

DOI:10.19651/j.cnki.emt.2518806

基于改进 ResNet-50 的智能流式细胞分析方法^{*}

许孔晨 汤怀锋 杨海倩 苏新 陆晓春

(河海大学信息科学与工程学院 常州 213022)

摘要: 流式细胞术是一种广泛应用于生命科学研究和临床诊断的高通量检测技术,然而传统流式细胞仪在面对复杂的数据维度和噪声干扰严重的情况下表现不佳,为了提高流式细胞术处理多参数,高维数据的能力,同时保证数据的及时性和准确性,提出了一套智能流式细胞分析系统。该系统涵盖流式细胞系统硬件设备设计、流式细胞系统软件设计和智能流式细胞分析系统算法设计。硬件层面设计了一个基于 FPGA 和 ARM 协同工作的实时数据采集系统,软件层面构建了一个嵌入式 Linux 架构,提出了一套预处理、解析、和批次归一化方法,在流式数据智能分析层面中,引入自组织映射算法进行数据降维,结合深度学习领域的改进残差网络,构建了一种 SE-ResNet-50 深度卷积神经网络模型。实验表明,SE-ResNet-50 模型与原模型 ResNet-50 相比,总体准确率提升了 4%,精确率提升了 3.8%。通过 SOM 与 SE-ResNet-50 的协同工作流程,有效地处理流式细胞仪采集的大量高维数据。实验结果证明了该方法的优越性。

关键词: 嵌入式系统;自组织映射;深度学习;残差网络;流式细胞术

中图分类号: TN929.52 文献标识码: A 国家标准学科分类代码: 520.60

Intelligent flow cytometry analysis system method based on improved ResNet-50

Xu Kongchen Tang Huaifeng Yang Haiqian Su Xin Lu Xiaochun

(College of Information Science and Engineering, Hohai University, Changzhou 213022, China)

Abstract: Flow cytometry is a high-throughput detection technique widely used in life science research and clinical diagnostics. However, conventional flow cytometers exhibit suboptimal performance when handling complex data dimensions and severe noise interference. To enhance the capability of flow cytometry in processing multi-parameter, high-dimensional data while ensuring timeliness and accuracy, this study proposes an intelligent flow cytometry analysis system. The system encompasses hardware design, software architecture, and algorithmic frameworks for flow cytometry. At the hardware level, a real-time data acquisition system was developed based on the cooperative operation of FPGA and ARM. On the software side, an embedded system with a Linux-based architecture was constructed, incorporating a suite of preprocessing, parsing, and batch normalization methods. For intelligent flow cytometry data analysis, a self-organizing mapping algorithm was introduced for dimensionality reduction, combined with an improved residual network from the field of deep learning, resulting in the development of an SE-ResNet-50 deep convolutional neural network model. Experimental results demonstrate that the SE-ResNet-50 model achieves a 4% improvement in overall accuracy and a 3.8% increase in precision compared to the original ResNet-50. The collaborative workflow integrating SOM and SE-ResNet-50 effectively processes the vast amounts of high-dimensional data acquired by flow cytometry. The findings validate the superiority of the proposed approach.

Keywords: embedded system; self-organizing mapping; deep learning; residual network; flow cytometry

0 引言

流式细胞分析技术实现了对单细胞群体物理特性及表

面标志物的高通量、多参数检测,同时具备目标细胞亚群分选功能,为免疫、肿瘤学等领域的定量研究提供了关键技术支撑。

流式细胞术作为一种多参数分析技术,通过流体动力学聚焦系统和光信息检测模块,对单细胞水平的物理特性如细胞体积和细胞内部颗粒复杂度以及生物化学指标如核算含量、蛋白表达量、酶动力学参数、粒子稳态、pH 梯度及膜点位变化等实现高通量同步检测。该技术通过荧光标记探针与特定的生物分子相结合,配合基于光电倍增管的多通道信号采集系统,可以精确获取到细胞群体的异质性数据,为血液系统疾病诊断、免疫表型分析、细胞功能研究提供客观实验数据^[1]。

流式细胞分析仪是一种基于流式细胞术的现代细胞分析仪器,结合了光学、电子信息学、流体动力学、细胞生物学和计算机技术。细胞或微粒通过流体动力学聚焦,依次经过激光探测区,激光激发产生的前向散射、侧向散射和多色荧光信号通过分光系统处理后,由高灵敏度的光强探测器同步检测,最后将提取的数据传输至计算机进行显示和统计分析。目前大多数流式细胞仪仍采用激光作为光源,这在一定程度上限制了其性能提升。目前市面上最先进的流式细胞仪主要集中在高参数分析、高速分选、自动化及多激光配置方面。

当前市场大多为传统流式、质谱流式与光谱流式三大技术线路,传统流式具备成熟的五参数检测体系(前向/侧向散射十三色荧光)在临床常规检测中保持主导地位;质谱流式借助金属同位素标记技术突破荧光光谱重叠限制,实现单细胞 50+ 标志物的同步检测;光谱流式则通过全光谱解析技术显著降低补偿误差,在免疫图谱绘制等复杂实验中展现独特优势。此外,成像流式细胞仪是一种结合传统流式细胞术和荧光显微镜成像技术的先进生物分析工具。与常规流式细胞仪仅能提供单一的散射光和荧光信号不同,成像流式细胞仪能够在高速流动的细胞上进行高分辨率成像,并结合多个荧光标记物,从而提供细胞的形态学信息和分子标记物的空间分布。这使得成像流式细胞仪能够同时获得细胞的定量数据和图像信息,极大地拓宽了流式细胞术在细胞分析中的应用范围。该技术在细胞生物学、免疫学、癌症研究等领域具有重要应用,尤其是在细胞亚群的精确分类、细胞功能研究和病理分析中,能够提供更为丰富和多维度的数据支持。通过高效处理海量细胞图像数据,成像流式细胞仪为高通量、细胞层面的精确研究提供了强有力的技术保障。

流式细胞分析术经过技术不断地迭代和系统的优化,但是在高维数据场景下面临分类精度不足、噪声干扰严重(如技术性信号波动和生物性非特异结合)、算法对稀疏群体敏感性低、样本制备标准化困难等挑战。国内外研究正通过 AI 算法优化(如深度学习降维)、多模态数据整合、硬件升级(如光谱流式)和标准化数据库建设(如人类细胞图谱)寻求突破,但计算资源依赖、模型可解释性及成本问题仍需解决。相比于国外流式细胞仪的长久发展,国内流式细胞仪才处于初步研发阶段,目前国内正在进行自主研究

工作的院校和研究所正逐步增多。

1 相关工作

史亮^[2]借助定量分析模块完成生物分子的绝对计数,显著提升临床诊断敏感性与科研数据标准化水平。赵书涛等^[3]融合了光电传感、流控工程、生物信息学等学科交叉技术,将原始数据导入智能分析平台进行可视化建模和量化解析。Dimitriadis 等^[4]指出流式技术正向更灵敏、更高通量的方向发展。曾希等^[5]提出新一代流式细胞仪可实现远程设备状态监控与异常预警。张子涵等^[6]提出更先进的激光管理技术和高灵敏度、高速度的传感器,流式细胞仪的检测能力将大幅提升,Rees 等^[7]通过告诉图像采集装置,可以在细胞高速流动过程中记录下多角度、动态的形体学信息,Kanno 等^[8]在不影响总荧光强度的采集的基础上,进一步获取荧光信号的时间维度信息,为细胞状态分析提供更加全面和精细的参数支撑,进一步提升其精度和可靠性。Yamamoto 等^[9]利用完全数字化的检测方法为流式细胞仪提供更精确的量化能力。Wong 等^[10]通过数字化技术,流式细胞仪实现信号的定量化。Lewis 等^[11]提出 AI 技术,能使流式细胞仪更加的智能化,能够实时分析复杂数据并自动化实验过程。方平龙^[12]设计了流式细胞仪中发光源和光束成形系统,保证了光源的稳定性和光束的精确性。雷天继^[13]解决了流式细胞仪激光器平台上光学器件受环境温度影响发生微小偏移的问题。袁宇鹏等^[14]聚焦小直径微生物的研究,Lugli 等^[15]提出传统的数据分析难以满足目前高效、精准的分析需求。为了解决这一问题,Li 等^[16]介绍深度学习是一种强大的大数据分析方法,能够自动提取大量数据中的复杂特征,根据这样的性质,为流式细胞术数据的分析提供了一种新的解决方案,针对高维复杂流式细胞数据的分析需求,Halpin^[17]说明了“机器智能”概念,Orhan 等^[18]介绍了感知机模型的实现奠定了神经网络的基础。随后出现 K 近邻算法^[19]、贝叶斯概率图模型等无监督学习方法^[20]。作为机器学习的重要分支,深度学习展现出显著优势^[21],其中 LeNet-5 的提出被认为卷积网络发展的里程碑^[22]。李宁等^[23]提出的方法能减少模型复杂度的同时,显著提升 FPGA 加速器的计算效率和资源利用率。目前市面上最先进的流式细胞仪主要集中在高参数分析、高速分选、自动化及多激光配置方面。如图 1 所示:BD FACSymphony™ S6 分选型流式细胞仪,其价格预估高达 120 w~150 w 美元。

根据以上突破性研究,为顺应流式细胞仪的迅速发展趋势,进一步提升流式细胞术面对复杂细胞数据和噪声影响严重时的实验效率和精度,本文旨在发挥嵌入式技术在实时性高、多任务处理能力强以及开发灵活性等方面的优势,针对流式细胞数据采集与处理硬件设备、流式细胞数据采集与处理系统软件、智能流式细胞分析系统算法等模块进行改进,设计并开发一套智能流式细胞分析系统。该系



图 1 BD FACSymphony™ S6 分选型流式细胞仪

Fig. 1 BD FACSymphony™ S6 cell sorter flow cytometer

统通过 ARM 与 FPGA 的协同工作, 基于机器学习与深度

学习相结合的智能分析框架, 实现智能流式细胞检测系统的信号采集以及数据的自动分析功能。此外, 本文基于 SOM 和改进 Res-Net 的血细胞亚型分类方法, 能够有效地处理流式细胞仪采集的大量高维数据, 并通过深度学习算法自动化实现 B 细胞淋巴瘤亚型的高效分类。实验结果证明了该方法的优越性, 为流式细胞技术在临床应用中的推广提供了有力的支持。

2 智能流式细胞分析系统模型

2.1 流式细胞数据采集与处理硬件设备设计

流式细胞数据采集与处理装备总体设计方案如图 2 所示, 设计方案充分考虑了流式细胞仪的数据采集、处理、显示和存储需求, 采用了 STM32、FPGA、i.MX6ULL 等多个核心模块进行协同工作。

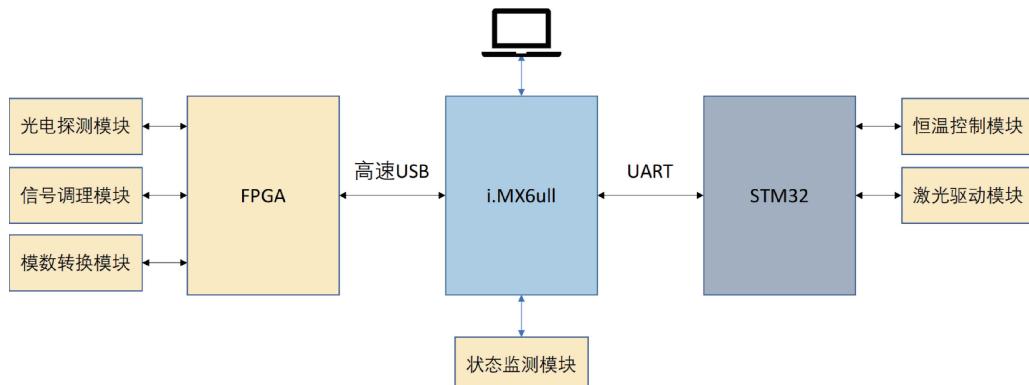


图 2 系统装备设计总体方案

Fig. 2 Overall plan for system equipment design

如图 2 所示, 主控电路选用 STM32F407 控制模块负责激光的驱动, FPGA 负责流式细胞仪中的信号采集与处理, i.MX6ULL 负责数据库存储与显示屏状态监测。激光驱动模块通过闭环控制确保激光二极管在稳定状态下工作, 从而提高检测准确性。光电探测模块选用 Thorlabs 公司的 FD05D 型号 InGaAs 光电二极管作为探头, 采用跨阻放大器(transimpedance amplifier, TA)实现电流-电压转换功能。另外, 采用 Analog Devices 公司生产的 AD8610 型运算放大器构建同相比例放大电路架构, 确保信号放大的线性度和稳定性。模数转换模块采用 AD7667 作为转换芯片, 在信号进入 AD7667 之前, 本文设计了一个由运算放大器和 RC 电路组成的滤波器, 用来抑制噪声的干扰, 同时还设计了一个电压跟随器, 并通过电阻分压的方式来使得输入电压调整到芯片允许承受的范围之内。最后, 面向恒温控制模块, 本文设计了基于热电制冷器(TEC)的精密温控方案, 通过 NTC 热敏电阻采集温度数据, 并结合外部模拟 PI 调节电路与专用温控芯片, 实现对系统关键部件的精确温度控制。

2.2 流式细胞数据采集与处理系统软件设计

为构建高效、稳定且具有良好扩展性的流式细胞数据

采集与处理系统软件平台, 实现流式细胞仪硬件资源与操作系统的高效对接, 确保数据采集、显示和实时监控功能, 本文基于 Linux 平台的系统软件总体架构做出以下设计。驱动程序设计中, 本文改进了 ARM 硬件设备包括 LCD 显示屏以及 USB 模块等。在液晶屏的驱动设计中, 系统将采集模块接收到的数据通过通信机制传输到 ARM 板子之后进行显示。为保证数据传输的连续性和实时性, 利用 USB 设备驱动程序, 优化设备分配的资源, 及时释放移除的 USB 设备所占用的资源。

在系统数据库设计中, 选用 SQLite 作为系统的数据库工具。该数据库能在嵌入式系统中进行高效的数据存储和查询。另外, 为降低复杂系统实现与仿真的难度, 采用 Xilinx 公司推出的 Vivado 集成开发环境作为设计平台。为确保 FPGA 内部逻辑正确运行, 选用 MMCM 时钟管理器, 利用其闭环调节机制确保输出时钟的长期稳定性。数据采样模块采用有限状态机(FSM)架构。数据缓存模块采用双通道 DDR3 存储架构, 基于硬件平台集成两片 Micron 公司的 MT41J256M16HA 存储器芯片, 存储器接口严格遵循 JEDEC DDR3-1600 规范, 并采用 1.5 V 核心电压供电, 确保与 FPGA 内嵌的 DDR3 控制器(通过 MIG

IP 核实现)的兼容性。最后,在数据处理模块中,光电监测系统的信号处理架构采用双级滤波方案,针对信号的频谱特征差异,系统采用混合信号处理策略。

3 智能流式细胞分析系统算法设计

3.1 基于 EMD 分解和小波去噪的信号降噪算法

考虑到数字滤波器处理后,信号的频域范围得到了进一步的规范,信号质量有所改善。然而,PMT 采集到的流式细胞信号通常较弱,且在采集和传输过程中不可避免的受到多种噪声的干扰,简单地滤波处理难以完全消除这些噪声。为解决这一问题,本文设计了一种降噪算法。用于非线性和非平稳信号的分析,利用经验模态分解算法,将原始信号分解为若干个本征模态函数 IMF。每组拆分后的 IMF 应同时满足以下两个条件:

1) 极值点与过零点的数量差值不超过 1。

$$|N_e - N_z| \leq 1 \quad (1)$$

2) 对信号进行三次样条拟合并得到极值信息后,所得到的由最大值点构成的上包络线和最小值点构成的下包络线在任意时刻的平均值为 0。

$$\frac{u(t) + l(t)}{2} = 0, \forall t \in [t_a, t_b] \quad (2)$$

经过经验模态分解,计算信号 $x(t)$ 的极大值和极小值,使用三次样条插值方法计算出局部极大值点和极小值点,根据极大值点计算得到上包络线 $u(t)$,根据极小值点计算出下包络线 $l(t)$ 。并求出平均包络线 $m(t)$ 。用原始信号 $x(t)$ 减去均值包络线 $m(t)$,得到:

$$h(t) = x(t) - m(t) \quad (3)$$

判断 $h(t)$ 是否满足 IMF 的两个约束条件,如果满足,则该 $h(t)$ 为一个 IMF 分量 $C_i(t)$;否则将该 $h(t)$ 代替原信号 $x(t)$,并重复上述步骤进行下一轮提取。每提取一个 IMF 后,都要从原信号 $x(t)$ 中减去这个 IMF,直到原信号剩下一个或多个残差信号 $r_n(t)$ 。经过 EMD 分解后,将得到若干个 IMF 和一个残差信号。将所有信号叠加,重建回原始信号 $x(t)$:

$$x(t) = \sum_{i=0}^N C_i(t) + r_n(t) \quad (4)$$

为实现信号特征与噪声成分的有效分离,利用小波变换阈值去噪,选取适当的阈值,对于大于阈值的小波系数保留;对于小于阈值的系数置零。常用阈值并未考虑噪声的具体程度,因此对阈值进一步修正,引入 σ_j 来对阈值进行调整:

$$\sigma_j = \text{median}(|d_j(i)|) / 0.6745 \quad (5)$$

小波阈值去噪方法中,硬阈值与软阈值格局特点,传统阈值方法在平衡噪声抑制与保真方面存在固有局限性,为新型自适应阈值函数的研发提供重要的研究切入点。

本文使用经典的小波阈值函数,应用 EMD 分解法对流式细胞信号分解成多个本征模态函数 IMF,然后对每个

IMF 分量进行小波阈值去噪。实验中通过加入 5 dB 的高斯白噪声来测试降噪算法的效果。

3.2 B 细胞淋巴瘤流式数据预处理方法

本文中用于 B 细胞淋巴瘤分类的数据集基于 MFC 数据,收集自 2016 年 1 月 1 日~2018 年 12 月 31 日之间的患者样本,这些样本来自 MLL 慕尼黑白血病实验室的公开数据集,下载自 Harvard DataVerse。数据集包含来自 8 种不同 B 细胞淋巴瘤亚型的样本:慢性淋巴细胞白血病(CLL)、前淋巴细胞白血病(PL)、滤泡性淋巴瘤(FL)、毛细胞白血病(HCL)、淋巴浆细胞淋巴瘤(LPL)、套细胞淋巴瘤(MCL)、边缘区淋巴瘤(MZL)和单克隆 B 细胞增多症(MBL)。数据集中还包含了健康对照样本。该数据集被分为训练集和测试集。训练集包含 18 274 个样本,测试集包含 2 348 个样本,这些测试集样本是在 2018 年 7 月 1 日之后收集的。所有的诊断都通过其他测试如组织学和荧光原位杂交(FISH)进行确认。

针对后续的数据分析,本文利用支持 FCS3.0 的解析工具提取出每个细胞的多参数数据,将这些数据通过补偿和归一化,使得数值矩阵在数值范围和分布上适合放入后续的训练 SOM 网络中。在流式细胞术数据分析中,使用 fcsmerge 工具将同一样本的多个 FCS 文件合并成统一的文件,fcsmerge 工具在处理大规模流式细胞数据时提供高效的整合方法,并能够基于现有数据为缺失的标记填充合理的值,保证合并后的一致性,可以直接用于后续的特征提取和深度学习建模。

利用自组织映射 SOM 神经网络模型,进行动态调整,在保留输入数据的拓扑结构的同时,捕捉输入数据的特征,实现高维数据的降维和可视化。本文算法思路就是将高维的流式细胞术数据映射到二维网格上并生成网格热图,每一个网格的节点代表一个细胞群体,节点的值可以是细胞数量或标记物的表达强度,生成的网格热图放入分类模型,最终输出的结果就是细胞的具体类别。

3.3 基于 SE-ResNet-50 的 B 细胞淋巴瘤亚识别方法

SE-Net 通常作为可嵌入的子模块集成到现有网络架构,SE-Net 通过全局信息压缩、通道权重学习、特征重校准三个阶段调整通道的重要性,使网络能够自适应地强化关键特征,从而提升模型在复杂任务中的性能。

全局信息压缩,通过全局平均池化(GAP)将特征图 U 的空间维度压缩为通道级统计量,生成通道描述符 $z \in \mathbf{R}^{1 \times 1 \times C}$,其数学表达式为:

$$z_c = F_{sq}(uc) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (6)$$

其中, Z_c 表示第 c 个通道的压缩结果, $uc(i, j)$ 为位置 (i, j) 的特征值。

在通道权重学习阶段,通过两层全连接网络学习通道间的非线性依赖关系,生成通道权重向量 $s \in \mathbf{R}^{1 \times 1 \times C}$,具体的计算公式表达如下:

$$s = F_{ex}(\mathbf{z}, \mathbf{W}) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z})) \quad (7)$$

其中, $\mathbf{W}_1 \in \mathbf{R}^{\frac{C}{r} \times C}$ 和 $\mathbf{W}_2 \in \mathbf{R}^{C \times \frac{C}{r}}$ 为全连接层权重矩阵 (r 为缩减比率), $\delta(\cdot)$ 表示 ReLU 激活函数, $\sigma(\cdot)$ 为 Sigmoid 函数, 将权重约束至 $[0, 1]$ 区间。

在特征重校准操作阶段, 将学习到的通道权重 s 与原始特征 U 逐通道相乘, 完成特征校准:

$$\tilde{U} = F_{scale}(U, s) = s \odot U \quad (8)$$

其中, \odot 表示通道逐通道乘法 (Hadamard 积), 最终输出 \tilde{U} 作为下一层的输出。

ResNet 解决梯度消失和网络退化的问题, 但在处理 B 细胞淋巴瘤细胞亚型分类任务时发现, 随着网络深度增加, ResNet 纯在特征提取瓶颈和过拟合风险, 因此本文对传统的 ResNet 架构进行针对性的改进, 在 ResNet-50 架构中引入 SE-Net, 形成 SE-ResNet-50 改进模型, SE-Net 模块被嵌入到每个残差组块的特定位置。该模块采用“压缩-激励-加权”的三阶段处理流程, SOM 的拓扑保持性压缩流式数据维度, SE 模块通过通道自适应加权增强了模型对关键标志物的敏感性。SOM 与 SE-ResNet-50 的结合在特征可解释性、无监督-监督混合学习、拓扑感知降维方面具有独特优势, 特别适合需要鲁棒性、小样本或动态降维的任务(如医疗影像、工业检测)。而其他方法(如 PCA、CBAM)更侧重单一维度的优化, 缺乏这种多层次协同能力。实际应用中需根据数据特点和计算资源权衡选择。

4 实验设计与分析

本文针对流式细胞采集与处理硬件设备、流式细胞采集与处理系统软件和智能流式细胞分析系统算法的有效性, 将改进的设计原理和流程进行展示, 并将改进的算法与传统算法进行数据实验和对比。

图 3 为恒温控制电路原理, R_x 代表负温度系数 (NTC) 热敏电阻的实时阻抗值, 该参数具有显著的温度依赖性。

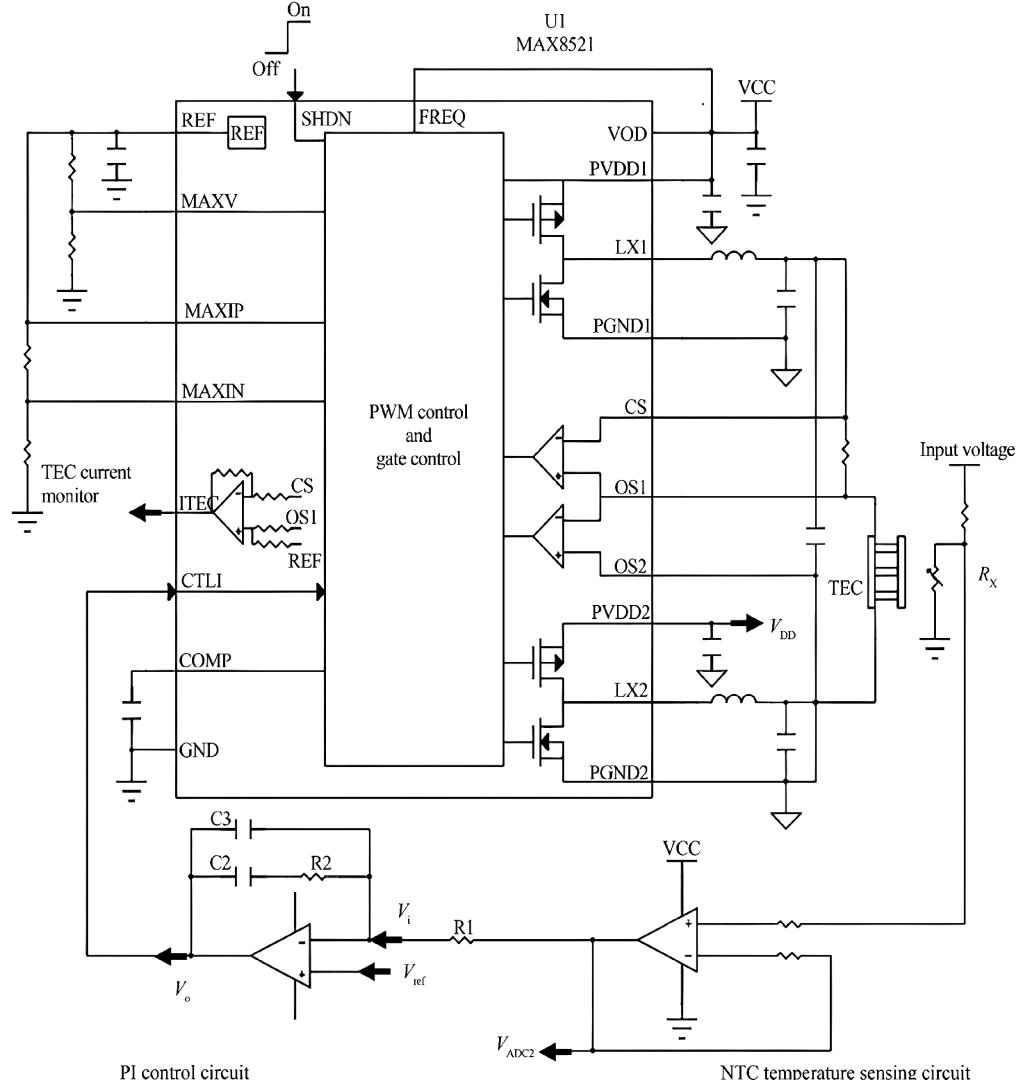


图 3 恒温控制电路原理图

Fig. 3 Schematic diagram of constant temperature control circuit

系统通过串联分压电路将阻抗变化线性转换为电压型信号。该电路创新性的进行以下措施：配置低输出阻抗的电压跟随器，有效消除信号传输损耗；采用高精度 ADC 模块进行信号数字化；通过 STM32 微控制器实现温度数据的实时处理，实现工作温度的连续监测与显示。此外，为提升热电制冷器（TEC）的温度控制精度与稳定

性，本设计采用基于串联补偿网络的拟比例-积分（PI）调节电路。

图 4 为智能流式细胞分析系统界面布局，该界面为用户提供实时信息，展示了流式细胞仪采集的 PMT（光电倍增管）信号，图中波动反应检测到的细胞信号强度，用户可以通过按钮进行数据的查询和上传。

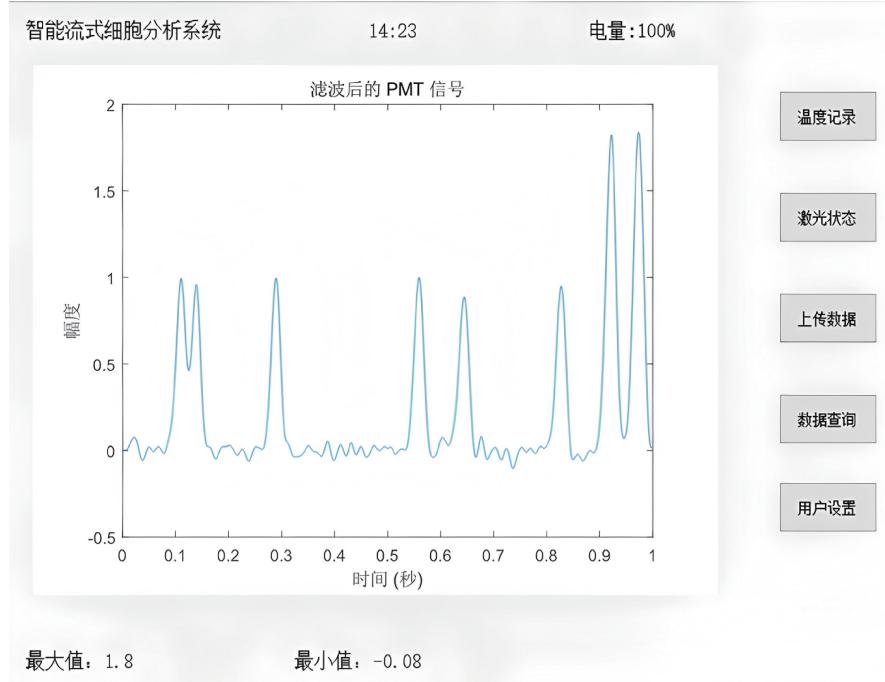


图 4 智能流式细胞分析系统界面布局

Fig. 4 Interface layout of intelligent flow cytometry analysis system

图 5 为 MMCM 内部框图，内部集成电压控制振荡器（VCO）、相位频率检测器（PFD）和多重反馈环路等关键电路，通过数字模拟混合设计实现高性能时钟管理。该混合模式时钟管理器相较于其他时钟管理器具备优异的抖动抑制特性和可编程滤波功能。

图 6 验证了汉宁窗作为 FIR 数字滤波器的窗函数能够实现更好的噪声抑制效果。其中，图 6(a)给出了 Matlab 中仿真加窗后的光信号频谱和荧光信号频谱，图 6(b)给出了向散射光信号和荧光信号中加入的汉宁窗频率响应。

图 7 给出了硬阈值函数和软阈值函数曲线图，硬阈值通过将低于设定阈值的小波系数直接归零处理，其算法实现简便且能保持较好的去噪效果。但该函数在临界点处存在不连续性，易导致重构信号产生震荡现象。软阈值函数通过引入连续收缩机制有效解决临界点不连续问题，然而其全局收缩特性会造成信号幅值衰减，导致重构信号与原始信号存在能量差异。

图 8 中第 1 行表示 PMT 采集到的流式细胞脉冲源信号，第 2 行表示加入 5 dB 高斯白噪声后的脉冲信号图，第 3 行是通过软阈值法降噪后的信号，第 4 行是通过硬阈值法降噪后的信号。

图 9 给出了 fcsmerge 工作原理，图中展示了两个不同的试管（Tube1 和 Tube2），每个试管包含不同的标记物（如 CD10、FMC7、CD5 等）。这些标记物在每个试管中的表现不同，但每个试管的数据文件（FCS 文件）中都包含了单独的细胞事件和特征。通过 fcsmerge 工具，将试管的数据合并成一个单一的 FCS 文件，确保标志物数据的一致性。此外，fcsmerge 能够基于现有数据为缺失的标志填充合理的值，从而确保了每个样本的标志物信息完整无缺。

本文创新性提出一种基于无监督学习的神经网络模型用于将多参数流式细胞数据降为二维矩阵，从而实现对细胞群体分布的可直观可视化。如图 10 所示：通道参考 SOM 权重平面图，通过将每个节点的某一通道（如 CD5）的数值映射为颜色深浅，可以生成热力图。图中亮色区域对应高表达，暗色区域对应低表达，这为后续细胞亚型的自动分类模型提供了输入。在训练完成后，每个节点的权重向量代表该区域内细胞的典型特征。可以通过选择某个标志物的值生成热力图。通过观察 SOM 热力图，可以识别出不同细胞群体的分布，还可以为后续的分类模型提供有效的特征输入，实现自动化的 B 细胞淋巴瘤亚型分类。

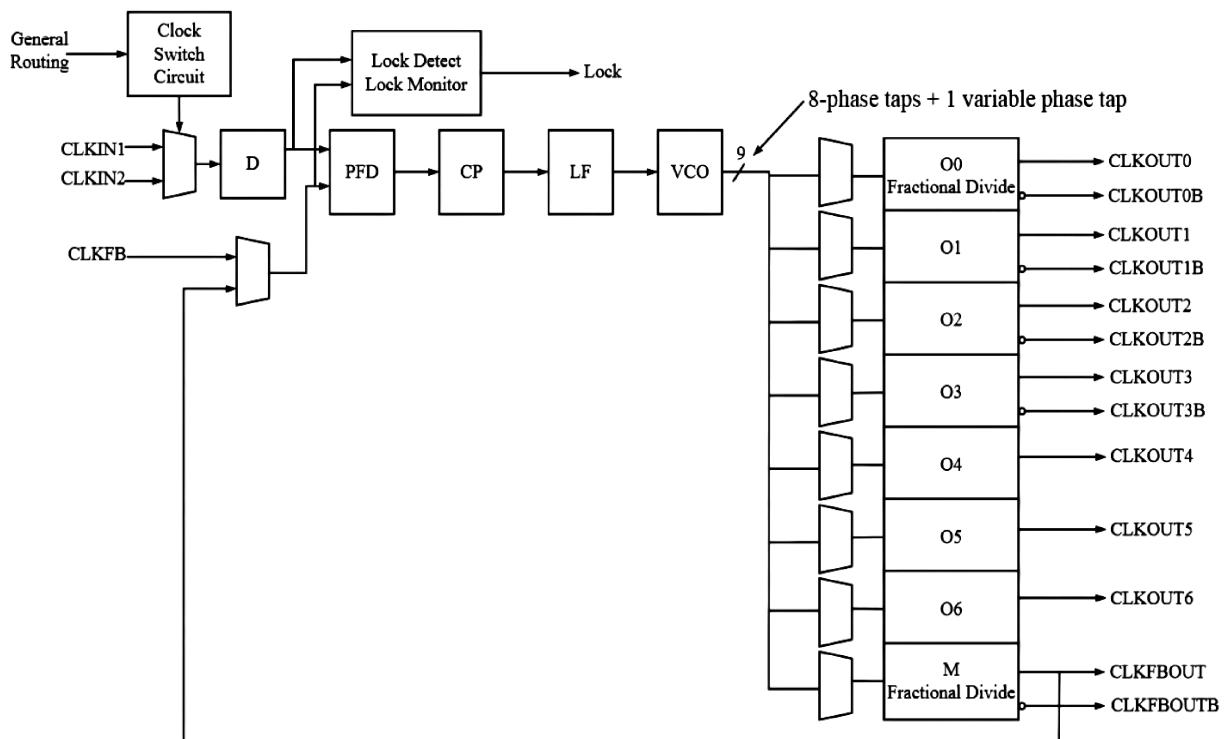


图 5 MMCM 内部框图

Fig. 5 Internal block diagram of MMCM

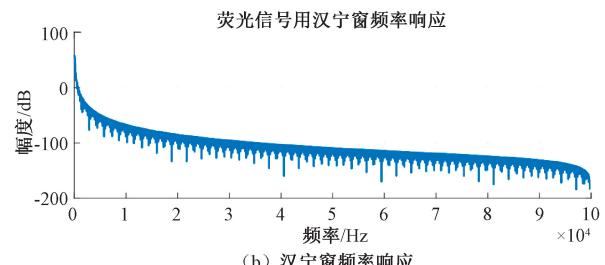
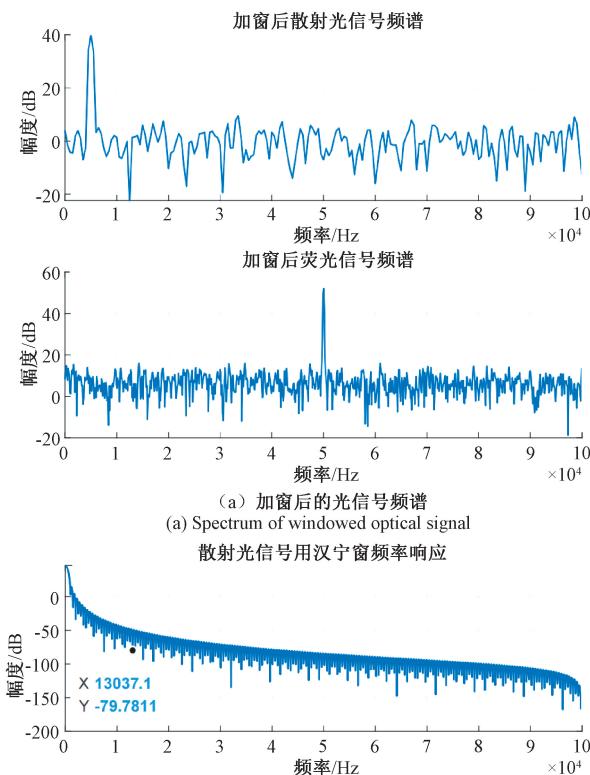


图 6 频谱与频率响应

Fig. 6 Spectrum and frequency response

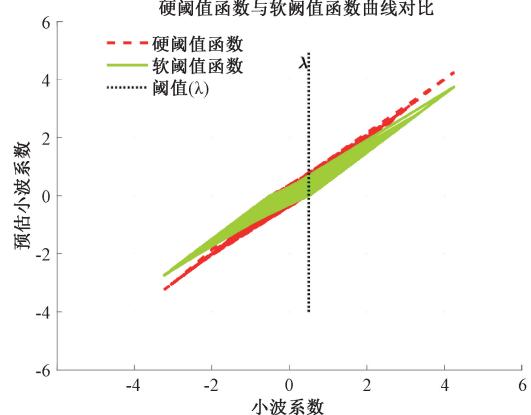


图 7 硬阈值函数和软阈值函数曲线图

Fig. 7 Curve plots of hard threshold function and soft threshold function

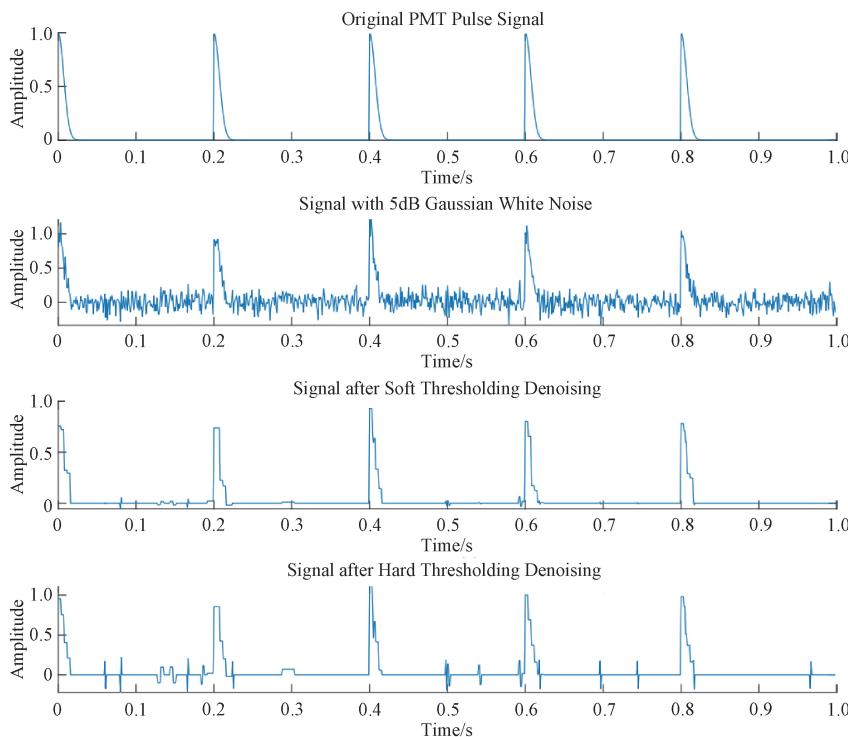


图 8 流式细胞信号重构图

Fig. 8 Flow cytometry signal reconstruction map

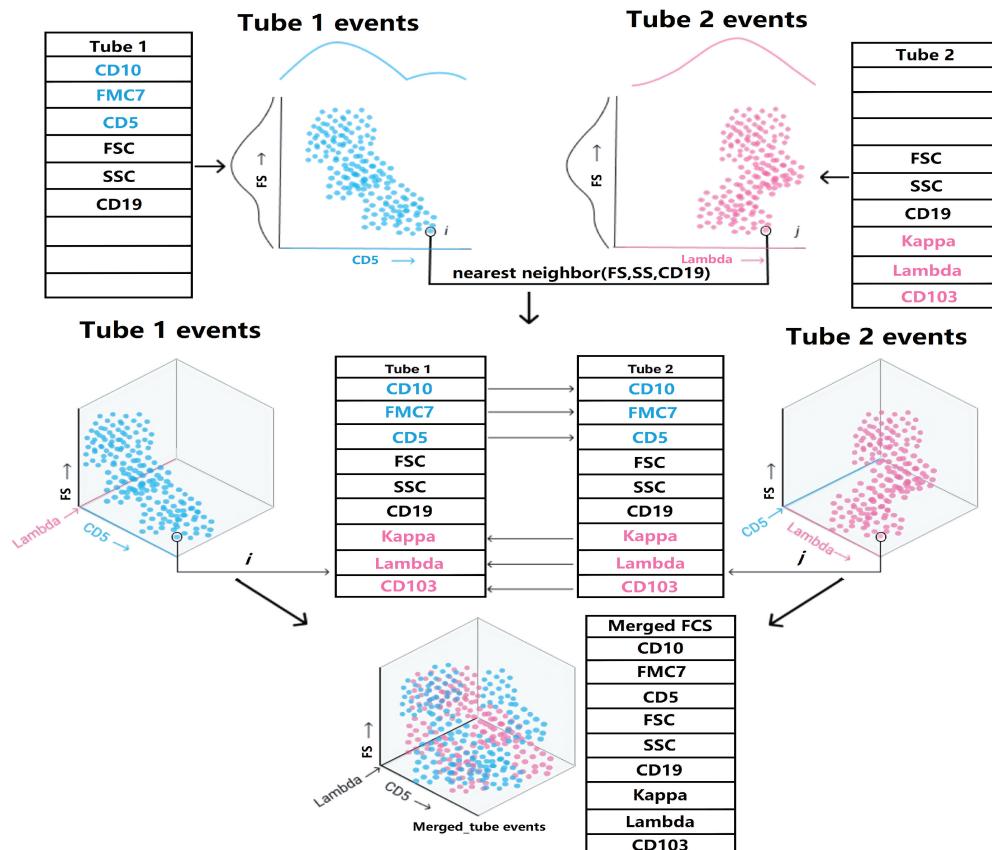


图 9 fcsmerge 工作原理示意图

Fig. 9 Schematic diagram of the working principle of fcsmerge

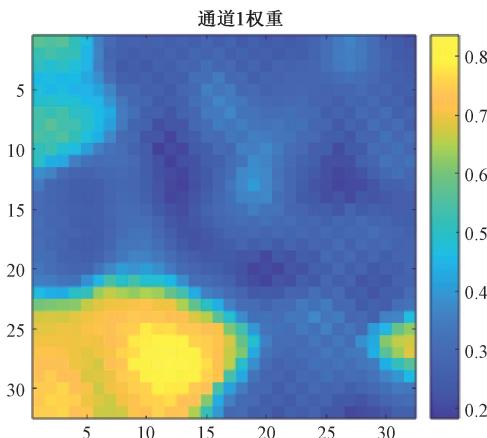


图 10 通道参考 SOM 权重平面图

Fig. 10 Channel reference SOM weight plane diagram

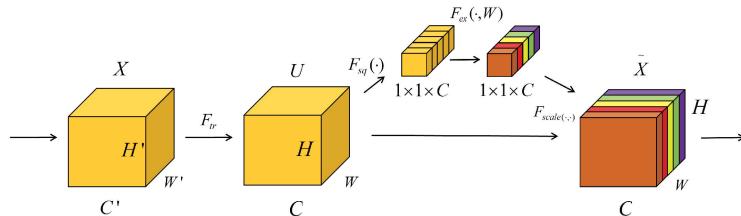


图 11 SE-Net 基本结构

Fig. 11 SE-Net basic structure

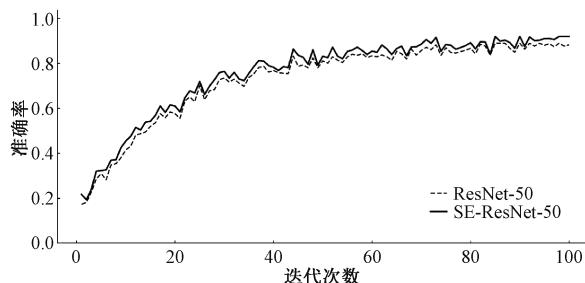


图 12 模型准确率曲线对比图

Fig. 12 Comparison chart of model accuracy curves

进,针对原模型在高阶特征表征能力受限和对稀有细胞亚型的识别率下降的问题上,在 ResNet-50 架构中引入 SE-Net,形成如图 13 所示:SE-ResNet-50 网络模型结构图。

表 1 给出了 ResNet-50 与 SE-ResNet-50 的模型对比

图 11 给出了 SE-Net 基本结构图,SE-Net 在保留主干网络拓扑的前提下,仅对特征图进行轻量级重校准。输入特征 X (维度为 $H' \times W' \times C'$)经卷积变换后输出中间特征 U (维度为 $H \times W \times C$)。该结构经过创新性改进,通过全局信息聚合和通道权重学习,增强判别性特征通道的响应,同时抑制冗余或噪声通道,从而提升网络的表征能力。

图 12 给出了 ResNet-50 与 SE-ResNet-50 模型的准确率曲线对比图,展示了 ResNet-50 网络模型与 SE-ResNet-50 网络模型在训练过程中准确率的变化曲线。虚线表示 ResNet-50 模型的准确率变化曲线,而实线则表示引入 SE-Net 的 SE-ResNet-50 模型的准确率变化曲线。从准确率的变化趋势对比中可以明显看出,SE-ResNet-50 模型在训练初期和最终的准确率均优于 ResNet-50 模型。

本文基于 SE-Net 模块对 ResNet 架构进行针对性改

数据,更加直观的呈现模型的精度表现,本表列出了训练过程中模型的总体准确率(OA)、精确率(Precision)、召回率(Recall)以及特异度(Specificity)等各项关键指标。

表 1 模型对比表

Table 1 Model comparison

模型	总体准确率	精确率	召回率	特异度
ResNet-50	87.7	88.3	87.8	99.4
SE-ResNet	91.7	92.1	91.8	99.6
模型提升	4.0	3.8	4.0	0.2

图 14 为 SE-ResNet-50 模型的混淆矩阵,经分析表明,该模型对 B 细胞淋巴瘤亚型整体分类效果良好,平均找回率高达 79%,其中,HCL(1.00)和正常细胞(0.99)近乎完美区分。

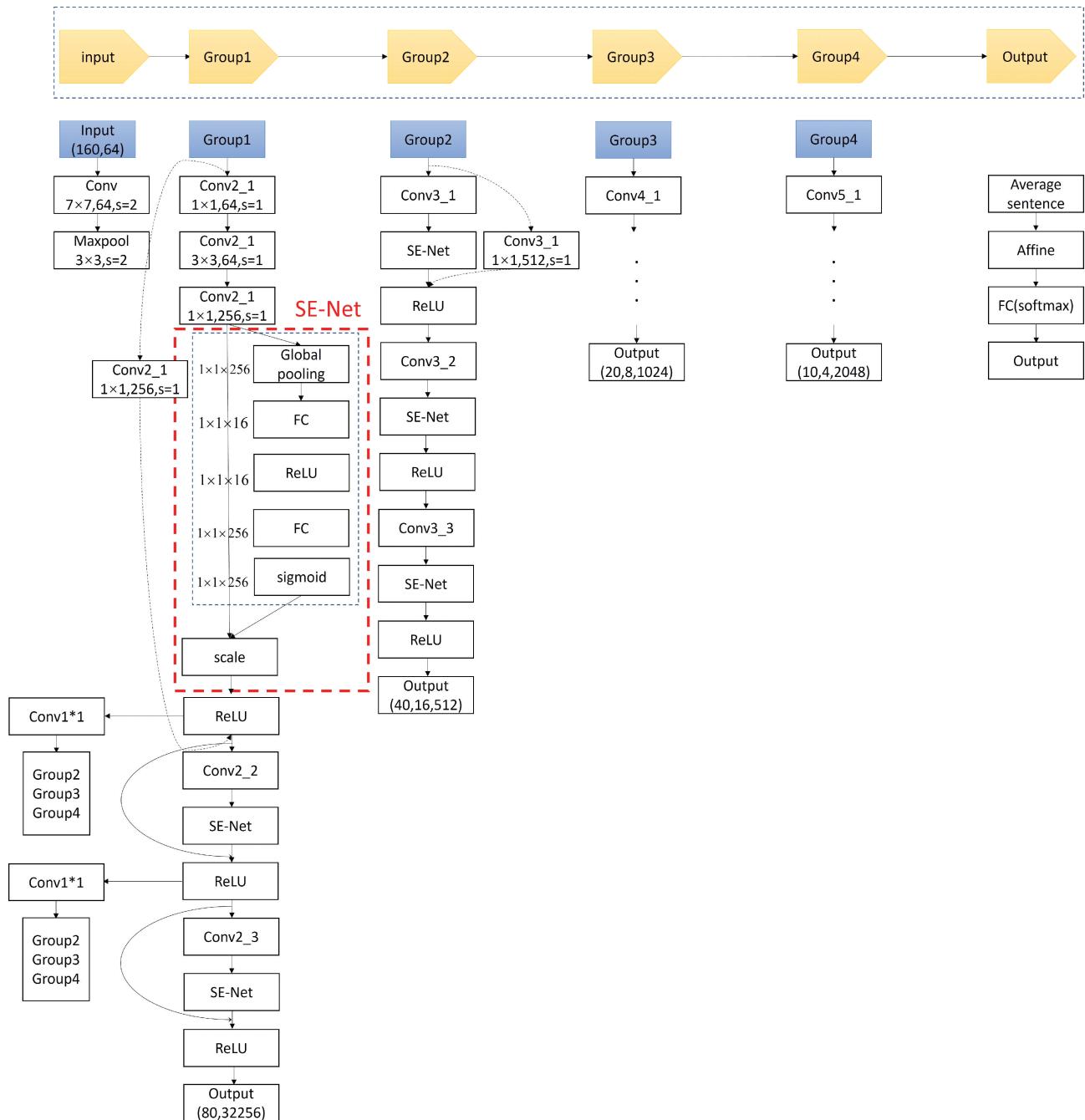


图 13 SE-ResNet-50 网络模型结构

Fig. 13 SE-ResNet-50 network model structure

		Confusion matrix								
		CLL	MBL	MCL	PL	LPL	MZL	FL	HCL	normal
True label	CLL	0.83	0.14	0.01	0.01	0.00	0.00	0.00	0.01	0.00
	MBL	0.10	0.82	0.02	0.04	0.00	0.00	0.00	0.00	0.02
	MCL	0.02	0.08	0.67	0.18	0.03	0.01	0.00	0.01	0.00
	PL	0.09	0.03	0.15	0.62	0.04	0.05	0.01	0.01	0.00
	LPL	0.00	0.01	0.05	0.01	0.65	0.13	0.04	0.00	0.11
	MZL	0.00	0.00	0.01	0.06	0.15	0.61	0.01	0.01	0.15
	FL	0.00	0.00	0.00	0.05	0.00	0.00	0.70	0.00	0.25
	HCL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
	normal	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.99
	CLL	MBL	MCL	PL	LPL	MZL	FL	HCL	normal	Predicted label

图 14 SE-ResNet-50 混淆矩阵
Fig. 14 SE-ResNet-50 confusion matrix

5 结 论

本文提出基于嵌入式系统和深度学习技术的流式细胞仪数据采集与自动化分析平台,设计了一套基于 FPGA 和 ARM 协同工作的实时数据采集系统,能够高效的采集和处理流式细胞仪的数据。基于嵌入式 Linux 平台构建完整的软件架构,涵盖了系统启动、设备驱动、数据管理与可视化等模块,解决了流式细胞术数据中的噪声和偏差问题。引入自组织映射(SOM)算法进行数据降维,并基于改进的残差网络(ResNet)构建了深度卷积神经网络模型,显著提高了细胞亚型分类的准确性。

参考文献

- [1] ADAN A, ALIZADA G, KIRAZ Y, et al. Flow cytometry: Basic principles and applications [J]. Critical Reviews in Biotechnology, 2017, 37(2): 163-176.
- [2] 史亮. 流式细胞仪的发展历史及其原理和应用进展研究[J]. 中国设备工程, 2021(12): 13-14.
SHI L. The development history of flow cytometry and its principle and application progress[J]. China Equipment Engineering, 2021(12): 13-14.
- [3] 赵书涛, 武晓东, 王策, 等. 流式细胞仪的原理, 应用及最新进展[J]. 现代生物医学进展, 2011, 11(22): 4378-4381.
- ZHAO SH T, WU X D, WANG C, et al. Principles of flow cytometry, applications and recent advances[J]. Advances in Modern Biomedicine, 2011, 11(22): 4378-4381.
- [4] DIMITRIADIS S, DOVA L, KOTSIANIDIS I, et al. Imaging flow cytometry: Development, present applications, and future challenges[J]. Methods and Protocols, 2024, 7(2): 28.
- [5] 曾希. 流式细胞仪在线监测系统的设计[J]. 医疗卫生装备, 2020, 41(12): 44-47.
ZENG X. Design of a flow cytometry online monitoring system [J]. Medical and Health Equipment, 2020, 41(12): 44-47.
- [6] 张子涵, 陈真诚, 陈望, 等. 基于流式细胞仪的激光器驱动系统研究与设计[J]. 光子学报, 2024, 53(6): 139-151.
ZHANG Z H, CHEN ZH CH, CHEN W, et al. Research and design of a laser-driven system based on flow cytometry [J]. Acta Photonica Sinica, 2024, 53(6): 139-151.
- [7] REES P, SUMMERS H D, FILBY A, et al. Imaging flow cytometry [J]. Nature Reviews Methods Primers, 2022, 2(1): 86.
- [8] KANNO H, HIRAMATSU K, MIKAMI H, et al. High-throughput fluorescence lifetime imaging flow

- cytometry [J]. Nature Communications, 2024, 15(1): 7376.
- [9] YAMAMOTO M, ROBINSON J P. Quantum approach for nanoparticle fluorescence by sub-ns photon detection [J]. Cytometry Part A, 2021, 99(2): 145-151.
- [10] WONG N, KIM D, ROBINSON Z, et al. K-means quantization for a web-based open-source flow cytometry analysis platform[J]. Scientific Reports, 2021, 11(1): 6735.
- [11] LEWIS J E, COOPER L A, JAYE D L, et al. Automated deep learning-based diagnosis and molecular characterization of acute myeloid leukemia using flow cytometry[J]. Modern Pathology, 2024, 37(1): 100373.
- [12] 方平龙. 流式细胞仪及光学系统设计[J]. 中国新技术新产品, 2024(14): 31-33.
- FANG P L. Flow cytometry and optical system design[J]. China's New Technology and New Products, 2024(14): 31-33.
- [13] 雷天继, 陈望, 陈真诚, 等. 多通道流式细胞仪激光器平台的热分析及设计[J]. 激光与光电子学进展, 2024, 61(21): 326-334.
- LEI T J, CHEN W, CHEN ZH CH, et al. Thermal analysis and design of multi-channel flow cytometry laser platform [J]. Advances in Lasers and Optoelectronics, 2024, 61(21): 326-334.
- [14] 袁宇鹏, 蒋卓雅, 李顺波, 等. 微生物聚焦的微流控芯片研究[J]. 压电与声光, 2025, 47(1): 75-80.
- YUAN Y P, JIANG ZH Y, LI SH B, et al. Research on microfluidic chip focusing by microorganisms[J]. Piezoelectric and Acousto-Optic, 2025, 47(1): 75-80.
- [15] LUGLI E, PINTI M, NASI M, et al. Subject classification obtained by cluster analysis and principal component analysis applied to flow cytometric data[J]. Cytometry Part A: The Journal of the International Society for Analytical Cytology, 2007, 71 (5): 334-344.
- [16] LI Y, MAHJOUBFAR A, CHEN C L, et al. Deep cytometry: Deep learning with real-time inference in cell sorting and flow cytometry[J]. Scientific Reports, 2019, 9(1): 11088.
- [17] HALPIN H. The semantic web: The origins of artificial intelligence redux [C] Third International Workshop on the History and Philosophy of Logic, Mathematics, and Computation(HPLMC-04 2005), 2004.
- [18] ORHAN U, HEKIM M, OZER M. EEG signals classification using the K-means clustering and a multilayer perceptron neural network model [J]. Expert Systems with Applications, 2011, 38 (10): 13475-13481.
- [19] TARLOW D, SWERSKY K, CHARLIN L, et al. Stochastic k-neighborhood selection for supervised and unsupervised learning[C]. International Conference on Machine Learning, 2013: 199-207.
- [20] LU J, PLATANIOTIS K N, VENETSANOPoulos A N. Face recognition using LDA-based algorithms [J]. IEEE Transactions on Neural networks, 2003, 14(1): 195-200.
- [21] 周飞燕, 金林鹏, 董军, 等. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
- ZHOU F Y, JIN L P, DONG J, et al. A review of convolutional neural networks[J]. Chinese Journal of Computers, 2017, 40(6): 1229-1251.
- [22] WEI G, LI G, ZHAO J, et al. Development of a LeNet-5 gas identification CNN structure for electronic noses[J]. Sensors, 2019, 19(1): 217.
- [23] 李宁, 肖昊. 基于FPGA的稀疏卷积神经网络加速器设计[J]. 电子测量技术, 2024, 47(5): 1-8.
- LI N, XIAO H. Design of sparse convolutional neural network accelerator based on FPGA [J]. Electronic Measurement Technology, 2024, 47(5): 1-8.

作者简介

许孔晨,硕士研究生,主要研究方向为智慧医疗、边缘计算、移动通信等。

汤怀锋,学士,主要研究方向为智慧医疗、边缘计算、移动通信等。

杨海倩,硕士研究生,主要研究方向为智慧医疗、边缘计算、移动通信等。

苏新,教授,硕士生导师,主要研究方向为智慧医疗、边缘计算、移动通信等。

陆晓春(通信作者),讲师,博士,主要研究方向为嵌入式人工智能、测量技术及仪器。

E-mail: Lu213022@163.com