

文献阅读报告一：生物图像信息学综述

黄磊 计 702 2022E8013282156

1. 摘要

成像技术的不断进步使得我们能观察到的生物图像有了质的飞跃，更多深层的、细微的、复杂的细胞或分子活动可以在图像中展现。但同时其中数据信息规模之巨大、复杂性之高已经远远超出人工处理能力范畴，人们转而使用计算机进行自动化处理，进而衍生出“生物图像信息学”，该学科涵盖对生物图像的处理以及信息管理与可视化。本文对生物图像信息学中的重要应用和关键技术进行了介绍和总结，并概述了一些现有的分析工具与可用资源平台。

2. 引入/背景介绍

在过去的几十年中，成像技术迅猛发展，从毫米级分辨率降低到纳米级单分子水平分辨率、从二维静态图像到三维动态时序图像，数据规模和信息规模的不断扩大导致人工分析数据的难度陡增，人们必须依靠计算机辅助数据分析。由于生物图像与常见图像存在巨大差异，常规方法无法满足分析需求，因此适用于生物领域的高效、高速、高通量的自动化数据分析手段成为一项关键性研究，新兴领域“生物图像信息处理”应运而生。

该文章从应用发展、关键技术、可用工具和资源等角度综合介绍生物图像信息处理领域。生物信息图像处理学科以应用为导向，关键研究方向在于图像处理与挖掘、图像数据库和可视化三方面，其中包含众多关键的数据分析技术。为了提升实用性，不少平台或工具集成了一些图像分析或可视化技术，方便研究者使用。生物图像信息处理已然成为计算机视觉与生物图像分析交叉的关键学科，未来将迎来蓬勃发展。

3. 文章内容整理与总结

文章首先介绍了生物图像信息处理的主要应用研究，之后介绍了该领域中的关键的计算机处理技术，最后概述了一些可用的集成分析工具以及其他相关资源。

3.1 主要应用研究

生物图像信息学是以应用驱动而不断发展的，一般而言，其关键应用研究有细胞表型的高通量和高含量分析、模型生物图集构建、细胞和活体的动态研究、3D 神经元结构和大脑连接模型的重建、以及生物图像信息学和其他生物信息学方法的联合分析等五大方向。

细胞表型的高通量和高含量分析对于确定基因功能、描述细胞通路、药物发现和癌症诊断都有巨大意义，更系统全面的细胞表型描述可以为系统生物学研究提供十分重要的信息。

模型生物图集构建主要是针对于一些实验生物，例如小鼠、果蝇等，构建他们的各种数字图谱，并进一步将其各自的解剖和本体知识整合进数据库中。

细胞和活体的动态研究在多个细分学科中都有重要作用，已经研究出很多方法用于该应用研究。例如图像分析方法用于研究基于细胞的蛋白质表达和转录表达模式，整体成像分析用于从系统生物学角度研究动物发育过程，流行学习用于计算重建这些模式的思维时空发展，三维基因表达的空间配准与比较用于重建基因发育动力学等。

3D 神经元结构和大脑连接模型的重建旨在探索大脑的工作原理，不仅需要重建神经元结构，还需要研究大脑结构以及功能连接。众多富有创造性的方法被提出，甚至卷积神经网络也被用于图像的纳米级重建。随着该研究的进行，成千上万个重建的神经元被整理到公开的数据集中，便于后续研究。

生物图像信息学与其他生物信息学方法联合研究可以更好地促进相关领域的发展，例如可以将胚胎基因表达模式与比较基因组学方法结合以预测具备调控功能的基因序列。

3.2 关键技术

生物图像信息处理的关键分析技术包括但不限于特征提取与筛选、图像分割、图像配准、聚类分类与注释、索引与检索、以及可视化等，此外还有信号处理、模型估计等技术在本文不做介绍。

1. 特征提取与筛选：有效提取特征能够以更低维度的数据表征该图像，例如点、边等纹理特征，角度、空间等运动特征等，常有的特征提取方法有基于域的方法和图像变换的方法；
2. 图像分割：其是许多生物图像处理的基本步骤之一，目的是将感兴趣的部分分割出来，该方法非常依赖于特征选取，也依赖于图像质量。一些集成好的工具或插件也能帮助处理图像分割；
3. 图像配准：配准用于比较不同条件下的多组图像，并且可以进行定量测量和可视化。但由于图像的复杂性、大体积以及低信噪比，3D 图像的配准仍难以达到理想效果；
4. 聚类、分类与注释：模式聚类和分类技术可以区分图像中不同的对象或板块，图像注释则是将图像内容信息转换为语义上有意义的文本或其他易于组织和搜索的信息；
5. 索引与检索：该技术主要是为了快捷精准地访问数据库中的图像数据，现有的主要方法是提供和整合文本描述符，以及根据原始数据和处理数据实现基于内容的访问；
6. 可视化：作为一般科学数据可视化的子学科，生物图像领域的可视化广泛使用的是体积、表面或流体的可视化以动态展示数据，方便研究者对数据直观分析。

3.3 现有工具

由于不同厂家图像格式可能不同，一些常用的格式转换工具有 ImageJ 以及 BioFormats，TIFF 格式可以使用开源库 libtiff 处理。

常用的图像分析工具有 ImageJ 以及 ITK，前者提供针对不同规模图像的分析工具，后者提供了许多分割和配准服务。V3D 可用于三维图像处理，还有一些工具针对于特定任务或场景。

其他工具还有 OME 注释工具、Amira 可视化工具等，在此不一一列举。

3.4 其他资源

其他资源包括服务于通用生物学科或者特定生物学科研究的图像数据集，并且有些涵盖二维和三维图像，并提供检索和共享服务。一些重要的会议、期刊和书籍也成为了前沿研究内容的重要获取途径。

4. 讨论

生物图像信息处理不仅仅包括图像的处理，也包括对其他生物图像或视频的数据分析。除了数据分析技术之外，提升图像计算方法、提升成像技术的质量和吞吐量也十分重要，二者共同发展才能带来生物图像领域的显著进步。生物图像信息学来源于实际应用，最终评估如何使用这些计算技术来增强我们对生物实体的理解和解决各自问题的能力。随着技术的不断发展完善，该领域也将蓬勃进步。

5. 参考文献

- [1] H. Peng, Bioimage informatics: a new area of engineering biology, Bioinformatics, vol. 24, pp. 1827-1836, 2008.
- [2] Emami N, Sedaei Z, Ferdousi R. Computerized cell tracking: Current methods, tools and challenges[J]. Visual Informatics, 2021, 5(1): 1-13.