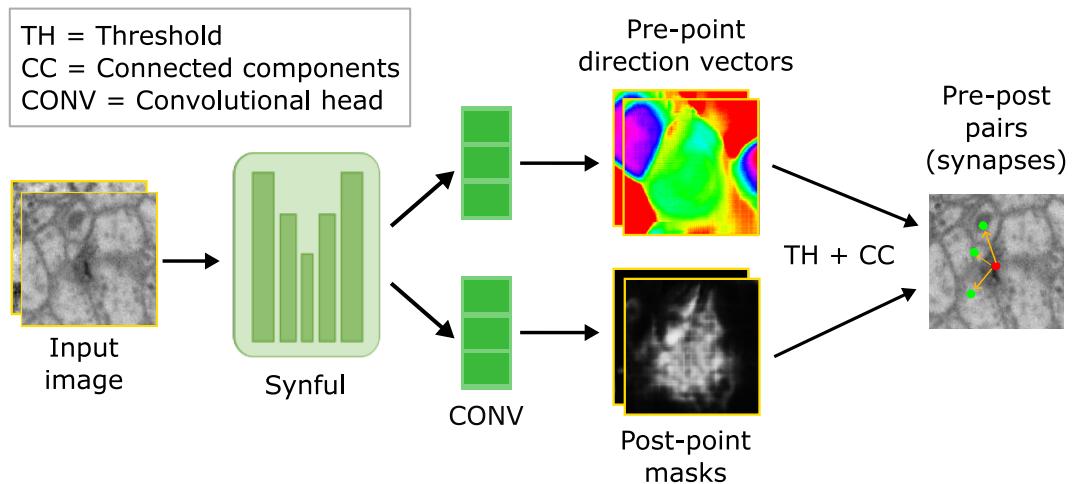


Figure S2: Synful schematic. Our Synful baseline uses a double-headed 3D unet architecture (MT1), which jointly predicts the post-synaptic masks and the pre-synaptic direction vectors. Final pre-post synaptic pairs are extracted after thresholding and grouping connected components.

Table S3: **Hyperparameter search for SIMPSYN, exploring the effect of input field of view on F1-score.** Larger patch sizes without downsampling along the Z-axis across all four levels of the Residual U-Net lead to improved performance.

Patch size	Z down.	Feature maps	Pre ↑	Post ↑
(128, 128, 128, 1)	✗	[16,32,64,128,256]	0.721	0.557
(64, 64, 64, 1)	✗	[16,32,64,128,256]	0.717	0.528
(64, 128, 128, 1)	✗	[16,32,64,128,256]	0.72	0.556
(80, 80, 80, 1)	✗	[32, 64, 128, 256]	0.711	0.546
(80, 80, 80, 1)	✓	[32, 64, 128, 256]	0.706	0.546
(128, 128, 128, 1)	✓	[32, 64, 128, 256]	0.738	0.564
(128, 128, 128, 1)	✗	[32, 64, 128, 256]	0.762	0.570

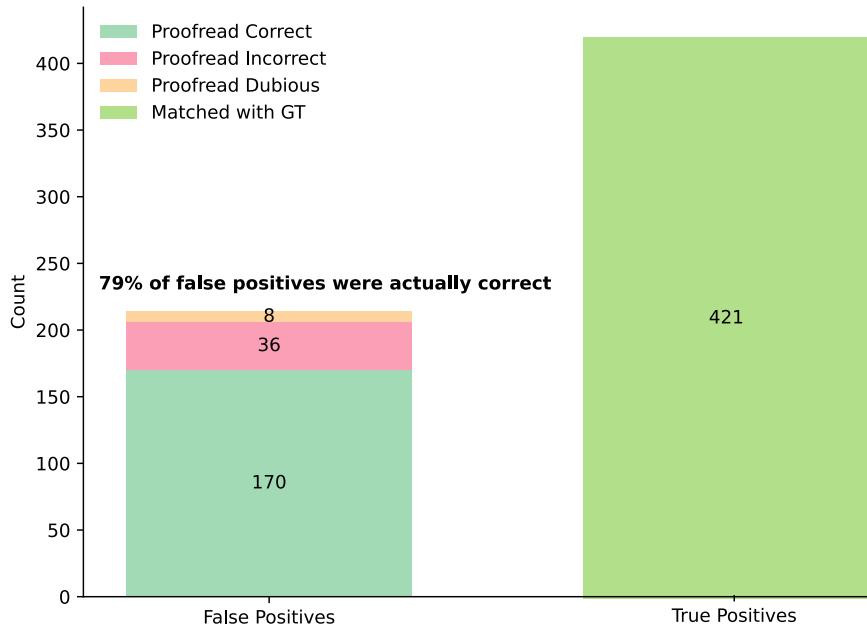


Figure S3: **Breakdown of synapse detection outcomes and proofreading results.** The left bar shows the composition of model-predicted 214 false positives in Octo. After manual proofreading by a domain expert : 170 (79%) were found to be valid synapses missed in the ground truth, while 36 were incorrect and 8 remained ambiguous. The right bar shows ground truth–matched true positives. These results underscore the inherent complexity of EM data and the resulting annotation incompleteness, which can lead to the penalization of correct predictions during evaluation.