

单分子定位显微镜数据分析方法的研究进展

杨 研, 谢 红

上海理工大学智能科技学院, 上海

收稿日期: 2025年1月13日; 录用日期: 2025年2月20日; 发布日期: 2025年2月28日

摘 要

单分子定位显微镜(Single-Molecule Localization Microscopy, SMLM)是一种突破传统光学显微镜分辨率极限的重要技术,能够在纳米尺度下提供单分子分辨率的定位信息。这项技术广泛应用于生物学和材料科学领域,为揭示复杂生物结构和分子相互作用提供了前所未有的细节。然而,SMLM数据具有高度稀疏性、非均匀性以及高维度等独特特性,数据定量和解析方法尚未完全跟上这一技术的进步,导致其在实际应用中的数据分析面临诸多挑战。在SMLM数据分析中,主要难点包括:如何准确识别并量化分子位置,如何解析复杂的分子分布模式,如何从噪声数据中提取有意义的生物信息,以及如何高效处理海量数据以确保分析的速度与精度。这些难点极大限制了SMLM技术在高精度定量研究中的应用。近年来,针对这些挑战,研究者开发了多种分析方法,例如空间描述性统计(用于描述分子分布特征)、聚类和分割算法(用于识别分子聚集模式),以及几何分析方法(用于解析分子结构形态),深度学习。这些方法在提高定位精度、揭示空间分布规律以及量化复杂生物结构等方面取得了显著进展。然而,现有方法仍存在一定的局限性,例如在处理大规模数据集时计算效率不足、对复杂分布模式的解析能力有限,以及对噪声和伪影的鲁棒性不足等问题。本文通过系统梳理SMLM数据分析的现有方法,总结了各方法的优势和适用场景,同时指出其局限性及改进方向。我们希望为研究者提供清晰的思路,帮助其根据研究目标选择最适合的分析策略,从而进一步提高SMLM数据分析的准确性和可靠性。

关键词

单分子定位显微镜, 数据分析, 聚类算法, 空间描述性统计, 几何形状分析, 深度学习

Research Progress of Data Analysis Methods for Single Molecule Localization Microscopy

Yan Yang, Hong Xie

School of Artificial Intelligence Science and Technology, University of Shanghai for Science and Technology, Shanghai

Received: Jan. 13th, 2025; accepted: Feb. 20th, 2025; published: Feb. 28th, 2025

文章引用: 杨研, 谢红. 单分子定位显微镜数据分析方法的研究进展[J]. 运筹与模糊学, 2025, 15(1): 613-636.
DOI: 10.12677/orf.2025.151055

Abstract

Single-Molecule Localization Microscopy (SMLM) is an important technique that breaks the resolution limit of traditional optical microscopy, and can provide single-molecule resolution localization information at the nanoscale. The technique is widely used in biology and materials science, providing unprecedented detail to reveal complex biological structures and molecular interactions. However, SMLM data has unique characteristics such as high sparsity, non-uniformity and high dimensionality, and data quantitative and analytical methods have not fully kept up with the progress of this technology, resulting in many challenges for data analysis in practical applications. In SMLM data analysis, the main difficulties include: how to accurately identify and quantify the molecular position, how to analyze complex molecular distribution patterns, how to extract meaningful biological information from noisy data, and how to efficiently process massive data to ensure the speed and accuracy of analysis. These difficulties greatly limit the application of SMLM technology in high-precision quantitative research. In recent years, a variety of analytical methods have been developed to address these challenges, such as spatial descriptive statistics (to describe molecular distribution characteristics), clustering and segmentation algorithms (to identify molecular aggregation patterns), geometric analysis methods (to analyze molecular structure morphology), and deep learning. These methods have made remarkable progress in improving localization accuracy, revealing spatial distribution laws, and quantifying complex biological structures. However, existing methods still have some limitations, such as insufficient computational efficiency in dealing with large-scale data sets, limited ability to resolve complex distribution patterns, and insufficient robustness to noise and artifacts. This paper systematically reviews the existing methods of SMLM data analysis, summarizes the advantages and application scenarios of each method, and points out its limitations and improvement directions. We hope to provide researchers with clear ideas to help them choose the most suitable analysis strategy according to the research objectives, so as to further improve the accuracy and reliability of SMLM data analysis.

Keywords

Single-Molecule Localization Microscopy, Data Analysis, Clustering Algorithm, Spatial Descriptive Statistics, Geometric Shape Analysis, Deep Learning

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

超分辨率显微镜(Super-Resolution Microscopy, SRM)是近年来生命科学领域的一项变革性技术,它突破了传统光学显微镜约 200 纳米的分辨率极限[1]-[3],能够实现亚细胞甚至单分子水平的精细结构成像,与原子力显微镜[4]、电子显微镜[5]-[7]等成像技术相比,超分辨率光学显微镜因其非接触性、对样品损伤较小以及多样化的成像机制等优点,已经成为生物医学研究中不可或缺的重要工具[8]-[10]。在这些技术中,单分子定位显微镜尤为引人注目。通过精确定位单个荧光分子的发射位置,单分子定位显微镜(Single-Molecule Localization Microscopy, SMLM)能够将空间分辨率推至 20 纳米甚至更低[11]-[13],从而为探索细胞内动态过程和纳米结构提供了前所未有的工具。

尽管单分子定位显微镜在结构生物学、细胞生物学等领域展现了广泛的应用前景,其在分子层面提供的高分辨率信息为探索复杂生物学过程和组织结构提供了前所未有的可能性,然而,伴随技术发展的

是数据处理和定量分析中面临的巨大挑战, 这些挑战的根源在于 SMLM 数据的特殊性和复杂性, 因为 SMLM 产生的数据量庞大且稀疏, 包含大量单分子定位信息, 这些信息不仅受到高噪声水平的干扰, 还存在定位误差、背景信号和样品伪影等问题。SMLM 数据处理的主要目的是在单分子分辨率的基础上还原样品的真实结构, 提取关键的生物学信息, 如何从这些复杂的数据中提取出高质量的空间分布、聚集模式以及定量化特征, 是推动 SMLM 在基础研究和实际应用中进一步发展的关键问题之一。提取这些关键的生物学信息, 有助于描述分子分布特征的空间描述性统计, 用于去噪和识别分子聚集模式的聚类分割算法, 也有用于解析分子结构形态的几何分析等处理方法。尽管成像硬件的快速发展显著提高了 SMLM 的成像能力, 但与之相比, 数据分析方法的完善显得相对滞后。现有方法在处理速度、适用性和对噪声的鲁棒性等方面尚有诸多不足。因此, 对现有数据分析方法进行讨论和方法比较, 不仅能够梳理当前技术的优劣, 还能为未来的研究方向提供重要参考。

本综述的目的在于总结和分类目前 SMLM 数据分析的主要方法, 从空间描述性统计、聚类分割、几何分析到深度学习等定量分析方法。我们将详细讨论每种分析方法的原理、适用场景、优缺点, 并针对不同的研究需求提出分析策略的选择指南。希望通过本文, 研究者能够深入理解 SMLM 数据分析的复杂性, 进而优化各类生物学实验的分析流程和结论的可靠性。

2. 单分子定位显微镜的原理

2.1. 荧光分子的光开关原理

单分子定位显微镜的核心是可切换的荧光染料[14]、荧光蛋白[15]或荧光量子点[16], 这些分子可以在不同的光照条件下在“亮态”和“暗态”之间切换。通常, 这些荧光分子在未激活状态下不发光, 激活后发出荧光, 这种开关行为可以通过化学或光物理机制实现[17], 如图 1 所示。在单个荧光分子持续发光过程中, 因其内在光物理或光化学过程而导致的发光强度随机变化现象称之为荧光闪烁现象。与传统荧光显微镜需要稳定且持续的荧光发射不同, SMLM 通过化学或光物理手段有效控制荧光闪烁, 使荧光分子能够在“亮态”和“暗态”之间灵活切换。

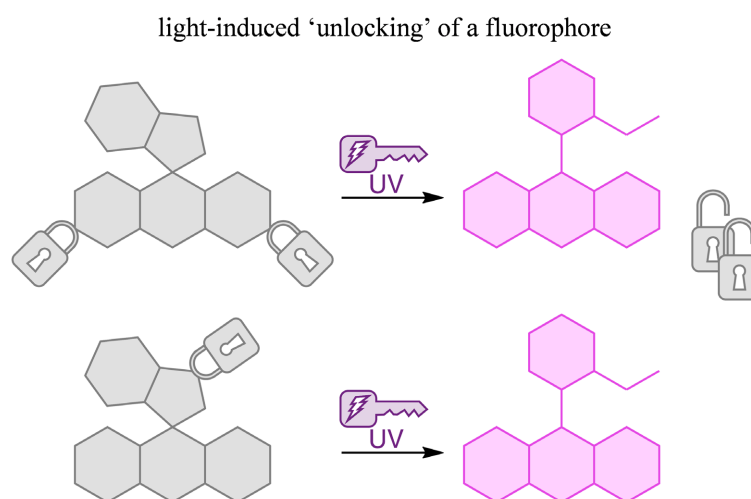


Figure 1. Light-induced fluorescent molecules [17]

图 1. 光诱导的荧光分子[17]

荧光蛋白的光开关机制[18]是指在特定波长光的作用下, 荧光蛋白能够从非荧光态转变为荧光态, 根据转换特性可以分为可逆和不可逆两种, 即荧光蛋白是否能够在交替波长光的作用下, 在荧光态和非荧

光态之间反复切换。荧光染料的光开关机制[19]通过三线态、负电激子态等机制, 实现荧光的可逆转换。荧光量子点则是基于俄歇复合或隧穿效应的闪烁特性控制荧光的开关[20]。SMLM 的核心在于利用荧光分子的光开关特性, 通过调控荧光分子的激发与灭活, 使每一时刻仅有少量分子发光。这种稀疏激发策略确保了空间上重叠的荧光信号在时间上分离, 一次少量稀疏分布的荧光信号为超分辨成像奠定了基础。

2.2. 单分子定位原理

根据阿贝衍射理论[21], 传统光学显微镜的分辨率极限为横向约 200 纳米, 轴向约 500 纳米, 对于理想无像差的光学系统, 光通过一个光圈(如显微镜或相机镜头的光圈)后, 由于衍射效应, 一个点光源不会成像为一个理想的点, 而是会扩展成一个具有明暗环状结构的图案, 中心部分最亮, 而点扩散函数(Point Spread Function, PSF)正是艾里斑的数学描述, 光学系统的分辨率是指该系统能够分辨两个相邻点或物体细节的最小间距, 可以用 PSF 的半高全宽(Full Width at Half Maximum, FWHM)来衡量 PSF 大小和光学系统分辨率。

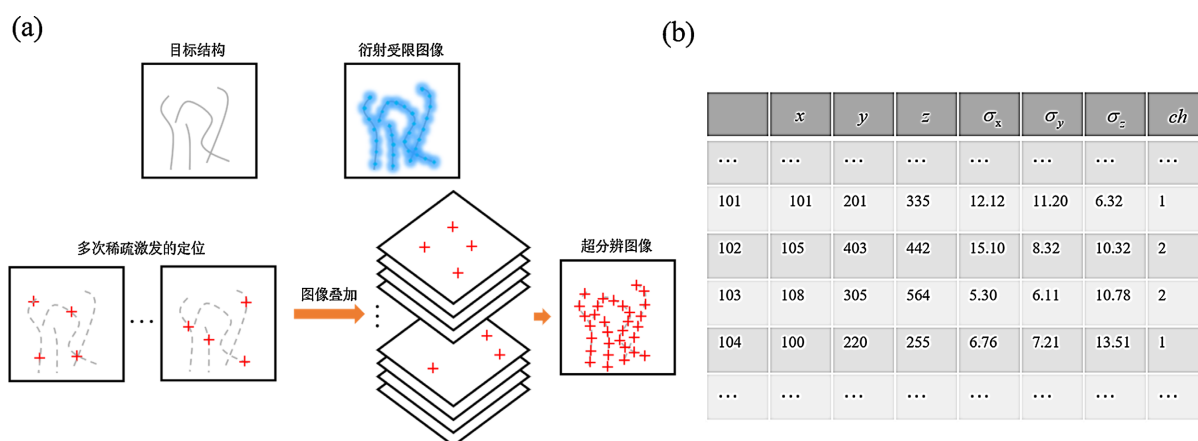


Figure 2. Single molecule localization principle (a) Schematic diagram of single molecule localization principle; (b) Components of coordinate based SMLM data

图 2. 单分子定位原理(a) 单分子定位原理示意图; (b) 基于坐标的 SMLM 数据的组成成分

在 SMLM 中, 用荧光分子探针标记样品后, 由于稀疏激发, 每一帧子图像中只有少量稀疏分布的单分子发射荧光, 通过精确测量每个荧光分子的发光位置, SMLM 能够以极高的精度(通常为 20~50 纳米)确定单个分子的坐标。这一过程通常涉及到将荧光分子的点扩散函数与图像中的荧光信号进行匹配, 利用统计学方法拟合出单个分子的确切位置, 这个过程称为定位。定位精度与 PSF 的标准偏差 s 和检测到的光子数量 N 成比例, 在低背景和小像素尺寸的情况下, 可以近似为 s/\sqrt{N} [22]。重复进行多次成像, 每次激活不同的荧光分子。通过收集到的所有激活分子的位置数据, 研究人员能够重建出高分辨率的整体图像, 如图 2(a)所示。这种图像不仅超越了传统显微镜的衍射极限, 而且能够提供细胞内结构的详细信息。例如, 随机光学重建显微镜 (Stochastic Optical Reconstruction Microscopy, STORM) [23] [24]、光激活定位显微镜(Photo-Activation Localization Microscopy, PALM) [25]和 DNA-Point Accumulation for Imaging in Nanoscale Topography (DNA-PAINT)技术[26]~[28]等其他技术[29] [30]都使用了这样的原理来成像, 其中 STORM 主要是使用光转换荧光染料(Cy3-Cy5 分子对)标记样品, 通过红光(633 nm)和绿光(532 nm)的交替使用实现荧光分子 Cy5 的 ON-OFF 态转换, 首先 Cy5 分子在高强度 633 nm 红光下被激发至 ON 态并迅速变为 OFF 态, 再切换成低强度的 532 nm 绿光使少量随机的 Cy5 分子恢复 ON 态, 最后再次切换至

633 nm 红光使其转变为 OFF, 循环往复至大部分荧光分子被漂白, 即可得到所有单分子定位。PALM 原理与 STORM 十分相似, 不同的是它是利用光激活荧光蛋白(PA-FP)标记样品, 该蛋白需要 405 nm 激光激活, 得到少量稀疏分布的荧光分子, 再利用 561 nm 激光使这些荧光分子处于 ON 态, 并将它们漂白, 多次循环并将得到的定位叠加即可得到超分辨图像。DNA-PAINT 技术则基于两条互补短链 DNA 之间的可逆瞬时结合, 通过这种动态的结合和解离过程产生荧光信号的开关效应。无论使用哪种方法, 生成的 SMLM 数据都是一个多维数组, 由每个定位的空间坐标和其他信息所组成, 这里的其他信息可以是定位精度, 亮度等。并且如果是更高级的成像, 例如 3D SMLM 和双色成像, 那么这个多维数组可以进一步扩展, 包括 z 轴坐标信息和对应的颜色通道信息, 如图 2(b)所示。

SMLM 的分辨率和质量主要受三个因素限制: 一是光子数及其统计学不确定性, 光子数越多, 定位精度越高; 二是荧光分子的密度与开关特性, 稀疏激活的荧光分子以及理想的开关性能有助于实现高分辨率成像。这两者决定了最终重建图像的精度和清晰度; 三是结构不仅需要高密度标记, 还需要检测大多数荧光团。这需要较长的测量时间, 从而导致在时间和空间分辨率之间进行权衡[31]。因此, SMLM 主要用于固定细胞或非常缓慢的细胞过程。活细胞成像则需要选择适合的荧光探针, 例如文章中提到的一种新型的近红外硅罗丹明(SiR)荧光探针[32][33]。另外要控制光照强度和曝光时间, 优化光学系统以减少光漂白和光毒性[34], 同时在时间分辨率和空间分辨率之间进行权衡。

3. 单分子定位显微镜中的数据分析

单分子定位显微镜的高分辨率使得我们可以看见传统光学显微镜无法看见的结构。为了更容易观察分子结构, 通常用高斯渲染方法[35]将 SMLM 实验中得到的定位数据(即每个荧光分子的坐标)转换成可视化图像, 生成类似传统光学显微镜图像的效果。但是有研究指出高斯渲染可能会导致分辨率损失高达 1.44 倍[35], 因为这种高斯渲染是通过计算每个子像素中的定位数量或者将每个定位渲染为大小与定位精度成正比的高斯分布来实现的, 它在原始数据的基础上引入了额外的模糊化处理, 这种处理可能会导致细节丢失, 产生虚假的结构。Nicovich 等人[36]强调 SMLM 数据不仅仅是一种超分辨率的图像, 还是多维数据集, 其包含了分子定位的空间坐标, 荧光强度等信息, 这种数据集能够提供更深入的见解, 不仅可以揭示分子结构, 还能解析分子之间的相互作用。因此直接使用 SMLM 原始定位数据来数据分析的算法, 将不会导致信息丢失, 例如径向分布函数, Ripley 的 K 函数, 聚类算法, 几何分析等方法对 SMLM 数据分析具有重要的意义和潜力, 如图 3 所示, 展示了应用于 SMLM 数据的数据分析方法。在数据分析时, 同时需要考虑分子标记物所带来的影响, 例如, 单分子使用分子标签(如 Halo tag)标记, 荧光在分子标签上, 是否需要考虑分子的运动和柔性。单分子使用抗体标记, 荧光在抗体上, 荧光间接显示分子位置, 分子之间是否存在位阻效应, 分子标记成功率等问题, 在面临这些问题时, 我们需要选择合适的分析方法和参数去处理数据。接下来我们将讨论单分子定位显微镜中的数据分析方法。

3.1. 空间描述性统计

空间描述性统计方法在单分子定位显微镜数据的分析中起到了至关重要的作用, 特别是在量化蛋白质聚集和分子分布的空间特征时。Ripley 的 K 函数和 L 函数[37]被广泛应用于不同尺度上评估蛋白质的聚集性, 能够揭示聚集体的大小和密度。径向分布函数(Radial Distribution Function, RDF) (即 $g(r)$ [38])则提供了一种通过距离依赖的相关性来识别蛋白质或分子聚集的概率分布的方法。这些方法具有无参数化、适应性强的优点, 能够用于多种实验条件和数据集, 并且都已被用于蛋白质聚集分析来实现对细胞内复杂过程的定量研究[37]-[39]。鉴于它们在分析 SMLM 数据方面的广泛应用, 我们接下来将详细介绍径向分布函数 RDF 和 Ripley 的 K 函数。

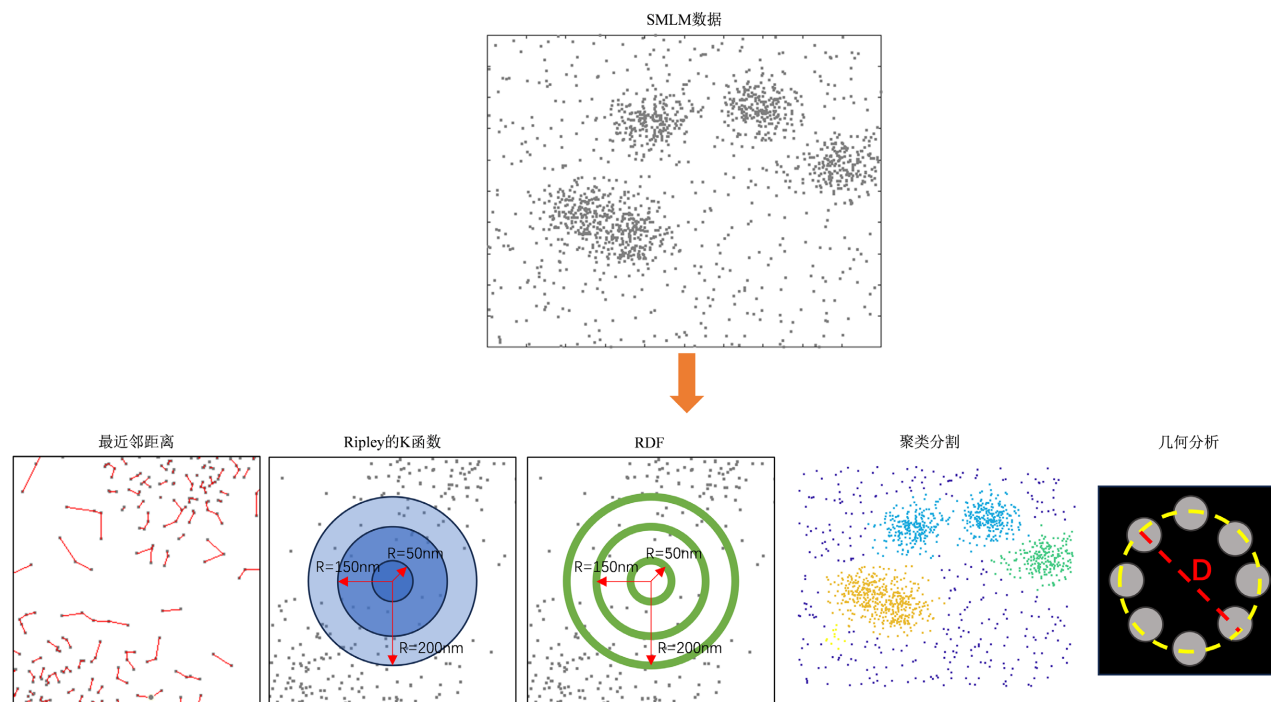


Figure 3. Data analysis methods applied to SMLM data

图 3. 应用于 SMLM 数据的数据分析方法

3.1.1. 径向分布函数 RDF

RDF 被广泛应用于 SMLM 点云数据, 它通常用于分析 2D 数据, 也可以扩展到 3D 空间数据, 以评估多个空间尺度的局部密度变化和聚类趋势。其过程是计算在距参考点 r 到 $r + dr$ 这样的同心环里的点数 N , 即计算每个同心环内的平均定位密度, RDF 由公式(1)给出。

$$g(r) = \frac{N}{\rho 4\pi r^2 dr} \quad (1)$$

其中 ρ 是体系的平均点密度, 即单位体积内的平均点数。最后通过拟合高斯函数或指数函数来确定聚类的大小或密度分布[38] [40], 如图 4 所示。其在数学上对应概率密度函数。蛋白质或其他生物分子的聚类往往具有关键的生物学功能, 如信号转导、受体激活和细胞通讯。正如研究中所强调的, RDF 通过揭示不同尺度上的密度波动, 帮助量化这些现象, 使研究人员能够区分随机分布和有生物学意义的聚类[38] [40]。RDF 的一个关键优势是其模型独立性, 这使得它能够在不预先假设簇的大小或形状的情况下检测聚类。

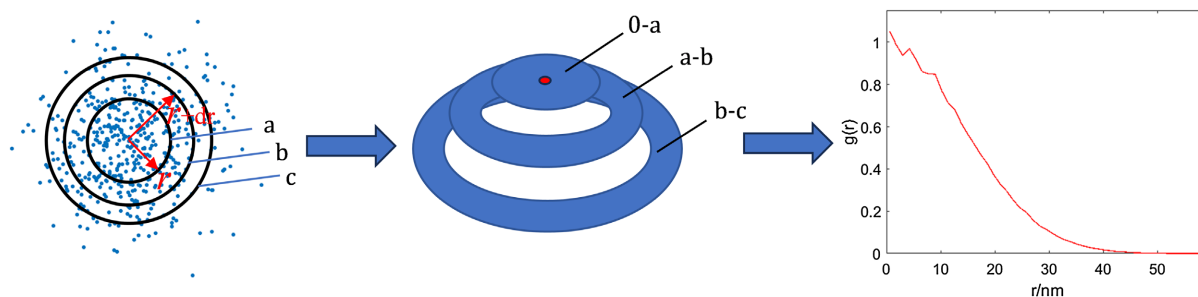


Figure 4. Schematic diagram of radial distribution function principle

图 4. 径向分布函数原理示意图

例如在 Sengupta 等人[38]的研究中, 径向分布函数 RDF($g(r)$)通过 PC-PALM 方法应用于分析质膜蛋白质在纳米尺度上的空间组织和聚集。该方法通过计算配对相关函数, 量化了在给定分子一定距离处找到第二个分子的可能性, 从而揭示了蛋白质是随机分布还是形成聚集, 他们通过计算 $g(r)$ 来评估像 GPI 锚定蛋白(PAGFP-GPI)等蛋白质在质膜上的空间分布。随机分布的 $g(r)$ 值为 1, 而聚集则导致 $g(r)$ 值大于 1, 表示蛋白质组织呈非随机分布, 他们基于 RDF 的分析还追踪了膜修改(如去除胆固醇或加入鞘磷脂酶)如何影响 PAGFP-GPI 等蛋白质的聚集。

3.1.2. Ripley 函数

公式 Ripley 的 K 函数能够定量描述某一给定距离内某个点附近的点的密度, 从而识别聚集特征。K 函数和它的变形 L 函数可以很好地揭示不同空间尺度上的蛋白质聚集模式[41] [42]。它和 RDF 一样可以用于分析 2D 数据和 3D 空间数据中某一点附近的点(如分子或蛋白质)的分布情况[42] [43], 并且在数学上与 RDF 互补, 对应于累积分布函数。通过计算每个点在不同半径范围内邻居点的密度, K 函数可以量化点的聚集程度或分布均匀性。理论上, K 函数可以由公式(2)给出[37] [44]-[48]。

$$K(r) = \frac{A}{n^2} \sum_{i=1}^n \sum_{j=1}^n I(d_{ij} < r) \quad (2)$$

对于需要其中 A 是分析区域的面积, n 是分析区域内的点的数量, d_{ij} 是第 i 个点和第 j 个点之间的距离, r 为半径, $I(d_{ij} < r)$ 是指当距离 d_{ij} 小于半径 r 时取值为 1, 否则为 0。这本质上是计算被以每个点为中心的同心环所环绕的其他点的数目[49]。在分析时, 为了便于解释, 通常将 K 函数线性化转化为 L 函数[50], 目的是让其在完全随机分布(Complete Spatial Randomness, CSR)的情况下呈现为一条直线。L 函数由公式(3)给出。

$$L(r) = \sqrt{K(r)/\pi} \quad (3)$$

对于随机分布的点, $L(r) = r$, 因此将 L 函数转化为 H 函数[51], H 函数由公式(4)给出

$$H(r) = L(r) - r \quad (4)$$

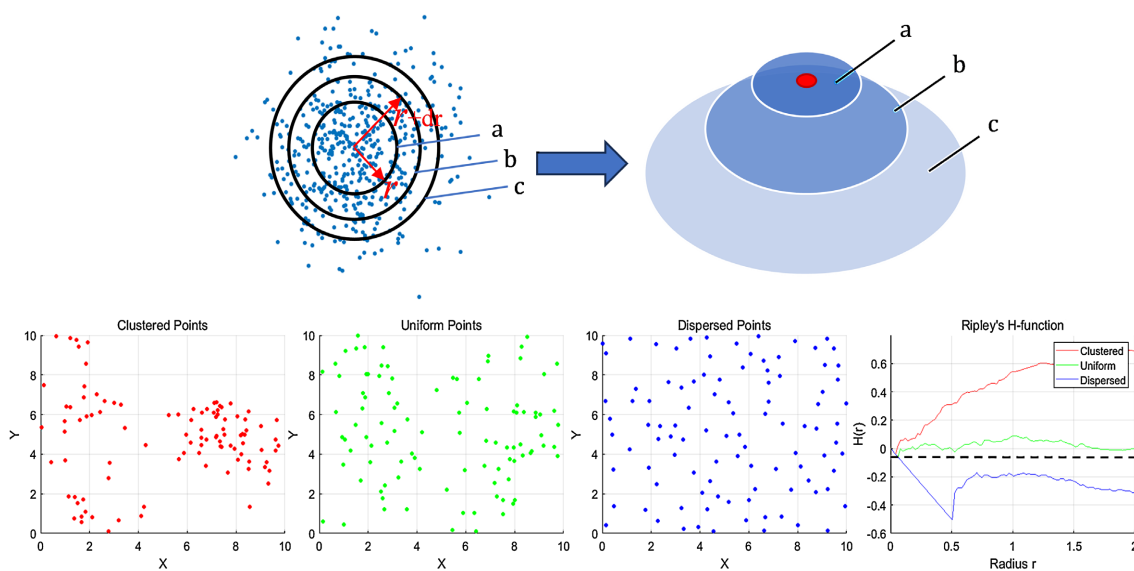


Figure 5. Schematic diagram of the principle of Ripley's K function

图 5. Ripley 的 K 函数原理示意图

因此对于 CSR 分布的点, H 函数值为 0; H 函数大于 0 的则点的分布呈现出聚集性, 反之呈现稀疏分布。图 5 中显示了三种点分布和相应的 Ripley H 函数曲线。对于均匀分布的点, H 函数的曲线在 0 附近波动, 对于聚类点, 则在 0 之上, 对于分散点, 则在 0 之下。在 Ehrlich 等人[51]的研究中, 他们利用 Ripley 的 K 函数分析网格蛋白涂层小体在形成过程中空间分布和聚集情况, K 函数揭示了网格蛋白聚集体的形成位置没有固定的模式, 这与假设网格蛋白涂层小体是随机启动的, 并通过稳定化(可能是通过货物的捕获)来确保成功内吞的观点相一致; 分析还突出了短命的网格蛋白聚集体的存在, 这些聚集体可能代表流产的网格蛋白涂层小体, 这些小体未能完成生命周期并过早地解离, Ripley 的 K 函数帮助识别了这些流产事件发生的区域, 为它们失败的机制提供了线索, 在这项研究中, Ripley 的 K 函数的分析中获得的见解对于理解细胞内吞过程的精细调控具有重要意义。另外, 还有研究提出一种模式提取方法(pattern extraction from relative positions of localizations, PERPL), 用于分析 3D 稀疏数据集中的纳米级别的有序结构, 不仅可以提取簇的大小, 还可以提取更复杂的重复结构特征, 例如 NPC 的半径和对称性[52]。

RDF 描述了系统中分子或粒子的空间分布特性, 主要用于分析粒子之间的局部排列和距离分布。适用于测量分子聚集情况或分析分子间的局部关系。而 Ripley 的 K 函数用于分析点过程的空间相关性, 特别是用于评估特定尺度下的空间聚集程度。它可以提供多个尺度下的空间模式, 适用于多尺度数据分析, 尤其在分析分子分布和模式时具有优势。这两种方法有不同的原理和功能, 我们对这两种方法使用模拟数据和真实数据进行了性能比较, 如图 6 所示, 在模拟数据上, 分析 RDF 和 Ripley 的 K 函数对不同数据模式(如随机分布、均匀分布、簇集分布等)的响应能力, Ripley 的 K 函数在应对不同数据模式时能够表现出明显的特征, 而 RDF 没有明显捕捉到不同模式的特征。同时使用实验中获得真实数据集进行验证, 展示两种方法对真实数据的分析结果, 反应了该数据的聚集性。因此, RDF 适用于局部分布特征的分析, 例如分子聚集、离散点的局部结构等。其局限性可能表现为无法很好地描述全局结构或处理大规模数据时的计算复杂性。而 Ripley 的 K 函数适用于多尺度的空间模式分析, 尤其对于分析群体聚集等现象具有优势。但它在处理非常复杂的空间模式或极高密度数据时, 可能会受到一些计算上的限制。

在数据分析时, 若需讨论研究对象的聚集程度变化或者周期性结构的分布, 上述方法具有良好的适用性, 他们都被用于研究细胞表面蛋白异质性, 例如 T 细胞质膜片上的 T 细胞受体等蛋白的分布[38]。另外在使用空间描述性统计进行定量分析时, 多个因素会影响数据的准确性。这些因素包括单分子荧光的闪烁、检测效率的限制、分子定位的不确定性等[53][54]。并且空间描述性统计只能描述全局的点分布特性, 无法直接识别或区分单个聚集体的具体位置或形状, 因此难以对特定的聚集体进行单独分析, 无法细致区分局部区域的特征或差异。然而, 可以使用聚类分割算法来识别分析复杂的局部分子聚集现象。

3.2. 聚类分割算法

在 SMLM 数据中, 我们经常观察到高密度聚集的点云。这些聚集的点云可以通过聚类算法进行分割与识别, 从而形成不同的簇。与空间描述性统计相反, 分割允许对同一视野内的各个簇进行研究和比较。然后可以很容易地计算基本参数, 例如作为大小度量的坐标的标准差、各个簇中的定位数量或各个簇的密度。同时, 通过对分割得到的簇进行更复杂的分析, 我们可以进一步探讨其尺寸分布、最近邻距离分布等特征, 从而揭示分子间的相互作用关系。接下来我们介绍一些聚类分割方法。

3.2.1. 基于密度的聚类算法

单分子定位显微镜在纳米尺度的分辨率下能够精确定位单个分子, 从而揭示超出传统光学显微镜能力范围的复杂生物结构。然而, 由于数据通常表现为大量噪声点以及非规则的空间分布, 如何有效地从这些点云数据中提取有意义的群体或结构成为了一个重要挑战。为了解决这一问题, 聚类算法在 SMLM

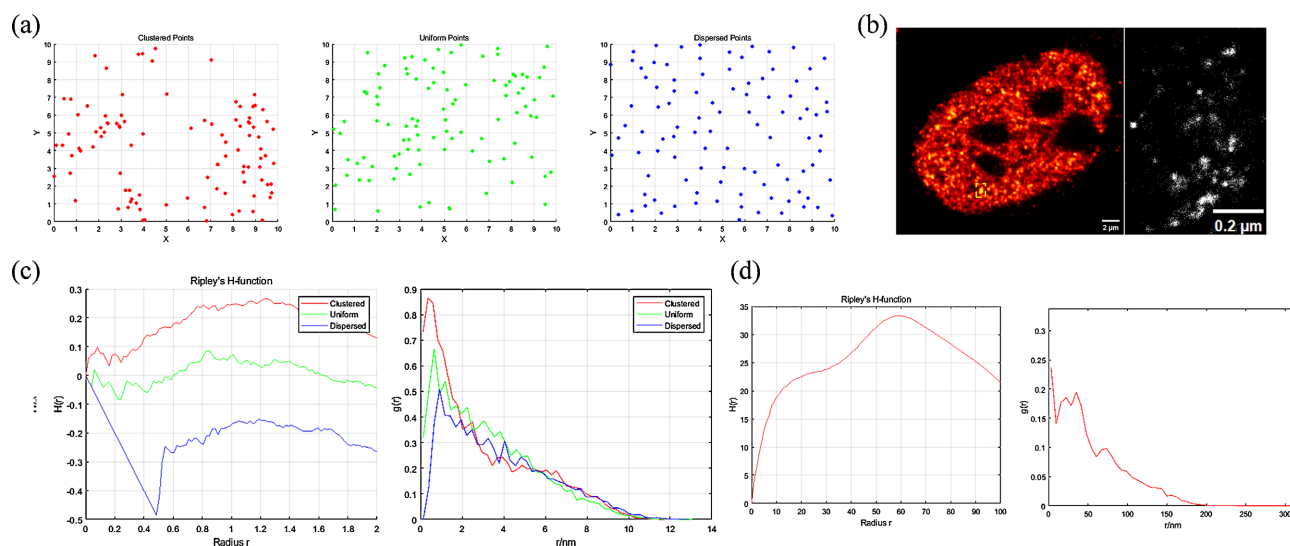


Figure 6. Performance comparison between Ripley's K function and RDF (a) Simulate SMLM data; (b) Real SMLM data; (c) The results of Ripley's K function and RDF on simulated data; (d) Ripley's K function and the results of RDF on real data

图 6. Ripley 的 K 函数和 RDF 的性能比较(a) 模拟 SMLM 数据; (b) 真实 SMLM 数据; (c) Ripley 的 K 函数和 RDF 在模拟数据上的结果; (d) Ripley 的 K 函数和 RDF 在真实数据上的结果

数据分析中得到了广泛应用, 特别是基于密度的聚类算法因其能够处理不规则形状的群体并对噪声具有较好的鲁棒性, 受到了广泛的关注。在众多基于密度的聚类算法中, **Density-Based Spatial Clustering of Applications with Noise (DBSCAN)** [55] [56]是最常用且经典的一种, 这种算法对有大量噪声和具有不同密度簇的数据具有良好的适用性, 并且不需要预先定义簇的数量。DBSCAN 通过定义高密度区域(即具有一定数量密集分布点的区域)作为聚类的核心, 能够识别出具有任意形状的聚类, 并有效地排除噪声点。其工作原理基于两个关键参数: Eps (ϵ , 表示聚类的半径范围)和 $MinPts$ (最小点数, 指形成核心点所需的邻域内点数)。当某个点的邻域(以 Eps 为半径)包含至少 $MinPts$ 个点时, 该点被定义为核心点, DBSCAN 通过扩展这些核心点的邻域来形成聚类, 直到到达 $MinPts$ 条件不再成立的边界点。否则, 该集群将被视为噪声。图 7 中展示了 DBSCAN 聚类方法的工作原理, 以及它的各个参数的含义。

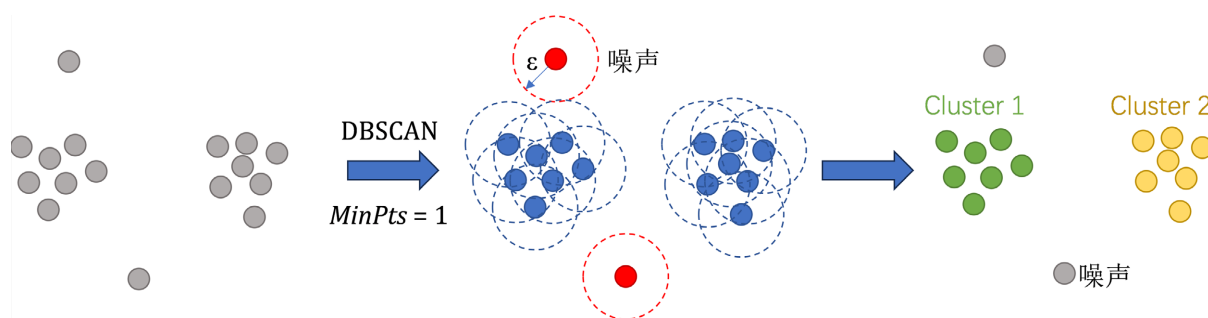


Figure 7. DBSCAN working principle diagram

图 7. DBSCAN 工作原理示意图

DBSCAN 的显著优势在于它不需要预先知道数据中的簇的数量, 并且能够识别出任意形状的簇, 这使得它在处理单分子定位显微镜中常见的复杂、非规则形状的结构时尤为有效。此外, DBSCAN 可以自动识别噪声点, 将那些不属于任何聚类的孤立点归为噪声, 从而进一步增强了对含有大量噪声数据的处理能力。

尽管 DBSCAN 在处理单分子定位显微镜(SMLM)数据中表现出了诸多优势, 尤其是在识别任意形状的聚类以及有效处理噪声方面, 但它并非没有局限性[57]。在实际应用中, 特别是面对大规模的 SMLM 数据集, DBSCAN 面临着一些不可忽视的挑战。首先, DBSCAN 的性能高度依赖于两个关键参数: Eps (半径)和 MinPts (最小点数)。然而, 选择这些参数往往具有主观性, 难以通过自动化方式确定最佳值。研究表明, DBSCAN 对这两个参数的选择非常敏感, 过高或过低的 Eps 和 MinPts 值可能导致错误的聚类结果, 尤其是当数据中存在显著噪声时。这种参数选择的依赖性在处理大规模数据集或复杂的 SMLM 图像时尤为突出, 需要反复实验以手动调整参数, 从而增加了分析的时间成本。其次, DBSCAN 在处理高密度噪声点时表现出了一定的局限性。SMLM 数据中, 荧光标签的闪烁现象和随机背景噪声会导致伪聚类的出现, 而 DBSCAN 有时难以有效区分这些伪聚类和真实的蛋白聚集体。此外, 由于 SMLM 数据中荧光信号的多次定位误差, 单个蛋白或荧光标签可能被错误识别为多个密集的定位点, 从而形成伪聚类。

在算法计算复杂度方面, 尽管相比于其他算法, DBSCAN 在大规模数据集上的效率较高, 但它的时间复杂度在最优情况下也为 $O(n \log(n))$ [57], 并且当分析超大规模 SMLM 数据时, 可能需要耗费数小时的计算时间。因此, 为了解决 DBSCAN 的局限性, Mazouchi 和 Milstein [57]提出了一种基于密度的聚类算法, 即用于定位的快速优化聚类算法(fast optimized cluster algorithm for localization, FOCAL)。它通过网格化处理方式对定位数据进行聚类, 并且它的时间复杂度是线性的 $O(n)$, 这显著加快了数据处理的速度。相比 DBSCAN 需要同时确定 Eps 和 MinPts 两个参数, FOCAL 仅需优化一个参数, 即最小密度阈值。但它也仍然存在一些局限性, 在需要精确检测小型或紧凑聚类时, 如果选择的网格大小过大, 可能会丢失部分精细的空间信息。此外, FOCAL 在高度密集的噪声环境中可能表现不佳。FOCAL 是一种对二维 SMLM 数据的分析方法, 为了能够分析三维 SMLM 数据, FOCAL3D [58]是 FOCAL 方法在三维数据上的扩展。

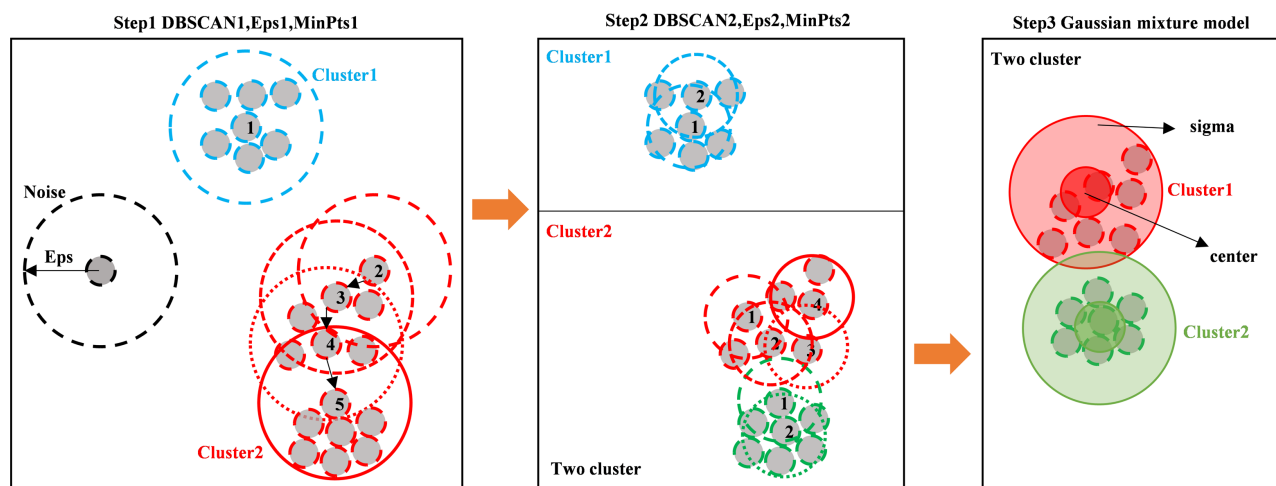


Figure 8. Schematic diagram of working principle of three-step clustering algorithm
图 8. 三步聚类算法工作原理示意图

在 SMLM 数据聚类分析中, DBSCAN 可以单独使用[57][59]-[61], 也可以与其他聚类算法结合[62]-[68], 形成更复杂的多步算法来改善分析效果。Jasmin K. Pape[68]等人提出了一种基于 minimal photon fluxes (MINFLUX) [68]-[72]定位数据的三步聚类算法, 结合了著名的 DBSCAN 算法和使用高斯混合模型的期望最大化方法, 用于将 2D 或 3D 定位云分配给荧光团位置, 该算法的第一步通过一个较大的搜索半径(大于定位精度)应用 DBSCAN 来预先分配定位点, 并识别噪声点。在第二步中, 算法对第一步中识别

出的分子进行进一步聚类,应用一个较小的搜索半径再次执行 DBSCAN。这一步通过缩小搜索范围来提高精度,识别更紧密的聚类。最后,使用高斯混合模型将第二步中的结果进一步优化,将定位点精确地分配给特定的分子,如图 8 所示。这个三步法的优势在于无需预先知道分子或聚类的数量,同时利用了每个分子定位云的高斯形状,从而提供了比单一使用 DBSCAN 更优的结果。并且 Jasmin K. Pape [68]等人利用 DBSCAN 的混合算法分析了位于线粒体内部的 MICOS 复合体的 Mic60 亚单位的聚集模式,发现其在线粒体嵴连接周围形成 40 到 50 纳米直径的环状结构。这些结果结合电子显微镜数据表明,这种环状分布可能是其稳定嵴连接的分子基础,并且通过该算法对不同亚单位(Mic10、Mic19 和 Mic60)间的相对位置进行了量化。结果显示, Mic19 与 Mic60 的空间关系更紧密,而 Mic10 和 Mic60 的相对位置具有较大的变异性。这种分析揭示了 MICOS 复合体结构的异质性和功能相关性,研究中通过结合 DBSCAN 的多层次分析,剔除了定位精度低或孤立的点,确保了分析的可靠性,能够在复杂的 3D 分子分布中区分个体分子位置,例如,在 200 纳米直径的模拟线粒体中精确定位密集分子聚集。此外,3D MINFLUX 技术 [68] [71] [72] 的出现为 SMLM 数据分析提供了突破性的方法,尤其是在解决投影伪影和伪簇问题上表现出色。传统的 SMLM 分析面临一个挑战,即定位精度的各向异性可能导致伪影,从而产生虚假的聚类。MINFLUX 则通过其各向同性的纳米级定位精度(可达到个位数纳米),有效消除了这些投影伪影和避免了由于同一分子多次定位产生的伪聚类现象。而在数据分析上,可以采用去重算法、校正漂移算法以及基于相关性的分析方法(对相关、自相关、互相关、共定位)对 SMLM 数据预处理来消除伪聚类现象。

3.2.2. 密度无关聚类算法

在单分子定位显微镜(SMLM)数据分析中,聚类算法是识别和分离分子簇的关键工具。尽管基于密度的聚类方法在处理高密度区域时表现出色,但它们通常在低密度或密度变化较大的区域遇到困难。除了 DBSCAN 和 FOCAL 这样的基于密度的聚类算法之外,还有密度无关聚类方法,例如 k-means 算法 [73] 也是聚类分析中常用的方法之一,特别是在处理大规模、高维度数据时, k-means 因其计算效率高和实现简单而受到广泛应用,对于各类簇以相似密度均匀分布、没有复杂形状、簇数量相对明确的数据,并且只需求得到初步的聚类信息, k-means 具有良好的适用性。k-means 通过将数据点划分为 k 个簇,每个簇通过一个质心(簇的平均位置)来定义。算法通过反复迭代,将数据点分配给距离最近的质心,然后重新计算质心,直到收敛,如图 9 展示了 k-means 的工作原理。k-means 的优势在于它的计算复杂度较低,能够快速处理大量定位数据,这在 SMLM 实验数据集中尤为关键。k-means 的计算复杂度较低,能够快速处理大量定位数据,这在 SMLM 实验数据集中尤为关键。其具有灵活的距离度量,在 SMLM 数据中,可以通过调整距离度量方式(如欧氏距离或曼哈顿距离)来适应不同类型的分子分布。然而,它也有一定的局限性,在实际 SMLM 数据中,分子簇的数量通常未知,需事先设定 k 值。这种需求增加了聚类参数优化的难度,可能影响最终结果的准确性。初始中心点的选择对结果有较大影响, k-means 可能会陷入局部最优解。SMLM 数据中可能存在噪声点, k-means 的距离最小化机制可能会导致噪声点被错误聚类,从而影响聚类中心的准确性。根据这些局限性, k-means 也可以结合其他算法形成多步算法来改善结果,可以用 DBSCAN 对 SMLM 数据初步聚类并去噪,再用 k-means 聚类或许会取得不错的效果。

在 Jaroslaw Jacak 的研究 [74] 中, K-means 聚类算法被用来分析大脑组织的 dSTORM 数据和荧光显微镜图像,重点是血清素受体相对于胶质细胞的分布。通过将图像分割成离散的强度层, K-means 有效地基于像素亮度对区域进行了分类。这种分割方法使得研究血清素受体与胶质结构之间的空间关系变得更加精确。该研究不仅加深了我们对大脑组织中受体分布的理解,还为在纳米级生物数据分析中使用聚类算法设定了一个先例。

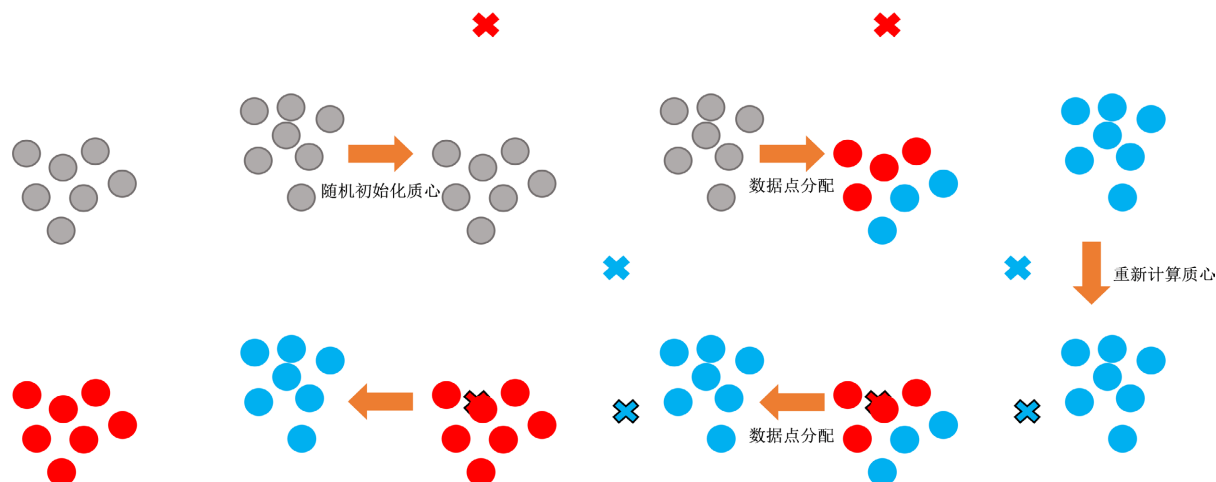


Figure 9. Schematic diagram of working principle of k-means

图 9. k-means 工作原理示意图

3.2.3. 基于 Voronoi 图的聚类算法

Voronoi 图[75]不仅在数学领域有着广泛的研究意义, 还因其独特的几何划分特性在许多应用场景中得到广泛应用, 例如生物学、天文学和城市规划等领域。Voronoi 图是一种将空间划分为多个区域的方法, 每个区域(Voronoi 单元)对应一个数据点, 并包含所有距离该点最近的空间位置。具体来说, 给定一组在空间中分布的点, 每个点的 Voronoi 单元由与其他点的距离关系决定。Voronoi 图的生成是基于“最近邻”原则: 空间中的每个位置都被分配给离它最近的一个点, 从而形成一系列非重叠的多边形区域。图 10(a)显示了 2D 空间中的一组点, 其中每个点可能代表一个分子定位, 它有两个聚集的簇和一些噪声。图 10(b)描述了图 10(a)中所示分子定位的 Voronoi 图。Voronoi 单元在图 10(c)中以不同的颜色显示, 并且越密集的区域有较深的颜色。Voronoi 图被广泛应用于 STORM、dSTORM 和 PLAM 数据的分子聚类分析[76]-[78], 例如基于 Delaunay 三角剖分和 Voronoi 图的方法用于研究心肌细胞中受体的分布, 识别和量化钙信号相关的蛋白簇, 揭示 Ryanodine 受体的聚类尺寸分布和相邻簇的紧密分布, 暗示了钙信号的协同释放机制[77]; 以及通过 Voronoi 分割, 能够识别单个 NPC, 并提供了其直径和密度的量化数据, 这些数据与已有的生物学信息一致[78]; 通过不同染色的蛋白质样本, 例如用 Alexa Fluor-555 和 Alexa Fluor-647 标记的微管蛋白, Voronoi 方法分析了两种标记之间的共定位, 量化了染色区域的重叠程度[78]。Voronoi 图为超分辨率显微镜数据的聚类分析和生物学问题的探索提供了强大的工具, 其在多种生物场景中的成功应用, 证明了其广泛的适用性和潜力。

Voronoi 图的引入为 SMLM 数据提供了一种几何分割的方法[78], 通过构建 Voronoi 单元, 可以有效地估算局部密度, 实现高分辨率数据的精确分割和聚类。Levet 等人[76]和 Andronov 等人[77]提出了使用 Voronoi 图对 SMLM 定位数据进行聚类的方法, 这两项研究都能够根据每个定位的局部密度来划分空间区域, 从而识别分子聚集的高密度区域, 并且通过将它们的 Voronoi 单元区域与空间均匀的参考分布进行比较来自动确定密度阈值对 SMLM 定位数据进行聚类, 越高密度区域的点划分到的 Voronoi 单元越小。这两个方法都适用于从大尺度的细胞水平到小尺度的分子簇级别的多种组织结构, 使其既能研究个体分子聚类, 又能分析更大范围的生物学模式, 能够在不同分子密度下实现精确的量化[76][77]。以上都是二维数据的分割和聚类, 而 Andronov 等人[79]扩展了 Voronoi 图在三维数据上的应用, 他们将 Voronoi 多面体作为 3D Voronoi 单元的特征来分割 3D SMLM 簇。根据 Andronov 等人的研究, Voronoi 图的方法可以防止边缘效应, 因为靠近边界的数据点具有大的或无限大的 Voronoi 单元, 因此局部密度较低, 不会对聚类做出贡献。

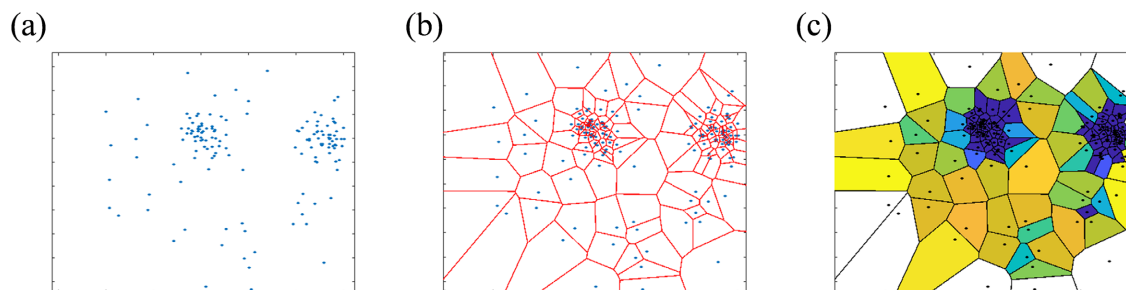


Figure 10. Clustering method of SMLM localizations data based on Voronoi diagram (a) Molecular localization; (b) molecular localization Voronoi diagram; (c) Voronoi diagrams displayed in different colors

图 10. 基于 Voronoi 图对 SMLM 定位数据进行聚类的方法(a) 分子定位; (b) 分子定位的 Voronoi 图; (c) 不同颜色显示的 Voronoi 图

因此,基于 Voronoi 单元的几何特性,Voronoi 图通过基于点的邻近性将空间划分为多边形或多面体,可以实现每个荧光定位点的精准局部密度估计,例如这些单元区域。一些数据点根据其单独的 Voronoi 单元区域被视为簇的一部分,这可以对这些数据点进行粗略的分割。此外,一些边界数据点由于局部密度较低可能被排除在簇的一部分之外,因为与局部密度较高的内部数据点相比,它们可能具有非常大的 Voronoi 单元区域(例如,图 10(c)中的白色单元区域),这使其特别适合识别分子聚类,自然地抑制了噪声。这种方法比依赖固定搜索半径的 Ripley 的 K 函数分析更为准确,且避免了由于参数选择(如搜索半径)引入的偏差。对于聚集区域较为紧密或存在重叠时,并且在 SMLM 数据中单个荧光团伪影多次闪烁,利用 Voronoi 图的方法可能无法准确分离这些簇,这是因为这些数据点的密度和簇与簇之间的距离对 Voronoi 图的分割有很大影响。基于 Voronoi 图的聚类方法也具有较大的计算复杂性,在三维空间中生成 Voronoi 多面体的计算量较大,尤其是对于高密度的 SMLM 数据。这种计算复杂度会对分析的实时性和处理效率带来挑战,可能需要高性能计算资源或优化算法。

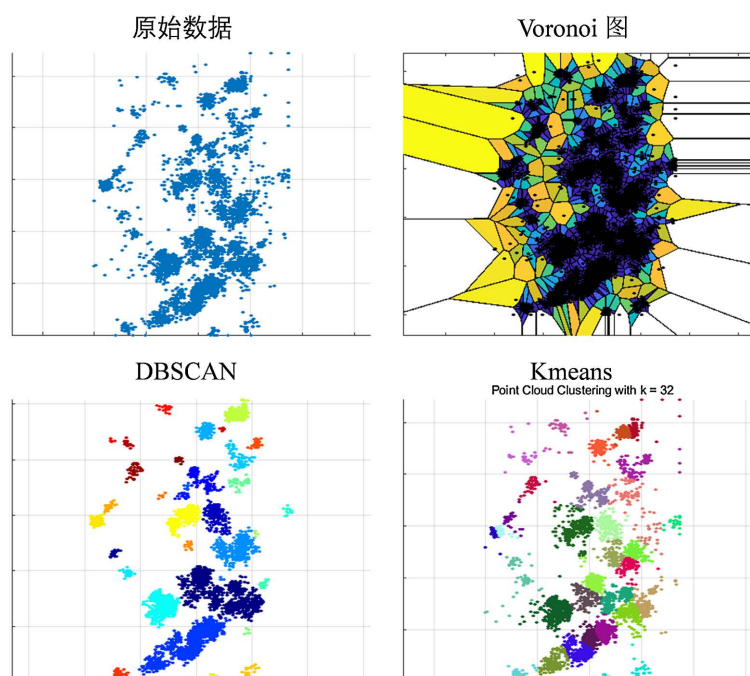


Figure 11. Comparison of different clustering methods

图 11. 不同聚类方法的对比

我们使用不同的聚类方法对同一数据进行处理, 以此比较他们的效果和性能, 如图 11 所示, Voronoi 图的方法能够很好的识别到噪声和分子聚类, 但可能无法准确分离这些簇, 并且相比于 DBSCAN 和 K-means, 它能避免由于参数选择(如搜索半径、簇的数量 K 值)引入的偏差。DBSCAN 可以很好的识别噪声和划分聚团, 但需要人为多次尝试选择到合适的参数。而 K-means 不能像 DBSCAN 那样很好的划分聚团, 同时也容易受到噪声的影响而错误的划分聚团, 在使用 K-means 时, 在已知簇的数量时将得到良好的应用。另外单一的聚类算法很难直接应用于 SMLM 数据分析, 针对不同的研究目的, 不同方法的结合或许能得到更好的聚类效果。

3.3. 几何形状分析

单分子定位显微镜提供了超高分辨率, 使得分子分布的几何结构研究成为可能。分子的几何分布结构不仅反映了分子自身的空间排列, 还揭示了这些分子与细胞内其他结构之间的相互作用。例如, 在微管中, 超分辨成像揭示了它的线状结构, 并结合线剖面分析解析出它的生物学参数[80]; 在神经元中, 超分辨显微镜揭示了轴突中的周期性环状结构, 这种结构由肌动蛋白和光谱蛋白相间排列而成, 具有明确的几何规律性[81]; 在 MKS 和 NPHP 复合物或者核孔复合体中, 超分辨成像结合几何建模进一步解析了他们的环状结构, 以及核孔复合体蛋白分布的八重对称特征[82]-[84]。这些研究不仅依赖于对单分子位置信息的高精度解析, 更得益于几何模型拟合方法的应用, 这些方法可以将分子位置的点云数据转化为具有生物学意义的几何参数。因此, 通过几何形状分析, 将分子的分布与细胞内复杂结构之间的关系建立起来, 是 SMLM 数据分析中的重要任务之一。

3.3.1. 线剖面分析

在单分子定位显微镜中, 数据分析是提取生物信息的关键步骤。由于 SMLM 能够以纳米级分辨率揭示细胞内结构的复杂性, 基于几何形状的分析方法在理解这些结构的生物学功能中发挥了重要作用。从 SMLM 数据中分析例如分子大小尺寸等附加信息可以通过假设一个近似的几何结构, 例如, 圆柱体, 球体, 圆环及其衍生物, 利用这样的几何结构对 SMLM 数据拟合, 统计拟合出的几何结构尺寸来确定分子的大小尺寸。这种几何分析可以不直接用于坐标, 例如用于在图中被线剖面框住的结构, 其对应的横向剖面定位被统计为直方图, 如图 12 所示, 然后通过曲线拟合来近似。而微管等管状结构的横向剖面通常通过两个高斯函数拟合描述[80], 得出两个峰之间的预期距离, 从而可以获取它们的直径。如此一来, 复杂的结构被简化得可以进行统计分析了。

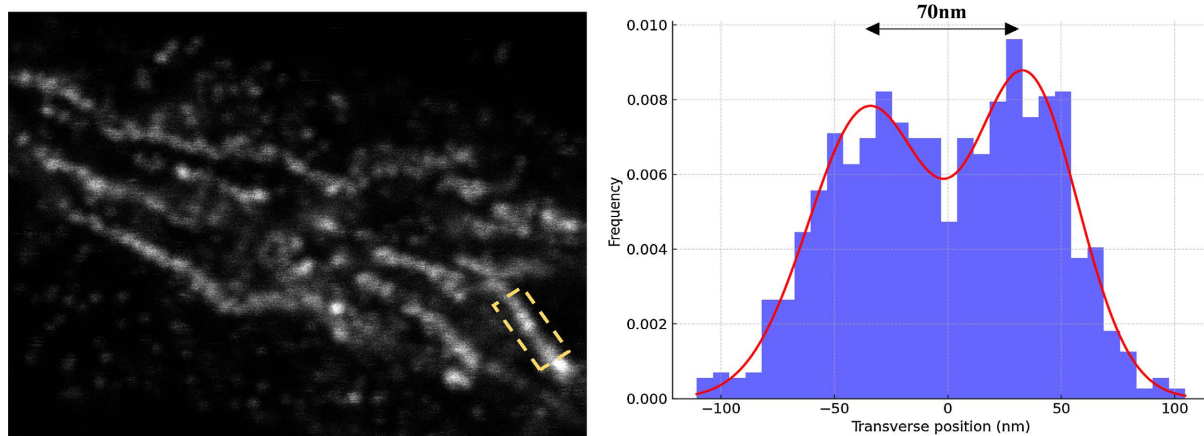


Figure 12. A schematic diagram of diameter parameters of tubular structures is obtained by using line profile analysis
图 12. 利用线剖面分析获取管状结构的直径参数的示意图

另外, Ke Xu 等人[81]的研究中, 轴突中的肌动蛋白丝和 β II-血影蛋白形成准一维周期性环状结构, 这些周期性结构由短的肌动蛋白丝环绕轴突周围来形成, 对于这种相邻重复的环状结构之间的距离, 他们采用肌动蛋白线剖面中相邻峰之间的距离来测量, 如图 13 所示。线剖面的方法成功揭示了神经元轴突中复杂的细胞骨架组织。通过分析这些结构的空分布和周期性, 研究者不仅对细胞骨架的组成有了更全面的理解, 还发现了它在维持细胞力学稳定性和调控功能方面的重要角色。这些研究成果为进一步探索轴突细胞骨架的分子机制及其在神经疾病中的潜在影响提供了理论基础。

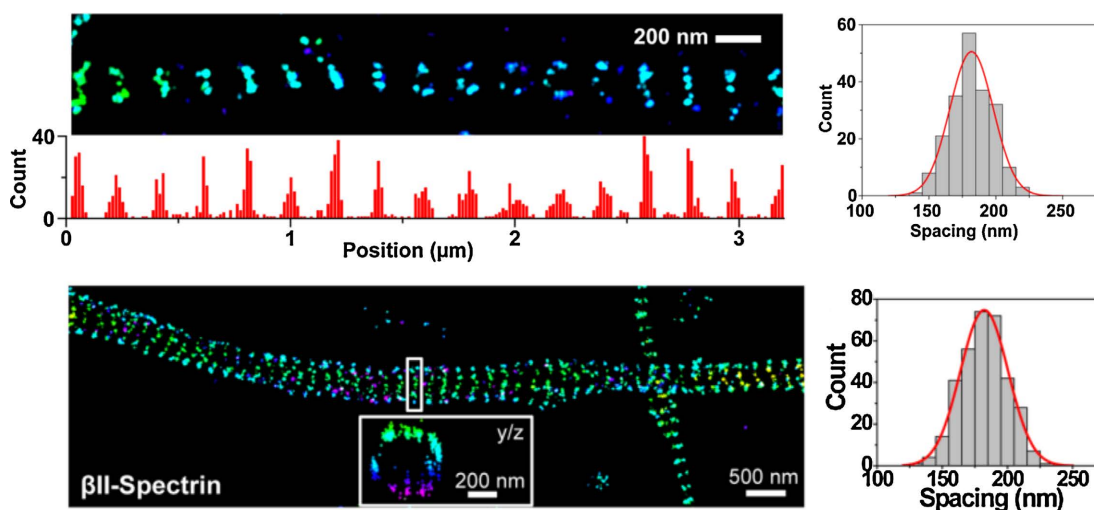


Figure 13. The distance between ring structures is obtained by line profile analysis [81]

图 13. 利用线剖面分析获取环状结构之间的距离[81]

但是这种线剖面通常是需要人为手动定义一个轴, 这个过程既费时又可能带来主观选择偏差, 另外, 由于数据被简化为一维, 也可能导致部分信息的丢失。这类问题可以通过使用更复杂的几何模型(2D, 3D 模型)来克服这些限制, 更准确地反映亚细胞或其他微观结构的实际分布和形态, 例如将许多亚细胞结构近似为环状结构。

3.3.2. 环状模型拟合分析

几何分析也可以直接基于 SMLM 坐标的数据进行灵活、定量和严格的分析, 其中就包括直接在 SMLM 坐标数据拟合环状模型, 在 Xiaoyu Shi 等人[82]的研究工作中, 他们对 MKS 和 NPHP 复合物进行环状拟合以获取环状结构的直径和中心位置。也有研究者将环状模型拟合到核孔复合体(nuclear pore complex, NPC) [83] [84]的数据上来获取相同的尺寸参数。另外, 也可以使用圆霍夫变换检测病毒颗粒的圆形结构, 并根据圆形边缘找到每个颗粒的中心位置[85], 如图 14 所示。霍夫变换是一种边缘检测技术, 圆霍夫变换能够识别图像中的圆形或接近圆形的结构。该变换通过在不同位置和半径的参数空间中投票, 最终选出最符合圆形特征的候选区域。

3.3.3. 复杂模型拟合分析

以上这些方法仅限于简单的几何形状或需要相同的结构, 最近, 有研究者提出了 Localization Model Fit (LocMoFit)用于将任意模型拟合到定位坐标[86], 他们用线性管模型对微管进行了拼接拟合, 另外, 还对单个 NPC 进行八重对称模型拟合, 并得到 3 种参数, 分别是环的半径, 间隔以及环之间的方位扭曲角, 如图 15 所示。因此使用更加复杂的几何模型去拟合到定位坐标, 能增加从 SMLM 数据中提取的信息内容。

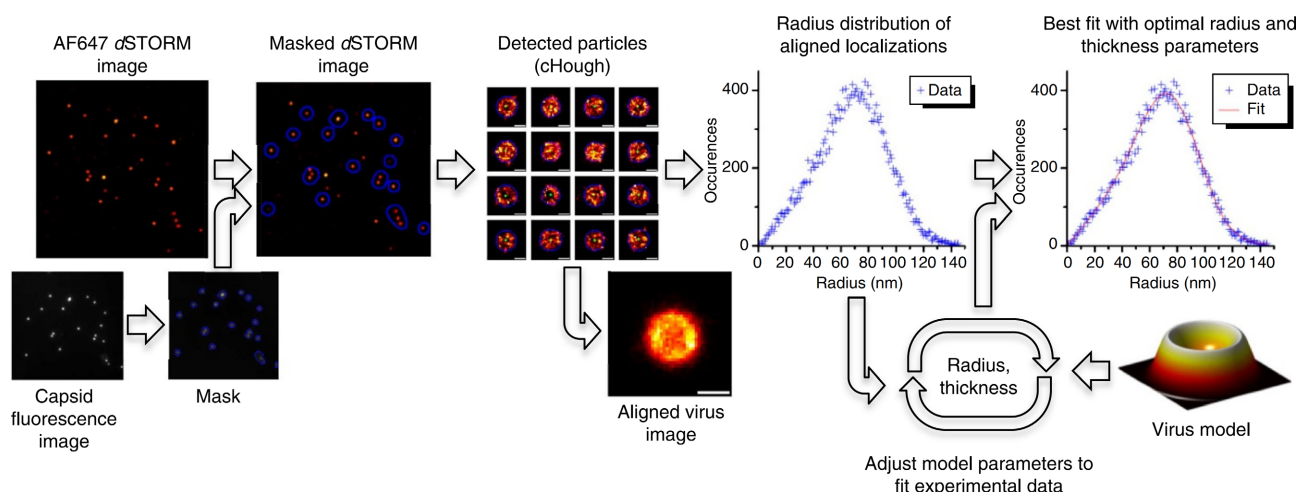


Figure 14. The radius of the ring structure is obtained by using circular Hough transform detection [85]

图 14. 利用圆霍夫变换检测获取环状结构的半径[85]

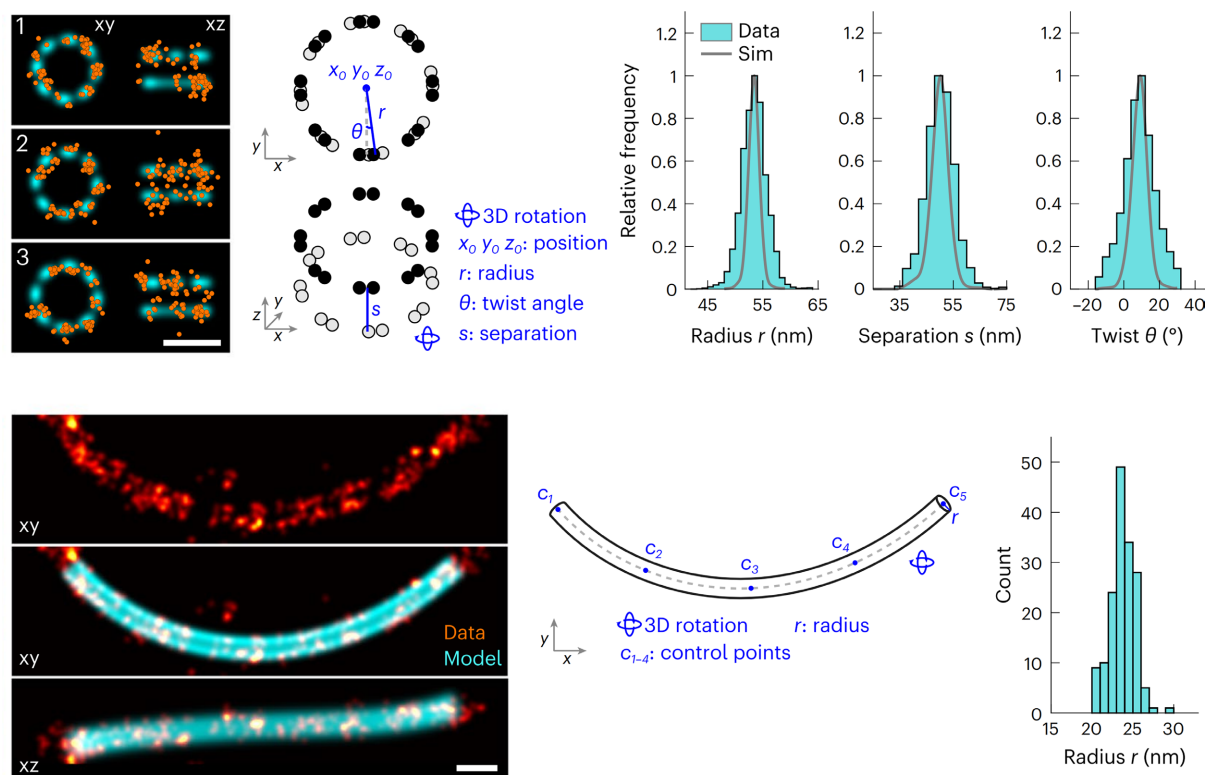


Figure 15. Application of LocMoFit model [86]

图 15. LocMoFit 模型的应用[86]

在实际应用中, 分析方法的选择应根据具体的研究问题和感兴趣的结构特点进行灵活调整。如果研究目标是获取结构的基本信息, 如半径和距离, 线剖面方法因其直观性和高效性通常是首选。这种方法能够沿指定方向提取感兴趣区域的空间分布数据, 快速提供结构参数的初步估计。然而, 当研究需要更深入地揭示结构的复杂性, 或当已有的信息能够为更精确的建模提供依据时, 构建复杂的几何模型便成为一种更优的选择。通过构建复杂的几何模型, 可以对单分子定位数据中的定位坐标进行拟合, 从而获

取更复杂且准确的结构参数。这种方法可以用于表征如环状、三维层状等更复杂的结构。与线剖面方法相比, 几何模型拟合不仅能够描述结构的全局特征, 还能帮助发现隐藏在定位数据中的微观规律, 如周期性、对称性或拓扑特征。因此, 在选择分析方法时, 应权衡研究目标的需求和数据的复杂性。例如, 当仅需验证某一假设中的几何参数时, 简单高效的线剖面分析已足够。而当研究涉及未知结构的揭示或对动态过程的细致描述时, 则需要采用更精确的几何模型拟合方法以获得全面且可靠的分析结果。这种选择的合理性不仅能够优化研究效率, 还能确保最终结果的科学性和准确性。

3.4. 深度学习

深度学习是机器学习的一个子集, 利用具有多层(因此称为“深”)的神经网络来建模数据中的复杂模式和结构。这些网络由多个互联的人工神经元组成, 能够自动学习从输入数据中提取分层特征。在图像分析领域, 例如 SMLM 数据, 深度学习方法擅长于识别微妙的模式、进行目标分割以及基于原始像素或坐标数据执行分类任务。与传统算法需要手动特征提取不同, 深度学习模型可以在训练过程中自动学习这些特征。

在 SMLM 的领域中, 深度学习技术已应用于数据分析的各个阶段, 从分割到聚类甚至特征提取。例如, 在分析与 caveolin-1 蛋白相关的 SMLM 数据时, 深度学习模型(如卷积神经网络(CNN)和 PointNet)已被用来分类并区分不同的生物结构, 且其准确性远高于传统方法[87][88]。这些方法通过实现自动化识别像 caveolae 和 scaffolds 等生物结构, 显著改善了传统显微技术的不足, 因为这些结构由于其纳米尺度大小难以通过标准显微镜技术分辨[87][89], 如图 16 所示。深度学习对 SMLM 的最大贡献之一是提高图像分辨率和减少噪声。传统的分子定位方法常常受到噪声和背景干扰的困扰, 但深度学习方法已被证明能够通过提高图像质量和处理速度来增加定位精度[90]。

尽管深度学习在 SMLM 中的聚类应用仍处于早期阶段, 但最近的工作显示出巨大的潜力。传统的聚类技术(如 DBSCAN)在处理大数据集或异质样本时存在挑战, 因为它们依赖于用户定义的参数[91]。相比之下, 基于深度学习的方法, 如图神经网络(GNN), 提供了一种更具可扩展性和自动化的方法。这些方法能够检测和量化颗粒聚集, 以显著更高的准确性和更少的计算成本解决集群的空间关系[92][93], 如图 17 所示。例如, 深度学习在 SMLM 中的一个重要应用是分析膜相关蛋白的纳米集群。与传统方法相比, 这些方法通过减少颗粒的错误识别并加速处理时间, 在处理大规模数据集时具有显著优势[93]。此外, 深度学习模型已被用于多色 SMLM, 使得多个蛋白质能够同时被检测, 从而为复杂的生物结构(如中心粒)提供更深入的见解[60]。

尽管取得了这些进展, 深度学习在 SMLM 分析中的整合仍然面临一些挑战。主要的难题之一是需要大量带注释的数据集来训练深度学习模型。这些数据集通常很难获取, 特别是在需要精确标注的专业生物学领域。此外, 训练深度学习模型的计算成本, 特别是在处理大规模三维数据集时, 仍然很高。另一个挑战是模型的泛化能力。深度学习技术通常是在特定数据集上训练的, 当应用于不同的实验设置或细胞类型时, 性能可能会下降, 这可能是因为当研究的目标分子不一样时, 所产生的 SMLM 数据显示不同的特征结构, 确保这些模型在不同生物学背景下的鲁棒性和可转移性是未来发展的关键领域。尽管面临这些挑战, 深度学习在 SMLM 数据分析中的潜力巨大。随着算法开发和计算能力的不断进步, 预计深度学习将在从复杂生物数据中提取有意义的见解方面发挥越来越重要的作用。随着更复杂的模型和更大、更具多样性的数据集的出现, 这些方法的准确性和效率将不断提高, 从而推动分子和细胞生物学的突破。

4. 展望

SMLM 成像作为一种相对较新的成像技术, 正在为我们探索和理解许多低于荧光显微镜衍射极限的

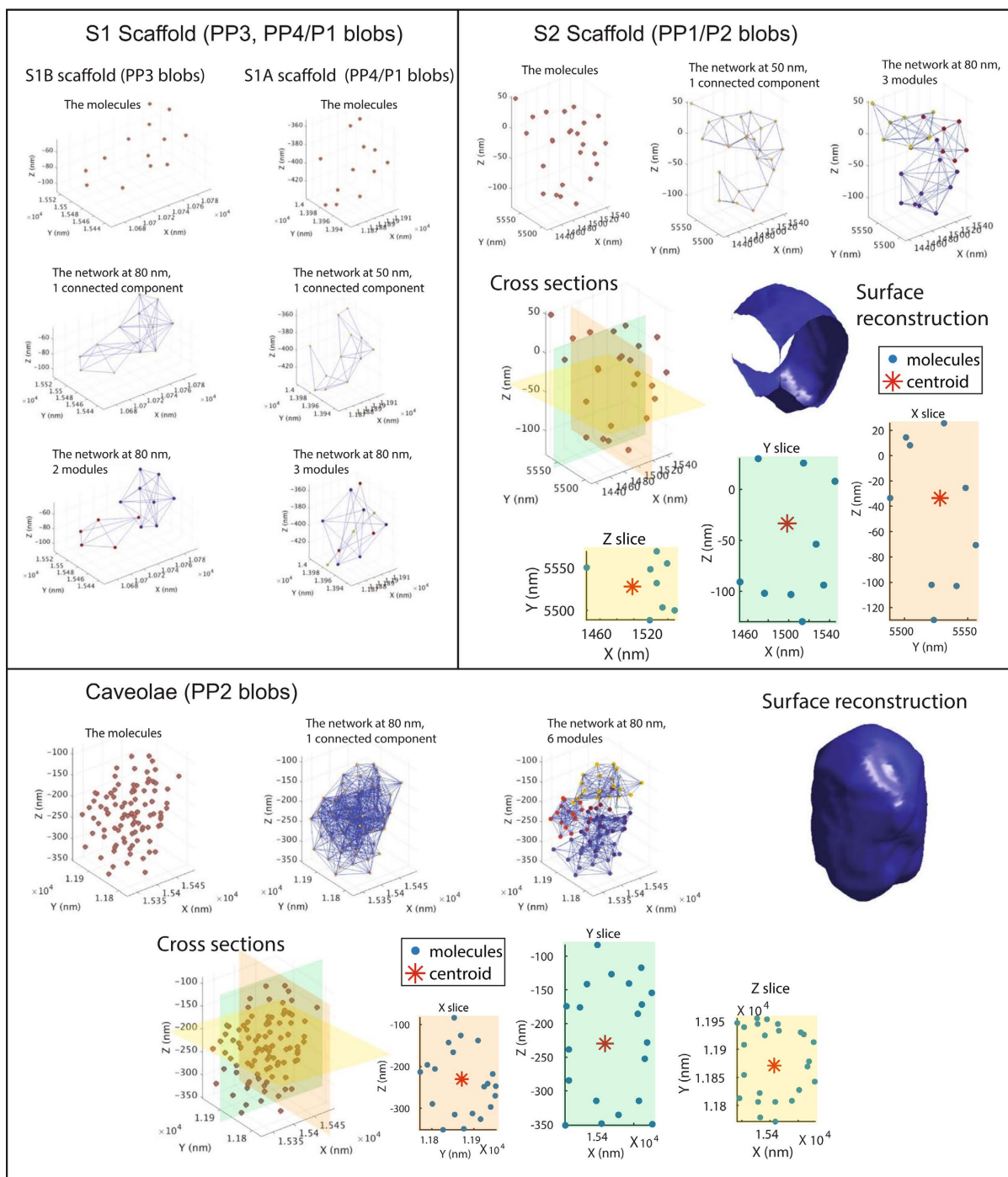


Figure 16. Application of deep learning to SMLM data analysis [88]

图 16. 深度学习在 SMLM 数据分析中的应用[88]

大分子复合物的结构和功能创造机会。其提供的高精度定位数据为科学发现带来了巨大的潜力。然而，目前的分析方法仍存在一定局限性，有待进一步发展，首先超分辨率成像技术的目标是获得准确的分子位置，但在实际成像过程中，会遇到标记物为间接标记(成像的荧光分子位置不直接代表真实的目标分子位置)、多次闪烁曝光、分子动态、SMLM 的噪音来源等问题，这些因素对数据分析提出了较高的要求。

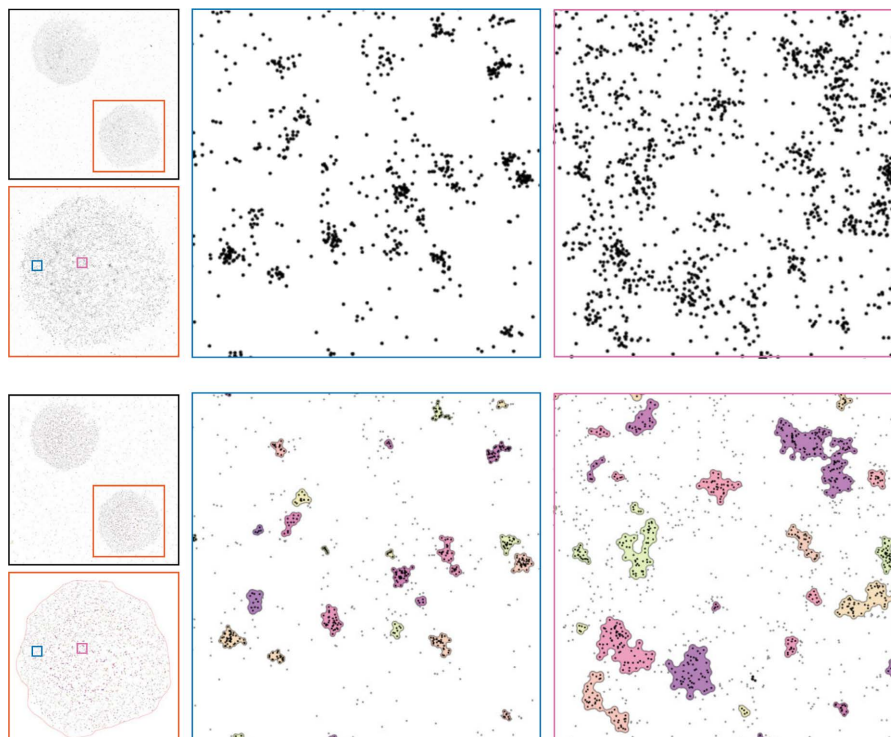


Figure 17. Analysis of dSTORM data by deep learning model [91]

图 17. 深度学习模型对 dSTORM 数据的分析[91]

另外, 如何准确地还原分子位置, 消除因信号重叠、定位误差和噪声引起的伪影, 仍然是当前分析方法中的一大难点。此外, 现有的分析方法在操作上繁琐, 且缺乏较高的集成度。不同分析方法之间往往缺乏兼容性, 导致中间步骤繁复, 效率低下。特别是对于大规模数据集, 手动分析不仅耗时且容易产生错误。因此, 迫切需要开发一种更为智能、自动化的分析工具来提升数据处理的效率和准确性。

针对这些挑战, 未来的研究方向应集中于以下几个方面:

- 1、开发用户友好的软件工具: 现有的 **SMLM** 数据分析工具大多数操作复杂, 且用户界面不够友好。因此, 开发一款集成度高、易于使用的分析软件将是关键。该工具应当能够自动化地完成数据预处理、定位精度优化、噪声过滤、分子位置还原等任务, 减少人为干预, 提升分析效率。
- 2、整合多种分析方法: 单一的分析方法往往难以应对 **SMLM** 数据中的多样性和复杂性。未来的研究应着重于整合现有的分析方法, 如径向分布函数(RDF)、Ripley 的 K 函数、DBSCAN 算法等, 采用集成分析策略, 结合几何模型拟合和聚类分析, 全面提升数据分析的精度和鲁棒性。
- 3、实现自动化分析流程: 为了进一步提高分析效率和准确性, 未来的研究应推动 **SMLM** 数据分析的自动化。通过机器学习和人工智能技术, 开发能够自我学习并自动选择最合适分析方法的系统, 降低分析过程中的人工干预, 并减少人为误差。
- 4、精确度提升与多维度信息整合: 除了常规的分子定位信息, 未来的分析方法应考虑整合显微镜参数、荧光团的光物理学特性(如再激活概率和定位误差分布)等多维度信息。这将有助于提升数据量化的精度和稳健性, 从而更准确地解析亚细胞结构中的分子组成和蛋白质相互作用。

综上所述, 未来的 **SMLM** 数据分析方法不仅需要解决现有的技术难题, 还应着眼于自动化和智能化的发展。通过开发集成度高、操作简便、精度更高的分析工具, 我们有望在未来实现更加深入和精确的细胞内部机制研究, 推动生命科学领域的重大突破。

参考文献

- [1] 毛峥乐, 王琛, 程亚. 超分辨远场生物荧光成像——突破光学衍射极限[J]. 中国激光, 2008, 35(9): 1283-1307.
- [2] 姚保利, 雷铭, 薛彬, 等. 高分辨和超分辨光学成像技术在空间和生物中的应用[J]. 光子学报, 2011, 40(11): 1607-1618.
- [3] 胡春光, 查日东, 凌秋雨, 等. 超分辨显微技术在活细胞中的应用与发展[J]. 红外与激光工程, 2017, 46(11): 1103002.
- [4] Giessibl, F.J. (1995) Atomic Resolution of the Silicon (111)-(7 × 7) Surface by Atomic Force Microscopy. *Science*, **267**, 68-71. <https://doi.org/10.1126/science.267.5194.68>
- [5] Browning, N.D., Chisholm, M.F. and Pennycook, S.J. (1993) Atomic-resolution Chemical Analysis Using a Scanning Transmission Electron Microscope. *Nature*, **366**, 143-146. <https://doi.org/10.1038/366143a0>
- [6] Sohda, Y., Yamanashi, H., Fukuda, M., *et al.* (2008) Scanning Electron Microscope. *Science*, **183**, 119.
- [7] 安莎, 但旦, 于湘华, 等. 单分子定位超分辨显微成像技术研究进展及展望(特邀综述) [J]. 光子学报, 2020, 49(9): 0918001.
- [8] Stephens, D.J. and Allan, V.J. (2003) Light Microscopy Techniques for Live Cell Imaging. *Science*, **300**, 82-86. <https://doi.org/10.1126/science.1082160>
- [9] Czirik, A., Rupp, P.A., Rongish, B.J. and Little, C.D. (2002) Multi-field 3D Scanning Light Microscopy of Early Embryogenesis. *Journal of Microscopy*, **206**, 209-217. <https://doi.org/10.1046/j.1365-2818.2002.01032.x>
- [10] Kobat, D., Horton, N.G. and Xu, C. (2011) *In Vivo* Two-Photon Microscopy to 1.6 mm Depth in Mouse Cortex. *Journal of Biomedical Optics*, **16**, Article ID: 106014. <https://doi.org/10.1117/1.3646209>
- [11] Yildiz, A., Forkey, J.N., McKinney, S.A., Ha, T., Goldman, Y.E. and Selvin, P.R. (2003) Myosin V Walks Hand-over-Hand: Single Fluorophore Imaging with 1.5 nm Localization. *Science*, **300**, 2061-2065. <https://doi.org/10.1126/science.1084398>
- [12] Aquino, D., Schönle, A., Geisler, C., Middendorff, C.v., Wurm, C.A., Okamura, Y., *et al.* (2011) Two-Color Nanoscopy of Three-Dimensional Volumes by 4Pi Detection of Stochastically Switched Fluorophores. *Nature Methods*, **8**, 353-359. <https://doi.org/10.1038/nmeth.1583>
- [13] Pertsinidis, A., Zhang, Y. and Chu, S. (2010) Subnanometre Single-Molecule Localization, Registration and Distance Measurements. *Nature*, **466**, 647-651. <https://doi.org/10.1038/nature09163>
- [14] Zheng, Q., Juette, M.F., Jockusch, S., Wasserman, M.R., Zhou, Z., Altman, R.B., *et al.* (2014) Ultra-Stable Organic Fluorophores for Single-Molecule Research. *Chemical Society Reviews*, **43**, 1044-1056. <https://doi.org/10.1039/c3cs60237k>
- [15] Shaner, N.C., Patterson, G.H. and Davidson, M.W. (2007) Advances in Fluorescent Protein Technology. *Journal of Cell Science*, **120**, 4247-4260. <https://doi.org/10.1242/jcs.005801>
- [16] Lidke, K.A., Rieger, B., Jovin, T.M. and Heintzmann, R. (2005) Superresolution by Localization of Quantum Dots Using Blinking Statistics. *Optics Express*, **13**, 7052-7062. <https://doi.org/10.1364/opex.13.007052>
- [17] Lincoln, R., Bossi, M.L., Rimmel, M., D'Este, E., Butkevich, A.N. and Hell, S.W. (2022) A General Design of Caging-Group-Free Photoactivatable Fluorophores for Live-Cell Nanoscopy. *Nature Chemistry*, **14**, 1013-1020. <https://doi.org/10.1038/s41557-022-00995-0>
- [18] Gould, T.J., Verkhusha, V.V. and Hess, S.T. (2009) Imaging Biological Structures with Fluorescence Photoactivation Localization Microscopy. *Nature Protocols*, **4**, 291-308. <https://doi.org/10.1038/nprot.2008.246>
- [19] Dempsey, G.T., Bates, M., Kowtoniuk, W.E., Liu, D.R., Tsien, R.Y. and Zhuang, X. (2009) Photoswitching Mechanism of Cyanine Dyes. *Journal of the American Chemical Society*, **131**, 18192-18193. <https://doi.org/10.1021/ja904588g>
- [20] Nirmal, M., Dabbousi, B.O., Bawendi, M.G., Macklin, J.J., Trautman, J.K., Harris, T.D., *et al.* (1996) Fluorescence Intermittency in Single Cadmium Selenide Nanocrystals. *Nature*, **383**, 802-804. <https://doi.org/10.1038/383802a0>
- [21] Abbe, E. (1873) Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv für Mikroskopische Anatomie*, **9**, 413-468. <https://doi.org/10.1007/bf02956173>
- [22] Mortensen, K.I., Churchman, L.S., Spudich, J.A. and Flyvbjerg, H. (2010) Optimized Localization Analysis for Single-Molecule Tracking and Super-Resolution Microscopy. *Nature Methods*, **7**, 377-381. <https://doi.org/10.1038/nmeth.1447>
- [23] Rust, M.J., Bates, M. and Zhuang, X. (2006) Sub-Diffraction-Limit Imaging by Stochastic Optical Reconstruction Microscopy (Storm). *Nature Methods*, **3**, 793-796. <https://doi.org/10.1038/nmeth929>
- [24] Heilemann, M., van de Linde, S., Schüttelpelz, M., Kasper, R., Seefeldt, B., Mukherjee, A., *et al.* (2008) Subdiffraction-Resolution Fluorescence Imaging with Conventional Fluorescent Probes. *Angewandte Chemie International Edition*, **47**, 6172-6176. <https://doi.org/10.1002/anie.200802376>

- [25] Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacio, J.S., *et al.* (2006) Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science*, **313**, 1642-1645. <https://doi.org/10.1126/science.1127344>
- [26] Ostersehl, L.M., Jans, D.C., Wittek, A., Keller-Findeisen, J., Inamdar, K., Sahl, S.J., *et al.* (2022) DNA-Paint Minflux Nanoscopy. *Nature Methods*, **19**, 1072-1075. <https://doi.org/10.1038/s41592-022-01577-1>
- [27] Jungmann, R., Steinhauer, C., Scheible, M., Kuzyk, A., Tinnefeld, P. and Simmel, F.C. (2010) Single-Molecule Kinetics and Super-Resolution Microscopy by Fluorescence Imaging of Transient Binding on DNA Origami. *Nano Letters*, **10**, 4756-4761. <https://doi.org/10.1021/nl103427w>
- [28] Li, H. and Vaughan, J.C. (2018) Switchable Fluorophores for Single-Molecule Localization Microscopy. *Chemical Reviews*, **118**, 9412-9454. <https://doi.org/10.1021/acs.chemrev.7b00767>
- [29] Hess, S.T., Girirajan, T.P.K. and Mason, M.D. (2006) Ultra-high Resolution Imaging by Fluorescence Photoactivation Localization Microscopy. *Biophysical Journal*, **91**, 4258-4272. <https://doi.org/10.1529/biophysj.106.091116>
- [30] van de Linde, S., Löschberger, A., Klein, T., Heidbreder, M., Wolter, S., Heilemann, M., *et al.* (2011) Direct Stochastic Optical Reconstruction Microscopy with Standard Fluorescent Probes. *Nature Protocols*, **6**, 991-1009. <https://doi.org/10.1038/nprot.2011.336>
- [31] Deschout, H., Zanacchi, F.C., Mlodzianoski, M., Diaspro, A., Bewersdorf, J., Hess, S.T., *et al.* (2014) Precisely and Accurately Localizing Single Emitters in Fluorescence Microscopy. *Nature Methods*, **11**, 253-266. <https://doi.org/10.1038/nmeth.2843>
- [32] Shroff, H., Galbraith, C.G., Galbraith, J.A. and Betzig, E. (2008) Live-Cell Photoactivated Localization Microscopy of Nanoscale Adhesion Dynamics. *Nature Methods*, **5**, 417-423. <https://doi.org/10.1038/nmeth.1202>
- [33] Lukinavičius, G., Umezawa, K., Olivier, N., Honigsmann, A., Yang, G., Plass, T., *et al.* (2013) A Near-Infrared Fluorophore for Live-Cell Super-Resolution Microscopy of Cellular Proteins. *Nature Chemistry*, **5**, 132-139. <https://doi.org/10.1038/nchem.1546>
- [34] Wäldchen, S., Lehmann, J., Klein, T., van de Linde, S. and Sauer, M. (2015) Light-Induced Cell Damage in Live-Cell Super-Resolution Microscopy. *Scientific Reports*, **5**, Article No. 15348. <https://doi.org/10.1038/srep15348>
- [35] Baddeley, D., Cannell, M.B. and Soeller, C. (2010) Visualization of Localization Microscopy Data. *Microscopy and Microanalysis*, **16**, 64-72. <https://doi.org/10.1017/s143192760999122x>
- [36] Nicovich, P.R., Owen, D.M. and Gaus, K. (2017) Turning Single-Molecule Localization Microscopy into a Quantitative Bioanalytical Tool. *Nature Protocols*, **12**, 453-460. <https://doi.org/10.1038/nprot.2016.166>
- [37] Owen, D.M., Rentero, C., Rossy, J., Magenau, A., Williamson, D., Rodriguez, M., *et al.* (2010) PALM Imaging and Cluster Analysis of Protein Heterogeneity at the Cell Surface. *Journal of Biophotonics*, **3**, 446-454. <https://doi.org/10.1002/jbio.200900089>
- [38] Sengupta, P., Jovanovic-Talisman, T., Skoko, D., Renz, M., Veatch, S.L. and Lippincott-Schwartz, J. (2011) Probing Protein Heterogeneity in the Plasma Membrane Using PALM and Pair Correlation Analysis. *Nature Methods*, **8**, 969-975. <https://doi.org/10.1038/nmeth.1704>
- [39] Cisse, I.I., Izeddin, I., Causse, S.Z., Boudarene, L., Senecal, A., Muresan, L., *et al.* (2013) Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells. *Science*, **341**, 664-667. <https://doi.org/10.1126/science.1239053>
- [40] Shivanandan, A., Unnikrishnan, J. and Radenovic, A. (2016) On Characterizing Protein Spatial Clusters with Correlation Approaches. *Scientific Reports*, **6**, Article No. 31164. <https://doi.org/10.1038/srep31164>
- [41] Ripley, B.D. (1977) Modelling Spatial Patterns. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **39**, 172-192. <https://doi.org/10.1111/j.2517-6161.1977.tb01615.x>
- [42] Dixon, P.M. (2006) Ripley's K Function.
- [43] Hansson, K., Jafari-Mamaghani, M. and Krieger, P. (2013) RipleyGUI: Software for Analyzing Spatial Patterns in 3D Cell Distributions. *Frontiers in Neuroinformatics*, **7**, Article No. 5. <https://doi.org/10.3389/fninf.2013.00005>
- [44] Wiegand, T. and A. Moloney, K. (2004) Rings, Circles, and Null-Models for Point Pattern Analysis in Ecology. *Oikos*, **104**, 209-229. <https://doi.org/10.1111/j.0030-1299.2004.12497.x>
- [45] Haase, P. (1995) Spatial Pattern Analysis in Ecology Based on Ripley's K-Function: Introduction and Methods of Edge Correction. *Journal of Vegetation Science*, **6**, 575-582. <https://doi.org/10.2307/3236356>
- [46] Marcon, E. and Puech, F. (2009) Generalizing Ripley's K Function to Inhomogeneous Populations. <https://shs.hal.science/halshs-00372631/document>
- [47] Baddeley, A.J., Møller, R.A., Howard, C.V. and Boyde, A. (1993) Analysis of a Three-Dimensional Point Pattern with Replication. *Applied Statistics*, **42**, 641-668. <https://doi.org/10.2307/2986181>
- [48] Goreaud, F. and Pélissier, R. (1999) On Explicit Formulas of Edge Effect Correction for Ripley's k-Function. *Journal of Vegetation Science*, **10**, 433-438. <https://doi.org/10.2307/3237072>

- [49] Kiskowski, M.A., Hancock, J.F. and Kenworthy, A.K. (2009) On the Use of Ripley's K-Function and Its Derivatives to Analyze Domain Size. *Biophysical Journal*, **97**, 1095-1103. <https://doi.org/10.1016/j.bpj.2009.05.039>
- [50] Besag, J. (1977) Comments on Ripley's Paper. *Journal of the Royal Statistical Society B*, **39**, 193-195.
- [51] Ehrlich, M., Boll, W., van Oijen, A., Hariharan, R., Chandran, K., Nibert, M.L., *et al.* (2004) Endocytosis by Random Initiation and Stabilization of Clathrin-Coated Pits. *Cell*, **118**, 591-605. <https://doi.org/10.1016/j.cell.2004.08.017>
- [52] Curd, A.P., Leng, J., Hughes, R.E., Cleasby, A.J., Rogers, B., Trinh, C.H., *et al.* (2020) Nanoscale Pattern Extraction from Relative Positions of Sparse 3D Localizations. *Nano Letters*, **21**, 1213-1220. <https://doi.org/10.1021/acs.nanolett.0c03332>
- [53] Pagoon, S.V., Tabarin, T., Yamamoto, Y., Ma, Y., Nicovich, P.R., Bridgeman, J.S., *et al.* (2016) Functional Role of T-Cell Receptor Nanoclusters in Signal Initiation and Antigen Discrimination. *Proceedings of the National Academy of Sciences*, **113**, E5454-E5463. <https://doi.org/10.1073/pnas.1607436113>
- [54] Deschout, H., Shivanandan, A., Annibale, P., Scarselli, M. and Radenovic, A. (2014) Progress in Quantitative Single-Molecule Localization Microscopy. *Histochemistry and Cell Biology*, **142**, 5-17. <https://doi.org/10.1007/s00418-014-1217-y>
- [55] Ester, M., *et al.* (1996) A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, Portland, 2-4 August 1996, 226-231.
- [56] Endesfelder, U., Finan, K., Holden, S.J., Cook, P.R., Kapanidis, A.N. and Heilemann, M. (2013) Multiscale Spatial Organization of RNA Polymerase in *Escherichia coli*. *Biophysical Journal*, **105**, 172-181. <https://doi.org/10.1016/j.bpj.2013.05.048>
- [57] Mazouchi, A. and Milstein, J.N. (2015) Fast Optimized Cluster Algorithm for Localizations (FOCAL): A Spatial Cluster Analysis for Super-Resolved Microscopy. *Bioinformatics*, **32**, 747-754. <https://doi.org/10.1093/bioinformatics/btv630>
- [58] Nino, D.F., Djayakarsana, D. and Milstein, J.N. (2020) FOCAL3D: A 3-Dimensional Clustering Package for Single-Molecule Localization Microscopy. *PLOS Computational Biology*, **16**, e1008479. <https://doi.org/10.1371/journal.pcbi.1008479>
- [59] Pengo, T., Holden, S.J. and Manley, S. (2014) PALMsiever: A Tool to Turn Raw Data into Results for Single-Molecule Localization Microscopy. *Bioinformatics*, **31**, 797-798. <https://doi.org/10.1093/bioinformatics/btu720>
- [60] Sieben, C., Banterle, N., Douglass, K.M., Gönczy, P. and Manley, S. (2018) Multicolor Single-Particle Reconstruction of Protein Complexes. *Nature Methods*, **15**, 777-780. <https://doi.org/10.1038/s41592-018-0140-x>
- [61] Barna, L., Dudok, B., Miczán, V., Horváth, A., László, Z.I. and Katona, I. (2015) Correlated Confocal and Super-Resolution Imaging by VividSTORM. *Nature Protocols*, **11**, 163-183. <https://doi.org/10.1038/nprot.2016.002>
- [62] Pagoon, S.V., Nicovich, P.R., Mollazade, M., Tabarin, T. and Gaus, K. (2016) Clus-DoC: A Combined Cluster Detection and Colocalization Analysis for Single-Molecule Localization Microscopy Data. *Molecular Biology of the Cell*, **27**, 3627-3636. <https://doi.org/10.1091/mbc.e16-07-0478>
- [63] Lagache, T., Grassart, A., Dallongeville, S., Faklaris, O., Sauvonnet, N., Dufour, A., *et al.* (2018) Mapping Molecular Assemblies with Fluorescence Microscopy and Object-Based Spatial Statistics. *Nature Communications*, **9**, Article No. 698. <https://doi.org/10.1038/s41467-018-03053-x>
- [64] Malkusch, S. and Heilemann, M. (2016) Extracting Quantitative Information from Single-Molecule Super-Resolution Imaging Data with LAMA—Localization Microscopy Analyzer. *Scientific Reports*, **6**, Article No. 34486. <https://doi.org/10.1038/srep34486>
- [65] Schnitzbauer, J., Wang, Y., Zhao, S., Bakalar, M., Nuwal, T., Chen, B., *et al.* (2018) Correlation Analysis Framework for Localization-Based Superresolution Microscopy. *Proceedings of the National Academy of Sciences*, **115**, 3219-3224. <https://doi.org/10.1073/pnas.1711314115>
- [66] Mollazade, M., Tabarin, T., Nicovich, P.R., Soeriyadi, A., Nieves, D.J., Gooding, J.J., *et al.* (2017) Can Single Molecule Localization Microscopy Be Used to Map Closely Spaced RGD Nanodomains? *PLOS ONE*, **12**, e0180871. <https://doi.org/10.1371/journal.pone.0180871>
- [67] Zhang, Y., Lara-Tejero, M., Bewersdorf, J. and Galán, J.E. (2017) Visualization and Characterization of Individual Type III Protein Secretion Machines in Live Bacteria. *Proceedings of the National Academy of Sciences*, **114**, 6098-6103. <https://doi.org/10.1073/pnas.1705823114>
- [68] Pape, J.K., Stephan, T., Balzarotti, F., Büchner, R., Lange, F., Riedel, D., *et al.* (2020) Multicolor 3D MINFLUX Nanoscopy of Mitochondrial MICOS Proteins. *Proceedings of the National Academy of Sciences*, **117**, 20607-20614. <https://doi.org/10.1073/pnas.2009364117>
- [69] Balzarotti, F., Eilers, Y., Gwosch, K.C., Gynnå, A.H., Westphal, V., Stefani, F.D., *et al.* (2017) Nanometer Resolution Imaging and Tracking of Fluorescent Molecules with Minimal Photon Fluxes. *Science*, **355**, 606-612.

- <https://doi.org/10.1126/science.aak9913>
- [70] Eilers, Y., Ta, H., Gwosch, K.C., Balzarotti, F. and Hell, S.W. (2018) MINFLUX Monitors Rapid Molecular Jumps with Superior Spatiotemporal Resolution. *Proceedings of the National Academy of Sciences*, **115**, 6117-6122. <https://doi.org/10.1073/pnas.1801672115>
- [71] Gwosch, K.C., Pape, J.K., Balzarotti, F., Hoess, P., Ellenberg, J., Ries, J., *et al.* (2020) MINFLUX Nanoscopy Delivers 3D Multicolor Nanometer Resolution in Cells. *Nature Methods*, **17**, 217-224. <https://doi.org/10.1038/s41592-019-0688-0>
- [72] Schmidt, R., Weihs, T., Wurm, C.A., Jansen, I., Rehman, J., Sahl, S.J., *et al.* (2021) MINFLUX Nanometer-Scale 3D Imaging and Microsecond-Range Tracking on a Common Fluorescence Microscope. *Nature Communications*, **12**, Article No. 1478. <https://doi.org/10.1038/s41467-021-21652-z>
- [73] Macqueen, J. (1967) Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, 1 January 1967, 281-297. <http://projecteuclid.org/euclid.bsmmsp/1200512992>
- [74] Jacak, J., Schaller, S., Borgmann, D. and Winkler, S.M. (2015) Characterization of the Distance Relationship between Localized Serotonin Receptors and Glia Cells on Fluorescence Microscopy Images of Brain Tissue. *Microscopy and Microanalysis*, **21**, 826-836. <https://doi.org/10.1017/s1431927615013513>
- [75] Okabe, A., Boots, B., Sugihara, K. and Chiu, S.N. (2000) Spatial Tessellations: Concepts and Applications of Voronoi Diagrams. John Wiley & Sons Ltd.
- [76] Levet, F., Hosy, E., Kechkar, A., Butler, C., Beghin, A., Choquet, D., *et al.* (2015) SR-Tesseler: A Method to Segment and Quantify Localization-Based Super-Resolution Microscopy Data. *Nature Methods*, **12**, 1065-1071. <https://doi.org/10.1038/nmeth.3579>
- [77] Andronov, L., Orlov, I., Lutz, Y., Vonesch, J. and Klaholz, B.P. (2016) ClusterViSu, a Method for Clustering of Protein Complexes by Voronoi Tessellation in Super-Resolution Microscopy. *Scientific Reports*, **6**, Article No. 24084. <https://doi.org/10.1038/srep24084>
- [78] Baddeley, D., Jayasinghe, I., Lam, L., Rossberger, S., Cannell, M.B. and Soeller, C. (2009) Optical Single-Channel Resolution Imaging of the Ryanodine Receptor Distribution in Rat Cardiac Myocytes. *Proceedings of the National Academy of Sciences*, **106**, 22275-22280. <https://doi.org/10.1073/pnas.0908971106>
- [79] Andronov, L., Michalon, J., Ouararhni, K., Orlov, I., Hamiche, A., Vonesch, J., *et al.* (2018) 3DClusterViSu: 3D Clustering Analysis of Super-Resolution Microscopy Data by 3D Voronoi Tessellations. *Bioinformatics*, **34**, 3004-3012. <https://doi.org/10.1093/bioinformatics/bty200>
- [80] Dempsey, G.T., Vaughan, J.C., Chen, K.H., Bates, M. and Zhuang, X. (2011) Evaluation of Fluorophores for Optimal Performance in Localization-Based Super-Resolution Imaging. *Nature Methods*, **8**, 1027-1036. <https://doi.org/10.1038/nmeth.1768>
- [81] Xu, K., Zhong, G. and Zhuang, X. (2013) Actin, Spectrin, and Associated Proteins Form a Periodic Cytoskeletal Structure in Axons. *Science*, **339**, 452-456. <https://doi.org/10.1126/science.1232251>
- [82] Shi, X., Garcia, G., Van De Weghe, J.C., McGorty, R., Pazour, G.J., Doherty, D., *et al.* (2017) Super-Resolution Microscopy Reveals That Disruption of Ciliary Transition-Zone Architecture Causes Joubert Syndrome. *Nature Cell Biology*, **19**, 1178-1188. <https://doi.org/10.1038/ncb3599>
- [83] Szymborska, A., de Marco, A., Daigle, N., Cordes, V.C., Briggs, J.A.G. and Ellenberg, J. (2013) Nuclear Pore Scaffold Structure Analyzed by Super-Resolution Microscopy and Particle Averaging. *Science*, **341**, 655-658. <https://doi.org/10.1126/science.1240672>
- [84] Thevathasan, J.V., Kahnwald, M., Cieřliński, K., Hoess, P., Peneti, S.K., Reitberger, M., *et al.* (2019) Nuclear Pores as Versatile Reference Standards for Quantitative Superresolution Microscopy. *Nature Methods*, **16**, 1045-1053. <https://doi.org/10.1038/s41592-019-0574-9>
- [85] Laine, R.F., Albecka, A., van de Linde, S., Rees, E.J., Crump, C.M. and Kaminski, C.F. (2015) Structural Analysis of Herpes Simplex Virus by Optical Super-Resolution Imaging. *Nature Communications*, **6**, Article No. 5980. <https://doi.org/10.1038/ncomms6980>
- [86] Wu, Y., Hoess, P., Tschanz, A., Matti, U., Mund, M. and Ries, J. (2022) Maximum-Likelihood Model Fitting for Quantitative Analysis of SMLM Data. *Nature Methods*, **20**, 139-148. <https://doi.org/10.1038/s41592-022-01676-z>
- [87] Khater, I.M., Aroca-Ouellette, S.T., Meng, F., Nabi, I.R. and Hamarneh, G. (2019) Caveolae and Scaffold Detection from Single Molecule Localization Microscopy Data Using Deep Learning. *PLOS ONE*, **14**, e0211659. <https://doi.org/10.1371/journal.pone.0211659>
- [88] Khater, I.M., Meng, F., Wong, T.H., Nabi, I.R. and Hamarneh, G. (2018) Super Resolution Network Analysis Defines the Molecular Architecture of Caveolae and Caveolin-1 Scaffolds. *Scientific Reports*, **8**, Article No. 9009. <https://doi.org/10.1038/s41598-018-27216-4>

-
- [89] Khater, I.M., Meng, F., Nabi, I.R. and Hamarneh, G. (2019) Identification of Caveolin-1 Domain Signatures via Machine Learning and Graphlet Analysis of Single-Molecule Super-Resolution Data. *Bioinformatics*, **35**, 3468-3475. <https://doi.org/10.1093/bioinformatics/btz113>
 - [90] Hyun, Y. and Kim, D. (2022) Development of Deep-Learning-Based Single-Molecule Localization Image Analysis. *International Journal of Molecular Sciences*, **23**, Article No. 6896. <https://doi.org/10.3390/ijms23136896>
 - [91] Williamson, D.J., Burn, G.L., Simoncelli, S., Griffié, J., Peters, R., Davis, D.M., *et al.* (2020) Machine Learning for Cluster Analysis of Localization Microscopy Data. *Nature Communications*, **11**, Article No. 1493. <https://doi.org/10.1038/s41467-020-15293-x>
 - [92] Saavedra, L.A., Mosqueira, A. and Barrantes, F.J. (2024) A Supervised Graph-Based Deep Learning Algorithm to Detect and Quantify Clustered Particles. *Nanoscale*, **16**, 15308-15318. <https://doi.org/10.1039/d4nr01944j>
 - [93] Lim, H., Kim, G.W., Heo, G.H., Jeong, U., Kim, M.J., Jeong, D., *et al.* (2024) Nanoscale Single-Vesicle Analysis: High-Throughput Approaches through Ai-Enhanced Super-Resolution Image Analysis. *Biosensors and Bioelectronics*, **263**, Article ID: 116629. <https://doi.org/10.1016/j.bios.2024.116629>