

# VESICLE: Volumetric Evaluation of Synaptic Interfaces using Computer Vision at Large Scale

William Gray Roncal<sup>1,2</sup> · wgr@jhu.edu

Michael Pekala<sup>2</sup> · mike.pekala@jhuapl.edu

Verena Kaynig-Fittkau<sup>3</sup> · vkaynig@seas.harvard.edu

Dean M Kleissas<sup>2</sup> · dean.kleissas@jhuapl.edu

Joshua T Vogelstein<sup>4,5</sup> · jovo@jhu.edu

Hanspeter Pfister<sup>3</sup> · pfister@seas.harvard.edu

Randal Burns<sup>1</sup> · randal@cs.jhu.edu

R Jacob Vogelstein<sup>2</sup> · jacobv@gmail.com

Mark A Chevillet<sup>2</sup> · mark.chevillet@jhuapl.edu

Gregory D Hager<sup>1</sup> · hager@cs.jhu.edu

<sup>1</sup> Johns Hopkins University, Baltimore, Maryland  
Department of Computer Science

<sup>2</sup> JHU Applied Physics Laboratory, Laurel, Maryland  
Research and Exploratory Development

<sup>3</sup> Harvard University, Cambridge, Massachusetts  
School of Engineering and Applied Sciences

<sup>4</sup> Johns Hopkins University, Baltimore, Maryland  
Department of Biomedical Engineering,

<sup>5</sup> Johns Hopkins University, Baltimore, Maryland  
Institute for Computational Medicine

An open challenge at the forefront of modern neuroscience is to obtain a comprehensive mapping of the neural pathways that underlie human brain function; an enhanced understanding of the wiring diagram of the brain promises to lead to new breakthroughs in diagnosing and treating neurological disorders. Inferring brain structure from image data, such as that obtained via electron microscopy (EM), entails solving the problem of identifying biological structures in large data volumes. Synapses, which are a key communication structure in the brain, are particularly difficult to detect due to their small size and limited contrast. Prior work in automated synapse detection has relied upon time-intensive, error-prone biological preparations (isotropic slicing, post-staining) in order to simplify the problem.

This paper presents VESICLE, the first known approach designed for mammalian synapse detection in anisotropic, non-poststained data. Our methods explicitly leverage biological context, and the results exceed existing synapse detection methods in terms of accuracy and scalability. We provide two different approaches - a deep learning classifier (VESICLE-CNN) and a lightweight Random Forest approach (VESICLE-RF), to offer alternatives in the performance-scalability space. Addressing this synapse detection challenge enables the analysis of high-throughput imaging that is soon expected to produce petabytes of data, and provides tools for more rapid estimation of brain-graphs. Finally, to facilitate community efforts, we developed tools for large-scale object detection, and demonstrated this framework to find  $\approx 50,000$  synapses in  $60,000 \mu m^3$  (220 GB on disk) of electron microscopy data.

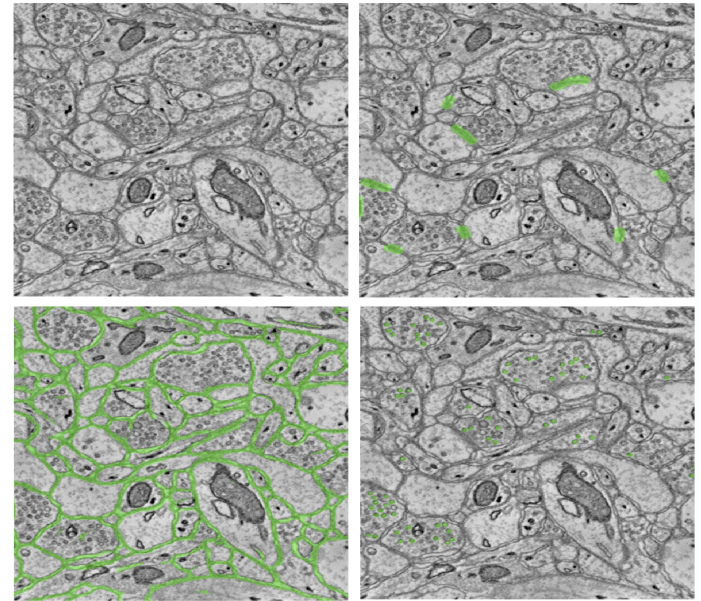


Figure 2: A single cross-section of EM data is shown (upper left). The detection task is to identify synapses shown in green (upper right). These synapses are known to exist at the interface of two neurons; these boundaries can be approximated by previously computed membranes, allowing us to restrict the evaluation regions to the green pixels (lower left). Clusters of vesicles are a good indicator of an axonal bouton, suggesting that one or more synaptic sites is likely nearby. Vesicles found by our automated detection step are highlighted in green (lower right).



Figure 1: Previous work on synapse detection has focused on isotropic post-stained data (left), which shows crisp membranes and dark fuzzy post synaptic densities (arrows) from all orientations. The alternative imaging technique of non post-stained, anisotropic data (middle, right) promises higher throughput, lack of staining artifacts, reduction in lost slices, and less demanding data storage requirements - all critically important for high-throughput connectomics. The XZ plane of a synapse in anisotropic data is shown (right), illustrating the effect of lower resolution. We address this more challenging environment, in which membranes appear fuzzier and are harder to distinguish from synaptic contacts. Data courtesy of Graham Knott (left) and Jeff Lichtman (middle, right).

In Figure 1, we provide examples demonstrating the challenging setting of our detection task. Figure 2 illustrates the importance of leveraging contextual information, especially neurotransmitter-containing vesicles and cell membranes. Finally, Figure 3 presents a summary of classifier performance; visualizations of the results are not shown here, but are available in the full paper and on our website. Our code and data are open source and available at: [openconnecto.me/vesicle](https://openconnecto.me/vesicle).

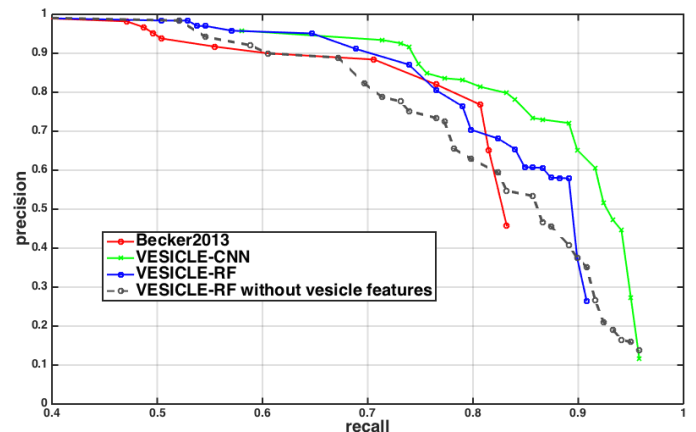


Figure 3: VESICLE-RF and VESICLE-CNN significantly outperform prior state-of-the art, particularly at high recall rates. The relatively abrupt endpoint of the Becker2013 method occurs because beyond this point, thresholded probabilities group into large detected regions rather than individual synapses, which are disallowed.