WASPSYN: A Challenge for Domain Adaptive Synapse Detection in Microwasp Brain Connectomes

Yicong Li, *Graduate Student Member, IEEE*, Wanhua Li, *Member, IEEE*,
Qi Chen, *Graduate Student Member, IEEE*, Wei Huang, Yuda Zou, Xin Xiao, Kazunori Shinomiya,
Pat Gunn, Nishika Gupta, Alexey Polilov, Yongchao Xu, Yueyi Zhang, *Member, IEEE*,
Zhiwei Xiong, *Member, IEEE*, Hanspeter Pfister, *Fellow, IEEE*, Donglai Wei,
and Jingpeng Wu, *Member, IEEE*

Abstract—The size of image volumes in connectomics studies now reaches terabyte and often petabyte scales with a great diversity of appearance due to different sample preparation procedures. However, manual annotation of neuronal structures (e.g., synapses) in these huge image volumes is time-consuming, leading to limited labeled training data often smaller than 0.001% of the large-scale image volumes in application. Methods that can utilize in-domain labeled data and generalize to out-of-domain unlabeled data are in urgent need. Although many domain adaptation approaches are proposed to address such issues in the natural image domain, few of them have been evaluated on connectomics data due to a lack of domain adaptation benchmarks. Therefore, to enable developments of domain adaptive synapse detection methods for large-scale connectomics applications, we annotated 14 image volumes from a biologically diverse set of *Megaphragma viggianii* brain regions originating from three different whole-brain datasets and organized the WASPSYN challenge at ISBI 2023. The annotations include coordinates of pre-synapses and post-synapses in the 3D space, together with their one-to-many connectivity information. This paper describes the dataset, the tasks, the proposed baseline, the evaluation method, and the results of the challenge. Limitations of the challenge and the impact on neuroscience research are also discussed. The challenge is and will continue to be available at https://codalab.lisn.upsaclay.fr/competitions/9169. Successful algorithms that emerge from our challenge may potentially revolutionize real-world connectomics research and further the cause that aims to unravel the complexity of brain structure and function.

Index Terms—Connectomics, deep learning, electron microscopy, domain adaptation, synapse detection.

I. Introduction

NEURONS are the basic functional units of the brain that can be long enough to span brain hemispheres and specifically connect to other neurons with nanometer-sized synapses. Synapses constrain the information flow in the brain and thus knowing synaptic connectivity is essential for understanding brain function and dysfunction. To investigate such connectivity, neuronal imaging methods with both a large field of view and nanometer resolution are needed. Fortunately, with the development of *Volume Electron Microscopy* [1], [2], [3], [4], [5], those requirements are met and as a result, many terabyte and petabyte-scale image volumes are being produced [6], [7]. Techniques involving machine learning [8], [9], especially *Deep Learning* [10], can label such large-scale image volumes automatically with good accuracy [11], provided that large-scale annotated data are available for training the model. However, manual annotation of neuronal structures (e.g., synapses) in these huge image volumes is time-consuming [12], leading to limited labeled training data often smaller than 0.001% of the large-scale image volumes in application. Moreover, image volumes across different brain samples may manifest a great diversity of appearances, making

it hard for deep learning models that are trained with data from one sample to generalize well on data from other different samples. In the natural image domain, a line of research called domain adaptation (DA) is well-suited to tackle the aforementioned issues but it is less evaluated on the synapse detection task in the connectomics field, due to a lack of domain adaptation benchmarks.

Based on limited training data, some challenges are hosted for developing state-of-the-art machine learning algorithms and most of them are designed for neurite tracing, such as SNEMI3D[1] and BigNeuron [13] in VEM and Light Microscopy images, respectively. In the neuron connectivity graph, neurons are the nodes and synapses are the edges. There is a lack of synapse detection challenges compared with neurite tracing. Here, we hosted a challenge, called WASPSYN, that aims to set a new benchmark to evaluate the effectiveness of current domain adaptation methods when tailored to the synapse detection task. The challenge data consists of 14 carefully annotated image volumes from a biologically diverse set of microwasp (*Megaphragma viggianii*) brain regions, originating from three different whole-brain datasets. Coordinates of pre-synapses and post-synapses in the 3D space, together with their one-to-many connectivity information, are included in the annotations. This paper gives an overview of the challenge, covering the dataset, the tasks, the proposed baseline, the evaluation method, and the analysis of the challenge results.

A. Related Work

In this section, we provide a comparison between the WASPSYN challenge and an existing challenge CREMI (Circuit Reconstruction from Electron Microscopy Images), as well as an overview of notable literature on synapse detection and domain adaptation.

1) Comparison With an Existing Challenge: CREMI[2] is a challenge at MICCAI 2016 conference. Volumes of adult *Drosophila melanogaster* brain with pre- and post-synapse annotations are provided in this challenge. It has substantially facilitated computer vision research in the connectomics community and helped to achieve accurate neuron segmentation and synapse detection results. However, CREMI still lacks coverage that we aim to address in the WASPSYN challenge:

- Volumes in CREMI are acquired using Serial Section Transmission Electron Microscopy (SS-TEM) with anisotropic voxel size. Thus, one missing task is to develop the analysis of another advanced imaging method, Focused Ion Beam Scanning Electron Microscopy (FIB-SEM), with isotropic voxel size. The voxel size of our volumes is $8 \times 8 \times 8$ nm compared with $4 \times 4 \times 40$ nm in CREMI. A detailed comparison of imaging methods can be found in [2].
- Our challenge focuses on testing generalization capability. In contrast with CREMI, whose data are from the same image stack where the test volumes are close to the training volumes, we provide data from three brain

samples, generating a diverse set of test volumes from different domains. In total, we have annotated 14 volumes compared to 6 volumes in the CREMI challenge.
- Different from CREMI in which synaptic clefts and post-synaptic density are easy to identify, such structures in our volumes are not clear due in part to smaller neurons and lower planar resolution. As a result, it is more challenging to detect post-synapses in our data.
- For each synapse, CREMI annotates multiple point pairs across the synaptic cleft and there are many such points in each pre-synapse. Since the post-synaptic densities are not always clearly visible but pre-synaptic motifs (also known as T-bar ribbons in the insect nervous system) are always visible in our volumes, we label one point in each T-bar ribbon in the bouton. Thus, the distance from the pre-synapse to the corresponding post-synapses is much longer than the cross-membrane distance in CREMI. This requires a larger field of view when developing the machine learning model. Besides, deciding pre- and post-synapse connectivity becomes harder in the WASPSYN challenge as our data exhibit one-to-many mapping while CREMI data exhibit one-to-one mapping crossing the post-synaptic density.
- The mushroom body neurons in insects have distinct synapse structures compared to other neurons. No such volumes are presented in the CREMI challenge.
- CREMI provides cell instance segmentation annotations for all the volumes, which could be used to help train the models. In contrast, we do not have those labels, resulting in a harder synapse detection task.

In short, an immense diversity of neuron and synapse textures exists in WASPSYN data and it is challenging to maintain consistent accuracy across different brain samples. We focus on testing the generalization capability of domain adaptation algorithms and hope that successful methods emerging from this challenge will reduce the required amount of manual annotations in real-world connectomics applications.

2) Synapse Detection: Synapse detection [14], [15] has gained increasing attention as it is a crucial task in connectomics. Early research methodologies [16], [17], [18], [19], [20] primarily centered around segmenting the synaptic cleft region with hand-crafted image features. Becker et al. [17] introduced the context cues features, which were computed in multiple image channels with several Gaussian kernels. Then AdaBoost was further employed to select the discriminative features. Jagadeesh et al. [18] presented an attribute-based descriptor for synapse classification and localization in SS-TEM images of the rabbit retina. Kreshuk et al. [19] proposed a two-stage training algorithm and performed pixel classification directly in 3D, where a Random Forest classifier was trained for final object classification. Due to the insufficient discriminative power of the hand-crafted features used in these methods, their performance still does not meet the practical demands.

With the development of deep learning techniques, the field of synaptic detection in electron microscopy volumes has witnessed significant advancements in recent years. Diverse strategies [21], [22] have emerged for the automated detection of synapses, while further investigations have delved into the

---

[1]https://snemi3d.grand-challenge.org/
[2]https://cremi.org/

intricacies of synaptic detection in more challenging model organisms. Significantly, in the context of vertebrates, recent methodologies have showcased a dependable ability to deduce synaptic connectivity. For example, Dorkenwald et al. [23] developed the SyConn framework for synapse detection, which attained great results on the electron microscopy data from zebrafish, mouse, and zebra finch using deep convolutional neural networks. Nonetheless, when it comes to insect brains, the process of recognizing synaptic connections is notably more intricate. This complexity arises from the relatively smaller size of synapses in contrast to those in vertebrates. Additionally, there's a frequent occurrence of polyadic synapses in insect brains, wherein a single pre-synaptic site establishes connections with multiple post-synaptic sites.

To tackle these challenges, Kreshuk et al. [24] proposed a probabilistic graphical model to learn the synaptic partner assignment, where random variables were used to represent adjacent neurite synaptic roles. Heinrich et al. [25] proposed a 3D U-Net architecture and training approach for synaptic cleft segmentation in non-isotropic SS-TEM of insect nervous systems, achieving significant advancements over previous methods on the CREMI challenge dataset. Huang et al. [26] introduced a two-step automated system that predicts synaptic connections in Drosophila brain images, first identifying T-bars and then predicting partnering post-synaptic densities, showing its effectiveness in accurately reconstructing complex synaptic connections and outperforming existing methods. Buhmann et al. [14] further introduced a single-step method, which simultaneously identifies pre-synaptic and post-synaptic sites and predicts their connectivity using a 3D U-Net. In this work, we further shift our attention to synapse detection within the microwasp brain, which has a much higher synapse density.

3) Domain Adaptation: Deep learning-based machine learning technology [27], [28], [29] has achieved remarkable success over the past decade. Along with its wide application, a significant challenge arises due to the domain shift problem [30], stemming from variations in distributions between the source or reference data and the target data. As a special case of transfer learning [31], domain adaptation has emerged as a promising solution to address the above issue, which aims to bridge the distribution gap that exists among different yet interconnected domains. Existing methods for domain adaptation can be categorized into supervised DA, semi-supervised DA, and unsupervised DA based on label availability in the target domain [32], [33]. Supervised DA usually assumes a small number of labeled data from the target domain are available for training the model. However, data annotation is time-consuming and labor-intensive, particularly in the context of medical image data [34]. Therefore, more challenging scenarios have been proposed: semi-supervised DA and unsupervised DA. For semi-supervised DA, a limited set of labeled data is augmented with a supplementary pool of unlabeled data from the target domain to facilitate the training process. Meanwhile, unsupervised DA exclusively relies on the utilization of unlabeled target data to train the adaptation model. In this paper, we focus on the most challenging setting where no labeled target domain data are available.

To bridge the domain gap, existing unsupervised DA methods often consider aligning at two distinct levels: images and features [34], [35]. The primary objective of feature-level alignment is to acquire domain-invariant features across various domains, accomplished through the utilization of specifically tailored deep neural networks. Yan et al. [36] proposed a discrepancy-based method, which fine-tuned the deep models on the target domain with pseudo labels. ADDA [37] presented an adversarial-based approach, which employed generative adversarial networks (GANs) to guarantee indistinguishability between the source and target domains. Some research [38] also explored reconstruction-based approaches, which leverage data reconstruction as an auxiliary task to create a shared representation between the source and target domains. While these works focus on feature-level alignment, another line of research performs image-to-image translation for unsupervised DA. For example, Pizzati et al. [39] employed an image-to-image translation network to bridge domain gaps and achieved great performance in semantic segmentation tasks. Li et al. [40] demonstrated improved neuron membrane segmentation performance after translation from the X-ray domain to the electron microscopy domain. Some of these DA methods are explored and utilized by participants in this challenge.

B. Main Contributions

The main contributions of this paper are three folds:
- We provide the first domain adaptation benchmark for synapse detection in the connectomics field, including 14 volumes of FIB-SEM images from microwasp brains with pre- and post-synapse annotations.
- We set up an online evaluation website that is accessible continuously for researchers to test and compare the performance of their proposed algorithms.
- We give a detailed description of the dataset, propose a reasonable baseline method, and conduct a comprehensive analysis of challenge results.

II. METHODS

A. Challenge Organization

The WASPSYN challenge is sponsored by the Simons Foundation and is associated with the 2023 IEEE 20th International Symposium on Biomedical Imaging[3] (ISBI 2023). It is also covered in media by Computer Vision News.[4] Participants can download the labeled training set and unlabeled test set for developing synapse detection algorithms. The use of additional datasets from other sources is not allowed. Results on the test set should be submitted to the challenge website at https://codalab.lisn.upsaclay.fr/competitions/9169 by participants for evaluation. The maximum number of submissions per day per team is 5 and the teams should submit all cases in the test set. A continuous evaluation is available at the challenge website but only the top 3 participants from submissions created before the challenge

---

[3]https://biomedicalimaging.org/2023/challenges/
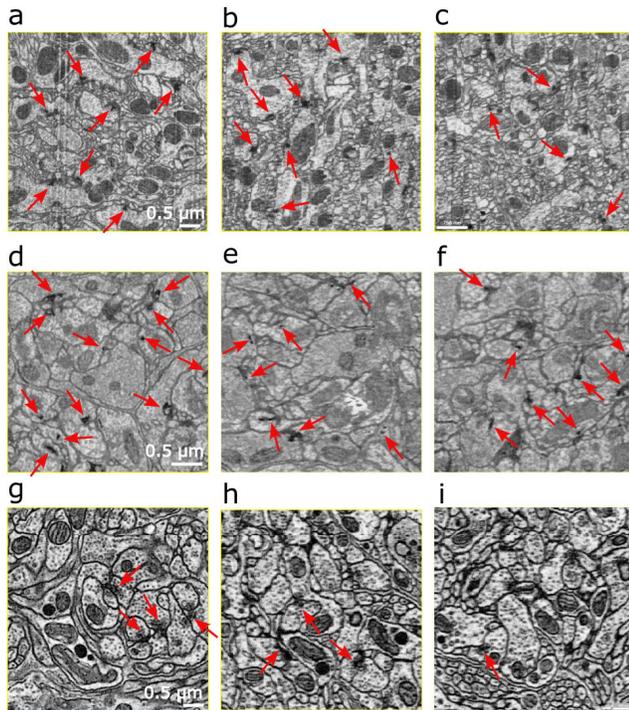[4]https://www.rsipvision.com/ComputerVisionNews-2023March/42/

Fig. 1. Image sections from three samples. Top row to bottom row: sections from sample 1, sample 2, and sample 3. Left column to right column: sections from X-Y, X-Z, and Y-Z plane. Arrows: synaptic sites identified in the section. Red arrows: synaptic sites of mushroom bodies specifically. Scale bar: 0.5 μm.

deadline were awarded prizes and invited to share their methods at ISBI 2023 WASPSYN challenge workshop. The dataset will also continue to be publicly available to everyone after the challenge under the CC-BY license. Participants are encouraged to perform an uncertainty or statistical analysis of their algorithms and report the results of the analysis in their publications. The code availability is voluntary but we strongly encourage the participants to open-source their code for reproducibility.

B. Dataset

WASPSYN dataset consists of 14 image volumes from *Megaphragma viggianii*, a type of insect with a small brain size. These wasps have evolved anuclate neurons, likely due to the selective pressure that has driven miniaturization [41]. The scientific significance is detailed in previous publications [41], [42], [43]. We present a detailed description of the WASPSYN dataset below.

1) Acquisition: The whole head of *Megaphragma viggianii* was firstly stained with heavy metal and embedded in resin [44]. Subsequently, the sample was imaged using an enhanced Focused Ion Beam Scanning Electron Microscope (FIB-SEM) [45], [46], [47] with an isotropic voxel size of $8 \times 8 \times 8$ nm.

2) Notable Features:
- Cross-sample variation: We imaged three brain samples and then annotated 14 image volumes from them. As shown in Fig. 1, significant appearance differences (i.e., domain differences) can be observed among three samples, posing generalization challenges to machine

learning models. Based on our observation, microwasp brain anatomy is conservative at the neural circuit level [48]. The variation among different samples is mostly from sample preparation and imaging parameter variations, i.e., the samples were prepared according to a protocol with parameter variations. Specifically, sample 2 was prepared according to the protocol described in [44] and embedded into Durcupan. Sample 1 and sample 3 were prepared using the same protocol as sample 2 but with minor differences. Sample 1 was prepared without ferrocyanide treatment and embedded in Epon. As for sample 3, the time of primary fixation was reduced and all washing stages were lengthened.
- Challenging cases: In the mushroom body, multiple Kenyon cell terminals connect to an output neuron terminal, exhibiting a rosette-like structure. Pre-synaptic terminals of Kenyon cells in a rosette lack platforms and are smaller than typical pre-synapses, making them harder to detect.

3) Annotations: Each image volume was initially annotated by one of five annotators (Section VI). These annotations were subsequently peer-reviewed by a different annotator. The annotators, who were either full-time or part-time technical staff, had specialized training in interpreting electron microscopy images and annotating ultrastructure in insect neurons for at least two years. We used CATMAID [49] and NeuTu [50] with DVID [51] to label the pre-synapses and post-synapses. In the brain of *Megaphragma viggianii*, a chemical synapse consists of a pre-synaptic terminal, accompanied by an electron-dense motif called T-bar, and multiple post-synaptic sites characterized by electron-dense regions. A T-bar consists of a platform, or "table-top", and a pedestal connecting the cell membrane and the platform. A pre-synapse point annotation (T-bar glyph) should be placed at the connecting point of the platform and the pedestal. A large platform may have contacts with more than one pedestal, in which case each contact point should be annotated as a separate pre-synapse. Neuronal processes are annotated as post-synapses if post-synaptic density is easy to identify (e.g., in sample 3). If post-synaptic density is not recognizable (e.g., in sample 1), all bodies within 40 nm from the edge of the platform are considered to have post-synapses (see Fig. 2 for details). We expect that the network trained using our ground truth would learn that only the touching neurite within a short distance of T-bar platform (Fig. 2) could be a post-synapse candidate. Fig. 3 shows an example of annotations for an image volume. It can be observed that each pre-synapse is connected to several post-synapses, annotated by points and lines representing their locations and connections. Specifically, the annotations are coordinates of pre-synapses' and post-synapses' locations in the 3D space, together with their one-to-many connectivity information.

4) Data Split: Table I provides detailed information on each image volume in the dataset. The training set includes 5 volumes from sample 3 while the test set includes 9 volumes (3 from each sample). Participants can access all 14 volumes but only the training set has ground truth annotations. In such a design, volumes from sample 3 are considered in-domain
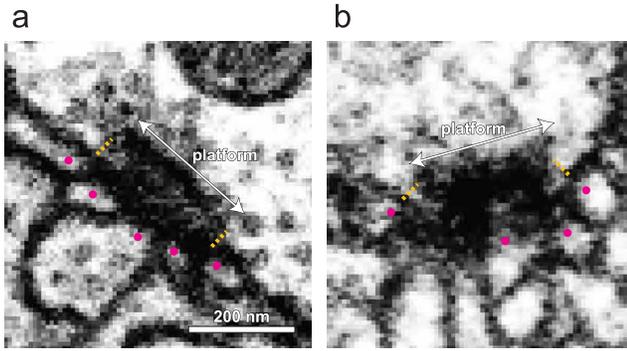
Fig. 2. The platform of a T-bar is a small structure that typically extends a few hundred nanometers in length. It is positioned parallel to the membrane and separated from it by about 40 nm of space, which is indicated by the dotted lines. The post-synaptic terminals, indicated by the pink dots, are marked by post-synaptic densities (PSDs; (a)), and their distribution corresponds to the extent of the platform in principle. If PSDs are not clearly visible, as in (b), all profiles within 40 nm from the tips of the platform are considered post-synaptic terminals. Scale bar: 200 *nm*.

(i.e., source domain) data, and volumes from the other two samples are considered out-of-domain (i.e., target domain) data, enabling us to test the generalization ability of machine learning models.

C. Tasks

In this challenge, we define two computational tasks following the data annotations:
- **Pre-synapse Detection (Task 1):** Participants are required to predict the locations of pre-synapses in the 3D space using the provided dataset.
- **Post-synapse Detection (Task 2):** Participants are required to predict the locations of post-synapses in the 3D space as well as the synaptic connectivity (i.e., IDs of the pre-synapses to which the post-synapses are connected) using the provided dataset.

D. Evaluation of Submissions

Participants are required to submit their detection results of pre-synapses and post-synapses including their connectivity for evaluation. To select the appropriate strategy for evaluation, we follow the newly established Metrics Reloaded[5] [52] framework. Based on the suggestion from the framework, the detection accuracy of a submission will be evaluated by solving an assignment problem [53] minimizing the Euclidean distance between detected synapses and ground truth synapses to find true matches and calculating the F1-score. Formally, given a set of detected synapses (D) by a participant and a set of ground truth synapses (G), we want to find a bipartite matching $f : D \rightarrow G$ with the Hungarian algorithm [53] to minimize the following cost function:

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

where $C(\cdot)$ denotes the Euclidean distance of a matched pair. Next, F1-score is defined as:

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

[5]https://metrics-reloaded.dkfz.de/

where TP is the true positive, FP is the false positive, and FN is the false negative.
- **Evaluation of Pre-synapse Detection:** Detected pre-synapses are considered to be potential matches to the ground truth pre-synapses. After solving the assignment problem, an unmatched detected pre-synapse will be counted as one FP, an unmatched ground truth pre-synapse will be counted as one FN, and a falsely matched pre-synapse pair will be counted as one FP and one FN. The pre-synapse detection accuracy will be expressed as the F1-score calculated using TPs, FPs, and FNs. Notably, we use a threshold to determine the falsely matched pairs. If the Euclidean distance between the detected pre-synapse and the matched ground-truth pre-synapse in a pair exceeds the threshold, this pair will be considered a falsely matched pair. Specifically, for each volume, we calculate the minimum Euclidean distance between two pre-synapses. Then, we get the average of these minimum distances across all volumes and set the half value of the average as the threshold, which is 88 nm for pre-synapse detection.
- **Evaluation of Post-synapse Detection:** Since our data involves one-to-many synapse connectivity, for each matched pre-synapse pair, we compare the post-synapses connected to it by solving the assignment problem mentioned above. An unmatched but detected post-synapse will be counted as one FP, an unmatched ground truth post-synapse will be counted as one FN, and a falsely matched post-synapse pair will be counted as one FP and one FN. The F1-score for post-synapse detection can be computed using TPs, FPs, and FNs. Similarly, we use a threshold of 52 nm for post-synapse detection.
- **Ranking Scheme:** For each test volume, we calculate an F1-score for pre-synapse detection (task 1) and an F1-score for post-synapse detection (task 2). Then, we calculate the arithmetic mean of these two scores to get the final score for each test volume. Lastly, we average the final scores over all test volumes to determine each participant's position on the leaderboard.

E. Baseline Method

To facilitate the challenge and provide the participants with a starting point for developing their own synapse detection algorithms, we propose a two-step method based on 3D U-Net [54], [55] as the challenge baseline which, to the best of our knowledge, is the first approach that aims to tackle the one-to-many synapse detection problem, as depicted in Fig. 4.

1) Training: At the training stage, the first step is to train a model for pre-synapse detection (Fig. 4-a). The input to the 3D U-Net is a small 3D image volume randomly sampled from the whole volume. Inspired by [56], the images are augmented by the following methods: brightness and contrast adjustment, Gamma transform, Gaussian noise or Gaussian blurring, random black boxes, perspective transformation in 2D, flipping, transpose, image misalignment by shifting 2D sections. The point annotations (3D coordinates) are transformed into voxel cubes with a size of $3 \times 3 \times 3$ to be used as the training target with a binomial cross-entropy
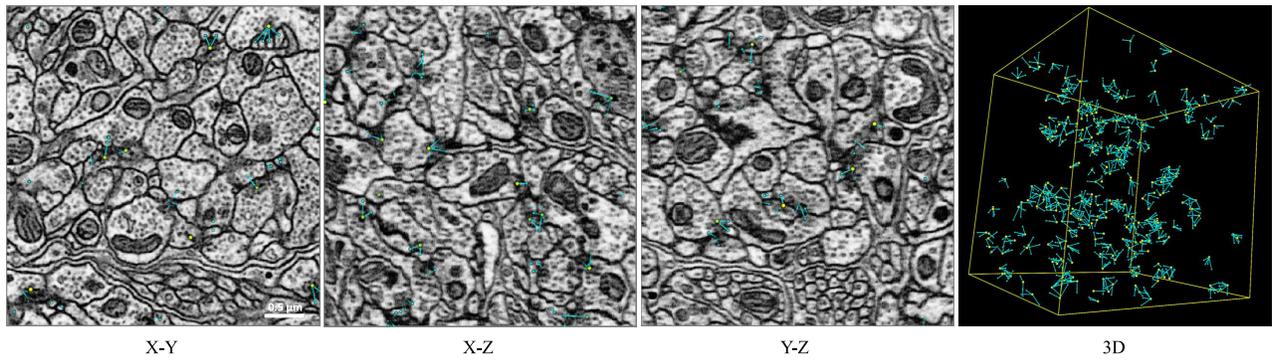
| X-Y | X-Z | Y-Z | 3D |

Fig. 3. Visualization of the ground truth annotations of an image volume from sample 3. Yellow dots: locations of pre-synapses. Cyan dots: locations of post-synapses. Cyan lines: synaptic connectivity. Scale bar: 0.5 μm.

TABLE I
DETAILED INFORMATION ON EACH IMAGE VOLUME IN THE DATASET

Training Volume	Voxel Size	Volume Size	# of Pre-synapses	# of Post-synapses	Brain Region	Label Availability
Sample 3 Volume 0	$8 \times 8 \times 8$ nm	$400 \times 400 \times 400$	295	880	MB-ML*	✔
Sample 3 Volume 1	$8 \times 8 \times 8$ nm	$416 \times 416 \times 416$	155	718	AL*	✔
Sample 3 Volume 2	$8 \times 8 \times 8$ nm	$416 \times 416 \times 416$	225	1504	PLP*	✔
Sample 3 Volume 3	$8 \times 8 \times 8$ nm	$416 \times 416 \times 416$	47	272	GNG*	✔
Sample 3 Volume 4	$8 \times 8 \times 8$ nm	$416 \times 416 \times 416$	261	878	CBL*	✔
Test Volume	**Voxel Size**	**Volume Size**	**# of Pre-synapses**	**# of Post-synapses**	**Brain Region**	**Label Availability**
Sample 1 Volume 0	$8 \times 8 \times 8$ nm	$600 \times 600 \times 600$	268	3441	GNG	✘
Sample 1 Volume 1	$8 \times 8 \times 8$ nm	$600 \times 600 \times 600$	164	1746	medulla, OCh1*	✘
Sample 1 Volume 2	$8 \times 8 \times 8$ nm	$600 \times 600 \times 600$	527	5264	medulla	✘
Sample 2 Volume 0	$8 \times 8 \times 8$ nm	$400 \times 400 \times 400$	151	1652	medulla	✘
Sample 2 Volume 1	$8 \times 8 \times 8$ nm	$400 \times 400 \times 400$	185	1296	AL	✘
Sample 2 Volume 2	$8 \times 8 \times 8$ nm	$600 \times 600 \times 600$	254	2732	MB calyx	✘
Sample 3 Volume 0	$8 \times 8 \times 8$ nm	$400 \times 400 \times 400$	188	1059	MB-VL*	✘
Sample 3 Volume 1	$8 \times 8 \times 8$ nm	$416 \times 416 \times 416$	197	1240	MB-ML protocerebrum	✘
Sample 3 Volume 2	$8 \times 8 \times 8$ nm	$416 \times 416 \times 416$	232	1607	OL*	✘

* MB-ML: mushroom body medial lobe. AL: antennal lobe. PLP: posterior lateral protocerebrum. GNG: gnathal ganglia (root of the cervical connective). CBL: lower unit of the central body (ellipsoid body). OCh1: first optic chiasm. MB-VL: mushroom body vertical lobe. OL: optic lobe.
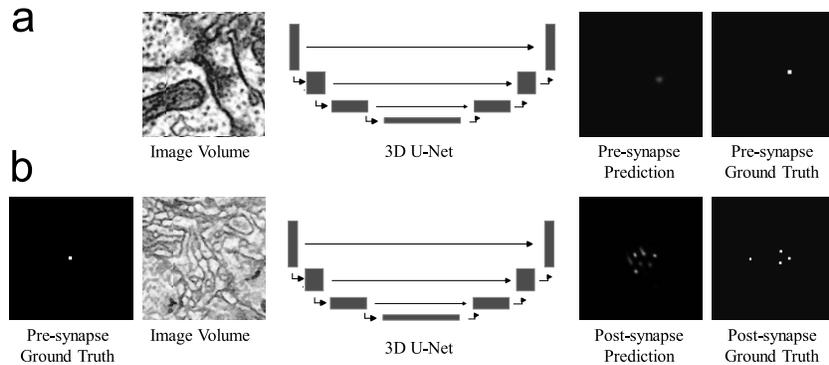

Fig. 4. Baseline method using 3D U-Net. (a) Training of T-bar detection network. The input image patches, subvolumes of ground truth image volumes, are randomly sampled so that some image patches might not contain any T-bar. (b) Training of post-synapse detection network. The T-bar is in the center of each input image patch and a fixed patch with a central point is used as a channel of the input. Note that the illustration is 2D while both the image patches and network are 3D and of isotropic size.

loss. In this way, the output of the model is a 3D pre-synaptic probability map, similar to a common semantic segmentation task. The second step is to train another model for post-synapse detection (Fig. 4-b). The input to the model includes a pre-synapse sampled from the ground truth annotations and a small image volume around that pre-synapse cropped from the whole volume. Similarly, the model aims to

predict a 3D post-synaptic probability map. The 3D U-Net architecture is modified from a previous synapse detection method [57]. All kernel sizes are correspondingly changed to isotropic.

2) Inference: At the inference stage, for pre-synapse detection, we process a whole image volume using a 3D sliding window. The window volumes overlap with each other
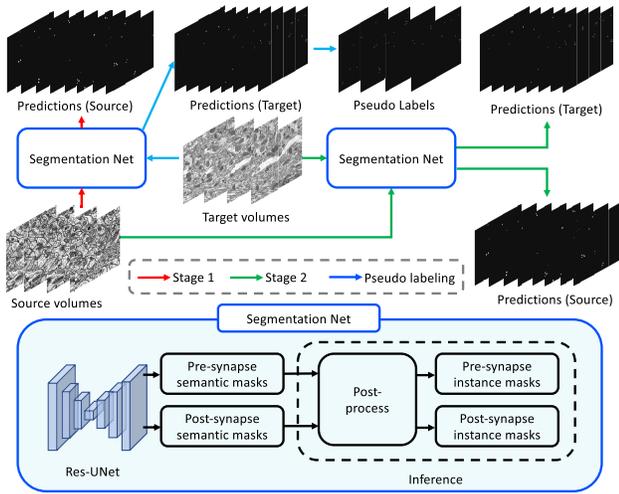
Fig. 5. The framework proposed by Team qicq1c [59]. In the first stage, the team trained an initial segmentation network using the source domain volumes with ground truth labels. In the second stage, they used the pre-trained network to generate pseudo labels for the target domain volumes. Then, the segmentation network is fine-tuned by incorporating both source labels and target pseudo labels as supervision.

by 50%, yielding eight-fold coverage of each voxel. The overlapping outputs from the model are blended together to produce a 3D pre-synaptic probability map. To find the exact pre-synapse locations, we detect local maxima in the probability map. As for post-synapse detection, we process each of the previously detected pre-synapses together with its surrounding image volume for doing the inference, producing a 3D post-synaptic probability map. Then, we detect local maxima in the probability map to find the exact post-synapse locations. The local maxima is filtered with a threshold of 0.3 and a minimum distance of 15 voxels. The inference is conducted using chunkflow [58].

III. CHALLENGE ENTRIES

Since the launching of the challenge, 48 teams have participated and over 255 submissions have been made on the challenge website. Below is a brief description of the methods proposed by the top 3 teams.

A. First Place: Team qicq1c

Team qicq1c presented a two-stage segmentation-based framework (AdaSyn) [59] for domain adaptive synapse detection with weak point annotations. They address the detection problem using an instance segmentation method. In the first stage, the team obtained the ground truth of synapse masks by expansion operation and trained a 3D segmentation network to predict synaptic regions. The network outputs two channels: one for pre-synaptic regions and another for post-synaptic regions. These predicted masks are then processed using connected component labeling to separate individual synapses. The location of each synapse is determined by calculating the center point coordinate of the corresponding mask. Then, the team assigned the nearest pre-synapse ID to each post-synapse using the nearest neighbor principle. In the second stage, to improve the generalization ability of the network, they adopted a model pre-trained on the source
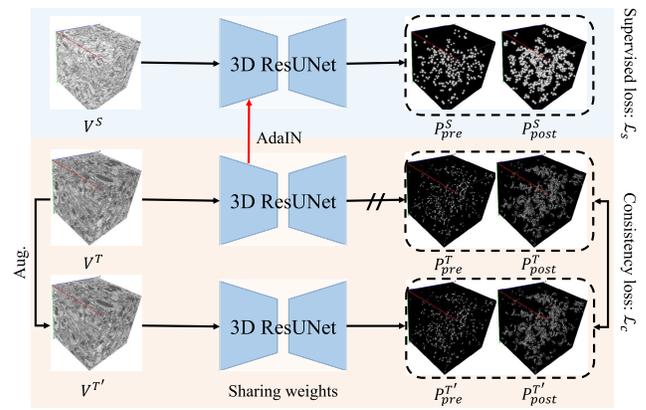

Fig. 6. The framework proposed by Team WeiHuang, consists of Adaptive Instance Normalization (AdaIN) and Consistency Learning (CL). $V^S/V^T/V^{T'}$: the volume from the source/target/augmented target domain. $P_{pre}^S/P_{pre}^T/P_{pre}^{T'}$ and $P_{post}^S/P_{post}^T/P_{post}^{T'}$: pre- and post-syanpse predictions from the source/target/augmented target volume. Note that the prediction branch of V^T is stop-gradient, i.e., without back-propagation.
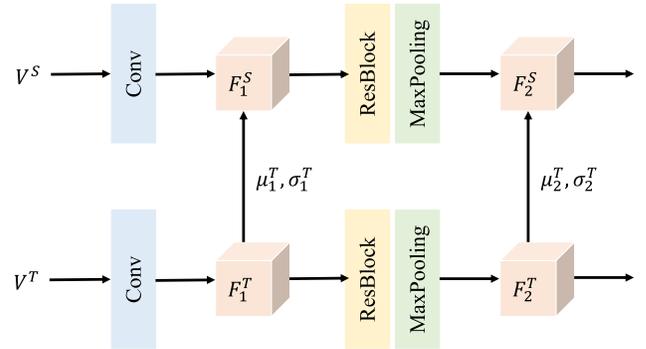

Fig. 7. The structure of AdaIN in the proposed framework by Team WeiHuang. F_i^S/F_i^T: the i-th layer feature in the source/target domain. μ_i^T and σ_i^T: the channel-wise mean and standard deviation of F_i^T.

data to generate pseudo labels for the target data. Finally, the segmentation network is fine-tuned using both source and target data. The framework is shown in Fig. 5.

B. Second Place: Team WeiHuang

The framework of the proposed method is depicted in Fig. 6. The team first converted the task of synapse detection into a segmentation task by transforming each synapse point annotation into a 3D Gaussian sphere of radius r. The center of the sphere has a value of 1 while the edge has a value of 0, and the rest of the values inside the sphere decrease with a Gaussian function from the center to the edge. Second, they adopted a 3D ResUNet [56] as a backbone to predict the paired pre- and post-synapses simultaneously. To tackle the domain adaptation problem, the team leveraged Adaptive Instance Normalization (AdaIN) and Consistency Learning (CL) to narrow the domain gap between the source and target domain, thereby improving the generalization ability of the model in the target domain. Inspired by previous style transfer methods [60], [61], AdaIN is applied on the shallow layers of the 3D ResUNet to transfer the feature distribution of the target domain into the source domain, as shown in Fig. 7. As for CL, following [62], the team

obtained the counterpart of the target volume by several augmentation operations, including Gaussian noise, Gaussian blurring, intensity variations, and cutout. Then, the target volume and its corresponding counterpart are sent to the network simultaneously.

C. Third Place: Team Melony

The framework of the proposed method is depicted in Fig. 8. Considering the clustered nature of post-synapses, the team proposed a Point Matching Network (PMN) to tackle the challenge. Specifically, they adopted the VGG-16 [64] model pre-trained on ImageNet as the backbone to extract abundant semantic representation. The feature map in stage 3 (downsampled 8 times) and that in stage 4 (downsampled 16 times) are concatenated and then fed into two parallel network branches. The two branches predict a classification score map of pre-synapses and post-synapses, respectively. In the training phase, the ground truth points are matched with the predicted point proposals based on the matching cost by the Hungarian algorithm. Regarding the loss function, the probability of all the point proposals and the distance between the matched point pairs are considered. In the testing phase, pre-synapses and post-synapses are obtained by filtering the classification scores and then matched with each other.

IV. RESULTS

In this section, we will present and discuss the quantitative results of the baseline and the methods proposed by the top 3 teams, followed by showing qualitative results highlighting true positive, false positive, and false negative cases.

A. Quantitative Results

Table II presents the overall results of synapse detection. Table III and Table IV present the results of pre- and post-synapse detection, respectively, on each volume. As expected, all methods generally work better on sample 3, which is in the same domain as the provided training data, compared to sample 1 and 2 results. Besides, the detection accuracy of pre-synapses is generally higher than that of post-synapses as there are many more post-synapses and their connectivity to pre-synapses also needs to be detected correctly. In Fig. 1, the texture difference (i.e., domain gap) between sample 3 and sample 2 is larger than that between sample 3 and sample 1. As such, volumes from sample 2 are more challenging, resulting in lower detection accuracy for all methods. According to Table II, III, and IV, the performance of the baseline method on sample 1 and sample 2 is much lower than that on sample 3, which fully reflects the challenges brought by the out-of-domain data caused by cross-sample variations. On the other hand, the leading methods developed by challenge participants show improved performance on sample 1 and sample 2 with the help of various domain adaptation algorithms, effectively alleviating the domain shift problem.

Challenge rankings can be unstable depending on how they are computed, therefore, inspired by [65], we perform a ranking uncertainty analysis by bootstrapping the challenge

TABLE II
RESULTS COMBINING PRE- AND POST-SYNAPSE DETECTION. THE NUMBERS SHOWN ARE F1-SCORES

Test Volume	Baseline	melony#3	WeiHuang#2	qicq1c#1
Sample 1 Volume 0	0.343	0.274	0.505	0.623
Sample 1 Volume 1	0.308	0.213	0.424	0.555
Sample 1 Volume 2	0.427	0.255	0.527	0.617
Sample 2 Volume 0	0.187	0.256	0.368	0.576
Sample 2 Volume 1	0.242	0.259	0.312	0.426
Sample 2 Volume 2	0.262	0.250	0.386	0.581
Sample 3 Volume 0	0.604	0.416	0.638	0.699
Sample 3 Volume 1	0.673	0.439	0.671	0.767
Sample 3 Volume 2	0.624	0.415	0.596	0.699
Overall Score	0.408	0.309	0.492	0.616

TABLE III
RESULTS OF PRE-SYNAPSE DETECTION. THE NUMBERS SHOWN ARE F1-SCORES

Test Volume	Baseline	melony#3	WeiHuang#2	qicq1c#1
Sample 1 Volume 0	0.359	0.307	0.620	0.796
Sample 1 Volume 1	0.259	0.212	0.499	0.650
Sample 1 Volume 2	0.439	0.249	0.598	0.637
Sample 2 Volume 0	0.153	0.288	0.586	0.757
Sample 2 Volume 1	0.293	0.312	0.500	0.624
Sample 2 Volume 2	0.293	0.334	0.665	0.837
Sample 3 Volume 0	0.687	0.461	0.787	0.851
Sample 3 Volume 1	0.762	0.450	0.758	0.851
Sample 3 Volume 2	0.738	0.464	0.753	0.808

TABLE IV
RESULTS OF POST-SYNAPSE DETECTION. THE NUMBERS SHOWN ARE F1-SCORES

Test Volume	Baseline	melony#3	WeiHuang#2	qicq1c#1
Sample 1 Volume 0	0.326	0.241	0.389	0.451
Sample 1 Volume 1	0.356	0.214	0.349	0.459
Sample 1 Volume 2	0.416	0.261	0.456	0.598
Sample 2 Volume 0	0.220	0.223	0.150	0.396
Sample 2 Volume 1	0.191	0.207	0.124	0.227
Sample 2 Volume 2	0.230	0.166	0.106	0.325
Sample 3 Volume 0	0.522	0.372	0.488	0.547
Sample 3 Volume 1	0.584	0.428	0.583	0.683
Sample 3 Volume 2	0.510	0.366	0.440	0.591

results using Rankings Reloaded toolkit.[6] Fig. 9 shows the results of the ranking uncertainty analysis. It suggests that, in post-synapse detection (Fig. 9-c), the uncertainty of ranking is slightly larger, but in general, the rankings of the four methods are stable in all three analyses.

B. Qualitative Results

Fig. 10 shows the qualitative results of pre- and post-synapse detection in sample 1, sample 2, and sample 3, from the first-place team's method. The results are consistent with what we have observed in Section IV-A. The detection of pre-synapses is generally easier than that of post-synapses since the proportion of true positive (magenta) cases in pre-synapse detection results is significantly larger. Results from sample 1 and sample 2 are worse than those of sample 3 due to the existing domain gaps. Although the majority of pre-synapses
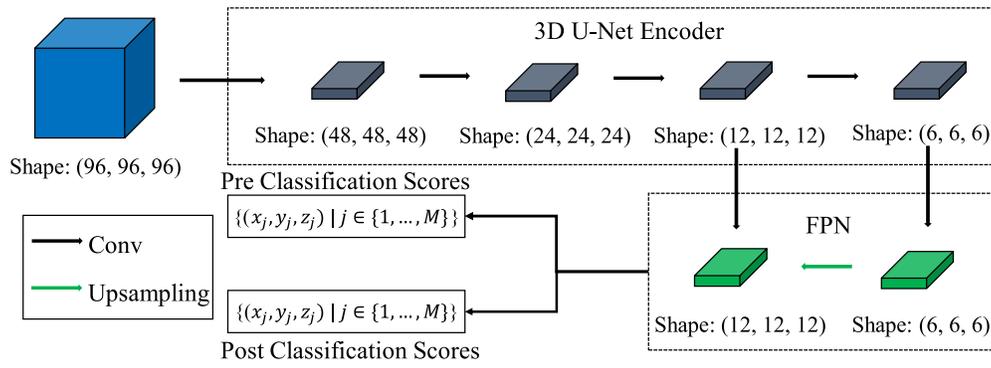
[6]https://www.rankings-reloaded.de/

Fig. 8. The framework proposed by Team melony. A pre-trained 3D VGG-16 model is used to extract multi-scale features, which are then fused through a Feature Pyramid Network (FPN) [63] module to cope with synapses with different scales. The classification maps of pre-synapses and post-synapses are generated through several simple convolutions.
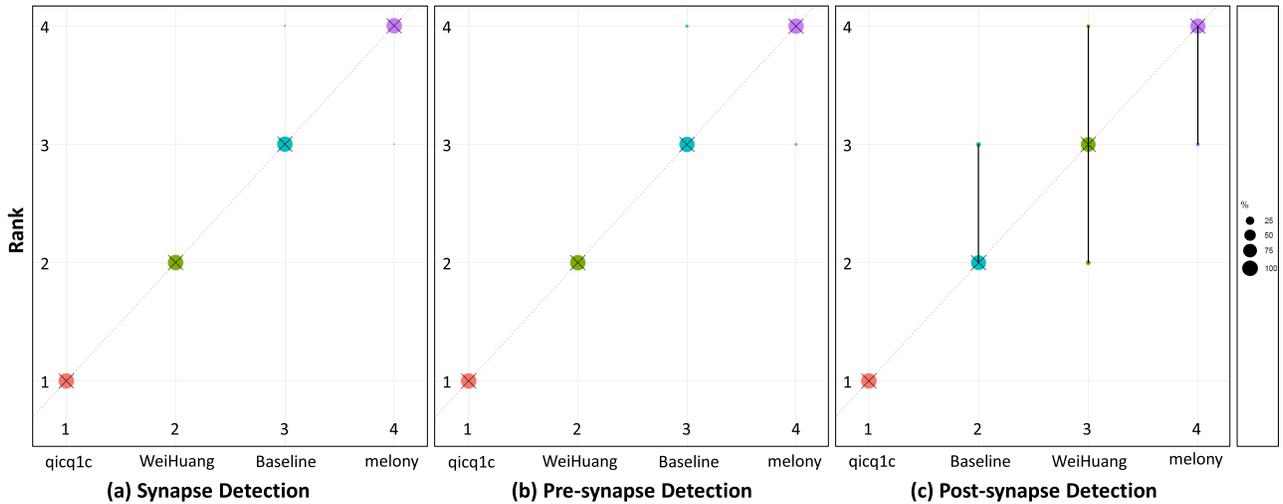

Fig. 9. Ranking uncertainty analysis. Methods are color-coded, and the area of each blob at position (method i, rank j) is proportional to the relative frequency method i achieved rank j across $b = 1000$ bootstrap samples. Black cross: the median rank for each method. Black lines: 95% bootstrap intervals across bootstrap samples.

in all three samples are accurately identified, the domain gaps appear to have a more detrimental effect on post-synapse detection based on the proportion of false negative (yellow) cases. This suggests that the inherent domain differences make it considerably challenging to detect post-synapses in sample 1 and 2. Specifically, in pre-synapse detection, false negative cases are often observed in locations that have ambiguous textures, and false positive cases are often observed in locations that resemble pre-synaptic regions. In post-synapse detection, false negative cases usually come from pre-synapses that have too many connected post-synapses, making it difficult for the model to cover all of them. False positive cases usually come from several closely located post-synapses that are wrongly assigned to only one pre-synapse but actually belong to two or more pre-synapses.

V. DISCUSSION

A. One-Step Approach v.s. Two-Step Approach

Upon reviewing the methods for synapse detection proposed in the challenge and in previous literature, we could roughly divide them into two categories: one-step approaches and two-step approaches. A one-step approach refers to detecting pre- and post-synapses in one inference step while a two-step approach refers to detecting pre-synapses first and then

detecting post-synapses for each pre-synapse during inference. In the challenge dataset, there are particular occasions when a post-synapse is connected to two pre-synapses, especially among mushroom bodies. One-step approaches may make mistakes on such occasions while two-step approaches are not negatively affected since they deal with the pre-synapses one by one. However, one-step approaches generally have the advantage of larger throughput and better computational efficiency because there are redundant overlaps of image volumes when sampling each detected pre-synapse for post-synapse detection in two-step approaches. Therefore, when making the choice between these two categories, one may need to consider the density of synapses in the data, i.e., choose two-step approaches when the density is low, and vice versa.

B. Limitations of the Challenge Design

We have identified a number of limitations that should be addressed in future studies. First, due to resource constraints, participants performed the algorithms' training offline. Additional data may potentially be used by certain teams that are not accessible to others and thus could introduce bias when comparing performance against each team. Enabling docker-based training of models directly on the challenge platform would be desirable. Second, the number of samples
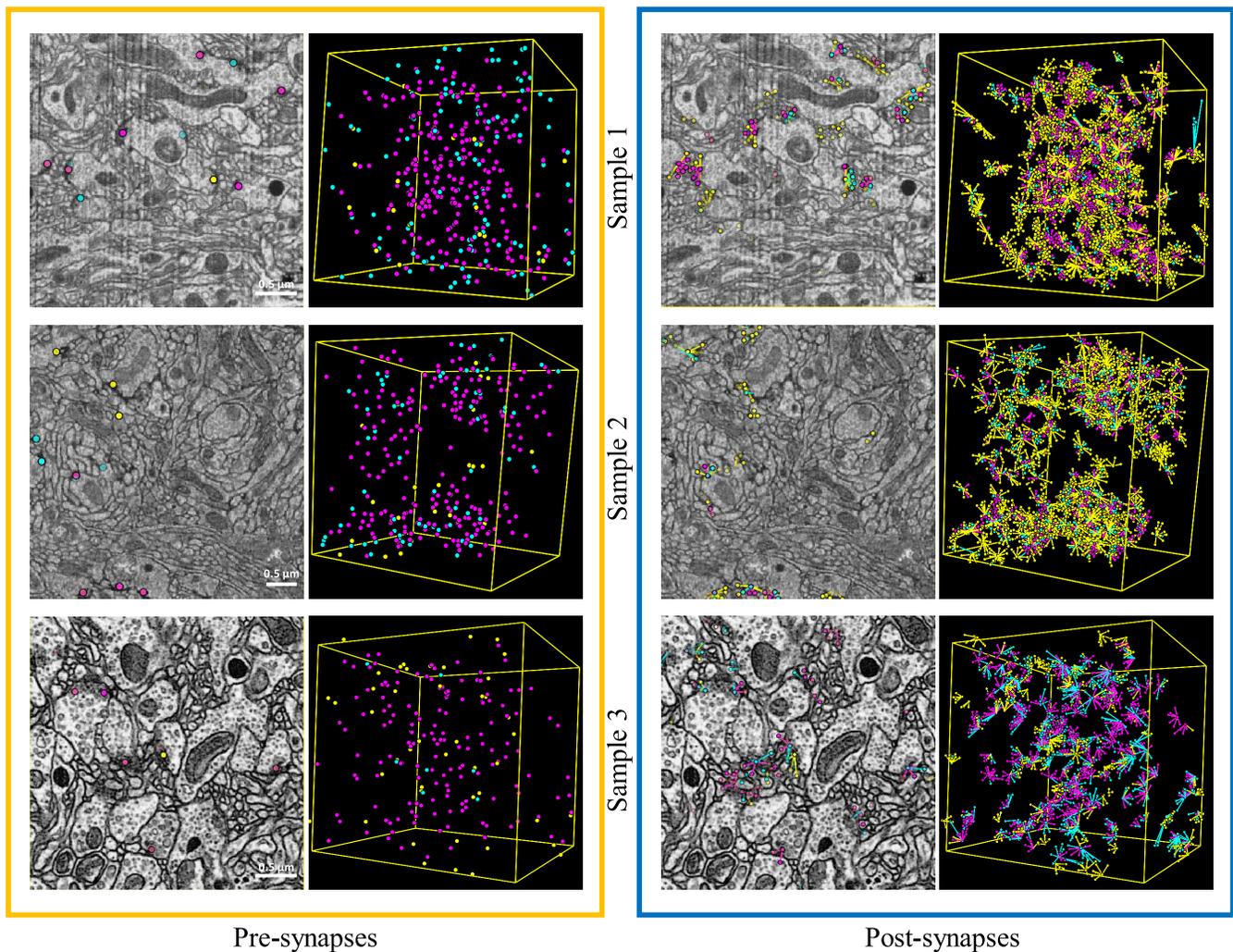
Fig. 10. Qualitative results of pre- and post-synapse detection in sample 1, sample 2, and sample 3, from the first-place team's method. Dots and lines: magenta-true positive, yellow-false negative, and cyan-false positive. Scale bar: 0.5 μm.

in the dataset, although already larger than that of the previous challenge CREMI, is still limited to some extent, preventing the developed algorithms from achieving higher accuracy. It would be possible to open source more data samples for the challenge but this is subject to the availability of resources needed to provide manual annotations. Third, a certain level of uncertainty exists in the ground truth annotations as annotators need to decide the exact point location even though we provided a guidance for annotators. Such fuzziness in point annotations could make the learning process harder than the previous segmentation mask style annotations. In the future, a desirable algorithm that emerges from this challenge should be able to incorporate the fuzziness into its modeling process. Last, this challenge considers the problem of out-of-domain generalization across different brain samples in microwasp. A more challenging and interesting research theme could be considering the generalization problem across species beyond microwasp, e.g., extending the dataset to include brain regions from fly, mouse, etc.

C. Impact on Future Research

In the WASPSYN challenge, different domain adaptation methods (either borrowed from the natural image domain or newly developed) are evaluated and compared on the synapse detection task, making it possible to identify methods that are superior when applied to connectomics research and helping explore novel ways to reduce the number of manual annotations needed to train a reasonably good machine learning model. Besides, this challenge also demonstrates that providing only point annotations as labels is also plausible for training machine learning models to detect synapses, a dataset preparation strategy that is more labor-efficient compared to producing segmentation mask style annotations (i.e., voxel painting of cleft regions). Although leading methods have demonstrated effective improvements, the scores are still not satisfactory, and the challenge of domain adaptation with weakly annotated data is still not well-solved. Therefore, the challenge website will continue to run, and we hope that some revolutionary algorithms will stand out in the future.

We believe that future research can be explored in the following aspects: 1) Models. Recent years have witnessed the huge success of foundation models. For example, SAM [66] is a foundation model for image segmentation and has been adapted for various domains. Utilizing and adapting existing powerful foundation models for domain adaptive synapse detection could be a promising direction. 2) Algorithms.

Developing new domain adaption algorithms is important as one of the biggest challenges for the WASPSYN benchmark is the domain gap. Studying more advanced domain adaption methods [67], [68] and applying them to the WASPSYN challenge would naturally bring performance improvements. 3) Data. The quality and size of data largely determine the performance of the model. While collecting and annotating real data is time-consuming and labor-intensive, high-quality data can be greatly expanded by synthesizing data. Recent research [69] has demonstrated the feasibility of using diffusion models as data engines.

D. Impact on Connectomics Community

After pre-processing each image volume and producing ground truth annotations for the training set, the challenge data is made publicly available and has facilitated scientific research in related fields. The starter code and baseline method are provided to help participants dive into the problem more easily. Besides, the online evaluation ensures that algorithms developed by participants are compared fairly against each other. Through these efforts, wide participation is achieved from research groups around the world.

VI. Conclusion

The WASPSYN challenge at ISBI 2023 is the first to evaluate a wide range of methods applied to the domain adaptive synapse detection problem in microwasp brain connectomes. The main goal is to provide a platform and a standardized benchmark for comparison of synapse detection methods under the out-of-domain setting. We establish a low entry barrier by producing electron microscopy volumes with manual annotations and enabling automatic evaluation on CodaLab platform. The dataset contains 14 image volumes from 3 different microwasp brain samples, exhibiting significant domain difference that is suitable for evaluating the generalization ability of participants' algorithms. A baseline using a two-step approach is also proposed to serve as a reference method for participants. Additionally, in this paper, we also discuss the quantitative and qualitative results of the challenge entries as well as the limitations and potential impact of the challenge itself. It is our hope that the WASPSYN challenge can help researchers in connectmoics field take a step further in disentangling the wiring diagram of the brain.

Acknowledgment

References

[1] C. J. Peddie and L. M. Collinson, "Exploring the third dimension: Volume electron microscopy comes of age," *Micron*, vol. 61, pp. 9–19, Jun. 2014.

[2] K. L. Briggman and D. D. Bock, "Volume electron microscopy for neuronal circuit reconstruction," *Current Opinion Neurobiol.*, vol. 22, no. 1, pp. 154–161, Feb. 2012.

[3] J. Kornfeld and W. Denk, "Progress and remaining challenges in high-throughput volume electron microscopy," *Current Opinion Neurobiol.*, vol. 50, pp. 261–267, Jun. 2018.

[4] C. J. Peddie et al., "Volume electron microscopy," *Nature Rev. Methods Primers*, vol. 2, no. 1, pp. 1–23, 2022.

[5] D. Kleinfeld et al., "Large-scale automated histology in the pursuit of connectomes," *J. Neurosci.*, vol. 31, no. 45, pp. 16125–16138, Nov. 2011.

[6] J. A. Bae et al., "Functional connectomics spanning multiple areas of mouse visual cortex," *BioRxiv*, 2021.

[7] A. Shapson-Coe et al., "A connectomic study of a petascale fragment of human cerebral cortex," *BioRxiv*, 2021.

[8] S. M. Plaza, L. K. Scheffer, and D. B. Chklovskii, "Toward large-scale connectome reconstructions," *Current Opinion Neurobiol.*, vol. 25, pp. 201–210, Apr. 2014.

[9] D. B. Chklovskii, S. Vitaladevuni, and L. K. Scheffer, "Semi-automated reconstruction of neural circuits using electron microscopy," *Current Opinion Neurobiol.*, vol. 20, no. 5, pp. 667–675, Oct. 2010.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[11] K. Lee, N. Turner, T. Macrina, J. Wu, R. Lu, and H. S. Seung, "Convolutional nets for reconstructing neural circuits from brain images acquired by serial section electron microscopy," *Current Opinion Neurobiol.*, vol. 55, pp. 188–198, Apr. 2019.

[12] A. Motta, M. Schurr, B. Staffler, and M. Helmstaedter, "Big data in nanoscale connectomics, and the greed for training labels," *Current Opinion Neurobiol.*, vol. 55, pp. 180–187, Apr. 2019.

[13] L. Manubens-Gil et al., "BigNeuron: A resource to benchmark and predict performance of algorithms for automated tracing of neurons in light microscopy datasets," *Nature Methods*, vol. 20, no. 6, pp. 824–835, Jun. 2023.

[14] J. Buhmann et al., "Automatic detection of synaptic partners in a whole-brain Drosophila electron microscopy data set," *Nature Methods*, vol. 18, no. 7, pp. 771–774, Jul. 2021.

[15] Z. Lin, D. Wei, J. Lichtman, and H. Pfister, "PyTorch connectomics: A scalable and flexible segmentation framework for EM connectomics," 2021, *arXiv:2112.05754*.

[16] A. Kreshuk et al., "Automated detection and segmentation of synaptic contacts in nearly isotropic serial electron microscopy images," *PLoS ONE*, vol. 6, no. 10, Oct. 2011, Art. no. e24899.

[17] C. Becker, K. Ali, G. Knott, and P. Fua, "Learning context cues for synapse segmentation," *IEEE Trans. Med. Imag.*, vol. 32, no. 10, pp. 1864–1877, Oct. 2013.

[18] V. Jagadeesh, J. Anderson, B. Jones, R. Marc, S. Fisher, and B. S. Manjunath, "Synapse classification and localization in electron micrographs," *Pattern Recognit. Lett.*, vol. 43, pp. 17–24, Jul. 2014.

[19] A. Kreshuk, U. Koethe, E. Pax, D. D. Bock, and F. A. Hamprecht, "Automated detection of synapses in serial section transmission electron microscopy image stacks," *PLoS ONE*, vol. 9, no. 2, Feb. 2014, Art. no. e87351.

[20] S. M. Plaza, T. Parag, G. B. Huang, D. J. Olbris, M. A. Saunders, and P. K. Rivlin, "Annotating synapses in large EM datasets," 2014, *arXiv:1409.1801*.

[21] Z. Lin et al., "Two stream active query suggestion for active learning in connectomics," in *Proc. 16th Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 103–120.

[22] P. J. Schubert et al., "SyConn2: Dense synaptic connectivity inference for volume electron microscopy," *Nature Methods*, vol. 19, no. 11, pp. 1367–1370, Nov. 2022.

[23] S. Dorkenwald et al., "Automated synaptic connectivity inference for volume electron microscopy," *Nature Methods*, vol. 14, no. 4, pp. 435–442, Apr. 2017.

[24] A. Kreshuk, J. Funke, A. Cardona, and F. A. Hamprecht, "Who is talking to whom: Synaptic partner detection in anisotropic volumes of insect brain," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. Cham, Switzerland: Springer, 2015, pp. 661–668.

[25] L. Heinrich, J. Funke, C. Pape, J. Nunez-Iglesias, and S. Saalfeld, "Synaptic cleft segmentation in non-isotropic volume electron microscopy of the complete Drosophila brain," in *Proc. 21st Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Granada, Spain. Cham, Switzerland: Springer, 2018, pp. 317–325.

[26] G. B. Huang, L. K. Scheffer, and S. M. Plaza, "Fully-automatic synapse prediction and validation on a large data set," *Frontiers Neural Circuits*, vol. 12, p. 87, Oct. 2018.

[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012.

[28] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015.

[29] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2961–2969.

[30] J. Donahue et al., "DeCAF: A deep convolutional activation feature for generic visual recognition," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 647–655.

[31] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Jan. 2009.

[32] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, Oct. 2018.

[33] V. M. Patel, R. Gopalan, R. Li, and R. Chellappa, "Visual domain adaptation: A survey of recent advances," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 53–69, May 2015.

[34] H. Guan and M. Liu, "Domain adaptation for medical image analysis: A survey," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 3, pp. 1173–1185, Mar. 2022.

[35] A. Farahani, S. Voghoei, K. Rasheed, and H. R. Arabnia, "A brief review of domain adaptation," in *Proc. ICDATA*, 2020, pp. 877–894.

[36] H. Yan, Y. Ding, P. Li, Q. Wang, Y. Xu, and W. Zuo, "Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 2272–2281.

[37] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7167–7176.

[38] M. Ghifary, W. B. Kleijn, M. Zhang, D. Balduzzi, and W. Li, "Deep reconstruction-classification networks for unsupervised domain adaptation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 597–613.

[39] F. Pizzati, R. D. Charette, M. Zaccaria, and P. Cerri, "Domain bridge for unpaired image-to-image translation and unsupervised domain adaptation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 2979–2987.

[40] Y. Li et al., "X-Ray2EM: Uncertainty-aware cross-modality image reconstruction from X-ray to electron microscopy in connectomics," 2023, *arXiv:2303.00882*.

[41] A. A. Polilov, "The smallest insects evolve anucleate neurons," *Arthropod Struct. Develop.*, vol. 41, no. 1, pp. 29–34, Jan. 2012.

[42] A. A. Makarova, A. A. Polilov, and D. B. Chklovskii, "Small brains for big science," *Current Opinion Neurobiol.*, vol. 71, pp. 77–83, Dec. 2021.

[43] A. A. Polilov, *At the Size Limit Effects of Miniaturization in Insects*. Cham, Switzerland: Springer, 2016.

[44] A. A. Polilov, A. A. Makarova, S. Pang, C. Shan Xu, and H. Hess, "Protocol for preparation of heterogeneous biological samples for 3D electron microscopy: A case study for insects," *Sci. Rep.*, vol. 11, no. 1, pp. 1–8, Feb. 2021.

[45] G. Knott, H. Marchman, D. Wall, and B. Lich, "Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling," *J. Neurosci.*, vol. 28, no. 12, pp. 2959–2964, Mar. 2008. [Online]. Available: https://www.jneurosci.org/content/28/12/2959

[46] C. S. Xu et al., "Enhanced FIB-SEM systems for large-volume 3D imaging," *eLife*, vol. 6, May 2017.

[47] C. S. Xu, S. Pang, K. J. Hayworth, and H. F. Hess, "Transforming FIB-SEM systems for large-volume connectomics and cell biology," in *Volume Microscopy*. Cham, Switzerland: Springer, 2020, pp. 221–243.

[48] N. J. Chua et al., "A complete reconstruction of the early visual system of an adult insect," *Current Biol.*, vol. 33, no. 21, pp. 4611–4623, Nov. 2023.

[49] S. Saalfeld, A. Cardona, V. Hartenstein, and P. Tomančák, "CATMAID: Collaborative annotation toolkit for massive amounts of image data," *Bioinformatics*, vol. 25, no. 15, pp. 1984–1986, Aug. 2009.

[50] T. Zhao, D. J. Olbris, Y. Yu, and S. M. Plaza, "NeuTu: Software for collaborative, large-scale, segmentation-based connectome reconstruction," *Frontiers Neural Circuits*, vol. 12, p. 101, Nov. 2018.

[51] W. T. Katz and S. M. Plaza, "DVID: Distributed versioned image-oriented dataservice," *Frontiers Neural Circuits*, vol. 13, p. 5, Feb. 2019.

[52] L. Maier-Hein et al., "Metrics reloaded: Recommendations for image analysis validation," *Nature Methods*, vol. 2024, pp. 1–18, Feb. 2024.

[53] D. F. Crouse, "On implementing 2D rectangular assignment algorithms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 4, pp. 1679–1696, Aug. 2016.

[54] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.

[55] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, pp. 424–432.

[56] K. Lee, J. Zung, P. Li, V. Jain, and H. Sebastian Seung, "Superhuman accuracy on the SNEMI3D connectomics challenge," 2017, *arXiv:1706.00120*.

[57] N. L. Turner, K. Lee, R. Lu, J. Wu, D. Ih, and H. S. Seung, "Synaptic partner assignment using attentional voxel association networks," in *Proc. IEEE 17th Int. Symp. Biomed. Imag. (ISBI)*, Jul. 2020, pp. 1–5.

[58] J. Wu, W. M. Silversmith, K. Lee, and H. S. Seung, "ChunkFlow: Hybrid cloud processing of large 3D images by convolutional nets," *Nature Methods*, vol. 18, no. 4, pp. 328–330, Apr. 2021.

[59] Q. Chen, W. Huang, Y. Zhang, and Z. Xiong, "Domain adaptive synapse detection with weak point annotations," in *Proc. IEEE 21st Int. Symp. Biomed. Imag. (ISBI)*, Jun. 2024.

[60] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2414–2423.

[61] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1501–1510.

[62] W. Huang et al., "Semi-supervised neuron segmentation via reinforced consistency learning," *IEEE Trans. Med. Imag.*, vol. 41, no. 11, pp. 3016–3028, Nov. 2022.

[63] T. Y. Lin, P. Dollàr, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2117–2125.

[64] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[65] M. Wiesenfarth et al., "Methods and open-source toolkit for analyzing and visualizing challenge results," *Sci. Rep.*, vol. 11, no. 1, p. 2369, Jan. 2021, doi: 10.1038/s41598-021-82017-6.

[66] A. Kirillov et al., "Segment anything," 2023, *arXiv:2304.02643*.

[67] Y. Cheng, F. Wei, J. Bao, D. Chen, and W. Zhang, "ADPL: Adaptive dual path learning for domain adaptation of semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 8, pp. 9339–9356, 2023, doi: 10.1109/TPAMI.2023.3248294.

[68] L. Hoyer, D. Dai, H. Wang, and L. Van Gool, "MIC: Masked image consistency for context-enhanced domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 11721–11732.

[69] S. Azizi, S. Kornblith, C. Saharia, M. Norouzi, and D. J. Fleet, "Synthetic data from diffusion models improves ImageNet classification," 2023, *arXiv:2304.08466*.