

SPECHT: Self-tuning Plausibility Based Object Detection Enables Quantification of Conflict in Heterogeneous Multi-scale Microscopy

Ben Cardoen ¹, Timothy Wong ², Parsa Alan ², Sieun Lee ², Joanne Aiko Matsubara ², Ivan Robert Nabi ², and Ghassan Hamarneh ²

¹Simon Fraser University

²Affiliation not available

October 30, 2023

Abstract

Self-tuning unsupervised analysis of STED super resolution of fluorescent labelled Caveolin-1, confocal microscopy of retina tissue. See manuscript for complete description.

SPECHT: Self-tuning Plausibility Based Object Detection Enables Quantification of Conflict in Heterogeneous Multi-scale Microscopy

Ben Cardoen *Student Member IEEE*, Timothy Wong, Parsa Alan, Sieun Lee, Joanne Aiko Matsubara, Ivan Robert Nabi*, Ghassan Hamarneh *Senior Member IEEE**

Abstract—Identifying objects in fluorescence microscopy is a non-trivial task burdened by parameter-sensitive algorithms. With experiments spanning multiple channels, datasets, operators, and microscopes, there is a clear need for an approach that adapts dynamically to changing imaging conditions. We introduce an adaptive object detection method that, given a microscopy image and an image level label, uses a kurtosis based matching of the distribution of the image differential to express operator intent in terms of recall or precision. Examples of image level labels include genome-based alteration of sub-diffraction limited cellular structures or pathological diagnosis based on image-based analysis of tissue section, where we wish to capture those aspects of the image that support the label, and to what extent. We show how a theoretical upper bound of the statistical distance in feature space enables application of belief theory to obtain statistical support for each detected object. We validate our method on 2 datasets: identifying Caveolin-1 labelled caveolae and scaffolds acquired by STED superresolution microscopy, and detecting amyloid- β deposits in confocal microscopy retinal cross sections of neuropathologically confirmed Alzheimer’s disease donor tissue. Our results show consistency with biological ground truth and with previous subcellular object classification results, yet adds insight into more nuanced object transition dynamics. We illustrate the novel application of belief theory to object detection in heterogeneous microscopy datasets and the quantification of conflict of evidence in a joint belief function. By applying our method successfully to confocal and superresolution microscopy, we demonstrate multi-scale applicability.

Index Terms—belief theory, object detection, statistical support, self-tuning, evidence conflict, live-cell microscopy, confocal and superresolution microscopy, diffraction limited

I. INTRODUCTION

SCIENTIFIC discovery in multiscale biomedical imaging is a difficult task confounded by multiple factors. Discovery must cope with data that has no, sparse, or conflicting annotation or supervision. Because discovery depends on the data without prior annotation, robustness to confounding factors such as changing acquisitions, datasets, and operators is a key design priority. However, as such robustness relies on manually adjusting a complex and large set of parameters, results are difficult to reproduce and replicate in different laboratories, such that researchers are burdened with time-consuming and elaborate parameter sensitivity tests. An ideal method to overcome these limitations is a self-tuning approach that configures itself for consistent results without user intervention. This allows for traceable and reproducible discovery that combines multiple sources of evidence to support the claimed observations of a novel phenomenon. Discovery studies using fluorescence microscopy typically span multiple cell lines, diverse fluorescent markers (imaged in parallel in channels), and various microscopes and operators, requiring a method that adapts automatically to this variability. In this work we propose a robust, adaptive, and self-tuning method that enables traceable biomedical knowledge discovery in heterogeneous multi-scale microscopy. The class of discovery problems we focus on is the identification of macro-molecular structures implicated in certain biological phenomena captured in an imaged specimen from subcellular super-resolution images to confocal tissue sections. The structures are labelled with fluorescent markers, where the aggregation (‘deposit’) of the marker is observed by the microscope (Fig. 1-A). The identified deposits are not always clearly distinguishable from the background or each other, for example, in the case of smaller modular structures gradually combining into larger. When deposits are indicative of a developing pathology, it is expected that they span a spectrum of appearances, rather than a single uniform identifiable instance. We focus on 2 specific use cases: identifying and distinguishing sub-diffraction limit (<100 nm diameter) caveolin-1 (Cav1) caveolae and scaffolds, and identifying retinal amyloid- β deposits associated with Alzheimer disease. In the first, superresolution (dSTORM) and network analysis are able to identify different types of Cav1

Correspondence to be addressed to: {bcardoen,hamarneh}@sfu.ca, irnabi@mail.ubc.ca

This work was supported in part by grants from the Canadian Institutes for Health Research (PJT-159845 and PJT-156424), Natural Sciences and Engineering Research Council, Simon Fraser University Big Data Scholarship, Canada Foundation for Innovation and British Columbia Knowledge Development Fund. Compute Canada provided computational resources and advice on high performance scientific computing.

* Ghassan Hamarneh and Ivan Robert Nabi contributed equally to this paper.

Ben Cardoen and Ghassan Hamarneh are with the Medical Image Analysis Laboratory, School of Computing Science, Simon Fraser University, BC Canada V5A 1S6. e-mail: {bcardoen, hamarneh}@sfu.ca

Sieun Lee is with the Department of Ophthalmology and Visual Sciences, University of British Columbia (UBC), Vancouver, BC, Canada V6T 1Z3, School of Engineering Science, Simon Fraser University, BC Canada V5A 1S6, and Precision Imaging Beacon, University of Nottingham, Nottingham, UK NG7 2UH. email: sieun.lee@nottingham.ac.uk

Joanne Matsubara is with the Department of Ophthalmology and Visual Sciences, Eye Care Centre, University of British Columbia, BC, Canada V5Z 3N9. email: jms@mail.ubc.ca

Ivan Robert Nabi, Timothy Wong and Parsa Alan are with the Department of Cellular and Physiological Sciences, Life Sciences Institute, University of British Columbia, BC Canada V6T 1Z3. e-mail: irnabi@mail.ubc.ca, tim.wong@alumni.ubc.ca, parsa.arabi@alumni.ubc.ca

The authors declare that this technology is submitted as a provisional patent: US PROV APP UBC-UILO File No. 20-041

structures in fixed cells [1], [2]. Here we use STimulated Emission Depletion (STED [3]) superresolution microscopy that can achieve <40 nm resolution and can be used for high speed live cell imaging [4]. In the second use case, we aim to distinguish automatically which retinal amyloid- β deposits are characteristic of Alzheimer disease labelled images, an open question in the field [5].

A. Problem Statement

Given a microscopy image I and its label L , we wish to identify the parts of the image that show evidence for L , or for a possibly divergent L' . L is not necessarily equal to $\overline{L'}$: $L \neq \overline{L'}$. The label L and L' are a subset of a largely unexplored label space \mathcal{L} . In the context of scientific discovery, we are, if successful, able to offer insight into the evidence based support for subsets of \mathcal{L} . Similarly, $L \cap L' \neq \emptyset$, for example $L \subset L'$ holds for a modular component L being a part of a larger whole L' [2]. With the contents of the image being the subject of scientific discovery, the label is obtained outside of the image domain. For instance, the image acquisition samples a tissue where a pathologist diagnoses the entire specimen, or the image captures the phenotype response to a genotype modification. It is not known if the information encoded in the image supports the label. We aim to model a function D that identifies subsets ('objects') o of the image and a function S_L that assigns to each object the statistical support for the image label L .

$$D : I \mapsto o \mid o \subset I. \quad (1)$$

$$S_L : o \mapsto e \mid e \in [0, 1], o \subset I, L \in \mathcal{L}. \quad (2)$$

In superresolution microscopy images of complex small-scale protein oligomer structures (Fig. 1-A), exact delineation of the object boundary is not always possible given the precision of the system. If the representation of each detected object o contains sufficient information to quantify its statistical support for the image-level label, the exact delineation is of secondary importance. In the remainder of this work, we use the notation $o \rightarrow L$ for the proposition that the object o supports the label L , and $S_L = S(o \rightarrow L)$ is the function S assigning a continuous support value to the proposition in the context of belief theory [6], a generalization of probability functions. A 'frame of discernment' $\Theta = \{o \rightarrow L \mid o \subset I, L \in \mathcal{L}\}$ is the set of all sources of evidence for the image I and label-space \mathcal{L} . When $|T| = 1 \quad \forall T \subset \Theta$, subsets of Θ are referred to as 'singletons'.

B. Requirements for a solution

In this section, we list what we believe are must-have features of an 'ideal' method for solving the stated problem.

1) *Adaptive*: In our problem setting, the diversity of datasets, microscopes, operators, channels, and even temporal information makes for a highly variable acquisition space. An ideal method is not insensitive to this variation, but rather adapts to it to ensure the information extraction is not interrupted or biased.

2) *Self-tuning*: The 'ideal' function D should retrieve objects from the image without requiring the operator to spend time on algorithm-specific parameter sensitivity studies. If exploring a parameter space is unavoidable, then ideally the parameters should be higher level, wherein the intent of the user in terms of precision or recall is automatically translated into lower level algorithm-specific parameters ensuring robust reproducible results even for unseen datasets.

3) *Unsupervised*: Provision of a per-object ground truth annotation of the sought-after new discovery is, by definition of discovery, contradictory. Annotations can bias knowledge discovery because supervision is based on pre-discovery knowledge that can be flawed [7] and can prevent extracting conflicting novel knowledge.

4) *Unhindered by choice of statistical method*: Modelling the statistical support (function S) should not be restricted by technical requirements induced by the chosen statistical framework, such as the additive property for Bayesian inference. A trivial example of a function S that models 'ignorance', or a vacuous belief [6], is a naive $S_L(I) = 1$ that only restates the knowledge about the given image label itself, and is unable to infer any relevant image parts, i.e. $S_L(o) = 0 \mid o \subset I$, and thus $\sum S_L(o) = 0 \mid o \subset I$. An alternative example is an image heavily perturbed by noise such that state-of-the-art object detectors fail to recover any objects containing signal to support L at all above a random baseline. Unless careful design ensures inference is robust to the choice of prior [8], one risks either invalidating the chosen inference technique or restricting the solution space. Bayesian inference requires assignment of a probability value to 'singletons', yet such a decomposition of Θ into singletons is not always possible due to lacking evidence or support. An assignment of uniform prior support can be used as a workaround, but this is not always justified or can violate domain restrictions. A similar constraint is Cromwell's rule [9], where a prior of one or zero results in a one or zero posterior.

5) *Adapts to conflicting evidence*: In the absence of perfect information, it is expected to find a single object $o \subset I$ that provides evidence for multiple partially conflicting labels, i.e. $S_L(o) \neq 1 - S_{L'}(o)$. An ideal method allows the quantification of such conflicting evidence.

6) *Supports a continuous support for the label space*: The evidence provided by an image may not lead to the exclusion of all labels but one. For example, an image of tissue where a pathology such as Alzheimer is present can have deposits that are indicative of healthy tissue as well as deposits that support, to a certain extent but not completely, the diagnosis. In this case, an 'ideal' function $S_L(o)$ should output the support of o for each L , either 'healthy' or 'disease', as a floating point value representing the position of the support on a continuous scale from uncertain (0) to certain (1).

7) *Uncertainty-encoding and bounded error*: The function S should, in addition to capturing support, encode the uncertainty involved in its computation. Given a label and a source or body of evidence for that label, uncertainty is defined as the difference between the maximal and minimal support for the label provided by the source [10] (Fig. 1-C.3 'r'). Our working definition of uncertainty requires lower and upper bounds on

the support for a label. A non-trivial theoretical bound (trivial: $0 \leq S_L(o) \leq 1$), in contrast to an empirical bound, transfers across datasets and enables more robust reproducibility in scientific discovery.

8) *Spans sparse heterogeneous data*: Obtaining data for biomedical knowledge extraction is a time and resource intensive process with limited experiment and replicate sizes that reduce statistical power. An optimal method should be able to report results on few images, without requiring retraining or transfer learning. Ideally, the approach should be decoupled from the acquisition, that is, for microscopy, the approach should work multi-scale, and be robust to device and operator variance.

C. Challenges in current approaches

The state-of-the-art methods tackling problems closest to our problem statement are broadly divided into (i) joint segmentation and classification, (ii) interpretable deep learning, and (iii) multiple instance learning (MIL).

(i) Joint or hybrid segmentation and classification decomposes the image into segments and classifies each segment, and have been applied successfully to microscopy [11], chromosome microscopy [12], breast biopsy [13], fundus images [14], and histopathology [15] to name a few. However, typically they require annotations, which we do not have, do not provide a theoretical upper bound to the label assigned to each segment, nor is it clear on how to apply the same method across heterogeneous datasets or quantify conflict.

(ii) Explainable AI, where neural networks can, for example, produce the regions of the image that provide the most decisive information supporting the predicted image level label, are covered in more detail in recent reviews [16], [17]. Recently, these approaches are accompanied by domain fusion, for example augmenting MRI of Alzheimer [18] patients with meta-data to learn the MRI signature of Alzheimer disease, or fusing diagnostic reports with image data [19] to offer interpretable improved diagnosis. While in such approaches the support each region has for a single label is found, it is not optimized to split those regions into smaller distinct objects, nor is the support bounded. Filtering the attention maps [20] to obtain a more precise delineation of which regions of an image support a label are one direction closing the gap towards granular object detection folded into interpretable AI.

(iii) In MIL terminology, a label exists for a ‘bag’ of instances. The ‘bag’ can refer to the image, where instances would be objects in the image. The standard MIL model has it that all bags with label L^- only contain instances with label L^- . Bags with label L^+ contain instances with at least one instance (‘witness’) with label L^+ . MIL has been adopted successfully for microscopy-specific tasks such as classifying and segmenting cells [21] with recent reviews [22], [23] detailing the different approaches. We are not aware of MIL methods that incorporate the explicit encoding of (conflicting) evidence and uncertainty in the context of evidence theory, nor do MIL approaches feature a theoretical bound on the support for each observed instance. In conclusion, we did not find a method that fulfils all our solution requirements.

D. Contribution

We introduce a 2-stage approach to model functions D and S (Sec. I-A), object detection and evidence-based object labelling, respectively, designed to satisfy all the aforementioned method requirements (Sec. I-B) and is novel in its combination into a single framework. The ability to distinguish scaffolds and caveolae in live cell superresolution microscopy and the application of belief theory to identify Alzheimer specific amyloid- β deposits in confocal microscopy is novel. Our contributions are:

- Adaptive and self-tuning object detection using the kurtosis of the Laplacian to match distributions across channels for fluorescence microscopy (Sec. II-A).
- Belief theory based labelling to quantify the support of the identified objects for an image-level label (Sec. II-B).

To the best of our knowledge, our use of kurtosis to estimate algorithm-specific parameters consistently across heterogeneous data, such as images from different devices, is novel. Furthermore, in this work, we do not have object-level annotation. Rather than an ad hoc method for specific blob detection, our work provides a novel, self-tuning, and robust framework for analyzing data without pre-existing annotations required for training and evaluation.

In Section II we detail both stages of our proposed method. Recognizing the importance of dataset, operator, and acquisition agnostic performance, we show that our method obtains stable results on a heterogeneous dataset when the operator and microscope used to generate the data differ from the dataset used for inference, and perform a parameter sensitivity study to confirm that our method obtains consistent results.

II. METHOD

In this section, we outline our method and verify that it satisfies the requirements listed in Section I-B.

A. Adaptive kurtosis aligned object detection

1) *Object detection principle*: While simple manual thresholding can balance a trade-off between precision and recall, finding the same consistent balance across images, channels, and datasets using manual thresholding requires a per-image threshold and is sensitive to high operator variance. We illustrate how we achieve this goal when thresholding V , the negative Laplacian. The image Laplacian ∇^2 , a measure of the second derivative of the image intensity, can be used to detect edges of objects where ∇^2 changes sign. In 2D microscopy images of 3D fluorescent deposits, we can leverage that connected components of $V = |\min(\nabla^2, 0)|$ (Alg. 1, line 5) coincide with the approximate outline of the objects, since the intensity curve of such observations is bell-shaped (Fig. 2) when the fluorescent marker is labelling complex spherical structures with a non-constant height. Even non-specific binding can maintain the same assumption given its tendency to self-organise [24] in concentrations of fluorescent label. More formally, the domain, use case, and acquisition allow us to state that the intensity profile for a single object can be approximated by a generalised normal distribution with

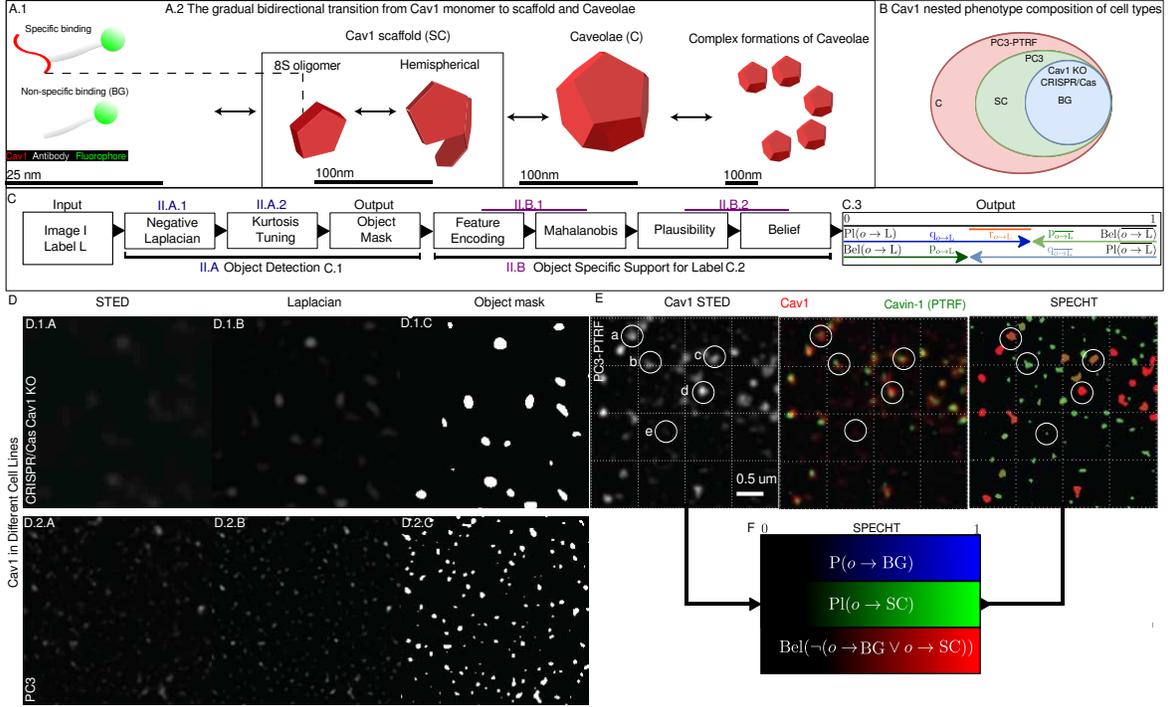


Fig. 1. Cav1 molecules imaged with fluorescent conjugated antibodies (A.1). A Cav1 molecule forms increasingly complex structures with oligomers grouping into scaffolds (SC) up to (networks of) caveolae (C, A.2). Our representation is deliberately simplified as this is still an active research question. A Venn diagram representation of the nested phenotypes in our problem statement (B). Cav1 KO cells only exhibit non-specific binding background labelling (BG) with the Cav1 gene knocked out. PC3 cells exhibit scaffolds (SC) and non-specific binding. PC3-PTRF cells exhibit all three. Panel C shows the control flow of SPECHT combining self-tuning object detection (Alg. 1, Sec. II-A) with belief function based labelling (Alg. 2, Sec. II-B). Examples for the object detection (D) for representative images of Cav1 KO (D.1.A, intensity increased for visualization) and PC3 (D.2.A) cells with corresponding Laplacian and object mask. It is hard to distinguish between Cav1 deposits in PC3 versus Cav1 KO, even in Laplacian space (D.1-2.B). The need to identify non-specific labelling in PC3 and PC3-PTRF is crucial to avoid mis-identification with scaffold-like structures. The PC3-PTRF image (E) shows Specht’s belief based identification of Cav1 objects, with high belief to be caveolae (E.d, red) showing an expected strong colocalization with PTRF (Cavin-1), a protein required for formation of Caveolae. Transient objects believed to be on the spectrum between scaffold and caveolae (E.a-c, orange) are colocalizing with lower PTRF. An object with high belief to be a scaffold shows a disjoint colocalization (E.b, green). An instance of identified non-specific labelling is shown in E.e (blue-green). Images in D, E are insets of STED superresolution images containing a single cell each with $1e3$ objects per image.

probability density function $\frac{\beta}{2\alpha\Gamma\beta-1}e^{-\frac{|x-\mu|}{\alpha}^\beta}$ with α scale, μ location, Γ the gamma function, and $1 \leq \beta \leq 3$. We apply a 2-stage Gaussian (Alg. 1-line 4) smoothing before and after V to ensure pixellation effects are minimized, with σ set at or below the precision of the system. This is related to the Laplacian of Gaussian (LoG) approach, underlying ‘blob’ detection in for example ‘scale-space’ object detection [25]. However, in the classical computer vision formulation of ‘blob’ detection, the object representation is assumed to have a constant or similar representation, not bell-shaped as is the case in our fluorescent microscopy use cases. The 2nd σ is used to smooth rectilinear effects by the Laplacian operator, the first can be omitted when the acquisition microscopy has a specialized deconvolution operator tuned to the point spread function used.

2) *Self-tuning adaptive detection*: Given an object detector that gives a higher response with respect to the location of the object, we need to threshold the response to obtain a binary mask serving as object detection. To unburden the practitioner and improve reproducibility as well as consistency across images and channels, a self-tuning approach is needed. The practitioner can be given the option to express their intent in favoring precision or recall (Fig. 2) and expect to have that intent translated consistently across heterogeneous datasets into corresponding values in the parameter space of

the object detection method. In order to express user intent consistently, we have to find a way to translate that parameter across distributions of V -space. We observe that an image with a few bright objects will have a long-tailed distribution in V -space, whereas an image with a high frequency of faint objects will have a short-tailed distribution in V -space (Fig. 2-B red, blue, respectively). The kurtosis of a distribution is a scalar value increasing with the length of a tail of the distribution. We illustrate this behavior in biological data in Fig. 1-D: compare Cav1 KO versus PC3 and PC3-PTRF; 3 cell lines that express markedly different image phenotypes (Fig. 1-B, Sec. III-A). We next use these insights to normalize V to $Z_V = \frac{V_i - E(X)}{\sqrt{\text{Var}(V)}}$ and then obtain an estimate $E'_z \sim E[Z]$ as a consistent threshold that can be scaled up or down consistently across images. While we can compute $E'_z \sim E[Z] = \int z f(z) dz$, this entails that we have a probability density function, which in practice involves fitting a parametric function, a process that is non-trivial to do consistently across datasets, and unless corrected will have a larger error at the tails of the distribution. Yet we have established that it is at those tails that most of the variation will occur. We can end up with $E'_{z_i} < E[Z_i]$, yet $E'_{z_j} > E[Z_j]$, defeating the purpose of obtain a consistent threshold in the

worst case. However, when we have $E'_{z_i} < E[Z_i] \forall i$, then we are assured that our estimate will be consistent even in the worst case, and can be scaled up or down consistently as well. We can derive such a strict lower bound by noting that $\text{kurtosis}(V) = E[Z_V^4]$ [26]. By a special case of the Cauchy-Schwartz inequality, we know that

$$\forall x_i \in \mathbb{R}^+ \quad \sum_{i=1}^n x_i^2 \leq \left(\sum_{i=1}^n x_i \right)^2 \leq n * \sum_{i=1}^n x_i^2 \text{ if } n < \infty \quad (3)$$

from which it then follows that $\sum_{i=1}^n Z_i^4 \leq \left(\sum_{i=1}^n Z_i \right)^4$. We can then derive:

$$\sqrt[4]{k(X)} \leq E(Z). \quad (4)$$

We now have a lower bound approximation E'_z to $E(Z)$ that allows us to express a threshold on the normalized Laplacian that scales with the shape of the distribution of the negative second derivative of the image, producing consistent results across images, channels and datasets. Moreover, by weighting the kurtosis, we can allow the user to alter the threshold in an distribution-aligned space. We use the ‘excess’ kurtosis ($k-3$) in our implementation. We scale the outcome by a floating point parameter ‘precision-recall (PRC)’ to complete our aim of intent-based self-tuning (Sec. I-B2) and adaptive method (Sec. I-B1). A value $\text{PRC} > 1$ leads to a distribution-aligned object extraction that favours recall, $\text{PRC} \leq 1$ favours precision. Appendix figure 2-B illustrates the scaling effect on in silico distributions. Figure 7 shows the aligned object detection in action on three markedly different images and contrasts against a fixed threshold approach in Z-space. However, our results illustrate the need for an auto-tuning approach where the object detection method retrieves objects consistent with the end-user intent by aligning the image differential distributions. For a value of $\text{PRC}=2$ (high recall), the associated $z' \leq E(Z)$ values are listed to illustrate the contrast with a fixed Z-value threshold. The complete algorithm to detect objects from a heterogeneous set of images is listed in Alg. 1.

Algorithm 1 Adaptive kurtosis-based self-tuning object detection

```

1: Input Set 2D images I, parameter  $\sigma_1, \sigma_2$ , precision-recall ratio (PRC)
2: Output Binary object masks  $M$ 
3: for  $I_j \in \mathbf{I}$  do
4:    $\nabla_j^2 \leftarrow \text{Gaussian}_{\sigma_1}(\text{Laplacian}(\text{Gaussian}_{\sigma_2}(I_j)))$ 
5:    $V_j \leftarrow |\min(\nabla_j^2, 0)|$ 
6:    $z_j \leftarrow \sqrt[4]{\text{kurtosis}(V_j)}$   $\triangleright$  Adaptive consistency across channels
7:    $V_j[V_j \leq \mu_g(V_j) * \sigma_g(V_j) \frac{z_j}{\text{PRC}}] \leftarrow 0$   $\triangleright$  Eq. 4
8:    $M_j \leftarrow \text{connectedcomponents}(V_j)$ 
9: end for

```

B. Probabilistic object labelling using belief functions.

The previous section gives us a function D (Eq. 1) that decomposes an image I with label L into objects ‘ o ’. Here we aim to find a function S (Eq. 2) that quantifies the evidence for the proposition $o \rightarrow L$ for each object.

Computing support for an image level label using belief theory: We model the problem of finding S for a label $L \in \mathcal{L}$ and image I :

$$S_L : o \rightarrow (p, q, r) \quad \{|o \mapsto L\} \subset \Theta, o \subset I, p, q, r \in [0, 1]. \quad (5)$$

The triplet (p, q, r) follows the notation of Dempster [10] where ‘ p ’ expresses the belief supported by probabilistic evidence that o supports the label L . ‘ q ’ is the belief o does not support L . ‘ r ’ is the uncertainty in measuring the respective beliefs. More formally a belief function on a set of propositions Θ is a function $\text{Bel} : 2^\Theta \mapsto [0, 1]$ such that $\text{Bel}(\Theta) = 1$, $\text{Bel}(\emptyset) = 0$, and $\text{Bel}\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{I \subset \{1, \dots, n\} \wedge I \neq \emptyset} (-1)^{|N|+1} \text{Bel}(A_i) \quad \forall A_i \subset \Theta$. Evidence can be encoded by a mass function $m(A) \rightarrow [0, 1] | A \subset \Theta$, where indivisible subsets A are referred to as ‘focal elements’, such that $\sum_{A \subset \Theta} m(A) = 1$. Probability functions and probabilities in Bayesian inference are a special case of belief functions where all focal elements are singletons. Unlike probability functions, for general belief functions $\text{Bel}(\bar{A}) \neq 1 - \text{Bel}(A)$. The ‘plausibility’ function is defined as $\text{Pl}(A) = 1 - \text{Bel}(\bar{A})$, and $\text{Pl}(A) \geq \text{Bel}(A) \forall A \subset \Theta$. In the (p, q, r) notation, we have that $p = \text{Bel}(A)$, $q = \text{Bel}(\bar{A})$, $r = \text{Pl}(A) - \text{Bel}(A)$. The reader can find a graphical illustration in Figure 1-C.3. For a more in-depth review of belief theory, we refer the interested reader to Yager et al. [27].

1) *Encoding evidence:* Given a set of images **J** with (singular) label L_J , and a set of images **I** with unknown label, we want to identify objects in images and assign to each object o a tuple (p, q, r) expressing the belief, plausibility and uncertainty on the proposition $o \rightarrow L_J$. In Alg. 2, we illustrate the steps we undertake to arrive at a belief based labelling of objects in images. The sets of images **J** and **I** can originate from different channels. The adaptive object detection stage ensures consistent results regardless of channel. After object detection (Alg. 1), we compute a feature descriptor for each object; in our experiments: intensity (sum), area (pixels) and Laplacian (V , sum), a simple, low-dimensional, with non-independent features. We next compute the statistical distance of any object o to the distribution of objects in images **J** in feature space using the Mahalanobis distance (Alg. 2-line 11) which accounts for co-dependent dimensions. The Mahalanobis distance range $([0, \infty))$ is not interpretable as a mass function.

2) *Inferring plausibility:* We want to be able to quantify both relative support and support for an individual label. We normalise the statistical distance (Alg. 2, line-13) so we can leverage Cantelli’s theorem [28]

$$\Pr[Z_i \geq z] \leq \frac{1}{1 + z^2} \quad (6)$$

to assign a theoretical upper limit to the probability that the object in question supports a label, which then becomes the plausibility $q_j = \text{Pl}(o \rightarrow L_J) \geq \text{Bel}(o \rightarrow L_J)$. From belief theory [10], we know that $\text{Bel}(\bar{A}) = 1 - \text{Pl}(A)$. For $o \subset \mathbf{I}$ we can formulate $p_i = \text{Bel}(o \rightarrow L_J) = 1 - q_j$. When we swap **I**, **J** we can obtain q_i and p_j , giving us

$r_i = q_i - p_i$ and $r_j = q_j - p_j$. Figure 1-C.3 illustrates the application of belief theory based labelling on object detection and the interplay between belief and plausibility. The resulting support function has no limiting specific priors or assumptions (Sec. I-B4), is continuous (Sec. I-B6), has a theoretical upper bound (Sec. I-B7, and requires no supervised training data (Sec. I-B3). When we are interested in relative support, comparing support for L_1 versus L_2 , the statistical distance can be sufficient without normalization. However, normalization allows us to compute plausibility and support for individual labels.

Algorithm 2 Probabilistic labelling algorithm

```

1: Input Images  $\mathbf{J}$  with label  $L_J$ , Images  $\mathbf{I}$ 
2: Output  $Pl_I$ , plausibility labelled objects for  $\mathbf{I}$ 
3:  $M_I \leftarrow \text{objectdetect}(\mathbf{I}, \sigma_1, \sigma_2, PRC)$  ▷ Alg. 1
4:  $M_J \leftarrow \text{objectdetect}(\mathbf{J}, \sigma_1, \sigma_2, PRC)$  ▷ Adapts to channel
5:  $F_J \leftarrow \{\text{features}(o_{j_i}) \mid o_{j_i} \in M_J[j], j \in [1, |\mathbf{J}|]\}$ 
6:  $\mu_J \leftarrow E(F_J), \Sigma_J \leftarrow \text{Cov}(F_J)$ 
7:  $D \leftarrow []$ 
8: for  $I_j \in \mathbf{I}$  do
9:   for  $o_k \in M_i[j]$  do
10:     $x_k \leftarrow \text{features}(o_k)$ 
11:     $D_j[k] \leftarrow \sqrt{(x_k - \mu_J)^T \Sigma_J^{-1} (x_k - \mu_J)}$  ▷ Mahalanobis
12:   end for
13:    $Z_j \leftarrow \frac{D_j - E(D_j)}{\sqrt{\text{Var}(D_j)}}$  ▷ Z-normalization
14:   for  $o_k \in M_i[j]$  do
15:     $Pl[o_k \rightarrow L_J] \leftarrow \frac{1}{1 + Z_j[k]^2}$  ▷ Eq. 6
16:   end for
17: end for

```

In Section III we apply our method to 2 use cases. First, we show how to apply our method on a hierarchical problem formulation where we differentiate between 3 nested labels $\{o \in L_{Cav1KO}\} \subset \{o \in L_{PC3}\} \subset \{o \in L_{PC3-PTRF}\}$ where a subset label is more specific as illustrated in Fig. 1-B). We validate our results with independent biological ground truth and previous art. We offer a parameter sensitivity study to quantify robustness. Second, we illustrate how to extend our method across heterogeneous small datasets and compute a joint belief function while quantifying the conflict between the composite belief functions. To the best of our knowledge using a belief theory based approach for object identification in images is novel.

In the following section we will apply our method to two distinct use cases to illustrate more advanced usage, in addition to validating the method.

III. USE CASES

The full description of the datasets used in this section is listed in the Appendix (Sec. A). Each subsection has a detailed breakdown of dataset structure, as this differs per use case. The use cases share that each is composed of 2D image / label pairs, where each image is a 2D observation of 3D fluorescent labelling.

A. Capturing the gradual construction of complex protein structures in live cells

Caveolae, subcellular structures composed of Cav1 protein complexes, are ~ 100 nm invaginations in the cell membrane

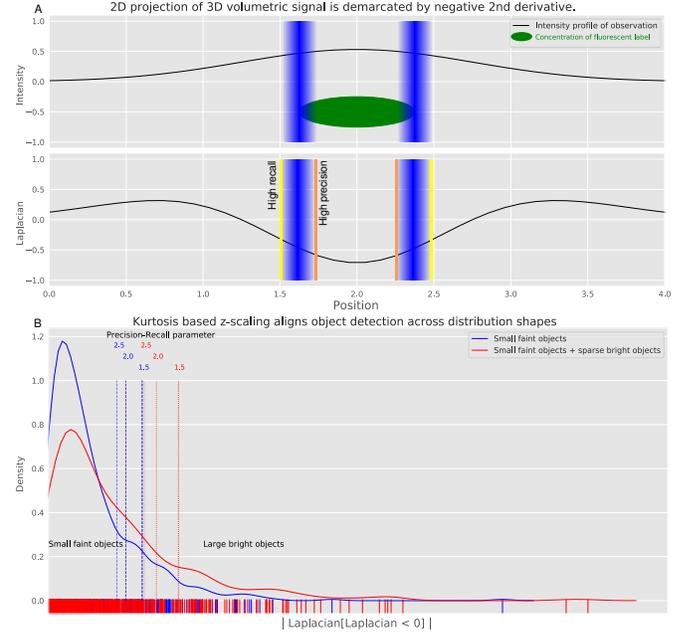


Fig. 2. The principle of object detection using negative Laplacian of a 2D observation of 3D fluorescent emissions (A). Kurtosis scaling estimates the required Z-threshold that aligns distributions across channels, cells, and operators (B). The PRC parameter works in kurtosis space, allowing the user to scale object detection consistently in favor of precision or recall across images, channels, and time without risking overfitting to a single image.

with a varied spectrum of functions [29]. Cav1 forms non-caveolar scaffolds (SC), including 8S oligomers that combine to form larger non-caveolar hemispherical scaffolds as well as caveolae (Fig 1-A.2) [1], [2], [30]–[32]. Scaffolds recruit Cavin1 (PTRF) to finally become caveolae (C) illustrated in Fig. 1-A.2. The gradual formation process enables a cell to respond to mechanical and hypotonic stress and is a factor in metastasis [29]. In fixed cells, superresolution (dSTORM) network analysis identifies individual scaffolds into separate subtypes whose modular similarity suggests that smaller scaffolds combine to form larger scaffolds and caveolae [2]. To allow investigation of the dynamics of caveolae biogenesis in live cells, we undertook to develop an approach to distinguish scaffolds and caveolae using an alternate super-resolution microscopy approach, STED, that is amenable to high speed live cell imaging. Cav1-labelled fluorescent deposits are identified in STED microscopy images and assigned a belief label describing where the identified concentration is on the spectrum between non-specific background labelling (BG, Fig. 1-A.1), scaffolds, or caveolae. BG deposits are fluorescent markers not associated with their biological target Cav1 molecules. BG can be considered background signal, but is differing from signal perturbing noise. BG fluorescent marker can have remarkable self-organising properties similar to free floating proteins [24]. Identifying BG allows us to exclude it from our biological targets. We study 3 cell lines: Cav1 CRISPR/Cas KO MDA-MB-231 cells with genetically disabled expression of Cav1, PC3 with genetically disabled expression of PTRF, and PC3-PTRF with PTRF and Cav1 enabled. In Cav1 KO we can only observe BG, in PC3 only SC and BG, in PC3-PTRF the

SC, BG and C are present (Fig. 1-A,B). Our label space \mathcal{L} is then $\{BG, SC, C\}$, with subsets $PC3=\{BG, C\}$ and $PC3\text{-PTRF}=PC3 \cup \{C\}$.

1) *Experimental procedure*: We detect fluorescent deposits (Alg. 1) in Cav1 KO and PC3 cell images and apply the belief function labelling (Alg. 2) to obtain $q_x = \text{Pl}(o \rightarrow L_x)$ and $p_{\bar{x}} = \text{Bel}(o \rightarrow L_x) = 1 - q_x$, where x is BG, PC3 respectively. Next, we process superresolution images of fluorescence labelled Cav1 deposits in PC3-PTRF (shorthand P3) cells. PC3 cells contain both BG and SC objects, or more formally $q_{PC3} = q_{SC} + q_{BG}$, therefore $q_{SC} = \max(q_{PC3} - q_{BG}, 0)$. The max formulation ensures the correct assignment to 0 plausibility when for outlier objects $q_{PC3} < q_{BG}$. The subtraction of plausibility functions represents the elimination of the maximum support of a subset (BG) from a superset (PC3) to correctly bracket the maximum support of the subset $SC = PC3 \setminus BG$.

We know that objects unique to PC3-PTRF cells are (formations of) caveolae (C), therefore $p_C = \text{Bel}(op_3 \rightarrow L_C) = \text{Bel}(op_3 \rightarrow L_{BG}) \wedge \text{Bel}(op_3 \rightarrow L_{SC}) = p_{BG} * p_{SC}$. We visualise the results for a single PC3-PTRF cell in Figure 1-E where blue, green, and red gradients correspond with q_{BG} , q_{SC} and p_{P3} , respectively. From visual inspection, we see correlation of colocalized PTRF with objects labelled with a high p_{P3} value, as expected (Fig. 1-E.a,d). More interestingly, we can now identify objects that are transitioning between SC and C (Fig. 1-E.c). To confirm this, we next perform extensive validation.

2) *Validation*: The scientific method obliges us to try, to the best of our ability, to falsify [33] our results. We use two independent sources of information, not leveraged during the design of the method. First, we know from previous art that the frequency of caveolae in the PC3-PTRF cell line has been reported at $\sim 20\%$ [2], when compared to other Cav1 structures. In Fig. 3-A, we show the cumulative distribution function (cdf) of p_C . The 20% caveolae frequency coincides with a sudden and sustained increase of the cdf ($p_C \sim 0.32$), or more formally the knee or elbow of the function [34]. Given the imbalanced frequency, we expect to see the classification boundary to coincide with a sudden increase in frequency of a probability label. The bi-modality of the right hand side of the probability density function (pdf) is indicative of 2 high deposits of object types, namely SC and BG. We observe that the network analysis [2] based decision boundary for caveolae in PC3-PTRF is consistent with our belief label, given that the cdf and pdf of our label show a marked and sustained increase in frequency ('elbow' of the cdf) coinciding with the network analysis decision boundary. Second, we know that caveolae can only form in presence of PTRF. Therefore we expect to see an increasing correlation of PTRF-Cav1 colocalization as p_{P3} increases. We compute PTRF colocalization P by measuring the mean PTRF intensity for each Cav1 object. The regression computes a linear model between p_{P3} and P for all objects, for all cells, per replicate (Fig. 3-C, replicate is a repeat experiment to ensure consistency). PTRF colocalization increases markedly when p_{P3} increases. In Fig. 4-C a LOWESS-regression [35] is computed to discover a more nuanced behavior in the correlation with PTRF association.

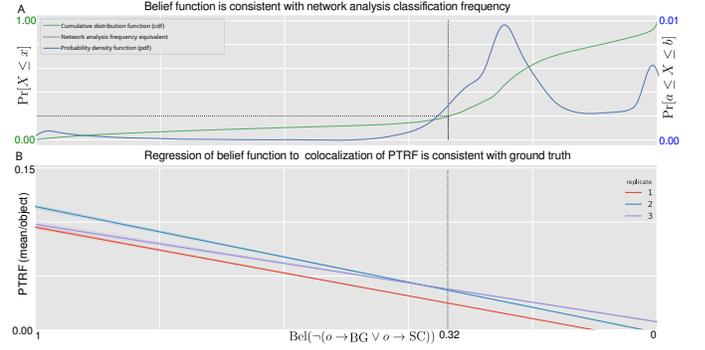


Fig. 3. Validation of the probabilistic labelling. The network analysis based classification reports a $\sim 20\%$ frequency of caveolae in PC3-PTRF cells. This threshold coincides with the inflection point of the cumulative density function ($\text{Bel}_Q(o) \geq x$) of our labelling (A) a label of ~ 0.32 ($\text{Bel}(\neg(o \rightarrow SC \vee o \rightarrow BG))$). Our belief function has a clear correlation with presence of PTRF (B), required for caveolae to form and unique to PC3-PTRF. Results are consistent across replicates ($n=3$).

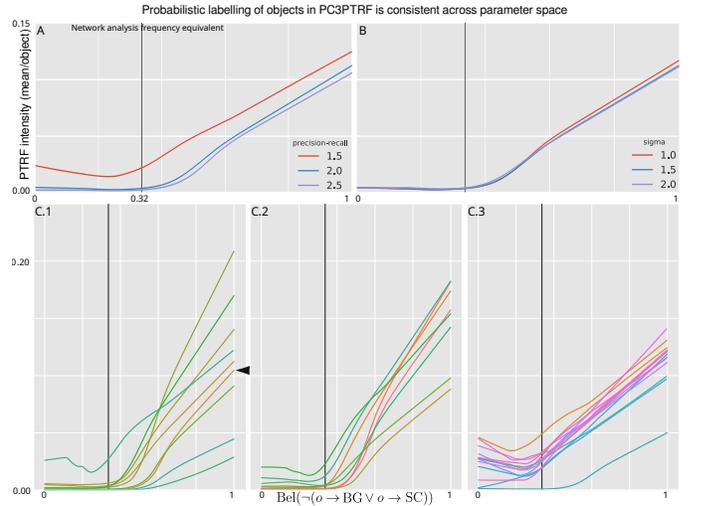


Fig. 4. Parameter sensitivity study shows increased correlation with mean PTRF is invariant to our parameter space. Precision-recall (A) and sigma (B) for cell 5 of replicate 1, the median of the validation results (C.1- ◀, are tested. Results for all cells, all replicates (C.1-3).

All cells show a consistent pattern across replicates. PTRF association markedly increases from non-responsive to a linear increase with the inflexion point of the LOWESS curve around ~ 0.35 , close to the network analysis based decision boundary equivalent of p_{P3} at ~ 0.32 . The linear increase beyond the inflection point reflects the expected gradual increase as PTRF is recruited to construct caveolae.

3) *Parameter sensitivity study*: Our method has 2 parameters: the Gaussian σ (std, Alg. 1) used in the smoothing and the precision-recall balance. Sigma should be at or below system precision to avoid creating artificially joined objects. For the Cav1 dataset, we omit the first Gaussian filter (σ_1 , Alg. 1), the sigma reported here is σ_2 . In superresolution microscopy, a deconvolution operation tuned to the acquisition specific point spread function is more accurate in restoring the signal. PRC is set at the users discretion; it is nonetheless important to document what its exact impact on the result can be. In Figure 4-A we compute the results for replicate 1, Cell 5, the

median of the trend (Fig. 4-C). A lower PRC (1.5) results in fewer, brighter objects dominating the selection. Fewer spots similar to non-specific Cav1 binding will be included explaining the upward shift of the curve while retaining the trend. When PRC is high (2.5) the inverse process occurs with BG spots driving the mean PTRF association lower. A larger sigma ($2 \leftrightarrow 1$) can lead to low intensity borders being included into the mask of an object. When those pixels are outside of the actual caveolae structure the expected PTRF association is not that of caveolae but of background, reducing the mean PTRF association, resulting in lowered correlation. We conclude that our parameter space does not invalidate our results with the two independent sources of information. Our method is therefore capable of extracting and identifying Cav1 structures in STED superresolution microscopy.

B. Identifying retinal amyloid- β deposits associated with Alzheimer disease

Identification and quantification of amyloid- β ($A\beta$) deposits in the retina in relation to Alzheimer disease (AD) is an open research problem [5]. In previous studies using confocal microscopy on post-mortem donor tissues, retinal $A\beta$ quantification was performed manually by blinded raters [36] or semi-automatically with manual segmentation. [37]. The resulting measurement of retinal $A\beta$ would be tested for its relationship with age, AD neuropathology, retinal regions, and other measures of interest. As the scarcity of postmortem retinal tissues from neuropathologically confirmed AD donors often limits the size of such data, variability from acquisition and manual raters can affect the quantification of retinal $A\beta$ and pose a challenge to achieving statistical significance. We illustrate how we can extend our method for measuring $A\beta$ across three heterogeneous sparse datasets of fluorescence confocal microscopy images of retinal cross-sections after $A\beta$ -specific immunohistochemistry, acquired using two different microscopes each operated by a different researcher. Rather than counting objects in the image, we use the belief function to identify which fluorescent marker deposits are more likely to be present in an AD image.

1) *Applying belief functions to identify AD across heterogeneous data:* We collected the following sets of images and labels:

- I_{H_1}, L_H : retinal tissue from healthy donors, microscope 1, $n=2$
- I_{H_2}, L_H : retinal tissue from healthy donors, microscope 2, $n=3$
- I_{D_1}, L_{AD} : retinal tissue from AD-confirmed donors, microscope 1, $n=3$.

We show an example image from each set in Figure 5. We identify fluorescent objects in all healthy images using Alg. 1 and obtain $q_{x_{L_H}} = \text{Pl}(o \rightarrow L_H)$ where x indicates which set of healthy images is used (1,2). Next, for each object detected in each AD image, we obtain as before $p_{\overline{x_{L_H}}} = 1 - q_{x_{L_H}} = p_{x_{L_{AD}}}$. In Figure 5-D,F we illustrate the outcome of $x=1, x=2$ respectively. We use Dempster's combination rule [38]:

$$m(A) = \frac{\sum m_1(B)m_2(C)|B \cap C = A}{\sum m_1(B)m_2(C)|B \cap C \neq \emptyset}, |A \subset \Theta \quad (7)$$

to define a joint belief function that combines the evidence from both sources to support a proposition $A (o \rightarrow L)$, while allowing the expression of the disagreement. Dempster's rule uses probability mass functions, which we can obtain from our belief functions by observing that our propositions ($o \rightarrow L$) are singleton focal elements, therefore in our case $\text{Bel}(A) = m(A) (m(A) = \sum_{B \subset A} (-1)^{|A \setminus B|} \text{Bel}(B))$. We enumerate in

Table I the intermediate results needed to compute the joint mass function for our use case. Let for a proposition $A = (o \rightarrow L)$ the probability mass $m_{H_1}(A) = t$ and $m_{H_2}(A) = s$ respectively. The table is indexed by subsets of all propositions (Θ) on which the belief functions are defined. An entry in the table on row B, column C represents $m_{H_1}(B \cap C) * m_{H_2}(B \cap C)$. The joint mass function $m_{H'}(A)$ is then given by:

TABLE I

DEMPTER COMBINATION RULE FOR ALZHEIMER USE CASE. $A, B, C \subset \Theta$

$B \cap C \rightarrow [0, 1]$	$m_{H_1}(A) = t$	$m_{H_1}(\overline{A}) = 1 - t$
$m_{H_2}(A) = s$	$A \rightarrow ts$	$\emptyset \rightarrow s(1 - t)$
$m_{H_2}(\overline{A}) = 1 - s$	$\emptyset \rightarrow (1 - s)t$	$\overline{A} \rightarrow (1 - s)(1 - t)$

$$m_{H'}(A) = \frac{ts}{1 - ((1 - s)t + (1 - t)s)} \quad (8)$$

Combining sources of evidence should be accompanied by a quantification of their disagreement or 'conflict' to allow a practitioner transparency in the construction and usage of the joint model. The weight (W) of conflict of the joint mass function, an expression of the disagreement between the two models, is given by the logarithm of the normalisation term $W = -\log(1 - ((1 - s)t + (1 - t)s))$. Combination is not meaningful when both sources are in complete contradiction, that is $(t, s) = (0, 1) \vee (1, 0)$. In such cases W is infinite, allowing the practitioner a sanity check for combination. By formulation a closed form expression for the joint model we satisfy the requirement to span heterogeneous data (Sec. I-B8).

2) *Results:* In Figure 5-D,E,F we illustrate the difference between the separate and joint belief functions. The individual belief functions are consistent in their results with respect to each other and the visually easily observable AB-deposits. The joint belief function combines both models to offer a weighted combination of the evidence provided by each model. In Figure 6 we plot the weight of conflict of the joint belief function for all 3 AD images. The weight of conflict is the smallest at both extrema of the joint belief function, indicating that the models from the two different microscopes agree the most for the objects that are strongly believed to be from in a healthy or AD retina by the joint belief function, while there is a greater disagreement for the objects without strong belief. A practitioner can use the weight of conflict for each object-prediction pair to quantify the agreement between multiple sources of evidence along with the output of joint evidence based on the joint belief function. (Sec. I-B5).

IV. DISCUSSION AND CONCLUSION

The motivation for this work was the need for a robust, adaptive, and self-tuning unsupervised probabilistic object

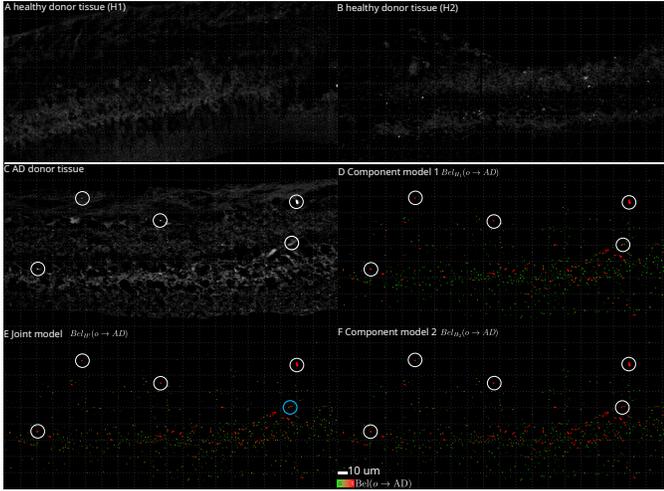


Fig. 5. Results of the separate and joint belief functions on Alzheimer’s retina study. Examples of the healthy retina images from different microscopes (H1, H2; panels A, B) and the AD retina image from microscope 1 (C). Visually easily observable amyloid- β deposits are marked by white circles in C. The AD image is analysed using the H1 (D), H2 (F) and joint H⁺ (E) models. Objects are labelled as a function of the $\text{Bel}(o \rightarrow \text{AD}^+) \mapsto [0, 1]$ function where 1 (red) indicates strong evidence support for the out of band AD image level label. Despite high inter-operator and acquisition variance, the predictions of all three models are consistent for visually observable amyloid- β objects (white circles) with disagreement exemplified by the blue marking.

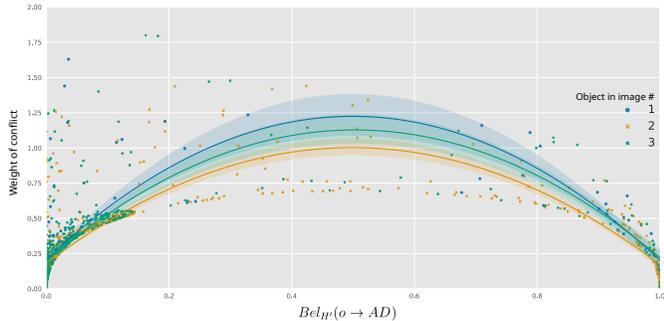


Fig. 6. The weight of conflict (y-axis, conflict between component belief functions) as a function of the joint belief function $\text{Bel}_{H^+}(o \rightarrow \text{AD})$ (x-axis) for the Alzheimer pilot study for all 3 AD images. A second order regression is fit through the points, with each point representing the label of an object in one of the 3 images. Minimum conflict occurs at either ends of the range of Bel_{H^+} , with minimal conflict maximal for $\text{Bel}_{H^+} \sim 0.5$

detection method applicable to heterogeneous multi-scale super-resolution microscopy. while it is feasible to use a larger number of more elaborate features to describe objects, e.g. deep learning, during development we found using simple low-dimensional features and statistical modelling obtained results validated by biological ground truth. Per Occam’s razor, the introduction of more complex models would not have been justified. We note that our formulation of belief functions makes them separable ($A \cap B$ is a focal element) and consonant ($A \subset B$ or $B \subset A$) support functions [6]. As a result, our p and q functions are equivalent to ‘necessity’ and ‘possibility’ functions from possibility theory [39]. We note that the joint model can also be formulated when objects from 2 different models overlap, for example, when we run our method with different σ and PRC values to obtain two models, one with

high recall, one with high precision. In such a joint model, we now have for each object o an inner, smaller object o' . One formulation for focal elements then can be: $A = \{(o \rightarrow L) \wedge (o' \rightarrow L)\}$, $B = \{(o \rightarrow L) \wedge \neg(o' \rightarrow L)\}$, leading to a more complex formulation for a joint model. A more interesting use case is when the object detection is fuzzy and allows for non-empty intersections. Due to space constraints we discuss the computation of uncertainty in Appendix C.

A. Limitations

When the intensity profile of the fluorescence diverges from a generalised normal distribution, our object detection will increasingly fail and split objects into parts; a different detection method is then warranted [40]. The Mahalanobis distance can be uninformative in high-dimensional space due to the ‘curse’ of dimensionality, however, this is only the case if the increase in dimensions is due to non-discriminatory features [41]. While the joining of belief functions by Dempster’s rule is not without criticism [42], we note that the preconditions [42] for its use are satisfied in our case with independent evidence sources and exclusive exhaustive hypotheses. In future work, we aim to adopt advances in evidence combination [43] to enable quantification of reliability of individual sources and make the joint model robust against unreliable sources.

B. Conclusion

We introduced a novel adaptive self-tuning method for object detection in 2D microscopy images of 3D fluorescent labelled proteins that enables consistent results across channels, and a novel method to assign each object a belief that expresses numerically the evidence encoded. We validated our method on superresolution data of Cav1 deposits where we showed agreement with related work and biological ground truth. We showed we are able to identify and characterize Cav1-labeled caveolae and scaffolds by STED superresolution microscopy, setting the stage for robust, reproducible temporal live cell analysis where consistency across images and channels is essential for scientific discovery. We applied our method on an Alzheimer pilot study illustrating the multiscale applicability. We illustrated with a closed form expression the capability to formulate a joint model spanning heterogeneous datasets while recording the conflict of evidence between the separate models as a reliability measure.

REFERENCES

- [1] I. M. Khater, F. Meng, T. H. Wong, I. R. Nabi, and G. Hamarneh, “Super resolution network analysis defines the molecular architecture of caveolae and caveolin-1 scaffolds,” *Scientific Reports*, vol. 8, no. 1, pp. 1–15, 2018.
- [2] I. M. Khater, Q. Liu, K. C. Chou, G. Hamarneh, and I. R. Nabi, “Super-resolution modularity analysis shows polyhedral caveolin-1 oligomers combine to form scaffolds and caveolae,” *Scientific Reports*, vol. 9, no. 1, pp. 1–10, 2019.
- [3] S. W. Hell and J. Wichmann, “Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy,” *Optics Letters*, vol. 19, no. 11, pp. 780–782, 1994.
- [4] G. Gao, C. Zhu, E. Liu, and I. R. Nabi, “Reticulon and climp-63 regulate nanodomain organization of peripheral er tubules,” *PLoS Biology*, vol. 17, no. 8, p. e3000355, 2019.

- [5] N. J. Hart, Y. Koronyo, K. L. Black, and M. Koronyo-Hamaoui, "Ocular indicators of alzheimer's: exploring disease in the retina," *Acta Neuropathologica*, vol. 132, no. 6, pp. 767–787, 2016.
- [6] G. Shafer, *A mathematical theory of evidence*. Princeton university press, 1976, vol. 42.
- [7] J. P. Ioannidis, "Why most published research findings are false," *PLoS Medicine*, vol. 2, no. 8, p. e124, 2005.
- [8] S. Andrews, N. Changizi, and G. Hamarneh, "The isometric log-ratio transform for probabilistic multi-label anatomical shape representation," *IEEE Transactions on Medical Imaging*, vol. 33, no. 9, pp. 1890–1899, 2014.
- [9] S. Jackman, *Bayesian analysis for the social sciences*. John Wiley & Sons, 2009, vol. 846.
- [10] A. P. Dempster, "The dempster-shafer calculus for statisticians," *International Journal of Approximate Reasoning*, vol. 48, no. 2, pp. 365–377, 2008.
- [11] A. BenTaieb, J. Kawahara, and G. Hamarneh, "Multi-loss convolutional networks for gland analysis in microscopy," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2016, pp. 642–645.
- [12] W. C. Schwartzkopf, A. C. Bovik, and B. L. Evans, "Maximum-likelihood techniques for joint segmentation-classification of multispectral chromosome images," *IEEE Transactions on Medical Imaging*, vol. 24, no. 12, pp. 1593–1610, 2005.
- [13] S. Mehta, E. Mercan, J. Bartlett, D. Weaver, J. G. Elmore, and L. Shapiro, "Y-net: joint segmentation and classification for diagnosis of breast biopsy images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 893–901.
- [14] F. Girard, C. Kavalec, and F. Chriet, "Joint segmentation and classification of retinal arteries/veins from fundus images," *Artificial Intelligence in Medicine*, vol. 94, pp. 96–109, 2019.
- [15] H. Qu, G. Riedlinger, P. Wu, Q. Huang, J. Yi, S. De, and D. Metaxas, "Joint segmentation and fine-grained classification of nuclei in histopathology images," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 900–904.
- [16] A. Singh, S. Sengupta, and V. Lakshminarayanan, "Explainable deep learning models in medical image analysis," *Journal of Imaging*, vol. 6, no. 6, 2020.
- [17] M. Reyes *et al.*, "On the interpretability of artificial intelligence in radiology: Challenges and opportunities," *Radiology: Artificial Intelligence*, vol. 2, no. 3, p. e190043, 2020.
- [18] S. Qiu *et al.*, "Development and validation of an interpretable deep learning framework for Alzheimer's disease classification," *Brain*, vol. 143, no. 6, pp. 1920–1933, 05 2020.
- [19] Z. Zhang, Y. Xie, F. Xing, M. McGough, and L. Yang, "Mdnet: A semantically and visually interpretable medical image diagnosis network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [20] S. A. Taghanaki *et al.*, "Infomask: Masked variational latent representation to localize chest disease," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 739–747.
- [21] O. Z. Kraus, J. L. Ba, and B. J. Frey, "Classifying and segmenting microscopy images with deep multiple instance learning," *Bioinformatics*, vol. 32, no. 12, pp. i52–i59, 2016.
- [22] M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple instance learning: A survey of problem characteristics and applications," *Pattern Recognition*, vol. 77, pp. 329–353, 2018.
- [23] G. Quellec, G. Cazuguel, B. Cochener, and M. Lamard, "Multiple-instance learning for medical image and video analysis," *IEEE Reviews in Biomedical Engineering*, vol. 10, pp. 213–234, 2017.
- [24] A. G. Vecchiarelli *et al.*, "Membrane-bound minde complex acts as a toggle switch that drives min oscillation coupled to cytoplasmic depletion of mind," *Proceedings of the National Academy of Sciences*, vol. 113, no. 11, pp. E1479–E1488, 2016.
- [25] T. Lindeberg, "Scale-space theory: A basic tool for analyzing structures at different scales," *Journal of Applied Statistics*, vol. 21, no. 1-2, pp. 225–270, 1994.
- [26] J. J. A. Moors, "The meaning of kurtosis: Darlington reexamined," *The American Statistician*, vol. 40, no. 4, pp. 283–284, 1986.
- [27] L. Liu and R. R. Yager, *Classic works of the Dempster-Shafer theory of belief functions: An introduction*. Springer, 2008, vol. 219.
- [28] F. P. Cantelli, "Sui confini della probabilita," in *Atti del Congresso Internazionale dei Matematici: Bologna del 3 al 10 de settembre di 1928*, 1929, pp. 47–60.
- [29] A. Hayer, M. Stoeber, C. Bissig, and A. Helenius, "Biogenesis of caveolae: stepwise assembly of large caveolin and cavin complexes," *Traffic*, vol. 11, no. 3, pp. 361–382, 2010.
- [30] S. Monier, R. G. Parton, F. Vogel, J. Behlke, A. Henske, and T. V. Kurzchalia, "Vip21-caveolin, a membrane protein constituent of the caveolar coat, oligomerizes in vivo and in vitro," *Molecular Biology of the Cell*, vol. 6, no. 7, pp. 911–927, 1995.
- [31] P. Lajoie, J. G. Goetz, J. W. Dennis, and I. R. Nabi, "Lattices, rafts, and scaffolds: domain regulation of receptor signaling at the plasma membrane," *Journal of Cell Biology*, vol. 185, no. 3, pp. 381–385, 2009.
- [32] M. Stoeber *et al.*, "Model for the architecture of caveolae based on a flexible, net-like assembly of cavin1 and caveolin discs," *Proceedings of the National Academy of Sciences*, vol. 113, no. 50, pp. E8069–E8078, 2016.
- [33] K. R. Popper, "Science as falsification," *Conjectures and Refutations*, vol. 1, pp. 33–39, 1963.
- [34] V. Satopaa, J. Albrecht, D. Irwin, and B. Raghavan, "Finding a" kneedle" in a haystack: Detecting knee points in system behavior," in *2011 31st International Conference on Distributed Computing Systems Workshops*. IEEE, 2011, pp. 166–171.
- [35] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, vol. 74, no. 368, pp. 829–836, 1979.
- [36] S. E. Perez, S. Lumayag, B. Kovacs, E. J. Mufson, and S. Xu, " β -amyloid deposition and functional impairment in the retina of the appwe/ps1 δ e9 transgenic mouse model of alzheimer's disease," *Investigative Ophthalmology & Visual Science*, vol. 50, no. 2, pp. 793–800, 2009.
- [37] S. Lee *et al.*, "Amyloid beta immunoreactivity in the retinal ganglion cell layer of the alzheimer's eye," *Frontiers in Neuroscience*, vol. 14, p. 758, 2020.
- [38] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," in *Classic Works of the Dempster-Shafer Theory of Belief functions*. Springer, 2008, pp. 57–72.
- [39] L. A. Zadeh, "Fuzzy sets as a basis for a theory of possibility," *Fuzzy Sets and Systems*, vol. 1, no. 1, pp. 3–28, 1978.
- [40] M. A. Mabaso, D. J. Withey, and B. Twala, "Spot detection methods in fluorescence microscopy imaging: A review," *Image Analysis & Stereology*, vol. 37, no. 3, pp. 173–190, 2018.
- [41] A. Zimek, E. Schubert, and H.-P. Kriegel, "A survey on unsupervised outlier detection in high-dimensional numerical data," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 5, no. 5, pp. 363–387, 2012.
- [42] F. Voorbraak, "On the justification of dempster's rule of combination," *Artificial Intelligence*, vol. 48, no. 2, pp. 171–197, 1991.
- [43] Y. Yang, D. Han, and C. Han, "Discounted combination of unreliable evidence using degree of disagreement," *International Journal of Approximate Reasoning*, vol. 54, no. 8, pp. 1197–1216, 2013.
- [44] T. H. Wong, I. M. Khater, B. Joshi, M. Shahsavari, G. Hamarneh, and I. R. Nabi, "Super-resolution network analysis shows cav1 csd mutant specifically alters intermediate scaffold structures," *Under preparation for submission to PLoS One*, vol. 0, no. 0, pp. W00–W00, 2020.
- [45] F. Meng, B. Joshi, and I. R. Nabi, "Galactin-3 overrides ptrf/cavin-1 reduction of pc3 prostate cancer cell migration," *PloS one*, vol. 10, no. 5, p. e0126056, 2015.
- [46] M. M. Hill *et al.*, "Ptrf-cavin, a conserved cytoplasmic protein required for caveola formation and function," *Cell*, vol. 132, no. 1, pp. 113–124, 2008.

APPENDIX

In this section we describe details that are not required for interpreting the results but necessary for their reproducibility. SPECHT is written as a Julia module leveraging the high level features this scientific programming language offers in combination with high performance. The source code is under embargo for patent filing, US provisional patent 63/071,557 filed 28th Aug 2020.

A. Data sets

1) *CAV1*: PC3, PC3-PTRF and CRISPR/Cas Cav1 KO MDA-MB-231 cells [1], [44]–[46] were cultured in RPMI-1640 medium (Thermo-Fisher Scientific Inc.) complemented with 10% fetal bovine serum (FBS, Thermo-Fisher Scientific Inc.) and 2 mM L-Glutamine (Thermo-Fisher Scientific Inc.) at 37 Celsius in a 5% CO₂/95% air incubator. Cells grown on 1.5H coverslips (Paul Marienfeld) were fixed with 3% paraformaldehyde (PFA), 15 min at room temperature, rinsed with PBS, permeabilized with 0.1% Triton X-100 in PBS plus 0.1 mM Ca²⁺ and 1 mM Mg²⁺ (PBS-CM) and blocked with 10% Goat Serum (Thermo Fisher Scientific, Waltham, MA) and 1% bovine serum albumin (Sigma, St. Louis, MO) in PBS-CM. Cells were incubated with the primary antibody (12h, 4 Celsius) and the secondary antibody (1h, room temperature). The primary and secondary antibodies were diluted in SSC (saline sodium citrate) buffer containing 1% BSA, 2% goat serum and 0.05% Triton X-100. Coverslips were mounted with Prolong Gold (Life Technologies, Thermo Fisher Scientific). Cells were washed after antibody incubations using SSC buffer containing 0.05% Triton X-100. Images were acquired with a 100x/1.4 Oil HC PL APO CS2 STED White objective of a Leica TCS SP8 3x STED microscope (Leica, Wetzlar, Germany) equipped with a white light laser, HyD detectors, time-gated fluorescence detection and Leica Application Suite X (LAS X) software. Acquisition was done at a scan speed of 600 Hz with a line average of 5. Pixel size is 20nm and resolution (precision) is around 70nm for the PTRF channel and 50nm for the Cav1 channel. GFP was excited at 488 nm and depleted at 592 nm. Alexa Fluor 647 was excited at 653nm and depleted at 775 nm. Huygens Professional software (Scientific Volume Imaging, Hilversum, NL) was used to deconvolve STED images, chromatic aberration correction was applied on PTRF images using the Cav1 channel as reference channel using the ‘correlation full’ method.

2) *Alzheimer*: Confocal microscopy images of retinal cross-sections after immunohistochemistry staining for amyloid- β . Tissues were obtained from control eyes from Eye Bank of BC, and AD eyes from donors with post-mortem neuropathological diagnosis of Alzheimer’s disease from UBC Department of Neurology. Tissues were processed as paraffin embedded cross-sections (5 μ m). BA4 primary antibody was used for specific binding for the first 2 amino acids of the AB peptide amino terminus, Cy3 secondary antibody was used to label BA4 in red fluorescence. Samples imaged at 543 nm wavelength using Zeiss LSM 510 at 0.44 μ m x 0.44 μ m pixel dimension over 450 μ m x 450 μ m area, and Zeiss Axio Imager M2 at 0.454 μ m x 0.454 μ m pixel dimension over 624.70 μ m

x 501.22 μ m area. LSM 510 images were resized to match the pixel dimension of those from Axio Imager M2. Images were manually segmented for the retinal layers, the vitreous and the region posterior to the outer nuclear layer were masked to reduce artefactual signals.

B. Self-tuning illustration

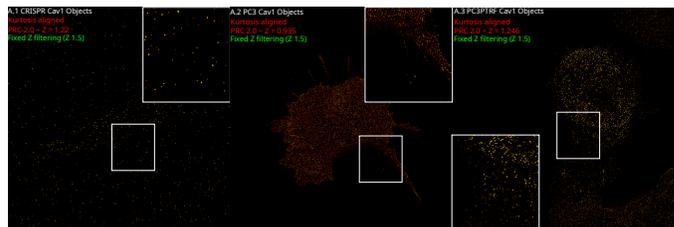


Fig. 7. Illustration of self-tuning object detection on the 3 different cell lines. A self-tuning (red) versus fixed Z-threshold (green) based object detection across cell lines illustrates how the different statistical signature of each image requires a self-tuning threshold (red) to have consistent results.

C. The value of uncertainty

1) *Defining uncertainty*: In this section, we briefly discuss the computation of ‘ r ’ (Eq. 5), the uncertainty in measuring the belief. For an object o , label x we have the plausibility $q_x = \text{Pl}(o \rightarrow L_x)$ and belief $p_x = \text{Bel}(o \rightarrow L_x) = 1 - q_{\bar{x}}$. Uncertainty, in this context, is defined as $r_x = q_x - p_x$. Intuitively this makes sense, one can interpret belief as the measurable support, whereas plausibility is the maximum potential support. When we divide the label space \mathcal{L} into ‘supports x ’ versus ‘does not support x ’, we have that \mathcal{L} is composed of two focal elements, x and \bar{x} . Note that while here x is a single label, \bar{x} is not. Examples where $|x| = |\bar{x}|$ are when x is ‘healthy’ versus ‘disease’. Given q_x , and our division of label space (and frame of discernment) into x and \bar{x} we can compute $p_{\bar{x}}$. Finally, we compute $r_x = q_x - p_x$, and $r_{\bar{x}} = q_{\bar{x}} - p_{\bar{x}}$. Our contribution gives us a way to compute both q_x and $q_{\bar{x}}$, therefore we can derive the belief functions, and with them the uncertainty, in essence a ‘top-down’ computation of belief functions.

2) *Interpreting uncertainty*: It is helpful to reflect what this ‘uncertainty’ actually means for the practitioners. Let us explore the most uncertain scenario, $q_x = 1$, and $p_x = 0$, with $r_x = 1$. Uncertainty increases with the inability of the belief function to obtain evidence ($p_x \rightarrow 0$). An inability to find evidence for a negative ($p_{\bar{x}} \rightarrow 0$), leads to $q_x \rightarrow 1$ and uncertainty increases. Let us now consider the converse, a scenario where uncertainty is minimal. Then it must be that $p_x = q_x$, and $p_{\bar{x}} = q_{\bar{x}}$. This can occur when the information measured is never neutral, the features always support or negate a label, but never both, and our capability to measure those features is perfect. In theory it is possible to obtain such a scenario, by increasing the dimensions of the feature space to infinity. However, one must ensure that the added dimensions (features) are maximizing information (support), otherwise we invoke the ‘curse’ of dimensionality [41]. A final analogy that can help is that of the balance between

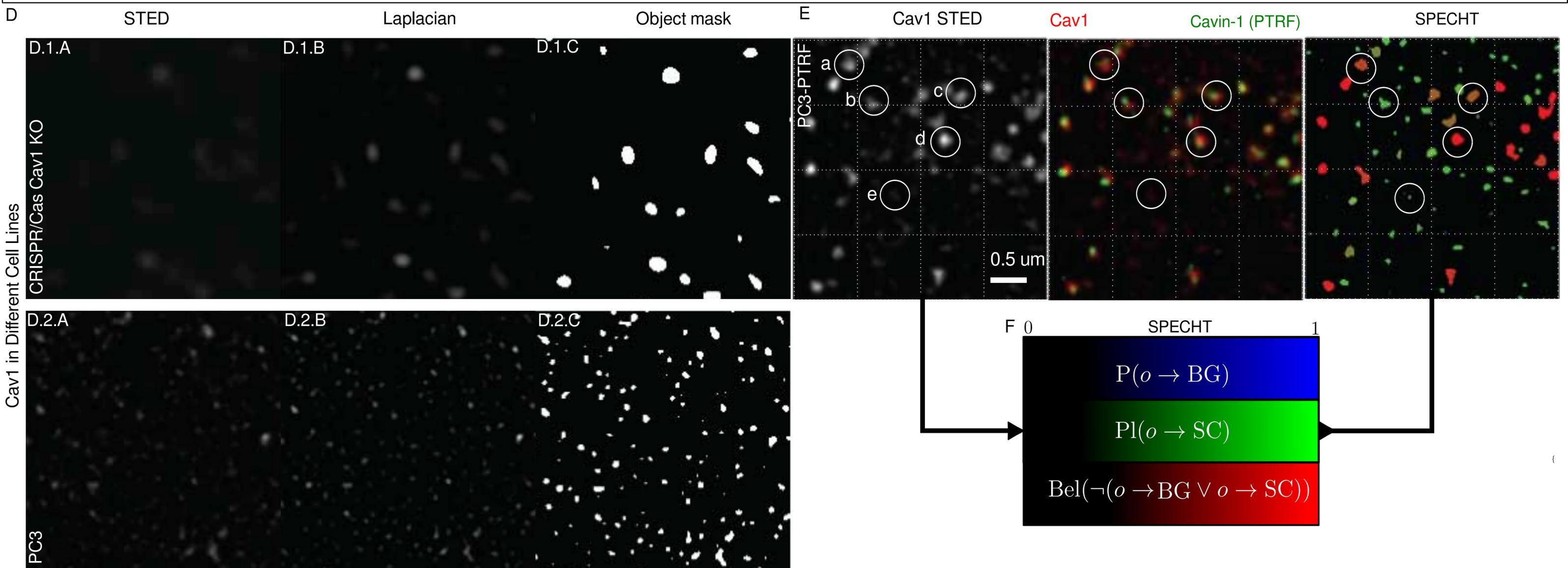
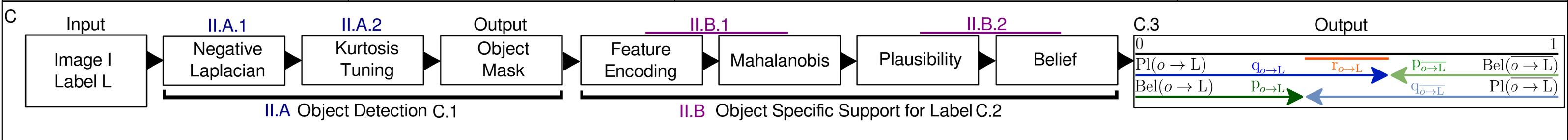
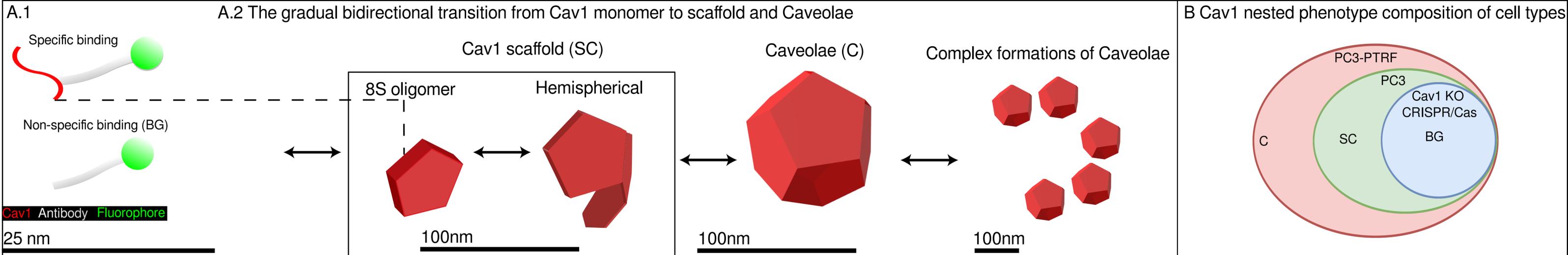
precision and recall. Inability to discern (believe in) the true support for a label would lead to low precision, and thus a belief tending to zero for label x . Low recall, conversely, is a belief function unable to discern support for a true negation of the label (\bar{x}). Uncertainty informs on both ($r_x, r_{\bar{x}}$, but in our setting is able to infer this without the need for annotation. If this is surprising, consider that quite often objects in an image support both the label x and its negation. When we label an image as ‘healthy’, versus ‘diseased’, we can be sure that both labels are supported by a confounded label of ‘tissue’, and perhaps the label of ‘background’ or non-tissue acquisition. The capture of the support for ‘healthy’ will include statistical support that is shared with ‘disease’, because not all tissue is affected equally, and before tissue is diagnosed as ‘diseased’ it has undergone a transition towards ‘diseased’. What we intuitively mean by ‘healthy’, is ‘healthy’ and never seen in ‘diseased’ and vice versa. In practice, this is not expressed in the labels, nor do we tend to encode this in the statistical learning methods employed. In these conditions, uncertainty can help to quantify exactly what we can capture. Due to space limitations and given that it is challenging to validate uncertainty, we have omitted empirical results on computing uncertainty on our datasets as they would be illustrative, rather than quantitative support for our method.

D. Practical limitations

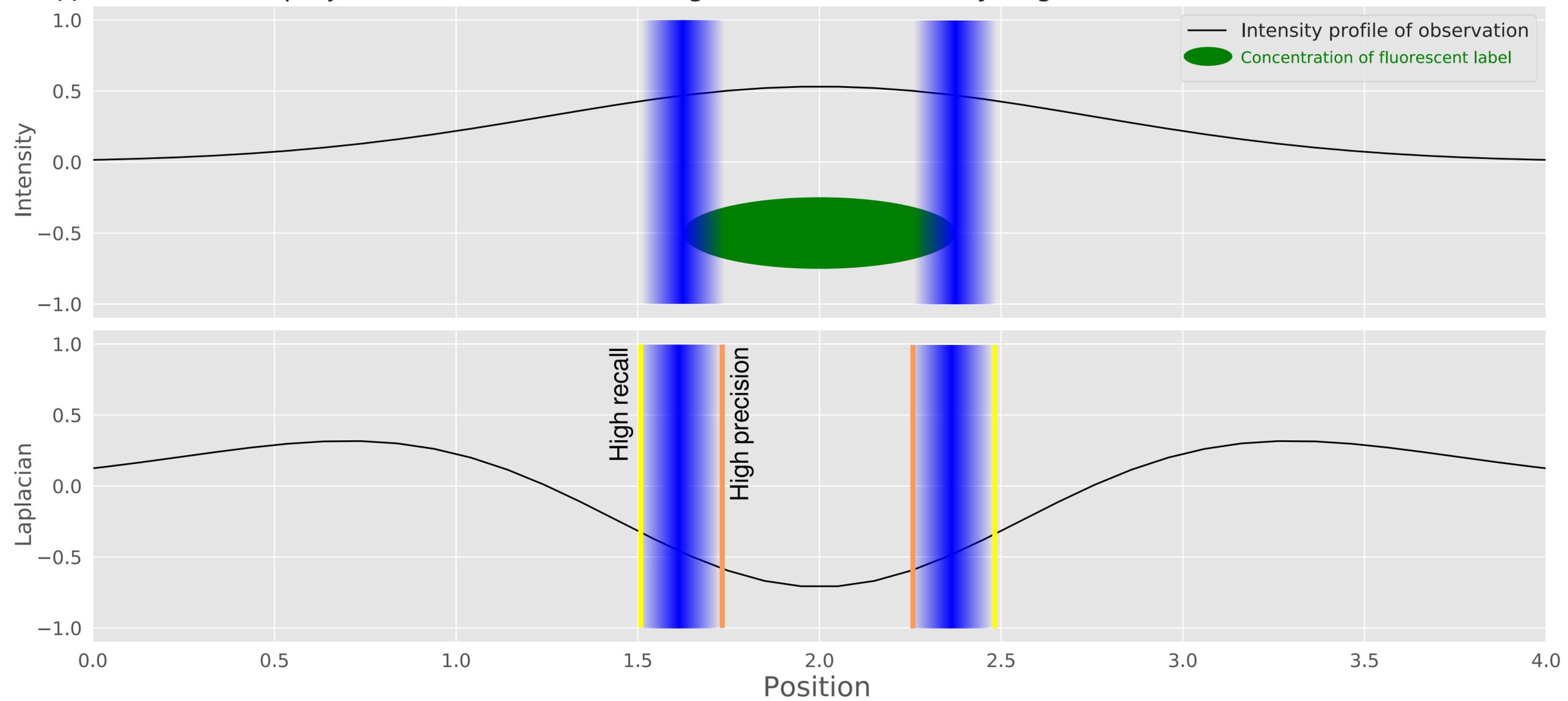
1) Numerical stability:

a) Combination rule: The closed form expression for the Alzheimer use case (Eq. 8) is sensitive to loss of precision due to catastrophic cancellation (loss of significance) when t, s are close to 0, we therefore correct t and s values to $\max(x, \epsilon)$ with ϵ the machine epsilon.

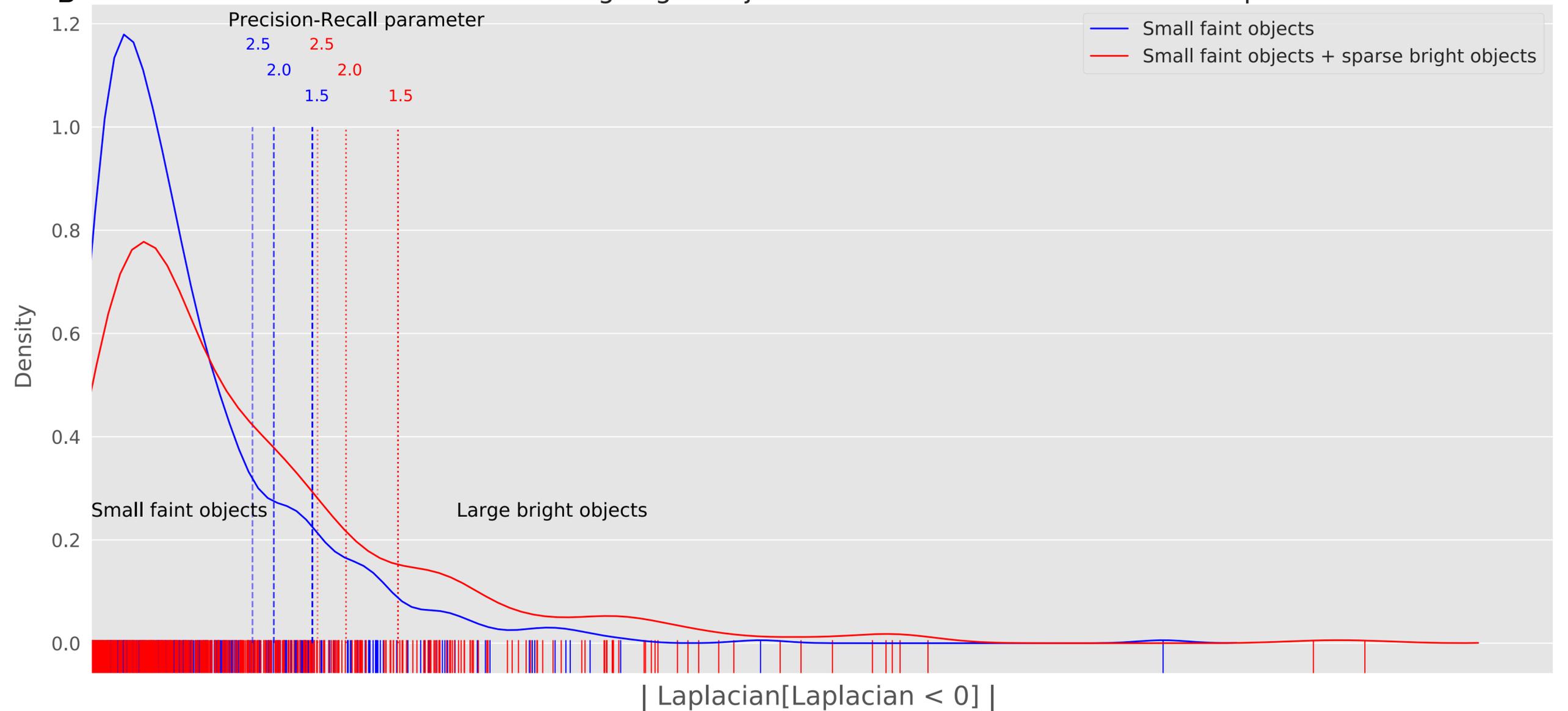
b) Kurtosis: Computing the 4th root of the kurtosis can be numerically unstable. However, we note that the alternative, for example, the geometric mean of V , more appropriate compared to the arithmetic mean given that V is a ratio, has similar issues in that it uses similar operations.



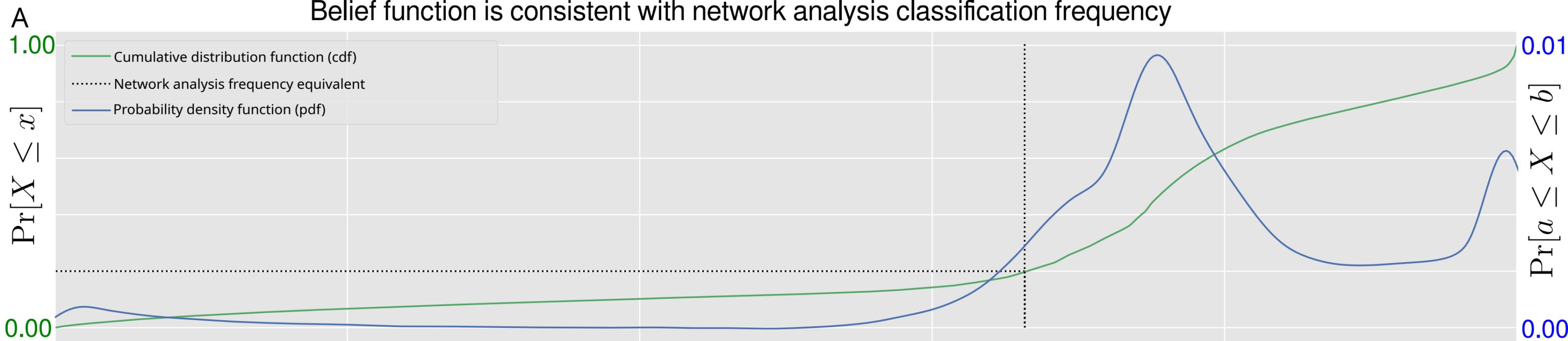
A 2D projection of 3D volumetric signal is demarcated by negative 2nd derivative.



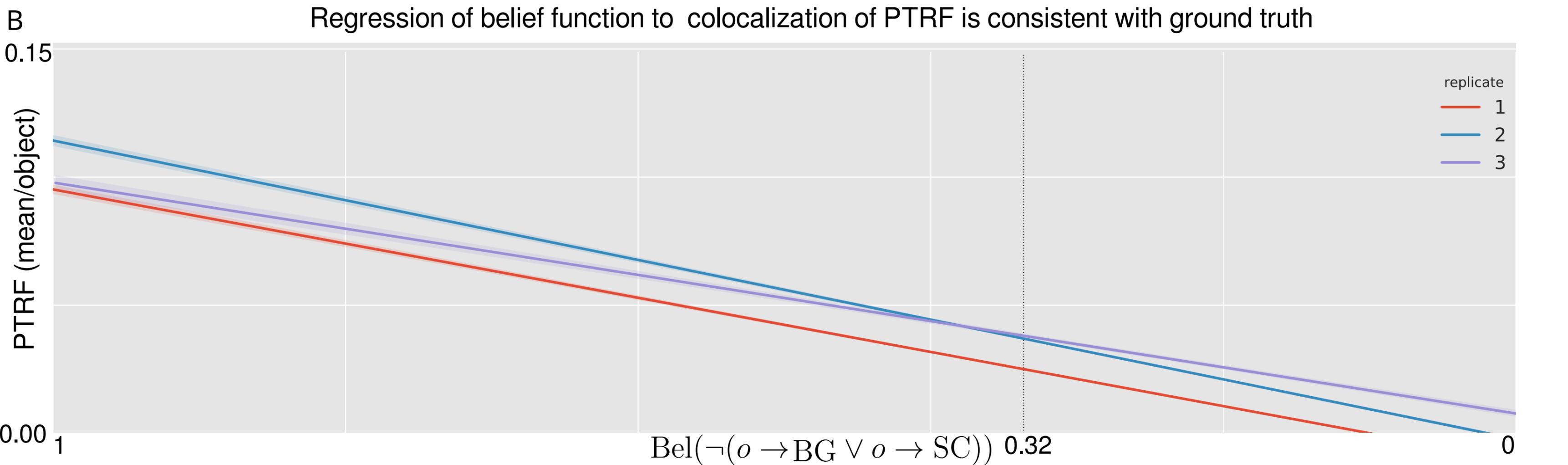
B Kurtosis based z-scaling aligns object detection across distribution shapes



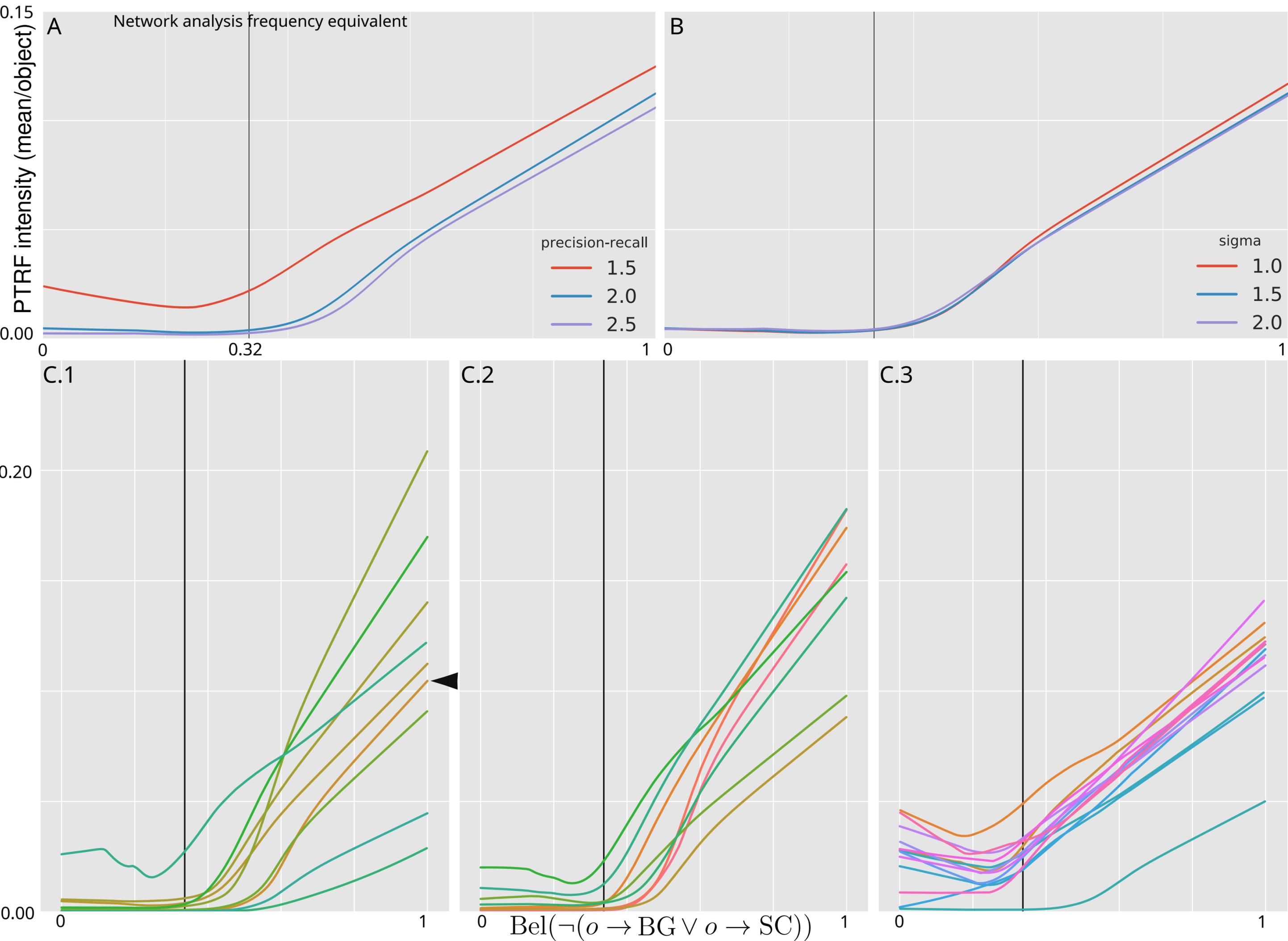
Belief function is consistent with network analysis classification frequency



Regression of belief function to colocalization of PTRF is consistent with ground truth



Probabilistic labelling of objects in PC3PTRF is consistent across parameter space



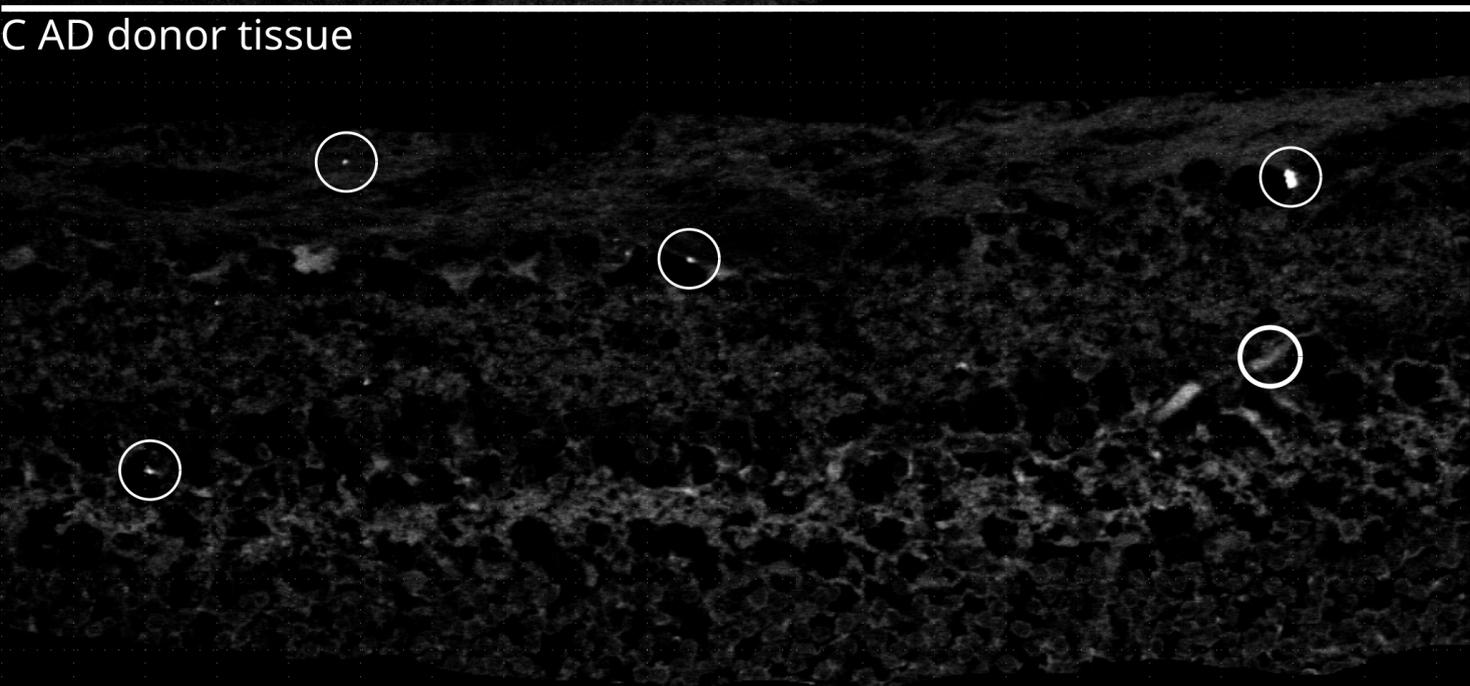
A healthy donor tissue (H1)



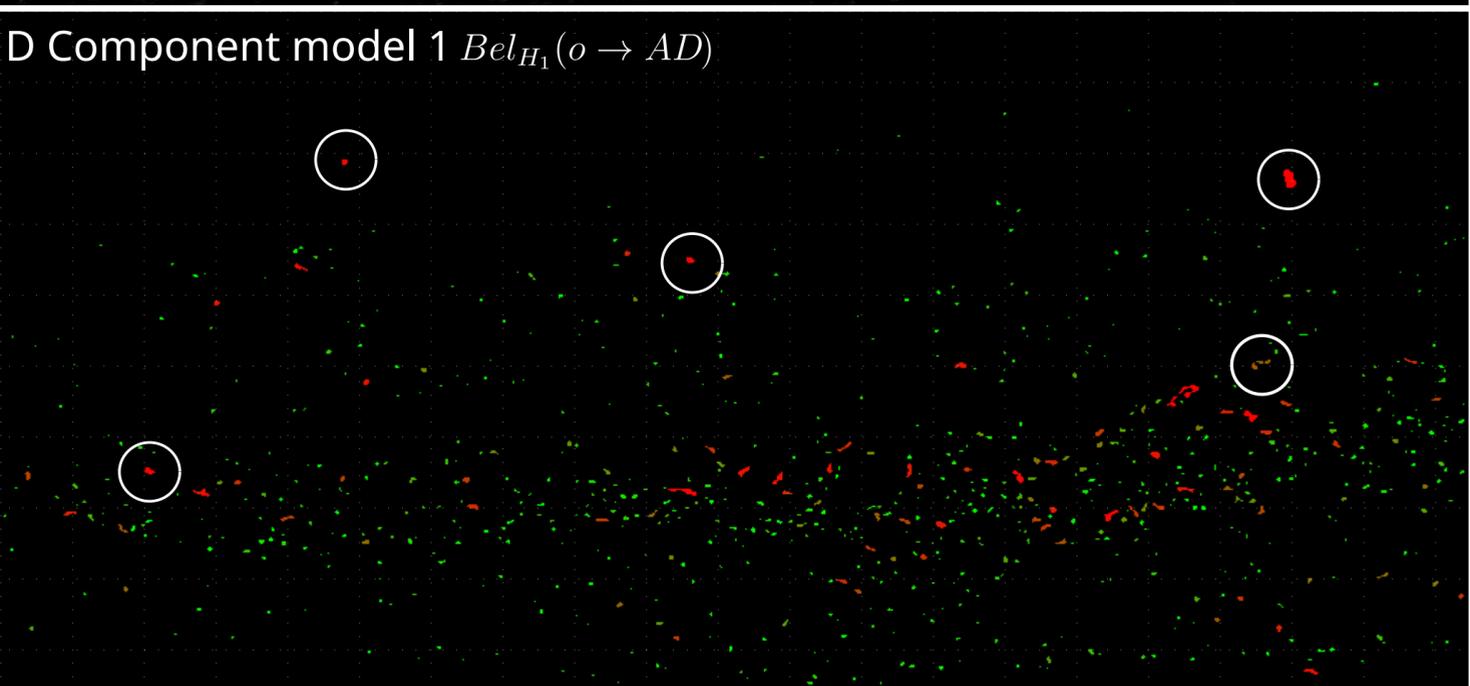
B healthy donor tissue (H2)



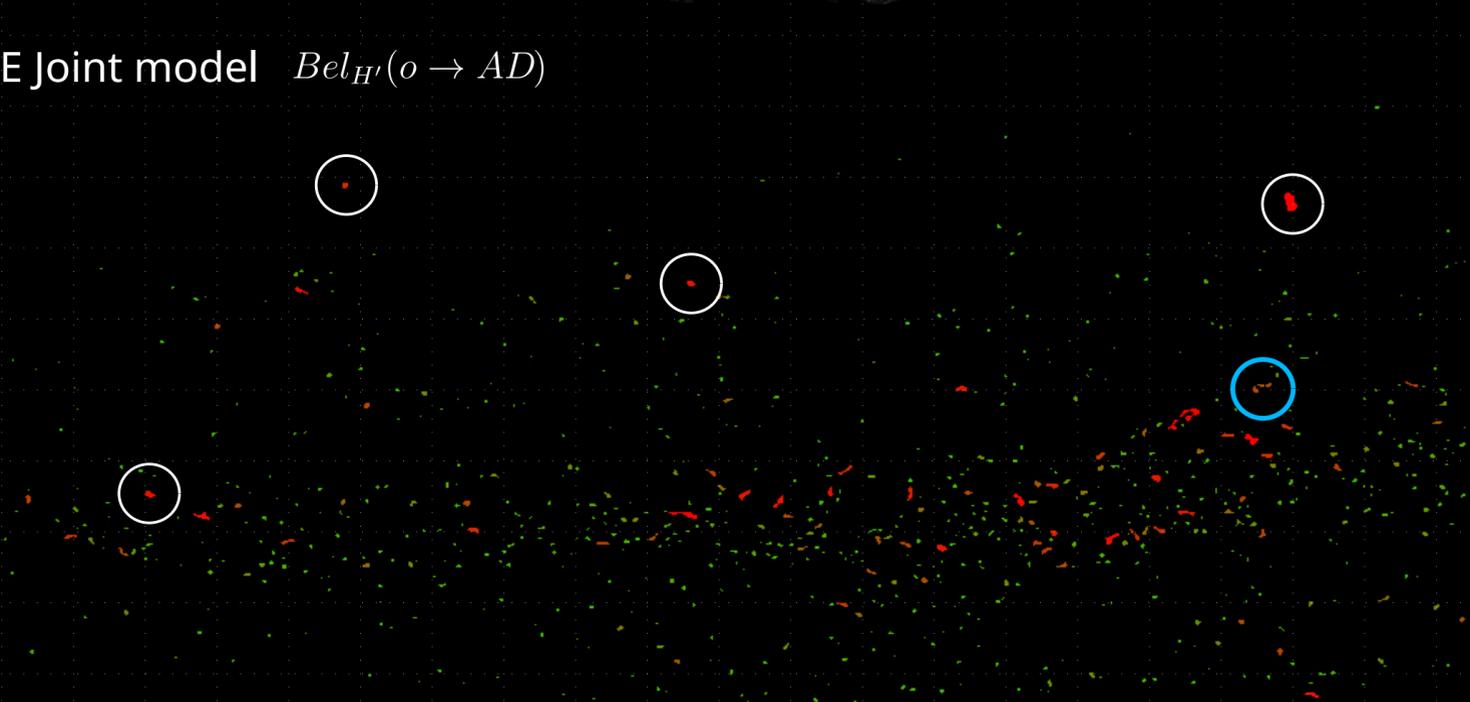
C AD donor tissue



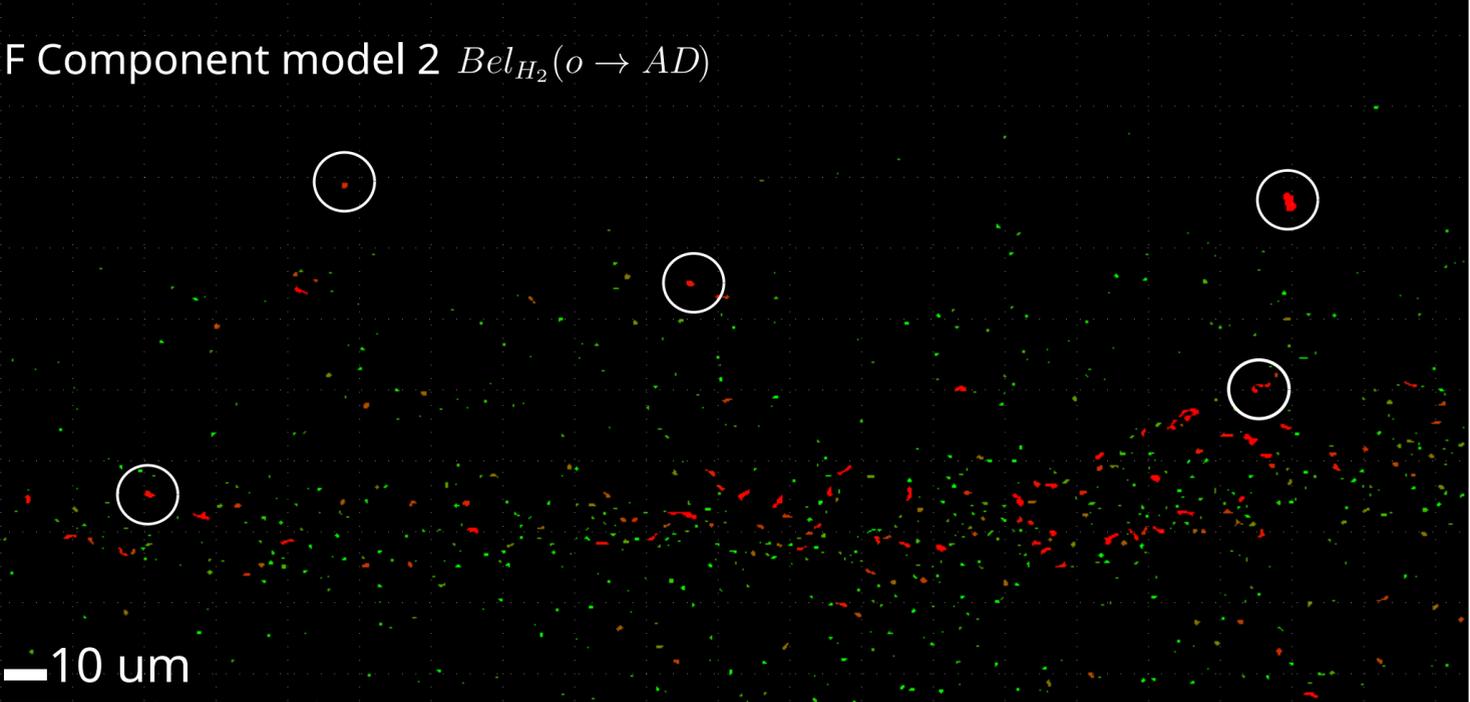
D Component model 1 $Bel_{H_1}(o \rightarrow AD)$



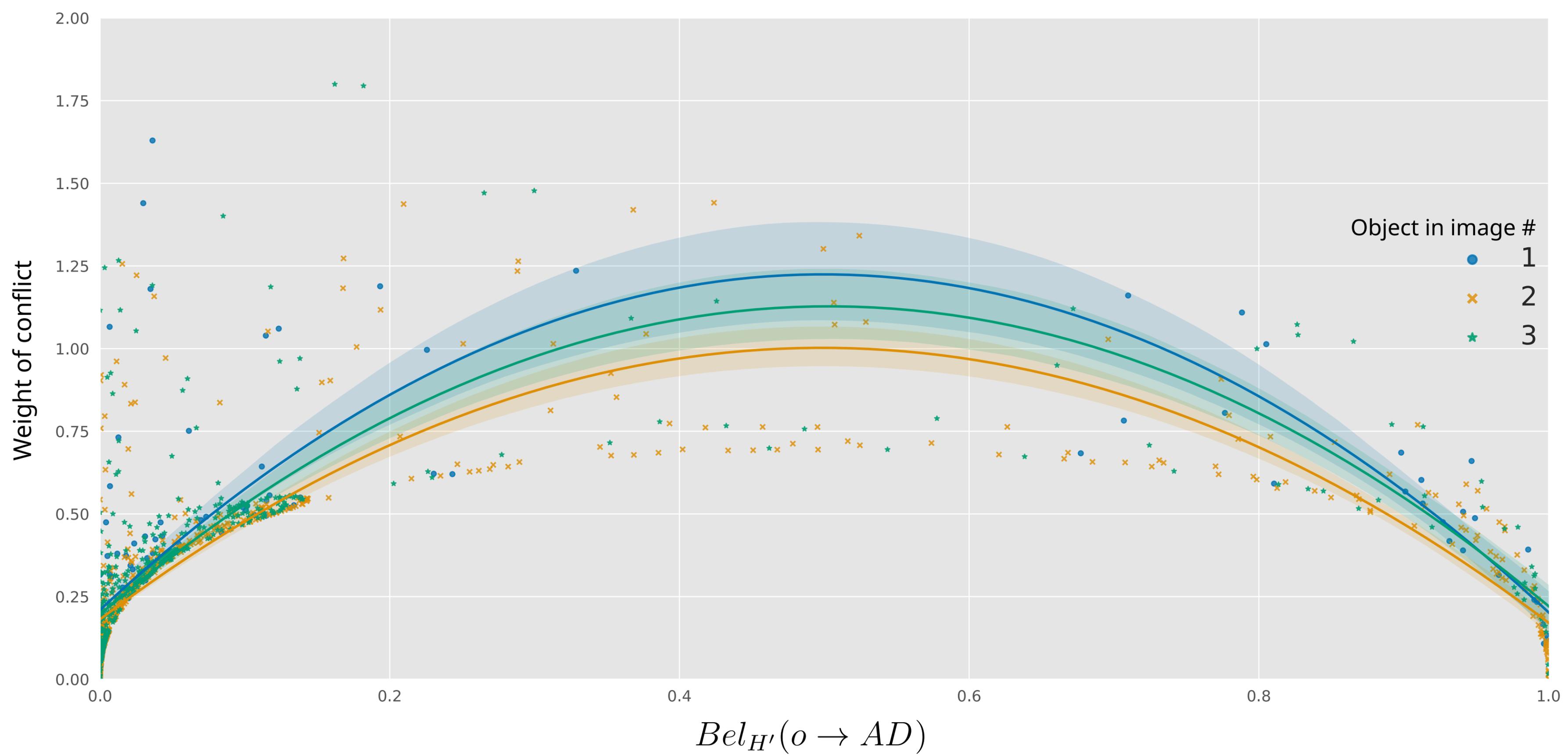
E Joint model $Bel_{H'}(o \rightarrow AD)$



F Component model 2 $Bel_{H_2}(o \rightarrow AD)$

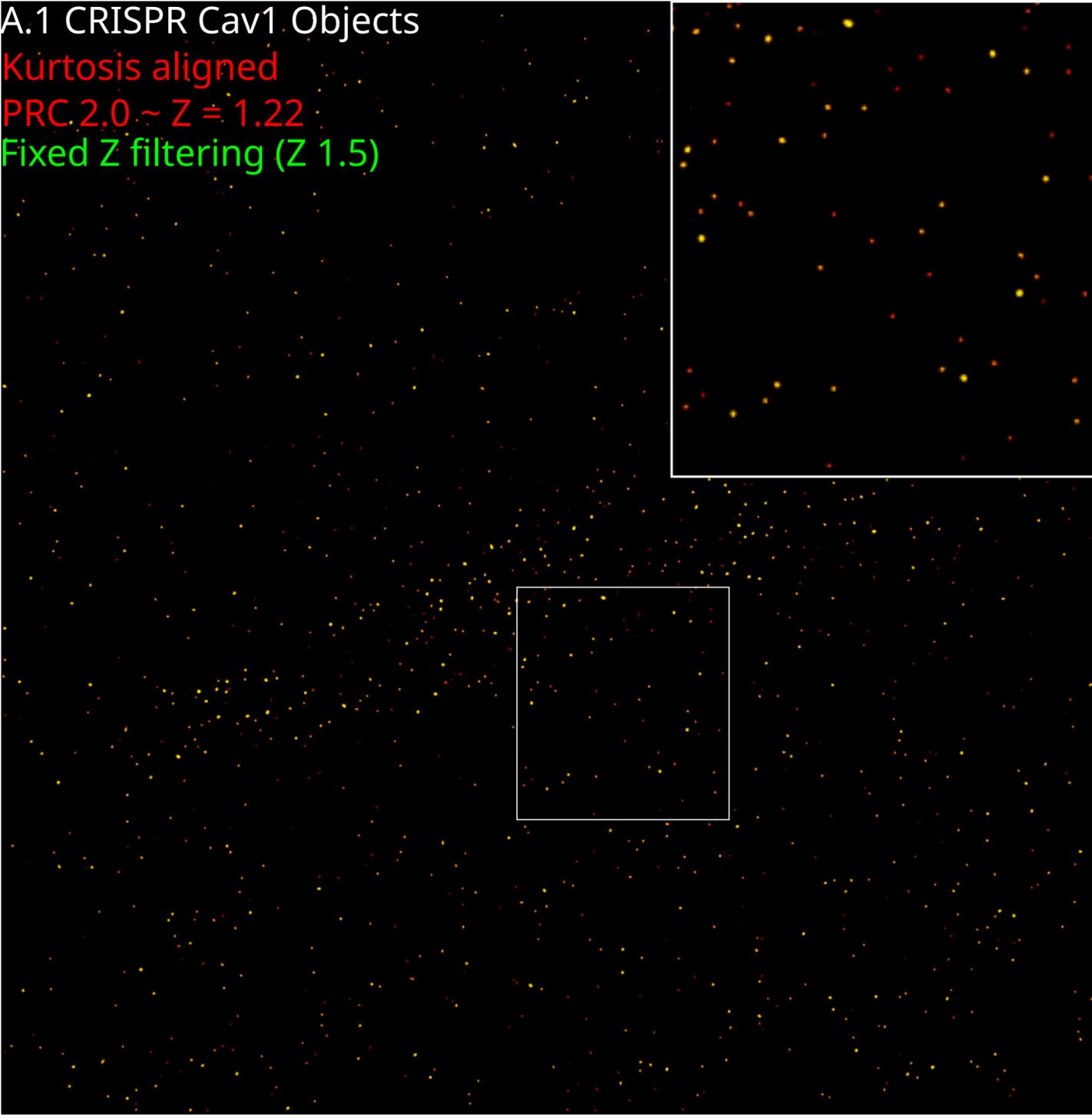


10 μ m
0 1 $Bel(o \rightarrow AD)$



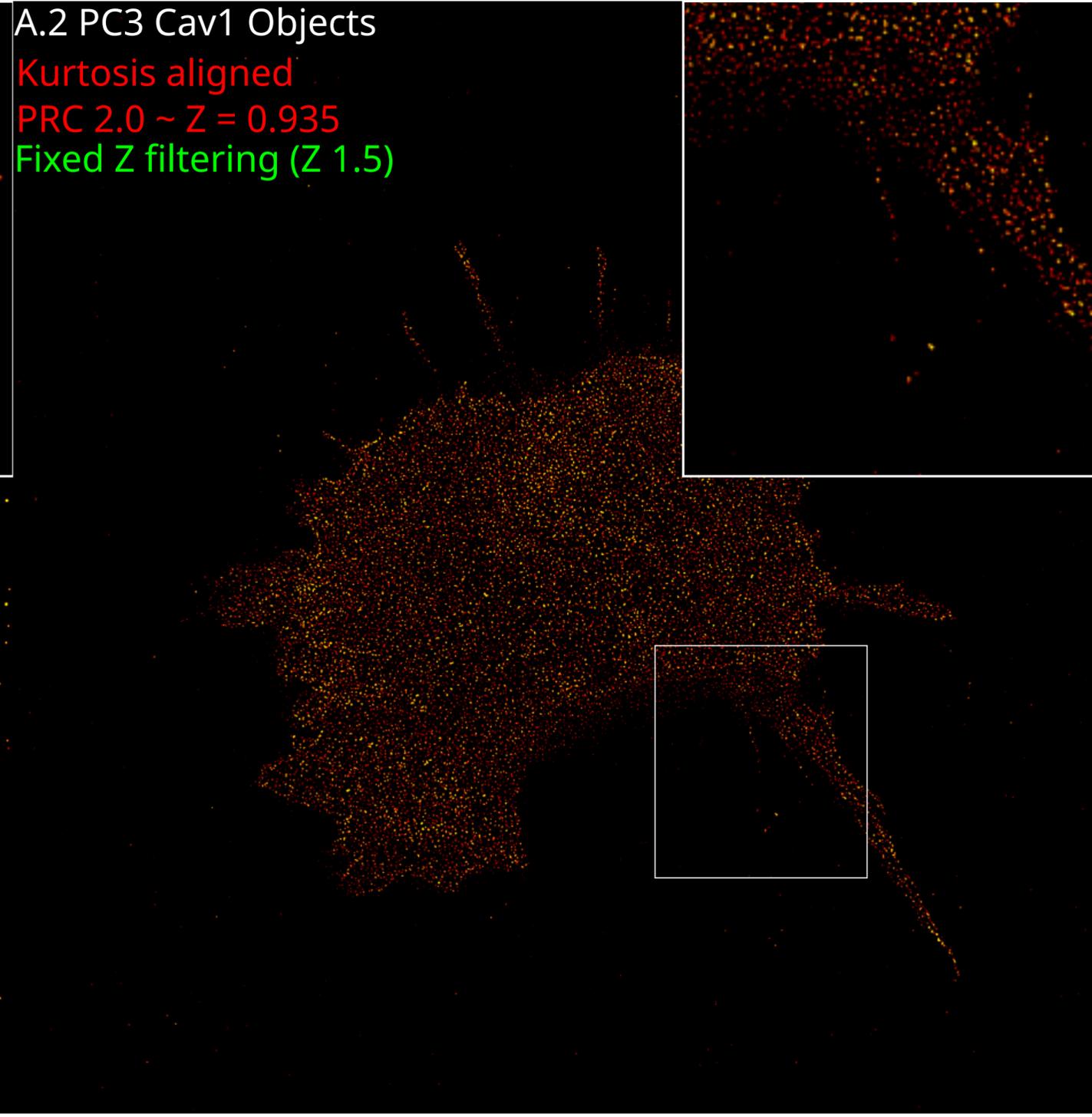
A.1 CRISPR Cav1 Objects

Kurtosis aligned
PRC 2.0 ~ Z = 1.22
Fixed Z filtering (Z 1.5)



A.2 PC3 Cav1 Objects

Kurtosis aligned
PRC 2.0 ~ Z = 0.935
Fixed Z filtering (Z 1.5)



A.3 PC3PTRF Cav1 Objects

Kurtosis aligned
PRC 2.0 ~ Z = 1.246
Fixed Z filtering (Z 1.5)

