SynQuant: An Automatic Tool to Quantify Synapses from Microscopy Images

Yizhi Wang[1,*], Congchao Wang[1,*], Petter Ranefall[2], Gerard Broussard[3,4], Yinxue Wang[1], Guilai Shi[3], Yue Wang[1], Lin Tian[3], and Guoqiang Yu[1,†]

[1]Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, USA, [2]Centre for Image Analysis and SciLifeLab, Uppsala University, Uppsala, Sweden. [3]Department of Biochemistry and Molecular Medicine, University of California Davis School of Medicine, USA. [4]Princeton Neuroscience Institute, Princeton University, United States.

* These authors contributed equally to this work. † To whom correspondence should be addressed.

Abstract

Motivation: Synapses are essential to neural signal transmission. Therefore, quantification of synapses and related neurites from images is vital to gain insights into the underlying pathways of brain functionality and diseases. Despite the wide availability of synapse imaging data, several issues prevent satisfactory quantification of these structures by current tools. First, the antibodies used for labeling synapses are not perfectly specific to synapses. These antibodies may exist in neurites or other cell compartments. Second, the brightness for different neurites and synapses is heterogeneous due to the variation of antibody concentration and synapse-intrinsic differences. Third, images often have low signal to noise ratio (SNR) due to constraints of experiments and availability of sensitive antibodies. The combination of these issues makes the detection of synapses challenging and necessitates developing a new tool to accurately and reliably quantify synapses.

Results: We present an automatic probability-principled synapse detection algorithm and integrate it into our synapse quantification tool SynQuant. Derived from the theory of order statistics, our method controls the false discovery rate and improves the power of detecting synapses. Through extensive experiments on both synthetic and real images in the presence of severe antibody diffusion, high heterogeneity, and large noise, our method was demonstrated to outperform peer specialized synapse detection tools as well as generic spot detection methods by a large margin. Finally, we show SynQuant reliably uncovers statistically significant differences between disease and control conditions in a neuron-astrocyte co-culture based model of Down Syndrome.

Availability: The Java source code, Fiji plug-in, and test data are available at https://github.com/yu-lab-vt/SynQuant.

Contact: yug@vt.edu

1 Introduction

The synapse is a critical structure in the nervous system that enables communication and interaction between neurons. Cognitive function hinges on proper wiring of synaptic connections within neural circuitry. Quantification of synapses, including synapse detection and synapse feature extraction, is thus an indispensable component of today's brain research. By measuring the properties of synaptic puncta such as density and size, as well as neurite properties like length and shape, under different phenotypes, researchers could gain insights into how brains function under normal and abnormal conditions. However, thorough and consistent analysis of synapses is a pre-requisite to this process of discovery (Lin and Anthony, 2010; Myers, 2012; Ullian et al., 2011).

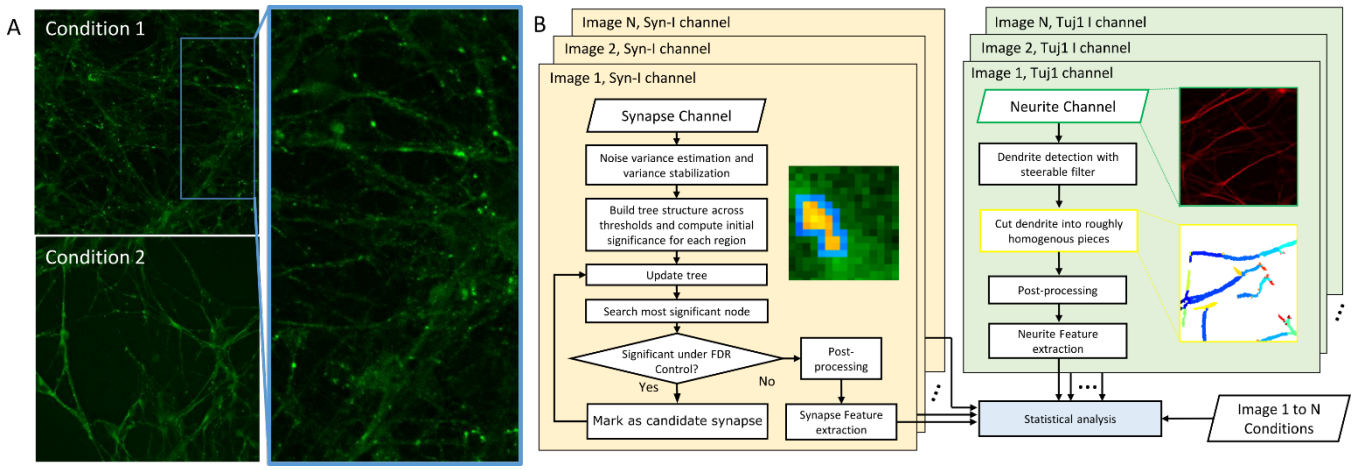


Figure 1. (A) Synapsin I labeled channel acquired by confocal microscopy. Green dots are potential synaptic puncta. Two images on the left panel belong to control (top) and case (bottom) groups, respectively. (B) Flowchart of SynQuant. Left panel: the synapse detection algorithm. The inset figure shows the choice of neighborhood pixels (blue) for a region (yellow pixels). Right panel: neurite features. Top sub figure shows the Tuj1 stained neurite channel. The bottom one shows the local homogenous cuts on the traced neurites. The extracted puncta and neurite features from N images are used for statistical analysis between conditions.

There are two main challenges in analyzing synapse images (Fig. 1A). First, different neurites and synapses show significant variations in terms of morphology and brightness. One reason for this is the inherent variation among neurons and neurites according to the role they play and the discrepancies in maturity. Second, localization of proteins of interest within synaptic puncta is not typically perfect. One possible reason is that there is actually Synapsin I at low concentrations present in the neurites, which will show a low level of positive staining. Another possibility is that staining procedures usually result in some amount of "non-specific" staining. This occurs when the antibody containing the fluorophore attaches to something other than the protein of interest. Besides, the bleach from the neurite channel is also a possible contributor, but this cannot be easily corrected by blind source separation (Supplemental Text S5).

As a result, this diffuse, non-homogenous signal interferes with synapse detection. For example, even the signal to noise ratio is high for some synapses, it could be much lower for many others in the same data set. The brighter synapses are more likely to be picked up, but this will introduce bias to the analysis. The non-specific antibodies make it hard to identify synapses purely based on intensity. Some diffused signals could even be stronger than some synapses. Therefore, the combination of synapse-intrinsic heterogeneity, imperfect protein localization to synapses, along with potentially low SNR, lead to great challenges in accurately and reliably detecting, segmenting, and quantifying synapses.

Synapse detection has been an active research topic in the recent years and quite a few methods were developed (Feng et al., 2012; Danielson and Sang, 2014; Schmitz et al., 2011; Simhal et al., 2017, Simhal et al., 2018). SynD (Schmitz et al., 2011) detected synapses by filtering pixels with a single intensity threshold and separating overlapped regions using an averaged synapse kernel. Feng et al. (2012) proposed BGM3d to separate clustered 3D synaptic puncta using a Bayesian Gaussian mixture model, which also used a single threshold to segment puncta from the background. Danielson and Sang (2014) segmented synapses within dendrites with multiple thresholds, but their method requires users to mark dendrite regions manually. Recently, Simhal et al. (2017) proposed a probabilistic method for quantifying synapse and Simhal et al.

(2018) designed an analysis framework based on it. But they did not consider either the imperfect antibody staining or the synapse heterogeneity issues.

Many image analysis tools for subcellular localization and spot detection have the potential to be repurposed to detect synapses. Coelho et al. (2013) tried to find the puncta-like nucleus or nucleoli within a cell by segmenting cells first. However, puncta do not necessarily lie within the stained portion of neurites. Zhang et al. (2007) proposed Multiscale Variance Stabilization Transform (MS-VST) to detect spots based on isotropic undecimated wavelet transform (IUWT) along with a Poisson Gaussian model. But IUWT also responds strongly to edges and tends to generate a large number of false positives around the neurite-like signals for our data. Maximum possible h-dome (MPHD) (Rezatofighi et al., 2012) can handle the neurite-like region; however, the proper *h* value is hard to set to get good regional maxima. MS-VST and MPHD were shown to be the best among the unsupervised spot detection methods (Smal et al., 2010).

We find the performance of existing algorithms is far from satisfactory, with either high rates of false positives or false negatives. For example, the rich patterns of synapse and non-specificity antibody reduce the performance for wavelet-based methods; morphology/thresholding based methods do not work well under inhomogeneous background; lack of reliable training data makes it hard to use supervised methods. More importantly, most of them cannot provide a rigorous statistical foundation to assess their output regions and thus no reliable method to distinguish true synapse from noises. Besides, the inhomogeneity is not considered in the quantification tools and the comparison between images under different conditions are not well calibrated.

In this work, we develop a probability-principled synapse detection method that considers the signal non-specificity, heterogeneity, and large noise. Then we integrate it into our quantification tool (SynQuant) that extracts neurites and synaptic features (Fig. 1B). To address the signal non-specificity and heterogeneity, we develop a model that is adaptive to localized region properties. If a region is a synapse, it is expected to be brighter than its surroundings, even though in the same image there may be brighter non-synaptic background regions that do not surround the synapse. Here are two major analytical problems: (1) how to choose the neighborhood pixels for localized modeling and (2) how to evaluate the difference between a candidate region and its surroundings, considering some difference may be purely due to noise. The choice of neighborhood pixels is crucial. For example, for a region inside the neurite, pixels in the non-neurite background which is much darker or in the other puncta which may be brighter should not be used as neighbors. The difference cannot be solely evaluated based on intensities contrast, because it ignores the number of pixels participating in the comparison: the more pixels, the more reliable the contrast is. Further, although the conventional *t*-test between a group of pixels and their neighbors can integrate the information from intensity contrast and number of pixels, the model is severely biased. The operation of choosing a candidate region and its neighbors has already implied that the candidate region is brighter than its surroundings.

Based on the reasoning above, SynQuant contains two key components. First, we propose to use order statistics (David and Nagaraja, 2003) to properly utilize the local information of puncta and fairly compare all synapse candidates. For a given candidate region, SynQuant integrates information from the average intensity inside the region, the average intensity of its neighbors, the size and ranking of all pixels in these two parts and their noise variance. The theory of order statistics provides a powerful tool to correct the bias introduced by the candidate choosing operation. To the best of our knowledge, this is the first time that the inherent bias for synapse detection has been rigorously modeled. Indeed, we suspect that

unawareness of the right model for the inherent bias was a major reason for the lack of rigorous statistical model in the field of synapse detection. Second, we propose an iterative updating strategy to identify appropriate neighbors of the synapse candidates for assessing their statistical significane. By this strategy, we will detect the smallest regions retaining statistical significance, which are more likely to be the synaptic puncta. In addition, our method uses the p-value reported by order statistics to control the false discovery rate, which can be pre-specified by the user. To make the software package comprehensive, we extract neurites by a steerable filter (Meijering et al, 2004) and cut them into roughly homogeneous pieces. For each neurite piece, their positions, neurite features, and corresponding synapse features are gathered.

Experiments show that our quantification framework obtains a large accuracy gain of synapse detection on both simulated and real image data. In addition, we build a regression model to study the relationship among those piece-level features and brain conditions or disease phenotypes. Since we gather the information based on each piece, the impact of inhomogeneity is modeled as a confounding factor. In addition, by removing the confounding factors, relationships between synapse density and disease phenotypes are reliably uncovered.

2 Method

We first estimate the noise model parameters and stabilize the noise variance of the image (Fig. 1B, left panel). After that, we choose candidate synapses by binarizing the image with multiple thresholds. Each threshold leads to some binary connected components, or regions (Fig. 2B). We build a tree structure where each region becomes a node. Each region is assigned an initial significance using order statistics. We iteratively search for possible synaptic puncta in the tree. Once we find the satisfactory node, we remove it. The remaining most significant region will be selected as a new candidate synapse. The significance of relevant regions is updated. The threshold to determine a synapse is controlled by the user-specified FDR threshold. The synaptic features are extracted after post-processing. The related neurite attributes are collected from the neurite channel after neurite tracing and are cut into segments (Fig. 1B, right panel). The features are then used for statistical analysis along with the disease phenotypic information of each image.

2.1 Noise estimation and variance stabilization

Application of order statistics theory requires the noise statistics of the pixels in a candidate region and its neighbor. Conventionally, the noise is modeled as following a Gaussian distribution which simplifies subsequent computations. However, the photon detector introduces noise whose variance is linearly dependent on the signal intensity. We apply the noise model proposed by Foi et al. (2008). The variance for pixel (i, j) is modelled as

$$\mathrm{var}(y_{i,j}) = a x_{i,j} + b. \qquad (1)$$

Here $\mathrm{var}(y)$ is the pixel noise variance. x is the underlying signal intensity, which is not observed but can be well approximated by the observed pixel intensity. The term ax models the Poisson type noise and the term b models the additive Gaussian noise. The model can be fit based on pixel data from a single image and the resulting a and b are used in the Anscombe transform to stabilize the noise (Foi et al., 2008), so that the noise variance associated with the new values after the transform is independent to the intensity itself and can be approximated by a single constant σ_{stab}^2.

2.2 Synapse significance score based on order statistics

In our adaptive tree search and updating algorithm, for each threshold, we get a set of isolated regions (nodes in the tree), each containing a set of pixels. These regions are potential candidates for synaptic puncta that need to be evaluated by statistical tests. The test for the individual region is based on the difference of this region and its neighbor pixels. A larger difference implies a larger possibility that this region is significantly different from the surroundings, which is a necessary (but not sufficient) condition for being a synapse. For each region, a group of neighbor pixels is selected. We assume there are M pixels $S = \{x_1, \dots, x_M\}$ in the region and N pixels $P = \{x_{M+1}, \dots, x_{M+N}\}$ in the neighbor. We may use a *t*-test to compare these two groups. However, due to the thresholding operation, all the M pixels must have higher intensities than the N neighbors. Even if there is no true signal, positive difference always exists between the means of the two groups for any candidate region considered. This positive difference is a bias and, if not corrected, will complicate the detection and result in a lot of false detections. Here, we are still interested in the difference between the candidate region and its neighbor pixels, and define the test statistic as the following,

$$L = \frac{x_1 + \cdots + x_M}{M} - \frac{x_{M+1} + \cdots + x_{M+N}}{N}, \qquad (2)$$

where $x_1 \geq \cdots \geq x_{M+1} \geq \cdots \geq x_{M+N}$. But the statistical significance is computed based on the theory of order statistics. As mentioned above, due to the thresholding operation, even without a true signal, L will be positive, and the exact value is determined by the noise variance, the sample size and the ratio of M and N. The theory of order statistics provides a formal approach to account for the bias by calculating the mean and variance of L under the null hypothesis that there is no true signal between the candidate region and its neighbor pixels. Let $n = M + N$, we can rewrite L as in (David and Nagaraja, 2003):

$$L = \frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) x_i. \qquad (3)$$

Here, $J(k)$ is a weight function corresponding to the coefficients for x_i in Eq. (2). For $1 \leq i \leq M$, $J(i/(n+1)) = n/M$, and for $M + 1 \leq i \leq M + N$, $J(i/(n+1)) = n/N$. We write

$$\mu(J, F) = \int_0^1 J(u) F^{-1}(u) du, \qquad (4)$$

and

$$\sigma^2(J, F) = \iint_{0<u_1<u_2<1} \frac{2J(u_1)J(u_2)u_1(1-u_2)}{f(F^{-1}(u_1))f(F^{-1}(u_2))} du_1 du_2. \qquad (5)$$

Then we have $E(L) = \mu(J, F)/\sqrt{n}$ and $var(L) = \sigma^2(J, F)/n$, when $n = M + N \to \infty$ (David and Nagaraja, 2003). Here f is the normal probability density function with zero mean and variance estimated as above. F^{-1} is the corresponding inverse normal cumulative distribution function, which depends on the stabilized noise variance σ_{stab}^2. The integration is computed by summation using all the n samples. Then we define the order statistic score z as a function f_{os}:

$$z \coloneqq f_{os}(S, P, \sigma_{stab}^2) = \frac{L - \mu(J,F)}{\sigma^2(J,F)}, \qquad (6)$$

where z follows asymptotically a standard Gaussian distribution and hence can be easily used to compute the statistical significance of any observed value of L.

We note that the statistical significance computed above is an accurate approximation only when the sample size is large enough, which may not be the case. With some typical image resolutions, one synapse may only contain about 10 or fewer pixels. Here we apply two corrections for the small sample size to improve the approximation. First, we notice for the double integration in $\sigma^2(J, F)$, the integration space is a triangle defined by $0 < u_1 < u_2 < 1$. Since we are using discrete samples, the boundary points will noticeably impact the integration results when the sample size is small. Therefore, half of the boundary points are incorporated in the integration and the other half are not.

Second, the integration over J is based on a uniform grid, which corresponds to the x values. However, the boundary points x_1 and x_n strongly deviate from this uniform assumption and the results will be affected when the sample size is small. We would like the integration to mimic the summation. Therefore, we compute the distribution of the largest (or smallest) sample and use the mean to get a new grid. This mean value d is computed by

$$d = 1 - F\big(E(x_1)\big) = 1 - F(n \int_0^1 F^{-1}(t) t^{n-1} dt). \qquad (7)$$

Here t should be densely sampled from 0 to 1. Then we get a new grid $[d, \dots, d + (i-1)(1-2d)/(n-1), \dots, 1-d]$.

2.3 Iterative detection, FDR control, and post-processing

Our iterative detection and segmentation scheme is driven by the statistical significance of each region as computed above (Algorithm 1 and Fig. 2). Assuming the image is stored in 8 bits, we threshold the image with all intensity values (0 to 255). For each threshold $thr \in \{0,1, \dots, 255\}$, we binarize the image I and get all connected regions. Suppose we have K regions. The set of all such regions is denoted by $V = \{S_1, \dots, S_K\}$. We may simply denote S_k as k, then $V = \{1, \dots, K\}$. We will build a tree T, whose nodes are V. We use E to denote the edge set describing the way to connect nodes (regions) in V.

Now each node k is associated with the region S_k, along with the threshold t_k under which it is generated. Then the directed edge set is defined as $E \coloneqq \{(i,j) | S_j \subseteq S_i, t_j = t_i + 1\}$, which means we link region i to a region j that is completely within it. However, not all inside regions should be linked. We link regions whose associated (more stringent) threshold is $t_i + 1$ (Fig. 2B). This structure is similar to conjunctive Bayesian networks (Beerenwinkel et al., 2007) and shares the similar principle as Mattes et al. (1999).

Each node k is also related to a neighbor pixel set P_k and a score z_k from order statistics. Since the computation of order statistics depends on the choice of neighbor pixels, z_k depends on P_k. Recalling Eq.6, we have $z_k = f_{os}(S_k, P_k, \sigma^2_{stab})$. On one hand, P_k should include neighbor pixel of S_k and thus will be within an ancestry node of k, which is defined by the tree and denoted as $An(k)$. The number of pixels in P_k needs to be carefully specified. If P_k is too large, many pixels far away from the candidate region S_k will be included and thus the comparison is not restricted to the local area. If P_k is too small, we lose the statistical power to assess the significance of the candidate region. We find that requiring P_k to have a similar size as the candidate region S_k is a good balance. In practice, we specify the neighbor region P_k by growing the candidate region S_k layer by layer until P_k is larger than S_k. On the other hand, not all neighbor pixels of S_k should be

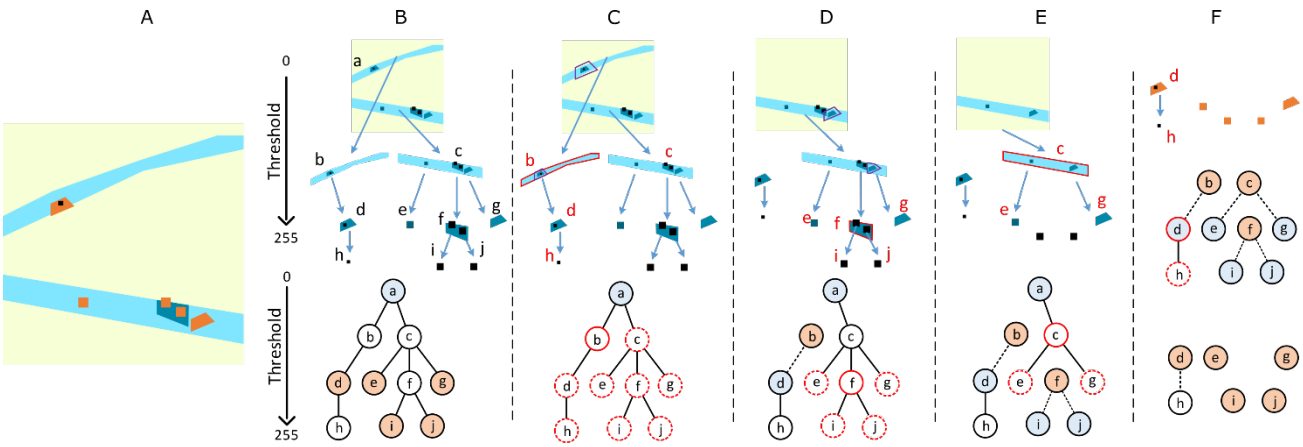


Figure 2. Joint synapse detection and segmentation by iterative tree searching and updating. ***(A)*** *Illustration for an image with neurite (light blue) and puncta (orange). The green background and black dots are both noises from the perspective of synapse detection.* ***(B)*** *Tree structure based on thresholding. Top: the original image is the root node a (Thr=0). Two branches (b and c) are the children of a with a higher Thr. Repeat this process, we get other nodes and edges. Bottom: tree representation. Blue nodes are the roots and orange ones are the synapses to be detected.* ***(C)*** *b is the current most significant node (red solid circle). The significance of all its descendants d and h, along with all nodes sharing the same ancestry with b are updated (red dashed circles). E.g., the neighborhood of d was originally chosen within a, but now they were chosen within b (purple boxes in a and b).* ***(D)*** *d becomes the root of a tree and b is the candidate region. As f is the most significant one now, i, j, g, e are chosen to be updated.* ***(E)*** *Now we have four trees with a, d, i, j as roots. Repeat this with node c.* ***(F)*** *Continue this process and we get the puncta. There are 5 significant synapses detected, d,e,g,i and j, respectively. Even though b, c and f are statistically significant regions, they are disqualified as synapses because they have children that are statistically significant. For the region d, it has a child h, but the region h is not statistically significant, so the region d remains as a synapse.*

included in P_k even though these pixels are close to S_k, because these pixels may belong to another synapse region. Yet, a synapse region should have a significant score. Therefore, we require P_k should not include any pixel of a significant region. Hence, P_k also depends on z_k as the significance of regions is determined by z_k, which leads to the iterative scheme as described below.

Our algorithm iteratively updates P_k and z_k for each node n on tree T. We initialize the root node (k $= 1$, whole image) as the candidate region. For all other nodes, we initialize $z_k = 0$. All the other nodes now choose neighbor pixels P_k within the image (Fig. 2B) and do not need to avoid any pixels, because there is no significant region. Based on the choice of P_k, we update z_k for all nodes (except for the root). Then we search for the most significant node k and update P_k for all the descendants of $An(k)$, except those that are already significant (Fig. 2C). After that, node k is removed from the tree as a candidate synapse and its children will become the new root of a new tree (Fig. 2D). Again, the updated P_k will give us new z_k. In later iterations, once any descendants of k becomes a new candidate, k is disqualified as a synapse. This drives the algorithm to avoid neurite-like structures (Fig. 2E-F).

As the mean and variance of order statistics under null hypotheses are given by Eq.3 and 4, we can calculate the scores of all relevant regions (Ω_{updt}). We pick the one with the highest score and we need to determine whether we can add it to the list of significant regions (Ω_{sig}). We want to keep the false discovery rate (FDR) lower than a given threshold among all synapses detected. The threshold (FDR level α) is a parameter specified by the user. A typical value is 0.05, which make the false discovery rate lower than or equal to 5%. In each iteration, we learn from the FDR control procedure whether

adding the newly selected region to the list of existing significant regions (Ω_{sig}) will make the FDR lower than the threshold. Because overlapped regions may be correlated, we use the general case introduced by Benjamini and Yosef, (1995). Details are given in Supplemental Text S7. The total number of iterations depends on the number of synapses (significant regions) in the image and the user-specified FDR threshold. The algorithm stops when no more candidate region is determined as statistically significant. During the algorithm, no new region will be generated. The region marked as significant will not be updated later. Therefore, the maximum possible number of iterations is the number of regions from all thresholds and the algorithm is guaranteed to terminate.

Several rules based on the prior knowledge of the size and relative positions of synapses are applied to post-process the synapse candidates found by the algorithm. For example, a synapse should not be too large or too long. Otherwise, it is likely to be areas with elevated background intensity, such as neurites. These characteristics can be simply measured by puncta's scale, aspect ratio and filling property (Uijlings et al., 2013). The remaining candidates are reported as synaptic puncta. The features of those detected puncta, such as size, brightness, and position, are then extracted.

Algorithm 1 Order statistics based iterative detection and segmentation

INPUT: Tree $T = \{V, E\}$ with K nodes, σ_{stab}^2, FDR level α, root node $1 \in V$

INIT: Significant nodes: $\Omega_{sig} \leftarrow \{1\}$. Candidate region: $\Omega_{cand} \leftarrow \{1\}$. Nodes to update: $\Omega_{updt} \leftarrow \{2,3,\dots,K\}$. Nodes not for neighbor pixels: $\Omega_{avoid} \leftarrow \phi$. Scores: $Z \leftarrow \{+\infty, -\infty, \dots, -\infty\}$.

WHILE 1

  FOREACH $k \in \Omega_{updt}$ // update neighbourhood pixels and score

    c $\leftarrow argmax_c\{t_c | c \in An(k), c \in \Omega_{sig}\}$ // t_c: region c threshold

    $X \leftarrow \{i | i \in S_c, i \notin S_k, k \in \Omega_{avoid}\}$ // S: pixels in region c

    $P_k \leftarrow \{i | i \in X, dist(i, S_k) < \epsilon\}$ // pixels close enough to S_k

    $z_k \leftarrow f_{os}(S_k, P_k, \sigma_{stab}^2)$

  END

  $z_{FDR} \leftarrow FDR_control(Z, \alpha)$ // corrected z-score threshold

  $j \leftarrow \text{argmax}_k\{z_k | z_k > z_{FDR}, k \notin \Omega_{sig}\}$ // significant node

  IF $j = \phi$ // no significant region remains, stop

    BREAK

  END

  $\Omega_{sig} \leftarrow \Omega_{sig} \cup \{j\}$

  $\Omega_{cand} \leftarrow \{j\} \cup \Omega_{cand} \setminus An(j)$ // disqualify ancestry as candidate

  $\Omega_{updt} \leftarrow \{k | t_k \geq t_j, k \neq j, k \sim j\}$ // ~: undirected graph reachability

  $\Omega_{avoid} \leftarrow \{k | k \in \Omega_{sig}, t_k > t_j\}$

  $V \leftarrow V \setminus \{j\}$

  $E \leftarrow E \setminus \{\{(j,k) | Parent(k) = j\} \cup (Parent(j), j)\}$

END

OUTPUT: Puncta index Ω_{cand} and regions $\{S_k | k \in \Omega_{cand}\}$

The iterative detection algorithm is implemented on a component tree structure (Najman et al., 2006), which makes it possible to create the tree in linear time instead of iterating over all threshold levels. When nodes are merged during the creation of the tree, features for test statistics of each node are recalculated the same way as in Ranefall et al. (2016).

2.4 Neurite quantification and statistical analysis

We use a steerable filter (Meijering et al, 2004) to trace the neurites. Unlike SynD (Schmitz et al., 2011), we relax the restriction that neurite must proceed from the soma to reduce false negatives for images with a large field of view. Details are given in Supplemental Text S11. This relaxation is desirable because such an image contains many neurons and some somas are not necessarily captured in it. Each detected neurite is cut into pieces. The intensity inside one piece is relatively homogeneous by cutting at branch points. Features including position, length, scale and mean intensity for those pieces are extracted.

Each neurite piece has a set of synaptic puncta located on it or near it. In addition, as we discussed before, quite a few features of those puncta can be extracted. Therefore, we can build a model which uses each neurite piece along with those puncta on or near it as an input sample. This is a local feature analysis which is more informative than aggregating the feature across images. Also, for each image, we know the phenotype related to it, such as disease or normal conditions. Putting all these together, we can analyze the image with local synapse, neurite, and phenotypic features and discern the underlying associations.

3 Results

We first tested SynQuant on both simulation and real data and compared it with four representative methods: SynD (Schmitz et al., 2011), MS-VST (Zhang et al., 2007), MPHD (Rezatofighi et al., 2012) and FDA (Smal et al., 2010). More comparison can be found in the supplementary information. Then, experiments were performed to measure the influence of Down syndrome on synaptic density while taking all confounding factors (extracted features) into consideration. We implemented the algorithm in Java as a Fiji plugin (Schindelin, Johannes, et al., 2012).

3.1 Synapse detection and segmentation on simulation data

We evaluated the performance by true positive rate (TPR) - modified false positive rate (FPR*) curve and its associated F-score. TPR is defined as TP/(TP+FN) and FPR* as FP/(TP+FN), where TP is the true positive, FP is the false positive, and FN is the false negative. FPR* is commonly used in spot detection problems since we cannot get true negative.

In spot detection problems, true positives (TP) are defined as those instances where the distance between the center of ground truth location and the detected spot is less than a threshold (Smal et al., 2010). This is not sufficient here since we need an accurate estimation of both puncta position and region to quantify their features. Hence, we use Intersection-over-Union (IoU) to infer TP, which is widely used in object detection problems. If the overlap of ground truth and the detected region is larger than 50% of their union, the detected punctum is viewed as a TP. IoU is more suitable when we want to jointly evaluate the detection and segmentation performance. The F score is $2 \times \text{precision} \times \text{recall}/(\text{precision} + \text{recall})$. The precision and recall are also defined by setting a 50% threshold on IoU.

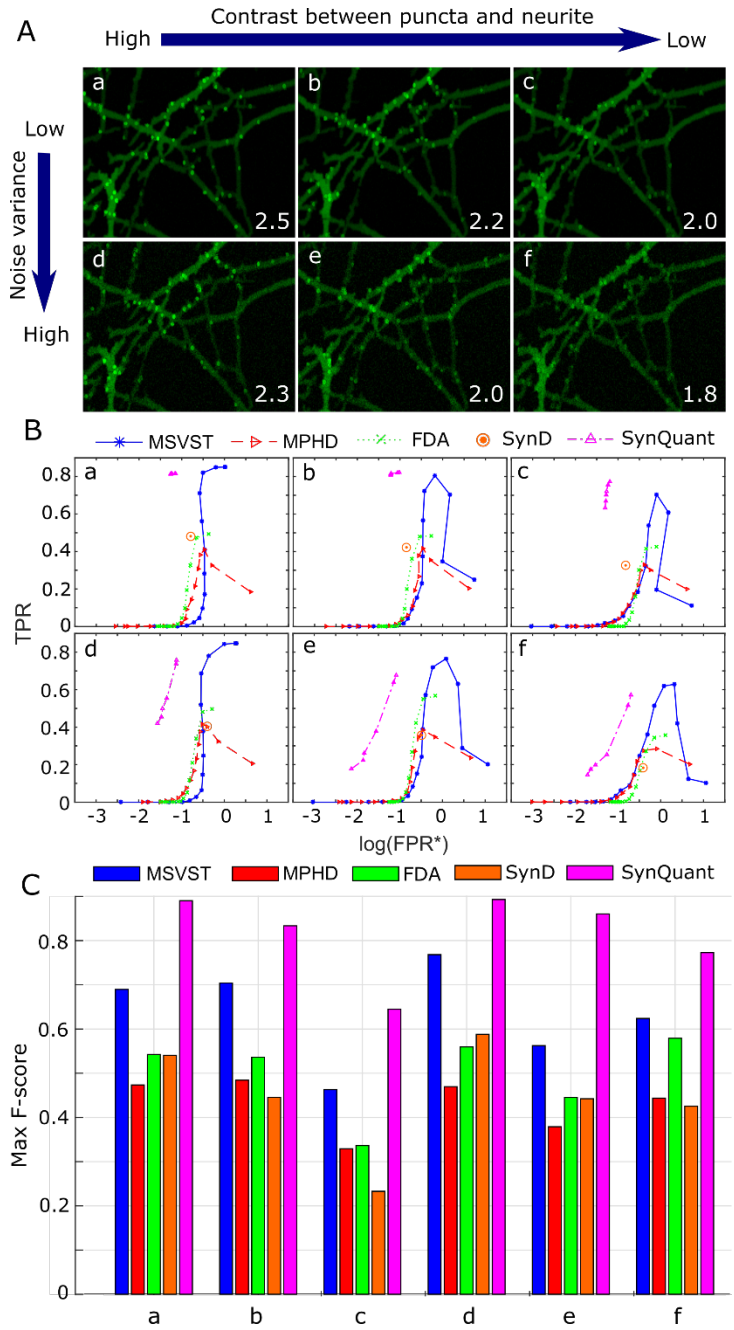


Figure 3. Simulation data and results. ***(A)*** *Six synthetic images. Synapse is shown as puncta with various sizes, brightness, and shapes. The signal to noise ratio (SNR) ranges from 1.8 dB to 2.5 dB (bottom right of each subplot), which is typical in microscopy images. Contrasts between synapses and neurites decrease from the left column to the right one. Noise variance increases from top row to bottom row. Since we have different contrast, the resulting SNR is different for different columns.* ***(B)*** *TPR-FPR* curve of five methods on simulation data. For images with low noise level (a-c), the parameters' impact on SynQuant is not obvious. Comparing with the other four methods, SynQuant always has many times less FPR* with the same TPR.* ***(C)*** *Maximum F-score comparison on the six images. SynQuant outperforms other methods on all the six images.*

Our simulation data consists of both synapse and neurite like signals to mimic realistic data (Supplemental Text S3). We simulated neurite-like signals by putting the segments of traced branches in the Tuj1 labeled channel in a real data set into the blank simulation image. After that, the image was smoothed by Gaussian filter. Then synapses are randomly cast on

the simulated dendrite with various shapes (either roughly circle or eclipse but always less than a 7×7 patch). The synapse intensity is assumed to be greater than the corresponding dendrite. Therefore, the contrast between synapse and its surrounding pixels is controlled by this intensity difference. Finally, Poisson-Gaussian noise was added. Examples of simulation data are given in Fig. 3A. We generate extra synthetic data sets and use the extracted features and ground truth labels to train the FDA model.

We compared the TPR-FPR* curves of these methods with different simulation parameters (Fig. 3B). SynQuant consistently outperforms peer methods. As our metric is based on IoU, too small or too large detected puncta will both lead to false positive. This makes the curves no longer monotonic. The maximum F-score from each of the six synthetic images with different methods is shown in Fig. 3C. For MS-VST and MPHD, the maximum F-scores were obtained by changing the signal threshold in the output signal map. For FDA, the output is a binary classification map and performance is evaluated with a threshold on size. For SynQuant, we changed the FDR significance level to test its performances. We found SynQuant performs best for all the six images with a large margin. The precision-recall curves are given in the Supplemental Text S8.

The simulation with pure noise (no synapses) shows t-test based score of regions is biased while order-statistics based method control the false positives (Supplemental Text S1). The order statistics are derived using vectors, but they are valid when applied to images by comparing the null distributions (Supplemental Text S6). Besides, we found a combination of the algorithms could be used, but it may not improve the results (Supplemental Text S9). Comparisons using location accuracy only is given in Supplemental Text S10.

3.2 Synapse detection on real data

We applied SynQuant and three unsupervised methods to images from a neuron-astrocyte co-culture. Supervised methods like FDA were not compared, because they require manual annotation of synapses which is time-consuming and may introduce bias. The image includes three channels, the nucleus channel (blue) is stained with Hoescht, the synapse channel (green) is labeled by Synapsin I and the neurite channel (red) is labeled with Tuj1. Here we utilize the green channel to detect putative pre-synaptic puncta. The red channel was used for neurite tracing and feature extraction.

The detection and segmentation results are shown in Fig. 4A. The puncta are drawn in red and overlapped with the green channel. The neurite mask (red channel) is used to filter out some puncta but it is not used in the detection. Under default settings, SynQuant detected 783 puncta with FDR level of 0.2. Compared with MS-VST, SynD, and MPHD, SynQuant produced less false positives and false negatives. SynD reported many more false positives and MPHD has higher false negatives. MS-VST performed better than SynD and MPHD, but still missed many puncta even under the optimal choice of threshold (see red circles, Fig. 4A). Since the features (used in Eq. 6) for p-value calculation are accumulative, the tree structure shown in Fig.2 can be built efficiently with adaptive local thresholding (Ranefall et al. 2016). Updating the nodes' order statistics scores consumes most of the time.

Generally, SynQuant takes ~20 seconds to detect and segment synapses on an image whose resolution is 512×512 pixels. All experiments are performed on a workstation with Xeon E5-2630 CPU. Besides, the accuracy of noise variance estimation is discussed in Supplemental Text S2. Results for the comparison to Simhal et al. (2017) on real data is given in Supplemental Text S12. It does not model the non-specificity and cannot distinguish synapses on or near neurites.

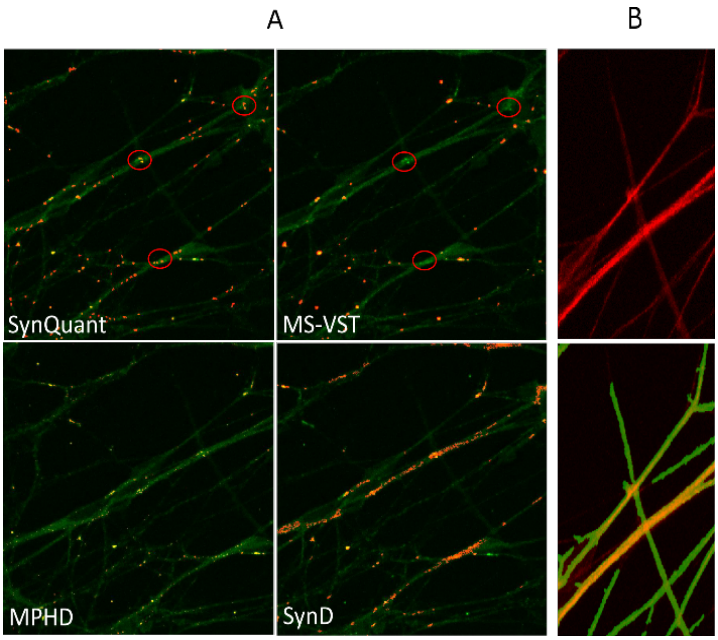


Figure 4. Performance on real data (A) Performance comparison of four methods on real data. The detected synapses are marked as red and overlapped on the green Syn-1 channel. MS-VST performs better than MPHD and SynD. Compared with SynQuant, MS-VST still has many puncta (red circles) missed. MS-VST can recall missing puncta by setting a lower threshold to its output score map, but generate a large number of false positives as well. (B) Neurite tracing results. Top: neurite channel. Bottom: traced neurite.

3.3 Synapse quantification

Conventionally, synapse density is defined as the number of synapses per unit length of the neurite. This is biased in an inhomogeneous image and does not utilize other neurite and synapse features. In this experiment, we consider the synapse and neurite features on the segment level. From the synapse channel, synapse features like mean intensity, size and position are extracted. Neurite tracing results from the neurite channel are shown in Fig. 4B. Neurites are then cut into pieces by their branch points and end points. In one image, we get hundreds of pieces and each of them is approximately homogenous inside. Each synapse is assigned to a corresponding neurite piece based on the distance.

We apply Poisson regression to investigate the relationship between disease status and synapse density using five groups of samples (Table 1). We use the neurite features as confounding factors. DS1 and DS2 are from co-culture samples containing two isogenic Down syndrome (DS) astroglial lines. DS1 astrocytes exhibit trisomy of chromosome 21, the defining hallmark of Down Syndrome. In contrast, the DS2 line, while derived in a manner similar to the DS1 line, instead exhibits disomy of chromosome 21. The DS2 line is thus expected to be more similar genetically to cells derived from non-DS sources. HA is a co-culture sample with human prenatal primary astrocytes. NO are samples from neuron monocultures. The control group is neurons co-cultured with astrocytes differentiated from normal induced pluripotent stem cells (iPSCs). Since we use binary coding for these four groups in the regression model, the control group is the baseline and is not shown in Table 1. Length, scale, and intensity are neurite features. First, we only used neurite length as a confounding factor (Table 1, column 2 and 3). We found that the NO, HA, and DS2 groups all show significant influence on synapse density. Then we used all three neurite features as confounding factors (Table 1, column 4 and 5). Under these analyses, only HA

and NO groups had significant coefficients and DS1 and DS2 did not have an obvious impact on synapse density. From the comparison in Table 1, we can see that quantification of synaptic density based only on synapse number and neurite length is prone to assigning a higher degree of statistical significance to all the different phenotypes. The same conclusion can be obtained when we use the size of synapses as a dependent variable (Supplemental Text S4).

The lack of astrocytes in NO group clearly shows the big impact of astrocytes on the formation of synapses (Boehler et al. 2007). Astrocytes in HA group plays important role in synapse formation as well (Reemst et al. 2016, Clarke et al. 2013), but these cells are less mature than the astrocytes in the control group. Therefore, it reduces the synapse density compared with the control group, but the reduction is less severe than the NO group. DS2 group is actually similar to non-DS groups, and it is expected to show no significant impact on synapse density once the confounding factors are controlled. The significant association with neurite length is expected as longer neurite is likely to harbor more synapses, however, the length only cannot account for all neurite heterogeneity. Previous studies have revealed diminished synapse formation and reduced dendritic ramification in infants with DS (Becker et al. 1991; Garner et al. 2012) and lower density of excitatory synapse also were observed in mouse models of DS (Kurt et al. 2004). However, currently, no studies have been able to give a measurement of synaptic density in human brains with DS. As shown here, without proper consideration of confounding factors, biased quantifications can result in misguided conclusions.

Table 1. Poisson regression using Down syndrome phenotypes and three neurite features as predictors and synapse puncta density as the response

Features	Neurite length only		All neurite features	
	Estimated Coefficients	**P-value**	**Estimated Coefficients**	**P-value**
Length	0.248	0	0.173	1.62e-34
Scale	N/A	N/A	0.079	9.43e-09
Intensity	N/A	N/A	0.263	1.45e-152
DS1	0.044	0.235	-0.058	0.115
DS2	0.156	**4.43e-06**	-0.060	0.086
HA	-0.366	**1.60e-25**	-0.287	**3.96e-16**
NO	-0.748	**1.16e-69**	-0.655	**1.26e-53**

4 Discussion

We have presented a new automatic synapse quantification framework (SynQuant) for segmentation and quantification of heterogeneous and noisy images of synapses and dendrites. Most existing spot and synapse detection methods fail under these conditions and the widely used (blind) linear un-mixing approaches work poorly and introduce artifacts. SynQuant is able to detect and segment synaptic puncta accurately. It can extract comprehensive features from both synapses and neurites and analyzes their relationships. Some extra functions supported are given in Supplemental Text S12.

The superior performance of SynQuant comes from the effective utilization of the local region-neighbor information. Enjoying the same principle as Hariharan et al. (2014), the probability principled iterative detection and segmentation algorithm uses the tree structure of regions to choose the correct neighborhood pixels. Order statistics provide an unbiased score to indicate the likelihood of each candidate region is a true synapse. The choice of neighborhood pixels and the computation of order statistics are iterated until no statistically significant regions can be found in the tree structure.

Compared with existing spot detection methods, SynQuant is able to extract accurate segmentation results, which allows access to important features for synapse studies. The main parameter needed for SynQuant is the FDR level to be controlled and the algorithm reports the P-value for each region detected.

Incorporated with properties extracted from the neurite channel, effects of cell types and disease phenotypes on synapse density can be revealed. With a large field of view, the number of neurons involved is large. As a result, different neurons and their corresponding neurites and synapses are at different development stages due to cell type and other stochastic effects. As such, these neurites and synapses must be treated locally and we should not aggregate the features in the image. In SynQuant we use the segmented neurite pieces as a sample in the regression analysis. Thus, the relationships we found are more comprehensive by including these local confounding factors and can more reliably reveal the relationship between disease and synapse features.

As we have the neurite channel as a reference, it is possible to either subtract the neurite-like background in synapse channel or use un-mixing techniques to separate these two channels (Neher et al., 2009; Zimmermann, 2005). However, the non-linear relationship between two channels makes blind un-mixing inapplicable (Supplemental Text S5).

SynQuant makes it possible to use antibodies that are not so specific. Most peer methods require a highly specific antibody. In some experiments, this is possible, as shown in peer methods, but this is not feasible in many experiment setups. In fact, to study synapse heterogeneity, the requirement of antibody specificity is even higher, which may limit the freedom to design experiments.

SynQuant can be extended to analyze 3D data. In addition, using both the pre-synaptic puncta and post-synaptic puncta, we can further improve the performance. Finally, SynQuant is a general framework to analyze images with a high level of non-specificity, we can adapt and apply it to bio-medical images beyond synapse staining, such as particle detection for the particle tracking problem.

References

Becker, L., Mito, T., et al. (1991). Growth and development of the brain in Down syndrome. Progress in Clinical and Biological Research, 373, 133–152.

Beerenwinkel, N., et al. (2007). Conjunctive Bayesian networks. *Bernoulli*, 893-909.Benjamini, Y., and Yosef H. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B* (Methodological): 289-300.

Bjornsson, C., et al. (2008) Associative image analysis: a method for automated quantification of 3D multi-parameter images of brain tissue. *Journal of neuroscience methods*, **170**, 165-178.

Boehler, MD., et al. (2007) "Added astroglia promote greater synapse density and higher activity in neuronal networks." Neuron glia biology 3.2: 127-140.

Clarke, L. E., and Barres, B. A. (2013). Emerging roles of astrocytes in neural circuit development. Nature reviews. Neuroscience, **14**(5), 311-21.

Coelho, L., et al. (2013) Determining the subcellular location of new proteins from microscope images using local features. *Bioinformatics* **29**, 2343-2349.

Danielson, E., and Sang L. (2014) SynPAnal: Software for Rapid Quantification of the Density and Intensity of Protein Puncta from Fluorescence Microscopy Images of Neurons. *PloS one* **9**, e115298.

David, Herbert Aron, and Haikady Navada Nagaraja. *Order statistics* 3rd edition. New York: Wiley-Interscience, 2003.

Feng, L. et al. (2012) Improved synapse detection for mGRASP assisted brain connectivity mapping. *Bioinformatics* **28**, i25-i3

Fish, K., et al. (2008) An automated segmentation methodology for quantifying immunoreactive puncta number and fluorescence intensity in tissue sections. *Brain research* **1240**, 62-72.

Fletcher, T., et al. (1991). The distribution of synapsin I and synaptophysin in hippocampal neurons developing in culture. *The Journal of Neuroscience*, **11**, 1617-1626.

Foi, A, et al. (2008) Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *Image Processing, IEEE Trans.*, **17**, 1737-1754

Garner, Craig C., and Daniel Z. (2012). Wetmore. Synaptic pathology of Down syndrome. Synaptic Plasticity. Springer Vienna. 451-468.

Hariharan, B., et al. (2014) Simultaneous detection and segmentation. In *Computer vision–ECCV*, Springer International Publishing, 297-312

Heck, N., et al. (2015) A new automated 3D detection of synaptic contacts reveals the formation of cortico-striatal synapses upon cocaine treatment in vivo. *Brain Structure and Function* **220**, 2953-2966.

Kurt, M. A., Kafa, M. I., et al. (2004). Deficits of neuronal density in CA1 and synaptic density in the dentate gyrus, CA3 and CA1, in a mouse model of Down syndrome. Brain Research, 1022 (1–2), 101–109.

Lin, Y., and Anthony K. (2010) Mechanisms of synapse and dendrite maintenance and their disruption in psychiatric and neurodegenerative disorders. *Annual review of neuroscience* **33**, 349.

Mattes, J., et al. (1999) Tree representation for image matching and object recognition. In *Discrete Geometry for Computer Imagery*, Springer Berlin Heidelberg, 298-309

Meijering E., et al. (2004) Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. *Cytometry A* **58**, 167–76.

Najman L. and Couprie M. (2006) Building the Component Tree in Quasi-Linear Time, IEEE Transactions on Image Processing, 15(11), 3531–3539.

Neher, R., et al. (2009) Blind source separation techniques for the decomposition of multiply labeled fluorescence images. *Biophysical Journal*, **96**, 3791-3800.

Myers, G. (2012) Why bioimage informatics matters. *Nature methods* **9**, 659-660.

Ranefall P., et al. (2016) Fast Adaptive Local Thresholding Based on Ellipse fit, Proceedings of the International Symposium on Biomedical Imaging (ISBI'16), Prague, Czech Republic.

Reemst, K., et al. (2016) "The indispensable roles of microglia and astrocytes during brain development." Frontiers in human neuroscience **10**, 566.

Rezatofighi, S., et al. (2012) A new approach for spot detection in total internal reflection fluorescence microscopy. *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, 860-863.

Smal, I., et al. (2008) A new detection scheme for multiple object tracking in fluorescence microscopy by joint probabilistic data association filtering. In *Biomedical Imaging: From Nano to Macro, 2008 IEEE International Symposium on*, IEEE, 264–267.

Smal, I., et al. (2010) Quantitative comparison of spot detection methods in fluorescence microscopy. *Medical Imaging, IEEE Trans.*, **29**, 282-301.

Sanders, J., et al. (2015) Learning-guided automatic three dimensional synapse quantification for drosophila neurons. *BMC bioinformatics* **16**, 1.

Schindelin, J., et al. (2012) Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676-682.

Schmitz, S., et al. "Automated analysis of neuronal morphology, synapse number and synaptic recruitment." *Journal of neuroscience methods* **195** (2011): 185-193.

Simhal, AK., et al. "Probabilistic fluorescence-based synapse detection." PLoS computational biology 13.4 (2017)

Simhal, AK., et al. "A Computational Synaptic Antibody Characterization Tool for Array Tomography." Frontiers in neuroanatomy 12 (2018).

Ullian, E., et al. (2001) Control of synapse number by glia. *Science* **291**, 657-661.

Uijlings, J., et al. (2013) Selective search for object recognition. *International journal of computer vision* **104**, 154-171.

Zhang, B., et al. (2007) Multiscale variance-stabilizing transform for mixed-Poisson-Gaussian processes and its applications in bioimaging. *Image Processing, 2007 IEEE International Conference on*, VI-233-VI-236.

Zimmermann, T. (2005). Spectral imaging and linear unmixing in light microscopy. In *Microscopy techniques*. Springer Berlin Heidelberg, 245-265

Supplementary material to "SynQuant: An Automatic Tool to Quantify Synapses from Microscopy Images"

S1. Synapse detection and segmentation on pure noise simulation

We compare Student's t-test with SynQuant (with order statistics) when evaluating the significance between a region and its neighborhood pixels (Supplemental Table 1). We first simulated an image of size 1024×1024 where all pixels have the same intensity. Then additive Gaussian noise with zeros mean and unit variance is introduced. The threshold varied from 1.2 to 1.8 with step size 0.2. Ideally, no synapse should be detected. First, we assume the noise variance is known. We count the total number of regions given by the thresholds and the number of candidates that are significant under the FDR level of 0.05. Totally we get 2574 candidates. SynQuant only reported 12 significant candidates while t-test generated 1310 false positives. So, t-test produces an inflated false positive in thresholding-based methods (Supplemental Table 1). We also observed that the control of the false positives for our method is conservative due to the correlation of scores between overlapped regions. Then we repeated the experiment with a poorly estimated noise variance (0.5 vs true 1.0). Then both tests generated many false positives. This confirms the importance of noise variance estimation in evaluating the significance level for order statistics.

Supplemental Table 1. The false positive rate on data with pure noise

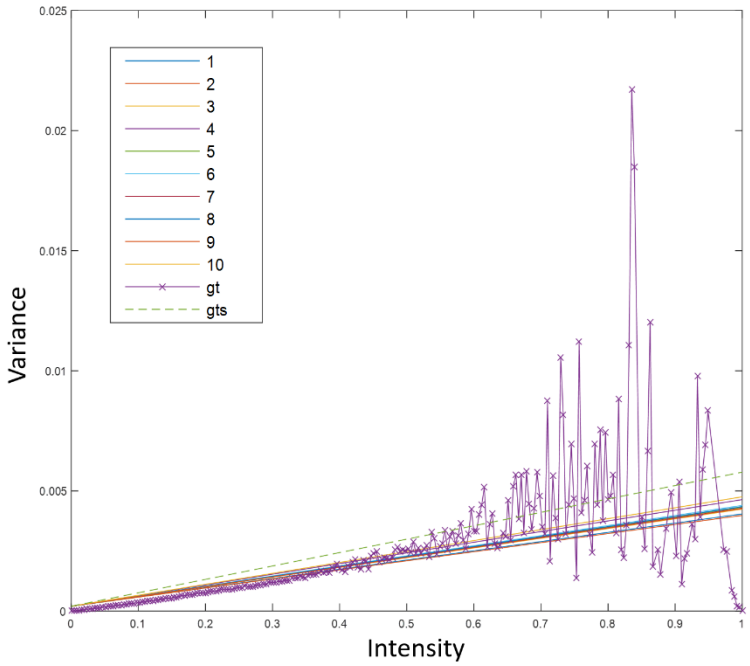
Noise variance used	***t*-test**	**Order Statistics**
True noise variance	0.543	0.0047
0.5 (a poor estimate)	0.539	0.424

S2. The accuracy of noise variance estimation

The noise of microscopy images consists of two parts: photon noise and camera readout noise. Therefore, it can be modeled by a Mixed-Poisson-Gaussian process (Zhang et al., 2007). The accuracy for the noise variance estimation is essential for the remaining steps in our algorithm. Usually, the Poisson part plays a more important role. Therefore, we first show the noise variance estimation accuracy on our microscopy images.

The model is fitted through the algorithm designed by Foi et al. (2008), though we do not consider the clipping and saturation effects. The outcome is a linear relationship between mean intensity and noise variance. Ground truth is obtained as follows. The image of one sample in the experiment is repeatedly taken for ten times under the identical conditions. Each pixel is observed 10 times. Then we can calculate the noise variance for each pixel and plot the relationship between its mean intensity and variance (curve 'gt' in Supplemental Figure 1).

From every single image, we estimated the noise and compared them with the ground truth. The comparison is shown in Supplemental Figure 1. It's clear that the estimated noise model is quite similar to the true value. Note that the linear regression line 'gts' in the figure is biased by the pixels with large means and variances. There are a relatively small amount of such pixels and are less reliable. On contrary, the results from single image estimation (line 1 to 10) are closer to the ground truth (curve 'gt').

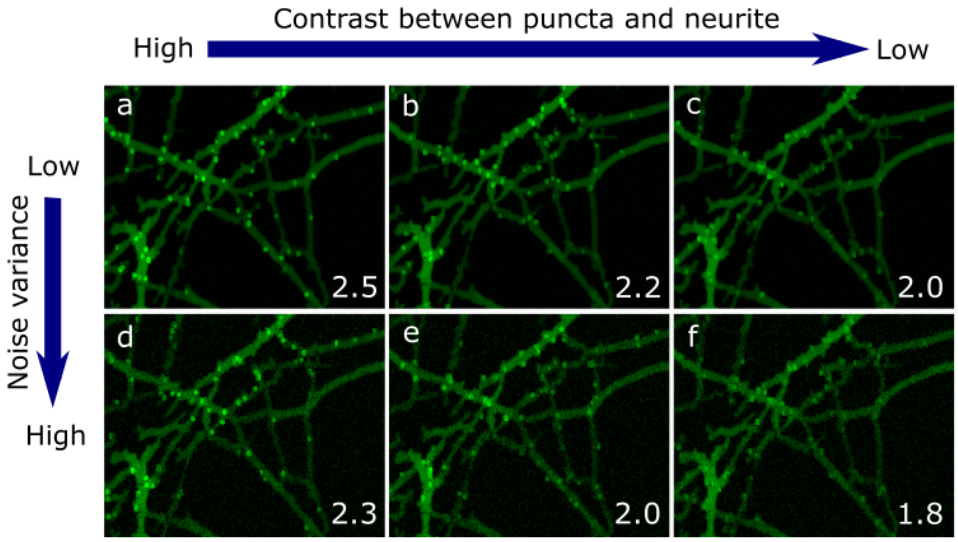


Supplemental Figure 1. Noise variance estimation results. Line 1-10 are 10 noise models estimated from 10 single images. gt and gts are the ground truth from 10 images and the linear regression result, respectively.

S3. Simulation data generation

The simulation data is generated with four steps. To fully simulate the conditions of neurite, we first extract the neurite from real data (neurite extraction method will be introduced in Supplemental Text S11). Based on the results of extraction, we cut them into small pieces (see Figure 1B in the main body). Pixels in each piece are viewed roughly homogeneous. The second step is generating neurite pieces on our simulation image at the same positions, with scales and brightness similar to the ones on the real image. For each neurite piece, we use a Gaussian kernel to smooth it. The third step is to cast synapses on each neurite piece. The number of synapses is based on the relationship between neurite features and synapse numbers obtained from real data. The synapse intensity is decided by the summation of three values: intensity of the neurite it falls on, a contrast parameter and a random noise value. The contrast parameter is used to define average contrast across the image between synapses and neurites they fall on. The random value is generated by multiplying the contrast parameter with a standard Gaussian random variable. Finally, Mixed-Poisson-Gaussian noise is added to the image.

We generate our simulation data in three scenarios as is shown in Supplemental Figure 2. Each scenario contains two images in one column with the same average intensity contrast between neurite and synapse. But their noise variances are different. Across the three scenarios, the contrast varies. With these three scenarios, it is easy for us the compare the robustness of algorithms under different noise variance and intensity contrast. Detail of parameters is also illustrated in the figures.



Supplemental Figure 2. The three scenarios of our simulation data. From left to right the contrast are decreased. From up to bottom, the noise variance is increased. The unit for SNR is dB.

S4. Relation test on synapse scale

Our algorithm does not separate overlapped synapses, because synapses seldom overlap in our data. It may be unfair to consider synapses with varies sizes equally under such condition. Here we use synapse's scale rather than the number as a dependent variable for the relationship test. The results are shown in Supplemental Table 2.

The results are consistent with those in the main paper. If we do not consider the confounding factors like neurite scale and intensity, the phenotype DS2 will show a significant effect on the synapse scale.

Supplemental Table 2. Poisson regression using Down syndrome phenotypes and three neurite features as predictors and synapse puncta scale as a response

Features	**Neurite length only**		**All neurite features**	
	Estimated Coefficients	**P-value**	**Estimated Coefficients**	**P-value**
Length	1.392	0	0.596	1.77e-12
Scale	N/A	N/A	0.934	1.36e-27
Intensity	N/A	N/A	1.022	3.45e-140
DS1	0.210	0.072	-0.136	0.241
DS2	0.677	**8.59e-10**	0.111	0.318
HA	-0.916	**5.95e-21**	-0.651	**2.55e-11**
NO	-1.437	**7.68e-44**	-1.091	**4.13e-26**

S5. Blind source separation for neurite-like signal removing

As we have access to both the synapse channel (green channel) and the neurite channel (red channel), we attempted to use the neurite channel to remove the neurite-like background in the synapse channel. This procedure is like blind source separation (BSS). Under the assumption that the neurite channel is mostly pure, whose signal mainly from true neurite, the synapse channel can be viewed as a mixed signal from the known neurite channel and unknown pure synapse signal. If we could extract the pure synapse signal, the difficulty of synapse detection and segmentation problem should be decreased.

For any two pixels i and j, the values in neurite channels are y_i and y_j. The values in synapse channels are x_i and x_j. If we assume signals in neurite channel go into the synapse channel, we will have $y_i = a_i x_i + b_i$ and $y_j = a_j x_j + b_j$. In order for blind source separation to work, we require $a_i = a_j$ and $b_i = b_j$. In this is true for any pixels, we say the relationship between these two channels are homogeneous. Otherwise, if $a_i \neq a_j$ for different pixels, the relationship is inhomogeneous and the blind source separation is not a valid choice.

Here we first show the inhomogeneous relationships using the real data. Then we list several BSS attempts based on regression and non-negative matrix factorization (NMF). Both the analysis and experiment results show that the synapse channel is not simply a linear combination of synapse and neurite signals and thus making the separation difficult to achieve.

Inhomogeneity of relationships between two channels

If y is the intensity of one pixel in the synapse channel, and the x is the intensity for the same pixel in the neurite channel, their relationship is given by a regression model

$$y = \alpha x + b. \tag{S1}$$

Here α is a random variable. We assume it follows Gaussian distribution $N(\mu, \sigma^2)$.

To study whether the relationship between these two channels is homogenous, we cut the neurite into small (homogeneous) pieces. The idea is that we study the relationship piece by piece. If the relationship is homogeneous, the results from difference neurite pieces should be similar.

Assume we have M pieces. Also, assume y_i is the intensity of any pixel in piece i in synapse channel and x_i is the intensity for the same pixel in the neurite channel. We have a new regression model

$$y_i = \alpha_i x_i + b_i \tag{S2}$$

For piece i, α_i consists of three components

$$\alpha_i = \mu + n_i + e_i. \tag{S3}$$

Here e_i is the noise caused by sample size (pixel number) in piece i and n_i is the term related to the position or other factors. Consider all pieces, we have the regression model for coefficient α

$$\alpha = \mu + n + e. \tag{S4}$$

If the relationship between these two channels are homogeneous, n is zero and thus all α are determined by μ and e

$$\alpha = \mu + e. \tag{S5}$$

Under the homogeneous condition, across piece variance $var(\alpha)$ should be equal to within piece variance $var(e)$, otherwise, $var(n)$ is non-zero. By calculating $var(\alpha)$ and $var(e)$ with real data, we can find whether the relationship between these two channels are homogeneous.

We stabilize synapse and neurite channels to make the noise variance un-related to intensity. Then in neurite channel, we cut neurite into small homogeneous pieces and remove those whose skeleton is less than 10 pixels. The pixels inside one piece in these two channels should be linearly related as is illustrated in Equation (S2). For each α_i, $var(\alpha_i) = \sigma_i^2$ is got from the linear regression. Taking the average of $\{\sigma_1^2 \dots \sigma_M^2\}$, we obtain $var(e)$. With $\{\alpha_1, \dots, \alpha_M\}$, we can estimate the sample variance $var(\alpha)$.

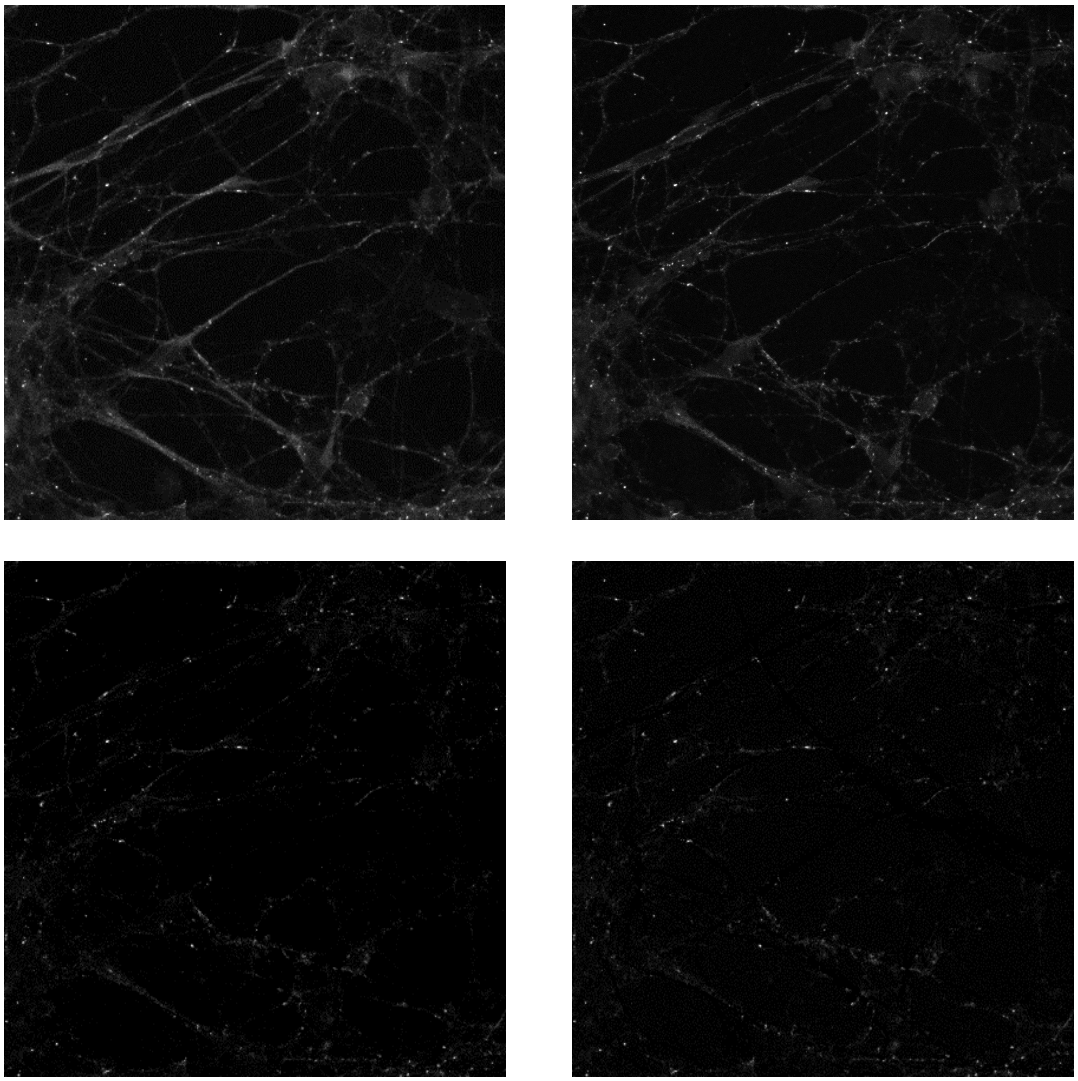
Using the two channels shown in Figure 1(B) in the main body, we get that $var(\alpha)$=0.064 and $var(e)$=0.015, which means $var(n)$ is not zero (0.049) and the relationship between these two channels are inhomogeneous. By Chi-Square test, the p-value of $var\ (\alpha)$=0.015 is 0, which means it's almost impossible to happen.

Linear regression

Here we test the relationship of neurite in these two channels. To reduce the influences of synapse signal, we remove most puncta before the test. Then linear regression is applied to the remaining pixels of the two channels. We subtract neurite channel based on the regression results, as shown in the bottom two figures of Supplemental Figure 3. Though it seems most neurite-like signals have been removed from observation, the remaining artifacts introduce more interference. The real data is manually labeled and then MSVST (Zhang et al., 2007), MPHD (Rezatofighi et al., 2012), and SynQuant are applied on the synapse channel. Results comparison is shown in Supplemental Table 3. The "Original" row shows results on the raw data and the other two rows show the results of these algorithms on the data after linear regression. FDA (Smal et al., 2010) needs training data and not applicable to a single image. SynD (Schmitz et al., 2011) detects synapse based on neurite position and thus also not applicable if no neurite exist. From the results, we can see that linear regression does not really help increase synapse detection accuracy. Also, we tried Lowess regression and segmented regression on the real data too, the results look similar to that of linear regression and introduce artifacts as well.

Non-negative matrix factorization (NMF)

NMF is a famous blind source separation technique. For our problem, we assume the synapse and neurite the channels are both composed of pure synapse and neurite signal. To apply NMF, we use the synapse and the neurite channels to form a 2 by N matrix V, where N is the number of pixels in each channel. In practice, for each channel, all horizontal rows of pixels are concatenated into a 1-D (column) vector and then be transformed into a 1-D (row) vector. These two vectors form the 2 by N matrix V. By NMF, V is factorized into a 2 by 2 matrix W and 2 by N matrix H. The two vectors in H should represent pure synapse and neurite signals. Results of NMF are shown in the right-up figure in Supplemental Figure 3. Because N >> 2, the un-mixing results are un-reliable. It's obvious that the synapse signal extracted from NMF is as noise as before.



Supplemental Figure 3. Results of BSS trials. The top-left is the original synapse channel. Top-right is the result got from NMF. It's obvious that NMF fails to handle most of the neurite-like signals. The bottom-left and bottom-right are results of linear regression with and without bias term. These two results are similar and looks like that most neurite-like signal has been removed. Our following experiments shows that this process introduces un-seen artifacts which does not truly help for detection.

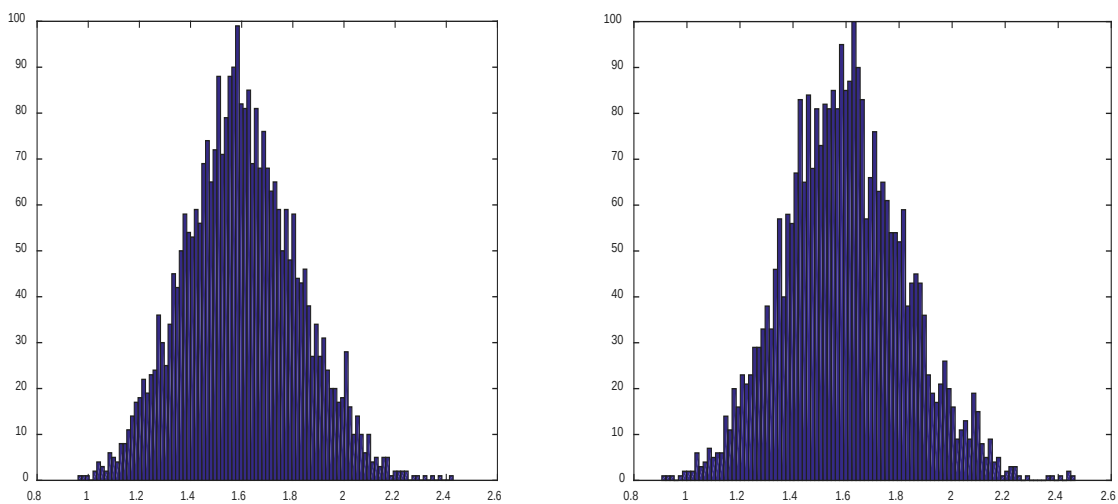
Supplemental Table 3. F-score comparison on real data

	MSVST	**MPHD**	**SynQuant**
Original	0.375	0.393	**0.744**
*y=k*x*	0.369	0.312	**0.526**
*y=k*x+b*	0.402	0.408	**0.544**

S6. Distribution from 1D and 2D null hypothesis

The distribution of the NULL hypothesis in our method is obtained by order statistics. In order statistic theory, if $\boldsymbol{x}$ is a random vector where each element follows I.I.D. Gaussian distribution, its test statistic (see Equation (2) in the main paper) follows a Gaussian distribution determined by M and N (see Equation (2) in the main paper). Here we test on images, where the signal is in 2D instead of 1D.

We generate an image that contains only Gaussian noises. Then all puncta whose intensities are larger than their surrounding pixels are chosen. Based on these puncta, we can draw the distribution of their test statistics. Supplemental Figure 4 shows this distribution, along with the distribution theoretically generated from 1D derivation. For simplicity, here we only show the distribution with 30 pixels (15 pixels with higher intensities vs. 15 neighboring pixels with smaller intensities). It's shown that for images, the order statistics derivation is also applicable.



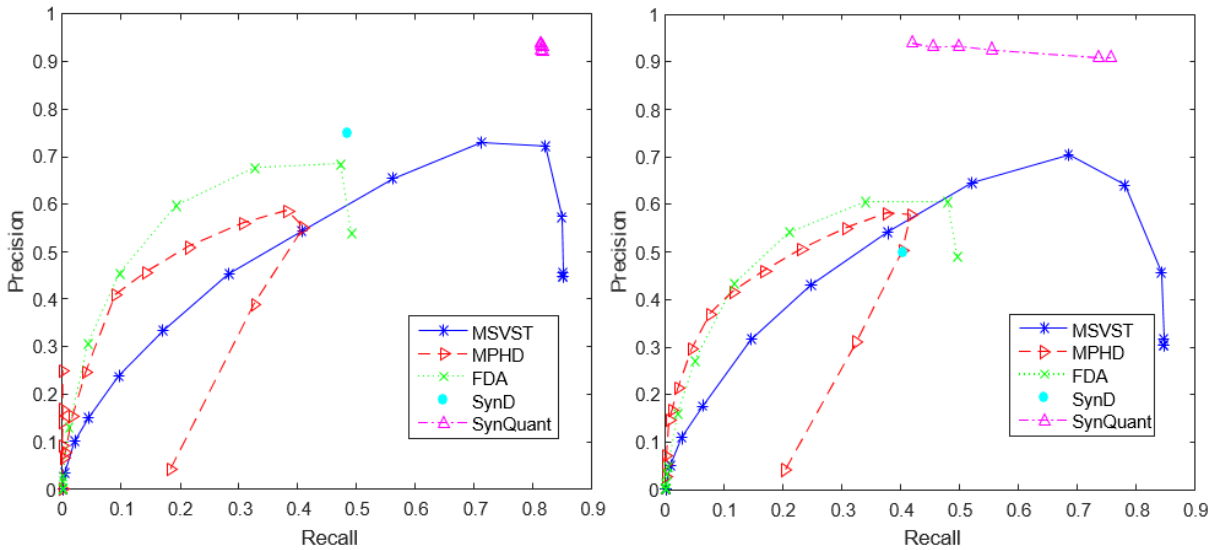
Supplemental Figure 4. Distributions of null hypothesis on image. The left one is the distribution got from one image with pure noises. The right is distribution got theoretically by generating random vectors. It's shown that order statistics is also applicable when confronted image data.

S7. FDR control

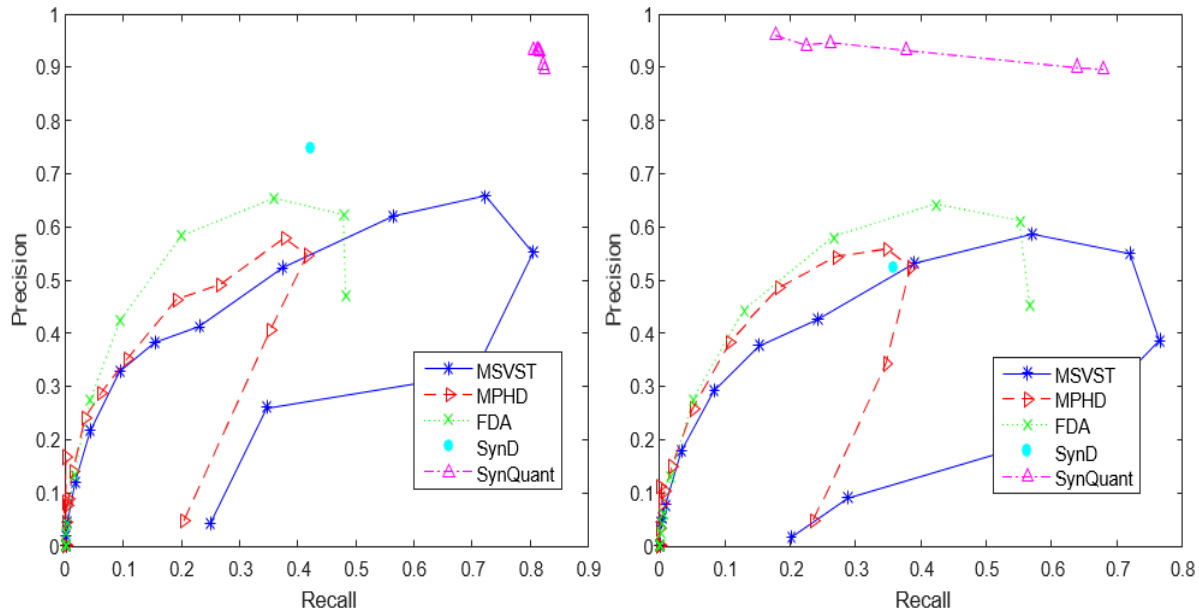
The FDR control offers a looser restriction over Bonferroni correction. We conduct FDR control in each iteration by testing whether all the un-selected nodes (regions) fail to meet the significance request. Otherwise, we pick up the most significant remaining one and continue to the next iteration until no such region exists. For example, as shown in Figure 2 (E) in the main paper, we have two candidate synapse puncta **b** and **f**. The next node with the highest significance value is **c**. Then FDR control will be conducted on all the nodes including **b** and **f** to see under such condition whether **c** is still significant. If so, set **c** as a candidate synapse punctum. Otherwise, stop and report all candidate synapse puncta we detected. Because for each iteration, the significance (order statistics score) of some nodes may change, FDR control is needed in each iteration. As overlapped regions may be highly correlated, for FDR control, we use the general case introduced by Benjamini and Yosef, (1995).

S8. Precision-recall curves for simulation

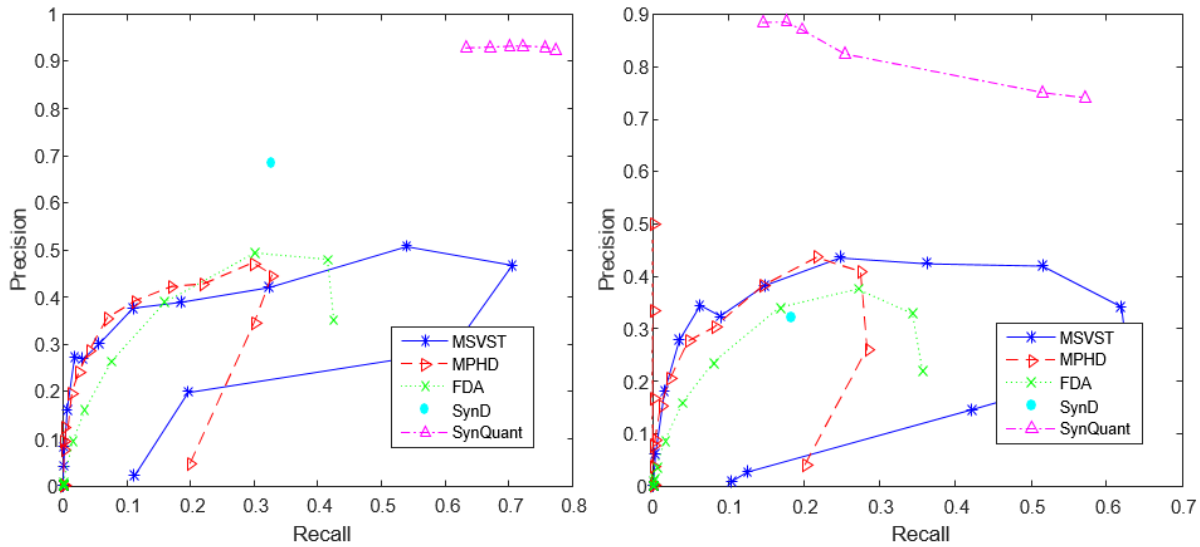
Here we give the complete precision-recall curves of SynQuant compared with the other four methods of synapse detection and segmentation (Supplemental Figure 5-7). From these curves, SynQuant always outperforms other algorithms by a large margin. With these curves, we calculate the maximum F-scores, which is shown in Figure 3 in the main paper. All the experiments are done on the six simulated images shown in Supplemental Figure 2 and the results are shown in the same order of the three scenarios there.



Supplemental Figure 5. Precision-recall curves for the algorithms in scenario 1. The left figure comes from image with lower noise. Because we use the IoU measurement, changing the parameters of algorithms may not lead to monotonic curves. From these two figures, it is obvious that SynQuant will not degrade when the noise variance is increased. For the image lower noise (left), SynQuant is not sensitive to parameter settings as is shown in the left figure.



Supplemental Figure 6. Precision-recall curves for the algorithms in scenario 2. Compared with results of scenario 1 in Supplemental Figure 5, we can see that other than SynQuant and SynD, all the other methods' precisions decrease. That's because the contrast between synapse and neurite is lower in this scenario. SynD is not largely influenced, but the performances are still far from satisfactory.

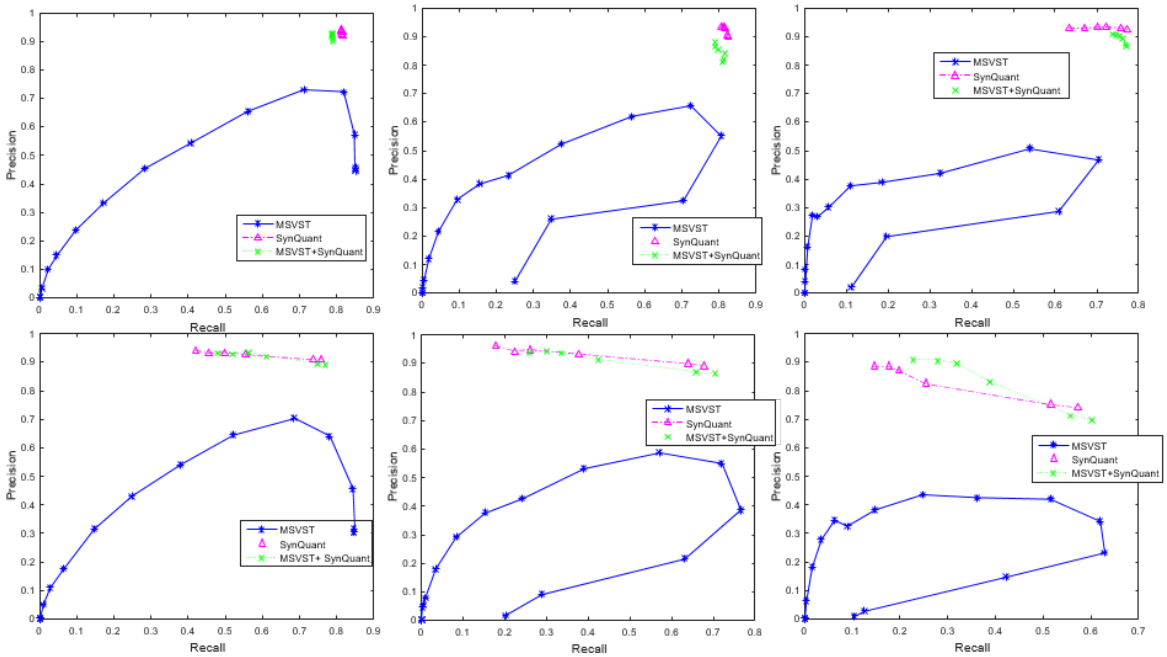


Supplemental Figure 7. Precision-recall curves for the algorithms in scenario 3. Compared with the former two scenarios, this scenario is the hardest to handle. Other than SynQuant, all other methods cannot get good performances. In spite of this, SynQuant still gets high precisions and recalls.

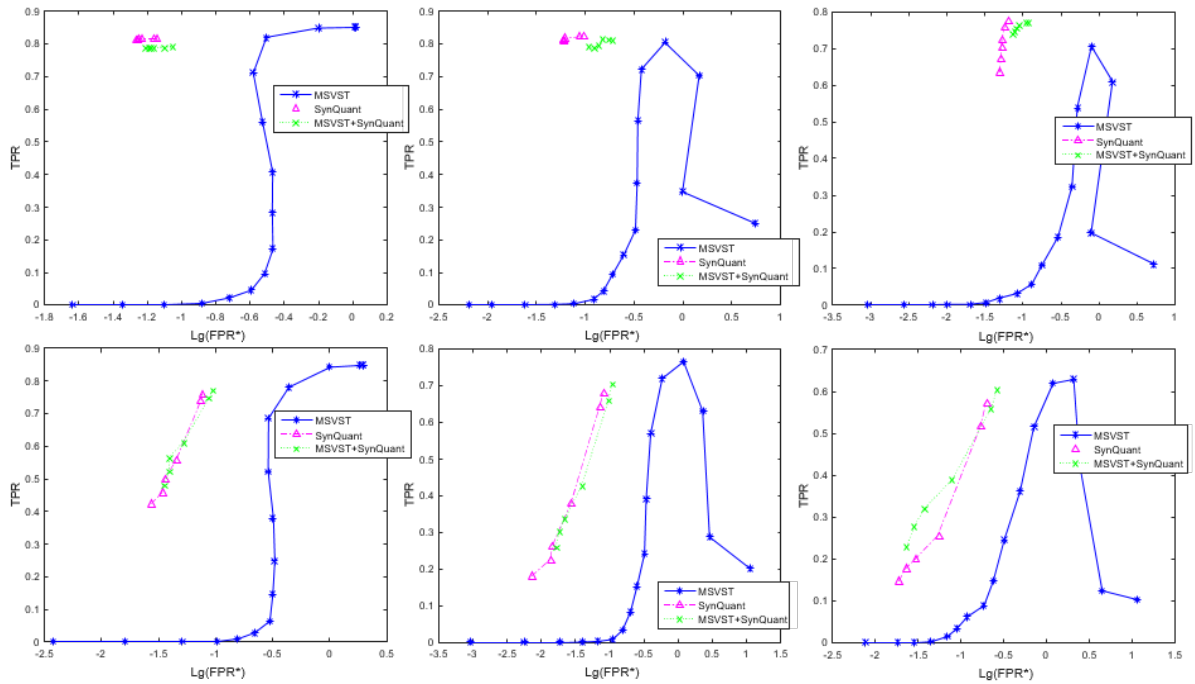
S9. Combine MS-VST and SynQuant

We are interested to see whether we can improve our performance by combining two methods. As MS-VST gives the probability of each pixel to be puncta, we seek to apply SynQuant on this probability map. Under such condition, most interference from the reference neurite channel has already been removed. However, because a large number of false positives in the probability map generated by MS-VST, we do not observe a clear improvement. Sometimes it is even worse than directly applying SynQuant on the original data.

Supplemental Figure 8-9 show the precision-recall and TPR-FPR* curves of SynQuant, MS-VST, and the joint method. Obviously, though MSVST and SynQuant perform best among the methods compared in the main paper, their combination fails to increase the accuracy.



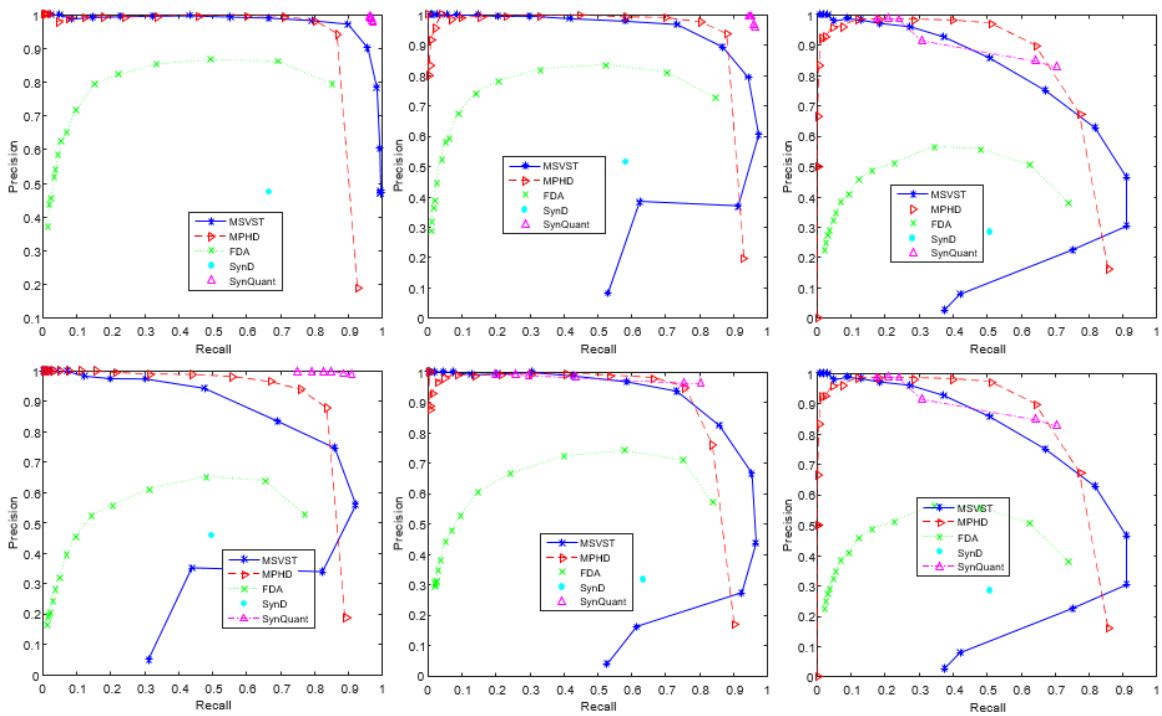
Supplemental Figure 6. The precision-recall curves of MSVST combined with SynQuant. Every two sub-figures in one column represent the results of images in one scenario. Scenario 1 to 3 are from left to right. Because SynQuant handles neurite signals well, even though MSVST removes most neurite-like signals, there is no obvious improvement on final results.



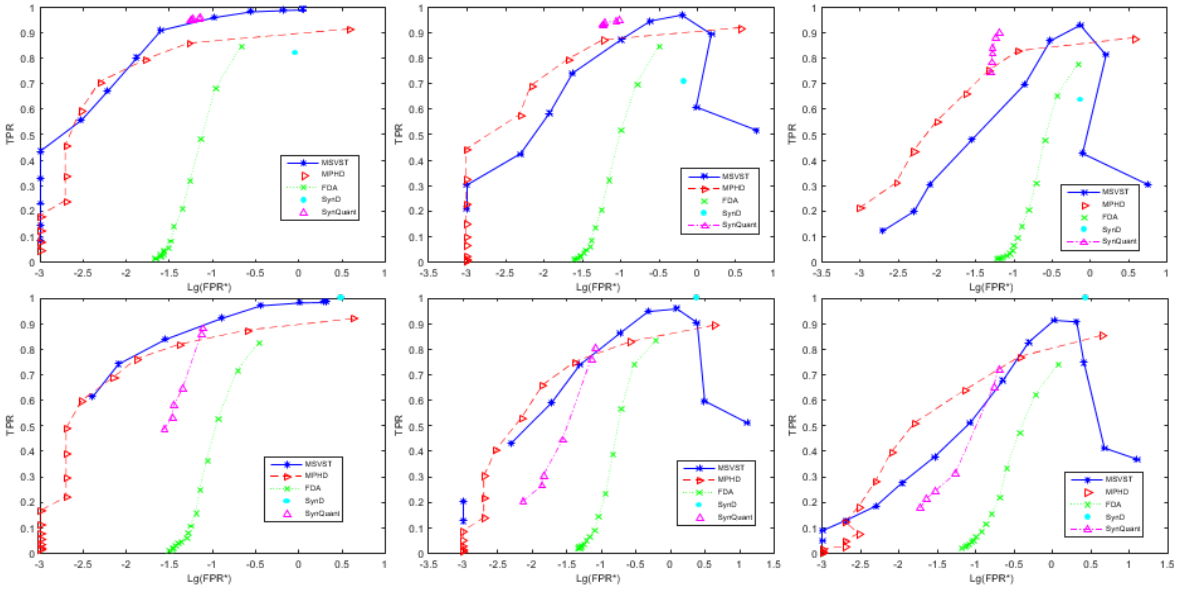
Supplemental Figure 7. The TPR-FPR curves of MSVST combined with SynQuant. The sub-figures are in the same order as Supplemental Figure 8. The results are consistent with Supplemental Figure 8. Besides, if some true synapses are missed by MSVST, the final results would be even worse than directly apply SynQuant on the original data.*

S10. Comparisons based on a location-based criterion

For the problem of spot detection, people sometimes care more about position accuracy rather than the pixel overlapping rate, which also measures the segmentation accuracy. Therefore, most spot detection algorithms (Zhang et al., 2007; Rezatofighi et al., 2012; Smal et al., 2010) use center distance between detected spots and ground truth to judge positive and negative. Here we also listed the comparisons based on that criterion. The precision-recall and TPR-FPR* curves are shown in Supplemental Figure 10-11. From these curves, we could tell that the performances are consistent with those under IoU measurement. However, this measurement only considers center distances between detected puncta and ground truth and does not consider overlapping ratio, which is quite unreliable in the synapse detection and segmentation problem. Under such condition, a large number of small points generated by noises around ground truth would be considered as true positives. What's more, the number of positive puncta detected would be much larger than reality. Such results are less useful in real application.



Supplemental Figure 8. Precision-recall curves for the algorithm under new criteria which was used by most spot detection methods. SynQuant still outperforms other methods. This kind of measurement only consider center distances between detected puncta and ground truth with no overlap ratio, which is unreliable in measuring the segmentation accuracy.



Supplemental Figure 9. TPR-FPR curves for the algorithm under the new criterion. Under such condition, a large number of small points generated by noises around ground truth will be consider as true positives. What's more, the number of detected puncta would be much larger than reality.*

S11. Neurite segmentation

For synapse quantification, it is important to collect synapse properties along with features from neurites. However, the neurite related features obtained from existing tools are not sufficient. For example, existing synapse quantification tools only utilize the length of neurite, with no other confounding factors in calculating synapse density. However, this density could be influenced by spatial inhomogeneous factors like neurite scale or sizes (Glynn et al., 2011; Klintsova and Greenough, 1999). When we test the relationship between synapse density and related disease phenotypes, if we do not consider such spatial related confounding factors of neurites, the conclusion might be unreliable. Therefore, analyzing the properties of neurites in details will help the study of the physiology and pathology of synapses.

Similar to SynD (Schmitz et al. 2011), we use a steerable filter (Meijering et al. 2004) to trace the neurite. The filter is based on the second-order derivatives of a Gaussian kernel:

$$f_{ij} = -\left(f * G_{ij}\right), \qquad G_{ij} = \frac{\partial^2 G}{\partial_i \partial_j}. \tag{S6}$$

Here * denotes convolution and index i and j are either x or y directions and f is the image. For each pixel, we get the Hessian matrix

$$H_f(t) = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}. \tag{S7}$$

Here $f_{xy} = f_{yx}$. The probability that pixel t belongs to a neurite is obtained from the eigenvalues and eigenvectors of this matrix. The cost for pixel q to be added to a neurite where pixel p is already on it is

$$C(p,q) = \gamma\, C_\lambda(q) + (1 - \gamma) C_v(p,q), \tag{S8}$$

where $\gamma \in [0,1]$ is a user-defined weight. In default we set it to be 0.5, so we have no preference for each term. $C_\lambda(q)$ and $C_v(p,q)$ are two normalized cost components. The former one is the cost of pixel q to be added to **any** neurite. Here

$$C_\lambda(q) = 1 - \rho(q), \qquad \rho(q) = \begin{cases} \lambda(q)/\lambda_{max} & \lambda > 0 \\ 0 & \lambda \le 0 \end{cases}. \tag{S9}$$

λ_{max} is the maximum of the largest eigenvalues obtained from the Hessian matrices of all pixels. $\lambda(q)$ is the largest eigenvalue of pixel q's hessian matrix. The latter one is the direction cost if we add pixel q to a neurite where pixel p is on it. It is defined as

$$C_v(p,q) = \frac{1}{2}\left\{\sqrt{1 - \omega(p,q)} + \sqrt{1 - \omega(q,p)}\right\}. \tag{S10}$$

$\omega(p,q) = |v(p) \cdot d(p,q)|$, which is the unit "link vector" from pixel p to q. $d(p,q) = (q-p)/||q-p||$ and v(q) is the normalized eigen vector of p. By taking the absolute value, this cost component ignores direction and only consider orientation similarity.

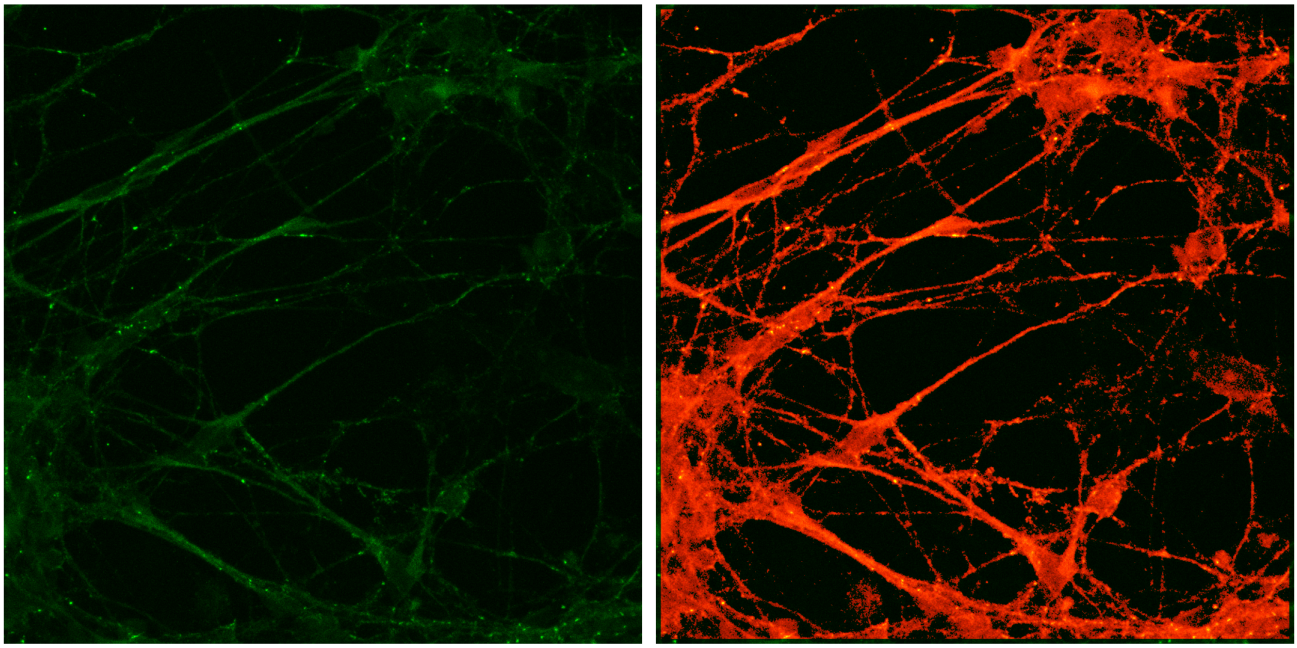
With Equation (S8), we can extract neurites from seed points by measuring if the cost is less than a threshold (0.9 in our implementation). Although neurite must grow from the soma in the SynD algorithm, this restriction is relaxed in our

experiments to avoid false negatives caused by missing somas. In our experiments, every unchecked point would be tested as a seed to find whether it belongs to a nearby neurite.

The detected neurite is cut by branch points and end points for further analysis. We assume the intensities of pixels in each piece is similar. Each piece corresponds to one or more synapses that stay on it. Concerning both synapse features and related neurite features, a comprehensive quantification of the synapse can be conducted.

S12. Results of probabilistic synapse detection (Simhal AK, 2017)

We use the code from https://github.com/aksimhal/probabilistic-synapse-detection. We modified the 'sample_runme' configuration file to use only one channel. We set blobSize = 2, search_win = 2 and edge_win = 8. This method cannot distinguish neurites and synapses (Supplemental Figure 12). Therefore, it can not be used in the presence of antibody non-specificity.



Supplemental Figure 10. Left: raw image with Synapsin labeling. Right: red channel shows detection results from Simhal, AK., et al 2017. Green channel shows the raw image. Majority of the neurites are detected as synapses.

S13. Other functions supported by SynQuant

Synaptic sites detection

Because researchers often want to compare locations of pre-synaptic and post-synaptic puncta (i.e. indicating the positions of synaptic sites), SynQuant allows users to simultaneously input two synapse channels (one for pre-synapse and another for post-synapse) and automatically reports the puncta's overlapping regions, which can be viewed as possible synaptic sites.

Synaptic sites quantification

Synaptic sites correspond to axon terminals, which are essential for neuron communication. It is meaningful to quantify synaptic sites and uncover the effects of some properties related to its density. Based on the synaptic sites detection results and dendrite extraction results, SynQuant could automatically finish this job with the three input channels (channels for pre-synapse, post-synapse and dendrite). The pipeline is similar to that of synapse quantification, but the dependent variable in the regression test needs to be changed from the number of synapses on one dendrite piece to the number of synaptic sites on one dendrite piece.

Supplemental references

Foi, Alessandro, et al. "Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data." Image Processing, IEEE Transactions on 17.10 (2008): 1737-1754.

Glynn, Marian W., et al. "MHCI negatively regulates synapse density during the establishment of cortical connections." Nature Neuroscience 14.4 (2011): 442-451.

Klintsova, A. Y., & Greenough, W. T. (1999) Synaptic plasticity in cortical systems. Current opinion in neurobiology, 9, 203-208.

Meijering E., et al. (2004) Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. Cytometry A 58, 167–76.

Rezatofighi, Seyed Hamid, et al. "A new approach for spot detection in total internal reflection fluorescence microscopy." Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on. IEEE, 2012: 860-863.

Schmitz, Sabine K., et al. "Automated analysis of neuronal morphology, synapse number and synaptic recruitment." Journal of neuroscience methods 195.2 (2011): 185-193.

Simhal, AK., et al. "Probabilistic fluorescence-based synapse detection." PLoS computational biology 13.4 (2017)

Smal, Ihor, et al. "Quantitative comparison of spot detection methods in fluorescence microscopy." Medical Imaging, IEEE Transactions on 29.2 (2010): 282-301.

Zhang, Bo, et al. "Multiscale variance-stabilizing transform for mixed-Poisson-Gaussian processes and its applications in bioimaging." Image Processing, 2007. ICIP 2007. IEEE International Conference on. Vol. 6. IEEE, 2007: VI-233-VI-236.