RESEARCH ARTICLE

A two-stage method for automated detection of ring-like endosomes in fluorescent microscopy images

Dongyun Lin[1], Zhiping Lin[1]*, Jiuwen Cao[2], Ramraj Velmurugan[3], E. Sally Ward[4,5,6], Raimund J. Ober[3,4,6]

1 School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore, **2** Key Lab for IOT and Information Fusion Technology of Zhejiang, Hangzhou Dianzi University, Hangzhou, China, **3** Department of Biomedical Engineering, Texas A&M University, College Station, TX, United States of America, **4** Department of Molecular and Cellular Medicine, Texas A&M University Health Science Center, College Station, TX, United States of America, **5** Department of Microbial Pathogenesis and Immunology, Texas A&M University Health Science Center, Bryan, TX, United States of America, **6** Cancer Sciences Unit, Centre for Cancer Immunology, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

* EZPLIN@ntu.edu.sg

OPEN ACCESS

Citation: Lin D, Lin Z, Cao J, Velmurugan R, Ward ES, Ober RJ (2019) A two-stage method for automated detection of ring-like endosomes in fluorescent microscopy images. PLoS ONE 14(6): e0218931. https://doi.org/10.1371/journal.pone.0218931

Editor: Nan Liu, Duke-NUS Medical School, SINGAPORE

Received: March 1, 2019

Accepted: June 12, 2019

Published: June 27, 2019

Data Availability Statement: All relevant data are available from https://zenodo.org/record/2577969#.XHnteegzaUk.

Funding: This work was supported in part by the National Institutes of Health (R01GM085575), to RJO (https://www.nih.gov/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. There was no additional external funding received for this study.

Abstract

Endosomes are subcellular organelles which serve as important transport compartments in eukaryotic cells. Fluorescence microscopy is a widely applied technology to study endosomes at the subcellular level. In general, a microscopy image can contain a large number of organelles and endosomes in particular. Detecting and annotating endosomes in fluorescence microscopy images is a critical part in the study of subcellular trafficking processes. Such annotation is usually performed by human inspection, which is time-consuming and prone to inaccuracy if carried out by inexperienced analysts. This paper proposes a two-stage method for automated detection of ring-like endosomes. The method consists of a localization stage cascaded by an identification stage. Given a test microscopy image, the localization stage generates a voting-map by locally comparing the query endosome patches and the test image based on a bag-of-words model. Using the voting-map, a number of candidate patches of endosomes are determined. Subsequently, in the identification stage, a support vector machine (SVM) is trained using the endosome patches and the background pattern patches. Each of the candidate patches is classified by the SVM to rule out those patches of endosome-like background patterns. The performance of the proposed method is evaluated with real microscopy images of human myeloid endothelial cells. It is shown that the proposed method significantly outperforms several state-of-the-art competing methods using multiple performance metrics.

1 Introduction

Fluorescent microscopy can produce snapshots of subcellular structures inside cells (e.g., see Fig 1a). Endosomes (shown in Fig 1b) are organelles which can be found in all eukaryotic cells

Competing interests: The authors have declared that no competing interests exist.

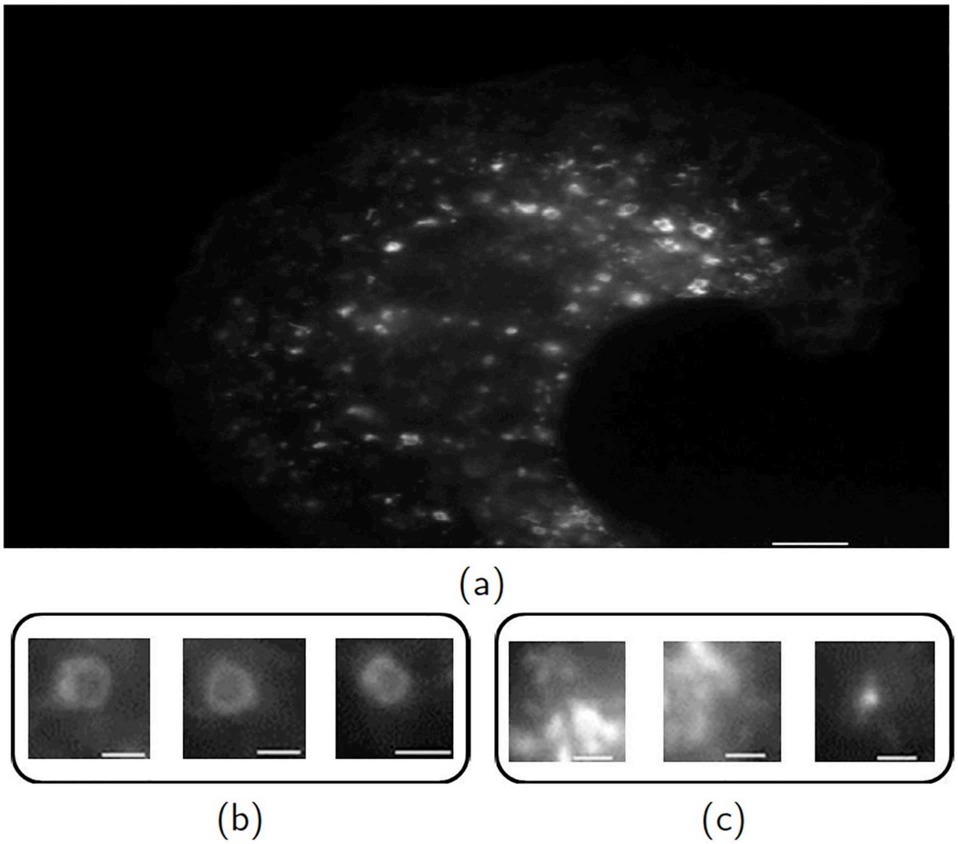


Fig 1. A microscopy images of a cell with multiple subcellular structures. (a) The whole microscopy image. (b) Three patches of ring-like endosomes. (c) Three patches of background patterns. The scale bar for (a) is 5 μm and that for (b) and (c) is 1 μm.

https://doi.org/10.1371/journal.pone.0218931.g001

and function as important transport compartments that shuttle proteins, nutrients and other materials inside cells [1]. The detection of ring-like endosomes from background patterns (e.g., see Fig 1c) is of significant biological interest relating to the analysis of interactions among proteins and organelles. For example, ring-like endosomes are relevant to the assessment of the effectiveness of a particular class of drugs called therapeutic monoclonal antibodies [2]. The annotation of endosomes is usually performed manually, which is time-consuming and inaccurate if carried out by inexperienced analysts. Hence, it is valuable to automate the process of endosome detection.

In the last decade, machine learning approaches have achieved great success in addressing tasks related to microscopy images (see e.g., [3–6]). For instance, the structures and functions of different proteins were studied based on their locations in certain organelles using image classification methods [7–12]. Motivated by these publications, this paper aims to propose a machine learning method which can automatically detect ring-like endosomes in microscopy images.

Several methods for detecting subcellular structures in fluorescent microscopy images have been proposed in the literature. Ref. [13] carried out a quantitative evaluation of multiple methods for spot detection. In general, detection methods can be divided into two groups: unsupervised and supervised methods. Unsupervised methods require some prior knowledge of the appearance model of the target and usually contain parameters to be tuned either manually or semi-automatically [13]. For instance, ref. [14] applied the fast radial symmetry transformation (FRST) [15] to detect cell nuclei based on the information that cell nuclei are approximately circular and symmetrical. On the other hand, supervised methods first "learn" the discriminative appearance information of the target from annotated patches, i.e., small regions of interest (ROIs), each of which contains either an isolated target (positive patch) or an irrelevant background pattern (negative patch). Then it searches in the test images for patches of targets from a classification perspective. An example of a supervised method is the Viola-Jones detector, i.e., the Adaboost classifier trained on Haar-like features [16, 17]. In general, most unsupervised methods are computationally efficient and perform well when the targets have simple and distinct features, e.g., distinguishing circular targets from highly non-circular background patterns. However, unsupervised methods normally fail to detect complicated targets especially when the adopted prior information is not sufficiently discriminative to distinguish the targets from the background patterns. In such scenarios, supervised methods generally outperform unsupervised methods as they better exploit the discriminative information between the targets and the background patterns. Therefore, we focus on supervised methods in this paper.

Supervised methods require a training phase before testing. In the training phase, a classifier is trained to discriminate between positive and negative patches. In the testing phase, given a microscopy image, sliding windows, i.e., overlapping small patches which scan over a relatively large image, are adopted to search all the locations in the microscopy image. Each window is fed to the pretrained classifier to determine whether it contains the target or not. In this paper, the aforementioned sliding window based supervised methods are called one-stage methods to distinguish them from our proposed two-stage supervised method. A disadvantage of one-stage supervised methods is the adoption of the sliding window search as it is computationally complicated in general and might miss the detection if the target is partially covered or "mismatched" with the sliding window, e.g., a small target staying inside a relatively large sliding window. Another important limitation of a one-stage supervised method is that the performance of its classifier highly depends on the selection of training patches that can be represented by discriminative feature vectors. Moreover, most one-stage supervised methods only exploit global information of each training patch, i.e., representing a training patch (as a whole) by a feature vector, which might deteriorate the detection performance.

In this paper, we propose a two-stage supervised method for the automated detection of ring-like endosomes in fluorescent microscopy images. The proposed method consists of a localization stage cascaded by an identification stage. Given a test microscopy image, the localization stage generates a voting-map by locally comparing the query endosome patches and the test image based on a bag-of-words model. Using the voting-map, a number of candidate patches of endosomes are determined. Subsequently, in the identification stage, a support vector machine (SVM) is trained from the endosome patches and the background pattern patches. Each of the candidate patches is classified by the SVM to rule out those patches of endosome-like background patterns. Compared with one-stage supervised methods, the proposed method has several advantages. In the localization stage, the proposed method directly localizes potential candidate patches of endosomes using the voting-map based approach, thereby avoiding the computationally complicated sliding window search. Each candidate patch locates a potential endosome at its center and the size of the patch can be automatically determined, which alleviates the "mismatched" effect caused by the sliding window search. Moreover, unlike one-stage methods which only exploit global information of training patches, the proposed method applies the local information of training patches in the localization stage and the global information in the identification stage. Both stages perform complementarily to improve the detection performance. This paper represents a significant development upon our preliminary works reported in [18].

The organization of the remaining parts of this paper is as follows. The dense SIFT based bag-of-words model is first introduced in Section 2.1 as both stages of the proposed method are based on this model. Then, the details of the localization and identification stages are presented in Sections 2.2 and 2.3, respectively. For experiments, Section 3.1 provides the specific information of the data sets. The performance metrics are introduced in Section 3.2. The experimental protocols and the parameter settings for the compared methods are specified in Section 3.3. In Section 3.4, the qualitative and quantitative experimental results are presented and analyzed, respectively. Section 4 concludes the entire paper.

2 Materials and methods

The proposed method consists of two stages: localization and identification. The localization stage is designed to provide location information of candidate patches. The identification stage is proposed to identify endosomes by further ruling out those candidate patches containing endosome-like background patterns from a classification perspective. As both stages exploit the dense SIFT based bag-of-words model for image representation, a brief review of it is given first.

2.1 Dense SIFT based bag-of-words model

The bag-of-words model is popular in image classification [19, 20] and retrieval [21, 22]. This model determines a set of representative local features (called the visual words) from a number of training images using unsupervised clustering, e.g., K-means clustering. An image can be represented as a bag-of-words histogram associated with the visual words. Here, we extract scale invariant feature transform (SIFT) features [23] from an image. SIFT features are used to describe the micro-structures of microscopy images as they remain robust across noise and illumination variation [23]. Particularly in this work, dense SIFT features are extracted from microscopy images. Unlike the original SIFT approach where a small number of keypoints are first determined from an image and SIFT features are extracted from the grids centering at these keypoints [23], dense SIFT features are extracted from dense and overlapping grids which cover the whole image [19]. Fig 2 illustrates the procedure of extracting dense SIFT features from an image. Based on this procedure, Fig 3 illustrates how to generate the bag-of-words visual words from several training images and how to represent an image by a bag-of-words histogram associated with these visual words. Note that the term “image” in both figures refers to its general meaning and it can mean a “small patch” or a “large microscopy image” depending on the context. Subsequently, note that both Figs 2 and 3 decompose the images into non-overlapping grids for easy illustration. In practice, overlapping grids for dense SIFT feature extraction are implemented.

2.2 Localization stage

The localization stage is proposed to determine candidate patches of endosomes from a microscopy image. Since endosomes are approximately circular with ring-like textures, a candidate patch of an endosome in a microscopy image can be determined by the centroid of the endosome (a pixel in the microscopy image) and a rectangle box that contains the endosome as shown in Fig 4a. To determine the centroids of the candidate endosomes, a voting-map, i.e., an image of the same size as the microscopy image, is generated. Each pixel in the voting-map quantifies the confidence of the centroid of an endosome. Fig 4b shows an example of the voting-map associated with the microscopy image in Fig 4a.

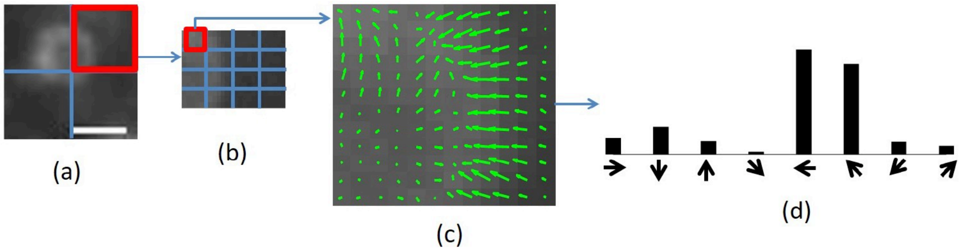


Fig 2. A diagram illustrating the procedure to extract dense SIFT features from an image. (a) Given an image, it is divided into 2 × 2 = 4 grids. (b) Each grid is then divided into 4 × 4 = 16 subgrids. Due to the border effects, the subgrids appeared at the border of each grid might have fewer pixels than the other subgrids. (c) For each subgrid, the magnitude and orientation of the gradient for each pixel are calculated. The green arrows denote the gradients where the length and direction of each arrow are the magnitude and the orientation of the gradient, respectively. (d) A histogram of gradients with respect to 8 orientation bins (in the range of 0 to 360 degrees) is built to summarize the gradients within each subgrid. A gradient is assigned to its nearest orientation bin by comparing the gradient's orientation with the 8 orientation bins. The weight of the assignment is the magnitude of the gradient. If a gradient has its orientation in the middle of 2 neighboring orientation bins, a bilinear interpolation is applied. By this procedure, each grid shown in (a) can be represented by concatenating 16 8-bin histograms, i.e., a 16 × 8 = 128 dimensional SIFT feature vector. The scale bar of Fig 2(a) is 1 μm.

https://doi.org/10.1371/journal.pone.0218931.g002

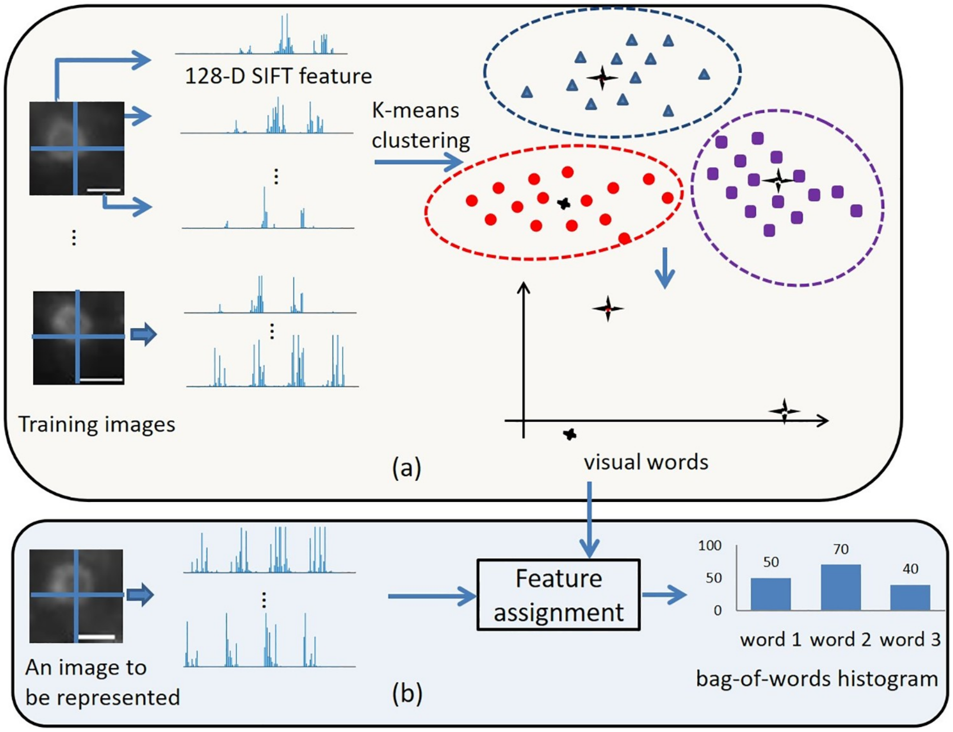


Fig 3. A diagram illustrating how to generate the bag-of-words visual words from several training images and how to represent an image by a bag-of-words histogram associated with these visual words. (a) The training images are decomposed into many SIFT features. By K-means clustering, these SIFT features are clustered to a predefined number of classes. The centroids of these clusters are referred as the visual words. In this example, three visual words are generated. The set containing all the visual words is referred as a dictionary. (b) Given an image to be represented by a bag-of-words model, it is first decomposed into several SIFT features. Each SIFT feature is assigned to its nearest visual word using Euclidean distance, i.e., vector quantization. Implementing such feature assignment for each SIFT feature, the image can be represented as a histogram of visual word frequencies, i.e., the counts of the occurrences of the SIFT features assigned to each visual word. Scale bar = 1μm.

https://doi.org/10.1371/journal.pone.0218931.g003

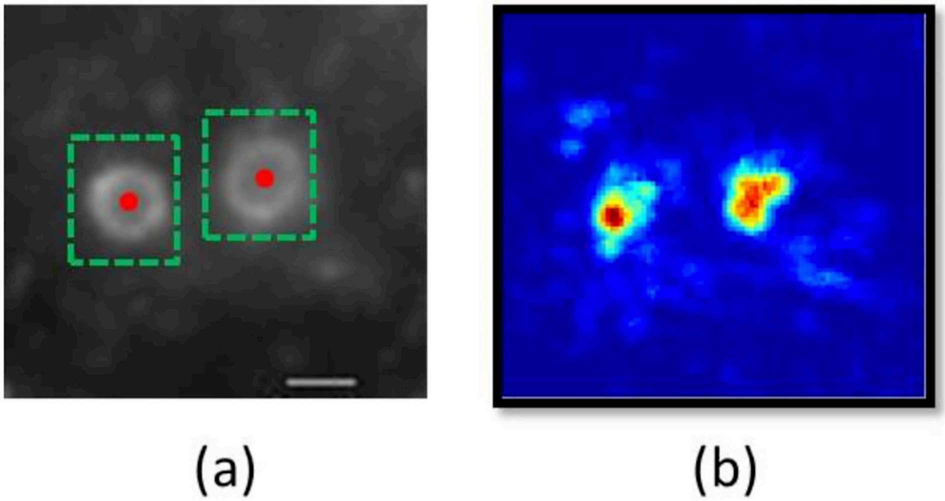


Fig 4. (a) The candidate patch of an endosome in a microscopy image can be determined by the centroid of the endosome (denoted by the red dot) and a rectangle box that contains the endosome (denoted by dashed green rectangles). (b) The voting-map for the microscopy image shown in (a). Scale bar = 1 μm.

https://doi.org/10.1371/journal.pone.0218931.g004

Motivated by the techniques proposed in [21] on natural object localization using one query patch of the object, we propose an improved category localization method for our application using multiple query patches of endosomes. Here, a query patch of an endosome is a patch of a standard endosome selected by hand. In this stage, each of the query patches and the test microscopy image are locally compared to determine the potential locations of endosome centroids. These locations are "voted" by a quantity called discriminative capability [24] to be defined later. After performing such a measurement between every query patch and the test microscopy image, a voting-map is generated. The remaining parts of this subsection show the technical details of the localization stage consisting of a training phase and a testing phase.

2.2.1 Training phase. In the training phase, dense SIFT features are first extracted from the query endosome patches and the background pattern patches. Then, a set of visual words is generated using the SIFT features. Based on the visual words, each patch is represented by a bag-of-words histogram. The location and assignment (to a visual word) for each SIFT feature are recorded. Finally, the discriminative capability of each visual word is determined. These quantities will be adopted in the testing phase for voting-map generation.

Specifically, we manually collect M (e.g., $M = 50$) patches of endosomes as a query set denoted by $\boldsymbol{Q}_{set} = \{\boldsymbol{Q}_1, \boldsymbol{Q}_2, \ldots, \boldsymbol{Q}_M\}$ and another set of N (e.g., $N = 200$) patches of background patterns denoted by $\boldsymbol{B}_{set} = \{\boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_N\}$. The size of each of these patches is of 45 × 45 pixels. The local SIFT features are extracted from each of the patches in $\boldsymbol{Q}_{set}$ and $\boldsymbol{B}_{set}$. Then a bag-of-words dictionary with K (e.g., $K = 1000$) visual words is generated from these SIFT features. With the dictionary, the bag-of-words histograms are calculated for the patches. For each query patch in $\boldsymbol{Q}_{set}$, the location of each of its SIFT features, i.e., the center pixel position of the grid where the SIFT is calculated, is recorded.

Subsequently, the discriminative capability [24] is calculated to quantify the significance of each visual word in discriminating the patches of endosomes from those of the background patterns. The discriminative capability $\boldsymbol{d}(k)$, $k = 1, 2, \ldots, K$, for the k_{th} visual word is defined as the ratio of the within-class similarity and the between-class similarity as:

$$\boldsymbol{d}(k) = \frac{\boldsymbol{\alpha}(k)}{\boldsymbol{\delta}(k)}. \tag{1}$$

The within-class similarity $\boldsymbol{\alpha}(k)$ associated with the endosome class ($\boldsymbol{Q}_{set}$) for the k_{th} visual word is defined as:

$$\boldsymbol{\alpha}(k) = \frac{1}{M \cdot (M-1)} \sum_{m=1}^{M} \sum_{l=1, l \neq m}^{M} \min\left(\boldsymbol{h}_{\mathrm{Q}_m}(k), \boldsymbol{h}_{\mathrm{Q}_l}(k)\right), \tag{2}$$

where $\boldsymbol{h}_{\mathrm{Q}_m}$ denotes the K-dimensional bag-of-words histograms of the m_{th} query patch $\boldsymbol{Q}_m$, $m = 1, 2, \ldots, M$. A large $\boldsymbol{\alpha}(k)$ indicates that the similarity among the patches from the same class is high for the k_{th} visual word, i.e., the k_{th} visual word is significant to represent the patches from the class.

On the other hand, the between-class similarity $\boldsymbol{\delta}(k)$, $k = 1, 2, \ldots, K$, for the k_{th} visual word is defined using the bag-of-words histograms of the patches from $\boldsymbol{Q}_{set}$ and $\boldsymbol{B}_{set}$ as:

$$\boldsymbol{\delta}(k) = \frac{1}{M \cdot N} \sum_{m=1}^{M} \sum_{n=1}^{N} \min\left(\boldsymbol{h}_{\mathrm{Q}_m}(k), \boldsymbol{h}_{\boldsymbol{B}_n}(k)\right), \tag{3}$$

where $\boldsymbol{h}_{\boldsymbol{B}_n}$ denotes the bag-of-words histograms of the n_{th} background patch $\boldsymbol{B}_n$, $n = 1, 2, \ldots, N$. A large $\boldsymbol{\delta}(k)$ shows that the similarity among the patches from different classes is high for the k_{th} visual word, i.e., the k_{th} visual word is significant to classify the patches from different classes.

From Eq (1), the discriminative capability $\boldsymbol{d}(k)$ for each visual word is determined. A large $\boldsymbol{d}(k)$ comes from either a large $\boldsymbol{\alpha}(k)$ or a small $\boldsymbol{\delta}(k)$. Therefore, a large $\boldsymbol{d}(k)$ indicates that the k_{th} visual word is significant not only to classify the patches from different classes but also to represent the patches from the same class.

2.2.2 Testing phase. With a test microscopy image $\boldsymbol{D}$, the goal is to search for the candidate patches within $\boldsymbol{D}$ which are visually similar to at least one query patch in $\boldsymbol{Q}_{set}$. To locate candidate patches, a voting-map based approach is exploited. Each pixel of the voting-map quantifies the confidence of an endosome centroid at the corresponding position of $\boldsymbol{D}$.

Specifically, the SIFT features are extracted from $\boldsymbol{D}$ and each of these SIFT features is assigned to a visual word generated in the training phase. To locally compare each query patch and the test microscopy image $\boldsymbol{D}$, a SIFT feature from a query patch is defined as 'matched' to another SIFT feature from $\boldsymbol{D}$ if both SIFT features are assigned to the same visual word. Assuming a pair of matched features $\boldsymbol{f}$ and $\boldsymbol{g}$ extracted from $\boldsymbol{Q}_m$ and $\boldsymbol{D}$ is assigned to the k_{th} visual word, the location of $\boldsymbol{f}$ and its relative bias vector with respect to the centroid $\boldsymbol{c}_{\mathrm{Q}_m}$ of the query patch $\boldsymbol{Q}_m$ can be determined geometrically by $L(\boldsymbol{f}) - L(\boldsymbol{c}_{\mathrm{Q}_m})$. $L(\boldsymbol{f})$ is the center pixel position of the grid where the SIFT feature $\boldsymbol{f}$ is calculated and $L(\boldsymbol{c}_{\mathrm{Q}_m})$ is the center pixel position of the query patch $\boldsymbol{Q}_m$. Subsequently, a possible location of the centroid $\boldsymbol{c}_{\boldsymbol{D}}$ of a candidate patch of an endosome in $\boldsymbol{D}$ is determined by $L(\boldsymbol{c}_{\boldsymbol{D}}) = L(\boldsymbol{g}) - s * (L(\boldsymbol{f}) - L(\boldsymbol{c}_{\mathrm{Q}_m}))$. $L(\boldsymbol{g})$ is the center pixel position of the grid where the SIFT feature $\boldsymbol{g}$ is calculated and s is the scaling factor. Since the endosomes appearing in $\boldsymbol{D}$ could be similar to a scaled version (by a scaling factor s) of the endosome in the query patch, we search through multiple scaling factors, i.e., $s_1, s_2, \ldots, s_{n_s}$, within the range of [0.5, 1.5] based on the knowledge of the potential sizes of endosomes in a microscopy image.

After determining $\boldsymbol{c_D}$, we calculate the pixel values of the voting-map $\boldsymbol{V}(s, \boldsymbol{Q}_m, \boldsymbol{D})$, which is defined as an image associated with s, $\boldsymbol{Q}_m$ and $\boldsymbol{D}$. First, all the pixel values of $\boldsymbol{V}(s, \boldsymbol{Q}_m, \boldsymbol{D})$ are initialized as zero. For a pair of matched features being assigned to the k_{th} visual word, $\boldsymbol{V}(s, \boldsymbol{Q}_m, \boldsymbol{D})$ is accumulated at the location $\boldsymbol{c_D}$ with a value of the discriminative capability $\boldsymbol{d}(k)$ which has been calculated in the training phase. By traversing all the matched feature pairs, the voting-map $\boldsymbol{V}(s, \boldsymbol{Q}_m, \boldsymbol{D})$ is generated. To integrate the information from multiple query patches, a general voting-map $\boldsymbol{V}(s, \boldsymbol{Q}_{set}, \boldsymbol{D})$ associated with s, $\boldsymbol{Q}_{set}$, and $\boldsymbol{D}$ can be determined as the summation of the voting-map associated with every query patch in $\boldsymbol{Q}_{set}$:

$$\boldsymbol{V}(s, \boldsymbol{Q}_{set}, \boldsymbol{D}) = \sum_{m=1}^{M} \boldsymbol{V}(s, \boldsymbol{Q}_m, \boldsymbol{D}). \tag{4}$$

To make the notation less clumsy, the notation for $\boldsymbol{V}(s, \boldsymbol{Q}_{set}, \boldsymbol{D})$ is simplified as $\boldsymbol{V}_s$. $\boldsymbol{V}_s(x, y)$ denotes the value of the pixel at the x_{th} row and the y_{th} column of $\boldsymbol{V}_s$. After repeating the same procedure for all the scaling factors $s_1, s_2, \ldots, s_{n_s}$, n_s voting-maps $\boldsymbol{V}_{s_1}, \boldsymbol{V}_{s_2}, \ldots, \boldsymbol{V}_{s_{n_s}}$ are generated. A final voting-map $\boldsymbol{V}_{final}$ is determined by selecting the optimal scaling factor for each pixel position. Specifically, for the position (x_v, y_v), the optimal scaling factor $s^*_{(x_v,y_v)}$ is determined by selecting the scaling factor which produces the maximum pixel value of $\boldsymbol{V}_{s_1}, \boldsymbol{V}_{s_2}, \ldots, \boldsymbol{V}_{s_{n_s}}$ at (x_v, y_v), i.e.,:

$$s^*_{(x_v,y_v)} = \arg \max_{s\in\{s_1,s_2,\ldots,s_{n_s}\}} \boldsymbol{V}_s(x_v, y_v). \tag{5}$$

The value of the final voting-map $\boldsymbol{V}_{final}$ at (x_v, y_v) is determined as:

$$\boldsymbol{V}_{final}(x_v, y_v) \triangleq \boldsymbol{V}_{s^*_{(x_v,y_v)}}(x_v, y_v). \tag{6}$$

The final voting-map $\boldsymbol{V}_{final}$ and the corresponding $s^*_{(x_v,y_v)}$ for each pixel in $\boldsymbol{V}_{final}$ will be adopted in the identification stage. Fig 5 illustrates the voting-map generation procedure.

2.3 Identification stage

After the localization stage, we generate the voting-map $\boldsymbol{V}_{final}$ and the optimal scaling factor corresponding to each pixel in $\boldsymbol{V}_{final}$. Candidate patches of endosomes centered at the locations with high voting-map pixel values can be obtained from the test microscopy image. The identification stage is proposed to train a support vector machine to identify endosome patches and rule out endosome-like background patches. The identification stage also consists of a training phase and a testing phase.

2.3.1 Training phase. In the training phase, a binary support vector machine (SVM) is trained to classify endosome patches from background pattern patches. Here, we treat endosomes as the positive class and background patterns as the negative class. The same query and background pattern patches used in the localization stage are adopted as the training patches in the identification stage.

To represent each training patch, we apply a modified bag-of-words model where locality-constrained linear coding (LLC) [20] is adopted. The LLC modifies the way that SIFT features are assigned to the visual words. Given a SIFT feature $\boldsymbol{f}$, the original feature assignment is to select the nearest visual word to $\boldsymbol{f}$ in Euclidean distance. This approach produces large quantization errors. To reduce the quantization errors, LLC encodes a SIFT feature with multiple visual words [20]. Specifically, LLC first selects C nearest visual words for $\boldsymbol{f}$ among all the visual words to form a local base $\boldsymbol{B_f} = [\boldsymbol{b_{1f}}, \boldsymbol{b_{2f}}, \ldots \boldsymbol{b_{Cf}}]$. Then it solves the following optimization

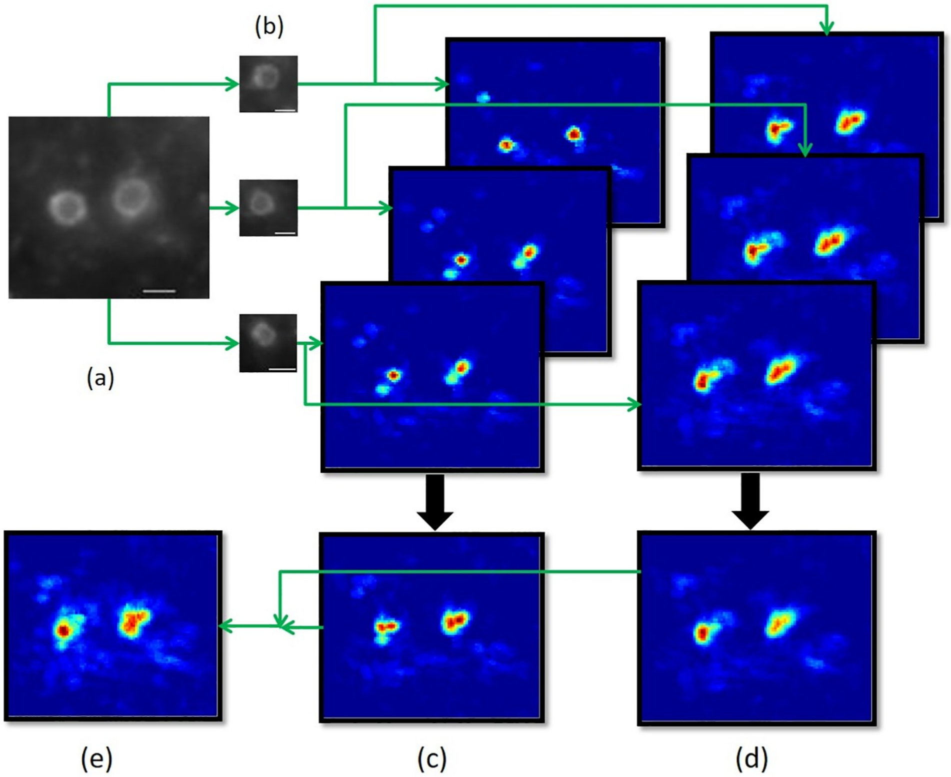


Fig 5. A diagram illustrating the procedure of voting-map generation. Three query patches and two scaling factors, i.e., 0.8 and 1, are adopted in this example. (a) The test microscopy image. (b) The three query patches. (c) The generation of the voting-maps with the scaling factor set at 0.8. For each query patch, a voting-map is generated by local feature matching between the query patch and the test image. Then a voting-map associated with the scaling factor of 0.8 is obtained by summing the three voting-maps. (d) The generation of the voting-maps with the scaling factor set at 1. (e) The final voting-map is determined by selecting the optimal scaling factor for each pixel position as described in the text. Scale bar = 1 μm.

https://doi.org/10.1371/journal.pone.0218931.g005

problem to determine a code vector $\boldsymbol{c_f} = [c_{f1}, c_{f2}, \ldots, c_{fC}]^T$ for $\boldsymbol{f}$:

$$\begin{aligned} &\min_{c_f} \|\boldsymbol{f} - \boldsymbol{B}_f \boldsymbol{c}_f\|^2, \\ &st.\ \mathbf{1}^T \boldsymbol{c}_f = 1, \end{aligned} \tag{7}$$

where **1** denotes a column vector with each entry value as 1. With each SIFT feature represented by an LLC code, the modified bag-of-words representation for the patch is determined by max-pooling these LLC codes, i.e., selecting the component-wise maximum value for each LLC code [20]. A binary SVM with a linear kernel is trained to classify the training patches represented by the modified bag-of-words histograms. The details of determining the optimal SVM hyperplane by solving a quadratic programming problem can be seen in [25].

Remark. *In the proposed method, we adopt vector quantization to encode the SIFT features with respect to the visual words in the localization stage and adopt LLC encoding in the identification stage. For the localization stage, we aim to measure the similarity between two SIFT features and therefore comparing the nearest visual word for these two features is a starightforward approach. For the identification stage, the goal is to build a dicriminative representation for a patch where using multiple visual words to represent each SIFT feature extracted from the patch has been proved more effective in* [20].

2.3.2 Testing phase. In the testing phase, the voting-map is first normalized to the range [0, 1] and a threshold γ is set for the voting-map $\boldsymbol{V}_{final}$ to rule out those locations with very low values. Consequently, the thresholded voting-map only retains the pixels whose values are larger than γ and sets other pixels to zero. Subsequently, the candidate patches of endosomes are determined. Specifically, for a non-zero pixel location $\{x_o, y_o\}$ in the thresholded voting-map, the optimal scaling factor $s^*_{(x_o,y_o)}$ which is recorded in the localization stage is retrieved. Assuming the width and height of each query patch being w and h, respectively, a candidate patch can be obtained and parameterized as $[x_{\min}, y_{\min}, x_{\max}, y_{\max}]$, where $(x_{\min}, y_{\min})$ and $(x_{\max}, y_{\max})$ denote the pixel positions of the upper-left and the bottom-right corners of the candidate patch, respectively. They are calculated as:

$$\begin{aligned} x_{\min} &= \lceil x_o - \tfrac{1}{2}s^*_{(x_o,y_o)} \times w \rceil, y_{\min} = \lceil y_o - \tfrac{1}{2}s^*_{(x_o,y_o)} \times h \rceil, \\ x_{\max} &= \lfloor x_o + \tfrac{1}{2}s^*_{(x_o,y_o)} \times w \rfloor, y_{\max} = \lfloor y_o + \tfrac{1}{2}s^*_{(x_o,y_o)} \times h \rfloor, \end{aligned} \tag{8}$$

where $\lceil . \rceil$ is the ceiling function which outputs the smallest integer larger than or equal to the input real number, and $\lfloor . \rfloor$ is the floor function which outputs the greatest integer which is smaller than or equal to the input real number. In this way, a candidate patch $[x_{\min}, y_{\min}, x_{\max}, y_{\max}]$ and its voting-map value $\boldsymbol{V}_{final}(x_o, y_o)$ is retrieved. Then, the non-maximum suppression (NMS) algorithm [26] is applied to handle the duplicated detection, i.e., for multiple candidate patches containing the same target, the patch with the highest voting-map value is selected and the others are removed (suppressed). Finally, the remaining candidate patches are classified by the SVM from the training phase as shown in Fig 6.

Algorithm 1 summarizes the overall procedure for the detection of ring-like endosomes in a test microscopy image. The algorithm consists of the localization stage and identification stage. The test microscopy image is fed to the two stages in sequence and the detection of endosomes is automatically determined.

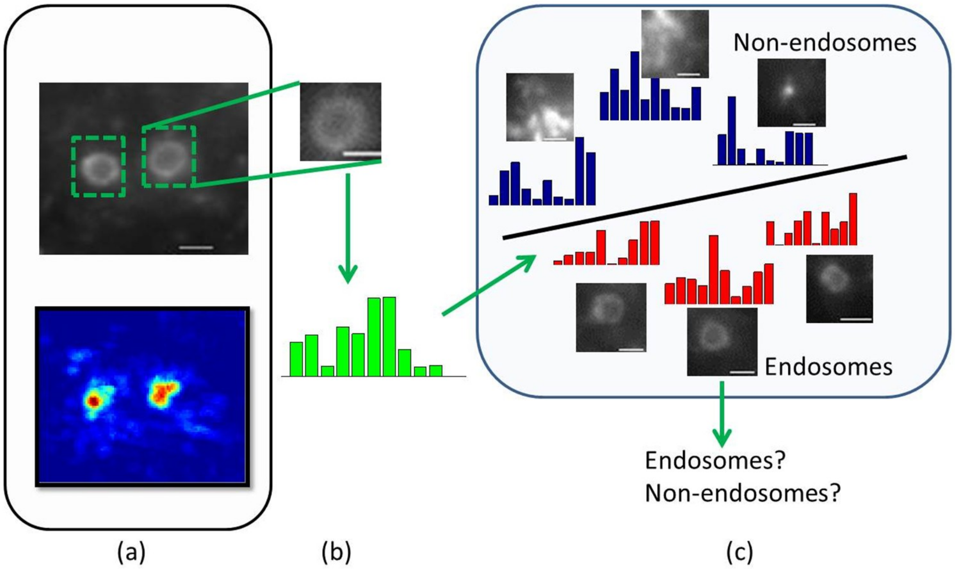


Fig 6. A diagram illustrating the identification stage using SVM. (a) After the voting-map thresholding and non-maximum suppression, the candidate patches of endosomes are obtained. (b) Each candidate patch is represented by the LLC-based bag-of-words histogram. (c) The histogram is classified by the SVM to determine whether the patch contains an endosome or a background pattern. Scale bar = 1 μm.

https://doi.org/10.1371/journal.pone.0218931.g006

Algorithm 1 The procedure for the detection of endosomes in a test microscopy image.

```
$\boldsymbol{D}$: the test image.
$\boldsymbol{Q}_m$: the $m_{th}$ query patch in the query set $\boldsymbol{Q}_{set}$.
$M$: the number of query patches.
$K$: the number of visual words.
$n_s$: the number of the elements in the scaling factor set $\boldsymbol{s} = \{s_1, s_2, ..., s_{n_s}\}$.
$L(\boldsymbol{f})$: the location of the SIFT feature $\boldsymbol{f}$.
$L(\boldsymbol{g})$: the location of the SIFT feature $\boldsymbol{g}$.
$L(\boldsymbol{c_D})$: the location of the centroid of the test image $\boldsymbol{D}$.
$L(\boldsymbol{c}_{\boldsymbol{Q}_m})$: the location of the centroid of the query patch $\boldsymbol{Q}_m$
Initialization: Initialize the pixel values for the voting-maps
$\boldsymbol{V}_{s_1}, \boldsymbol{V}_{s_2}, ..., \boldsymbol{V}_{s_{n_s}}$ as zeros.
Localization Stage:
for $s = s_1$ to $s_{n_s}$ do
  for visual word $k = 1$ to $K$ do
    for query patch $m = 1$ to $M$ do
      for local features $\boldsymbol{f} \in \boldsymbol{Q}_m$ and $\boldsymbol{g} \in \boldsymbol{D}$ do
        if $\boldsymbol{f}$ and $\boldsymbol{g}$ are both assigned to the $k_{th}$ visual word then
          $L(\boldsymbol{c_D}) = L(\boldsymbol{g}) - s * (L(\boldsymbol{f}) - L(\boldsymbol{c}_{\boldsymbol{Q}_m}))$.
          $\boldsymbol{V}_s(L(\boldsymbol{c_D})) = \boldsymbol{V}_s(L(\boldsymbol{c_D})) + \boldsymbol{d}(k)$.
        end if
      end for
    end for
  end for
end for
for Each location $\{x_v, y_v\}$ do
  Determine the optimal scaling factor $s^*_{(x_v, y_v)}$ by (5).
  Determine the values of the final voting-map $\boldsymbol{V}_{final}$ by (6).
end for
Identification Stage:
Normalize the values of the final voting-map $\boldsymbol{V}_{final}$ into the range
[0, 1] and threshold the final voting-map by a given threshold $\gamma$.
for Each location $\{x_o, y_o\}$ with non-zero thresholded voting-map value
$\boldsymbol{V}_{final}(x_o, y_o)$ do
  The optimal scaling factor $s^*_{(x_o, y_o)}$ and $\boldsymbol{V}_{final}(x_o, y_o)$ are retrieved.
  A candidate patch $\boldsymbol{P}$ centered at $\{x_o, y_o\}$ is determined by (2.3.2).
end for
Apply the non-maximum suppression (NMS) algorithm to remove the duplicated detection patches.
for Each candidate patch $\boldsymbol{P}$ do
  Calculate the bag-of-words histogram from $\boldsymbol{P}$.
  Apply the SVM to classify the histogram and determine if $\boldsymbol{P}$ contains
an endosome or not.
end for
```

3 Experiments

In this section, we first describe the dataset in Subsection 3.1. In Subsection 3.2, several metrics to evaluate the detection performance are introduced. The experimental protocol and parameter settings are specified in Subsection 3.3. The experimental results are presented and analyzed in Subsection 3.4.

The competing methods include two one-stage supervised learning methods and one unsupervised learning method. For the supervised learning methods, one is the AdaBoost classifier trained on Haar-like features (AB-Haar). This method was originally proposed in [16] to systematically build a face detector and was extended for the detection of molecular particles in

Table 1. The dataset information.

Total magnification	100.8X	157.5X
# of microscopy images	30	20
# of annotated endosomes	765	581

https://doi.org/10.1371/journal.pone.0218931.t001

live cells in [17]. The second is a linear discriminant analysis (LDA) model trained on simple pixel features (LDA-Pixel) which was proposed in [13] for spot detection in microscopy images. For the competing unsupervised method, we choose the method proposed in [14] for cell nuclei detection mainly based on image morphological operations and fast radial symmetric transform (MO-FRST).

3.1 Dataset description

The detection performance of the compared methods is evaluated using the experimental microscopy images of human myeloid endothelial cells generated in our laboratory. To generate the dataset, the human myeloid endothelial cells were maintained, transfected with DNA plasmids that express green fluorescent protein-tagged neonatal Fc receptor (FcRn-GFP), plated in glass-bottom dishes and imaged as described in [2, 27]. The images were acquired using a Zeiss AxioObserver.Z1 microscopy with widefield arc lamp illumination, a 63X 1.4NA Plan Apochromat oil immersion objective, a GFP-specific filterset (GFP-3035D-000, Semrock) and a monochromatic CCD camera (Orca ER, Hamamatsu). The dataset consists of two groups of microscopy images with different total magnifications, i.e., 100.8X and 157.5X, respectively. In total, there are 50 microscopy images which contain 1346 endosomes annotated by the analysts. Table 1 summarizes the specific information of the dataset.

3.2 Performance metrics

The detection performance is evaluated based on an overlay of the detected patches and the ground truth patches annotated by the analysts. A detected patch is defined as correct if the area of overlap (α_0) between the detected patch b_p and the ground truth patch b_{gt} exceeds 50%, where α_0 is defined as

$$\alpha_0 = \frac{area(b_p \cap b_{gt})}{area(b_p \cup b_{gt})}, \tag{9}$$

where $b_p \cap b_{gt}$ denotes the intersection of the detected and ground truth patches and $b_p \cup b_{gt}$ denotes their union [28]. Based on this criterion, three basic performance statistics, namely N_{FN}, N_{TP} and N_{FP} are determined. N_{FN} is the number of false negative detection, i.e., the number of patches which are in the ground truth but not detected by the method. N_{TP} is the number of true positive detection, i.e., the number of patches which are in the ground truth and also detected by the method. N_{FP} is the number of false positive detection, i.e., the number of patches which are detected by the method but not shown in the ground truth.

Using these three statistics, several performance metrics are defined. For classifier based detection methods, the confidence of each detected patch is quantified by a score output from the classifier, e.g., posterior probabilities. By setting thresholds to the confidence scores, we can plot (using log-log plots) the miss rate (MR) against the false positives per image (FPPI) curves [29, 30] to show the overall detection performance, where the miss rate (MR) is calculated by $MR = \frac{N_{FN}}{N_{FN}+N_{TP}}$ and the false positives per image (FPPI) is the average value of N_{FP} over the test microscopy images. Specifically, given a threshold, a pair of values of the false positives per image (FPPI) and the corresponding miss rate (MR) are determined. By varying the threshold, an MR-FPPI curve can be plotted. A better detection method produces a lower miss rate and a smaller number of the false positive detection and therefore its MR-FPPI curve stays closer to the left-bottom corner [29].

Quantitatively, four performance metrics are calculated, namely the log-average miss rate, the precision, the recall and the F_1 score. The log-average miss rate is defined based on the MR-FPPI curve. It is calculated by averaging miss rate at nine FPPI rates evenly spaced in log-space in the range 10^{-2} to 10^{0} [29]. The definitions of the precision (P), the recall (R) and the F_1 score are as follows:

$$\mathrm{P} = \frac{N_{TP}}{N_{FP} + N_{TP}}, \mathrm{R} = \frac{N_{TP}}{N_{FN} + N_{TP}}, F_1 = \frac{2 \times P \times R}{P + R}. \tag{10}$$

The precision is the ratio of the number of the correctly detected patches to that of all the detected patches. The recall is the ratio of the number of the correctly detected patches to that of all the ground truth patches. The F_1 score combines the information from both the precision and the recall. The values of log-average miss rate, the precision, the recall and the F_1 score are within [0, 1]. A better detection method produces higher values for the precision, the recall and the F_1 score while a lower value for the log-average miss rate.

3.3 Experimental protocol and parameter settings

We conduct 500 rounds of independent experiments to evaluate the detection performance of the compared methods. For each round, we randomly select 20% of the microscopy images for training and the remaining 80% of the images for testing. This experimental protocol is consistently applied to the two groups of microscopy images with different total magnifications, respectively. All the performance metrics are calculated as the average (with standard deviations) of the 500 rounds of experiments.

For the proposed method, in the localization stage, we select 50 query patches of endosomes together with 200 patches of the background patterns to determine the voting-map. The size of these patches is set at 45 × 45 (pixels). For bag-of-words model generation, each training patch (and each test microscopy image) is divided into 16 × 16 (pixels) overlapping grids and the grid space of two neighboring grids is set at 2 pixels. A SIFT feature is extracted from each grid. The number of visual words for the bag-of-words dictionary is set at 400. The range for the scaling factor searching is [0.5, 0.6, . . ., 1.5] and therefore n_s is set at 11. The threshold γ for the normalized voting-map is set at 0.2. For locality-constrained linear coding (LLC) [20], the number of visual words selected to encode a SIFT feature is set at 5. For SVM training, the regularization parameter C of the SVM is selected through a 5-fold cross validation within the range of $[2^{-8}, 2^{-7}, \ldots, 2^{7}, 2^{8}]$.

For the two competing supervised methods, namely, AB-Haar and LDA-Pixel, the positive and negative patches are the same as those adopted in the proposed method. For the AB-Haar method, each training patch is represented by the Haar-like features adopted in [13]. For the LDA-Pixel method, each training patch is represented by a feature vector generated through concatenating the horizontal rows of pixels [13]. The sliding window size is set at 45 × 45 (pixels) and the two neighboring sliding windows have 14 pixels in overlapping for both horizontal and vertical directions.

For the unsupervised method, i.e., MO-FRST [14], the searching range of the radius of FRST transformation is set at [16, 18, . . ., 30] (pixels) to cover the possible radius of an endosome. The radial strictness parameter in FRST is set at 2, the scaling factor is set at $k_\rho = 10$ and the size of the pixels for the non-maximum suppression (NMS) is set at 20 [14].

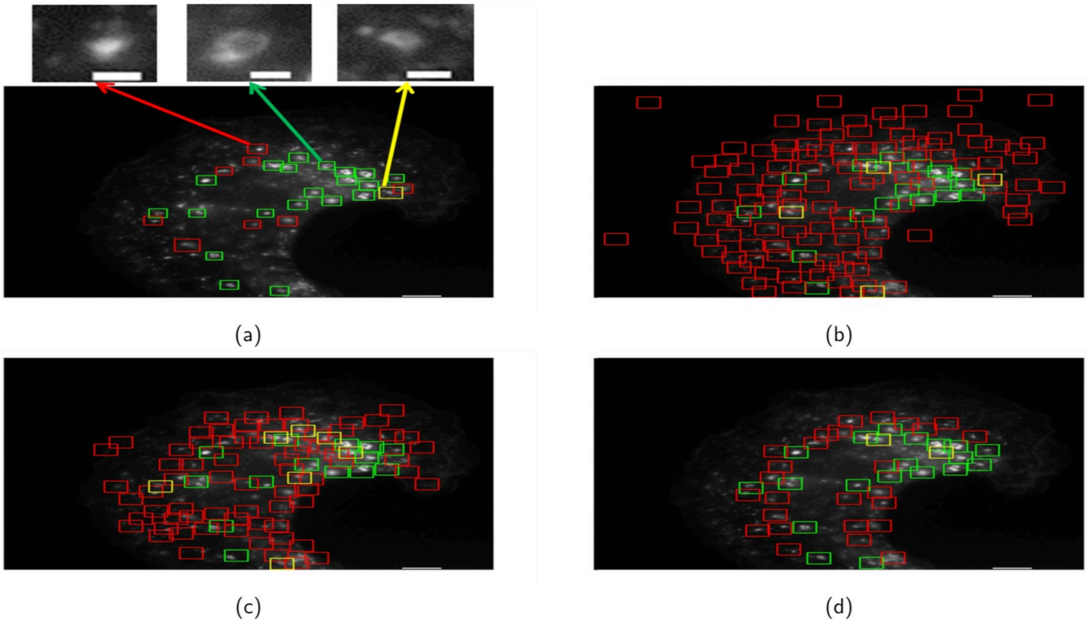


Fig 7. The detection results for an example microscopy image with the total magnification 100.8X. (a) the proposed method, (b) AB-Haar, (c) Pixel-LDA, (d) MO-FRST. The true positives, the false positives and the false negatives are represented by green, red, and yellow patches, respectively in the color version of this paper. In (a), we show three enlarged patches which are a false positive (indicated by a red arrow), a true positive (green arrow) and a false negative (yellow arrow). The scale bar for the large microscopy images is 5 μm and that for the three small patches shown in (a) is 1 μm.

https://doi.org/10.1371/journal.pone.0218931.g007

3.4 Experimental results and analysis

Figs 7 and 8 show the qualitative detection performance for two test microscopy images with different total magnifications. As shown in Figs 7a and 8a, the proposed method approximately localizes each endosome at the center of the detected patch and automatically determines the size of each detected patch. However, Both AB-Haar and LDA-Pixel methods suffer from the "mismatch" between the sliding window and the endosome (see Figs 7b and 7c, 8b and 8c). The MO-FRST can localize the centroid of all the circular structures but fail to further distinguish the endosomes and background patterns that are circularly shaped (see Figs 7d and 8d). The proposed method produces the highest number of true positive detection and the lowest number of false positive and false negative detection among all the compared methods. It is also noted that the proposed method can better handle the detection of the endosomes which stay very close to or even touching each other since the proposed localization stage locally compares the SIFT features. Hence, even when an endosome only partially appears, there are still some local features being determined as matched to the local features extracted from the query patches. However, it is observed that the performance of the competing methods significantly deteriorate in this scenario.

Quantitatively, the detection performance of the compared methods is shown by the MR-FPPI curves and the four metrics introduced in Section 3.2. Fig 9a and 9b show the MR-FPPI curves of the compared methods for the two groups of microscopy images with different total magnifications, respectively. Each curve is plotted based on one typical trial out of 500 rounds of experiments. It is noted that we have similar observations from the MR-FPPI curves of different experimental trials. It is observed that the proposed method outperforms the competing methods as its MR-FPPI curve stays the closest to the left-bottom corner. Table 2 presents the quantitative performance metrics of the compared methods (the results

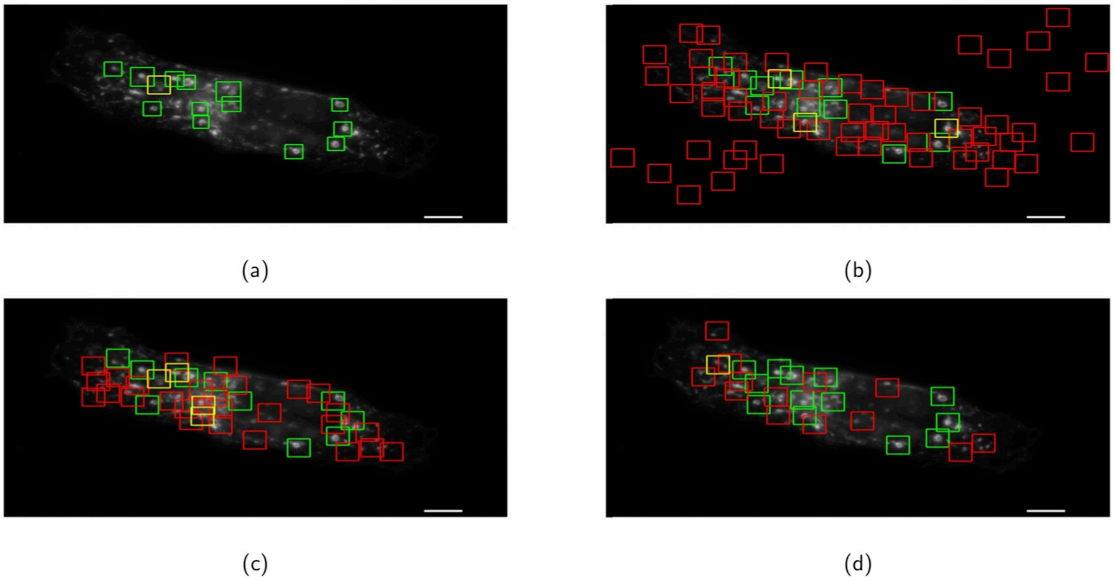


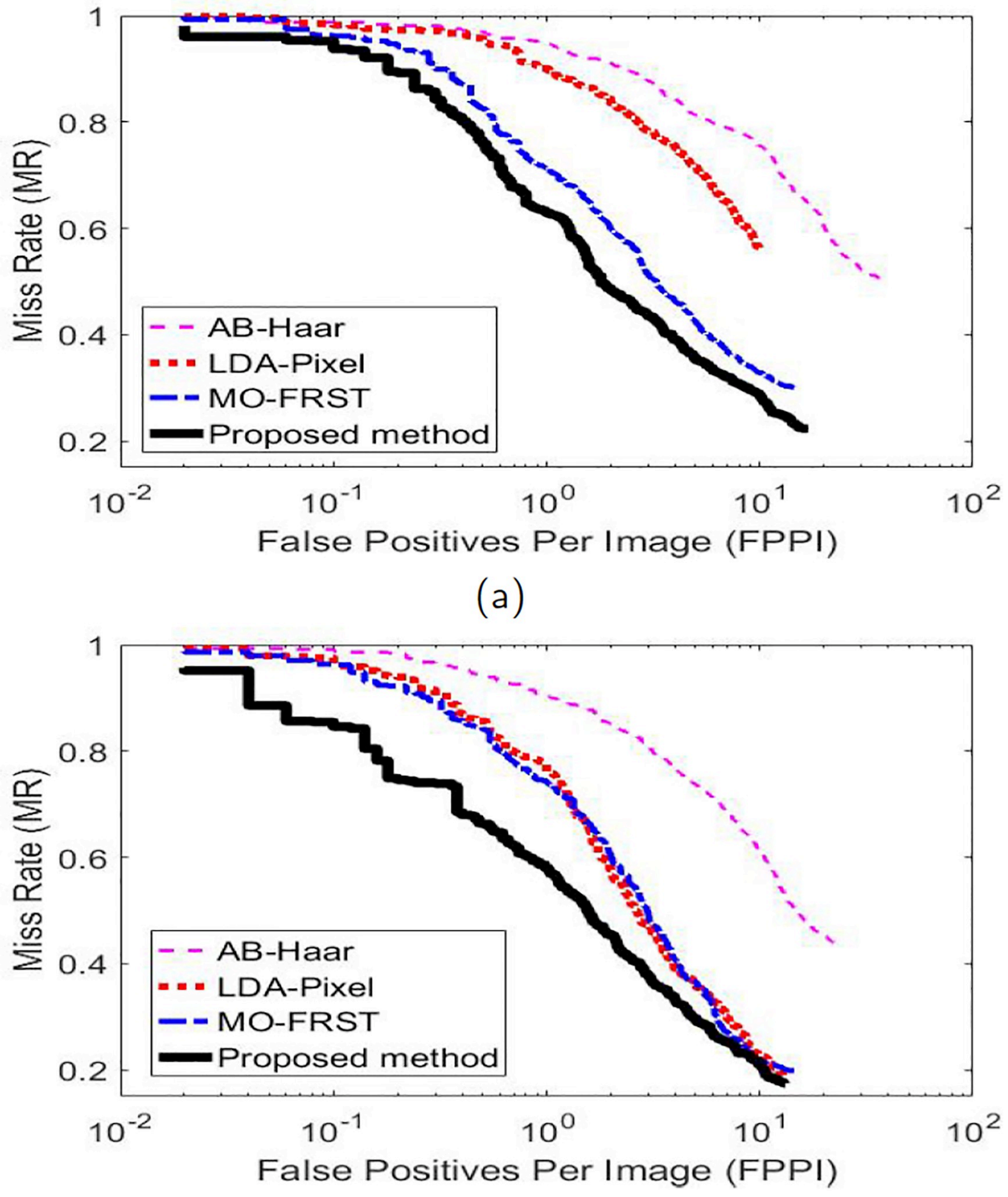
Fig 8. The detection results for an example microscopy image with the total magnification 157.5X. (a) the proposed method, (b) AB-Haar, (c) Pixel-LDA, (d) MO-FRST. Here, the notations are the same as Fig 7. The true positives, the false positives and the false negatives are represented by green, red, and yellow patches, respectively. Scale bar = 5 μm.

https://doi.org/10.1371/journal.pone.0218931.g008

for the best method are bolded). It is observed from the table that the proposed method significantly outperforms all the competing methods in terms of producing the lowest log-average miss rate and the highest precision, recall and F_1 score, respectively. This indicates the proposed method produces more accurate detection of endosomes and lower missing detection than the competing methods.

More specifically, for those one-stage supervised methods, since ring-like endosomes are not just simple "bright structures" in the dark background like spots, features like Haar-like features or pixel intensities adopted for bright spot detection might not be sufficiently discriminative to describe the micro-structures of endosomes. On the other hand, the proposed method extracts SIFT features summarizing local gradients which can robustly capture the distinct local structures of endosomes despite of the influences of noise and illuminations in the microscopy images. Therefore, the proposed method outperforms the two one-stage supervised methods by extracting more discriminative features. Moreover, the proposed two-stage detection structure utilizes both local and global information from the endosome patches and the background pattern patches. On the one hand, searching for matched SIFT features in the localization stage provides the location of each candidate endosome. On the other hand, the adoption of the bag-of-words model modified by LLC to represent the patches for SVM training applies the global information of each training patch as a whole to identify endosomes. Hence, the proposed method outperforms the one-stage supervised detection methods whose detection performance highly relies on the classifiers as only the global information is incorporated.

For the comparison with the unsupervised method, MO-FRST [14] was specifically formulated to detect cell nuclei which are circular and symmetrical "bright structures". However, endosomes are only approximately circular and symmetrical. Moreover, they have ring-like textures. MO-FRST produces high false positive detection since it only captures all circular and symmetrical patterns but fails to further identify the ring-like structures. In contrast, the proposed method exploits distinct textures of ring-like endosomes in both stages and hence produces much lower false positive detection rate.



(a)

Fig 9. The MR-FPPI curves of the compared methods. (a) the microscopy images with the total magnification of 100.8X. (b) the microscopy images with the total magnification of 157.5X.

https://doi.org/10.1371/journal.pone.0218931.g009

4 Conclusion

In this paper, we have proposed a two-stage method for automatically detecting ring-like endosomes in fluorescent microscopy images. The proposed method is a supervised method and therefore it does not require much prior knowledge as unsupervised methods do. Once the training phase in the two stages is performed off-line, the method can achieve fully automated detection of ring-like endosomes. Essentially, the method exploits local SIFT feature

Table 2. Average performance metrics (and standard deviations) for the compared methods.

Methods	Log-average miss rate	Precision	Recall	F_1 score
The images with a total magnification of 100.8X				
AB-Haar [17]	0.529 (0.006)	0.296 (0.012)	0.660 (0.007)	0.409 (0.011)
LDA-Pixel [13]	0.344 (0.005)	0.615 (0.025)	0.636 (0.008)	0.625 (0.010)
MO-FRST [14]	0.269 (0.009)	0.519 (0.011)	0.761 (0.005)	0.617 (0.007)
Proposed method	**0.138 (0.007)**	**0.812 (0.009)**	**0.807 (0.005)**	**0.773 (0.006)**
The images with a total magnification of 157.5X				
AB-Haar [17]	0.457 (0.022)	0.338 (0.009)	0.695 (0.007)	0.455 (0.008)
LDA-Pixel [13]	0.235 (0.023)	0.475 (0.013)	0.846 (0.009)	0.608 (0.011)
MO-FRST [14]	0.241 (0.015)	0.441 (0.018)	0.833 (0.005)	0.577 (0.016)
Proposed method	**0.097 (0.013)**	**0.819 (0.012)**	**0.850 (0.015)**	**0.770 (0.009)**

https://doi.org/10.1371/journal.pone.0218931.t002

matching to locate candidate endosome patches and then identifies each of the candidate patches using a pretrained SVM from a classification perspective. The experiments on the real microscopy images show that the proposed method can produce significantly better detection results compared with several competing methods in terms of multiple detection performance metrics.

Author Contributions

Conceptualization: Dongyun Lin, Zhiping Lin, Jiuwen Cao, Raimund J. Ober.

Formal analysis: Dongyun Lin.

Funding acquisition: Raimund J. Ober.

Investigation: Dongyun Lin, Zhiping Lin.

Methodology: Dongyun Lin, Zhiping Lin, Jiuwen Cao.

Project administration: E. Sally Ward, Raimund J. Ober.

Resources: Ramraj Velmurugan.

Software: Dongyun Lin.

Supervision: Zhiping Lin, E. Sally Ward, Raimund J. Ober.

Validation: Dongyun Lin, Ramraj Velmurugan.

Visualization: Dongyun Lin.

Writing – original draft: Dongyun Lin.

Writing – review & editing: Dongyun Lin, Zhiping Lin, Jiuwen Cao, Raimund J. Ober.

References

1. Mellman I. Endocytosis and molecular sorting. Annual review of cell and developmental biology. 1996; 12(1):575–625. https://doi.org/10.1146/annurev.cellbio.12.1.575 PMID: 8970738
2. Devanaboyina SC, Lynch SM, Ober RJ, Ram S, Kim D, Puig-Canto A, et al. The effect of PH dependence of antibody-antigen interactions on subcellular trafficking dynamics. In: MAbs. vol. 5. Taylor & Francis; 2013. p. 851–859.
3. Sommer C, Gerlich DW. Machine learning in cell biology–teaching computers to recognize phenotypes. J Cell Sci. 2013; 126(24):5529–5539. https://doi.org/10.1242/jcs.123604 PMID: 24259662

4. Calderon CP. Data-driven techniques for detecting dynamical state changes in noisily measured 3D single-molecule trajectories. Molecules. 2014; 19(11):18381–18398. https://doi.org/10.3390/molecules191118381 PMID: 25397733

5. Ozaki Y, Yamada H, Kikuchi H, Hirotsu A, Murakami T, Matsumoto T, et al. Label-free classification of cells based on supervised machine learning of subcellular structures. PLOS ONE. 2019; 14(1):1–20. https://doi.org/10.1371/journal.pone.0211347

6. Shimahara Y, Sugawara K, Kojo KH, Kawai H, Yoshida Y, Hasezawa S, et al. IMACEL: A cloud-based bioimage analysis platform for morphological analysis and image classification. PLOS ONE. 2019; 14 (2):1–11. https://doi.org/10.1371/journal.pone.0212619

7. Boland MV, Murphy RF. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. Bioinformatics. 2001; 17 (12):1213–1223. https://doi.org/10.1093/bioinformatics/17.12.1213 PMID: 11751230

8. Murphy RF, Velliste M, Porreca G. Robust numerical features for description and classification of subcellular location patterns in fluorescence microscope images. Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology. 2003; 35(3):311–321. https://doi.org/10.1023/B:VLSI.0000003028.71666.44

9. Johnson GR, Li J, Shariff A, Rohde GK, Murphy RF. Automated learning of subcellular variation among punctate protein patterns and a generative model of their relation to microtubules. PLoS Comput Biol. 2015; 11(12):e1004614. https://doi.org/10.1371/journal.pcbi.1004614 PMID: 26624011

10. Coelho LP, Peng T, Murphy RF. Quantifying the distribution of probes between subcellular locations using unsupervised pattern unmixing. Bioinformatics. 2010; 26(12):i7–i12. https://doi.org/10.1093/bioinformatics/btq220 PMID: 20529939

11. Hu Y, Osuna-Highley E, Hua J, Nowicki TS, Stolz R, McKayle C, et al. Automated analysis of protein subcellular location in time series images. Bioinformatics. 2010; 26(13):1630–1636. https://doi.org/10.1093/bioinformatics/btq239 PMID: 20484328

12. Shariff A, Murphy RF, Rohde GK. A generative model of microtubule distributions, and indirect estimation of its parameters from fluorescence microscopy images. Cytometry Part A. 2010; 77(5):457–466.

13. Smal I, Loog M, Niessen W, Meijering E. Quantitative comparison of spot detection methods in fluorescence microscopy. IEEE Transactions on Medical Imaging. 2010; 29(2):282–301. https://doi.org/10.1109/TMI.2009.2025127 PMID: 19556194

14. Ram S, Rodríguez JJ. Size-Invariant Detection of Cell Nuclei in Microscopy Images. IEEE Transactions on Medical Imaging. 2016; 35(7):1753–1764. https://doi.org/10.1109/TMI.2016.2527740 PMID: 26886972

15. Loy G, Zelinsky A. Fast radial symmetry for detecting points of interest. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2003; 25(8):959–973. https://doi.org/10.1109/TPAMI.2003.1217601

16. Viola P, Jones MJ. Robust real-time face detection. International Journal of Computer Vision. 2004; 57 (2):137–154. https://doi.org/10.1023/B:VISI.0000013087.49260.fb

17. Jiang S, Zhou X, Kirchhausen T, Wong ST. Detection of molecular particles in live cells via machine learning. Cytometry Part A. 2007; 71(8):563–575. https://doi.org/10.1002/cyto.a.20404

18. Lin D, Lin Z, Velmurugan R, Ober RJ. Automatic endosomal structure detection and localization in fluorescence microscopic images. In: 2017 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE; 2017. p. 1–4.

19. Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). vol. 2. IEEE; 2006. p. 2169–2178.

20. Wang J, Yang J, Yu K, Lv F, Huang T, Gong Y. Locality-constrained linear coding for image classification. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2010. p. 3360–3367.

21. Shen X, Lin Z, Brandt J, Wu Y. Spatially-constrained similarity measure for large-scale object retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2014; 36(6):1229–1241. https://doi.org/10.1109/TPAMI.2013.237

22. Sivic J, Zisserman A. Video Google: A text retrieval approach to object matching in videos. In: Ninth IEEE International Conference on Computer Vision, 2003. Proceedings. IEEE; 2003. p. 1470–1477.

23. Lowe DG. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision. 2004; 60(2):91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

24. Cao J, Chen T, Fan J. Landmark recognition with compact BoW histogram and ensemble ELM. Multimedia Tools and Applications. 2016; 75(5):2839–2857. https://doi.org/10.1007/s11042-014-2424-1

25. Burges CJ. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery. 1998; 2(2):121–167. https://doi.org/10.1023/A:1009715923555

26. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). vol. 1. IEEE; 2005. p. 886–893.

27. Gan Z, Ram S, Vaccaro C, Ober RJ, Ward ES. Analyses of the recycling receptor, FcRn, in live cells reveal novel pathways for lysosomal delivery. Traffic. 2009; 10(5):600–614. https://doi.org/10.1111/j.1600-0854.2009.00887.x PMID: 19192244

28. Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A. The Pascal visual object classes (VOC) challenge. International Journal of Computer Vision. 2010; 88(2):303–338. https://doi.org/10.1007/s11263-009-0275-4

29. Dollar P, Wojek C, Schiele B, Perona P. Pedestrian detection: An evaluation of the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012; 34(4):743–761. https://doi.org/10.1109/TPAMI.2011.155 PMID: 21808091

30. Satpathy A, Jiang X, Eng HL. LBP-based edge-texture features for object recognition. IEEE Transactions on Image Processing. 2014; 23(5):1953–1964. https://doi.org/10.1109/TIP.2014.2310123 PMID: 24690574