A REVIEW OF SUPER-RESOLUTION SINGLE MOLECULE LOCALIZATION MICROSCOPY CLUSTER ANALYSIS AND QUANTIFICATION METHODS

A PREPRINT

Ismail M. Khater *
Medical Image Analysis Lab
School of Computer Science
Simon Fraser University
Burnaby, BC V5A 1S6, Canada
ikhater@sfu.ca

Ivan Robert Nabi
Department of Cellular and
Physiological Sciences
Life Sciences Institute
University of British Columbia
Vancouver, BC V6T 1Z3, Canada
irnabi@mail.ubc.ca

Ghassan Hamarneh
Medical Image Analysis Lab
School of Computer Science
Simon Fraser University
Burnaby, BC V5A 1S6, Canada
hamarneh@sfu.ca

April 8, 2020

ABSTRACT

Single molecule localization microscopy (SMLM) is a relatively new imaging modality, winning the 2014 Nobel Prize in Chemistry, and considered as one of the key super-resolution techniques. SMLM resolution goes beyond the diffraction limit of light microscopy and achieves resolution on the order of 10–20 nm. SMLM thus enables imaging single molecules and studying the low-level molecular interactions at the subcellular level. In contrast to standard microscopy imaging that produce 2D/3D pixel/voxel grid data, SMLM generates big-data of 2D/3D point-clouds with millions of localizations and associated uncertainties. This unprecedented breakthrough in imaging helps researchers employ SMLM in many fields within biology and medicine, such as studying cancerous cells, cell-mediated immunity, and accelerating drug discovery. However, SMLM data quantification and interpretation methods have yet to keep pace with the rapid advancement of SMLM imaging. Researchers have been actively exploring new computational methods for SMLM data analysis to extract biosignatures of various biological structures and functions. In this survey, we describe the state-of-the-art clustering methods adopted to analyze and quantify SMLM data and examine the capabilities and shortcomings of the surveyed methods. We classify the methods according to (i) the biological application (i.e. the imaged molecules/structures); (ii) the data acquisition (imaging modality, dimension, resolution, number of localizations, etc.); (iii) the analysis details (2D vs. 3D, field-of-view vs. region-of-interest, use of machine learning and multi-scale analysis, biosignature extraction, etc.). We observe that the majority of the methods that are based on second order statistics are sensitive to noise and imaging artifacts, have not been applied to 3D data, do not leverage machine learning formulations, and are not scalable for big-data analysis. Finally, we summarize state-of-the-art methodology, discuss some key open challenges, and identify future opportunities for better modeling and design of integrated computational pipeline to address the key challenges.

Keywords Super-resolution · Nanoscopy · Single Molecule Localization Microscopy · SMLM · Cluster analysis · Quantification · Biological structures · Molecular complexes · Point clouds

1 Introduction

Cells are the structural and functional units of living organisms. Studying the cell requires an understanding of its different compartments and their relationship to each other inside and outside the cell. With the aid of microscopes,

*Corresponding Author and Lead Contact

researchers can visualize, identify, and study cell organelles and molecular components, which is critical to understand cell function in health and malfunction in different diseases. The recent advent of super-resolution microscopy, which provides an order of magnitude improvement in resolution compared to light microscopy allows visualization and quantification of the organization of proteins to form macromolecular complexes ***in situ*** in intact cells, facilitating our understanding of molecular interactions in different biological structures that drive cell behavior. This revolutionary discovery advances science to better understand cellular function and the machinery of its subcellular compartments.

1.1 Resolution Limit

Optical microscopy (also known as a light microscopy) uses visible light and series of lenses to image and magnify cell and tissue samples. Light is usually used in non-invasive imaging of cells ***in vitro*** and ***in vivo*** for various imaging applications, which helps researchers magnify and visualize biological structures within the optical resolution of the imaging system. The resolution of the imaging system is affected by physical phenomena such as lens misalignment and the diffraction of light, known as the diffraction limit of the imaging system (i.e. the microscope). The diffraction limit of the light microscope, which was theoretically described by Abbe in 1873 [1], is proportional to the wavelength of the light (λ) being observed and inversely proportional to the numerical aperture (NA) of the objective lens. The formulation of Abbe's diffraction limit is given in the equation 1 [2, 3].

$$d = \frac{\lambda}{2\text{NA}_{obj}} \tag{1}$$

For example, if we use visible light with a wavelength of 500 nm (green light) and an objective lens with NA_{obj} of 1.0, then the minimum theoretical distance d will be 250 nm that defines two separable objects [2]. However, practically, the resolution can be less than that due to various experimental reasons. Abbe's diffraction limit of the light microscopy (i.e. 250 nm) is a major barrier that has prevented researchers from studying multiple biological structures and macromolecular complexes below the diffraction barrier. Breaking this barrier has been achieved finally with development of super-resolution microscopes (Section 2) and improvements in fluorescent probes and labelling techniques for super-resolution microscopy [4].

2 Super-resolution Nanoscopy Methods

The invaluable contribution of super-resolution microscopy was acknowledged by the 2014 Nobel Prize in Chemistry awarded jointly to three scholars for their contribution to bringing the light microscopy into the nano-meter scale (i.e. nanoscopy). Eric Betzig, William E. Moerner, and Stefan W. Hell shared the prize [3,5,6]. Betzig and Moerner developed the principles of single molecule localization microscopy (SMLM) [7, 8] and Hell worked on stimulated emission depletion (STED) microscopy [9]. Another microscopy method, structured illumination microscopy (SIM) [10], developed by Mats Gustafsson lately gained rapid popularity [6]. Another super-resolution method, super-resolution optical fluctuation imaging (SOFI) [11], has been developed to overcome the diffraction limit of light. SOFI is faster than SMLM but has a lower resolution. Schidorsky et al. [12] combined SOFI and SMLM to improve the overall imaging performance. They showed that by rejecting common background sources, SOFI-assisted SMLM can be used to improve image reconstruction. Figure 1 depicts the various super-resolution methods developed to break the diffraction limit barrier of the light microscopy.

SMLM methods include: photoactivated localization microscopy (PALM) [7, 13], stochastic optical reconstruction microscopy (STORM) [14], *direct* STORM (*d*STORM) [15], ground state depletion (GSD) [16], DNA-based point accumulation for imaging in nanoscale topography (DNA-PAINT) [17], and MINFLUX [18]. SMLM achieved the highest resolution among the super-resolution methods (Fig. 1). The lateral resolution of SMLM could be from 10 to 30 nm (minflux achieves 2 nm resolution). The STED lateral resolution reaches 60–100 nm, while it is about 100–120 nm for the SIM. On the other hand, the analysis complexity of SMLM is ranked as the most complex according to Owen et al. [19], followed by SIM as intermediate in the analysis complexity and, finally, STED as the simplest. Wegel et al. [20] experimentally studied the super-resolution methods, including SMLM, and applied them to image various subcellular structures. They showed the weakness and strength of each method on the studied structures (e.g. vesicles, filaments).

Super-resolution microscopy has allowed for unprecedented high resolution visualization of various biological structures such as microtubules, actin, clathrin-coated pits, mitochondria, chromatin complexes, neurons, endoplasmic reticulum, and focal adhesion complexes [21]. However, the initial demand for high resolution images of biological structures has been replaced by a need for quantitative methods and analysis [22]. SMLM imaging methods produces spatial coordinates of molecular localizations, called 'point clouds' in this survey, that are ideally suited for the application of cluster analysis algorithms and tools. We focus this survey on state-of-the-art super-resolution SMLM cluster analysis

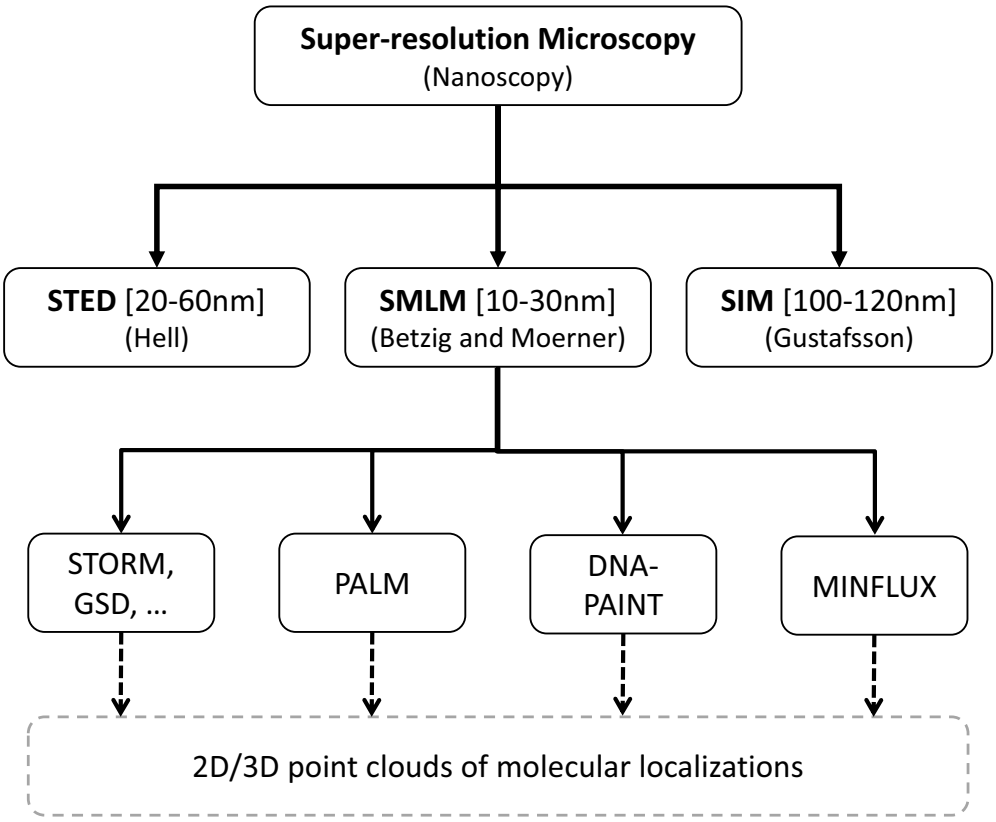

Figure 1: The classification of super-resolution nanoscopy methods. All SMLM methods generate localizations as 2D/3D point clouds.

methods and their capabilities and shortcomings. Note that we do not intend to survey all possible data clustering methods but rather limit the presented works to those methods that have been applied to SMLM data clustering.

2.1 From Imaging to Quantification

Figure 2 gives an overview of the imaging-to-quantification pipeline for SMLM, which starts with fluorescent labelling of the target molecule, determining molecular localization from the acquired SMLM images and ends with post-processing and quantification of the imaged and localized proteins. Our focus in this survey will be on cluster analysis and quantification of SMLM data.

2.2 Acquisition and Localization

The first step in imaging or tracking a protein is to label the target protein with a fluorescent dye. The labelling process varies depending on the SMLM imaging technique. For example, for STORM imaging the target molecule is labeled via antibodies conjugated to organic dyes (e.g. Alexa 647 molecules)(Fig. 2). In PALM, genetically modified fluorescent proteins (FPs) are used in the labelling of the target proteins (e.g. mEos2). Also, the labelling strategy may depend on the binding proteins/antibodies used in the labelling. We show the primary-secondary labelling strategy in Figure 2A as an example. Other labeling strategies, including using fragment antigen-binding (Fab) antibodies (that can be obtained for both primary and secondary labelling), nanobody-labelling, etc., might be used to reduce the size of the fluorescent probe and improve the resolution [25–28].

Although the exact implementation may vary, all SMLM methods fundamentally rely on temporal separation of the emissions of the excited fluorophores [4, 29], where the fluorophores are sparsely activated and forced to switch between **on** "bright state" and **off** "dark state" stochastically during the imaging session. Stochastic “blinking” of non-overlapping point spread functions (PSFs), formed due to diffraction of light, are recorded by the imaging system. Positional localization of individual fluorophores is approximated to be at the centre of a Gaussian fitting of the PSF, as shown in Figure 3, resulting in significantly improved ($\sim$10X) resolution. By repeating this process thousands of times

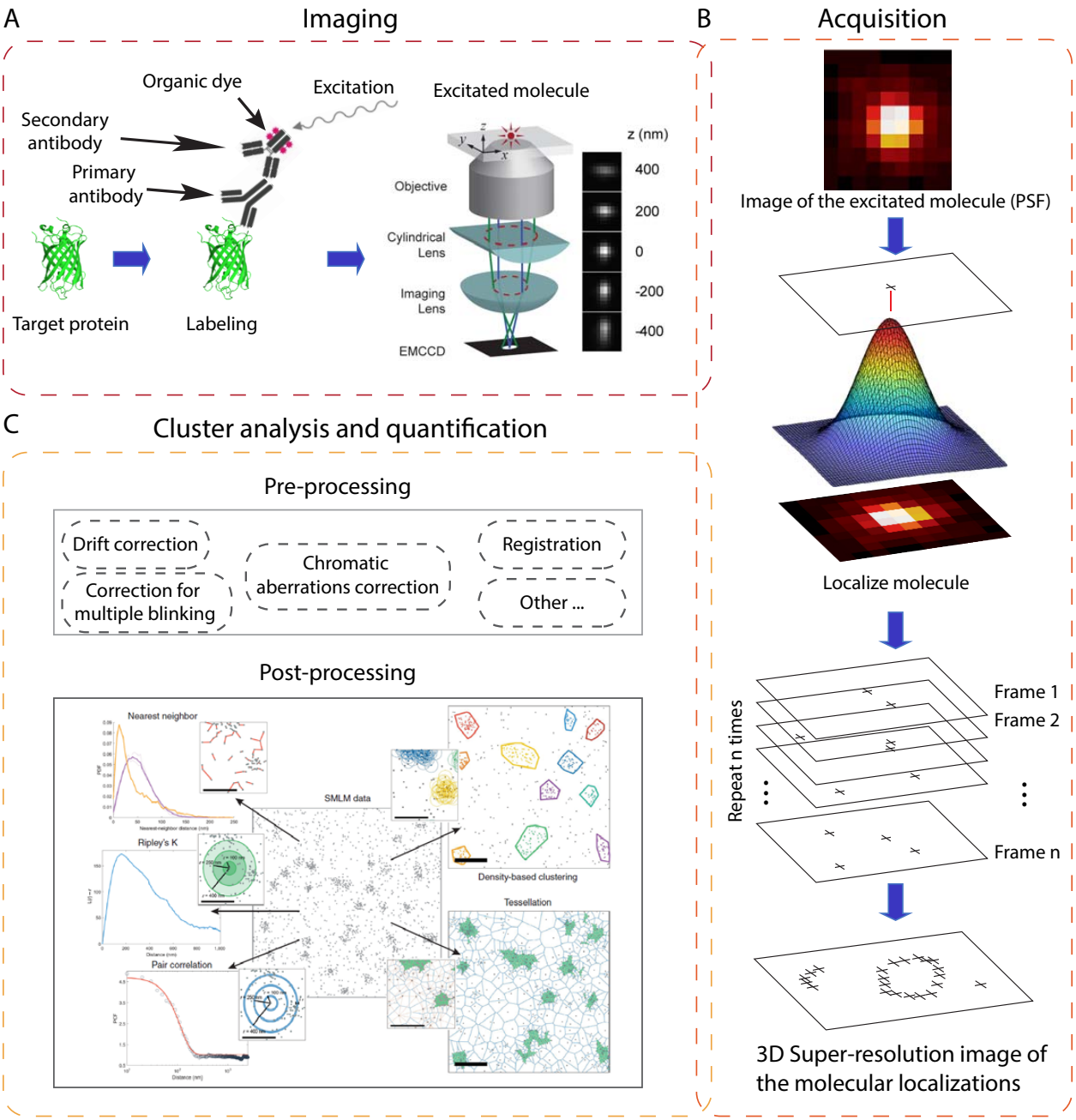

Figure 2: An overview of the whole framework from imaging to quantification of the super-resolution SMLM data: (A) 3D SMLM imaging of the target protein [23], (B) acquiring the protein localizations and getting a map for the molecular coordinates, and (C) analyzing the super-resolved image to quantify the SMLM clusters. Part (C) is produced with permission from [24].

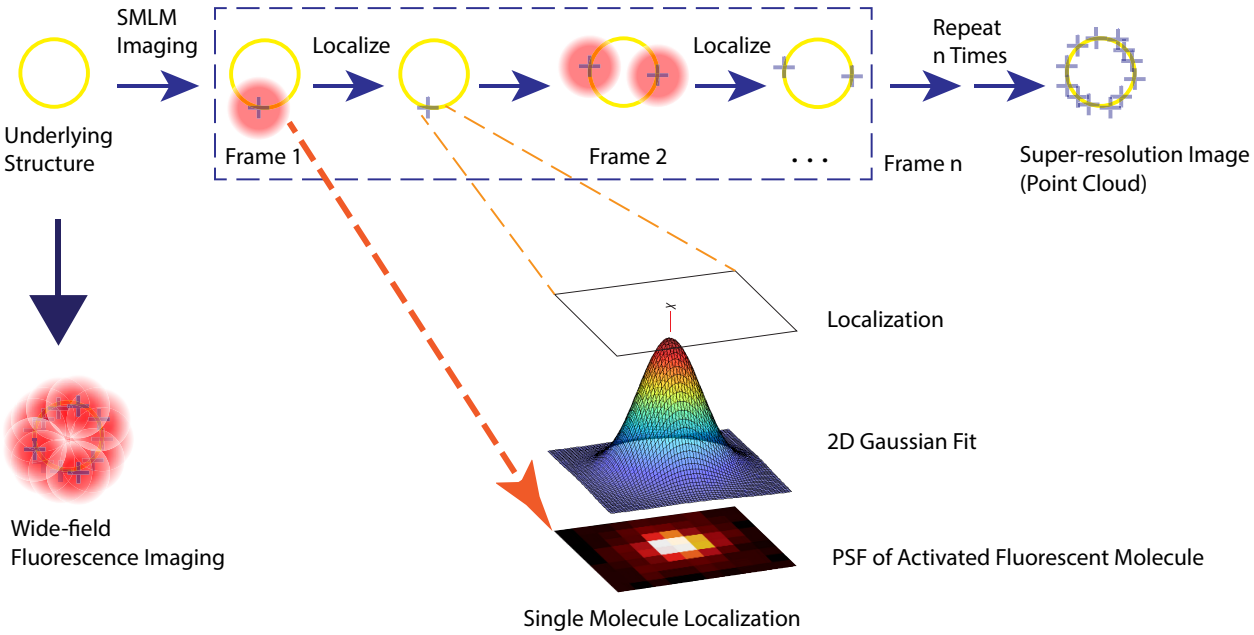

Figure 3: Illustration of the SMLM imaging principle. Labelling the yellow circle (i.e. biological structure below the diffraction limit) efficiently with fluorescent dye to be imaged with a fluorescent microscope. The conventional diffraction-limited wide-field microscope produces a blurred image. The SMLM imaging produces a super-resolved image that is constructed from a set of time-separated images, wherein each time frame image a sparse set of excited labelled-proteins can be localized using Gaussian PSF to form the final point cloud super-resolution image for the structure.

and compiling the fluorophore localizations from all the acquired frames, we obtain a high-resolution image. This is in contrast to diffraction limited fluorescent microscopy in which, due to the single-shot approach, the PSFs of molecules at distances below the diffraction limit overlap (Fig. 3) resulting in reduced resolution of the image.

*d*STORM is based on the use of standard fluorophores that are commercially available conjugated to a wide range of antibodies and therefore applicable to common immunofluorescent labeling of multiple cellular constituents. For dSTORM, fluorophores are induced to enter a weakly emissive or dark state by high powered laser illumination from which fluorophores will spontaneously return to the ground state and emit fluorescence [2]. Choice of fluorophore is based on photon output, as higher photon output improves localization accuracy, as well as the relative time the fluorophore is in the dark and bright states (duty cycle) and how many times the fluorophore can cycle between the dark and bright states (switching cycle). Fluorophore blinking is enhanced using buffers containing thiol reducing agents and oxygen scavengers. The dSTORM dye of choice is Alexa 647 that exhibits high levels of blinking and photon yield that are critical for analysis approaches described later in this review.

In addition to localizing the photon events of the excited fluorophores in the plane, i.e. X and Y coordinates, introducing a cylindrical lens in the light path of the imaging system will deform the PSF according to the depth (i.e. Z) of the molecule within the imaged sample. Arriving at the depth value of a single molecule involves fitting a multi-variate Gaussian PSF to the deformed shape of PSF [23]. Other 3D localization methods include biplane [30], PSF engineering as in (i) double-helix PSF [31], (ii) phase ramp [32], (iii) Zernike Optimized Localization Approach in 3D (ZOLA-3D) [33], Dual opposed objective interferometry as (i) iPALM [34], (ii) 4Pi detection scheme [35], and supercritical-angle fluorescence recovery [36].

The imaging process is repeated thousands of times using a Gaussian PSF fit to determine the localization of the individual molecules at high resolution. Every molecular location might have other information such as the localization uncertainty (fitting error of the Gaussian PSF), frame number, number of photons, etc. The final super-resolved image is formed by recombining all the localizations from all the imaged frames. A number of methods have been designed for this purpose, such as ThunderSTORM [37], QuickPALM [38], RapidSTORM [39], RainSTORM [40], etc. (see EPFL SMLM Software Directory [2]). As shown in Figure 2B, the localization of the molecules is obtained from every frame image individually enabling the user to map locations in 2D or 3D coordinate space. The set of molecular

[2]http://bigwww.epfl.ch/smlm/software/index.html

localizations and their associated meta-data are known as point clouds, events-list, pointillist, etc. The point clouds representation is used as input to the cluster analysis and quantification as shown in Figure 2C. We guide the reader's attention to many of the excellent references and reviews on super-resolution microscopy, especially the SMLM imaging techniques [2, 7, 13, 14, 22, 26]. In Figure 2A, we show 3D STORM imaging as an example of imaging a single molecule [23].

2.3 Imaging Artifacts

Quantification of super-resolution SMLM data might be biased due to some imaging artifacts. Some of the artifacts are challenging [41] and should be accounted for before analyzing the data. There are common pitfalls in super-resolution microscopy specimen preparation and imaging acquisition that should be avoided and optimized to ensure data reproducibility [42]. However, there are computational methods to address some of the artifacts and mitigate their effects to produce artifact-free super-resolution images. To know more about the super-resolution imaging artifacts and challenges that facing the cluster analysis and quantification, we guide the reader's attention to the following references [41–47]. We list here the main artifacts as they appear in recent papers [41–43].

- **Labelling errors.** Labeling of the protein of interest in SMLM is done primarily by expression of a photoactivatable FP directly linked to the protein of interest or via an antibody conjugated fluorescent organic dye by means of immunolabeling. The photo-activable FP, used in PALM, has a large size and may alter the localization and function of the protein of interest. For the immunolabeling approach adopted for STORM, the dye is conjugated to an antibody specific for the protein of interest. This method might include unspecific labelling and antibody specificity for the protein of interest should be validated [41]. In both labeling methods, the location of the imaged dye/FP can differ from the true location of the protein of interest and in a random direction. These localization errors create limitations to the quantification methods such that the protein clusters appear enlarged.
- **Detection efficiency.** Several methods have been proposed to quantify the percentage of proteins that are properly active. Not all the photoconvertible FP and molecules/fluorophores used in protein labelling are mature or successfully photoconvert. Hence, no algorithm can count the proteins that never appear [43].
- **Localization uncertainty.** Several methods [48] have been used to determine the position of the emitting molecule. The localization algorithms estimate the localization of the formed PSF of the fluorescent molecule. For example, in the Gaussian PSF model, the localization uncertainty is inversely proportional to the square root of the number of collected photons from the molecule.
- **Blinking.** The blinking artifact, also known as multiple blinking of a single fluorophore, is considered a serious artifact and has been studied extensively [45, 49–53]. Multiple blinking affects molecular counting and creates pseudo-clusters. For example, in PALM imaging, according to the four-state photokinetic model for photoswitchable fluorescent protein (Fig 4), once the fluorescent probe is activated, it can switch between non-fluorescent and fluorescent state before photobleaching irreversibly occurs [50], resulting in overcounting.
- **Drift.** Super-resolution SMLM images consist of thousands of stacked frames collected over time. Changes in temperature, the vibration of the microscope base, air current, etc. might cause sample drift in both lateral and axial directions [41]. Hence, the consequent drift can introduce localization (or spatial translation) errors by dozens of nanometers for different molecules relative to each other during the data acquisition.
- **Chromatic aberrations.** This occurs in multi-color imaging where light undergoes wavelength-dependent distortions. Motion artifacts (e.g. mechanical movements) and imperfections in the optical imaging system are the main sources of the chromatic aberration artifacts that affect co-localization of fluorophores of different colors [54]. The chromatic aberration correction is thus required on top of the drift.

2.4 Cluster Analysis and Quantification

Using point cloud representation for SMLM data analysis is not trivial. Point cloud representation is fundamentally different from the intensity grid valued pixel/voxel image representation used in conventional microscopy. Consequently, the computational tasks such as image processing, segmentation, registration, etc., applied for SMLM data analysis are different [55] from the ones applied for the conventional microscopy data analysis. Researchers have been working for decades to develop computational methods designed for conventional microscopy image analysis. However, these methods are not necessarily applicable to the point cloud data. Analysis of SMLM point cloud data is more complex [19, 55] and therefore requires new approaches. In particular, cluster analysis methods are most appropriate for analysis of super-resolution SMLM data point clouds generated by the localization methods that produce the data.

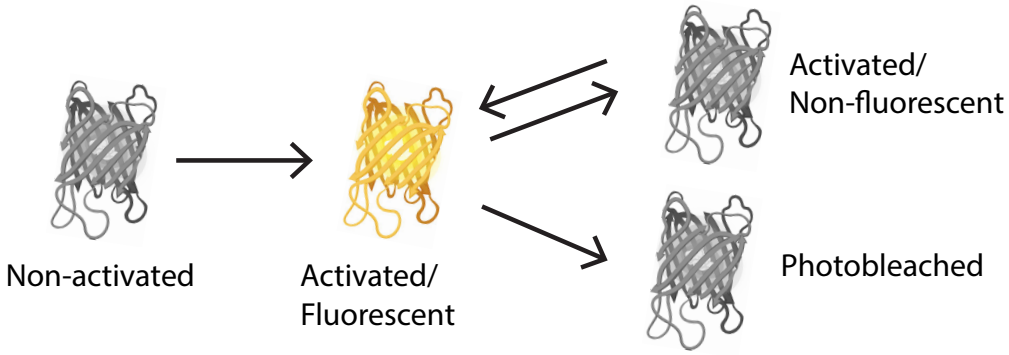

Figure 4: The four-state photokinetics model for photoswitchable fluorescent proteins. The image used in this illustration is adapted from Frick et al. [50].

The SMLM data analysis literature uses pre-processing and post-processing terms interchangeably. We believe that a clear distinction between the two terms should be made. Hence, in Figure 2C, we have included the methods used to correct for imaging artifacts in a pre-processing sub-box and the analysis methods used to quantify the biological clusters in a post-processing sub-box. For better quantification and interpretation of biological clusters, as opposed to artifactual pseudo-clusters, we believe that pre-processing should be applied first to obtain artifact free data. In this survey, we will focus on the post-processing methods used for SMLM cluster analysis as in Section 3. Specifically, this work focuses on reviewing the post-processing methods used to cluster and analyze point cloud SMLM data. However, some other methods, i.e. image-based cluster analysis, have been utilized to analyze the SMLM data and do not utilize the intrinsic pointillist properties of SMLM data. Examples of image-based cluster analysis methods which include extracting statistical measures are [56–60].

3 SMLM Cluster Analysis Methods

Protein-to-protein interactions produce heterogeneous and dynamic multi-molecular protein complexes [61]. The complex arrangements might consist of multiple molecules that vary in sizes; ranging from few to tens of nanometers. Studying protein cluster structure and organization is important to determine their function in the cell. Figure 5 shows how the protein molecules could cluster together in many ways to form more complex structures. In this section, we focus on the cluster analysis methods used to specifically understand the molecular clusters in super-resolution SMLM data.

Not all the clusters in SMLM data are related to biological structures. Some of the clusters in SMLM data are due to imaging artifacts (i.e. pseudo-clusters) caused by the uncertainty of the photophysical properties of the fluorescent reporters [27] as mentioned in Section 2.3 as well as labeling of the target molecule by more than one antibody probe. Pseudo-clusters could bias the quantification and the interpretation of detected molecular clusters. Hence, pre-processing analysis is required to correct for multi-blinking artifacts.

3.1 SMLM Molecular Localizations Clustering Task

SMLM data is a point cloud in 2D or 3D coordinate space where fluorophore events, or localizations, of the labelled target protein are output as an eventlist or pointillist $\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_L\}$, where L is the total number of localizations. Each $\mathbf{p}_i$ records the reconstructed spatial location of the event as a coordinate vector (x_i, y_i) in 2D or (x_i, y_i, z_i) in 3D. Additionally, meta-data associated with each fluorophore event are recorded, which can vary from SMLM imaging technique to another (e.g. STORM, PALM, GSD) or from microscope to another. Examples of fluorophore event $\mathbf{p}_i$ meta-data are $photon\text{-}count_i$, $frame\text{-}ID_i$, $error\text{-}in\text{-}x_i$, and $error\text{-}in\text{-}y_i$. This meta-data may be leveraged during the pre-processing of the eventlist. For example, $photon\text{-}count$ could be used as a confidence measure for an event where events with low photon counts are discarded. Following attempts to correct for imaging artifacts (e.g. diminishing the multiple blinking that causes pseudo-clusters and the offsets due to the length of antibody chain), one can obtain a reconstructed point cloud that provides estimates of molecular localizations. Each reconstructed molecular localization $\mathbf{m}_i$ is represented as a 2D or 3D spatial coordinate. Mathematically, we write $\mathbf{M} = \mathcal{R}(\mathbf{P})$, where $\mathbf{M}$ is a set of reconstructed molecular localizations, $\mathcal{R}$ represents the correction process, and $\mathbf{m}_i \in \mathbf{M}, i = 1, 2, \cdots, N$, where N is the total number of reconstructed molecular localizations (i.e. $N = |\mathbf{M}|, L = |\mathbf{P}|$). $\mathbf{m}_i$ has coordinate vector $(x_i, y_i) \in \mathbb{R}^2$ for a 2D point or $(x_i, y_i, z_i) \in \mathbb{R}^3$ in 3D. Notice that typically $N \leq L$.

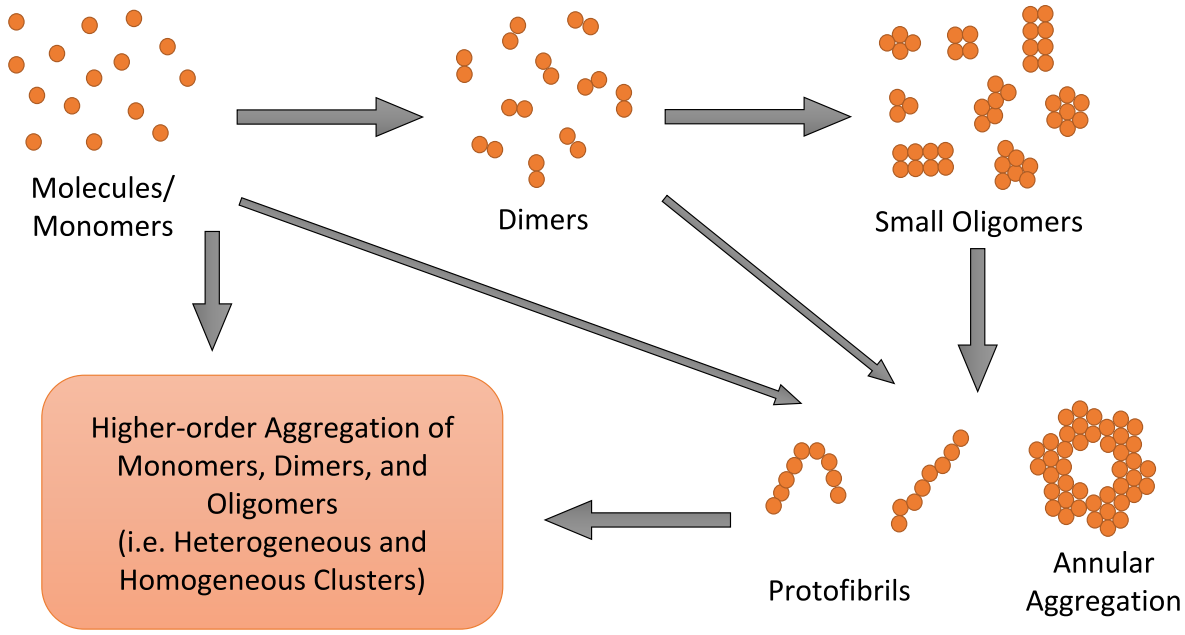

Figure 5: Illustration of how the protein molecules cluster together to form complexes. Monomers aggregate to form dimers which aggregate to form the small oligomers. Monomers could also cluster directly to form the large mutants and oligomers.

The molecular cluster analysis task is to apply various mathematical operators to find relations, patterns, curves (e.g. Ripley's H-function that shows a cluster of molecular localizations), etc. over the entire set $\mathbf{M}$ to detect, segment, classify, etc. the molecular clusters that represent the different protein complexes in the imaged SMLM data. Formally, a cluster of molecular localizations is a dense group of molecules that are, loosely speaking, closer in some way to each other than to the localizations outside the cluster. Clustering analysis is the task of decomposing a given point cloud of molecular localizations $\mathbf{M}$ into smaller disjoint (non-overlapping) subsets, such that their union covers the whole set $\mathbf{M}$. Each subset (i.e. cluster/blob b_i) implies some underlying macromolecular biological structure. Sometimes the process of isolating a given set of point clouds of molecular localizations $\mathbf{M}$ into smaller disjoint subsets is called segmentation.

Formally, let $\mathcal{C}$ represents the clustering operation applied to the set of localizations $\mathbf{M}$. $\mathcal{C}$ takes every localization in $\mathbf{M}$ and bins into one of K bins or clusters denoted $b_j, j = 1, 2, \cdots, K$. If the set of K clusters or bins is donated $\mathbf{B}$, and when $\mathbf{m}_i$ can only belong to one cluster (i.e. disjoint, non-intersecting bins), we write the mapping $\mathcal{C} : \mathbf{M} \rightarrow \mathbf{B}$ as $\mathcal{C}(\mathbf{m}_i) = b_j \in \mathbf{B}, \cup_j b_j = \mathbf{M},\ b_j \cap b_k = \phi, \forall j, k$, with $\cup$ and $\cap$ are union and intersection, respectively.

Sometimes, the cluster analysis is applied directly to the raw SMLM data (i.e. $\mathbf{P}$).

In this section, we will dig deeper into the cluster analysis methods used to quantify the molecular clusters (i.e. post-processing) obtained through super-resolution SMLM. The methods include statistical, Bayesian, density-based, Voronoi tessellation-based, and graph-based approaches.

3.2 Statistical Methods

Over the past few years, researchers have started to apply statistical methods that are based on second order statistics and spatial point analysis methods to quantify the SMLM clusters. The statistical methods have been applied to ecological spatial data and adopted for SMLM analysis. In this section, we will cover the main statistical methods used in the literature to analyze super-resolution SMLM data as listed in Table 2. Given its large popularity in analyzing SMLM data, we will describe Ripley's functions next as well as some of its variants. We will then describe a class of statistical methods that are based on correlation techniques.

3.2.1 Ripley's Functions

Ripley's K-, H-, and L-functions are gaining popularity in cluster analysis of the SMLM membrane proteins. These functions are used increasingly due to the point cloud nature of SMLM data (localization of molecules). Ripley [62] studied the stochastic models that have been proposed for spatial point patterns. Ripley's K-function is a tool used

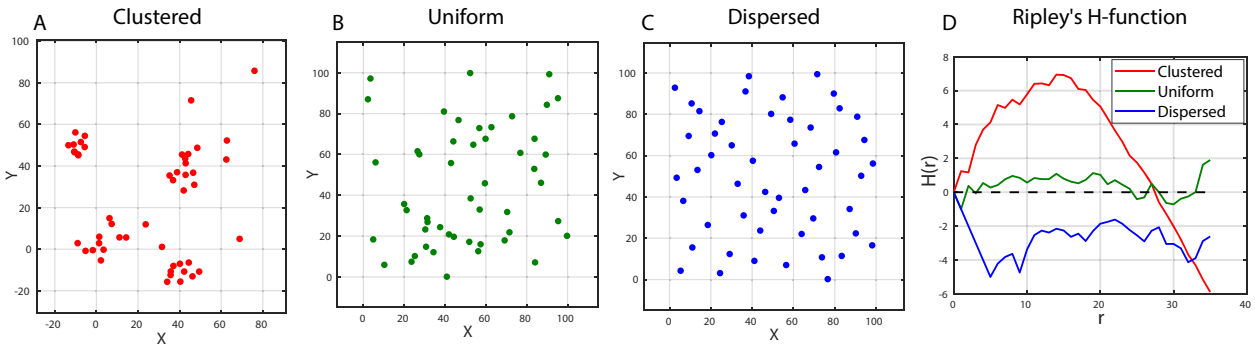

Figure 6: An example that shows the behavior of the $H(r)$ function for the different distributions of spatial point patterns. The $H(r)$ function (D) has positive values for clustered points (A), fluctuates around 0 for uniformly distributed (random) points (B), and has negative values for dispersed points (C). The generated data consists of 50 points for each one of the patterns shown in (A–C).

for analyzing spatial point process data [62, 63]. It is usually used for analyzing 2D data, but may be used to analyze locations along a line (1D) or may be extended to 3D spatial data [63, 64].

The density of points in an area (number of points per unit area) is known as the first moment property [65]. The second moment property (also known as a bivariate or multivariate generalization) used to describe the relationships between two or more point patterns by finding the expected number of points N within a distance r of another point [63, 65]. Ripley's K-function [62] is a second moment property (second-order statistics). Theoretically, the K-function is given in equation 2 [63].

$$K(r) = \lambda^{-1} E[number\ of\ points\ within \\ distance\ r\ of\ randomly\ chosen\ point], \tag{2}$$

where λ is the density normalization of points (number of points per area A, $\lambda = N/A$). Formally, the K-function is given in equation 3 [62, 65].

$$K(r) = \frac{1}{n} \sum_{i=1}^{n} N_{p_i}(r)/\lambda, \tag{3}$$

where p_i is the i-th point in the data and the sum is taken over n points. For homogeneous **Poisson process**, which is know as *complete spatial randomness* (CSR), the expected value of $K(r)$ function is given in the equation 4 [63].

$$K(r) = \pi r^2, \tag{4}$$

notice that deviation from the CSR expected value indicates scales of clustering and dispersion. So, Ripley's K-function is typically used to find the level of clustering by comparing a given distribution with a random distribution.

Other Ripley's functions can be derived from the K-function. The complete derivation for all the other functions can be found in these papers [63, 65]. The L-function was proposed by Besag [66] as a normalization for K-function as seen in equation 5.

$$L(r) = \sqrt{K(r)/\pi} \tag{5}$$

The L-function and it's derivative can be used to identify the radius of the clusters [65]. Normalizing L-function will produce another function called H-function [67]. Hence, H-function is a further normalization of the original K-function. The H-function is given in equation 6.

$$H(r) = L(r) - r \tag{6}$$

Notice that H-function may result in a positive value which indicates clustering over the spatial scale, on the other hand, the negative value indicates dispersion. The value is zero when we have CSR distributed points (not clustered nor dispersed points). This is because for a CSR distribution, $L(r) = r$ for all values of r. Figure 6 shows three cases of spatial point patterns and the corresponding Ripley's H-function behavior. The pattern of the H-function fluctuates around zero for uniformly distributed points, above zero for clustered points and below zero for dispersed points.

To estimate $K(r)$, the numerator of equation 3 can be written as $N^{-1} \sum_i \sum_{j \neq i} I(d_{ij} < r)$, [63, 68]. Where, d_{ij} is the distance between the i-th and j-th points. $I(.)$, is an indicator function that is equal to 1 if $d_{ij} \leq r$ and is *zero* otherwise.

In its current formulation, K-function does not consider the effect of the points close to the border of the study area. This issue, which is called the edge effect of Ripley's K-function, causes underestimation of K [63, 68–70]. Hence, Ripley's

K-function requires more elaborate methods for edge correction. Many methods have been proposed to correct the edge effect of the Ripley's K-function [69–72]. Generally, the corrected K-function ($\hat{K}(r)$) can be written as equation 7.

$$\hat{K}(r) = \frac{A}{N} \sum_{i} \sum_{j \neq i} \frac{I(d_{ij} < r)}{w_{ij}}, \tag{7}$$

where w_{ij} is a weight function that provides the edge correction.

Ripley's functions are becoming increasingly popular in analyzing SMLM data. The functions have been utilized in many biological applications to find the level of molecular clustering. They are either used alone or in combination with the other cluster analysis methods [55, 73–97]. We summarize how the methods are adopted for SMLM cluster analysis of the different biological applications in Table 2.

3.2.2 Getis and Franklin's (G&F) Local Point Pattern

Getis and Franklin published a paper and proposed a new $L(r)$ function. Their function is a variant of Ripley's K-function, called second-order neighborhood analysis. The goal of their $L(r)$ function is to quantify the clustering of the points (molecules in SMLM context) at various spatial scales [98]. The values of $L(r)$ function are calculated for each point as described in equation 8.

$$L(r)_j = \sqrt{A \sum_{i=1}^{n} \frac{\delta_{ij}}{n-1} / \pi}, \tag{8}$$

where A is the region area (e.g. rectangular region) under study, n is the total number of points in the region. The indicator function delta $\delta_{ij} = 1$ if the distance between point i and point j is $< r$ and 0 otherwise. $\sum_{i=1}^{n} \delta_{ij}$ is the summation over all points within distance r from point j (i.e. all points within a circle of radius r centered at localization j). Thus, $L(r)$ is another way of normalizing Ripley's K-function by finding the local point patterns normalized by the average point density in the whole analyzed region. The property of $L(r) = r$ for all values of r when having a CSR distribution, still holds.

G&F function could be used to find the local descriptor (patterns) of every localization, while Ripley's function is used to describe all the points in the region globally. Therefore, the G&F function has been used in combination with Ripley's K-function to analyze the localization of molecules from SMLM data [75, 80, 83, 85, 87–93, 99]. For example, G&F point pattern was used in double protein-labelling analysis [99] to investigate the co-clustering of membrane proteins. It was also used for generating a topographic map of the level of clustering to determine the heights of peaks in the map across a region [90] and then, using the relative heights of the peaks to determine the clustering characteristics and avoiding the inaccurate thresholding.

3.2.3 Correlation-based Methods

Correlation-based analysis methods (pair correlation, auto-correlation, cross-correlation, co-localization, etc.) have been applied to super-resolution SMLM data for both pre-processing and post-processing quantification (Fig. 2C). Pre-processing methods [53, 100–102] address imaging artifacts such as the multiple blinking of a single fluorophore that may cause molecular overcounting. Overcounting, as well as the other implications of imaging artifacts, might bias SMLM cluster analysis and should be corrected before post-processing the SMLM data for quantifying the biological clusters. For example, Malkusch et al. [101] used a correlation coefficient framework, coordinate-based co-localization, to analyze every single localization within a certain radial distance and assign to it a score ranging from -1 to 1. In their formulation, -1 is assigned to perfectly segregated localizations, 0 for uncorrelated (randomly distributed), and $+1$ for perfect co-localization. The same coordinate-based co-localization idea has also been used to analyze the protein localizations of biological clusters [82, 88, 96]. Many other correlation functions have been used to quantify the localization clustering for post-processing SMLM data [82, 88, 96, 99, 102–104]. The proposed correlation functions used to analyze biological clusters rather than the biologically irrelevant pseudo-clusters (also known as nano-clusters). For example, Schnitzbauer et al. [102] derived a cross-correlation function for the localization coordinates inspired by the translational cross-correlation function in pixel-based image representation. They mathematically showed that the point-to-point distance distribution in super-resolution SMLM is equivalent to the pixel-based correlation function. Then, they extended cross-correlation to quantify spatial relationship between complicated structures by considering the point-to-set distance.

3.3 Bayesian Methods

The SMLM cluster analysis methods usually depend on a set of user-defined parameters. Sometimes, the subjectivity and the ambiguity of selecting the parameters affect the performance of the SMLM clustering task. The main goal

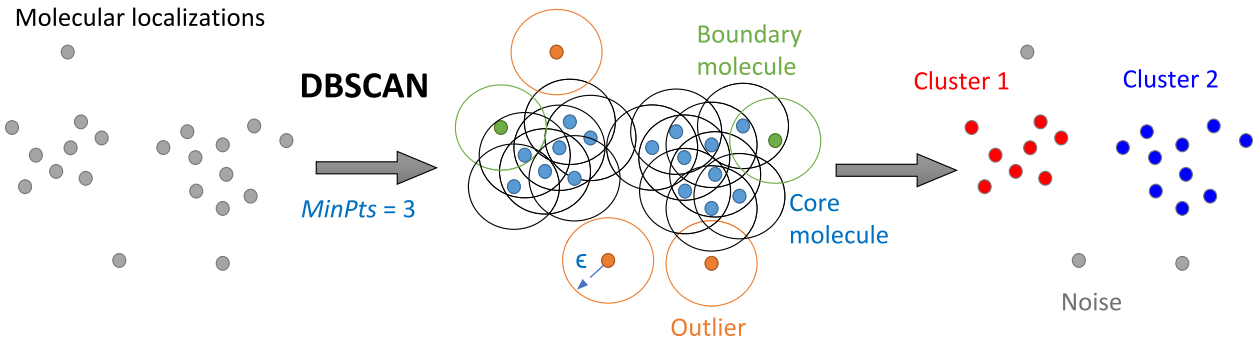

Figure 7: An example that illustrates the density-based DBSCAN clustering method applied to SMLM data. For instance, DBSCAN algorithm is applied when using ϵ and $MinPts = 3$ parameters. Sometimes, the subjectivity of selecting the parameters might change the clustering results dramatically. For example, pseudo-clusters in SMLM complicate the selection of the algorithm parameters.

of the Bayesian approach for super-resolution SMLM data is to design a clustering method that alleviates the need for arbitrary user-selected analysis parameters. A well defined Bayesian prior will replace the arbitrary user-selected parameters. Bayesian is a model-based approach which is used for spatial point clustering generated by SMLM. The model is used to evaluate the assignment of every molecule to clusters by its marginal posterior probability. The posterior probability is computed based on a specified model for the molecular data and their uncertainties [78]. So, the mechanism is to select clusters from a set of generated clustering proposals. Usually, clustering proposals are generated with variable spatial scale and threshold using statistical methods such as Ripley's K-function [83] or using G&F function [78]. Then, after generating thousands of candidate proposals per region of interest (ROI), the optimum number of proposals are selected by scoring them against the Bayesian model [55, 78, 83, 105]. For example, in the Griffié et al. [83] generative model, they considered an ROI containing clustered and non-clustered localizations. The user set the probability that localization is non-clustered and this is the prior parameter for the model. They also have an assumption that the molecular positions in the cluster are following spherical Gaussian distribution. The radius (Gaussian standard deviation) of the cluster is drawn from a user-specified histogram of sizes. They claim that the aforementioned model reflects the *a priori* knowledge of the molecular distribution. The Bayesian approach is not limited to quantifying 2D and it was extended by Griffié et al. to analyzing 3D SMLM data [83].

3.4 Density-based Methods

Density-based clustering methods are popular in data mining and spatial data clustering. Ester et al. proposed density-based spatial clustering of applications with noise (DBSCAN) [106], a density-based clustering algorithm that is capable of discovering clusters of arbitrary shapes. It could be used to filter out noisy events from the SMLM data when its parameters are set correctly.

DBSCAN is based on two parameters for detecting and segmenting the clusters in SMLM data. It requires a neighborhood radius ϵ and the minimum number of localizations/points ($MinPts$) within ϵ to qualify as a cluster (Fig. 7). The algorithm can start from any molecular localization that has not been visited. The connectivity of the molecules of the qualified clusters should maintain the $MinPts$ condition within ϵ while propagating from one molecule to another within the same cluster until reaching the boundary molecules, where the $MinPts$ condition no longer holds. Otherwise, the cluster is considered as an outlier. It is clear that the clustering is conditioned on the minimum density of molecules within neighborhood radius ϵ. Figure 7 shows an example of how the DBSCAN clustering method works and the required parameters to cluster the localizations.

According to Mazouchi et al. [107], leveraging DBSCAN to analyze super-resolution SMLM data has certain limitations. The algorithm is slow scaling with the number of localizations and it has $O(n\ log(n))$ at best. The ambiguity and subjectivity in selecting the algorithm parameters affect its performance and makes the algorithm general. The imaging artifacts and the multiple blinking of a single fluorophore cause the formation of pseudo-clusters. DBSCAN, may not be sufficient to differentiate between clusters of localizations due to protein (i.e. biological clusters) and the non-biologically relevant pseudo-clusters [107]. To address the limitations of DBSCAN, Mazouchi et al. [107] proposes a density-based clustering algorithm, fast optimized cluster algorithm for localizations (FOCAL). FOCAL is a grid-based method optimized for fast analysis of SMLM data. It has one parameter that needs to be optimized, density threshold ($minL$). It has a linear time complexity ($O(n)$). FOCAL has limitations in dealing with small clusters and requires

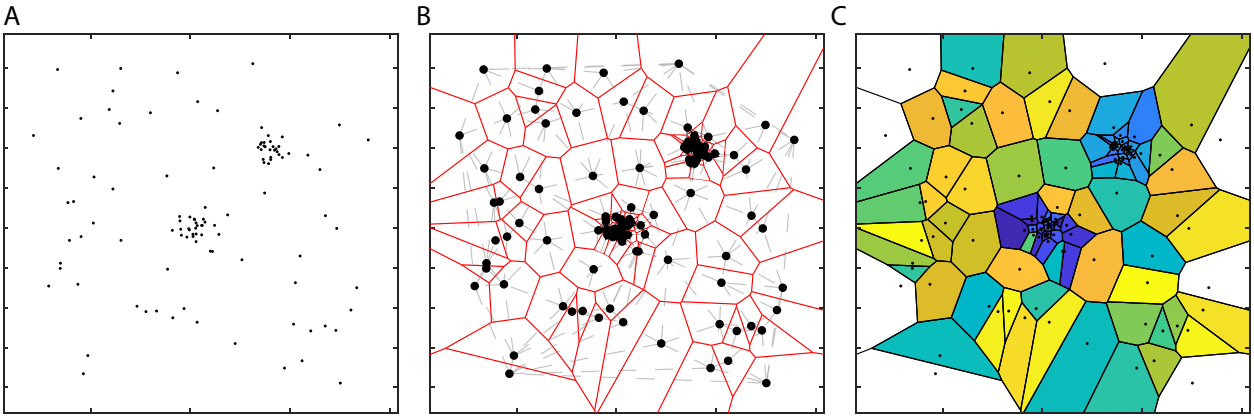

Figure 8: Voronoi tessellation-based method used to segment the clustered SMLM molecular localizations. (A) The input space of molecular localization. It has two clusters and noisy/background localizations. (B) The Voronoi tessellation and partitioning the space into polygonal regions (Voronoi cells) in red. The Delaunay triangulation (dual of Voronoi) is shown in grey dashed connections. (C) The Voronoi cells colored with different colors. The white regions are the Voronoi cells with open regions.

setting fine grids. The problem becomes more severe with high levels of noise. Moreover, FOCAL has issues with high-density SMLM data and the overlapped clusters. FOCAL3D [108] is an extension for the FOCAL method that is capable of analyzing 3-dimensional SMLM data.

In SMLM data cluster analysis, DBSCAN is used either alone [107, 109–111] or in combination with other clustering algorithms [82, 95, 96, 102, 112, 113] to quantify the SMLM clusters.

3.5 Voronoi Tessellation-based Methods

Voronoi diagram or tessellation of point clouds has been used in many applications for various goals, including computational geometry, computational physics, astrophysics, computational chemistry, biology, etc. In SMLM, a Voronoi diagram is a method to partition the input space of molecular localizations into regions according to the Euclidean distance between the seed points (i.e. molecules). The resultant polygonal regions are called Voronoi cells, where each cell is centered around one of the molecules. Figure 8A shows a set of points in 2D space, where each point might represent a molecular localization. Figure 8B depicts the Voronoi diagram for the molecular localizations shown in (A). The Voronoi cells are shown in different colors in Figure 8C. Notice that the Voronoi edges are equidistant from the two nearest molecules. That is, the projected perpendicular line from every molecule to any one of its Voronoi cell edges is the shortest distance between every neighboring pair of molecules. Hence, there is no intersection between any Voronoi cells. To learn more about the methods used to find the Voronoi polygons, we guide the reader's attention to the work of Okabe et al. [114].

Segmenting SMLM molecular clusters using geometric properties (e.g. area, shape) of the Voronoi cell was done by [115, 116]. The Voronoi geometric characteristics could be used to describe the neighborhood of the molecules. The Voronoi geometric characteristics for every molecule might be different based on the density and data organization. Both works [115, 116] depended on the Voronoi cell area to segment the SMLM molecular clusters, where the Voronoi cell area is inversely proportional with the density of the molecules (regions with high molecular densities are composed of Voronoi cells with smaller areas). Both works designed a method to cluster SMLM localizations by comparing their Voronoi cell areas with a reference distribution chosen to be either a spatially uniform [115] or CSR distribution [116]. Figure 8 explains the concept of Voronoi diagram method when applied to point cloud data. Notice that the clustered points (Fig. 8A) have Voronoi cell areas (Fig. 8C) that are smaller than the non-clustered points.

Utilizing Voronoi tessellation for clustering molecular localization has been applied to structures with various shapes, such as tubular shaped structures (e.g. microtubules, filamentous, fibrous) [115, 116]. Voronoi tessellation was also adopted in other super-resolution SMLM cluster analysis applications [84, 86, 97].

Delaunay triangulation is the dual of the Voronoi tessellation (Fig. 8B). It has also been used for analyzing and quantifying super-resolution SMLM data [86, 117]. However, Voronoi tessellation methods are different from the Delaunay triangulation in the former's ability to provide a direct estimation of the region of influence [118] and hence is more preferable for analyzing SMLM data. Recently, a 3D extension of the Voronoi tessellation method has been

proposed by Andronov et al. [118]. They used Voronoi volumes as a characteristic of 3D Voronoi cells to segment 3D SMLM clusters. Andronov et al. [118] claim that the Voronoi based methods are able to handle the edge effect (see Section 3.2.1). This is because the border molecules have larger or infinite Voronoi cell areas which prevent them from contributing to the clustering.

Clustering molecular localizations is therefore based on the geometrical properties of the Voronoi cells such as the cell area. Some molecules are considered as part of a cluster based on their individual Voronoi cell areas. This leads to crude segmentation for the clusters. Moreover, some of the border molecules might be excluded from being part of a cluster as they might have very large Voronoi cell areas (e.g. white cells in Figure 8C) compared with inner molecules. In addition, the Voronoi-tessellation based methods might fail in extracting the true molecular clusters from SMLM data with multiple blinking of single fluorophore artifact. Leveraging Voronoi cell area for segmenting clusters from SMLM data with varying cluster densities might be another problem in such methods because the cell areas hugely depend on the underlying molecular densities and the closeness of the nearby clusters.

3.6 Graph-based Methods

Graphs are strong mathematical structures employed to model the interaction between objects or entities of a system. The entities are represented as graph nodes and their interactions are represented as edges [119]. Hence, the graphs are considered powerful and rich data structures that encode the connectivity relationships between the different entities of a system. In real-world problems, graphs are frequently complex networks because they have many properties that make them different than other types of graphs such as random graphs [120]. For example, real-world networks have many subgraphs, modules, patterns, small-worldness, etc. that are not frequent in other types of graphs. Networks are ubiquitous and they are used to study and model many real-world problems effectively. Recently, network analysis methods have been successfully adopted in many fileds of study such as the brain, social, computer, road, metabolic, Internet, etc. [121–125].

Leveraging graph theory to analyze SMLM data is infrequent in the surveyed literature. Few recent works have been proposed to utilize complex networks and graphs for cluster analysis of super-resolution SMLM data. Various neighborhood networks/graphs could be constructed from the spatial SMLM data. Figure 9 shows two types of neighborhood graphs that can be adopted for analyzing the SMLM data.

Networks have been adopted to analyze large data sets of prostate cancer data as well as cardiac data from various SMLM modalities [126, 127]. A graph-based network method has been proposed by Khater et al. to model SMLM data [126], where the nodes represent the molecules and the molecular interactions are represented as edges connecting the nodes. Various network features have been leveraged to pre-process the SMLM data (e.g. correcting for multiple blinking of a single fluorophore artifact) as well as post-processing (e.g denoising the SMLM data and extracting its constituent molecular clusters). A combination of network analysis and the mean-shift algorithm are leveraged to segment the clusters and obtain 3D representation of molecular localization of diffraction limited cellular structures, in this case plasma membrane invagination called caveolae. Khater et al. [128] also proposed a graph-based method to extract graphlet features form SMLM molecular clusters for automatic identification and quantification of various biological structures. Network community detection and modularity analysis have been proposed to decipher the architecture of the molecular clusters [129]. Communities represent sub-clusters of molecules within a larger cluster. An example of modular detection within caveolae is shown in Figure 10.

Some other very recent graph-based methods to analyze and quantify SMLM data are posted as preprints and still unpublished yet. We cover them briefly in this survey and categorize them under the graph-based methods. Researchers are exploring new computational methods to analyze the SMLM data. For example, community detection has also been exploited for extracting SMLM clusters [130]. A segmentation protocol based on persistence homology and DBSCAN has been employed to segment and quantify the topological structure within SMLM data [131]. In this persistence homology method, the density modes were constructed from a graph that connecting all the localizations within the same search radius.

3.7 Machine Learning-based Methods

Machine learning algorithms (including deep learning) are data-driven approaches. Deep learning approaches typically require relatively large data that could capture the variations in the dataset. Supervised machine learning approaches require ground truth data for training the different learning models, which is hard to obtain in SMLM data. Machine learning models can be trained to perform various computational tasks such as predicting, segmenting, classifying, etc. of the molecular complexes.

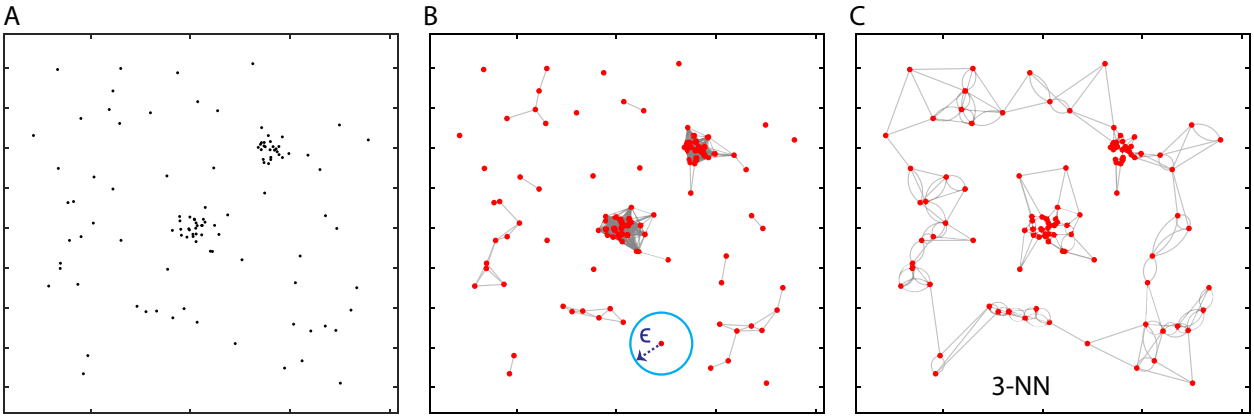

Figure 9: Network/Graph-based method used to model the SMLM molecular localizations for cluster analysis. (A) The input space of molecular localization. It has two clusters and noisy/background localizations. (B) The ϵ-graph used to construct the network, where every node is connected to all the other nodes within the proximity distance ϵ. (C) The kNN graph used to construct the network, where every node is connected to only the k closest neighboring nodes. We constructed the 3-NN graph for illustration.

We have not witnessed a large amount of work leveraging machine learning for super-resolution SMLM data for the aforementioned computational tasks. However, some works have applied deep learning for the localization and data acquisition [132–134] but not for cluster analysis.

Khater et al. [126, 128] designed a graph-based machine learning method to automatically identify the class of the molecular complexes from super-resolution SMLM data. They leveraged machine learning for many computational tasks such as determining the scale of clustering, finding the biosignatures for several biological structures, and identifying patterns of the isolated and multiple antibody proteins. They also leveraged deep learning for the biological structures classification task [135] applied to several SMLM data representations. Lately, Sieben et al. [110] used machine learning to identify the class of the biological structures from SMLM data. Another recent work that utilized machine learning to detect clustered and unclustered (background) molecules was proposed by Tobin et al. [103]. Williamson et al. [136] proposed a supervised machine learning method that is capable to classifying all the localizations from microscopy datasets into clustered or not-clustered classes. Their model is trained on several simulated clustered datasets.

3.8 Validation

Validation of the cluster analysis method is critical when applying the various algorithms to find the biological clusters from SMLM data. There is no publicly available dataset with ground truth class labels for the membership of the localizations to the various cluster types. Hence, most of the methods are unsupervised approaches, where the ground truth is not provided along with the data. The ground truth might include information such as the number of clusters and their features (e.g. sizes, densities). Also, comparing the different clustering methods requires having benchmark SMLM data with known cluster features. We summarize the main methods used to validate the super-resolution SMLM clustering methods in the following subsections.

3.8.1 Computer Simulations (*in silico*)

Synthetic data
Generating synthetic data with known cluster features (density, size/volume, shape, etc.) has been widely used to mimic SMLM data. Background and noise signals with known distributions are also generated along with the synthetic clusters. Some methods generate synthetic data that is based on specific assumptions, such as generating Gaussian clusters and the minimum distance between the generated clusters should be greater than some threshold value [65, 118].

Given synthetic data, a clustering method is tested on extracting the synthetic clusters first. Then, it could be used for cluster analysis in experimental SMLM data. To assess the quality of a clustering method, the extracted clusters and their features are compared with the known clusters used in the data generation.

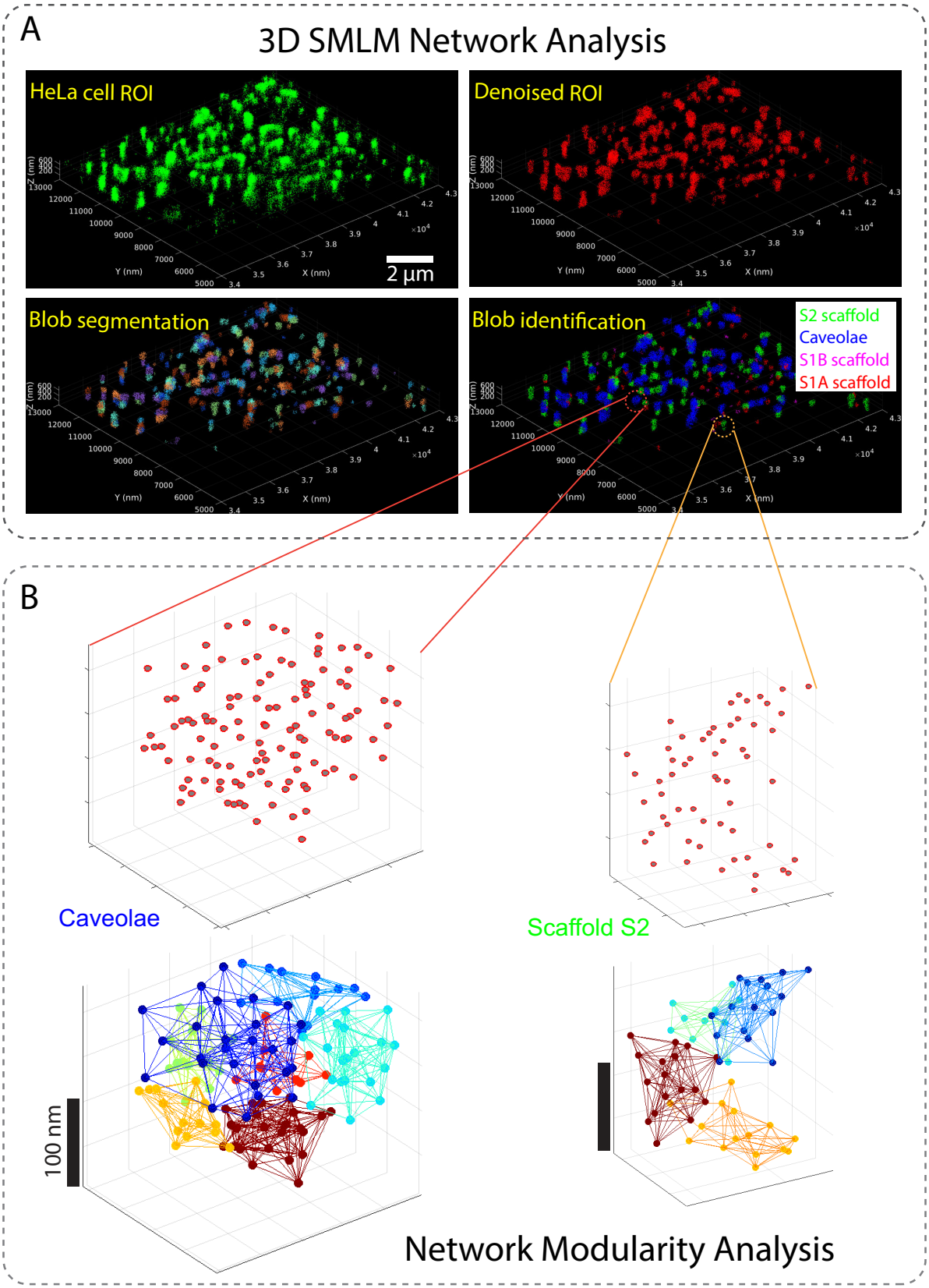

Figure 10: Graph-based network analysis methods for SMLM data proposed by Khater et al. [126, 129]. (A) They proposed *3D SMLM Network Analysis* pipeline [126] to correct for multiple blinking of a single fluorophore, filter out noisy localizations, segment the biological structures into clusters/blobs, and identify the cluster/blob classes. (B) Network community/modularity analysis [129] detecting the modules within caveola and S2 scaffold domains.

Simulated data
The simulation could be used to mimic super-resolution SMLM imaging for known biological structures. The simulators have the ability to imitate the SMLM imaging by varying several parameters (labelling strategy, labelling efficiency, epitope length, number of frames, imaging time, density, background, etc.) which might be useful for optimizing the imaging of the experimental sample. Hence, simulation gives more control to study all the possible scenarios that might lead to less imaging artifacts in the data. Moreover, the simulation could help in assessing the quality of the adopted clustering analysis methods.

Recently, many SMLM simulators have been developed and posted as publically available software tools. The simulators facilitate generating the data for using it in cluster analysis applied to various biological structures. Some popular SMLM simulators software are SuReSim [137], TestSTORM [138], and SMeagol [139]. Synthetic data generation offers controlled creation of clusters and background with known density distributions (e.g. Gaussian, uniform) and cluster shapes (e.g. circular, tubular). For example, Levet et al. [140] generated synthetic 2-color 2D/3D clusters of circular and square shapes. They also simulated multiple scenarios by varying the number of clusters, their relative positions, their diameters, their density ratios, and background/noise levels.

Data simulators, on the other hand, are designed to mimic realistic labeling and imaging conditions [46,47]. SMLM data simulation considers the inner-workings of the SMLM imaging technique and labelling parameters (e.g. epitope length, labeling efficiency, localization precision, number of frames, blinking events per frame, etc.) in the data generation, but does not give direct control of the resulting data. For example, Spahn et al. [141] used the SuReSim simulator [137] with specific SMLM imaging parameters (e.g. cluster diameter of 100 nm and and various epitope number per cluster) to generate an image, with a field of view of $15 \times 15\ \mu m^2$, of some biological structures (e.g. clathrin-coated pits). Sieben et al. [110] also used simulation to validate their work. They mimicked their real experimental conditions to generate ground truth models by controlling the labelling efficiency, localization positions, noise molecules, fluorophore parameters (e.g. distributions for photon count, localization precision), etc. To evaluate their ERGO emitter density estimation method, Cardoen et al. [134] used the *in silico* sequence of 2500 frames (each a 64×64 grayscale pixels corresponding to a 2D view of $100 \times 100\ nm^2$) from [48], which simulates a realistic acquisition of microtubules labelled with the commonly used Alexa 647 fluorophore. They verified their approach on real world data [142] with a markedly different microscope configuration where they showed that aligning the intensity distribution between training and real world data is sufficient to obtain consistent results without retraining.

3.8.2 Validation via Physical Phantoms

DNA-origami and nanorulers
DNA origami and nanorulers have been developed to validate many of the SMLM imaging and analysis methods. They are used in super-resolution imaging and microscope calibration [47]. The DNA origami is designed to allow placing of a known number of fluorescent molecules to nanostructures in defined geometries [143]. Also, DNA origami has been used to quantify the protein copy number in the cells using super-resolution microscopy [144].

3.8.3 Validation via Knowledge of Biology and Other Imaging Modalities

Real experimental super-resolution SMLM data could be used in clustering methods validation if the studied clusters of biological complexes have been studied before with other imaging modalities. Biological structures, imaged using Electron Microscopy (EM), with known size and number of molecules could be used as ground truths for super-resolution cluster analysis methods. Generally, researchers use simulation or synthetic data to validate their methods then, apply their methods to real experimental data. For example, Sieben et al. [110, 145] used EM imaging to validate their multicolor 3D SMLM reconstruction and analysis method. They used dual-color SMLM to image around 300 centrioles per field of view. Then, they used masking to segment the localizations and DBSCAN to separate adjacent particles. The 3D volumes were reconstructed by EM routines and classified by applying 2D clustering. Khater et al. [126, 129] used known information about the cell surface invagination, caveolae, to validate their work. For example, they compared their findings with known topology, size, and number of predicted proteins per segmented structure.

4 Summary and Discussion

In this paper, we surveyed the state-of-the-art cluster analysis and quantification works applied to super-resolution SMLM. We depended on various criteria to study the papers and tabulate them in Table 2 according to (i) the biological application of the study; (ii) the data acquisition; (iii) and the data analysis technique adopted. Then, we categorized the different clustering methods for easy reference and comparison and identified the pros and cons of these categories in Table 1. Looking at the various methods/algorithms listed in Table 2, we note the following:

- **2D/3D analysis:** Some algorithms have been used only for 2D super-resolution SMLM data analysis, while some other algorithms were used for 2D and then extended to 3D. Dealing with 3D SMLM data is challenging as the axial resolution is worse than the lateral resolution. Also, some biological structures depict structural properties evident in 3D (e.g. hollow structures) so the analysis methods should be designed with care to handle such 3D structures in the denoising, clustering, and identification stages.
- **Pre-processing:** Few methods could effectively handle some of the imaging artifacts, such as the multiple blinking of a single fluorophore artifact (e.g. graph-based, statistical methods), while some other methods could not (e.g. Voronoi tessellation-based, density-based methods).
- **Localization uncertainties:** Few methods utilized the localization uncertainties (e.g. Bayesian methods) in the analysis, while the majority of the methods did not.
- **Parameterization:** The majority of the methods have parameters, while the Bayesian methods are claimed to be parameter-free models. However, Bayesian methods are relatively much slower (e.g. Griffié et al. [105] reported that the processing time for one dataset consisting of 30 small 2D ROIs is $\sim$19 hours with user input). Voronoi tessellation-based clustering is parameter-free method if the segmentation threshold is determined by Monte-Carlo simulations [116].
- **Intra-cluster analysis:** The intra-cluster features (features of molecular interaction within a cluster and its sub-clusters such as network analysis of the molecules, modularity analysis, sub-networks, etc.) lead to understanding the architecture of the biological complexes. Very few methods are equipped with capabilities to extract the intra-cluster features (e.g. graph-based), while the majority of the methods do not have this capability.
- **Machine learning integration:** Most of the methods are not equipped to be integrated with machine learning approaches for further analysis. Machine learning approaches require associating features with samples/clusters to train a model. Subtle features could be extracted from graph-based methods (e.g. network measures) and used to train a machine learning model. Also, some recent deep learning approaches for graphs [146, 147] could be leveraged for analyzing SMLM data.
- **Big-data analysis:** The majority of the methods do not scale well to handle the big data generated from the super-resolution SMLM imaging techniques, while some methods are highly efficient and scale-up efficiently with big-data (e.g. graph-based and Voronoi tessellation-based methods).
- **Cluster shape variation:** Some of the methods could discover the clusters with various shapes (e.g. density-based methods). Some methods are more suitable for identifying tubular-like shapes (e.g. Voronoi tessellation-based methods).
- **FOV/ROI analysis:** The majority of the surveyed methods were used to analyze small regions of interest (ROIs) rather than the whole field of view (FOV). Also, in most methods, the ROIs were either selected manually or randomly from the whole FOV. We believe selecting a small ROI is not a good strategy and bias the cluster analysis. Analyzing ROI is dependent on its location in the cell. For example, selecting an ROI very close to the periphery of the cell could have structures that are different than those in an ROI in the middle of the cell. Because the structures at the periphery might have different functions (e.g. focal adhesion) than the structures in the middle of the cell.
- **Software:** Some published software is designed to visualize SMLM data with very limited analysis capabilities such as ViSP [148]. Some published software is limited to analyzing 2D regions of the SMLM data, is unstable, cannot handle the whole FOV but is limited to small ROIs, and is not robust to noise. Some software packages implement more than one method for analyzing SMLM data. In general, we noticed limited work on automatic quantification and analysis methods applied to super-resolution SMLM data.
- **2D methods for 3D data:** We noticed that some methods acquire 3D super-resolution SMLM images, then, in order to leverage existing 2D super-resolution SMLM cluster analysis methods, they project the 3D data to 2D [81, 86]. Projecting 3D data to 2D for analysis is not a good idea. Processing data in its 3D native format is much better to (i) avoid artifacts, and (ii) filter the noisy background localizations. About 78% of the super-resolution SMLM cluster analysis methods are 2D as shown in Figure 11B. Around 49% of the SMLM imaging used STORM-based techniques as depicted in Figure 11A.
- **Validation:** Various ways to validate the methods were used, such as using DNA origami and nanorulers, synthetic data generation, SMLM simulators, and finally, using real experimental cellular data with known biosignatures. Validation and evaluation of the different methods remains a challenging task for all surveyed methods as no public dataset is available for benchmarking and assessing the performance of different (post-localization) analysis methods.

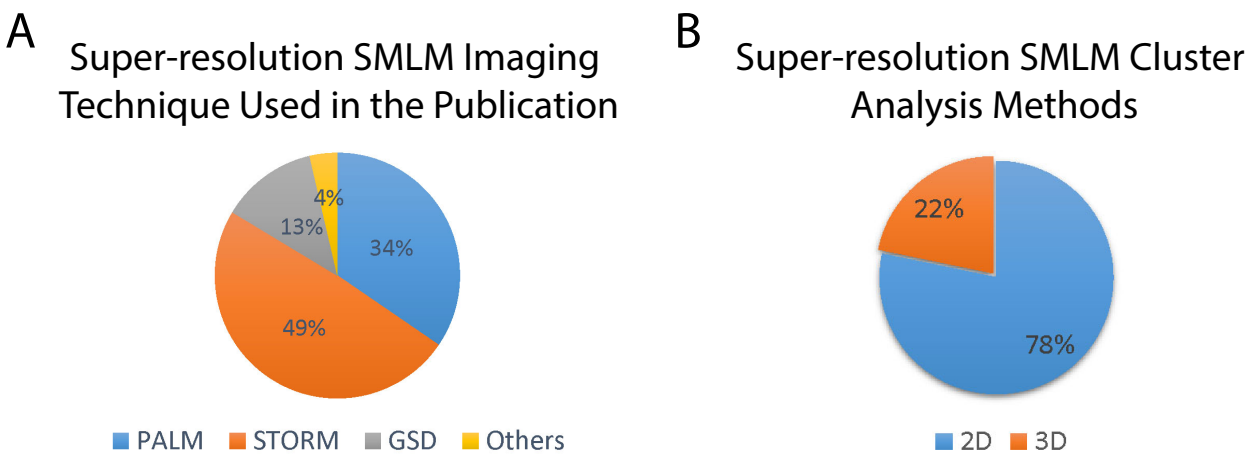

Figure 11: (A) The distribution of the publications based on the super-resolution SMLM imaging technique used in the study. (B) The distribution of the publications based on the dimensionality of the super-resolution SMLM cluster analysis method.

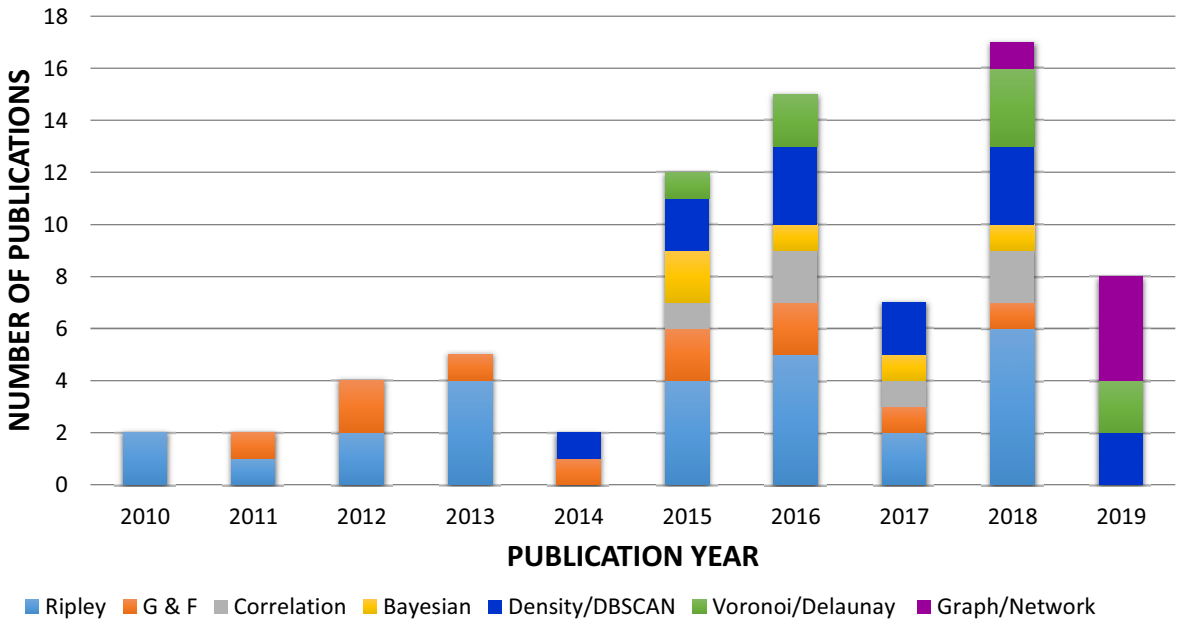

Figure 12: Number of of publications per year (starting from 2010) categorized according to super-resolution SMLM cluster analysis methods used in the study.

We summarized the number of publications (listed in Table 2) per year categorized according to super-resolution SMLM cluster analysis methods used in the study in Figure 12. Our survey shows, that before 2014, only a couple of methods (i.e statistical) and few studies address analysis of SMLM data. After 2014, researchers started exploring new clustering methods and, at the same time, the number of publications per year started growing, except for 2017. Graph-based cluster analysis methods applied to super-resolution SMLM data started appearing in 2018. We expect more clustering methods based on graphs will appear in 2020 and onward. Ripley's functions are the most popular methods used for super-resolution SMLM cluster analysis over the years as depicted in Figure 12. Further, given their successes in analyzing other imaging modalities, we anticipate an a growth in the number of methods that leverage and adapt machine (and particularly deep) learning methods for SMLM analysis.

5 Conclusions

The SMLM imaging modality is relatively new and is creating exciting opportunities to understand the structure and function of many macromolecular complexes below the diffraction limit of fluorescent microscopy. The data it provides can enable discoveries, but we note there is still a need and an opportunity to develop methods and tools that can read data from different super-resolution microscopes and pre-process the data to handle image acquisition artifacts, be able to provide different visualization alternatives, can analyze a large number of datasets, each in its 3D entirety, to extract and quantitatively describe the structural geometry and interaction of the underlying biological structures, and be able to do that in a way that is either robust to parameter setting, or provide intuitive descriptions of parameters easily communicated to and understandable by the end user.

Our observations from conducting this review revealed that studying the various methods adopted for cluster analysis and quantification requires a benchmarking dataset and evaluation measures for the assessment of the quality of clusters. The benchmark dataset should be available to validate both 2D and 3D cluster analysis methods. It should have several types of clusters with various densities, shapes, sizes, and noise levels.

Our review also motivated us to highlight the importance of strong interdisciplinary collaborations between computational scientists, biophysicists, biochemistries, and biologists for novel breakthrough discoveries. Furthermore, artificial intelligence algorithms (e.g. machine learning) needs to be incorporated in the analysis to (i) get rid of the subjectivity and bias; (ii) robustly analyze the generated big-data; (iii) automatically identify the distinct biological structures and their constituent biosignatures.

In this paper, we summarized and compared the various computational methods for SMLM cluster analysis and quantification. We note that network/graph-based methods have more capabilities such that they could be used for pre-processing (e.g. correcting for multiple blinking of a single fluorophore artifact) and post-processing (e.g. filtering, segmenting, identifying the biological structures) of the SMLM data. Graph-based methods could be applied to extract per-point and per-cluster features for analyzing 3D SMLM data, analyze whole filed of view, process big-data (millions of localizations), and extract heterogeneous clusters. Graph-based methods well-suited to extract intra-cluster features and they could be integrated with the machine learning algorithms for automatic analysis and quantification of the underlying clusters.

In the end, we believe that synergy and harmony between cross research disciplines such as biophysics, biology, imaging, biochemistry, artificial intelligence (AI), etc. are required to improve our understanding of the underlying protein cluster structure and function. The extraordinary SMLM imaging modality and elegant computational methods will result/lead to a better understanding of protein interactions in several subcellular structures. Consequently, enhancing the modeling of the antibodies and drugs design for better diseases therapies.

Table 1: The summary for the main categories of the super-resolution cluster analysis and quantification methods. The cons (+) and pros (−) for each of the clustering analysis method are also shown for comparison.

Method	**Pros (+)**	**Cons (−)**
Statistical	• Simple and easy to implement. • Could be used to detect the level of clustering. • Could be used for both pre- and post-processing.	• Restricted to analyze homogeneous clusters. • Edge effect. • Some normalization based on Poisson point process assumption.
Bayesian	• Can handle SMLM localizations and their associated uncertainties. • Parameter-free model.	• Very slow. • Sensitive to the prior settings. • Used in combination with other methods. • Require generation of thousands of cluster proposals. • Very sensitive to the imaging artifacts.
Density-based	• Efficient in noise removal. • Could be used to discover clusters with various shapes.	• Clustering is conditioned on the minimum density of molecules within neighborhood radius. • Slow scaling with the number of localizations. • The ambiguity and subjectivity in selecting the algorithm parameters affect its performance. • Cannot deal with varying cluster densities and hollow clusters.
Voronoi-based	• Fast and scalable to handle big-data. • Efficient in noise removal. • Sensitive to clusters of specific geometry (e.g. tubular-like structures).	• Might cause problems in segmenting data with non-isotropic distribution. • Limited multi-scale capabilities. • Might not be good for segmenting hollow clusters.
Graph-based	• Fast and scalable to handle big-data. • Easy to be integrated with machine/deep learning. • Robust to noise. • Capable to extract per-point and per-cluster features. • Topological graph is invariant to the dimensionality of the data. • Extracting heterogeneous clusters capability. • Could be used for both pre- and post-processing. • All the other methods can be derived from the graph-based method.	• Graph construction is not straightforward for big-data. • Might cause problems in segmenting data with non-isotropic distribution. • Clustering results are highly dependent on the graph construction method.

Table 2: Super-resolution SMLM cluster analysis and quantification methods (post-processing). The nomenclatures used in this table are as follows. Ref: reference to the method; App: application; ImgMeth: imaging method; Dim: dimensionality; Res: resolution; #FPI: number of frames per image; #Loc: number of localizations; DataSz: dataset size; CAM: cluster analysis method; A/V: area/volume; WAV: whole area visualization; SRIC: surface reconstruction for individual clusters; CP: cluster properties; ICA: intra-cluster analysis; ML & CC: machine learning and cluster classification; MSA: multi-scale analysis; SW: software. NR: the info is not reported. *info is reported for simulated data; #code available upon request. † data acquisition is 3D but the analysis is applied to the projected 2D data.

Study			**Acquisition**						**Analysis**								
Ref	Year	App	ImgMeth	Dim	Res (nm)	#FPI	#Loc	DataSz	CAM	A/V(μm^2 /μm^3)	WAV	SRIC	CP	ICA	ML & CC	MSA	SW
[73]	2010	Lck and Src in T cells	PALM dSTORM	2D	$\sim$20	15K	1500/μm^2	NR	Ripley	2×2	✓		2–5			✓	
[74]	2010	TCR and Lat in T cells	hsPALM	2D	$\sim$25	1K	140–150/μm^2	5–10 cells per exp. (3 exp.)	Ripley	NR	✓		2–5			✓	
[93]	2011	Lat in T cells	PALM dSTORM	2D	NR	150–200K	NR	3–25 exp.	Ripley G&F	3×3			3–4			✓	
[75]	2012	HIV-1 amtrix in HIV-1 virus	dSTORM	2D	15–20	20K	NR	5–6 cells per exp. (2 exp.)	Ripley; G&F	10×10	✓		2–5			✓	
[92]	2012	LAT in T; HeLa cells	PALM	2D	NR	15K	NR	7 cells per cond. (4 cond.)	Ripley G&F	3×3			2			✓	
[76]	2013	NKG2D in NK T cells	PALM GSD	2D	20–30	20–25K	1140–1920/μm^2	16–23 cells (2–4 exp.)	Ripley	2×2 3×3	✓		2–5			✓	
[77]	2013	Lck and CD45 in T cells	PALM dSTORM	2D	21	15–20K	NR	10–13 cells (3 exp.)	Ripley	3×3 4×4	✓		2–5			✓	
[80]	2013	DiI,Lat vesicles in T cells	HILO-PALM	3D	NR	NR	NR	NR	Ripley G&F	2×2	✓	✓	3–5				
[81]	2013	HIV-1 Gag polyprotein in T cells	dSTORM	$2D^{\dagger}$	$\sim$20	4–10K	NR	NR	Ripley NN	2×2	✓		2–3			✓	
[99]	2014	Lck and CD45 in T cells	PALM dSTORM	2D	20–30	NR	NR	NR	G&F	2×2*	✓		2			✓	
[89]	2015	CD37,β2-integrin in HL-60 cells	dSTORM	2D	NR	8K	NR	20 cells	Ripley G&F	3×3	✓						
[104]	2015	Lyn kinase,BCR in CH27 cells	STORM	2D	NR	NR	NR	5 cells	steady state cross-correlation	NR	✓					✓	Matlab func.
[85]	2016	STAT1,STAT3 in HeLa cells; CENP-A in U2OS cells	dSTORM	2D	29	5K	NR	20 cells per cond. (5 exp.)	Ripley G&F	4×4	✓		2–3			✓	
[91]	2016	KIR,KIR2DL1, KIR2DS1 in NKL cells	GSD	2D	NR	20K	NR	14–35 cells per exp. (24 exp.)	Ripley G&F	3×3	✓		8				
[149]	2017	TLR4 in HEK 293 cells	PALM	2D	50	NR	NR	9–10 cells per cond.	NN histogram	NR	✓						
[88]	2017	FcγRI,FcγRII, SIRPα in T,B cells	dSTORM	2D	25	5K	NR	10–30 cells	Ripley G&F CBC	5×5	✓		2–3			✓	
[87]	2018	NKG2D in T cells	dSTORM	2D	NR	5K	NR	12–500 cells	Ripley G&F	5×5	✓		2–3			✓	
[94]	2018	actin cytoskeleton in T cells	iPALM dSTORM	2D 3D	NR	50–100K	NR	NR	Angular Ripley	3×3			2–3			✓	Matlab func.
[78]	2015	CD3 in T cells	PALM dSTORM	2D	10–30	20K	2K*	30 ROIs (per cond.)	Bayesian; Ripley	3×3	✓		4			✓	R func.

Study			**Acquisition**						**Analysis**								
Ref	Year	App	ImgMeth	Dim	Res (nm)	#FPI	#Loc	DataSz	CAM	A/V(μm^2 /μm^3)	WAV	SRIC	CP	ICA	ML & CC	MSA	SW
[90]	2015	LFA-1 in T cells	dSTORM	2D	NR	NR	NR	10 cells	Ripley G&F Bayesian	3×3* 2×2			4				
[105]	2016	ZAP-70 in T cells	PALM dSTORM	2D	20–30	NR	15–20K	12 cells	Bayesian	2×2 3×3			3			✓	Matlab,R
[83]	2017	LAT vesicles in T cells	iPALM	3D	10–30	30K	3K 2K*	5 cells per cond.	Ripley Bayesian	3×3×0.6* 2×2×0.6			5			✓	
[55]	2018	CD4 in T cells	Live-cell PALM	2D	NR	NR	4K	6 cells	Ripley Bayesian	3×3* 2×2			3				
[109]	2014	GaG data	PALM	2D	NR	NR	NR	NR	DBSCAN	NR	✓	✓					PALMsiever
[79]	2015	HIV,Nef, PACS-1, LAMP1, etc. in HeLa cells	GSD	2D	20	5–30K	NR	5 cells per exp. (3 exp.)	Density-based; Ripley	1.5×1.5 3.5×3.5 4×4	✓		4			✓	MIiSR
[107]	2015	RNAP II in cortex cells; H-NS in E.coli	PALM dSTORM	2D	∼10	NR	NR	NR	Density-based; DBSCAN	1×1 10×10*	✓		2			✓	FOCAL
[82]	2016	TCR, VD45 in T cells	dSTORM	2D	20–30	20K	NR	NR	Ripley DBSCAN	4×4	✓		13			✓	Clus-DoC
[96]	2016	HIV,gag,env in T cells	SMLM	2D	NR	NR	NR	1 cell	DBSCAN Ripley OPTICS	NR			5			✓	LAMA
[111]	2016	mitochondrial protein Tom20 in neuroblast brain cells	STORM	3D	40	20K	NR	NR	DBSCAN	NR	✓		3				VividSTORM
[112]	2017	RGD peptides	dSTORM	2D	16	20K	NR	NR	DBSCAN NND	5×5*	✓		3			✓	
[113]	2017	Salmonella typhimurium mutants in bacterial cells	FPALM	2D 3D	∼35	6–18K	NR	58–600 cells	DBSCAN wavelet	NR	✓		3			✓	
[150]	2017	CD4 glycoprotein mutants in T cells	SOFI; simulated PALM	2D	NR	5K	NR	20 cells per cond.	SOFI density analysis	3×3	✓		3				code#
[102]	2018	cis-Golgi,GRASP65, GM130,trans-Golgi, TGN46 in RPE cells	STORM	2D	NR	NR	NR	NR	DBSCAN histogram correlation	NR	✓		2–3				Python func.
[95]	2018	Synapsin,VGLUT in primary hippocampal neurons	SIM STORM	3D	NR	30K	0.5M	NR	DBSCAN Ripley SODA	20×20×2	✓		3–5				Icy Plugin
[110]	2018	Proteins within centrioles and procentrioles in KE37 cells	STORM	3D	32–65	30–60K	NR	NR	DBSCAN	2×2	✓	✓	8		✓		SPARTAN
[103]	2018	Trastuzumab (HER2 receptor)in breast cancer cell lines (BT-474,SK-BR-3, MDA-MB-468)	dSTORM	2D	NR	20–40K	230–360 locs/μm^2	17–23 cells	density pair-correlation	20×20	✓		6		✓		ClusterOccupancy
[108]	2019	nuclear pore complex (NPC) in U-2 human OS cells	dSTORM	3D	NR	NR	166,274*	NR	Density-based; DBSCAN	2×2 5×5×5*	✓		2			✓	FOCAL3D

Study			Acquisition						Analysis								
Ref	Year	App	ImgMeth	Dim	Res (nm)	#FPI	#Loc	DataSz	CAM	A/V(μm^2 /μm^3)	WAV	SRIC	CP	ICA	ML & CC	MSA	SW
[151]	2019	DSB foci in Human (U2Os) cells	dSTORM	2D	NR	NR	NR	NR	KDE DBSCAN Voronoi	NR			3				SMoLR
[115]	2015	Microtubules in COS7 cells; GluA1,Tubulin integrin-β3 in neuronal cells	PALM dSTORM	2D	33.9	NR	0.024–0.277M	3 cells per cond.	Voronoi	15×15*	✓		4				SR-Tesseler
[116]	2016	Microtubules, Chromatin in HeLa cells	GSD	2D	NR	NR	0.230M	NR	Voronoi	4×4*	✓		5				ClusterVisu
[84]	2016	β-tubulin in HeLa cells TPR	GSD	2D	20	NR	NR	NR	Ripley Voronoi	1.5×1.5	✓					✓	SharpViSu
[118]	2018	β-tubulin in HeLa cells; CENP-A in U2OS cells	dSTORM	3D	NR	28–50K	NR	3 exp.	Voronoi	ROIs from 18×18	✓	✓	2–4				3DClusterViSu
[86]	2018	RAD51,RPA in HeLa cells; HPNE,LN9 EUFA423	dSTORM	$2D^{\dagger}$	∼30–40	25K	NR	5 exp.	Ripley Delaunay Voronoi	NR	✓		2–5				Grafeo
[97]	2018	F-actin in T cell; microtubule network in fixed HeLa cell	dSTORM	2D	60	100K	NR	3–5 cells	Voronoi; angular Ripley	3×3	✓		4				
[140]	2019	nuclear pore complex; microtubules; actin cytoskeleton regulators;	DNA-PAINT dSTORM PALM	2D/3D	20-60	40K	20K-8.3M	3-18 cells per cond.	Voronoi	NR	✓		4				Coloc-Tesseler
[126]	2018	Cav1,Cavin-1 in PC3 cells	GSD	3D	20–50	32–40K	0.45–1.2M	9–11 cells per cond. 2 cond. (4 exp.)	Network graph mean-shift	18×18×0.8	✓	✓	28	✓	✓	✓	
[128]	2019	Cav1,Cavin-1 in PC3 cells	GSD	3D	20–50	32–40K	1.17–1.43M	10 cells per cond. (2 cond.)	Network graph mean-shift graphlet	18×18×0.8	✓		11–47	✓	✓	✓	

The criteria details used in Table 2 are as below:

- **Application**: the main biological application of the related studies. The imaged protein and biological model.
- **Imaging method**: imaging technique applied in the study (the study may apply more than one imaging method, we are here focusing on the SMLM super-resolution methods only):
  - PALM: photoactivated localization microscopy
  - hsPALM: high-speed version of PALM
  - iPALM: interferometric PALM

  - FPALM: Fluorescence PALM
  - STORM: stochastic optical reconstruction microscopy
  - dSTORM: direct STORM
  - GSD: ground state depletion
  - GSDIM: GSD followed by individual molecule return
  - HILO: highly inclined and laminated optical sheet
  - SOFI: super-resolution optical fluctuation imaging
  - SIM: structured-illumination microscopy
- **Dimensionality**: is either 2D data analysis (for 2D acquisition) or 3D data analysis (for 3D acquisition) applied for the provided method. † Means that the data acquisition is 3D but the analysis is applied to the projected 2D data.
- **Resolution / Localization precision**: The resolution of the images on which the reported analysis is applied. (We report the localization precision if the resolution is missing). The localization precision is much smaller than the spatial image resolution.
- **Area/Volume**: the total area or volume where the analysis is applied.
- **#frames per image**: number of frames per super-resolution image (in thousands: K).
- **#localizations (blinks, events, pointillist, etc.)**: the total number or density of acquired blinks/molecules (in PALM-based methods)/localization events as described in the study.
- **Dataset size**: data set size (how many cells, how many experiments used in the work).
- **Cluster analysis method**: the clustering analysis method(s) used in the work.
- **Whole-area visualization**: is cluster visualization of the whole analyzed area provided (after cluster analysis)?
- **Surface reconstruction for individual clusters**: is surface reconstruction for individual clusters provided?
- **Cluster properties/descriptors**: does the study provide analysis on any cluster properties? Yes or No. If yes, how many features have been used in the analysis?
- **Intra-cluster analysis**: does the study provide detailed analysis at intra-cluster levels (network analysis of the molecules, modularity analysis, sub-networks, etc.)?
- **Machine learning and cluster classification**: is machine learning used to automatically classify the clusters?
- **Multi-scale analysis**: is multi-scale analysis supported and used by the method?
- **Software**: is a software available for the method/algorithm? Yes (software name) or No.

Acknowledgments

Supported by grants from the CIHR (PJT-156424, PJT-159845), CHRP, NSERC and CFI/BCKDF (IRN, GH). We thank the anonymous reviewers for their extensive feedback and insightful suggestions that were invaluable in improving this review.

Author Contributions

Conceptualization, I.M.K., I.R.N., and G.H.; Methodology, I.M.K., I.R.N., and G.H.; Software, I.M.K.; Formal Analysis, I.M.K, I.R.N., and G.H.; Investigation, I.M.K., I.R.N., and G.H.; Writing – Original Draft, I.M.K.; Writing – Review & Editing, I.M.K., I.R.N., and G.H.; Visualization, I.M.K.; Funding Acquisition, I.R.N. and G.H.; Resources, I.R.N. and G.H.; Supervision, I.R.N., and G.H.

Declaration of Interests

An international patent PCT/CA2018/051553 covering the material presented in the manuscript has been submitted by the authors: "Methods for Analysis of Single Molecule Localization Microscopy to Define Molecular Architecture", US Patent Application No. 62/594,642, Dec 5, 2018.

References

[1] Ernst Abbe. Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung. *Archiv für mikroskopische Anatomie*, 9(1):413–418, 1873.

[2] Erdinc Sezgin. Super-resolution optical microscopy for studying membrane structure and dynamics. *Journal of Physics: Condensed Matter*, 29(27):273001, 2017.

[3] Sviatlana Shashkova and Mark C Leake. Single-molecule fluorescence microscopy review: shedding new light on old problems. *Bioscience reports*, 37(4):BSR20170031, 2017.

[4] Teresa Klein, Sven Proppert, and Markus Sauer. Eight years of single-molecule localization microscopy. *Histochemistry and cell biology*, 141(6):561–575, 2014.

[5] Daniel Choquet. The 2014 Nobel Prize in Chemistry: a large-scale prize for achievements on the nanoscale. *Neuron*, 84(6):1116–1119, 2014.

[6] Romain F Laine, Gabriele S Kaminski Schierle, Sebastian van de Linde, and Clemens F Kaminski. From single-molecule spectroscopy to super-resolution imaging of the neuron: a review. *Methods and applications in fluorescence*, 4(2):022004, 2016.

[7] Eric Betzig, George H Patterson, Rachid Sougrat, O Wolf Lindwasser, Scott Olenych, Juan S Bonifacino, Michael W Davidson, Jennifer Lippincott-Schwartz, and Harald F Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645, 2006.

[8] William E Moerner and Lothar Kador. Optical detection and spectroscopy of single molecules in a solid. *Physical review letters*, 62(21):2535, 1989.

[9] Stefan W Hell and Jan Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics letters*, 19(11):780–782, 1994.

[10] Mats GL Gustafsson. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of microscopy*, 198(2):82–87, 2000.

[11] Thomas Dertinger, Ryan Colyer, Gopal Iyer, Shimon Weiss, and Jörg Enderlein. Fast, background-free, 3d super-resolution optical fluctuation imaging (sofi). *Proceedings of the National Academy of Sciences*, 106(52):22287–22292, 2009.

[12] Shachar Schidorsky, Xiyu Yi, Yair Razvag, Julia Sajman, Kobi Hermon, Shimon Weiss, and Eilon Sherman. Synergizing superresolution optical fluctuation imaging with single molecule localization microscopy. *Methods and applications in fluorescence*, 6(4):045008, 2018.

[13] Samuel T Hess, Thanu PK Girirajan, and Michael D Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical journal*, 91(11):4258–4272, 2006.

[14] Michael J Rust, Mark Bates, and Xiaowei Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods*, 3(10):793, 2006.

[15] Mike Heilemann, Sebastian Van De Linde, Mark Schüttpelz, Robert Kasper, Britta Seefeldt, Anindita Mukherjee, Philip Tinnefeld, and Markus Sauer. Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. *Angewandte Chemie International Edition*, 47(33):6172–6176, 2008.

[16] Jonas Fölling, Mariano Bossi, Hannes Bock, Rebecca Medda, Christian A Wurm, Birka Hein, Stefan Jakobs, Christian Eggeling, and Stefan W Hell. Fluorescence nanoscopy by ground-state depletion and single-molecule return. *Nature methods*, 5(11):943–945, 2008.

[17] Joerg Schnitzbauer, Maximilian T Strauss, Thomas Schlichthaerle, Florian Schueder, and Ralf Jungmann. Super-resolution microscopy with DNA-PAINT. *Nature protocols*, 12(6):1198, 2017.

[18] Francisco Balzarotti, Yvan Eilers, Klaus C Gwosch, Arvid H Gynnå, Volker Westphal, Fernando D Stefani, Johan Elf, and Stefan W Hell. Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. *Science*, 355(6325):606–612, 2017.

[19] Dylan Myers Owen and Katharina Gaus. Imaging lipid domains in cell membranes: the advent of super-resolution fluorescence microscopy. *Frontiers in plant science*, 4:503, 2013.

[20] Eva Wegel, Antonia Göhler, B Christoffer Lagerholm, Alan Wainman, Stephan Uphoff, Rainer Kaufmann, and Ian M Dobbie. Imaging cellular structures in super-resolution with sim, sted and localisation microscopy: A practical comparison. *Scientific reports*, 6(1):1–13, 2016.

[21] Bo Huang, Hazen Babcock, and Xiaowei Zhuang. Breaking the diffraction barrier: super-resolution imaging of cells. *Cell*, 143(7):1047–1058, 2010.

[22] David Baddeley and Joerg Bewersdorf. Biological insight from super-resolution microscopy: What we can learn from localization-based images. *Annual review of biochemistry*, 87:965–989, 2018.

[23] Bo Huang, Wenqin Wang, Mark Bates, and Xiaowei Zhuang. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science*, 319(5864):810–813, 2008.

[24] Philip R Nicovich, Dylan M Owen, and Katharina Gaus. Turning single-molecule localization microscopy into a quantitative bioanalytical tool. *Nature protocols*, 12(3):453, 2017.

[25] Stefan W Hell, Steffen J Sahl, Mark Bates, Xiaowei Zhuang, Rainer Heintzmann, Martin J Booth, Joerg Bewersdorf, Gleb Shtengel, Harald Hess, Philip Tinnefeld, et al. The 2015 super-resolution microscopy roadmap. *Journal of Physics D: Applied Physics*, 48(44):443001, 2015.

[26] Markus Sauer and Mike Heilemann. Single-molecule localization microscopy in eukaryotes. *Chemical reviews*, 117(11):7478–7509, 2017.

[27] Ottavia Golfetto, Devin L Wakefield, Eliedonna E Cacao, Kendra N Avery, Victor Kenyon, Raphael Jorand, Steven J Tobin, Sunetra Biswas, Jennifer Gutierrez, Ronald Clinton, et al. A platform to enhance quantitative single molecule localization microscopy. *Journal of the American Chemical Society*, 140(40):12785–12797, 2018.

[28] Steffen J Sahl, Stefan W Hell, and Stefan Jakobs. Fluorescence nanoscopy in cell biology. *Nature reviews Molecular cell biology*, 18(11):685, 2017.

[29] Bartosz Turkowyd, David Virant, and Ulrike Endesfelder. From single molecules to life: microscopy at the nanoscale. *Analytical and bioanalytical chemistry*, 408(25):6885–6911, 2016.

[30] Manuel F Juette, Travis J Gould, Mark D Lessard, Michael J Mlodzianoski, Bhupendra S Nagpure, Brian T Bennett, Samuel T Hess, and Joerg Bewersdorf. Three-dimensional sub–100 nm resolution fluorescence microscopy of thick samples. *Nature methods*, 5(6):527–529, 2008.

[31] Sri Rama Prasanna Pavani, Michael A Thompson, Julie S Biteen, Samuel J Lord, Na Liu, Robert J Twieg, Rafael Piestun, and WE Moerner. Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proceedings of the National Academy of Sciences*, 106(9):2995–2999, 2009.

[32] David Baddeley, Mark B Cannell, and Christian Soeller. Three-dimensional sub-100 nm super-resolution imaging of biological samples using a phase ramp in the objective pupil. *Nano Research*, 4(6):589–598, 2011.

[33] Andrey Aristov, Benoit Lelandais, Elena Rensen, and Christophe Zimmer. Zola-3d allows flexible 3d localization microscopy over an adjustable axial range. *Nature communications*, 9(1):1–8, 2018.

[34] Gleb Shtengel, James A Galbraith, Catherine G Galbraith, Jennifer Lippincott-Schwartz, Jennifer M Gillette, Suliana Manley, Rachid Sougrat, Clare M Waterman, Pakorn Kanchanawong, Michael W Davidson, et al. Interferometric fluorescent super-resolution microscopy resolves 3d cellular ultrastructure. *Proceedings of the National Academy of Sciences*, 106(9):3125–3130, 2009.

[35] Daniel Aquino, Andreas Schönle, Claudia Geisler, Claas v Middendorff, Christian A Wurm, Yosuke Okamura, Thorsten Lang, Stefan W Hell, and Alexander Egner. Two-color nanoscopy of three-dimensional volumes by 4pi detection of stochastically switched fluorophores. *Nature methods*, 8(4):353, 2011.

[36] Nicolas Bourg, Céline Mayet, Guillaume Dupuis, Thomas Barroca, Pierre Bon, Sandrine Lécart, Emmanuel Fort, and Sandrine Lévêque-Fort. Direct optical nanoscopy with axially localized detection. *Nature Photonics*, 9(9):587, 2015.

[37] Martin Ovesnỳ, Pavel Křížek, Josef Borkovec, Zdeněk Švindrych, and Guy M Hagen. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics*, 30(16):2389–2390, 2014.

[38] Ricardo Henriques, Mickael Lelek, Eugenio F Fornasiero, Flavia Valtorta, Christophe Zimmer, and Musa M Mhlanga. QuickPALM: 3D real-time photoactivation nanoscopy image processing in mageJ. *Nature methods*, 7(5):339, 2010.

[39] Steve Wolter, Anna Löschberger, Thorge Holm, Sarah Aufmkolk, Marie-Christine Dabauvalle, Sebastian Van De Linde, and Markus Sauer. rapidSTORM: accurate, fast open-source software for localization microscopy. *Nature methods*, 9(11):1040, 2012.

[40] Eric J Rees, Miklos Erdelyi, Gabriele S Kaminski Schierle, Alex Knight, and Clemens F Kaminski. Elements of image processing in localization microscopy. *Journal of Optics*, 15(9):094012, 2013.

[41] A Shivanandan, H Deschout, Marco Scarselli, and A Radenovic. Challenges in quantitative single molecule localization microscopy. *FEBS letters*, 588(19):3595–3602, 2014.

[42] Talley J Lambert and Jennifer C Waters. Navigating challenges in the application of superresolution microscopy. *J Cell Biol*, 216(1):53–63, 2017.

[43] Antony Lee, Konstantinos Tsekouras, Christopher Calderon, Carlos Bustamante, and Steve Pressé. Unraveling the thousand word picture: an introduction to super-resolution data analysis. *Chemical reviews*, 117(11):7276–7330, 2017.

[44] Ulrike Endesfelder and Mike Heilemann. Art and artifacts in single-molecule localization microscopy: beyond attractive images. *Nature methods*, 11(3):235, 2014.

[45] Paolo Annibale, Stefano Vanni, Marco Scarselli, Ursula Rothlisberger, and Aleksandra Radenovic. Quantitative photo activated localization microscopy: unraveling the effects of photoblinking. *PloS one*, 6(7):e22678, 2011.

[46] Siân Culley, David Albrecht, Caron Jacobs, Pedro Matos Pereira, Christophe Leterrier, Jason Mercer, and Ricardo Henriques. Quantitative mapping and minimization of super-resolution optical imaging artifacts. *Nature methods*, 15(4):263, 2018.

[47] Richard J Marsh, Karin Pfisterer, Pauline Bennett, Liisa M Hirvonen, Mathias Gautel, Gareth E Jones, and Susan Cox. Artifact-free high-density localization microscopy analysis. *Nature methods*, 15(9):689–692, 2018.

[48] Daniel Sage, Thanh-An Pham, Hazen Babcock, Tomas Lukes, Thomas Pengo, Jerry Chao, Ramraj Velmurugan, Alex Herbert, Anurag Agrawal, Silvia Colabrese, et al. Super-resolution fight club: assessment of 2d and 3d single-molecule localization microscopy software. *Nature methods*, 16(5):387–395, 2019.

[49] Paolo Annibale, Stefano Vanni, Marco Scarselli, Ursula Rothlisberger, and Aleksandra Radenovic. Identification of clustering artifacts in photoactivated localization microscopy. *Nature methods*, 8(7):527, 2011.

[50] Franziska Fricke, Joel Beaudouin, Roland Eils, and Mike Heilemann. One, two or three? probing the stoichiometry of membrane proteins by single-molecule localization microscopy. *Scientific reports*, 5:14072, 2015.

[51] Christos Karathanasis, Franziska Fricke, Gerhard Hummer, and Mike Heilemann. Molecule counts in localization microscopy with organic fluorophores. *ChemPhysChem*, 18(8):942–948, 2017.

[52] Florian Baumgart, Andreas M Arnold, Konrad Leskovar, Kaj Staszek, Martin Fölser, Julian Weghuber, Hannes Stockinger, and Gerhard J Schütz. Varying label density allows artifact-free analysis of membrane-protein nanoclusters. *Nature methods*, 13(8):661, 2016.

[53] Prabuddha Sengupta, Tijana Jovanovic-Talisman, Dunja Skoko, Malte Renz, Sarah L Veatch, and Jennifer Lippincott-Schwartz. Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. *Nature methods*, 8(11):969, 2011.

[54] Miklos Erdelyi, Eric Rees, Daniel Metcalf, Gabriele S Kaminski Schierle, Laszlo Dudas, Jozsef Sinko, Alex E Knight, and Clemens F Kaminski. Correcting chromatic offset in multicolor super-resolution localization microscopy. *Optics express*, 21(9):10978–10988, 2013.

[55] Juliette Griffié, Garth L Burn, David J Williamson, Ruby Peters, Patrick Rubin-Delanchy, and Dylan M Owen. Dynamic Bayesian cluster analysis of live-cell single molecule localization microscopy datasets. *Small Methods*, page 1800008, 2018.

[56] Mickaël Lelek, Francesca Di Nunzio, Ricardo Henriques, Pierre Charneau, Nathalie Arhel, and Christophe Zimmer. Superresolution imaging of hiv in infected cells with flash-palm. *Proceedings of the National Academy of Sciences*, 109(22):8564–8569, 2012.

[57] Anna Szymborska, Alex De Marco, Nathalie Daigle, Volker C Cordes, John AG Briggs, and Jan Ellenberg. Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science*, 341(6146):655–658, 2013.

[58] Raquel Salvador-Gallego, Markus Mund, Katia Cosentino, Jale Schneider, Joseph Unsay, Ulrich Schraermeyer, Johann Engelhardt, Jonas Ries, and Ana J García-Sáez. Bax assembly into rings and arcs in apoptotic mitochondria is linked to membrane pores. *The EMBO journal*, 35(4):389–401, 2016.

[59] Markus Mund, Johannes Albertus van der Beek, Joran Deschamps, Serge Dmitrieff, Philipp Hoess, Jooske Louise Monster, Andrea Picco, François Nédélec, Marko Kaksonen, and Jonas Ries. Systematic nanoscale analysis of endocytosis links efficient vesicle formation to patterned actin nucleation. *Cell*, 174(4):884–896, 2018.

[60] Leonid Andronov, Khalid Ouararhni, Isabelle Stoll, Bruno P Klaholz, and Ali Hamiche. Cenp-a nucleosome clusters form rosette-like structures around hjurp during g1. *Nature communications*, 10(1):1–8, 2019.

[61] Eilon Sherman. Resolving protein interactions and organization downstream the T cell antigen receptor using single-molecule localization microscopy: a review. *Methods and Applications in Fluorescence*, 4(2):022002, 2016.

[62] Brian D Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(2):172–192, 1977.

[63] Philip M Dixon. Ripley's K function. *Wiley StatsRef: Statistics Reference Online*, 2014.

[64] Kristin Hansson, Mehrdad Jafari-Mamaghani, and Patrik Krieger. RipleyGUI: software for analyzing spatial patterns in 3D cell distributions. *Frontiers in neuroinformatics*, 7:5, 2013.

[65] Maria A Kiskowski, John F Hancock, and Anne K Kenworthy. On the use of Ripley's K-function and its derivatives to analyze domain size. *Biophysical journal*, 97(4):1095–1103, 2009.

[66] JE Besag. Comments on Ripley's paper. *Journal of the Royal Statistical Society B*, 39:193–195, 1977.

[67] Marcelo Ehrlich, Werner Boll, Antoine Van Oijen, Ramesh Hariharan, Kartik Chandran, Max L Nibert, and Tomas Kirchhausen. Endocytosis by random initiation and stabilization of clathrin-coated pits. *Cell*, 118(5):591–605, 2004.

[68] Thorsten Wiegand and Kirk A. Moloney. Rings, circles, and null-models for point pattern analysis in ecology. *Oikos*, 104(2):209–229, 2004.

[69] Peter Haase. Spatial pattern analysis in ecology based on Ripley's K-function: Introduction and methods of edge correction. *Journal of vegetation science*, 6(4):575–582, 1995.

[70] Eric Marcon and Florence Puech. Generalizing Ripley's K function to inhomogeneous populations. 2009.

[71] AJ Baddeley, RA Moyeed, CV Howard, and A Boyde. Analysis of a three-dimensional point pattern with replication. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 42(4):641–668, 1993.

[72] François Goreaud and Raphaël Pélissier. On explicit formulas of edge effect correction for Ripley's K-function. *Journal of Vegetation Science*, 10(3):433–438, 1999.

[73] Dylan M Owen, Carles Rentero, Jérémie Rossy, Astrid Magenau, David Williamson, Macarena Rodriguez, and Katharina Gaus. Palm imaging and cluster analysis of protein heterogeneity at the cell surface. *Journal of biophotonics*, 3(7):446–454, 2010.

[74] Björn F Lillemeier, Manuel A Mörtelmaier, Martin B Forstner, Johannes B Huppa, Jay T Groves, and Mark M Davis. TCR and lat are expressed on separate protein islands on T cell membranes and concatenate during activation. *Nature immunology*, 11(1):90, 2010.

[75] Cândida F Pereira, Jérémie Rossy, Dylan M Owen, Johnson Mak, and Katharina Gaus. HIV taken by STORM: super-resolution fluorescence microscopy of a viral infection. *Virology journal*, 9(1):84, 2012.

[76] Sophie V Pageon, Shaun-Paul Cordoba, Dylan M Owen, Stephen M Rothery, Anna Oszmiana, and Daniel M Davis. Superresolution microscopy reveals nanometer-scale reorganization of inhibitory natural killer cell receptors upon activation of NKG2D. *Science signaling*, 6(285):ra62–ra62, 2013.

[77] Jérémie Rossy, Dylan M Owen, David J Williamson, Zhengmin Yang, and Katharina Gaus. Conformational states of the kinase lck regulate clustering in early T cell signaling. *Nature immunology*, 14(1):82, 2013.

[78] Patrick Rubin-Delanchy, Garth L Burn, Juliette Griffié, David J Williamson, Nicholas A Heard, Andrew P Cope, and Dylan M Owen. Bayesian cluster identification in single-molecule localization microscopy data. *Nature methods*, 12(11):1072, 2015.

[79] Fabiana A Caetano, Brennan S Dirk, Joshua HK Tam, P Craig Cavanagh, Maria Goiko, Stephen SG Ferguson, Stephen H Pasternak, Jimmy D Dikeakos, John R de Bruyn, and Bryan Heit. MIiSR: molecular interactions in super-resolution imaging enables the analysis of protein interactions, dynamics and formation of multi-protein structures. *PLoS computational biology*, 11(12):e1004634, 2015.

[80] Dylan M Owen, David J Williamson, Lies Boelen, Astrid Magenau, Jérémie Rossy, and Katharina Gaus. Quantitative analysis of three-dimensional fluorescence localization microscopy data. *Biophysical journal*, 105(2):L05–L07, 2013.

[81] Sebastian Malkusch, Walter Muranyi, Barbara Müller, Hans-Georg Kräusslich, and Mike Heilemann. Single-molecule coordinate-based analysis of the morphology of HIV-1 assembly sites with near-molecular spatial resolution. *Histochemistry and cell biology*, 139(1):173–179, 2013.

[82] Sophie V Pageon, Philip R Nicovich, Mahdie Mollazade, Thibault Tabarin, and Katharina Gaus. Clus-DoC: a combined cluster detection and colocalization analysis for single-molecule localization microscopy data. *Molecular biology of the cell*, 27(22):3627–3636, 2016.

[83] Juliette Griffié, Leigh Shlomovich, David J Williamson, Michael Shannon, Jesse Aaron, Satya Khuon, Garth Burn, Lies Boelen, Ruby Peters, Andrew P Cope, et al. 3D Bayesian cluster analysis of super-resolution data reveals LAT recruitment to the T cell synapse. *Scientific reports*, 7(1):4077, 2017.

[84] Leonid Andronov, Yves Lutz, Jean-Luc Vonesch, and Bruno P Klaholz. SharpViSu: integrated analysis and segmentation of super-resolution microscopy data. *Bioinformatics*, 32(14):2239–2241, 2016.

[85] Jing Gao, Feng Wang, Junling Chen, Jianzhong Wang, Mingjun Cai, Haijiao Xu, Junguang Jiang, and Hongda Wang. Super-resolution imaging of STAT3 cellular clustering during nuclear transport. *RSC Advances*, 6(59):54597–54607, 2016.

[86] Kalina T Haas, MiYoung Lee, Alessandro Esposito, and Ashok R Venkitaraman. Single-molecule localization microscopy reveals molecular transactions during RAD51 filament assembly at cellular DNA damage sites. *Nucleic acids research*, 46(5):2398–2416, 2018.

[87] Štefan Bálint, Filipa B Lopes, and Daniel M Davis. A nanoscale reorganization of the IL-15 receptor is triggered by NKG2D in a ligand-dependent manner. *Sci. Signal.*, 11(525):eaal3606, 2018.

[88] Filipa B Lopes, Štefan Bálint, Salvatore Valvo, James H Felce, Edith M Hessel, Michael L Dustin, and Daniel M Davis. Membrane nanoclusters of FcγRI segregate from inhibitory SIRPα upon activation of human macrophages. *J Cell Biol*, pages jcb–201608094, 2017.

[89] Janet L Wee, Keith E Schulze, Eleanor L Jones, Louisa Yeung, Qiang Cheng, Candida F Pereira, Adam Costin, Georg Ramm, Annemiek B van Spriel, Michael J Hickey, et al. Tetraspanin CD37 regulates β2 integrin–mediated adhesion and migration in neutrophils. *The Journal of Immunology*, page 1402414, 2015.

[90] Juliette Griffié, Lies Boelen, Garth Burn, Andrew P Cope, and Dylan M Owen. Topographic prominence as a method for cluster identification in single-molecule localisation data. *Journal of biophotonics*, 8(11-12):925–934, 2015.

[91] Anna Oszmiana, David J Williamson, Shaun-Paul Cordoba, David J Morgan, Philippa R Kennedy, Kevin Stacey, and Daniel M Davis. The size of activating and inhibitory killer Ig-like receptor nanoclusters is controlled by the transmembrane sequence and affects signaling. *Cell reports*, 15(9):1957–1972, 2016.

[92] Dylan M Owen, David J Williamson, Astrid Magenau, and Katharina Gaus. Sub-resolution lipid domains exist in the plasma membrane and regulate protein diffusion and distribution. *Nature communications*, 3:1256, 2012.

[93] David J Williamson, Dylan M Owen, Jérémie Rossy, Astrid Magenau, Matthias Wehrmann, J Justin Gooding, and Katharina Gaus. Pre-existing clusters of the adaptor lat do not participate in early T cell signaling events. *Nature immunology*, 12(7):655, 2011.

[94] R Peters, J Griffié, DJ Williamson, J Aaron, S Khuon, and DM Owen. Development of 2-colour and 3D SMLM data analysis methods for fibrous spatial point patterns. *Journal of Physics D: Applied Physics*, 52(1):014005, 2018.

[95] Thibault Lagache, Alexandre Grassart, Stéphane Dallongeville, Orestis Faklaris, Nathalie Sauvonnet, Alexandre Dufour, Lydia Danglot, and Jean-Christophe Olivo-Marin. Mapping molecular assemblies with fluorescence microscopy and object-based spatial statistics. *Nature communications*, 9(1):698, 2018.

[96] Sebastian Malkusch and Mike Heilemann. Extracting quantitative information from single-molecule super-resolution imaging data with LAMA–LocAlization Microscopy Analyzer. *Scientific reports*, 6:34486, 2016.

[97] Ruby Peters, Juliette Griffié, Garth L Burn, David J Williamson, and Dylan M Owen. Quantitative fibre analysis of single-molecule localization microscopy data. *Scientific reports*, 8(1):10418, 2018.

[98] Arthur Getis and Janet Franklin. Second-order neighborhood analysis of mapped point patterns. *Ecology*, 68(3):473–477, 1987.

[99] Jérémie Rossy, Edward Cohen, Katharina Gaus, and Dylan M Owen. Method for co-cluster analysis in multichannel single-molecule localisation data. *Histochemistry and cell biology*, 141(6):605–612, 2014.

[100] Sarah L Veatch, Benjamin B Machta, Sarah A Shelby, Ethan N Chiang, David A Holowka, and Barbara A Baird. Correlation functions quantify super-resolution images and estimate apparent clustering due to over-counting. *PloS one*, 7(2):e31457, 2012.

[101] Sebastian Malkusch, Ulrike Endesfelder, Justine Mondry, Márton Gelléri, Peter J Verveer, and Mike Heilemann. Coordinate-based colocalization analysis of single-molecule localization microscopy data. *Histochemistry and cell biology*, 137(1):1–10, 2012.

[102] Joerg Schnitzbauer, Yina Wang, Shijie Zhao, Matthew Bakalar, Tulip Nuwal, Baohui Chen, and Bo Huang. Correlation analysis framework for localization-based superresolution microscopy. *Proceedings of the National Academy of Sciences*, 115(13):3219–3224, 2018.

[103] Steven J Tobin, Devin L Wakefield, Veronica Jones, Xueli Liu, Daniel Schmolze, and Tijana Jovanović-Talisman. Single molecule localization microscopy coupled with touch preparation for the quantification of trastuzumab-bound HER2. *Scientific reports*, 8(1):15154, 2018.

[104] Matthew B Stone and Sarah L Veatch. Steady-state cross-correlations for live two-colour super-resolution localization data sets. *Nature communications*, 6:7347, 2015.

[105] Juliette Griffié, Michael Shannon, Claire L Bromley, Lies Boelen, Garth L Burn, David J Williamson, Nicholas A Heard, Andrew P Cope, Dylan M Owen, and Patrick Rubin-Delanchy. A Bayesian cluster analysis method for single-molecule localization microscopy data. *nature protocols*, 11(12):2499, 2016.

[106] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, volume 96, pages 226–231, 1996.

[107] A Mazouchi and JN Milstein. Fast optimized cluster algorithm for localizations (FOCAL): a spatial cluster analysis for super-resolved microscopy. *Bioinformatics*, 32(5):747–754, 2015.

[108] Daniel Nino, Daniel Djayakarsana, and Joshua N Milstein. Focal3d: A 3-dimensional clustering package for single-molecule localization microscopy. *bioRxiv*, page 777722, 2019.

[109] Thomas Pengo, Seamus J Holden, and Suliana Manley. PALMsiever: a tool to turn raw data into results for single-molecule localization microscopy. *Bioinformatics*, 31(5):797–798, 2014.

[110] Christian Sieben, Niccolo Banterle, Kyle M Douglass, Pierre Gönczy, and Suliana Manley. Multicolor single-particle reconstruction of protein complexes. *Nature methods*, 15(10):777, 2018.

[111] László Barna, Barna Dudok, Vivien Miczán, András Horváth, Zsófia I László, and István Katona. Correlated confocal and super-resolution imaging by VividSTORM. *Nature protocols*, 11(1):163, 2016.

[112] Mahdie Mollazade, Thibault Tabarin, Philip R Nicovich, Alexander Soeriyadi, Daniel J Nieves, J Justin Gooding, and Katharina Gaus. Can single molecule localization microscopy be used to map closely spaced RGD nanodomains? *PloS one*, 12(7):e0180871, 2017.

[113] Yongdeng Zhang, María Lara-Tejero, Jörg Bewersdorf, and Jorge E Galán. Visualization and characterization of individual type III protein secretion machines in live bacteria. *Proceedings of the National Academy of Sciences*, 114(23):6098–6103, 2017.

[114] Atsuyuki Okabe, Barry Boots, Kokichi Sugihara, and Sung Nok Chiu. *Spatial tessellations: concepts and applications of Voronoi diagrams*, volume 501. John Wiley & Sons, 2009.

[115] Florian Levet, Eric Hosy, Adel Kechkar, Corey Butler, Anne Beghin, Daniel Choquet, and Jean-Baptiste Sibarita. SR-Tesseler: a method to segment and quantify localization-based super-resolution microscopy data. *Nature methods*, 12(11):1065, 2015.

[116] Leonid Andronov, Igor Orlov, Yves Lutz, Jean-Luc Vonesch, and Bruno P Klaholz. ClusterViSu, a method for clustering of protein complexes by voronoi tessellation in super-resolution microscopy. *Scientific reports*, 6:24084, 2016.

[117] David Baddeley, Isuru D Jayasinghe, Leo Lam, Sabrina Rossberger, Mark B Cannell, and Christian Soeller. Optical single-channel resolution imaging of the ryanodine receptor distribution in rat cardiac myocytes. *Proceedings of the National Academy of Sciences*, 106(52):22275–22280, 2009.

[118] Leonid Andronov, Jonathan Michalon, Khalid Ouararhni, Igor Orlov, Ali Hamiche, Jean-Luc Vonesch, and Bruno P Klaholz. 3DClusterViSu: 3D clustering analysis of super-resolution microscopy data by 3D voronoi tessellations. *Bioinformatics*, 1:9, 2018.

[119] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[120] Jongkwang Kim and Thomas Wilhelm. What is a complex graph? *Physica A: Statistical Mechanics and its Applications*, 387(11):2637–2652, 2008.

[121] Albert-Laszlo Barabasi and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. *Nature reviews genetics*, 5(2):101, 2004.

[122] Andrea Baronchelli, Ramon Ferrer-i Cancho, Romualdo Pastor-Satorras, Nick Chater, and Morten H Christiansen. Networks in cognitive science. *Trends in cognitive sciences*, 17(7):348–360, 2013.

[123] Danielle S Bassett and Olaf Sporns. Network neuroscience. *Nature neuroscience*, 20(3):353, 2017.

[124] Olaf Sporns. Structure and function of complex brain networks. *Dialogues in clinical neuroscience*, 15(3):247, 2013.

[125] Mark Newman. *Networks: an introduction*. Oxford university press, 2010.

[126] Ismail M Khater, Fanrui Meng, Timothy H Wong, Ivan Robert Nabi, and Ghassan Hamarneh. Super resolution network analysis defines the molecular architecture of caveolae and caveolin-1 scaffolds. *Scientific reports*, 8(1):9009, 2018.

[127] Ismail M Khater, David RL Scriven, Edwin DW Moore, and Ghassan Hamarneh. Sub-cellular network analysis of ryanodine receptor positioning in control and phosphorylated states. In *2016 Computing in Cardiology Conference (CinC)*, pages 821–824. IEEE, 2016.

[128] Ismail M Khater, Fanrui Meng, Ivan Robert Nabi, and Ghassan Hamarneh. Identification of Caveolin-1 Domain Signatures via Machine Learning and Graphlet Analysis of Single Molecule Super-Resolution Data. *Bioinformatics*, 02 2019.

[129] Ismail M Khater, Qian Liu, Keng C Chou, Ghassan Hamarneh, and Ivan Robert Nabi. Super-resolution modularity analysis shows polyhedral caveolin-1 oligomers combine to form scaffolds and caveolae. *Scientific reports*, 9(1):9888, 2019.

[130] Joshua Scurll, Libin Abraham, Reza Tafteh, Keng Chou, Michael R Gold, Daniel Coombs, et al. StormGraph: An automated graph-based algorithm for quantitative clustering analysis of single-molecule localization microscopy data. *bioRxiv*, page 515627, 2019.

[131] Jeremy A Pike, Abdullah O Khan, Chiara Pallini, Steven G Thomas, Markus Mund, Jonas Ries, Natalie S Poulter, and Iain B Styles. Topological data analysis quantifies biological nano-structure from single molecule localization microscopy. *Bioinformatics*, 36(5):1614–1621, 10 2019.

[132] Elias Nehme, Lucien E Weiss, Tomer Michaeli, and Yoav Shechtman. Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica*, 5(4):458–464, 2018.

[133] Nicholas Boyd, Eric Jonas, Hazen P Babcock, and Benjamin Recht. Deeploco: Fast 3D localization microscopy using neural networks. *BioRxiv*, page 267096, 2018.

[134] Ben Cardoen, Hanene Ben Yedder, Anmol Sharma, Keng C Chou, Ivan Robert Nabi, and Ghassan Hamarneh. Ergo: Efficient recurrent graph optimized emitter density estimation in single molecule localization microscopy. *IEEE Transactions on Medical Imaging*, 2019.

[135] Ismail M. Khater, Stephane T. Aroca-Ouellette, Fanrui Meng, Ivan Robert Nabi, and Ghassan Hamarneh. Caveolae and scaffold detection from single molecule localization microscopy data using deep learning. *PLOS ONE*, 14(8):1–17, 08 2019.

[136] David J Williamson, Garth L Burn, Sabrina Simoncelli, Juliette Griffié, Ruby Peters, Daniel M Davis, and Dylan M Owen. Machine learning for cluster analysis of localization microscopy data. *Nature Communications*, 11(1):1–10, 2020.

[137] Varun Venkataramani, Frank Herrmannsdörfer, Mike Heilemann, and Thomas Kuner. SuReSim: simulating localization microscopy experiments from ground truth models. *Nature methods*, 13(4):319, 2016.

[138] Tibor Novák, Tamás Gajdos, József Sinkó, Gábor Szabó, and Miklós Erdélyi. TestSTORM: Versatile simulator software for multimodal super-resolution localization fluorescence microscopy. *Scientific reports*, 7(1):951, 2017.

[139] Martin Lindén, Vladimir Ćurić, Alexis Boucharin, David Fange, and Johan Elf. Simulated single molecule microscopy with SMeagol. *Bioinformatics*, 32(15):2394–2395, 2016.

[140] Florian Levet, Guillaume Julien, Rémi Galland, Corey Butler, Anne Beghin, Anaël Chazeau, Philipp Hoess, Jonas Ries, Grégory Giannone, and Jean-Baptiste Sibarita. A tessellation-based colocalization analysis approach for single-molecule localization microscopy. *Nature communications*, 10(1):1–12, 2019.

[141] Christoph Spahn, Frank Herrmannsdörfer, Thomas Kuner, and Mike Heilemann. Temporal accumulation analysis provides simplified artifact-free analysis of membrane-protein nanoclusters. *Nature methods*, 13(12):963, 2016.

[142] Yiming Li, Markus Mund, Philipp Hoess, Joran Deschamps, Ulf Matti, Bianca Nijmeijer, Vilma Jimenez Sabinina, Jan Ellenberg, Ingmar Schoen, and Jonas Ries. Real-time 3d single-molecule localization using experimental point spread functions. *Nature methods*, 15(5):367, 2018.

[143] Mario Raab, Ija Jusuk, Julia Molle, Egbert Buhr, Bernd Bodermann, Detlef Bergmann, Harald Bosse, and Philip Tinnefeld. Using DNA origami nanorulers as traceable distance measurement standards and nanoscopic benchmark structures. *Scientific reports*, 8(1):1780, 2018.

[144] Francesca Cella Zanacchi, Carlo Manzo, Angel S Alvarez, Nathan D Derr, Maria F Garcia-Parajo, and Melike Lakadamyali. A DNA origami platform for quantifying protein copy number in super-resolution. *Nature methods*, 14(8):789, 2017.

[145] Christian Sieben, Kyle M Douglass, Paul Guichard, and Suliana Manley. Super-resolution microscopy to decipher multi-molecular assemblies. *Current opinion in structural biology*, 49:169–176, 2018.

[146] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[147] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in neural information processing systems*, pages 3844–3852, 2016.

[148] Mohamed El Beheiry and Maxime Dahan. ViSP: representing single-particle localizations in three dimensions. *Nature methods*, 10(8):689, 2013.

[149] Carmen L Krüger, Marie-Theres Zeuner, Graeme S Cottrell, Darius Widera, and Mike Heilemann. Quantitative single-molecule imaging of TLR4 reveals ligand-specific receptor dimerization. *Sci. Signal.*, 10(503):eaan1308, 2017.

[150] Tomáš Lukeš, Daniela Glatzová, Zuzana Kvíčalová, Florian Levet, Aleš Benda, Sebastian Letschert, Markus Sauer, Tomáš Brdička, Theo Lasser, and Marek Cebecauer. Quantifying protein densities on cell membranes using super-resolution optical fluctuation imaging. *Nature communications*, 8(1):1731, 2017.

[151] Maarten W Paul, H Martijn de Gruiter, Zhanmin Lin, Willy M Baarends, Wiggert A van Cappellen, Adriaan B Houtsmuller, and Johan A Slotman. Smolr: visualization and analysis of single-molecule localization microscopy data in r. *BMC bioinformatics*, 20(1):30, 2019.