

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

Ben Cardoen^{1*}, Timothy Wong², Parsa Alan², Sieun Lee^{3,4}, Joanne Aiko Matsubara³, Ivan Robert Nabi^{2‡*}, Ghassan Hamarneh^{1‡*},

1 Medical Image Analysis Laboratory, School of Computing Science, Simon Fraser University, Burnaby, British Columbia, Canada

2 Department of Cellular and Physiological Sciences, Life Sciences Institute, University of British Columbia, British Columbia, Canada

3 Department of Ophthalmology and Visual Sciences, Eye Care Centre, University of British Columbia, Vancouver, British Columbia, Canada

4 Mental Health & Clinical Neurosciences, School of Medicine, University of Nottingham, Nottingham, United Kingdom

‡ These senior authors contributed equally to this work.

* Correspondence to be directed to: bcardoen@sfu.ca, irnabi@mail.ubc.ca, hamarneh@sfu.ca

Abstract

Identification of objects in fluorescence microscopy is a non-trivial task burdened by parameter-sensitive algorithms, for which there is a clear need for an approach that adapts dynamically to changing imaging conditions. Here, we introduce an adaptive object detection method that, given a microscopy image and an image level label, uses kurtosis-based matching of the distribution of the image differential to express operator intent in terms of recall or precision. 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, capturing those aspects of the image that support the label, and to what extent. We validate our method on 2 datasets : distinguishing sub-diffraction limit caveolae and scaffold by stimulated emission depletion (STED) super-resolution microscopy; and detecting amyloid- β deposits in confocal microscopy retinal cross-sections of neuropathologically confirmed Alzheimer's disease donor tissue. Our results are consistent with biological ground truth and with previous subcellular object classification results, and add insight into more nuanced class 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 diffraction-limited confocal imaging of tissue sections and super-resolution microscopy of subcellular structures, we demonstrate multi-scale applicability.

Author summary

We introduce a novel method that is able to localize fluorescent labelled objects in multi-scale 2D microscopy, and is robust to highly variable imaging conditions.

Localized objects are then classified in a novel way using belief theory, requiring only the image level label. Each object is assigned a ‘belief’ that describes how likely it is to appear in an image with a given set of labels. We apply our method successfully to identify amyloid- β deposits, associated with Alzheimer’s disease, and to discover caveolae and their modular components in superresolution microscopy. We illustrate how our approach allows the fusion or combination of models learned across markedly different datasets. We show how we can compute the ‘conflict’, or disagreement between the models, an insight that allows the domain expert to interpret the composite model.

Introduction

Fluorescence microscopy is a robust experimental tool for the study of biological samples. Applications range from micrometer scale labelling of tissues to super-resolution nanometer scale analysis of molecular components of cells [1]. These analyses typically encompass multiple samples, diverse fluorescent markers (often imaged in parallel in different fluorescent channels), and various microscopes and operators. With experiments spanning multiple channels, datasets, operators, and microscopes, there is a clear need for image analysis methods that adapt dynamically to changing imaging conditions.

Given a set of microscopy images $\{I\}$ from different groups (label L : e.g. normal, control, treated, infected, mutation, wildtype): we identify the parts of the image that show evidence for L , i.e. weakly supervised object detection (WSOD), and quantify the confidence in the identification.

Unlike the classical WSOD problem statement where an object in nature is either, say, a cat or horse, objects identified using SPECHT (Self-tuning Plausibility Based Object Detection Enables Quantification of Conflict in Heterogeneous Multi-scale Microscopy) can assign support for multiple labels to a single object. To overcome per image variation in acquisition, we introduce an adaptive approach that detects objects by thresholding the Laplacian of each image, using its kurtosis to ensure the threshold scales with the distribution of each specific image. Using belief theory we then assign to each identified object the support or numerical evidence it has for a set of labels, and introduce how belief calculus can offer the user interpretable information on the conflict and agreement of composite models learned on heterogeneous datasets. Here, we apply SPECHT to two use cases of distinct scale, identification of intermediate stages in the construction of more complex subcellular structures using super-resolution microscopy and detection of a gradual pathology from fluorescent confocal images of tissue sections.

First, caveolae, subcellular structures composed of CAV1 protein complexes, are ~ 100 nm invaginations in the cell membrane with a varied spectrum of functions [2]. CAV1 forms non-caveolar scaffolds (SC), including 8S oligomers that combine to form larger non-caveolar hemispherical scaffolds as well as caveolae [1, 3–6]. In the absence of CAVIN1, CAV1 forms non-caveolar scaffold structures that contain few CAV1 molecules and those can only be distinguished from caveolae by super-resolution microscopy [1]. Caveolae flattening functions as a membrane buffer, protecting cells from membrane breakage in response to mechanical stress, and scaffolds have been shown to be pro-metastatic in prostate cancer [7, 8]. In fixed cells, superresolution network analysis identifies individual scaffolds into separate subtypes whose modular similarity suggests that smaller scaffolds combine to form larger scaffolds and caveolae [1, 5]. This represents an example of hierarchical modelling in cell biology in which a larger modular protein structure is composed of smaller subunits, with both having distinct functions and properties, e.g. the smaller structure(s) can exist as independent, functional units [5]. We show that application of SPECHT to an alternate super-resolution microscopy approach, STED [9], amenable to high speed live cell imaging [10], is

capable of distinguishing these sub-diffraction limited sub-cellular structures.

Second, identification and quantification of amyloid- β (A β) deposits in the retina in relation to Alzheimer disease (AD) is an open research problem [11]. In previous studies using confocal microscopy on post-mortem donor tissues, retinal A β quantification was performed manually by blinded raters [12] or semi-automatically with manual segmentation [13]. The resulting measurement of retinal A β 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 β and pose a challenge to achieving statistical significance.

Finally, when we present the detected and identified objects to the domain expert, be they cell biologist or medical practitioner, we want to be able to report how ‘sure’ our method is in its computation. With biomedical data typically scarce and acquired with differently configured devices, we would like to have a method that can work across these datasets, to maximize reuse and reproducibility. If models computed on different dataset disagree in the identification of objects, we want to report a mathematically grounded quantification of this disagreement or conflict.

Problem Statement We aim to model a function D that identifies subsets of pixels (objects o) of the image and a function S_L that assigns to each object the statistical support, or evidence e , for the image label L . Statistical support is the numerically quantified evidence distilled by a statistical (learning) method from feature descriptors and their distribution.

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

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

The notation $f: A \mapsto B$ indicates that the function f has domain A and range B . 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 ($e \in [0, 1]$) support value to the proposition in the context of belief theory [14], a generalization of probability functions. In other words, if an object o has high support for image label L , it can be considered likely or typical to appear in images with label L . 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} . Labels can be hierarchically organized, for example where a the label disease further is specialized into specific conditions, or sub-labels. An advantage of using belief theory to model this problem is that it does not require each proposition to be a singleton. Belief theory allows us to specify only the support we can compute. If we do not have measurable evidence to compute the support an object has for a specific sub-label, then we do not need to assign it an arbitrary evidence score, as long as we can assign some evidence to its superset. This enables us to extend our model as more information becomes available, and lends itself to hierarchical, modular, and nested labels. Given that our use cases have a nested label space, this is a key advantage, yet with belief theory a superset encompassing probability functions we do not lose functionality by selecting this framework.

Related work 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, (iii) multiple instance learning (MIL), and (iv) weakly supervised object detection and localization (WSOD/L). However, each presents

deficiencies with respect to application of the method across data sets as well as to object classification.

Joint or hybrid segmentation and classification In joint or hybrid segmentation and classification learning, the tasks of image classification and segmentation at the same time higher accuracy can be obtained in the classification task, while class specific priors can then be leveraged to improve segmentation [15]. Similar approaches have been applied to chromosome microscopy [16], breast biopsy [17], fundus images [18], and histopathology [19] to name a few. However, typically they require object level annotations, which we do not have. In contrast with their object level classification, we aim to capture a gradual transition between classes, e.g. discrete versus continuous object level annotations. In addition, they do not always provide a theoretical upper bound to the support assigned to each segment, meaning that in principle the label assigned to the object can take on extreme values not warranted by the data. It is unclear how to apply the same method across heterogeneous datasets, or to quantify conflict between models learned from such datasets.

Explainable AI In explainable AI 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 [20,21]. 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. In addition there are no non-trivial (0,1) bounds on the support that each object is assigned, potentially leading to high uncertainty. Filtering the attention maps [22] 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. Recent work adds the computation of uncertainty to the ‘importance’ of features in interpretable AI [23], however, the output for image features is still region based, rather than object based.

Multiple Instance Learning 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^+ . Alternatively, the MIL formulation can be adapted to learn the distribution of labels over a bag [24]. MIL has been adopted successfully for microscopy-specific tasks such as classifying and segmenting cells [25,26] with recent reviews [27,28] 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.

Weakly Supervised Object Detection and Localization A complete review of WSOD/L methods has been presented recently [29], and is indicative by the sharp rise in deep learning based WSOD/L approaches. However, subtle yet critical differences are present between our problem statement and the problem statement addressed in WSOD/L methods. Uncertainty, in our context, is interpreted as the margin of error in assigning a certain support to an object. In WSOD/L the uncertainty refers to the noise or variability induced by the human made image level label annotation [29], domain shift, image noise, variability in object appearance, or imprecise localization annotations (e.g. size of enclosing bounding box). In our use case we have no ground truth localizations or bounding boxes to compare against. Deep learning based WSOD/L methods can be sensitive to small dataset sizes with long tail

distributions, which, unfortunately, matches our datasets and is common in biomedical data for scientific discovery. We do have the advantage of leveraging acquisition specific information to inform the object detection stage. WSOD/L methods do, to the best of our knowledge, not quantify or report the evidence based on object features in the end results, nor do they feature joint models across heterogeneous datasets, which we require. Interestingly, it has been suggested that such refining of evidence is a key attribute of biological vision systems [30], and likely to drive adoption towards reinforcement learning based methods for WSOD/L.

Proposed Contribution We introduce here SPECHT for object detection and evidence-based object labelling. SPECHT involves two stages:

- Adaptive and self-tuning object detection using the kurtosis of the Laplacian to match distributions across channels for fluorescence microscopy.
- Belief theory-based labelling to quantify the non-crisp evidence each identified object has for a set of image-level labels.

Use of kurtosis enables estimation of algorithm-specific parameters consistently across heterogeneous data in the absence of object-level annotation, providing a novel, self-tuning, and robust framework for analyzing images. The class of problems we address here is identification of fluorescently labeled structures from background and fuzzy classification of these structures from each other. The algorithm is illustrated in Fig. 1-A. We use SPECHT to, first, identify and distinguish sub-diffraction limit (<100 nm diameter) caveolin-1 (CAV1) domains using STED [9] super-resolution microscopy. Next, we show that SPECHT can automatically distinguish amyloid-deposits characteristic of Alzheimer disease in retinal scans. Distinctive retinal amyloid deposits are associated with Alzheimer disease, however, their identification requires expert analysis [11]. These two use cases demonstrate the ability of SPECHT to provide adaptive object detection and classification to multi-scale fluorescent microscopy data sets. The kurtosis scaling and belief theory based object identification are not restricted to microscopy use cases, or to the features we use. To the best of our knowledge we are the first to employ belief theory based object identification for WSOD/L, enabling the usage of belief calculus for more general applications, as well as leveraging it to span heterogeneous datasets. By separating the object detection stage from the identification and labelling, we ensure that our approach will detect objects that have no or weak discriminative information for an image level label. This capability is important because it enables a more comprehensive quantitative analysis of images by recording both rare, common, unique, and distinguishing objects in a variety of images. Learning to distinguish common objects in sets of images allows our models to learn to identify those in new data where exactly those objects are altered by disease, dysfunction, or genomic modification. Without decoupling the object detection stage, we would run the risk of overfitting a learned model only on discriminative information in the current set of images. The unbiased identification also enables frequency analysis, where we measure the change in frequency of ‘discriminative’ objects with respect to ‘common’, rather than being bound to counting of discriminative objects alone.

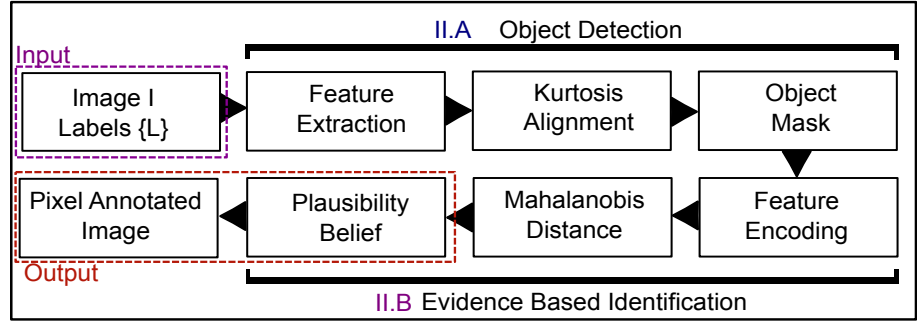


Fig 1. SPECHT algorithm. Adaptive identification of subcellular structures in superresolution microscopy in tandem with belief based labelling of each object's support for the cell level genotype. Each image can have a set of labels, SPECHT then computes for each identified object o the support it has per individual label.

Method

In this section, we outline the design of our proposed method. To help the reader unfamiliar with some of the domain specific terms we have 3 glossaries of imaging (S1. Table), belief theory (S2. Table), and biology (S3. Table) related terms to help understanding of the contribution and ensure terms are unambiguously defined.

Adaptive kurtosis aligned object detection

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 operator variance. The image Laplacian ∇^2 , a measure of the second derivative of the image intensity, can be used to detect edges of objects wherever ∇^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.

Non-specific binding can, given its tendency to self-organise [31] in concentrations of fluorescent label, can have a similar intensity profile. 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 generalized normal distribution with probability density function $\frac{\beta}{2\alpha\Gamma\beta-1}e^{-\frac{|x-\mu|}{\alpha}^\beta}$ with scale α , location μ , the gamma function Γ , 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 [32]. 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 imaging point spread function.

Object detection by thresholding Laplacian of Gaussian

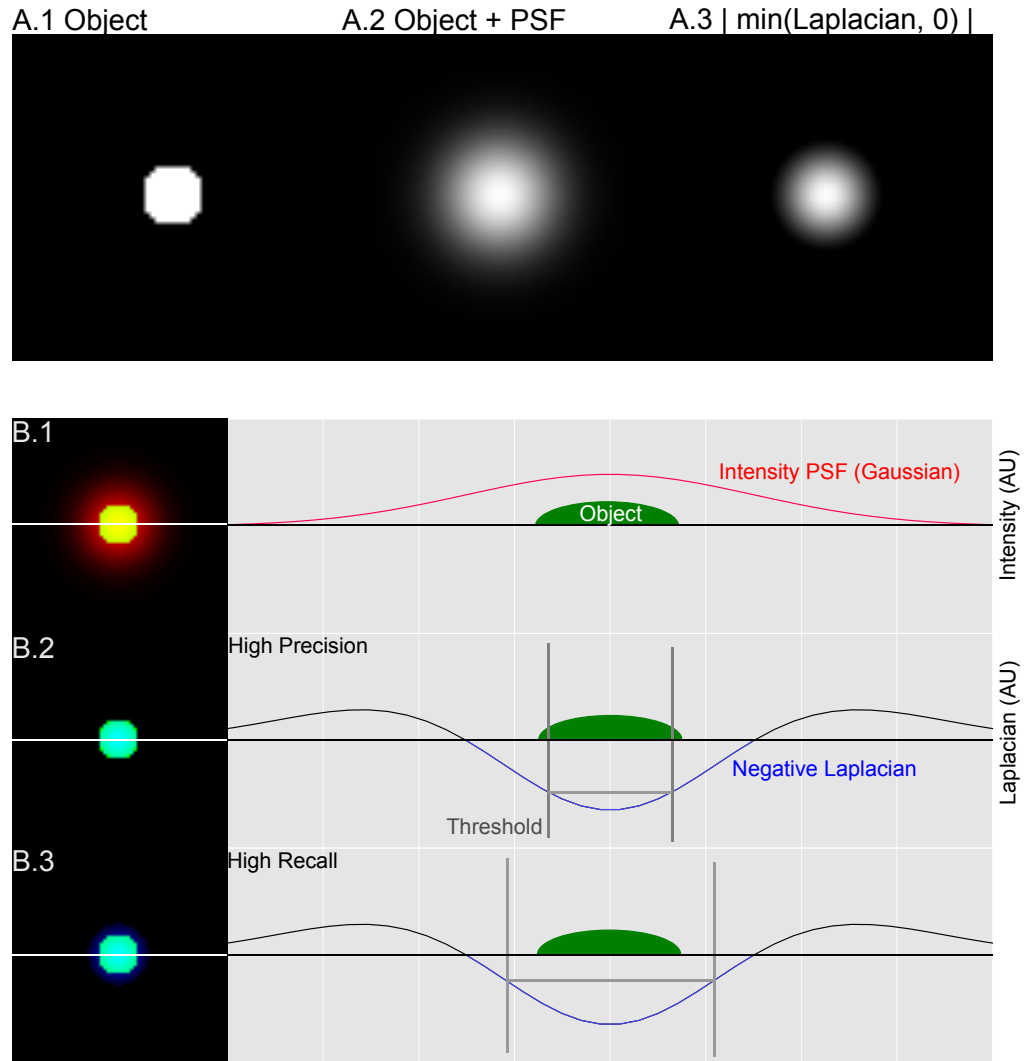


Fig 2. Object detection principle. The negative Laplacian (A.3-V, B.1)) can be leveraged to detect Gaussian 2D observations of 3D fluorescent objects. Thresholding V is a balance between high precision (B.2) and high recall (B.3)

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. 3), in terms of object retrieval, 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 parameters across distributions of V -space (negative Laplacian). 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. 3 red, green, respectively). The kurtosis of a distribution is a scalar value increasing with the ‘tailedness’ of the distribution. We illustrate this by means of in silico data (Fig. 3), where the threshold in kurtosis space scales with the shape of the two different distributions (Fig. 3-C.2-green versus red). An increase in frequency of objects in an image will lead to higher V values, and with them a change in the tailedness of the distribution. Conversely, a decrease will lead to a less tailed distribution, given that most V values are due to the image background. If we can find a thresholding method that adapts to the tailedness of the distribution of V , then we are more likely to obtain consistent across images. We next use these insights to normalize V to $Z_V = \frac{V_i - E(V)}{\sqrt{\text{Var}(V)}} \forall V_i$ and then obtain an estimate $E'_z \sim E[Z]$ (the expected z-score), as a consistent threshold, that can be scaled up or down automatically across images. While we can compute $E[Z] = \int z f(z) dz$, 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. Due to inaccurate estimation we can end up with estimates that for one image underestimate $E[Z]$, yet for another overestimate. We then have results that vary per image in precision and recall, rather than being consistent in their results. If we aim for a lower bound on $E[Z]$ then we maximize consistency. If a preference for precision over recall is preferred, one can weight the lower bound, while retaining consistency across images. We can derive such a strict lower bound by noting that $\text{kurtosis}(V) = E[Z_V^4]$ [33]. 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. We use the ‘excess’ kurtosis ($k-3$) in our implementation. Moreover, by weighting the kurtosis, we can allow the user to alter the threshold in a distribution-aligned space. We scale the outcome by a floating point parameter ‘precision-recall (PRC)’ to fulfill our aim of an intent-based self-tuning and adaptive method. A value $\text{PRC} > 1$ leads to a distribution-aligned object extraction that favours

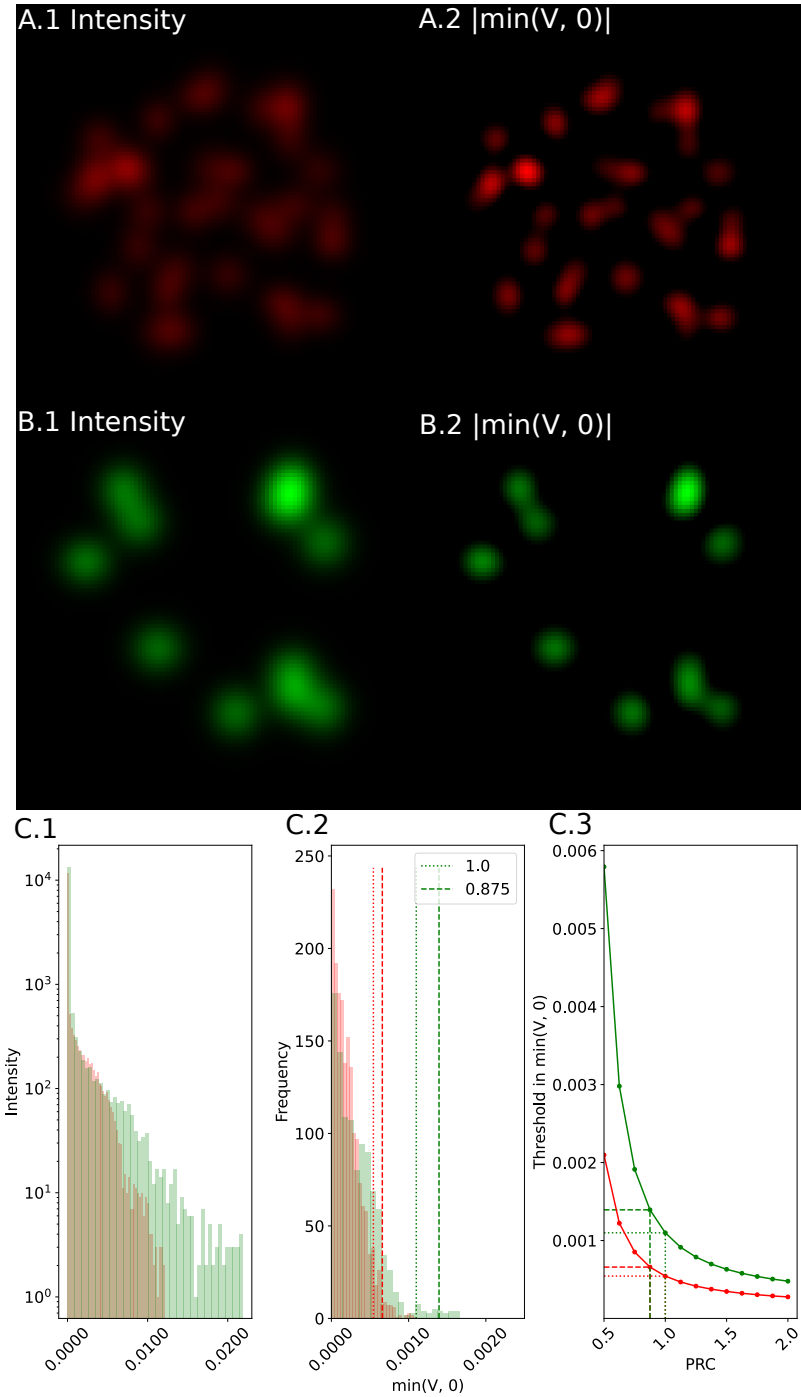


Fig 3. Adaptive thresholding. Kurtosis based thresholding illustrated on two in silico images. A: $N=35$, $\sigma = \frac{10+5 \times \text{rand}()}{3}$, B: $N=12$, $\sigma = \frac{15+5 \times \text{rand}()}{1}$, sources randomly placed, with isotropic PSF. A.2 and B.2 show the negative Laplacian, and illustrate how it is less susceptible to intensity differences. C.1: The intensity distribution of both images. C.2: The distribution of the Laplacian of both images. C.3: The automatically derived threshold based on kurtosis can be scaled in favor of precision (PRC < 1) or recall (PRC > 1). The plot shows how kurtosis space thresholding follows the different shapes of the distributions.

recall, $\text{PRC} \leq 1$ favours precision. Fig. 3 illustrates the scaling effect on in silico distributions. 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 distributions of image differentials.

We note that even though we apply the kurtosis scaling on the negative Laplacian, there are no constraints to extending this approach to multinomial distributions of arbitrary features, as long as those have finite moments. Computing kurtosis is non-parametric, e.g. does not expect a certain family of distribution, hence easily generalizes to new applications.

Finally, we binarize the image where a pixel is 1 if and only if the corresponding negative laplacian exceeds the dynamic threshold. The binarized image is then decomposed by using the connected components algorithm treating the 2D image as a graph. 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  $\mathbf{I}$ , 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:    $k_j \leftarrow \mathbb{E} \left( \frac{V_j - \mu_{V_j}}{\sigma_{V_j}} \right)^4$  ▷ Kurtosis, 4th moment
7:    $z_j \leftarrow \sqrt[4]{k_j}$  ▷ Adaptive consistency across channels
8:    $V_j[V_j \leq \mu_g(V_j) * \sigma_g(V_j)^{\frac{z_j}{\text{PRC}}}] \leftarrow 0$  ▷ Eq. 4
9:    $M_j \leftarrow \text{connectedcomponents}(V_j)$ 
10: end for

```

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 :

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

The triplet (p, q, r) follows the notation of Dempster [34] 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

$$\text{Bel}(\Theta) = 1, \text{Bel}(\emptyset) = 0, \text{ and } \text{Bel} \left(\bigcup_{i=1}^n A_i \right) \geq \sum_{I \subset \{1, \dots, n\} \wedge I \neq \emptyset} (-1)^{|I|+1} \text{Bel}(A_i) \quad \forall A_i \subset \Theta.$$

Evidence can be encoded by a mass function $m(A) \rightarrow [0, 1] | A \subset \Theta$, where 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

$Pl(A) \geq Bel(A) \forall A \subset \Theta$. In the (p, q, r) notation, we have that $p = Bel(A), q = Bel(\bar{A}), r = Pl(A) - Bel(A)$. The reader can find a graphical illustration in Fig. 4. For a more in-depth review of belief theory, we refer the interested reader to Yager et al. [35].

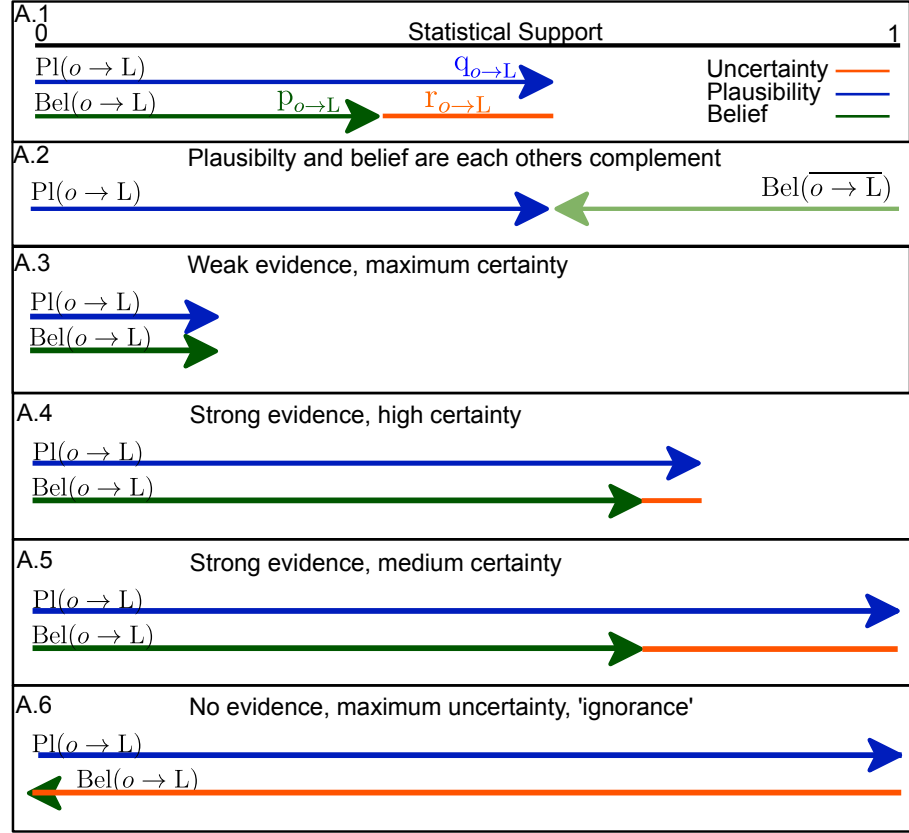


Fig 4. Belief theory. Graphical illustration of the concept of plausibility, belief, and uncertainty in the context of belief theory and as used in the remainder of this manuscript. A.2: Plausibility and belief can be expressed as the complement of their respective support. A trivial, or naive, model has a plausibility of 1, belief 0, and uncertainty 1. A3-6: Illustrates the flexibility of belief theory based modelling. Weak, but certain evidence (A.3) occurs when belief and plausibility are equal, yet small. Conversely strong evidence can be certain (A.5), but does not need to be (A.4). Finally, absence of quantifiable evidence is mapped to ‘ignorance’, maximal uncertainty, where belief is 0, plausibility 1.

Encoding evidence

Given a set of images \mathbf{J} with label L_J , and a set of images \mathbf{I} , we want to identify objects in the images and assign to each object o a tuple (p, q, r) expressing the belief, plausibility and uncertainty of the proposition $o \rightarrow L_J$ for objects in images in \mathbf{I} . We illustrate in the results section that our method can be applied to distinguish objects from a nested hierarchy $L_J \subset L_B \subset L_A$. In Alg. 2, we illustrate the steps we undertake to arrive at a belief based labelling of objects in images. The sets of images \mathbf{J} and \mathbf{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 \mathbf{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.

Inferring plausibility

We want to be able to quantify both relative support and support for an individual label. Since the Mahalanobis distance has a range $[0, \infty)$, we normalize the statistical distance (Alg. 2, line-13) so we can leverage Cantelli's theorem [36]

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

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 [34], we know that $\text{Bel}(\overline{A}) = 1 - \text{Pl}(A)$. For $o \in \mathbf{I}$ we can formulate $p_i = \text{Bel}(o \rightarrow L_J) = 1 - q_j$. When we swap \mathbf{I} and \mathbf{J} , we can obtain q_i and p_j , giving us $r_i = q_i - p_i$ and $r_j = q_j - p_j$. Fig. 7-C illustrates the application of belief theory based labelling on object detection and the interplay between belief and plausibility (Fig. 4-A). The resulting support function has no limiting specific priors or assumptions, is continuous, has a theoretical upper bound, and requires no supervised training data. 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. Fig. 4-A1-6 provides a graphical illustration of the flexibility of the belief theory framework, and can help the reader understand the definitions of ‘uncertainty’, strong versus weak ‘evidence’, and ‘ignorance’ or absence of information.

Algorithm 2 Probabilistic labelling algorithm

```

1: Input Images  $\mathbf{J}$  with label  $L_J$ , Images  $\mathbf{I}$ 
2: Output  $\text{Pl}_I$ , 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:  $F_J \leftarrow \{\text{features}(o_{ji}) \mid o_{ji} \in M_J[j], j \in [1, |\mathbf{J}|]\}$ 
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:     $\text{Pl}[o_k \rightarrow L_J] \leftarrow \frac{1}{1 + Z_j[k]^2}$  ▷ Eq. 6
16:   end for
17: end for

```

Next 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 $\{o \in L_{CAV1KO}\} \subset \{o \in L_{PC3}\} \subset \{o \in L_{PC3-CAVIN1}\}$ where a subset label is more

specific as illustrated in Fig. 7-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. While belief theory based combination has been used for histopathology classification [37], our usage for individual object detection in microscopy is to the best of our knowledge novel.

Using belief calculus to express joint models spanning heterogeneous data

Especially in the case of human tissue data of patients, data is sparse and usually acquired by different institutions, with operators, acquisition, and protocols varying. Using a single sparse dataset degrades statistical power. Here we show how a practitioner can leverage belief calculus to compute an interpretable joint model over such datasets. 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 $\text{Bel}(A) = m(A)$ ($m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \text{Bel}(B)$). We enumerate in Table 1 the intermediate results needed to compute the joint mass function for our use case. Let for

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

Table 1. Dempster combination enabling the expression of a joint model. $A, B, C \subset \Theta$

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:

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

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 of a closed form expression for the joint model, allows us to span heterogeneous data. Using Dempster’s combination rule to combine models has been shown to be a robust method to combine multiple heterogeneous object detector models on natural images (‘Dynamic Belief Fusion’) [39], where it outperformed both Bayesian fusion and weighted sum approaches. In addition, the application was detection of discrete classes of objects in a supervised setting, e.g.

detecting a 'chair' in a natural image. More importantly, the computation and reporting of conflict was not leveraged, as is the application across heterogeneous datasets. These are an important distinctions with respect to applications in microscopy, where object types are fuzzy or continuous, and heterogeneous data adds further complexity to the fusing of models, as well as necessitating the reporting of conflict to the end user.

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.

Results

Next, we evaluate SPECHT on in silico, and real world data. The full description of the real world datasets used in this section is listed in the Appendix (Sec. S5 Text). 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.

Evaluation of object detection on in silico data

Consistency across datasets

In order for the belief labelling stage to function with minimal bias, it is critical that the object detection stage performs consistent, and predictable across datasets. It is indeed possible to design or train a method to perform optimal on a single chosen dataset, but compromise performance to an unknown extent on future datasets. A major source of variance across datasets, in fluorescence microscopy, is the distribution in size and brightness of labelled objects. Variance of fluorescence is not only a common obstacle across datasets, but also appears in multiple channel analysis, given that two fluorescently marked targets are rarely exhibiting the same distribution, even in the same cell. We need this consistency, given that we have high variance both across cells and channels, in our real world datasets. We simulate images of 512×512 pixels, with a Gaussian and Poisson noise model [40] of respectively σ and λ set to 0.062 (in 8-bit grayscale). In each image, k bright and j dim objects are randomly placed, where $k \in [1, 25, 50]$, and $j \in [50, 25, 1]$. Bright objects are modelled with a Gaussian PSF ($\sigma = 3$), whereas dim objects have $\sigma = 6$ and with their intensity reduced by a factor of 4. For our specific use case, where no ground truth is available and little domain knowledge can be exploited, we compare against 2 tried and true approaches: automatic scale space detection [32, 41], and Otsu thresholding [42]. More advanced object detection methods have become available for fluorescence microscopy [43, 44], but these invariably require parameters related to the objects of interest, e.g. scale range. In our use case we do not have that information, and estimating it would risk propagating size-based biased information to the belief based labelling. The scale space algorithm is pre-configured with the range of σ 's of objects to detect, which in our real data is not available. In addition, the output of the scale space detection is filtered by an Otsu-filtering stage to remove false positives. SPECHT's PRC is set to 2, with σ set to the PSF σ . Figure 5 illustrates the results. Objects are considered correctly reconstructed when the detected object overlaps with the ground truth (green). False positives are marked in red. False negatives in blue. Under these varying conditions, SPECHT is not always optimal (middle row, Otsu), but is very consistent in object retrieval. In contrast, the two reference methods can be optimal for a given distribution, but vary markedly. If domain specific information is available, or ground truth data,

more targeted object detection stages can be used in place of our adaptive method. However, reliance on a consistent, adaptive object detection stage provides the capability to obtain similar results on future unseen datasets, without having to worry about potential parameter sensitive bias.

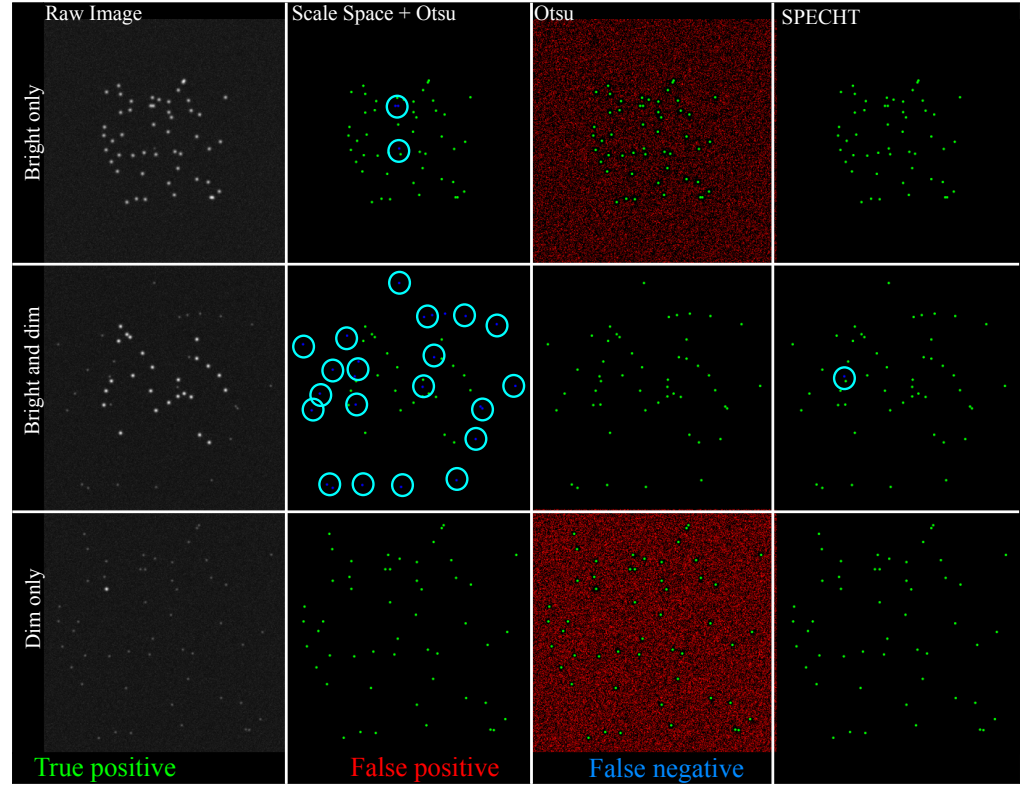


Fig 5. Consistency compared to existing methods. We simulate 3 markedly different in silico scenarios where light sources are either dominated by bright, dim, or are a mixture of both. Note that SPECHT is not always optimal, but does produce consistent results across these variable conditions.

Robustness to noise

Noise from different sources is unavoidable in fluorescence microscopy data. The Laplacian operator is sensitive to noise. The classical sequence of Gaussian-Laplacian-Gaussian mitigates noise amplification by smoothing. However, at severe noise levels the smoothing step itself can introduce artifacts, that then lead to false positives or skew the Laplacian operator output. By pruning SPECHT's output with a local maxima heuristic, we can mitigate introduction of false positives. Recovery of signal that is below background noise is infeasible. To measure the effectiveness of our algorithm under increasingly noisy conditions, we simulated a mixture of bright and dim light sources (Gaussian PSF), then added both Gaussian and Poisson noise. In Figure 6 we see that at severe levels of noise ($\sigma = \lambda = 96/255$, or 0.37 in 8 bit grayscale images), SPECHT starts to introduce false negatives and omit faint light sources. However, note that even at intermediate values ($\sigma = \lambda = 64/255$), recovery of faint objects is not compromised. Robust object detect is highly relevant to our application, given that fluorescence labelling can vary across targets, channels, and datasets. In ideal

settings, one would preprocess data with specifically developed denoising algorithms, but it is nonetheless important to measure SPECHT’s sensitivity to low signal-to-noise ratios (SNR).

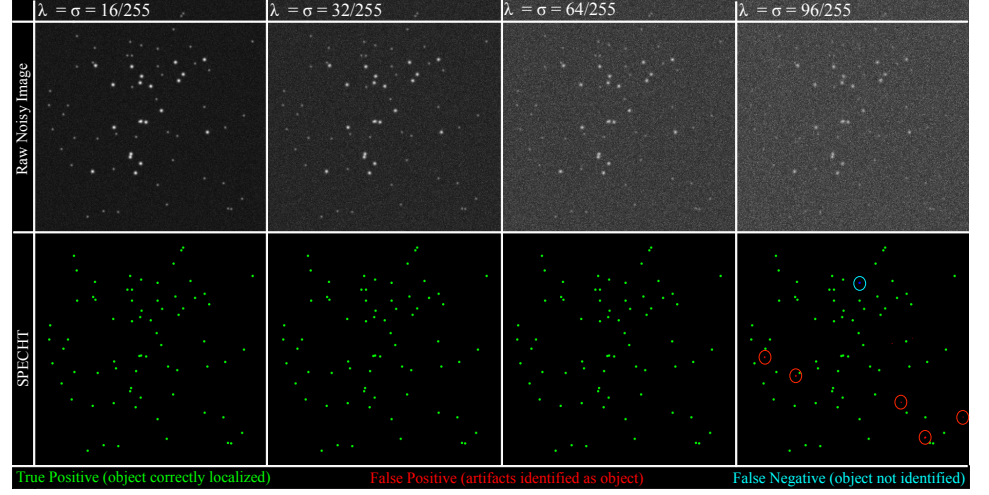


Fig 6. Sensitivity to noise. SNR decreases rapidly as the parameters of both noise sources (Gaussian, Poisson) are increased, yet SPECHT’s recovery of faint objects remains stable under moderate noise conditions. At severe noise levels, as is expected, artifacts appear, while sources with intensity lower than background intensity can no longer be recovered.

Capturing the gradual construction of complex protein structures in STED

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. 7-C), scaffolds (SC), or caveolae (C). 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 [31]. 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 CAVIN1, and PC3-CAVIN1 with CAVIN1 and CAV1 enabled [45]. In CAV1 KO we can only observe BG, in PC3 only SC and BG, in PC3-CAVIN1 the SC, BG and C are present (Fig. 7-B). Our label space \mathcal{L} is then $\{BG, SC, C\}$, with subsets $PC3 = \{BG, C\}$ and $PC3-CAVIN1 = PC3 \cup \{C\}$.

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 $p_{\bar{x}} = \text{Bel}(\overline{o \rightarrow L_x}) = 1 - q_x$, where x is BG, PC3 respectively. Next, we process superresolution images of fluorescence labelled CAV1 deposits in PC3-CAVIN1 (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

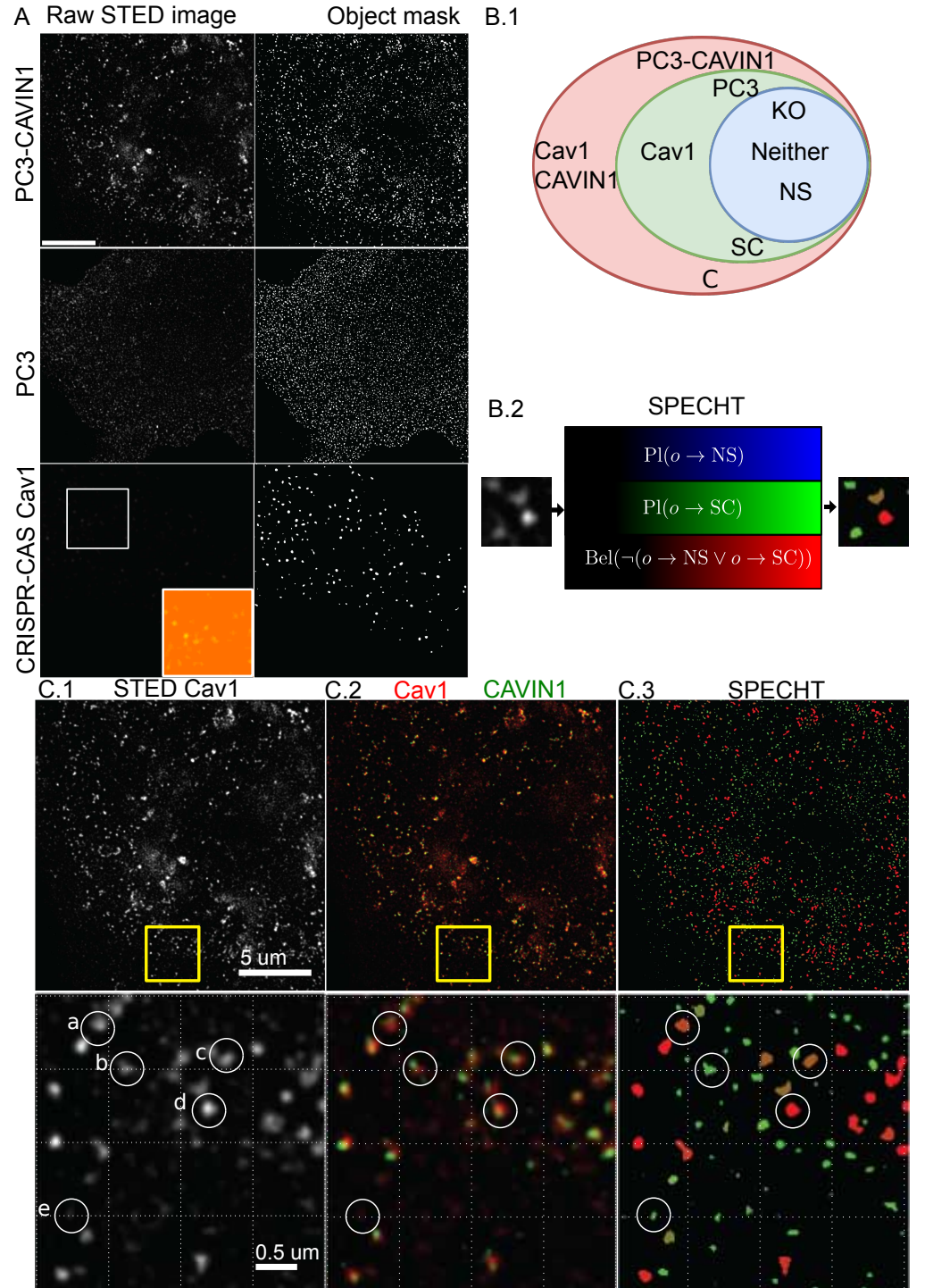


Fig 7. Visualization of results on CAV1 datasets. A: Object detection results on the 3 cell lines with a markedly different intensity profile. B.1: A Venn diagram illustrating how we differentiate between different genotypes. B.2 SPECHT labelling function assigning each object 3 values representing the belief that the object is evidence for either of the 3 object types. C: Illustration of the results on a PC3-CAVIN1 cell.

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-CAVIN1 cells are (formations of) caveolae (C), therefore $p_C = \text{Bel}(o_{P3} \rightarrow L_C) = \text{Bel}(\overline{o_{P3}} \rightarrow \overline{L_{BG}}) \wedge \text{Bel}(\overline{o_{P3}} \rightarrow \overline{L_{SC}}) = p_{\overline{BG}} * p_{\overline{SC}}$. We visualise the results for a single PC3-CAVIN1 cell in Fig. 7-C where blue, green, and red gradients correspond with q_{BG} , q_{SC} and p_{P3} , respectively. From visual inspection, we see correlation of colocalized CAVIN1 with objects labelled with a high p_{P3} value, as expected (Fig. 7-C.3.a, d). More interestingly, we can now identify objects that are transitioning between SC and C (Fig. 7-C.3.c). To confirm this, we next perform extensive validation.

Validation

Given that there is no object-level ground truth available, a direct evaluation is impossible, nor is there to the best of our knowledge a method that discriminates between caveolae and scaffolds in 2D STED. Therefore, the only feasible validation is using previous work on caveolae detection in dSTORM, and colocalization of CAVIN1, essential for formation of caveolae, 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-CAVIN1 cell line has been reported at $\sim 20\%$ [5], when compared to other CAV1 structures. SPECHT computes a belief (p_C) for each detected object that it likely caveolae. In Fig. 8-A, we show the cumulative distribution function (cdf) of that belief function. We observe a tri-modal distribution, as expected for each of the 3 labels (C, SC, BG). The label distribution shows a long left tail, corresponding with 20% of the data, demarcated at the sudden rise of the cdf ($p_C \sim 0.32$), matching a transition into the 2nd mode of the trimodal distribution. In other words, if we threshold the belief label at 0.32, the point where the belief function splits into major and minor part, we find the exact same frequency of caveolae-like objects as previously reported. Second, we know that caveolae can only form in the presence of CAVIN1. Therefore, we expect to see an increasing correlation of CAVIN1-CAV1 colocalization as p_{P3} increases. We compute CAVIN1 colocalization P by measuring the mean CAVIN1 colocalization 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. 8-B, replicate is a repeat experiment to ensure consistency). CAVIN1 colocalization increases markedly when p_{P3} increases. In Fig. 8-B a LOWESS-regression [46] is computed to discover a more nuanced behavior in the correlation with CAVIN1 association. All cells show a consistent pattern across replicates. More importantly, the belief value where the colocalization of CAVIN1 suddenly increases, matches the threshold found when comparing to previous work (0.32), confirming the belief function is consistent with biological ground truth and prior work. The SPECHT color legend is overlayed for ease of interpretation. In conclusion, SPECHT's label indicating an object is likely to be caveolae is consistent with the expected frequency of caveolae in PC3-CAVIN1, and the colocalization of CAVIN1.

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 user's discretion; it is nonetheless important to document what its exact impact on the result can be. In Fig. 8-B.2 we

compute the results for replicate 1, Cell 5, the median of the trend (Fig. 8-B1.1). 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 CAVIN1 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 the actual caveolae structure the expected CAVIN1 association is not that of caveolae but of background, reducing the mean CAVIN1 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.

Identifying retinal amyloid- β deposits associated with Alzheimer disease

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.

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 AD+ image in Fig. 9-B.1. 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 Fig. 9 we illustrate the idea and visual results as well as the quantification of conflict that is offered to the end user. 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 Fig. 9-C 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.

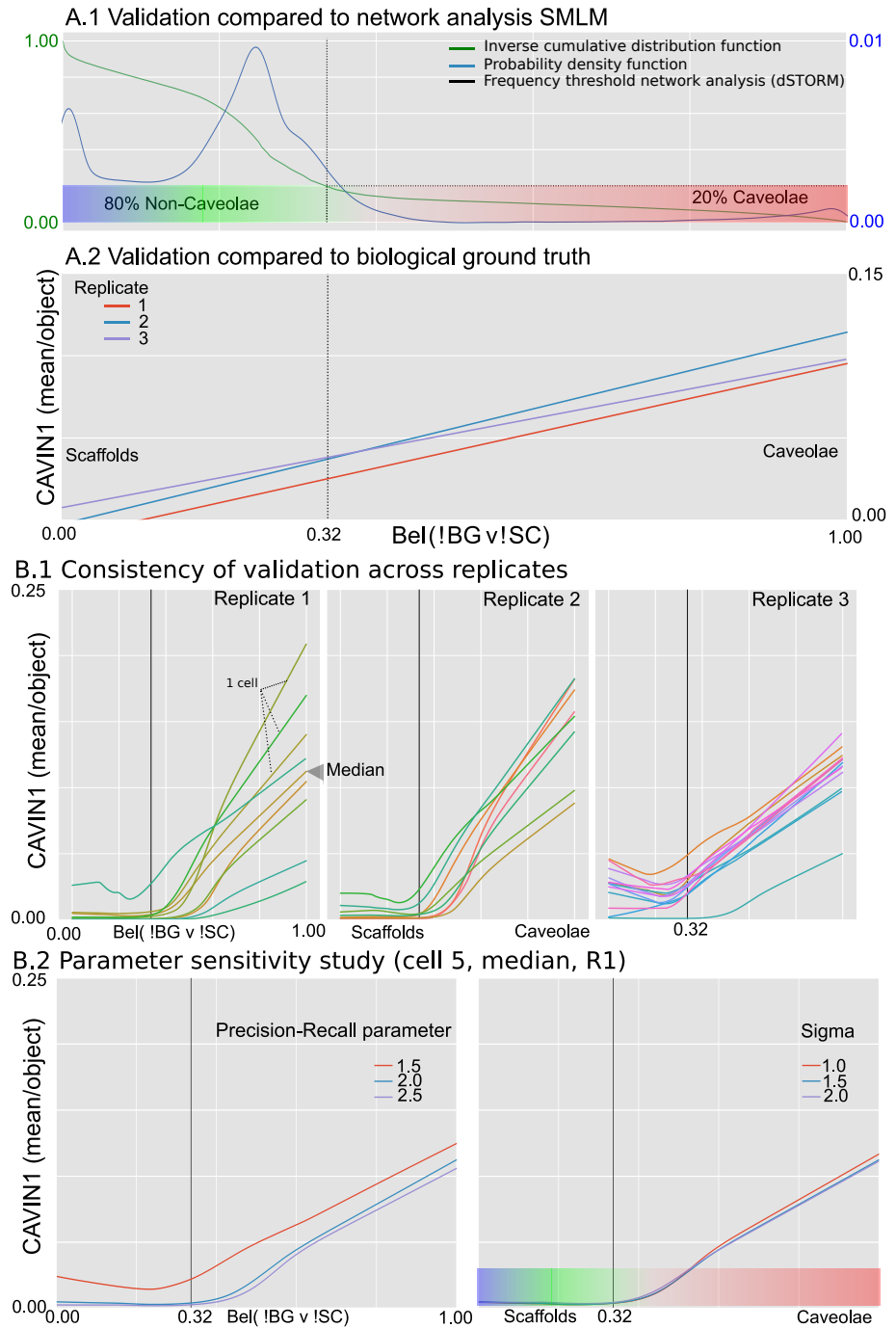
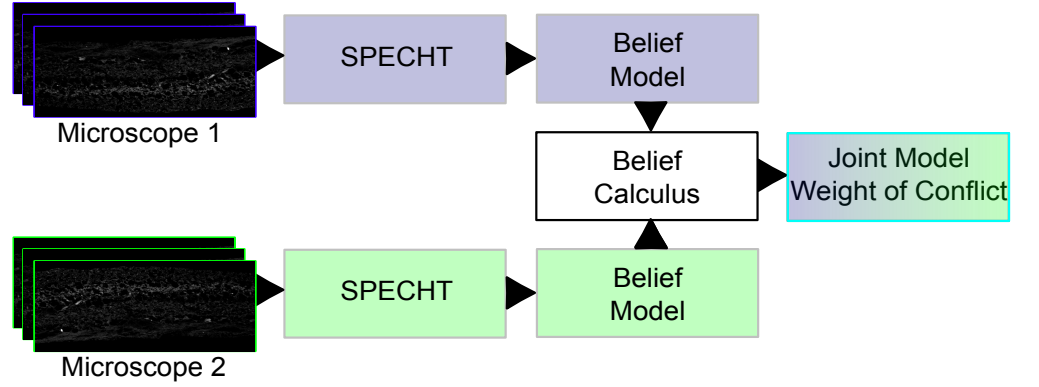
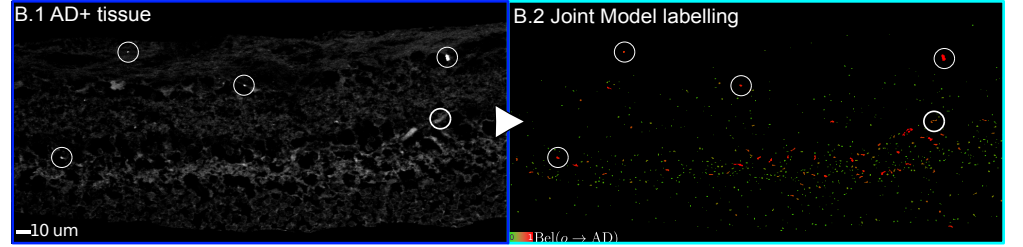


Fig 8. Results on CAV1 dataset. A: Validation with respect to previous art (A.1) and biological ground truth (A.2). A: The distribution of SPECHT's label (A.1, x-axis: $P[\text{object}] = \text{caveolae}$) shows a distinct long left tail, containing 20% of the data. The frequency division matches previous art in dSTORM analysis. Caveolae only form in the presence of CAVIN1, therefore the probability of an object being Caveolae should correlate with the colocalization of CAVIN1 (A.2), which is what we observe. B.1: The detection threshold (~ 0.35 , A.1-2) matches the sudden rise in colocalization when we use a LOWESS regression, rather than a linear regression, and results are consistent across 3 replications (30 cells total, each line represents a single cell). B.2: Varying hyperparameters does not alter the consistency of the result with respect to biological ground truth (colocalization CAVIN1).

A. Belief Calculus Enables Joint Model



B. Illustration of Joint Model



C. Weight of Conflict of Joint Model

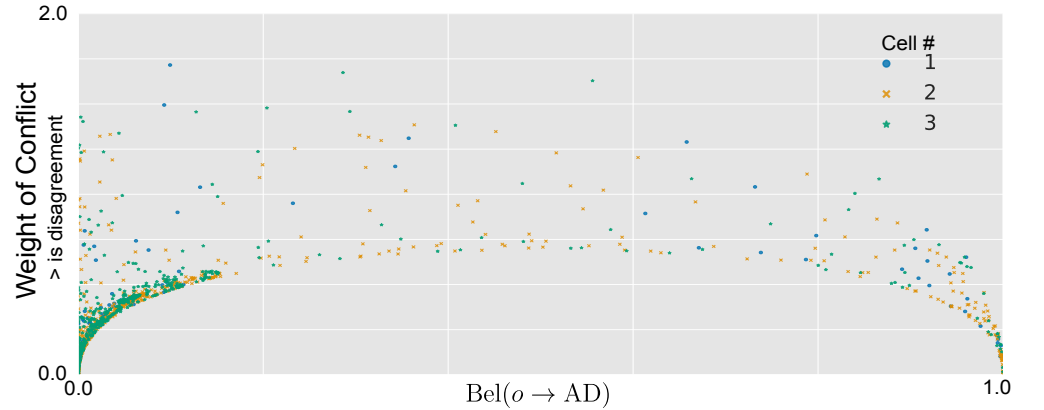


Fig 9. Results on Alzheimer data. A: Belief calculus enables the combination of models learned data originating from different microscopes. B: We visualize how the joint model operates on a single image of retinal tissue stained for amyloid- β , sourced from an AD+ positive patient. Object marked in red express a high belief in being AD+ specific. C: We offer the end user a per-object expression of the conflict between the 2 models that create the joint model. An increased weight of conflict (Y-axis) indicates the models disagree on the labelling for a specific object. We illustrate the visualization here for 3 AD+ images. Observe that for objects where both models are uncertain ($\sim .5$) their minimal conflict is higher than it is for objects that have a higher support for being either AD+ specific or healthy.

Discussion

The motivation for this work was the need for a robust, adaptive, and self-tuning unsupervised probabilistic object detection method applicable to heterogeneous multi-scale superresolution 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. 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 [14]. As a result, our p and q functions are equivalent to ‘necessity’ and ‘possibility’ functions from possibility theory [47]. 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 S6 Text. The belief theory framework allows us to work with nested or hierarchical labels, as well as leverage the mathematically sound concepts of conflict and joint models. The object detection stage is designed to be robust to long tail distributions and high variations in density.

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 [48]. In addition, when image quality degrades to low signal to noise values (SNR), the intensity distribution can cause negative adjusted kurtosis values. In this case raising such a negative value to a fractional power is a domain error. A deconvolution or task specific denoising is recommended to recover or improve SNR before analysis, and will typically be part of an image processing pipeline. 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 [49]. While the joining of belief functions by Dempster’s rule is not without criticism [50], we note that the preconditions [50] 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 [51] to enable quantification of reliability of individual sources and make the joint model robust against unreliable sources.

Conclusion

We introduced a novel adaptive self-tuning method for object detection in 2D microscopy images of 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.

578
579
580

References

1. Khater IM, Meng F, Wong TH, Nabi IR, Hamarneh G. Super resolution network analysis defines the molecular architecture of Caveolae and Caveolin-1 scaffolds. *Scientific Reports*. 2018;8(1):1–15.
2. Parton RG, Del Pozo MA. Caveolae as plasma membrane sensors, protectors and organizers. *Nature reviews Molecular cell biology*. 2013;14(2):98–112.
3. Monier S, Parton RG, Vogel F, Behlke J, Henske A, Kurzchalia TV. VIP21-caveolin, a membrane protein constituent of the caveolar coat, oligomerizes in vivo and in vitro. *Molecular Biology of the Cell*. 1995;6(7):911–927.
4. Lajoie P, Goetz JG, Dennis JW, Nabi IR. Lattices, rafts, and scaffolds: domain regulation of receptor signaling at the plasma membrane. *Journal of Cell Biology*. 2009;185(3):381–385.
5. Khater IM, Liu Q, Chou KC, Hamarneh G, Nabi IR. Super-resolution modularity analysis shows polyhedral Caveolin-1 oligomers combine to form scaffolds and Caveolae. *Scientific Reports*. 2019;9(1):1–10.
6. Stoeber Mea. 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*. 2016;113(50):E8069–E8078.
7. Sinha B, Köster D, Ruez R, Gonnord P, Bastiani M, Abankwa D, et al. Cells respond to mechanical stress by rapid disassembly of caveolae. *Cell*. 2011;144(3):402–413.
8. Moon H, Lee CS, Inder KL, Sharma S, Choi E, Black DM, et al. PTRF/cavin-1 neutralizes non-caveolar caveolin-1 microdomains in prostate cancer. *Oncogene*. 2014;33(27):3561–3570.
9. Hell SW, Wichmann J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Letters*. 1994;19(11):780–782.
10. Gao G, Zhu C, Liu E, Nabi IR. Reticulon and CLIMP-63 regulate nanodomain organization of peripheral ER tubules. *PLoS Biology*. 2019;17(8):e3000355.
11. Hart NJ, Koronyo Y, Black KL, Koronyo-Hamaoui M. Ocular indicators of Alzheimer's: exploring disease in the retina. *Acta Neuropathologica*. 2016;132(6):767–787.
12. Perez SE, Lumayag S, Kovacs B, Mufson EJ, Xu S. β -amyloid deposition and functional impairment in the retina of the APP^{swe}/PS1 Δ E9 transgenic mouse model of Alzheimer's disease. *Investigative Ophthalmology & Visual Science*. 2009;50(2):793–800.
13. Lee S, Jiang K, McIlmoyle B, To E, Xu QA, Hirsch-Reinshagen V, et al. Amyloid beta immunoreactivity in the retinal ganglion cell layer of the Alzheimer's eye. *Frontiers in neuroscience*. 2020;14:758.

14. Shafer G. A mathematical theory of evidence. vol. 42. Princeton university press; 1976.
15. BenTaieb A, Kawahara J, Hamarneh G. Multi-loss convolutional networks for gland analysis in microscopy. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). IEEE; 2016. p. 642–645.
16. Schwartzkopf WC, Bovik AC, Evans BL. Maximum-likelihood techniques for joint segmentation-classification of multispectral chromosome images. *IEEE Transactions on Medical Imaging*. 2005;24(12):1593–1610.
17. Mehta S, Mercan E, Bartlett J, Weaver D, Elmore JG, Shapiro L. 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. p. 893–901.
18. Girard F, Kavalec C, Cheriet F. Joint segmentation and classification of retinal arteries/veins from fundus images. *Artificial Intelligence in Medicine*. 2019;94:96–109.
19. Qu H, Riedlinger G, Wu P, Huang Q, Yi J, De S, et al. Joint Segmentation and Fine-Grained Classification of Nuclei in Histopathology Images. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019); 2019. p. 900–904.
20. Singh A, Sengupta S, Lakshminarayanan V. Explainable Deep Learning Models in Medical Image Analysis. *Journal of Imaging*. 2020;6(6).
21. Reyes Mea. On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities. *Radiology: Artificial Intelligence*. 2020;2(3):e190043.
22. Taghanaki SAea. Infomask: Masked variational latent representation to localize chest disease. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2019. p. 739–747.
23. Li X, Zhou Y, Dvornek NC, Gu Y, Ventola P, Duncan JS. Efficient Shapley Explanation for Features Importance Estimation Under Uncertainty. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2020. p. 792–801.
24. Doran G, Ray S. Multiple-instance learning from distributions. *The Journal of Machine Learning Research*. 2016;17(1):4384–4433.
25. Kraus OZ, Ba JL, Frey BJ. Classifying and segmenting microscopy images with deep multiple instance learning. *Bioinformatics*. 2016;32(12):i52–i59.
26. Saberian MS, Moriarty KP, Olmstead AD, Nabi IR, Jean F, Libbrecht MW, et al. DEEMD: Drug Efficacy Estimation against SARS-CoV-2 based on cell Morphology with Deep multiple instance learning. *arXiv preprint arXiv:210505758*. 2021;.
27. Carbonneau MA, Cheplygina V, Granger E, Gagnon G. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition*. 2018;77:329–353.

28. Quéllec G, Cazuguel G, Cochener B, Lamard M. Multiple-instance learning for medical image and video analysis. *IEEE Reviews in Biomedical Engineering*. 2017;10:213–234.
29. Zhang D, Han J, Cheng G, Yang MH. Weakly Supervised Object Localization and Detection: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2021; p. 1–1. doi:10.1109/TPAMI.2021.3074313.
30. Larochelle H, Hinton GE. Learning to combine foveal glimpses with a third-order boltzmann machine. *Advances in neural information processing systems*. 2010;23:1243–1251.
31. Vecchiarelli AGEa. 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*. 2016;113(11):E1479–E1488.
32. Lindeberg T. Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of Applied Statistics*. 1994;21(1-2):225–270.
33. Moors JJA. The meaning of kurtosis: Darlington reexamined. *The American Statistician*. 1986;40(4):283–284.
34. Dempster AP. The Dempster–Shafer calculus for statisticians. *International Journal of Approximate Reasoning*. 2008;48(2):365–377.
35. Liu L, Yager RR. Classic works of the Dempster-Shafer theory of belief functions: An introduction. vol. 219. Springer; 2008.
36. Cantelli FP. Sui confini della probabilita. In: *Atti del Congresso Internazionale dei Matematici: Bologna del 3 al 10 de settembre di 1928; 1929*. p. 47–60.
37. George K, Faziludeen S, Sankaran P, Paul JK. Deep Learned Nucleus Features for Breast Cancer Histopathological Image Analysis based on Belief Theoretical Classifier Fusion. In: *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*. IEEE; 2019. p. 344–349.
38. Dempster AP. Upper and lower probabilities induced by a multivalued mapping. In: *Classic Works of the Dempster-Shafer Theory of Belief functions*. Springer; 2008. p. 57–72.
39. Lee H, Kwon H, Robinson RM, Nothwang WD, Marathe AM. Dynamic belief fusion for object detection. In: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*; 2016. p. 1–9.
40. Axelrod D. Total internal reflection fluorescence microscopy. *Methods in cell biology*. 1989;30:245–270.
41. Lindeberg T. Feature detection with automatic scale selection. *International journal of computer vision*. 1998;30(2):79–116.
42. Otsu N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*. 1979;9(1):62–66.
43. Basset A, Boulanger J, Salamero J, Bouthemy P, Kervrann C. Adaptive spot detection with optimal scale selection in fluorescence microscopy images. *IEEE Transactions on Image Processing*. 2015;24(11):4512–4527.

44. Smal I, Loog M, Niessen W, Meijering E. Quantitative comparison of spot detection methods in fluorescence microscopy. *IEEE transactions on medical imaging*. 2009;29(2):282–301.
45. Wong TH, Khater IM, Joshi B, Shahsavari M, Hamarneh G, Nabi IR. Single molecule network analysis identifies structural changes to caveolae and scaffolds due to mutation of the caveolin-1 scaffolding domain. *Scientific reports*. 2021;11(1):1–14.
46. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*. 1979;74(368):829–836.
47. Zadeh LA. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*. 1978;1(1):3–28.
48. Mabaso MA, Withey DJ, Twala B. SPOT DETECTION METHODS IN FLUORESCENCE MICROSCOPY IMAGING: A REVIEW. *Image Analysis & Stereology*. 2018;37(3):173–190.
49. Zimek A, Schubert E, Kriegel HP. A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*. 2012;5(5):363–387.
50. Voorbraak F. On the justification of Dempster’s rule of combination. *Artificial Intelligence*. 1991;48(2):171–197.
51. Yang Y, Han D, Han C. Discounted combination of unreliable evidence using degree of disagreement. *International Journal of Approximate Reasoning*. 2013;54(8):1197–1216.
52. Meng F, Joshi B, Nabi IR. Galectin-3 overrides PTRF/Cavin-1 reduction of PC3 prostate cancer cell migration. *PloS one*. 2015;10(5):e0126056.
53. Hill MM, Bastiani M, Luetterforst R, Kirkham M, Kirkham A, Nixon SJ, et al. PTRF-Cavin, a conserved cytoplasmic protein required for caveola formation and function. *Cell*. 2008;132(1):113–124.

Supporting information

S1. Table Imaging terms glossary.

Term	Description
I	2D fluorescence microscopy image
L	Image level label (e.g. image of patient with Alzheimer positive diagnosis: $L = AD+$)
\bar{L}	The complement of the label L, for example if $L=AD+$, then \bar{L} expresses that the patient is healthy.
\mathcal{L}	The label space, superset of all labels, defined by the domain.
$a \overset{\square}{\neq} b$	Mathematical operator denoting that a 'is not necessarily equal to' b.
$D(I)$	Function decomposing an image into objects ($o \subset I$) based on their fluorescent marker
o	Detected object, a set of adjacent pixels, in image I ($I = \{\cup o \mid \forall o \subset I\}$)
∇^2	Image Laplacian, measure of the second derivative of image intensity.
V	The absolute value of the negative part of the image Laplacian: $V = \min(\nabla^2, 0) $
σ	The standard deviation of a normal (Gaussian) distribution, used here as a parameter of Gaussian (smoothing) filter.
Z_V	Normalized negative Laplacian $Z_V = \frac{V_i - E(V)}{\sqrt{\text{Var}(V)}} \mid \forall V_i$
PRC	Precision ReCall, parameter used to scale object detection in our algorithm consistently across channels. A higher value (>1) favors recall, a lower value favors precision
STED	Stimulated Emission Depletion, a superresolution microscopy technique, amenable to live cell imaging.
PSF	Point Spread Function, the measured response or intensity distribution in 2/3D of an imaging system to a light source (point). A common mathematical model, in absence of device specific information, is the Gaussian function.
Deconvolution	In our context, using precise knowledge of the imaging system's PSF in combination with a deconvolution restores (improves) image quality.

S2. Table Statistics and belief theory glossary.

Term	Description
$k(X)$	Kurtosis of the distribution of a random variable X : $k(X) = E[(\frac{X-\mu}{\sigma})^4]$. Kurtosis of a normal distribution is 3, “excess” kurtosis is $k-3$. $k(X)$ increases as outliers are more extreme and frequent (longer tail of the distribution)
L	Categorical label assigned to an image. A single image can have multiple labels assigned.
\mathcal{L}	Set of all subsets of all labels.
\bar{L}	Set-complement of the label L , with respect to a label space \mathcal{L} . $\bar{L} = \mathcal{L} \setminus L$. E.g. in the AD+ use case, $\overline{AD+} = \text{healthy}$, however for more complex label spaces the complement is not equal to the negation.
$S_L(o)$	Function computing the statistical support, or evidence e the object o has for the image level label L ($o \subset I$)
$Bel(o \rightarrow L)$	The minimum statistical support, or belief, an object o has for the image level label L .
$p_L(o)$	Shorthand for the belief function $Bel(o \rightarrow L)$
$Pl(o \rightarrow L)$	The maximum statistic support, or plausibility, an object o has for the image level label L
$q_L(o)$	Shorthand for the plausibility function $Pl(o \rightarrow L)$
$r_L(o)$	The uncertainty in evidence based support of object o for label L : $r_L(o) = q_L(o) - p_L(o)$. See Fig. 4 for graphical illustration of the relation of uncertainty to evidence.
Θ	‘Frame of discernment’ $\Theta = \{o \rightarrow L \mid o \subset I, L \in \mathcal{L}\}$
Focal element	Subset $f \subset \Theta$ such that f does not contain any smaller subsets of Θ .
cdf/pdf	The cumulative distribution function (cdf) of a random variable X : $F(x) = P(X \leq x)$ and the probability density function (pdf) as the derivative of $F(X)$: $\int_a^b f(x)dx$.
Evidence	Statistical information that can be leveraged to compute statistical support. See Fig. 4 for graphical illustration of strong versus weak evidence, in the context of uncertainty.
Conflict	The disagreement the statistical measures (e.g. belief functions) have on the support that an object o has for an image level label L .
Weight of conflict (W)	A numerical value measuring the disagreement (conflict) between multiple belief functions.
Combination	Given 2 or more belief functions, computing the joint (combined) support as well as conflict that a single object o has for image level label L .
Separable support function	Support function where for all focal elements A, B holds that $A \cap B$ is a focal element.
Consonant support function	Support function where for all focal elements A, B holds that $A \subset B$ or $B \subset A$.

S3. Table Biology glossary.

Term	Description
CAVIN1	Cavin-1 protein, necessary to form Caveolae.
PC3	Human prostate cancer cell line, here we use the shorthand PC3 to refer to the cells with CAVIN1 expression disabled.
PC3-CAVIN1	PC3 with expression of CAVIN1 enabled.
CAV1	Caveolin-1 protein, main component of Caveolae.
Caveolae (C)	Sub-cellular structures composed of CAV1 protein complexes, ~100 nm invaginations in the cell membrane.
Oligomer	Molecule consisting of a few identical units, in our contexts proteins, e.g. CAV1 oligomer contains several CAV1 proteins.
Scaffolds (SC)	Non-caveolar scaffolds (SC), including 8S oligomers that can combine to form larger non-caveolar hemispherical scaffolds as well as caveolae.
Non-specific labelling (BG)	Fluorescent marker that is not marking its intended target, contributing to semantic “noise”.
KO	‘KnockOut’, a cell line where a particular gene’s expression is disabled using the CRISPR/Cas technique.
CAV1 KO MDA-MB-231	MDA-MB-231 is a breast cancer cell line, here with CAV1 expression disabled (KO).
Amyloid- β	A protein whose excess deposits in neuron cells have a complex role in expression of neurodegenerative disease, such as Alzheimer.

S4 Text Code and datasets SPECHT is written as a Julia module leveraging the high level features this scientific programming language offers in combination with high performance. Datasets and code are released under open source license (CC By SA 4.0, Affero GPLv3 respectively) at <https://github.com/bencardoen/SPECHT.jl>.

S5 Text Data acquisition and cultures. PC3, PC3-CAVIN1 and CRISPR/Cas CAV1 KO MDA-MB-231 cells [1, 45, 52, 53] 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 CAVIN1 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 CAVIN1 images using the CAV1 channel as reference channel using the ‘correlation full’ method. 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.

S6 Text Discussion on uncertainty in context of belief theory. 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. 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 [49]. 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.

S8 Deploying SPECHT in heavily degraded SNR conditions While the kurtosis scaling already guards the object detection stage against perturbations in the presence of noise, in certain acquisition conditions a SNR nearing 1 cannot be avoided due to, for example, non-specific labelling filling the cell, or capturing labelling that is sensitive to the health of its target. In such conditions the error margin of object detection methods diverges quickly leading to large numbers of false positives or false negatives. Achieving consistency under these conditions is non-trivial. Here we document a heuristic that can be used to deploy SPECHT in such conditions. We note that recovering missed objects is not feasible, so the first stage is ensuring maximal recall, by setting the PRC parameter to e.g. 4. In order to then screen out false positives, we can make use of the observation that labelled object in fluorescence microscopy will likely contain a local intensity maxima. In very low SNR conditions we note that the intensity distribution will be dominated, by definition, by the noise values. We can then use these two observations to define a heuristic that detects false positives. We first compute all local maxima on the raw image, optionally preprocessed with a small smoothing factor to reduce an excessive number of local maxima. Given that noise now forms the mass of the intensity distribution, by elimination the local maxima of signal will be found in the outliers of the distribution. We can detect such outliers by computing the interquartile range and setting an intensity threshold $TH \leftarrow Q1 + 1.5 \times IQR$. We then verify that each detected object contains a local intensity maxima $> TH$. Objects that do not, are removed. As SNR tends to 0 all object detection will fail, but we show that with this heuristic we can use SPECHT on a wider range of SNR values. In addition, consistency across channels is still ensured because the IQR based threshold will scale with SNR. The entire procedure is listed in Algorithm 3.

S9 Text Numerical stability. The closed form expression for combination rule in 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. 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.

Algorithm 3 False positive filtering heuristic for use in severely degraded SNR images. This is an optional postprocessing step using the output of Algorithm 1

```

1: Input Raw image  $I$ , object mask  $M$ ,  $\sigma$ 
2: Output Binary object mask  $OM$ 
3:  $I_g \leftarrow \text{Gaussian}_\sigma(I)$ 
4:  $Q1, Q3 \leftarrow \text{quantile}(I_g, [0.25, 0.75])$ 
5:  $TH \leftarrow Q3 + 1.5(Q3 - Q1)$ 
6:  $M_I \leftarrow \text{localmaxima}(I_g)$ 
7:  $OM \leftarrow \text{copy}(M)$ 
8:  $C \leftarrow \text{connectedcomponents}(M)$ 
9: for  $M_i \in M_I$  do
10:    $x, y \leftarrow \text{index}(M_i)$ 
11:   if  $M_i > TH$  then
12:     if  $M[x, y] > 0$  then
13:        $c_j \leftarrow C[x, y]$  ▷ Component id
14:        $OM[C == c_j] \leftarrow 1$ 
15:     end if
16:   end if
17: end for

```
