

填表说明

一、本报告中相关的技术或数据如涉及知识产权保护、军工项目保密等内容，请作脱密处理。

二、请用宋体小四字号撰写本报告，可另行附页或增加页数，A4纸双面打印。

三、表中所涉及的签名都必须用蓝、黑色墨水笔，亲笔签名或签字章，不可以打印代替。

四、同行专家业内评价意见书编号由工程师学院填写，编号规则为：年份4位+申报工程师职称专业类别(领域)4位+流水号3位，共11位。

一、个人申报

(一) 基本情况【围绕《浙江工程师学院（浙江大学工程师学院）工程类专业学位研究生工程师职称评审参考指标》，结合该专业类别(领域)工程师职称评审相关标准，举例说明】

1. 对本专业基础理论知识和专业技术知识掌握情况(不少于200字)

我在单细胞数据分析领域积累了扎实的专业基础和实践经验。在人工智能技术方面，我掌握了机器学习核心算法，包括支持向量机、随机森林等分类算法，线性回归、岭回归等回归算法，以及K-means、层次聚类等聚类算法。同时，我对生成对抗网络（GAN）、变分自编码器（VAE）、自编码器（AE）、Transformer架构及大语言模型等前沿技术有深入理解。

在药学专业领域，我具备扎实的药理学、病理学、药物制剂学和药物化学知识体系，这为我在单细胞测序数据的生物学解读提供了坚实的理论基础。在技术实践层面，我精通单细胞测序数据的全流程分析，包括数据质量控制与预处理、高变基因筛选、降维分析（PCA、t-SNE、UMAP）、细胞聚类及基于专业数据库的细胞类型注释。我熟练运用多种先进的单细胞分析工具，如Harmony、Seurat v5、scGPT、CODE-AD、SCALEX、SingleR等，这些工具显著提升了数据分析的效率和准确性。

基于专业积累，我成功开发了一款基于选择性对抗网络的自动化注释工具。该工具创新性地整合了细胞标记基因表达特征和药物作用靶点信息，不仅显著提高了稀有细胞亚群的识别准确率，还实现了潜在新细胞类型的预测功能，为单细胞数据分析提供了新的解决方案。

2. 工程实践的经历(不少于200字)

我参与了浙江大学智能创新药物研究院的单细胞数据批次归一化处理分析项目，该项目于2023年03月01日开始，至2024年06月01日结束，共计458天。在此期间，我专注于开发和优化基于选择性对抗网络UniMap的单细胞数据批次归一化处理方法。

作为算法开发成员，我负责多重选择性对抗网络UniMap的设计与实现工作，包括编写和优化代码。我成功完成了初步版本的开发，并在内部测试中验证了其基本功能。同时，我还承担了数据预处理和特征提取工作，确保数据质量和一致性，为算法评估提供了支持数据。

在结果评估阶段，我参与了算法性能的评估和结果分析，编写了对比分析报告并提出了改进建议。我们发现算法在某些数据集上表现出不稳定性，计算效率在处理大规模数据时较低，且不同来源的数据异质性影响了批次归一化效果。针对这些问题，我建议进一步优化网络UniMap结构以提高稳定性，改进计算流程以提升处理效率，并引入更多的数据预处理技术以减少数据异质性的影响。

这段工程实践使我深入理解了单细胞数据的特性和批次效应处理的挑战，提升了我的算法开发和数据分析能力，培养了我解决复杂生物信息学问题的实践经验。

3. 在实际工作中综合运用所学知识解决复杂工程问题的案例(不少于1000字)

项目背景

随着单细胞测序技术的飞速发展，研究者能够在单细胞分辨率下研究生物学问题。然而，不同实验条件下收集的单细胞数据往往存在显著的批次效应，这严重影响了数据整合和分析结果的可靠性。浙江大学智能创新药物研究院位于杭州钱塘区医药港小镇，是浙江大学与杭州

钱塘区联合共建的创新药物研究机构。作为多学科协作平台，研究院需要处理来自多个实验室、不同测序平台的单细胞数据，因此批次效应问题尤为突出，并且不同数据集之间存在大量不相同的细胞类型会造成“数据集部分重合”的问题，传统基于消除全局批次效应的单细胞自动注释方法并不适用。

项目概述

本项目从2023年03月01日开始，至2024年06月01日结束，历时458天。核心任务是开发和优化基于选择性对抗网络UniMap的单细胞数据批次归一化处理方法，以提高部分重合场景下的单细胞注释性能，在达到去除数据批次效应的同时，尽可能增加模型保留不同数据集的生物差异的能力。根据对问题的深入分析，我们发现最核心的任务是让模型能够识别到各个不同数据集之间的干扰细胞类型，并降低这部分数据对模型训练的干扰。

数据预处理与特征提取

由于单细胞数据普遍存在高噪声、高维度和高稀疏性的特点，本研究参考Scanpy推荐的标准数据标准化流程，设计了一套针对单细胞数据特性的预处理流程如下：

质量控制：首先通过质量控制将表达基因数低于200的细胞剔除，同时过滤掉在少于3个细胞中检测到表达的基因，以降低测序噪声的影响；在数据标准化阶段，对过滤后的数据进行对数归一化处理。具体而言，将每个细胞的基因计数标准化为CPM的1,000,000倍，并通过对数变换确保细胞间计数的可比性。

高变基因筛选：为有效识别不同数据来源中的HVGs，本项目在控制基因表达均值的前提下，采用变异系数作为基因变异度的度量指标。基于变异系数的计算结果，分别从参考数据和查询数据中选取变异系数最高的1200-2000个基因作为高变基因特征集合。

这一预处理流程有效提高了后续批次归一化的效果，为算法提供了高质量的基础数据。

算法开发与优化

作为项目的核心技术创新，我参与设计并实现了多重选择性对抗网络UniMap。该算法通过以下机制解决批次效应问题：

多重选择性对抗模块：构建多个专门针对不同类型批次效应的对抗模块，每个模块关注不同标签的数据的特征分布情况。

自适应权重机制：引入自适应权重，动态调整各对抗模块的贡献度。

生物学信息保护层：设计特殊的损失函数，在消除批次效应的同时保留关键的生物学信号。我们的模型的独特优势在于能够同时处理多种来源的批次效应，并保持数据的生物学一致性，避免不同数据集之间的标签不平衡。在实现过程中，我编写了超过10000行Python代码，使用了PyTorch深度学习框架，并针对GPU加速进行了优化，确保算法在实际应用中的高效运行。

性能评估与验证

在项目实施过程中，我们遇到了几个关键技术难点：算法稳定性问题、计算效率瓶颈以及数据异质性影响。初版选择性对抗网络UniMap在某些数据集上表现不稳定，波动较大；处理大规模数据时（>10万细胞）计算效率低下；而不同来源数据质量差异也显著影响了归一化效果。

针对这些问题，我主导了三轮主要的迭代改进：

- (1) 设计了基于数据质量的权重分配机制，使算法能够自适应地处理不同质量的数据批次，针对每一个不同的细胞类型，设计一个单独的配备有权重系数的判别器用于判断数据的重要性，提高了逐类别的批次去除能力了，特别是对于复杂异质性数据，具有明显效果提升。
- (2) 添加了边界损失和Focal loss，补充了模型在识别细胞亚型差异方面的能力；
- (3) 统一修改了模型的参数与架构优化，让模型结构与Torch2.0更加适配，提高模型运算速率。

项目成果

经过几轮迭代优化，UniMap在多个测试数据集上展现出优异性能：

在6个公开单细胞数据集上，UniMap在生物差异保留方面平均排名第一，批次效应排名第二。

与现有主流方法相比，UniMap能够在部分数据集中有最优秀的表现。

算法目前已经部署于浙江大学智能创新药物研究院，用于帮助研究人员进行快速的单细胞注释。

经验总结

通过这个复杂工程问题的解决过程，我深刻体会到理论知识与工程实践相结合的重要性：

数据质量控制和预处理：往往比算法本身更能影响最终结果。

定制化算法：针对特定领域问题的定制化算法通常优于通用算法。

结合领域知识的评估指标：对于算法的有效改进至关重要。

未来展望

展望未来，我计划进一步优化UniMap网络结构以提高其稳定性，改进计算流程以提升处理效率，并引入更多的数据预处理技术，以扩展算法在更广泛单细胞数据分析场景中的应用，为生物医学研究和药物开发提供更有力的数据分析支持。并且继续开放便于科研人员使用。

通过本项目的实施，我不仅提升了自己的技术能力，还深刻理解了如何在实际工作中综合运用所学知识解决复杂工程问题。这一经验将对我未来的职业发展产生深远影响。

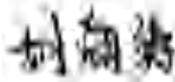
(二) 取得的业绩(代表作)【限填3项, 须提交证明原件(包括发表的论文、出版的著作、专利证书、获奖证书、科技项目立项文件或合同、企业证明等)供核实, 并提供复印件一份】

1. 公开成果代表作【论文发表、专利成果、软件著作权、标准规范与行业工法制定、著作编写、科技成果获奖、学位论文等】

成果名称	成果类别 [含论文、授权专利(含发明专利申请)、软件著作权、标准、工法、著作、获奖、学位论文等]	发表时间/授权或申请时间等	刊物名称/专利授权或申请号等	本人排名/总人数	备注
UniMap: Type-Level Integration Enhances Biological Preservation and Interpretability in Single-Cell Annotation	TOP期刊	2025年02月27日	Advanced Science	1/11	SCI期刊收录
基于选择性域鉴别器的单细胞自动注释方法及装置	发明专利申请	2023年12月28日	2023118302274	2/7	导师1作
基于协同注意力的可解释性抗癌药物协同预测方法及装置	授权发明专利	2024年04月26日	专利号: ZL2023 11155808.2	5/5	

2. 其他代表作【主持或参与的课题研究项目、科技成果应用转化推广、企业技术难题解决方案、自主研发设计的产品或样机、技术报告、设计图纸、软课题研究报告、可行性研究报告、规划设计方案、施工或调试报告、工程实验、技术培训教材、推动行业发展中发挥的作用及取得的经济社会效益等】

无

(三) 在校期间课程、专业实践训练及学位论文相关情况	
课程成绩情况	按课程学分核算的平均成绩： 86 分
专业实践训练时间及考核情况(具有三年及以上工作经历的不作要求)	累计时间： 1.3 年（要求1年及以上） 考核成绩： 86 分
本人承诺	
<p>个人声明：本人上述所填资料均为真实有效，如有虚假，愿承担一切责任，特此声明！</p> <p style="text-align: right;">申报人签名： </p>	

浙江大学研究生院
攻读硕士学位研究生成绩单

学号: 22260236	姓名: 胡海涛	性别: 男	学院: 工程师学院	专业: 生物与医药	学制: 2.5年						
毕业时最低应获: 24.0学分		已获得: 27.0学分		入学年月: 2022-09	毕业年月:						
学位证书号:			毕业证书号:		授予学位:						
学习时间	课程名称	备注	学分	成绩	课程性质	学习时间	课程名称	备注	学分	成绩	课程性质
2022-2023学年秋季学期	工程技术创新前沿		1.5	90	专业学位课	2022-2023学年秋冬学期	研究生英语		2.0	83	专业学位课
2022-2023学年秋季学期	新时代中国特色社会主义思想理论与实践		2.0	90	专业学位课	2022-2023学年春季学期	数学建模		2.0	85	专业选修课
2022-2023学年秋季学期	人工智能算法与系统		2.0	100	跨专业课	2022-2023学年春季学期	自然辩证法概论		1.0	77	专业学位课
2022-2023学年秋冬学期	高阶工程认知实践		3.0	89	专业学位课	2022-2023学年春季学期	研究生英语基础技能		1.0	67	公共学位课
2022-2023学年冬季学期	新药发现理论与实践		2.0	88	专业学位课	2022-2023学年夏季学期	工程师创新创业思维		2.0	95	专业选修课
2022-2023学年秋冬学期	工程伦理		2.0	94	专业学位课	2022-2023学年夏季学期	药品创制工程实例		2.0	86	专业学位课
2022-2023学年秋冬学期	研究生论文写作指导		1.0	77	专业选修课		硕士生读书报告		2.0	通过	
2022-2023学年冬季学期	产业技术发展前沿		1.5	89	专业学位课						

说明: 1. 研究生课程按三种方法计分: 百分制, 两级制 (通过、不通过), 五级制 (优、良、中、及格、不及格)。
2. 备注中“*”表示重修课程。

学院成绩校核章:

成绩校核人: 张梦依

打印日期: 2025-03-20



UniMap: Type-Level Integration Enhances Biological Preservation and Interpretability in Single-Cell Annotation

Haitao Hu, Yue Guo, Fujing Ge, Hao Yin, Hao Zhang, Zhesheng Zhou, Fangjie Yan, Qing Ye, Jialu Wu, Ji Cao,* Chang-Yu Hsieh,* and Bo Yang*

Integrating single-cell datasets from multiple studies provides a cost-effective way to build comprehensive cell atlases, granting deeper insights into cellular characteristics across diverse biological systems. However, current data integration methods struggle with interference in partially overlapping datasets and varying annotation granularities. Here, a multiselective adversarial network is introduced for the first time and present UniMap, which functions as a “discerner” to identify and exclude interfering cells from various data sources during dataset integration. Compared to other state-of-the-art methods, UniMap emphasizes type-level integration and proves to be the best model for preserving biological variability, achieving noticeably higher accuracy in single-cell automated annotation under various circumstances. Additionally, it enhances interpretability by revealing shared and domain-specific cell types and providing prediction confidence. The efficacy of UniMap is demonstrated in terms of identifying new cell types, creating high-resolution cell atlases, annotating cells along developmental trajectories, and performing cross-species analysis, underscoring its potential as a robust tool for single-cell research.

microenvironments.^[4] Recently, several cutting-edge large-scale studies have aimed to create comprehensive reference maps of all human cells, other species, or specific disease contexts.^[5–7] However, due to the high costs associated with such studies, a more economical approach involves integrating existing datasets to create more comprehensive atlases for specific, less-studied biological context.

The major issue is how to effectively align Single cell RNA sequencing (scRNA-seq) data from different domains, given the substantial batch effects across datasets acquired from diverse studies, sequencing methods, and sample origins.^[8] Current data integration methods aim to align various datasets within a low-dimensional space, while enabling automated cell annotation for unannotated datasets through label transfer in the shared feature space, thereby advancing research in the field.

For example, classic machine learning integration methods, such as Harmony,^[9] Seurat v5,^[10] fastMNN,^[11] Scanorama,^[12] and BBKNN,^[13] use strategies like soft clustering or mutual nearest neighbor techniques to integrate similar cells across different datasets. Subsequently, deep learning models based on auto-encoders or variational auto-encoders, including scVI,^[14] scANVI,^[15] trVAE,^[16] Scalex,^[17]

1. Introduction

Single-cell sequencing technologies, with their unprecedented resolutions, have fundamentally revolutionized our understanding of basic biological processes and various disease states, such as cell differentiation,^[1,2] cell communication,^[3] and the tumor

H. Hu, Y. Guo, F. Ge, H. Yin, H. Zhang, Z. Zhou, F. Yan, J. Cao, B. Yang
Institute of Pharmacology and Toxicology
Zhejiang Province Key Laboratory of Anti-Cancer Drug Research
College of Pharmaceutical Sciences
Zhejiang University
Hangzhou 310058, China
E-mail: caoji88@zju.edu.cn; yang924@zju.edu.cn

H. Hu, H. Yin, H. Zhang
Polytechnic Institute of Zhejiang University
Zhejiang University
Hangzhou 310015, China

Q. Ye, J. Wu, C.-Y. Hsieh
College of Pharmaceutical Sciences
Zhejiang University
Hangzhou, Zhejiang 310058, P. R. China
E-mail: kimhsieh@zju.edu.cn

J. Cao, C.-Y. Hsieh, B. Yang
The Innovation Institute for Artificial Intelligence in Medicine
Zhejiang University
Hangzhou 310018, China

J. Cao, B. Yang
Engineering Research Center of Innovative Anticancer Drugs
Ministry of Education
Hangzhou 310000, China

J. Cao
Center for Medical Research and Innovation in Digestive System Tumors
Ministry of Education
Hangzhou 310020, China

B. Yang
School of Medicine
Hangzhou City University
Hangzhou 310015, China

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/advs.202410790>

© 2025 The Author(s). Advanced Science published by Wiley-VCH GmbH. This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/advs.202410790

经检索《Web of Science》的《Science Citation Index Expanded (SCI-EXPANDED)》数据库，下述论文被《SCI-EXPANDED》收录。（检索时间：2025年3月17日）

第 1 条，共 1 条

标题:UniMap: Type-Level Integration Enhances Biological Preservation and Interpretability in Single-Cell Annotation
作者:Hu, HT(Hu, Haitao);Guo, Y(Guo, Yue);Ge, FJ(Ge, Fujing);Yin, H(Yin, Hao);Zhang, H(Zhang, Hao);Zhou, ZS(Zhou, Zhesheng);Yan, FJ(Yan, Fangjie);Ye, Q(Ye, Qing);Wu, JL(Wu, Jialu);Cao, J(Cao, Ji);Hsieh, CY(Hsieh, Chang-Yu);Yang, B(Yang, Bo);

来源出版物:ADVANCED SCIENCE 提前访问日期:FEB 2025 DOI:10.1002/advs.202410790 出版年:2025 FEB 27 2025

入藏号:WOS:001433962100001

文献类型:Article; Early Access

地址:

[Hu, Haitao; Guo, Yue; Ge, Fujing; Yin, Hao; Zhang, Hao; Zhou, Zhesheng; Yan, Fangjie; Cao, Ji; Yang, Bo] Zhejiang Univ, Inst Pharmacol & Toxicol, Coll Pharmaceut Sci, Zhejiang Prov Key Lab Anticanc Drug Res, Hangzhou 310058, Peoples R China.

[Hu, Haitao; Yin, Hao; Zhang, Hao] Zhejiang Univ, Polytech Inst, Hangzhou 310015, Peoples R China.

[Ye, Qing; Wu, Jialu; Hsieh, Chang-Yu] Zhejiang Univ, Coll Pharmaceut Sci, Hangzhou 310058, Zhejiang, Peoples R China.

[Cao, Ji; Hsieh, Chang-Yu; Yang, Bo] Zhejiang Univ, Innovat Inst Artificial Intelligence Med, Hangzhou 310018, Peoples R China.

[Cao, Ji; Yang, Bo] Minist Educ, Engr Res Ctr Innovat Anticanc Drugs, Hangzhou 310000, Peoples R China.

[Cao, Ji] Minist Educ, Ctr Med Res & Innovat Digest Syst Tumors, Hangzhou 310020, Peoples R China.

[Yang, Bo] Hangzhou City Univ, Sch Med, Hangzhou 310015, Peoples R China.

通讯作者地址:

Cao, J; Yang, B (通讯作者), Zhejiang Univ, Inst Pharmacol & Toxicol, Coll Pharmaceut Sci, Zhejiang Prov Key Lab Anticanc Drug Res, Hangzhou 310058, Peoples R China.; Hsieh, CY (通讯作者), Zhejiang Univ, Coll Pharmaceut Sci, Hangzhou 310058, Zhejiang, Peoples R China.; Cao, J; Hsieh, CY; Yang, B (通讯作者), Zhejiang Univ, Innovat Inst Artificial Intelligence Med, Hangzhou 310018, Peoples R China.; Cao, J; Yang, B (通讯作者), Minist Educ, Engr Res Ctr Innovat Anticanc Drugs, Hangzhou 310000, Peoples R China.; Cao, J (通讯作者), Minist Educ, Ctr Med Res & Innovat Digest Syst Tumors, Hangzhou 310020, Peoples R China.; Yang, B (通讯作者), Hangzhou City Univ, Sch Med, Hangzhou 310015, Peoples R China.

电子邮件地址:caoji88@zju.edu.cn; kimhsieh@zju.edu.cn; yang924@zju.edu.cn

IDS号:Y7O2E

eISSN:2198-3844

注:

1. 以上检索结果来自 CALIS 查收查引系统。
2. 以上检索结果均得到委托人及被检索作者的确认。



教育部科技查新工作站(2019)

检索人(姓名): 李佳

2025年3月17日



(19) 国家知识产权局



(12) 发明专利申请



(10) 申请公布号 CN 117789837 A

(43) 申请公布日 2024.03.29

(21) 申请号 202311830227.4

(22) 申请日 2023.12.28

(71) 申请人 浙江大学

地址 310058 浙江省杭州市西湖区余杭塘路866号

(72) 发明人 杨晓春 胡海涛 曹戟 何俏军

严芳洁 郭越 尹昊

(74) 专利代理机构 杭州求是专利事务有限公司 33200

专利代理人 赵航丽

(51) Int. Cl.

G16B 50/10 (2019.01)

G16B 40/20 (2019.01)

G16B 25/10 (2019.01)

G16B 50/30 (2019.01)

权利要求书2页 说明书6页 附图3页

(54) 发明名称

基于选择性域鉴别器的单细胞自动注释方法及装置

(57) 摘要

本发明提供基于选择性域鉴别器的单细胞自动注释方法及装置。该方法对包含批次信息的单细胞参考数据和单细胞查询数据进行统一处理,获得低维编码特征;通过选择性鉴别器,将不同来源的低维编码特征对齐,达到降低数据噪音的目的;以参考数据的注释信息为真值标签训练预测模型;将查询数据输入到训练好的模型中,获得与参考数据注释信息对应的查询数据的预测标签。本发明方法能够充分有效地利用多个来源的参考数据信息,自动对未知细胞类型的单细胞数据进行精准注释,有助于提高单细胞注释的效率和准确性。本发明方法有效解决单细胞测序数据的分析和注释问题。



CN 117789837 A



国家知识产权局

310013

浙江省杭州市西湖区古墩路701号紫金广场C座1506室 杭州求是
专利事务所有限公司
赵杭丽(13606643397)

发文日:

2023年12月28日



申请号: 202311830227.4

发文序号: 2023122801175700

专利申请受理通知书

根据专利法第28条及其实施细则第38条、第39条的规定,申请人提出的专利申请已由国家知识产权局受理。现将确定的申请号、申请日等信息通知如下:

申请号: 2023118302274

申请日: 2023年12月28日

申请人: 浙江大学

发明人: 杨晓春,胡海涛,曹戟,何偲军,严芳洁,郭越,尹昊

发明创造名称: 基于选择性域鉴别器的单细胞自动注释方法及装置

经核实,国家知识产权局确认收到文件如下:

权利要求书 1份2页,权利要求项数: 5项

说明书 1份7页

说明书附图 1份3页

说明书摘要 1份1页

专利代理委托书 1份2页

发明专利请求书 1份5页

实质审查请求书 文件份数: 1份

申请方案卷号: 赵-231-98

提示:

- 1.申请人收到专利申请受理通知书之后,认为其记载的内容与申请人所提交的相应内容不一致时,可以向国家知识产权局请求更正。
- 2.申请人收到专利申请受理通知书之后,再向国家知识产权局办理各种手续时,均应当准确、清晰地写明申请号。

审查员: 自动受理
联系电话: 010-62356655

审查部门: 初审及流程管理部



200101 纸件申请,回函请寄: 100088 北京市海淀区蓟门桥西土城路6号 国家知识产权局专利局受理处收
200210 电子申请,应当通过专利业务办理系统以电子文件形式提交相关文件。除另有规定外,以纸件等其他形式提交的文件视为未提交。

证书号第6946397号



发明专利证书

发明名称：基于协同注意力的可解释性抗癌药物协同预测方法及装置

发明人：杨波;郭越;曹戟;何俏军;胡海涛

专利号：ZL 2023 1 1155808.2

专利申请日：2023年09月08日

专利权人：浙江大学

地址：310058 浙江省杭州市西湖区余杭塘路866号

授权公告日：2024年04月26日

授权公告号：CN 117275608 B

国家知识产权局依照中华人民共和国专利法进行审查，决定授予专利权，颁发发明专利证书并在专利登记簿上予以登记。专利权自授权公告之日起生效。专利权期限为二十年，自申请日起算。

专利书记载专利权登记时的法律状况。专利权的转移、质押、无效、终止、恢复和专利权人的姓名或名称、国籍、地址变更等事项记载在专利登记簿上。



局长
申长雨

申长雨



证书号 第6946397号

专利权人应当依照专利法及其实施细则规定缴纳年费。本专利的年费应当在每年09月08日前缴纳。未按照规定缴纳年费的，专利权自应当缴纳年费期满之日起终止。

申请日时本专利记载的申请人、发明人信息如下：

申请人：

浙江大学

发明人：

杨波;郭越;曹戟;何俏军;胡海涛