

分类号: _____

密级: 公开

UDC _____

学号: 2022023414

华南师范大学

South China Normal University

硕士学位论文

(专业学位)

FRET 双杂交分析数据处理软件的设计与开发

学位申请人: 魏智强

专业名称: 光电信息工程

研究方向: 荧光图像分析

所在院系: 光电科学与工程学院

导师姓名及职称: 胡敏 实验师

2025 年 3 月 10 日

FRET 双杂交分析数据处理软件的设计与开发

专业名称： 光电信息工程

申请者： 魏智强

导师： 胡敏 实验师

摘 要

福斯特共振能量转移 (Förster Resonance Energy Transfer, FRET) 技术被广泛用于探究活细胞中生物大分子的动态相互作用过程, 如生物大分子的构象变化、蛋白质相互作用、信号通路中蛋白质间的调节机制等, 在研究生命科学基础问题、疾病的发生发展和药物研发等方面具有广阔的应用前景。FRET 双杂交分析不仅可以在活细胞内原位实时测量蛋白质间真实的相互作用, 还能给出蛋白质结合的相对亲和力及化学计量比。目前主流的活细胞 FRET 双杂交分析方法是基于以供体 / 受体为中心的 FRET 效率 (E_D 或 E_A) 与自由受体 / 供体浓度 (A_{free} 或 D_{free}) 之间关系, 即 $E_D - A_{free}$ 或 $E_A - D_{free}$ 进行朗缪尔拟合的方法 (本文称为朗缪尔拟合方法 L-FRET), 但是其鲁棒性较差, 对细胞样本制备和数据分析的要求极为苛刻。为此, 有团队发展了基于 FRET 效率与受供体浓度比 (R_C), 即 $E_D - R_C$ 或 $E_A - 1/R_C$ 线性分离的方法 (本文称为线性拟合方法 DC-FRET)。实验证明, DC-FRET 具有较好的鲁棒性, 并且降低了对细胞样本的要求。

关键词：FRET; FRET 定量分析; 荧光图像; FRET 双杂交分析; 自动数据处理

Design and Development of Data Processing Software for FRET Two-Hybrid Analysis

Major: Optoelectronic Information Engineering

Name: Zhiqiang Wei

Supervisor: Min Hu

ABSTRACT

Urban flooding is one of the major environmental challenges affecting human society. Understanding the mechanisms of how geographic factors influence flooding is crucial for developing effective disaster mitigation strategies. This study innovatively integrates street-view imagery data with machine learning methods to extract surface features such as building ratio (SVB), vegetation ratio (SVV), road ratio (SVR), and terrain ratio (SVT), while incorporating variables like population density and land use. Comparative analyses were conducted using XGBoost, RandomForest, LightGBM, and Linear Regression models. The results indicate that: (1) The XGBoost model performed the best, with an MAE of 0.2089, R^2 of 0.0596, and RMSE of 0.4083, significantly outperforming other models and demonstrating its advantages in capturing the nonlinear relationships of flood factors. (2) SHAP value analysis revealed that impervious surface ratio and population density are key factors influencing flood susceptibility, while surface features extracted from street-view imagery also showed significant contributions, further highlighting the spatial heterogeneity of urban flooding. By introducing street-view imagery data, this study effectively improves the accuracy and interpretability of flood susceptibility prediction, providing new technical tools and decision-making support for urban planning and flood risk management.

KEY WORDS: FRET, FRET quantitative analysis, FRET two-hybrid assay, Automated data processing

目 录

摘 要	I
ABSTRACT	III
1 绪论	1
1.1 福斯特共振能量转移 (FRET)	1
1.1.1 FRET 原理概述	1
1.1.2 FRET 在生物学中的应用	2
1.1.3 基于受体敏化发射的 FRET 定量测量方法 (E-FRET)	3
1.2 FRET 双杂交分析	5
1.2.1 基于 Langmiur 模型曲线拟合的 FRET 双杂交分析 (L-FRET)	5
1.2.2 基于 FRET 效率 E_D 和受供体浓度比 R_C 线性分离的 FRET 双杂交分析 (DC-FRET)	7
1.3 FRET 数据处理分析方法现状	8
1.3.1 传统的数据处理	8
1.3.2 基于 MATLAB 的 FRET 双杂交辅助数据处理软件	9
1.3.3 基于深度学习的 FRET 数据处理方法	10
1.4 本文的工作内容和意义	11
1.5 本文的章节安排	12
2 FRET 双杂交分析数据处理软件 (Fretha) 的设计和开发	13
2.1 引言	13
2.2 需求分析和总体设计	13
2.2.1 需求分析	13
2.2.2 模块划分	14
2.2.3 软件总体框架	14
2.2.4 开发技术选型	16
2.3 Fretha 的开发实现	17
2.3.1 成像参数设置模块	17
2.3.2 数据完备性检验模块	18
2.3.3 FRET 图像处理模块	19

2.3.4	数据控制模块	23
2.3.5	结果可视化模块	26
2.3.6	FRET 双杂交计算模块	26
2.4	本章小结	30
3	自动 FRET 双杂交分析	33
3.1	引言	33
3.2	材料与方法	33
3.2.1	细胞培养和转染	33
3.2.2	FRET 成像系统	34
3.2.3	基于明度和均匀度的自动 ROI 选取算法 (LURS)	34
3.2.4	DC-FRET 方法的数据范围选取	38
3.3	实验结果	39
3.3.1	标准质粒验证实验	39
3.3.2	模型质粒验证实验	40
3.4	本章小结	41
4	软件的应用和测试	43
4.1	引言	43
4.2	材料与方法	43
4.3	结果	45
4.3.1	使用 Fretha 手动处理数据	45
4.3.2	使用 Fretha 自动处理数据	45
4.3.3	与深度学习方法对比	45
4.4	软件的测试	45
4.4.1	成像参数设置测试	45
4.4.2	数据导入导出测试	46
4.4.3	结果保存测试	47
4.4.4	软件的性能测试	47
4.4.5	软件的稳定性测试	48
4.5	本章小结	49
5	总结与展望	51
5.1	本文的主要内容和总结	51

5.2 未来发展展望	51
参考文献	53

表 目 录

表 2.1	FRET 成像参数	18
表 2.2	FRET 图像视图类型	22
表 2.3	FRET 圈点状态栏显示内容	23
表 2.4	FretDataPiece 数据类型	25
表 3.1	FRET 成像系统参数	34
表 3.2	对标准质粒进行 3 ³ -FRET 和 E-FRET 测量的结果	39
表 3.3	模型质粒的 FRET 双杂交分析结果	41
表 4.1	切换参数对 C32V 质粒和 C4Y 质粒的 E-FRET 分析结果	46

图 目 录

图 1.1	FRET 过程示意图	1
图 1.2	基于 Matlab 的 FRET 双杂交辅助数据处理软件	9
图 2.1	FRET 双杂交分析数据处理流程	14
图 2.2	Fretha 软件总体架构	15
图 2.3	Fretha 数据层实体关联关系图	16
图 2.4	参数设置模块业务流程图	19
图 2.5	FRETscopeII 数据文件层级结构树形图	20
图 2.6	Fretha 数据完备性检验业务主流程图	20
图 2.7	Fretha 图像处理模块界面示意图	21
图 2.8	Fretha 自动圈点模块流程图	24
图 3.1	LURS 算法流程图	38
图 3.2	模型质粒验证实验结果	40

1 绪论

1.1 福斯特共振能量转移 (FRET)

1.1.1 FRET 原理概述

1948 年, Förster 首次阐述了福斯特共振能量转移 (Förster Resonance Energy Transfer, FRET) 理论。FRET 是指处于激发态的供体分子 (Donor) 通过偶极子间相互作用将一部分能量以非辐射的形式转移给邻近处于基态的受体 (Acceptor) 分子 (供受体之间的距离 R 在 0-10 nm)。FRET 的发生会使得供体的荧光淬灭和受体的荧光增强, 其中由于发生 FRET 而增强受体荧光称为敏化发射荧光。FRET 技术突破传统观测局限, 精准解析分子间相互作用与距离。在细胞生物学微观环境里, FRET 技术如同精密“分子尺”, 捕捉纳米尺度生物分子动态, 助力剖析分子结构与功能。因而, FRET 技术在细胞生物学多领域, 如细胞信号转导、蛋白质相互作用探测等广泛应用。

FRET 双杂交分析是目前唯一能够在活细胞内实现生物大分子“滴定”实验的技术。蛋白质之间相互作用的化学计量比是阐明蛋白质间相互作用机制的重要参数, FRET 双杂交分析通过结合受体中心的 33-FRET 和供体中心的 E-FRET 方法, 能简单稳健地得到活细胞中生物复合物化学计量比和相对亲和力常数。

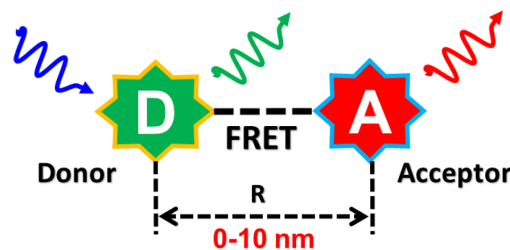


图 1.1 FRET 过程示意图

理论和实验证明, 当供受体荧光分子的距离为 r 时, 它们之间的能量转移速

率由下式给出：

$$k_T(r) = \frac{1}{\tau_D} \left(\frac{R_0}{r} \right)^6, \quad (1.1)$$

其中 τ_D 是供体荧光寿命， R_0 为 Förster 半径，由下式表示：

$$R_0^6 = 8.79 \times 10^{-5} (n^{-4} Q_D J(\lambda) \kappa^2), \quad (1.2)$$

上式中， n 表示介质的折射率， Q_D 表示供体的量子产率， $J(\lambda)$ 是光谱重叠积分， κ^2 是方向因子，表示供、受体偶极子的相对方向，一般在自由状态下取 $\kappa^2 = 2/3$ 。FRET 发生需要满足三个条件：（1） r 的数值在 R_0 附近，约 0-10nm；（2）供体的发射谱与受体的吸收谱有超过 30% 的重叠；（3）供、受体偶极子方向不互相垂直。

FRET 效率（ E ）定义为供体转移给受体的能量与供体发射的总能量的比例，是用来衡量 FRET 程度的指标，其主要和分子间距和荧光团光谱的重叠度有关。光谱有部分重叠的供受体分子间距越小，能量转移就越高效。其计算式为：

$$E = \frac{k_T(r)}{k_T(r) + \tau_D^{-1}}, \quad (1.3)$$

将1.1代入，可以看出 E 与 r 的六次方成反比的关系：

$$E = \frac{R_0^6}{R_0^6 + r^6} = \frac{1}{1 + (r/R_0)^6}. \quad (1.4)$$

1.1.2 FRET 在生物学中的应用

FRET 技术作为一项强大的工具，在生物学领域展现出了广泛且关键的应用价值。FRET 发生的条件是供、受体分子之间的距离 r 在 0-10nm 之间，这使得 FRET 技术成为一种“纳米尺”，能够有效地测量纳米级的分子间距，是的 FRET 技术在研究生物蛋白相互作用、大分子构象变化、信号通路中的蛋白质调节机制等方面得到了广泛的应用。

在细胞生物学层面，FRET 为研究细胞内蛋白质等大分子相互作用提供了独特视角。2016 年，Ben-Johny 等人利用 FRET 双杂交分析算法研究了钙离子通道

钙离子通道与钙调蛋白之间的结合。研究发现当细胞内钙离子浓度较低时，一个钙调蛋白与一个钙离子通道结合；当细胞内钙离子浓度较高时，一个钙离子通道可以同时结合两个钙调蛋白。

在疾病诊断方面，FRET 技术除了在癌症早期检测中有突出表现，在药物疗效评估等方面同样具有重要意义。2014 年，Bozza 等人通过设计特定的 FRET 生物传感器，能够直接反映癌症药物诱导癌细胞凋亡的效果，为药物的研发和筛选提供了有力的工具^[1]。相较于传统的细胞生存力测定法（Viability Assay），这种基于 FRET 的检测方法能更准确地体现药物对癌细胞的杀伤作用，避免了因无法区分细胞死亡和生长抑制而导致的结果偏差。

荧光蛋白的发展使得 FRET 技术进一步应用到各种生物研究中。1962 年，第一种绿色荧光蛋白（Green Fluorescent Protein, GFP）在维多利亚水母中被发现，由于荧光蛋白的无毒性以及稳定遗传性，可以利用基因转导技术将荧光蛋白接到感兴趣的蛋白质分子上，借助显微成像技术实时观察活细胞中目标蛋白的转位以及信号传递等生物问题。随着基因技术的发展，最先发现的 GFP 蛋白被改造出了多种 GFP 突变体，多种荧光蛋白的出现为同时追踪多个蛋白质分子间的相互作用和多种蛋白质的共定位等复杂的生物问题提供了必要条件，使得基于 FRET 显微成像技术在分子生物学和以及生物物理学的活细胞在体研究得到了广泛的应用。

1.1.3 基于受体敏化发射的 FRET 定量测量方法（E-FRET）

基于受体敏化发射的通道方法（E-FRET）具有测量速度快、无损伤的特性，被公认是最适合用于活细胞动态监测的 FRET 定量检测技术。E-FRET 方法需要对实验系统响应和荧光团的光学性质进行严格的校准，一般通过提前测量多个参照样本，然后保持在整个实验过程中系统的稳定和实验条件的一致。E-FRET 方法需要 3 个 Cube 组成的 3 个通道，分别实现供体激发供体收集（DD 通道）、供体激发受体收集（DA 通道）和受体激发受体收集（AA 通道）。受体的敏化荧光从 DA 通道测得（ I_{DA} 图像），但实际上 I_{DA} 图像包含有受体的敏化荧光、直接激发受体的荧光和供体的发射荧光这 3 个部分。为了消除后两部分串扰是 FRET 指标不依赖于荧光强度，额外的 2 个通道的光图像 I_{AA} 和 I_{DD} 需要收集。通过减掉串

扰得到敏化荧光 F_C ，由如下公式得到：

$$F_C = I_{DA} - a(I_{AA} - cI_{DD}) - d(I_{DD} - bI_{AA}). \quad (1.5)$$

其中 a, b, c, d 是串扰系数，由单转的供体样本和单转的受体样本得到，其计算公式如下：

$$a = \frac{I_{DA}(A)}{I_{AA}(A)}, \quad (1.6)$$

$$b = \frac{I_{DD}(A)}{I_{AA}(A)}, \quad (1.7)$$

$$c = \frac{I_{AA}(D)}{I_{DD}(D)}, \quad (1.8)$$

$$d = \frac{I_{DA}(D)}{I_{DD}(D)}. \quad (1.9)$$

其中， $I_{DA}(A)$ 表示单转受体（A）样本在 DA 通道测得的荧光强度，其他参量意义类似。

E-FRET 还需要测量得到荧光强度由 DD 通道转换为 DA 通道的转换因子 G ，在仪器系统参数保持不变时， G 因子可以通过敏化发射荧光 F_C 在受体光漂白后荧光减少与受体光漂白后供体在 DD 通道的荧光恢复的比值确定，其定义如下：

$$G = \frac{F_C - F_C^{post}}{I_{DD}^{post} - I_{DD}}. \quad (1.10)$$

其中 I_{DD}^{post} 是受体光漂白后受体敏化发射的荧光强度， I_{DD}^{post} 是受体光漂白后供体的荧光强度。获得敏化淬灭转化因子 G 和敏化发射荧光强度 F_C 后，供体角度的 FRET 效率 E_D 可以通过如下公式计算：

$$E_D = \frac{F_C}{F_C + G \cdot I_{DD}}. \quad (1.11)$$

为了确定待测样本中受体与供体的浓度比例，需要首先通过测量受体与供体比例为 1:1 的 FRET 固定质粒样本来确定发生 FRET 时相等浓度的供体荧光和受

体荧光的比例:

$$k = \frac{I_{DD} + F_C/G}{I_{AA}}. \quad (1.12)$$

然后使用 k 测量待测样本中的受供体浓度比 R_C , 计算方式为:

$$R_C = \frac{[A]}{[D]} = \frac{k \cdot I_{AA}}{I_{DD} + F_C/G}. \quad (1.13)$$

1.2 FRET 双杂交分析

1.2.1 基于 Langmiur 模型曲线拟合的 FRET 双杂交分析 (L-FRET)

蛋白质之间相互作用的化学计量比是阐明蛋白质间相互作用机制的重要参数, 确定化学计量比能够进一步评估蛋白质间的生物相关性, 并且能够确定其病理角色。传统的体外生化方法往往都需要从细胞中分离并且提纯大分子复合物才能进行测量, 这类体外实验方法无法在活细胞中获得结果, 而且一些大分子的复合物不容易进行分离提纯或者体外重建, 这就限制了这些体外方法的应用。FRET 双杂交分析则可以揭示细胞内大分子复合物的化学计量比。

FRET 过程改变了供受体复合物荧光发射谱的两个方面: (1) 供体荧光淬灭; (2) 受体荧光增强。从这两个方面出发, FRET 效率也可以从两种方式进行测量: (1) 通过 E-FRET 方法从供体角度测量的 FRET 效率 E_D , 即供体转移给受体的能量占供体总能量的比例; (2) 通过 3^3 -FRET 方法从受体角度测量的 FRET 效率 E_A , 即受体敏化发射的荧光量占有所有供体能量转移给受体时受体的荧光发射总量的比例。 3^3 -FRET 方法中, E_A 可由如下公式给出:

$$E_A = \frac{F_C}{a \cdot I_{AA}} \frac{\varepsilon_A(\lambda)}{\varepsilon_D(\lambda)}. \quad (1.14)$$

其中, $\varepsilon_A(\lambda)$ 是受体和供体的摩尔消光系数, a 是光谱串扰系数, 由公式1.6确定。

FRET 双杂交分析是目前唯一可以在活细胞内进行生物大分子结合“滴定”实验的技术。FRET 双杂交分析实验中, 通过不断增加受体的浓度使得每个供体都结合有受体, 从而测出饱和结合时供体角度探测的最大 E_D ($E_{D,max}$)。同样的方

法，通过不断增加供体的浓度使得每个受体都结合有供体，从而测出饱和结合时受体角度探测的最大 E_A ($E_{A,max}$)。在存在自由供体、自由受体和以 $n_D : n_A$ 比例结合的受供体复合物 ($n_D D - n_A A$, n_D 和 n_A 是供体和受体在复合物中的数量)，当所有供体都被受体结合时，每个供体预期的能量转移效率为

$$E_{D,max} = \frac{1}{n_D} \sum_{i=1}^{n_D} \sum_{j=1}^{n_A} E_{ij}, \quad (1.15)$$

当所有受体都被供体结合时，每个受体预期的能量转移效率为

$$E_{A,max} = \frac{1}{n_A} \sum_{i=1}^{n_D} \sum_{j=1}^{n_A} E_{ij}. \quad (1.16)$$

于是， $E_{A,max}$ 与 $E_{D,max}$ 的比值即为供受体的化学计量比：

$$\nu = \frac{n_D}{n_A} = \frac{E_{A,max}}{E_{D,max}}. \quad (1.17)$$

2016 年，Elisabeth S Butz 等人提出将 FRET 效率和自由受供体浓度按照 Langmiur 模型进行拟合的 FRET 双杂交分析方法。对于包含自由供体、自由受体和供受体复合物中， E_D 和 E_A 可分别与自由受体浓度、自由供体浓度相关联：

$$E_A = E_{A,max} \frac{D_{free}}{D_{free} + K_{d,EFF}}, \quad (1.18)$$

$$E_D = E_{D,max} \frac{A_{free}}{A_{free} + K_{d,EFF}}. \quad (1.19)$$

其中， $K_{d,EFF}$ 为相对解离常数， D_{free} 是自由供体的浓度， A_{free} 是自由受体的浓度。从公式 (1.18) 和公式 (1.19) 可以看出，与体外滴定实验类似，用 3³-FRET 法可得到 E_A 随自由供体浓度变化的动态滴定曲线，并得到 $E_{A,max}$ ，用 E-FRET 方法也可以得到 E_D 随自由受体浓度的动态滴定曲线，并得到 $E_{D,max}$ 。供受体复合物的化学计量比计算与公式 (1.17) 相同。

1.2.2 基于 FRET 效率 E_D 和受供体浓度比 R_C 线性分离的 FRET 双杂交分析 (DC-FRET)

传统的 L-FRET 方法存在一定的局限性。从实验条件上来看, L-FRET 需要得到 FRET 效率与自由供体 / 受体间的饱和滴定曲线, 这意味着实验人员需要精心准备不同供受体浓度比的样本, 并且这些样本的供受体浓度分布要尽可能均匀。例如, 在实际操作中, 可能需要配置一系列不同比例的供体和受体溶液, 然后分别对其进行检测和分析, 这增加了实验样本的数量。同时, 在实验数据处理时, 需要对每一个样本进行细致的处理和测量, 这极大地增加了实验人员的工作量和操作难度。另一方面, 从公式 (1.18) (1.19) 来看, A_{free} 和 D_{free} 是拟合过程中更新的中间量, 在实际的计算分析过程中, 由于这些中间量的不确定性以及数据的复杂性, 很容易出现拟合失败的情况。一旦拟合失败, 就需要重新进行实验或者调整参数, 进一步增加了实验成本。

为了解决这些问题, 在 2018 年, 本研究小组的杜孟艳等人创新性地发展了基于 FRET 效率 (E_D) 和受供体浓度比 (R_C) 线性分离的 FRET 双杂交分析方法, 简称为线性拟合方法或 Du-Chen-FRET (DC-FRET)。DC-FRET 只关注分析供体饱和结合和受体饱和结合的数据。当供体完全饱和时, 即 $D_{free} \gg K_d$, 此时受体被完全结合, 以下公式成立:

$$E_A = E_{A,max}, \quad (1.20)$$

$$E_D = E_{A,max} \cdot R_C. \quad (1.21)$$

此时 E_D 与 R_C 成正比, 且斜率为 $E_{A,max}$ 。当受体饱和时, 即 $A_{free} \gg K_d$, 此时供体被完全结合, 以下公式成立:

$$E_D = E_{D,max}, \quad (1.22)$$

$$E_A = E_{D,max} \cdot 1/R_C. \quad (1.23)$$

此时 E_D 与 R_C 成正比, 且斜率为 $E_{A,max}$ 。求得 $E_{A,max}$ 和 $E_{D,max}$ 后, 供受体复合物

中的化学计量比计算同公式 (1.17)。

DC-FRET 方法相比 L-FRET 方法有着显著的优势。L-FRET 由于数据来源的不确定性以及拟合过程的复杂性，可能会导致拟合结果的偏差较大。在 DC-FRET 方法中，实验人员只需要准备 R_C 比较大的样本和 R_C 比较小的样本即可，无需像 L-FRET 那样准备处于中间分布的大量样本。这样的样本选择方式极大地简化了实验流程，减少了样本的制备和处理时间，从而大大减少了实验人员的工作量。在 DC-FRET 中，用于线性拟合的 E_D 和 R_C 数据是由 E-FRET 方法定量测量得到的。E-FRET 方法具有较高的准确性和稳定性，能够提供可靠的数据，通过这种方式得到的数据进行线性拟合，其结果也更加稳定可靠。

1.3 FRET 数据处理分析方法现状

1.3.1 传统的数据处理

传统 FRET 双杂交数据处理过程操作复杂^[2]，操作较为费时。科研人员总共需要经过图像分析、FRET 定量计算、FRET 双杂交分析处理等步骤，且每一步处理均较为费时。FRET 双杂交图像分析依赖人工处理图像，需要人工选取上百个左右的典型荧光区域作为 ROI，再计算每个 ROI 的灰度值得到荧光信号值和背景值等信息，耗时 1-3 h；根据标定的灰度值进行异常值滤去和确定 FRET 效率需要经历 5 个步骤，耗时约 30 min - 1 h；使用 Excel 的规划求解对测得的 FRET 数据拟合 Langmuir 模型需要经历 14 个步骤，耗时约 30 min。

传统的数据处理需要在多种不同的科研软件工具之间频繁进行数据转移，增加了科研人员的学习成本。在图像处理和人工标记荧光区域时，研究人员一般使用显微镜公司提供的图像处理软件如蔡司 ZEN、滨松 HCLImage 等；在 FRET 定量计算时，需要从图像处理软件中导出来自 FRET 三通道的荧光强度等数据，再通过 Excel 进行公式设定完成批量 FRET 定量计算；在双杂交分析计算中，工作人员使用 Excel 的规划求解功能，设定好数据拟合的误差评估函数以及约束条件，从而使用规划求解功能进行数据拟合，更新迭代参数减小误差，通过 Matlab 或 Python 等语言编写解析来自 Excel 的数据，然后使用

在 FRET 计算中，工作人员使用 Excel 预定各个单元格的计算公式，在手动圈

点采集前景和背景荧光强度之后，输入数据到 Excel 表格中，后续单元格会按照绑定的计算公式自动计算敏化发射荧光和 FRET 效率等数据。此外，人工处理尤其是手动圈点步骤的操作标准规范相对模糊，不同科研人员在操作过程中因为主观因素影响到数据的可复现性。人工数据处理也无法对原始图片充分挖掘数据信息，同一样本可能需要多个拍摄多个视野才能保证数据量满足后续数据处理分析需求。

传统的 FRET 双杂交分析计算依赖 Excel 等专业数据处理软件，这类软件功能丰富强大，但是没有针对 FRET 双杂交技术数据处理而简化或优化，往往要求科研人员拥有丰富的数据处理经验，增加了科研人员的学习成本和处理时间，限制了 FRET 双杂交分析的应用和推广。

1.3.2 基于 MATLAB 的 FRET 双杂交辅助数据处理软件

2019 年，孙晗等人在研发自动化多模态 FRET 显微成像仪^[3] 时，开发了一款用于 FRET 双杂交分析的辅助数据处理软件。该软件将 FRET 双杂交算法封装为函数，可自动处理输入的 FRET 数据，完成 FRET 双杂交分析计算。

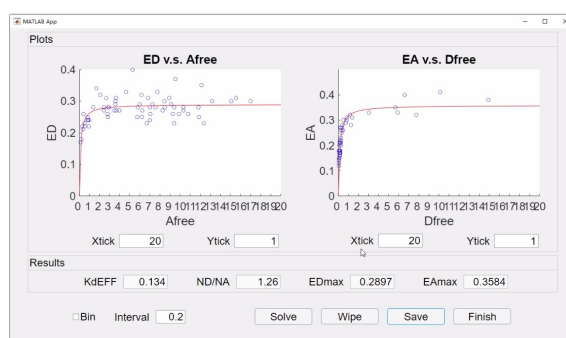


图 1.2 基于 Matlab 的 FRET 双杂交辅助数据处理软件

具体操作为，用户首先要把从自动 E-FRET 测量模块中所获取的原始数据导入到双杂交分析软件中。紧接着，需要在软件界面上，通过人工操作选取感兴趣的区域（Region of Interest, ROI）。这一操作是后续数据分析的关键步骤，所选区域将生成用于 FRET 双杂交分析的数据。随后，软件通过封装的函数算法，自动进行规划求解计算，最终为用户输出蛋白间的化学计量比以及亲和力等信息。

然而，这款软件在实际应用中也存在一定的局限性。作为 E-FRET 测量模块的后续辅助数据处理软件，它高度依赖前置数据处理模块的数据导入，自身无法

直接对 FRET 三通道原始图像进行数据处理工作。在处理结果展示方面，该软件仅以较为简单的方式显示计算结果信息，无法满足丰富的可视化展示或者科研绘图等需求。此外，数据处理环节目前依然依赖人工手动进行 ROI 标注，这种人工操作方式不仅耗时费力，而且标注的准确性可能会受到操作人员主观因素的影响，整体处理速度较慢，难以满足大规模数据快速处理的需求。该款软件也不包括 DC-FRET 等后续发展的分析方法。

1.3.3 基于深度学习的 FRET 数据处理方法

近年来，随着机器学习不断发展，越来越多的研究者着手将此类方法应用于 FRET 数据分析，以求提升荧光强度提取的准确性与效率。

Ge 等人研发出一种基于 U-Net^[4] 模型的深度学习方法进行高效 FRET 分析，该模型基于一个包含 230 个手动标注的感兴趣区域 (ROIs) 的数据集展开训练，随后通过旋转操作将数据集扩充，最终得到 2760 个样本^[5]。U-Net 模型是图像分割领域广泛运用的卷积神经网络架构，它能够有效捕捉荧光图像的特征，精准地提取 ROIs 的荧光强度信息。通过手动标注一定数量的 ROIs，再利用旋转等数据增强手段扩充数据集，如此便能让模型得到更充分的训练，进而提升其在荧光强度提取方面的泛化能力与准确性。

Feldmann 等人则采用了 ilastik^[6] 这一用于生物医学图像分析的开源交互式工具，以极少的标注工作量来描绘细胞及背景区域。ilastik 提供了直观的图形用户界面，使用户能以相对简易的操作交互式地标注并分割图像。它将机器学习算法与用户交互相结合，不但能减轻手动标注的工作量，还能让分割结果更为精准可靠，从而更出色地提取目标区域的荧光强度。

然而，这些由机器学习驱动算法存在一定的局限性。其一，它们的效能取决于数据集的规模与质量，而这在很大程度上依赖于手动标注。规模更大、质量更高的数据集可为模型训练提供更充裕的信息，使模型更好地捕捉数据的特征与模式，进而提升荧光强度提取的性能。但手动标注既耗时又费力，在实际应用中往往难以获取大量精准标注的数据集。数据不足可能引发模型过拟合等问题，致使模型的泛化能力与准确性下滑。其二，就可解释性而言，这些算法通常较为晦

涩。源于机器学习的神经网络模型常常捕捉到特定数据集的细微差别，却未能提供清晰的数学框架来阐释这些模式。尽管这些模型在荧光强度提取方面能够达成良好效果，但用户很难理解模型是怎样依据输入数据做出决策与预测的。神经网络的内部结构与运行机制繁杂，不存在直观、明晰的数学公式或规则用以解释输入与输出之间的关联，这给模型的深入理解与进一步优化带来了一定困难。

1.4 本文的工作内容和意义

专门设计的 FRET 双杂交数据处理软件可以显著提高 FRET 双杂交数据处理工作的效率和准确性，推广 FRET 双杂交分析技术的应用。FRET 双杂交数据处理软件专门针对 FRET 双杂交数据处理而开发，内嵌了完整的 FRET 双杂交计算算法，因此用户不需要使用其他数据处理软件提供的数据处理功能，可以一键式导入数据，计算出 FRET 双杂交分析的结果，提高数据处理的工作效率。使用全自动数据处理功能则可以一键式完成由拍摄得到数据到最后得到计算结果的全部数据处理工作，包括图像预处理、自动背景选取、自动 ROI 选取、FRET 定量计算、FRET 双杂交分析计算等。手动数据处理引入的人为因素使得数据处理的计算结果相对缺乏稳定，这一缺陷在采集荧光信号时尤为明显。使用图像处理等程序算法构建荧光信号的评价函数，能够定量评价荧光信号的质量，并以此为根据发展自动信号采集算法。使用我们的软件能够为不同批次实验提供定量规范的信号采集和数据计算功能，从而大大提高了 FRET 双杂交实验数据分析的定量程度和可复现性。我们还会丰富数据计算结果的输出形式，加强数据的可视化输出和数据的兼容性，方便对软件计算的结果数据进行对比分析。

面对海量样本数据的活细胞高通量药物筛选应用，自动化的 FRET 双杂交分析是必然的选择。高通量药物筛选数据量庞大，且对数据处理速度要求更高，人工手动数据处理远远无法满足要求。通过计算机强大的数据处理能力，高度自动化的 FRET 双杂交数据处理软件能够快速计算得到科学家们关心的参数指标，大大提高实验速度，从而提高整套系统的效率。自动化程度高、计算速度快的 FRET 双杂交数据处理能力能够计算产生大量的生物学数据指标，为药物筛选系统后续建立完备的数据库提供充足的数据量，进一步丰富了药物筛选数据分析的可能性。

针对实验室搭建的用于活细胞高通量药物筛选的 FRET 双杂交分析系统设计专用的离线数据处理软件，能够提供。在该系统中，我们拥有灵敏快速的检测仪器和强大的计算机控制系统等硬件设备，进而将 FRET 双杂交实验流程高度实现自动化，实验仪器的自动化要求相配套的数据处理自动化程度的提高。为了提高 FRET 双杂交分析技术的数据处理效率和准确性，本文提出了一种基于计算机视觉的自动化数据处理方法，该方法构造了荧光信号的评价函数，可以快速地从原始图像中提取出 FRET 信号，并对其进行校正和归一化，从而得到可靠的 FRET 效率值。通过供体中心的 FRET 效率和受体中心的 FRET 效率，我们组还对数据特点进行分析，发展出了基于数据预处理的 FRET 双杂交分析和基于 FRET 效率 E_D 受供体浓度比 R_C 线性分离的 FRET 双杂交分析等多种方法，从而比较验证得到稳定可靠的化学计量比、相对亲和力常数等重要参数，提高了 FRET 双杂交分析的稳定性。

1.5 本文的章节安排

第一章，绪论。本章简要概述了定量 FRET 技术和 FRET 双杂交分析技术的理论基础和发展，然后介绍了 FRET 数据分析处理方法的研究现状。同时介绍了本文的工作内容和结构安排。

第二章，FRET 双杂交分析数据处理软件的设计和开发。本章基于 FRET 双杂交分析数据处理流程和 FRET 多模态显微成像系统的特点进行需求分析，然后设计了 Fretha 软件的总体架构和模块划分，最后详细介绍了各个模块的开发实现。

第三章，自动 FRET 双杂交分析数据处理算法。本章分析了 FRET 双杂交分析时的痛点，提出了基于明度和均匀度的自动 ROI 选择算法，并结合 DC-FRET 方法实现了全自动的 FRET 双杂交分析数据处理算法，并在模型质粒上进行了验证。

第四章，软件的应用和测试。本章应用 Fretha 进行高通量药物筛选应用实验，成功快速地测量了在 MCF-7 细胞中加药后 Bcl-XL 和 Bak 结合比例的下降，并与其他几种数据处理方法进行比较。同时，对软件的性能指标进行了测试和分析。

第五章，总结与展望。本章总结了本文的工作内容和在相关领域的贡献，对后续研究内容和方法进行展望。

2 FRET 双杂交分析数据处理软件（Fretha）的设计和开发

2.1 引言

在现代生物医学研究中，FRET 双杂交分析作为一种关键技术，用于检测生物分子间的相互作用，为揭示生物过程的分子机制提供了重要手段。然而，传统的 FRET 双杂交分析数据处理过程繁琐且效率低下，严重依赖人工操作，容易引入误差。开发专门针对 FRET 双杂交分析数据处理的配套软件，能够实现数据的自动化采集、标准化处理以及高效的分析统计，对于提升研究效率和准确性具有重要意义。本章节首先根据我们组自研的用于高通量药物筛选的 FRET 显微成像系统（FRETscopeII）的硬件参数和 FRET 数据处理及 FRET 双杂交分析数据处理功能进行深入的需求分析。基于此需求分析，对软件按照功能进行科学合理的模块划分和总体框架设计。接着，详细介绍软件的开发技术选型以及每个模块的具体实现方式，旨在全面展示 Fretha 软件从设计构思到开发实现的全过程。

2.2 需求分析和总体设计

2.2.1 需求分析

FRETscopeII 是本课题组自研的适用于 3^3 -FRET、E-FRET、Pb-FRET 等多种定量 FRET 分析方法的 FRET 多模态显微成像系统。该系统具备高分辨率成像、多通道同步采集等先进特性，能够获取高质量的 FRET 三通道图像数据。

使用 FRETscopeII 对制备好的样本进行图像采集得到若干视野的 FRET 三通道图像后，通过 FRET 双杂交分析方法定量计算细胞中的结合比等信息需要如下分析步骤：首先需要对 FRET 数据进行数据完整性校验，确保每个视野下存在完整的三通道图像和文件完整可读，然后对每个视野进行荧光信号提取，通过在图像上绘制 ROI 并计算其灰度均值作为 FRET 分析计算重要的荧光信号；根据 E-FRET 和 3^3 -FRET 方法将上述荧光信号代入对应的计算公式求取 E_A 、 E_D 、 R_C 等

FRET 数据；对这些数据进一步依据物理含义或者数据分析进行异常值去除等数据预处理；最后是通过优化算法拟合 Langmiur 模型或者线性模型计算相关的参数。具体流程如图2.1所示。



图 2.1 FRET 双杂交分析数据处理流程

2.2.2 模块划分

结合 FRET 双杂交分析的数据处理流程与功能逻辑，Fretha 软件主要划分为以下核心模块：成像参数设置模块：负责设置 FRET 成像过程中的关键参数，确保成像数据的准确性和一致性。数据完备性检验模块：对输入数据进行严格的完整性和合法性检验，避免异常数据导致的分析错误。FRET 图像处理模块：支持手动图像处理和 ROI 选取，满足用户对数据的精细化处理需求。自动 ROI 圈点模块：结合 LURS 算法实现自动 ROI 选取，显著提高数据处理效率。结果可视化模块：将分析结果以直观的图表形式展示，并提供数据保存功能，方便用户进一步分析和应用。

2.2.3 软件总体框架

Fretha 架构采用分层设计，由顶层向下依次分别为：表现层、业务层、数据访问层和数据层，如图2.2所示。

表现层处于整个架构的最外层，直接面向用户，是用户与系统进行交互的主要界面。它负责接收用户输入的各种指令和数据，并以直观、友好的方式展示系统的处理结果。Fretha 的表现层主要包括参数设置界面、图像处理界面、ROI 圈点界面、数据记录界面、可视化界面等，用户能够轻松地与系统进行交互，完成 FRET 双杂交分析的数据处理操作。

业务层是整个架构的核心逻辑处理部分，承担着对系统业务规则和流程的实现。它接收来自表现层的请求，根据预设的业务逻辑对数据进行处理和转换。Fretha 的业务层封装了包括参数读 / 写业务、数据完备性检验业务、自动 ROI 选取业务、数据导入 / 导出业务、FRET 双杂交分析业务、数据绘图 / 可视化业务



图 2.2 Fretha 软件总体架构

等。FRET 定量计算器和 FRET 图像分析业务是业务层中进行 FRET 定量计算的业务，分别可以对 FRET 原始灰度数据和 FRET 图像数据进行定量 FRET 计算分析。Fretha 中这两个业务为其他业务提供了关于 FRET 定量计算的数据处理能力，使得 FRET 定量分析及 FRET 双杂交分析数据处理过程中的科学计算逻辑进一步独立，减少了软件架构的冗余度，提高了编码以及程序运行的效率。

数据访问层实现对数据的访问和操作，它将业务层与数据层进行隔离。数据访问层提供了统一的数据访问接口，业务层通过调用这些接口来获取和存储数据，而无需关心数据的具体存储方式和位置。通过设置数据访问层，能够使得在复杂的业务处理时避免对数据的直接操作和影响，从而提高了数据存储的安全性。Fretha 的数据访问包括 FRET 图像数据访问、FRET 数值数据访问和 FRET 双杂交数据访问的接口。

最后是数据层，作为架构的最底层，数据层负责存储系统的所有数据。Fretha 数据层包括系统静态数据、FretImage 类、FretDataPiece 类和 FretRecord 类。系统

静态数据是在软件运行时的环境参数，只需要在指定步骤运行前提前设置好即可，如成像参数、文件目录等。FretImage 类、FretDataPiece 类和 FretRecord 类用来表示数据处理时的各种动态数据。在 FRET 双杂交分析数据处理中，一组 FRET 三通道图像中可以提取并计算出若干条 FRET 数据，由若干条 FRET 数据作为一个批次只能解析出一条 FRET 双杂交分析结果。因此，在设计上三种数据实体类型存在关联关系，FretImage 和 FretDataPiece 之间存在一对多的关系，FretDataPiece 和 FretRecord 之间存在一对多关系，如图2.3所示。

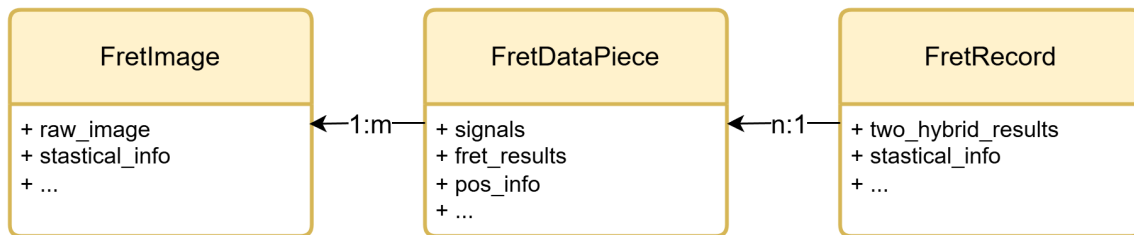


图 2.3 Fretha 数据层实体关联关系图

2.2.4 开发技术选型

Qt 5.15.2 是 Qt 官方发布的长期支持 (LTS) 版本，具有卓越的跨平台性能。该版本能够在 Windows、Linux、macOS 以及嵌入式系统等多种平台上稳定运行。开发者只需对一套代码进行简单配置，即可实现不同平台的适配，这种特性大幅降低了软件开发与维护的成本。FRETscopeII 的控制软件系统选择了 Qt 5.15.2 版本作为技术选型。为了确保 Fretha 和 FRETscopeII 在开发过程中的一致性和可融合性，Fretha 同样选用该版本作为软件开发的基础技术。这一选择不仅有利于代码的共享、复用和维护，还能显著降低二次开发以及未来合并开发过程中的成本和风险。

在计算机视觉处理方面，Fretha 选用了 OpenCV (Open Source Computer Vision Library)。OpenCV 是一个开源的计算机视觉与机器学习软件库，具备强大的跨平台能力，可在 Windows、Linux、macOS、Android、iOS 等多种操作系统上运行。其功能丰富多样，广泛涵盖了图像和视频处理的各个方面，例如图像的读取、滤波、边缘检测等。在性能上，OpenCV 经过 SIMD 指令集优化以及多线程并行计算等技术的优化，处理效率得到了显著提升，能够充分满足实时性的需求。此外，

OpenCV 拥有一个活跃的社区，汇聚了大量的开发者。社区中提供了丰富的算法、示例代码和实际案例，为开发者提供了有力的技术支持和学习借鉴的资源。基于以上诸多优势，OpenCV 完全契合本项目中计算机视觉任务的需求，因此被选定为本项目的重要技术工具。

在最优化计算方面，Fretha 选择了 C++ 库 Dlib。Dlib 库集成了基于梯度的优化算法，并采用了自适应学习率机制，能够实现快速且稳定的收敛，有效降低了陷入局部最优解的概率。该库还采用了牛顿法与拟牛顿法，在保证计算精度的同时兼顾了计算效率，能够显著缩短大规模训练所需的时间。此外，Dlib 在处理约束优化问题方面表现出色，例如可以运用内点法来解决资源分配中的约束难题，为研究工作提供了可靠的支持。

2.3 Fretha 的开发实现

2.3.1 成像参数设置模块

FRET 定量分析中，测量并设置成像过程时的成像参数至关重要。Fretha 作为离线数据处理软件，在数据处理前需要设置好 FRET 定量计算过程中必须的参数。

FRET 成像参数在 Fretha 中以静态参数保存在软件内存中，是数据处理时的环境参数。其中， a 、 b 、 c 、 d 、 G 、 k 和 Y 是 FRET 成像系统的光学参数，在前文中已介绍； $ExpTimeAA$ 、 $ExpTimeDD$ 和 $ExpTimeDA$ 是成像时三个探测通道的曝光时间，在 FRET 定量计算时需要根据曝光时间参数在各个通道归一化，然后才能进行计算。Fretha 中包括的所有成像参数如表2.1所示。

成像参数设置模块在业务设计上遵循以下原则：（1）设置时所有参数应一同更新；（2）每个参数需要在其有意义的范围内，避免无意义的值。具体来说，在点击“更新参数”按钮时，若无法从界面中的每个参数输入框都解析到合法的数字，那么本次更新参数就会失败。这是因为 FRET 成像参数是一批参数，参数间存在依赖关系，如测量参数 G 、 k 、 Y 就依赖参数 a 、 b 、 c 、 d ，这是因为测定参数 G 、 k 、 Y 时需要计算敏化发射荧光 F_C ，根据公式1.5所示， F_C 的确定与 a 、 b 、 c 、 d 密切相关。强制所有参数一同更新可以避免用户单独设置某一参数而导致参数之间不匹配等问题。

表 2.1 FRET 成像参数

参数名	参数意义	典型范围	单位
a	供体激发时在 DA 通道的串扰系数	(0, 1)	无
b	所有的一切都在这里面。	(0, 1)	无
c	模板类文件。	(0, 1)	无
d	受体激发时在 DA 通道的串扰系数	(0, 1)	无
G	供体猝灭和受体荧光增强的比值	(0, $+\infty$)	无
k	受体浓度和供体浓度相同时的荧光比值	(0, $+\infty$)	无
Y	供受体在激发光条件下的消光系数之比，也记作 $\varepsilon_{YFP}(\lambda)/\varepsilon_{CFP}(\lambda)$	(0, $+\infty$)	无
$ExpTimeAA$	AA 通道下的成像曝光时间	(0, $+\infty$)	毫秒 (ms)
$ExpTimeDD$	DD 通道下的成像曝光时间	(0, $+\infty$)	毫秒 (ms)
$ExpTimeDA$	DA 通道下的成像曝光时间	(0, $+\infty$)	毫秒 (ms)

FRET 成像参数一般比较稳定，一般 2 到 3 个月才需要重新测量，因此需要持久化到本地，以供多次处理数据时使用。Fretha 的本地参数文件保存为可执行程序同级目录下的“config.ini”中。在软件初始化阶段，会自动检测并应用本地配置文件中的参数。因此，用户可通过保存多套配置文件，在使用时替换目标配置文件，快速进行参数配置的切换。

参数设置的业务流程如图2.4所示。

2.3.2 数据完备性检验模块

FRET 双杂交分析需要处理一批 FRET 图像文件，因此需要对输入数据的数据完备性进行检验识别。该模块的作用有以下两个方面：一方面通过模式识别 FRET 合法数据，避免了异常输入导致的运行错误；另一方面，在这一模块会将 FRET 批数据的视野子文件夹进行解析和类型识别，为后续数据处理提供对子文件夹的不同操作。

Fretha 的数据识别检验模块匹配识别 FRETscopeII 的数据格式，从而保证数据处理能够正常开始。FRETscopeII 的数据结构由上层到下层依次为：（1）批数据根目录；（2）视野子目录；（3）图片数据和参数文件。其结构如图2.5所示。

数据完备性检验业务的流程如图2.6所示。其中，检查子文件夹类型是通过

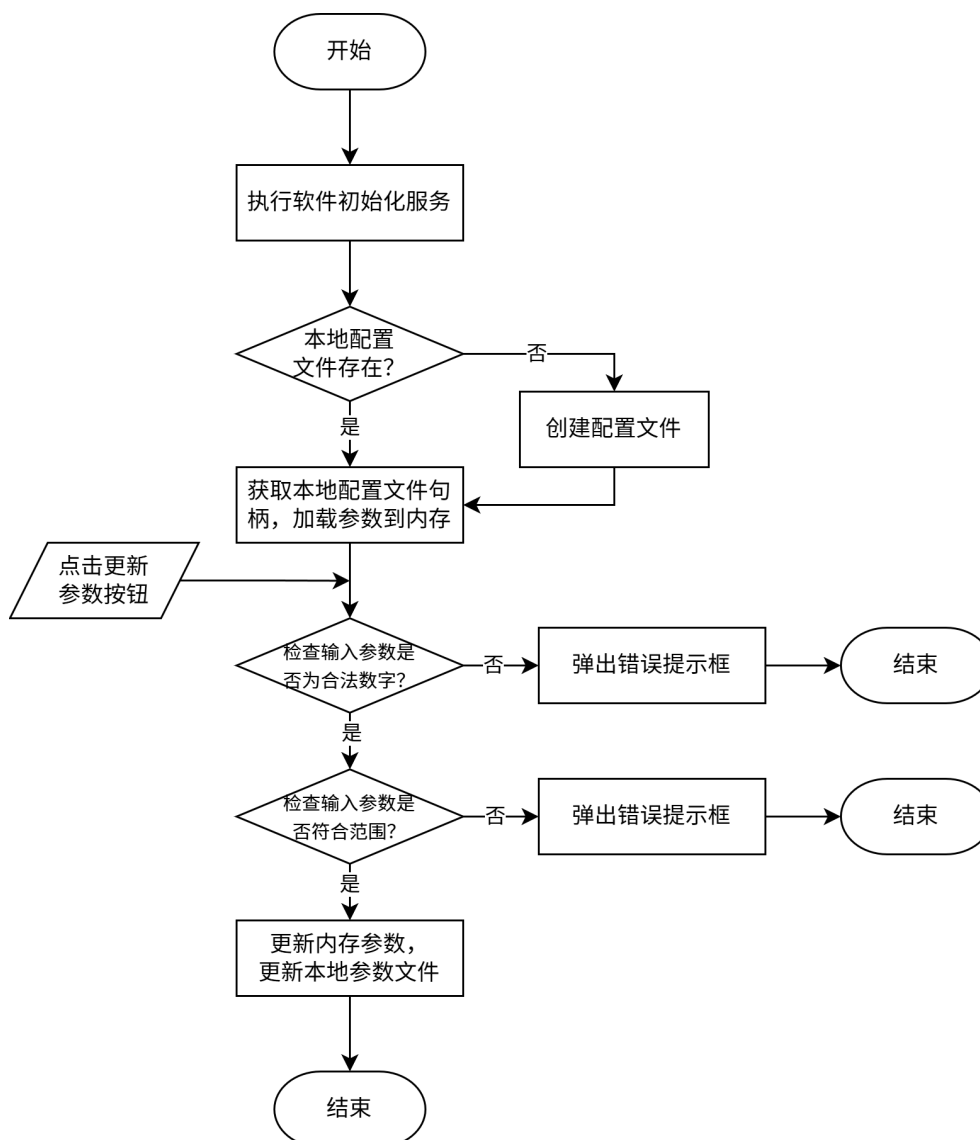


图 2.4 参数设置模块业务流程图

图2.5进行匹配的，当且仅当子文件夹中同时存在“DA.tif”、“DD.tif”和“AA.tif”图片文件时，当前子文件夹会被识别为 FRET 视野，并在视野表格模型中记录。其他情况的子文件夹会被记作“Unknown”文件夹，在后续 FRET 图像处理或者自动处理中被跳过。

2.3.3 FRET 图像处理模块

FRET 图像处理模块是对 FRET 三通道图像进行 ROI 提取等图像处理分析的模块，包括手动 ROI 圈点和自动 ROI 圈点等。其界面如图2.7所示：

在 FRET 图像处理工作里，手动选取感兴趣区域 ROI 是一个基础且关键的操

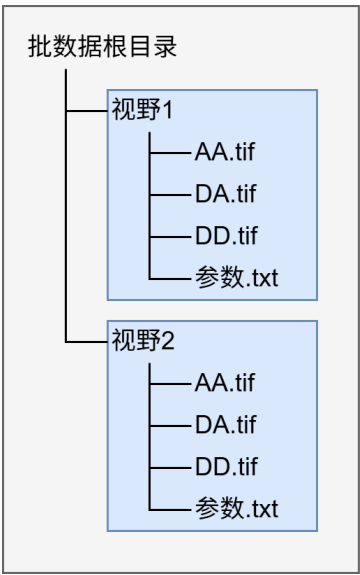


图 2.5 FRETscopeII 数据文件层级结构树形图

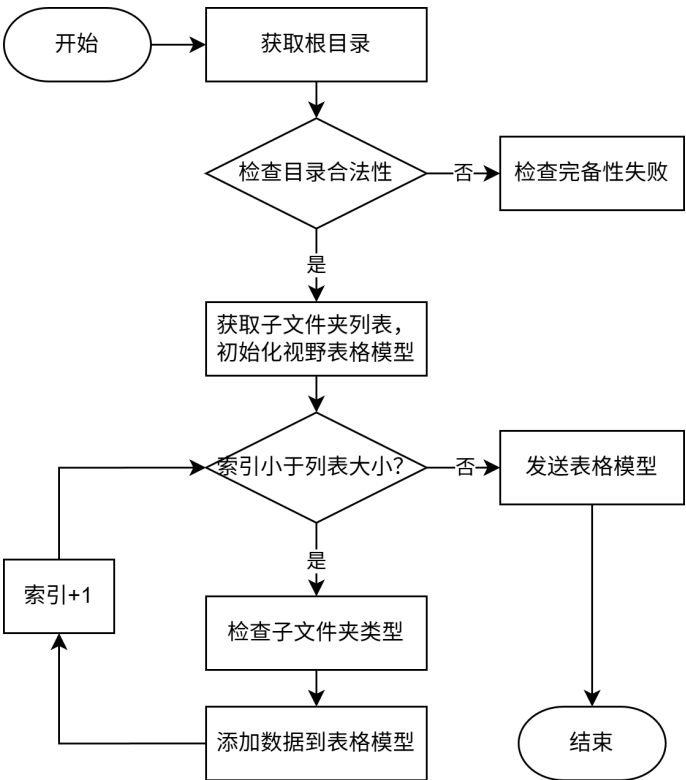


图 2.6 Fretha 数据完备性检验业务主流程图

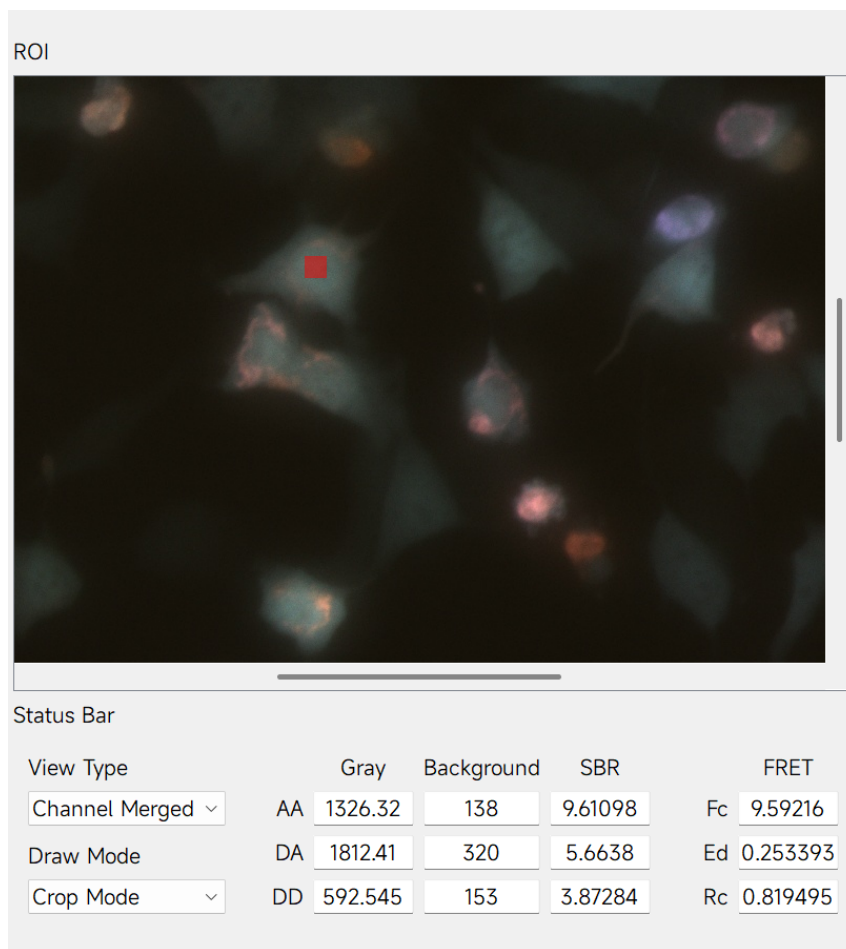


图 2.7 Fretha 图像处理模块界面示意图

作，它能帮助分析人员聚焦于特定区域，进行更精确的数据处理与分析。为达成图像处理过程中 ROI 的手动选取功能，Fretha 借助 Qt 框架提供的 `QGraphicsView` 类，开发了自定义的 `FretGraphicsView` 类。`QGraphicsView` 是 Qt 用于可视化和交互处理二维图形场景的重要类，具备丰富的功能和良好的可扩展性，为 `FretGraphicsView` 的实现提供了有力支撑。

`FretGraphicsView` 类主要用于支持复杂的 ROI 编辑操作。该类中新增了一个 ROI 成员，这一设计使 `FretGraphicsView` 具备了更完善的 ROI 管理能力。它能够在视图中直观呈现 ROI 的大小和位置，方便用户确认所选区域。同时，通过状态栏更新 ROI 内的数据，并为用户提供具有参考意义的 FRET 信息，有助于用户更深入地理解图像数据。

在交互机制上 `FretGraphicsView` 通过自定义鼠标释放事件的信号，实现了与数据访问层的数据交互。当用户完成 ROI 选取并释放鼠标时 `FretGraphicsView` 会将

ROI 的坐标和大小信息传递给数据访问层。数据访问层依据当前视野索引，从数据模型中读取相应的灰度数据。随后，这些灰度数据会被输送至 FretCalculator 进行定量 FRET 计算。计算完成后，原始灰度数据和 FRET 数据会更新到界面上，方便用户获取处理结果，为后续的数据分析工作提供依据。这种基于 QGraphicsView 扩展的设计，有效提升了 ROI 选取的灵活性和准确性，实现了图像处理与数据计算的有机结合，为图像处理应用开发提供了一种可行的技术方案。

Fretha 的状态栏中提供了视图类型切换选项，在数据处理中可以辅助圈点，主要包括归一化增强图，和 FRET 伪彩图两类，支持的视图类型如表2.2所示。归

表 2.2 FRET 图像视图类型表

信息	说明
Channel Merged	三通道归一合成图
DD Normalized	DD 通道的归一化增强图
DA Normalized	DA 通道的归一化增强图
AA Normalized	AA 通道的归一化增强图
R_C Pseudo	R_C 逐像素数值伪彩图
E_D Pseudo	E_D 逐像素数值伪彩图

一化增强图包括 DD、DA、AA 通道分别归一化的增强视图以及三通道归一化合成图。FRET 三通道原始图片由于测量系统的灰度量程为 0 至 65535，在计算机显示时，灰度为 65536 的像素显示为纯白色，灰度为 0 的像素显示为纯黑。然而，FRET 信号往往达不到量程的最大值，因此图片会显得灰暗，无法识别到图像中的细胞等问题。图像增强的计算公式如公式2.1所示。

$$y = \frac{x - \min_{old}}{\max_{old} - \min_{old}} \times (\max_{new} - \min_{new}) + \min_{new}. \quad (2.1)$$

其中， x 是原图像中的像素值， y 是经过归一化处理后的像素值， \min_{old} 和 \max_{old} 分别是原图中像素灰度值的最小值和最大值， \min_{new} 和 \max_{new} 分别是目标像素值中像素灰度值的最小值和最大值。在图像显示时，我们将其设置为易于观察的

0255, 因此公式2.1实际计算为:

$$y = \frac{x - \min_{old}}{\max_{old} - \min_{old}} \times 255 \quad (2.2)$$

除了视图辅助, Fretha 状态栏能够清晰展示当前视野及 ROI 的状态信息, 如当前视野的三通道背景灰度值、ROI 信号的三通道信号背景比等, 还可以显示使用当前 ROI 提供的扣除背景灰度值后的 I_{DD} 、 I_{DA} 和 I_{DD} 。Fretha 状态栏的所有显示内容如表2.3所示。当对 ROI 调整完毕后, 点击“添加数据”可以将 ROI 内的数据记录到软件右侧的表格中。点击左侧“自动圈点”按钮, 可以使用 LURS 算法进

表 2.3 FRET 圈点状态栏显示内容

信息	说明
AA 信号	AA 通道的 ROI 内灰度均值
DA 信号	DA 通道的 ROI 内灰度均值
DD 信号	DD 通道的 ROI 内灰度均值
AA 背景	AA 通道的视野背景灰度值
DA 背景	DA 通道的视野背景灰度值
DD 背景	DD 通道的视野背景灰度值

行自动 ROI 圈点, 并记录到右侧的数据记录表中。自动圈点功能通过 FRET 双杂交求解器中封装的多线程服务, 在不影响 Fretha 前台界面显示和操作的前提下, 实现了一键式线程分离的自动数据处理。自动 ROI 圈点模块的业务流程及架构如图2.8所示。架构层内的流程流转以实线表示, 不同架构层之间的流程流转以虚线表示。在实现自动圈点功能的流程转移中, 发生在架构不同层之间的转移占更多数, 而层内的流程转移相对较少, 分层架构设计使得每个层内实现较好的封装, 因此能够在实现复杂业务功能时, 只需要专注于层和层之间的数据和流程切换即可。图中关于 LURS 算法和自动 ROI 圈点的算法原理将会在第三章具体说明。

2.3.4 数据控制模块

数据控制模块支持对数据的实时操作, 可以对于图像处理模块获得的数据, 进行异常数据筛选、数据导入、数据导出、数据追踪、计算入口等数据控制功能。

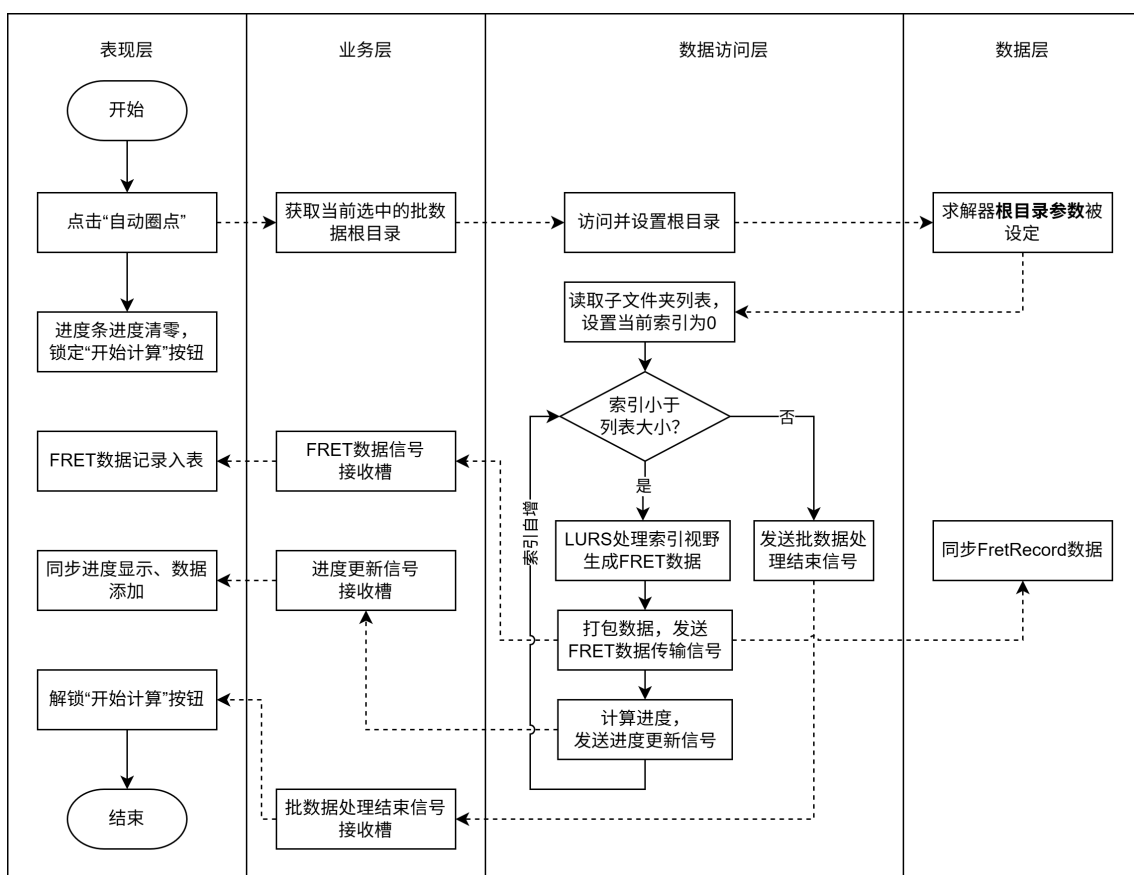


图 2.8 Fretha 自动圈点模块流程图

在数据模型上，我们将一条数据类型 FretDataPiece 定义为：

异常数据筛选分为两个阶段：

- (1) 物理意义筛选。根据数据项的物理属性设定合理定义域。例如， I_{DD} 应满足 $I_{DD} > 0$ 。对于数值型数据 x_i ，删除条件可表示为：

$$\text{删除条件} = \begin{cases} 1, & \text{若 } x_i \notin [L_i, U_i] \\ 0, & \text{其他情况} \end{cases} \quad (2.3)$$

其中 L_i 和 U_i 为第 i 个变量的定义域上下限，删除条件为 1 表示需要删去该数据。

- (2) 统计离群检测。对通过物理筛选后的数据计算均值 μ 和标准差 σ ，构建离群

表 2.4 FretDataPiece 数据类型

信息	说明
I_{DD}	ROI 在 DD 通道扣除背景后的信号强度
I_{DA}	ROI 在 DA 通道扣除背景后的信号强度
I_{AA}	ROI 在 DD 通道扣除背景后的信号强度
E_D	根据 I_{DD} 、 I_{DA} 、 I_{AA} 计算的 E_D
R_C	根据 I_{DD} 、 I_{DA} 、 I_{AA} 计算的 R_C
E_A	根据 I_{DD} 、 I_{DA} 、 I_{AA} 计算的 E_A
$1/R_C$	根据 I_{DD} 、 I_{DA} 、 I_{AA} 计算的 $1/R_C$
x	ROI 的横坐标
y	ROI 的纵坐标
w	ROI 的宽度
h	ROI 的高度
$view$	ROI 隶属的视野

点判定准则：

$$\text{离群阈值} = \mu \pm 3\sigma \quad (2.4)$$

超出该范围的数据点视为离群点。具体步骤为：

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i, \quad \sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2} \quad (2.5)$$

首先根据各项数据的物理意义进行筛选，删除其中超出定义域的数据项。然后，删除超出数据中超过 3 倍标准差的离群数据。

数据导入和导出用于保存数据处理时的 ROI 信息。在一次 FRET 双杂交分析数据处理后，需要记录本次数据处理的处理数据，方便后续查看。导出的数据信息包括：

数据追踪是指在数据表格中点击某一条数据，软件会根据这条数据的位置和形状信息，在图像显示上显示该条数据对应的 ROI 位置，以方便识别数据中潜在的错误。实现这一功能，主要基于 Qt 提供的信号与槽机制。将数据中的数据模型被单机的事件与 FretGraphicsView 控件绑定后，FretGraphicView 模型可以接收到

来自所选数据的信息，并将当前活动的 ROI 按照所选数据的位置和形状重新绘制。

2.3.5 结果可视化模块

FRET 双杂交分析的数据结果包括 FRET 数据进行双杂交计算的参数结果和拟合曲线图。FRET 双杂交分析数据拟合结果的参数是用户得到的核心结果，其中 n_D/n_A 、 $K_{d,EFF}$ 等参数是生物学家最关心的生物大分子之间相互作用的参数。FRET 双杂交分析的拟合曲线图是将双杂交拟合曲线和实验采集到的散点同时绘制在一张图上，用于评估双杂交计算结果的拟合效果，以进一步判断 FRET 双杂交分析结果的可信度。

点击“保存结果”按钮后，上述数据结果和图片结果都将被保存，还将保存可在其他绘图软件上制图的原始数据，其中相比较圈点的数据来说，主要新增了 A_{free} 和 D_{free} 数据点，这部分信息的计算依赖于 FRET 双杂交分析的参数结果，在结果展示时才能确定。

结果可视化模块的界面主要包括三个部分：FRET 双杂交分析算法选择；FRET 双杂交分析图表结果；结果保存按钮。

2.3.6 FRET 双杂交计算模块

FRET 双杂交计算模块是对采集到 FRET 批数据 E_D 、 E_A 、 R_C 、 A_{est} 和 D_{est} 等数据进行最优化计算，以求取使得预测结果和测量结果之间误差最小时的关键参数。这部分模块主要封装了进行 FRET 双杂交计算时需要的优化算法库等。

(1) **DC-FRET 线性拟合算法的封装**。根据公式1.20和1.22，DC-FRET 拟合斜率时截距项为 0，假设拟合的线性方程形式为

$$E_D = E_{A,max} \cdot R_C, \quad (2.6)$$

其中， E_D 是自变量， R_C 是自变量， $E_{A,max}$ 是斜率。线性拟合的目标是找到合适的参数 $E_{A,max}$ ，使得方程预测的 E_D 值与实际观测到的 E_D 值之间的误差尽可能小。

通常使用最小二乘法，其原理是最小化观测值与预测值之间的误差平方和，即

$$S = \sum_{i=1}^n (E_{D_i} - (E_{A,max} R_{C_i}))^2 \quad (2.7)$$

其中， n 是数据点的数量， R_{C_i} 和 E_{D_i} 分别是第 i 个数据点的自变量和因变量的值。为了找到 S 最小的 $E_{A,max}$ 值，对 S 关于 $E_{A,max}$ 求偏导，并另其等于 0。首先，展开误差 S ：

$$S = \sum_{i=1}^n (E_{D_i}^2 - 2E_{A,max} R_{C_i} E_{D_i} + E_{A,max}^2 R_{C_i}^2), \quad (2.8)$$

$$\frac{\partial S}{\partial E_{A,max}} = \sum_{i=1}^n (-2R_{C_i} E_{D_i} + 2E_{A,max} R_{C_i}^2), \quad (2.9)$$

令 $\frac{\partial S}{\partial E_{A,max}} = 0$ ：

$$\sum_{i=1}^n (-2R_{C_i} E_{D_i} + 2E_{A,max} R_{C_i}^2) = 0, \quad (2.10)$$

$$-2 \sum_{i=1}^n R_{C_i} E_{D_i} + 2E_{A,max} \sum_{i=1}^n R_{C_i}^2 = 0, \quad (2.11)$$

最后求解 $E_{A,max}$ ：

$$E_{A,max} = \frac{\sum_{i=1}^n R_{C_i} E_{D_i}}{\sum_{i=1}^n R_{C_i}^2}. \quad (2.12)$$

公式2.12可用于直接求解斜率 $E_{A,max}$ 和 $E_{D,max}$ ，从而避免了基于迭代的线性拟合求解算法的消耗。

(2) Langmiur 模型拟合算法的封装。Langmiur 模型具有模型的非线性和参数的复杂性，无法通过简单的解析式解析其求解公式。因此，我们引入基于 `dlib` 计算库的双杂交系统参数拟合。

首先，基于 `dlib` 定义了用于拟合的数据类型 `ColumnVector` 类型。此类型用于存储双精度浮点数的列向量，在后续的计算和优化过程中承载参数向量和中间计算结果。定义 `ExperimentalData` 结构体用于整合实验数据。该结构体包含四个 `std::vector<double>` 类型的成员变量：`aest` 存储与变量 `a` 相关的估计值序列，`dest`

存储与变量 d 相关的估计值序列, ea_corr 存储变量 a 的相关误差序列, ed_corr 存储变量 d 的相关误差序列。这种数据结构的设计便于在不同函数间传递和处理实验数据, 提高代码的模块化和可读性。

Listing 2.1 基于 dlib 的 FRET 双杂交求解器代码片段

```

1  #include <iostream>
2  #include <vector>
3  #include <cmath>
4  #include <dlib/optimization.h>
5
6  // 定义列向量类型
7  typedef dlib::matrix<double, 0, 1> ColumnVector;
8
9  // 定义数据结构体, 用于存储实验数据
10 struct ExperimentalData {
11     std::vector<double> aest;
12     std::vector<double> dest;
13     std::vector<double> ea_corr;
14     std::vector<double> ed_corr;
15 };

```

CalculateError 函数用于计算单个数据点的误差。其数学定义为 $Error(x, y) = (x - y)^2$, 通过将输入值 x 和 y 进行平方差求和运算, 返回误差值。此函数基于最小二乘法原理, 将误差的平方作为衡量预测值与真实值偏离程度的指标, 在后续的损失函数计算中被广泛应用。CalculateLoss 函数负责计算模型在整个数据集上的整体损失。对于给定的 ExperimentalData 实例 $data$ 和参数向量 $parameters$ 。

ObjectiveFunction 函数是优化过程的目标函数。它接受参数向量 $parameters$ 和指向 ExperimentalData 结构体的指针 $data_ptr$ 作为输入。函数内部将指针转换为 ExperimentalData 类型, 然后调用 CalculateLoss 函数, 以计算并返回基于当前参数和实验数据的损失值。该目标函数为优化算法提供了评估参数优劣的标准, 优化算法通过不断调整参数来最小化该目标函数的值。

Listing 2.2 基于 dlib 的 FRET 双杂交求解器代码片段

```

1  // 计算误差
2  double CalculateError(double x, double y) {
3      return (x - y) * (x - y);
4  }
5
6  // 计算整体损失
7  double CalculateLoss(const ExperimentalData& data, const
8      ColumnVector& parameters) {
9      double total_error = 0.0;

```

```

9   for (size_t i = 0; i < data.aest.size(); ++i) {
10      double d_free = ((data.dest[i] - parameters(0) - data.
11          aest[i] * parameters(1)) +
12          std::sqrt(std::pow(data.dest[i] - parameters
13              (0) - data.aest[i] * parameters(1), 2) +
14              4 * parameters(0) * data.dest[i])) /
15              2;
16      double a_free = data.aest[i] - (data.dest[i] - d_free) /
17          parameters(1);
18      double ea_pred = parameters(2) * d_free / (d_free +
19          parameters(0));
20      double ed_pred = parameters(3) * a_free / (a_free +
21          parameters(0) / parameters(1));
22      total_error += CalculateError(data.ea_corr[i], ea_pred) +
23          CalculateError(data.ed_corr[i], ed_pred);
24  }
25  return total_error;
26 }

// 目标函数，用于优化过程
27 double ObjectiveFunction(const ColumnVector& parameters, void*
28     data_ptr) {
29     auto data = static_cast<ExperimentalData*>(data_ptr);
30     return CalculateLoss(*data, parameters);
31 }

```

TwoHybridSolver 函数实现了整个参数拟合的优化过程，并对外提供调用接口。它接受四组实验数据向量 `aest_data`、`dest_data`、`ea_corr_data` 和 `ed_corr_data` 作为输入。首先初始化优化的起始点 `starting_point` 为一个包含四个元素的列向量 `[1, 1, 0.5, 0.5]`，分别对应为 $K_{d, EFF}$ 、 n_D/n_A 、 E_{Amax} 、 $E_{D,max}$ 的初始值，这个起始点的初始值基于实际物理意义的先验知识，虽然它可能不是最优解，但为优化算法提供了一个初始起点。接着创建 `ExperimentalData` 结构体实例，并使用输入的四组数据向量进行初始化。这一步确保了实验数据能够被正确地整合和传递给后续的计算函数。然后在 `try - catch` 块中进行优化操作。在 `try` 块内，定义了目标函数包装器 `objective_wrapper`，它通过捕获外部的 `data` 变量，调用 `ObjectiveFunction` 函数，为优化算法提供了一个可调用的目标函数对象。随后使用 `dlib` 库的 `find_min_using_approximate_derivatives` 函数进行优化。该函数采用 `bfgs_search_strategy` 搜索策略，该策略在优化过程中通过近似计算目标函数的梯度来确定搜索方向，具有收敛速度较快的优点；同时使用 `objective_delta_stop_strategy` 停止策略，当目标函数值的变化小于指定阈值（这里为 10^{-7} ）时，认为优化过程已收敛，停止迭代。优化过程的目标函数为 `objective_wrapper`，起始点为之前初始

化的 `starting_point`。如果在优化过程中捕获到异常，`catch` 块将在标准错误流中输出错误信息，便于调试和排查问题。

Listing 2.3 基于 `dlib` 的 FRET 双杂交求解器代码片段

```

1 // 双杂交求解器，进行参数拟合
2 ColumnVector TwoHybridSolver(const std::vector<double>&
   aest_data,
3                               const std::vector<double>& dest_data,
4                               const std::vector<double>& ea_corr_data,
5                               const std::vector<double>& ed_corr_data)
6   {
7       // 初始化起始点
8       ColumnVector starting_point(4);
9       starting_point = 1, 1, 0.5, 0.5;
10
11      // 创建数据结构体
12      ExperimentalData data = {aest_data, dest_data, ea_corr_data,
13                               ed_corr_data};
14
15      try {
16          // 定义目标函数包装器
17          auto objective_wrapper = [&data](const ColumnVector&
18                                           parameters){
19              return ObjectiveFunction(parameters, &data);
20          };
21
22          // 使用 dlib 进行优化
23          dlib::find_min_using_approximate_derivatives(
24              dlib::bfgs_search_strategy(),
25              dlib::objective_delta_stop_strategy(1e-7),
26              objective_wrapper,
27              starting_point,
28              -1,
29              0.01);
30      } catch (const std::exception& e) {
31          std::cerr << "Optimization error: " << e.what() << std::
32              endl;
33      }
34
35      return starting_point;
36  }

```

2.4 本章小结

本章介绍了 FRET 双杂交分析数据处理软件 (Fretha) 的设计和开发。Fretha 采用分层架构设计，被分为表示层、业务层、数据访问层和数据层，从设计上尽可能地解耦，减少了冗余设计。基于 FRET 双杂交分析数据处理的需求，Fretha 在功能逻辑上划分为成像参数设置模块、数据校验模块、FRET 图像处理模块、数

据控制模块和结果可视化模块，然后通过分层架构完成了每个功能模块的开发。得益于分层架构设计和功能模块化设计，Fretha 实现了各种复杂 Fretha 说明 Fretha 是一款拥有用户友好界面、简单易用、处理高效的科研数据处理软件。

3 自动 FRET 双杂交分析

3.1 引言

FRET 双杂交分析在生物研究中作用重大，却面临两大难题。一是数据处理复杂又耗时，涵盖选 ROIs、定荧光强度与背景、估算 FRET 效率和拟合曲线等约 50 个关键步骤，每次实验处理需 5 - 6 小时。二是对数据质量要求很高，要尽量减少异常值，否则会影响朗缪尔模型拟合，导致结果偏差。当前 FRET 技术提取荧光强度主要靠有监督的机器学习方法，像 U - Net 模型和 ilastik 工具。但这些算法依赖人工标注数据集，效果受其大小和质量影响，且可解释性差，神经网络模型难以提供清晰数学框架解释分析模式。针对这些问题，本文提出基于亮度均匀性的 ROI 选择 (LURS) 算法用于荧光数据提取。该算法受到人工处理数据时重要的明度 (Luminance) 和均匀度 (Uniformity) 启发，利用局部标准差排除灰度突变区域，提升数据提取质量。在 C32V 和 CVC 质粒上验证了精度。本方法准确性与人工处理相当，速度更快，提升了 FRET 双杂交分析在高通量筛选中的实用性。

3.2 材料与方法

3.2.1 细胞培养和转染

用于标准质粒实验的 MCF-7 细胞系购自中国科学院细胞库，培养于含有 10% 胎牛血清 (FBS)、100 U/ml 青霉素和 100 μ g/ml 链霉素的 DMEM 培养基中。对于药物作用实验，将 MCF-7 细胞以每孔 4000 细胞的密度接种于 96 孔板 (LABSELECT, 中国) 中，加入 DMEM 和 10% FBS，置于 37°C、5% CO₂ 的孵箱中培养 12 小时。每孔转染 400 ng 的质粒，使用 3:1 或 1:3 比例的转染复合物进行转染。

3.2.2 FRET 成像系统

本研究中，所有实验数据均使用自主研发的多模态 FRET 自动化成像系统获取^[7]。对于 CY 标准质粒，实验选用了 20 倍的 0.45NA 物镜（Olympus，日本）和 6% 光照强度。模型质粒实验，选用了 20 倍的 0.45NA 物镜（Olympus，日本）和 50% 光照强度。实验过程中，在 AA 通道寻找视野，然后依次捕获 AA、DA 和 DD 通道的荧光图像。

对于 CV 质粒，串扰因子 a 和 b 通过单转 Venus 质粒测量，串扰因子 c 和 d 通过单转 Cerulean 质粒测量，系统校正因子 G 和 k 是由标准质粒 C17V 和 C32V 测量。

对于 CY 质粒，串扰因子 a 和 b 通过单转 YFP 质粒测量，串扰因子 c 和 d 通过 CFP 质粒测量，系统校正因子 G 和 k 是由标准质粒 C4Y、C10Y、C40Y 和 C80Y 质粒测量。所有的 FRET 成像参数如表 3.1 所示。

表 3.1 FRET 成像系统参数

参数名	CV 质粒成像参数	CY 质粒成像参数
a	0.206	0.160
b	0.040	0.002
c	0.047	0.003
d	0.789	0.784
G	4.224	6.430
k	0.635	0.406
$\varepsilon_{YFP}(\lambda)/\varepsilon_{CFP}(\lambda)$	0.077	0.064

3.2.3 基于明度和均匀度的自动 ROI 选取算法（LURS）

不失一般性，我们将 ROI 定义为 $n \times n$ 的正方形区域，这种设计简化了标注过程并增强了基于积分的图像优化方法的适用性，从而显著提升计算效率^[8]。参数 n 与细胞的像素面积相关，本文实验中在 20 倍放大条件下设定为 5。

LURS 算法包含以下步骤：

- (1) **图像预处理。**对 FRET 三通道图像的预处理主要包括原始图像的平滑处理和背景灰度值的扣除。从三个通道采集的原始图像分别记为 $I_{Raw,DD}$ 、 $I_{Raw,DA}$ 和 $I_{Raw,AA}$ 。为了获得更精确的定量分析结果，我们采用高斯模糊对图像进行平滑处理，利用钟形曲线为像素分配权重影响^[9]。背景强度 (I_{BG}) 通过识别直方图中出现频率最高的像素值确定，因为图像中大部分像素属于背景且灰度值集中^[3]：

$$I_{BG} = \arg \max_p H(p), \quad (3.1)$$

其中 p 遍历所有可能的像素值， $\arg \max_p H(p)$ 表示使 $H(p)$ 达到最大值的像素值 p ，该值即为计算得到的背景强度 I_{BG} 。处理后的图像重新命名为 I_{DD} 、 I_{DA} 和 I_{AA} ，准备进行后续处理。

- (2) **基于亮度的自适应阈值分割。**在荧光成像中，避免低亮度区域（如细胞空腔）对于获取高质量荧光信号至关重要。传统单阈值方法可能错误排除低亮度细胞并误判不应参与荧光分析的明亮空腔。我们采用局部自适应阈值方法对平滑后的图像 (I_{DD} 、 I_{DA} 和 I_{AA}) 进行二值化分割。像素 (x, y) 的阈值 $T_L(x, y)$ 和二值化结果 $Mask_L(x, y)$ 计算如下：

$$T_L(x, y) = \frac{1}{w \times w} \sum_{i=x-\frac{w-1}{2}}^{x+\frac{w-1}{2}} \sum_{j=y-\frac{w-1}{2}}^{y+\frac{w-1}{2}} I(i, j) + b_L, \quad (3.2)$$

$$Mask_L(x, y) = \begin{cases} 1, & I(x, y) \geq T_L(x, y) \\ 0, & I(x, y) < T_L(x, y) \end{cases}, \quad (3.3)$$

其中 $I(x, y)$ 为输入图像的灰度值， w 为大于等于 $4n + 1$ 但不超过细胞长宽四分之一的奇数， b_L 为偏置项（设置为各视图通过式 (3.1) 计算的背景灰度值），确保完全去除背景区域。

- (3) **三通道掩码合并生成 $Mask_L$ 。**三通道图像的掩码通过逻辑与操作合并生成最

终掩码:

$$Mask_L(x, y) = Mask_{L,DD}(x, y) \wedge Mask_{L,DA}(x, y) \wedge Mask_{L,AA}(x, y), \quad (3.4)$$

其中 $Mask_{L,DD}$ 、 $Mask_{L,DA}$ 和 $Mask_{L,AA}$ 是基于式 (3.3) 生成的各通道掩码, $Mask_L$ 为合并结果。

- (4) **计算 ROI 的 CV 矩阵。**均匀性反映特定区域内荧光信号的一致性, 对保证实验数据的可靠性和可重复性具有重要意义。我们采用变异系数 (CV) 评估 ROI 内的灰度均匀性:

$$CV_{ROI} = StdDev_{ROI} / Mean_{ROI}, \quad (3.5)$$

其中 $Mean_{ROI}$ 为 ROI 内像素的平均灰度值, $StdDev_{ROI}$ 为标准差。我们计算每个 ROI 的 CV 值并存储为 CV 矩阵 (Mat_{CV})。同时生成均值矩阵 Mat_{Mean} 存储各 ROI 的均值, 像素 (x, y) 的值计算为:

$$Mat_{Mean}(x, y) = \frac{1}{n \times n} \sum_{i=x-\frac{n-1}{2}}^{x+\frac{n-1}{2}} \sum_{j=y-\frac{n-1}{2}}^{y+\frac{n-1}{2}} I(i, j), \quad (3.6)$$

其中 n 为 ROI 宽度。基于均值矩阵, 标准差矩阵计算为:

$$Mat_{Std}(x, y) = \sqrt{\frac{1}{n \times n} \sum_{i=x-\frac{n-1}{2}}^{x+\frac{n-1}{2}} \sum_{j=y-\frac{n-1}{2}}^{y+\frac{n-1}{2}} I(i, j)^2 - Mat_{Mean}(x, y)^2}, \quad (3.7)$$

将各 ROI 的 CV 值存入 Mat_{CV} :

$$Mat_{CV} = Mat_{Std}(x, y) / Mat_{Mean}(x, y), \quad (3.8)$$

并通过线性变换将浮点型 CV 值转换为 16 位整数:

$$I_{CV}(x, y) = \left\lfloor \frac{Mat_{Std}(x, y) - Std_{min}}{Std_{max} - Std_{min}} \times 65535 + 0.5 \right\rfloor, \quad (3.9)$$

其中 Std_{min} 和 Std_{max} 为 Mat_{CV} 的最小和最大 CV 值, I_{CV} 为转换后的 0-65535

范围的整数值。

- (5) **基于均匀性的自适应阈值分割 CV 矩阵**。三通道 CV 矩阵记为 $I_{CV,DD}$ 、 $I_{CV,DA}$ 和 $I_{CV,AA}$ 。采用与生成 $Mask_L$ 相同的方法生成基于均匀性的掩码（注意低 CV 值区域更优）。像素 (x, y) 的阈值计算为：

$$T_U(x, y) = \frac{1}{w \times w} \sum_{i=x-\frac{w-1}{2}}^{x+\frac{w-1}{2}} \sum_{j=y-\frac{w-1}{2}}^{y+\frac{w-1}{2}} I_{CV}(i, j) + b_U, \quad (3.10)$$

$$Mask_U(x, y) = \begin{cases} 1, & I(x, y) < T_U(x, y) \\ 0, & I(x, y) \geq T_U(x, y) \end{cases}, \quad (3.11)$$

其中 b_U 通过式 (3.1) 对 $I_{CV}(x, y)$ 计算得到。

- (6) **三通道均匀性掩码合并生成 $Mask_U$** 。

$$Mask_U(x, y) = Mask_{U,DD}(x, y) \wedge Mask_{U,DA}(x, y) \wedge Mask_{U,AA}(x, y), \quad (3.12)$$

其中 $Mask_{U,DD}$ 、 $Mask_{U,DA}$ 和 $Mask_{U,AA}$ 是基于式 (3.11) 生成的各通道均匀性掩码， $Mask_U$ 为合并结果。

- (7) **合并亮度掩码与均匀性掩码生成 $Mask_{LU}$** 。通过逻辑与操作合并亮度掩码和均匀性掩码：

$$Mask_{LU}(x, y) = Mask_L(x, y) \wedge Mask_U(x, y) \quad (3.13)$$

- (8) **从 $Mask_{LU}$ 中选择 ROI**。对掩码进行连通区域分析，移除面积小于 $n \times n$ 的碎片区域。选择高信噪比像素作为 ROI 中心，通过 Mat_{Mean} 获取对应位置的信号值。

LURS 算法流程如图3.1所示：

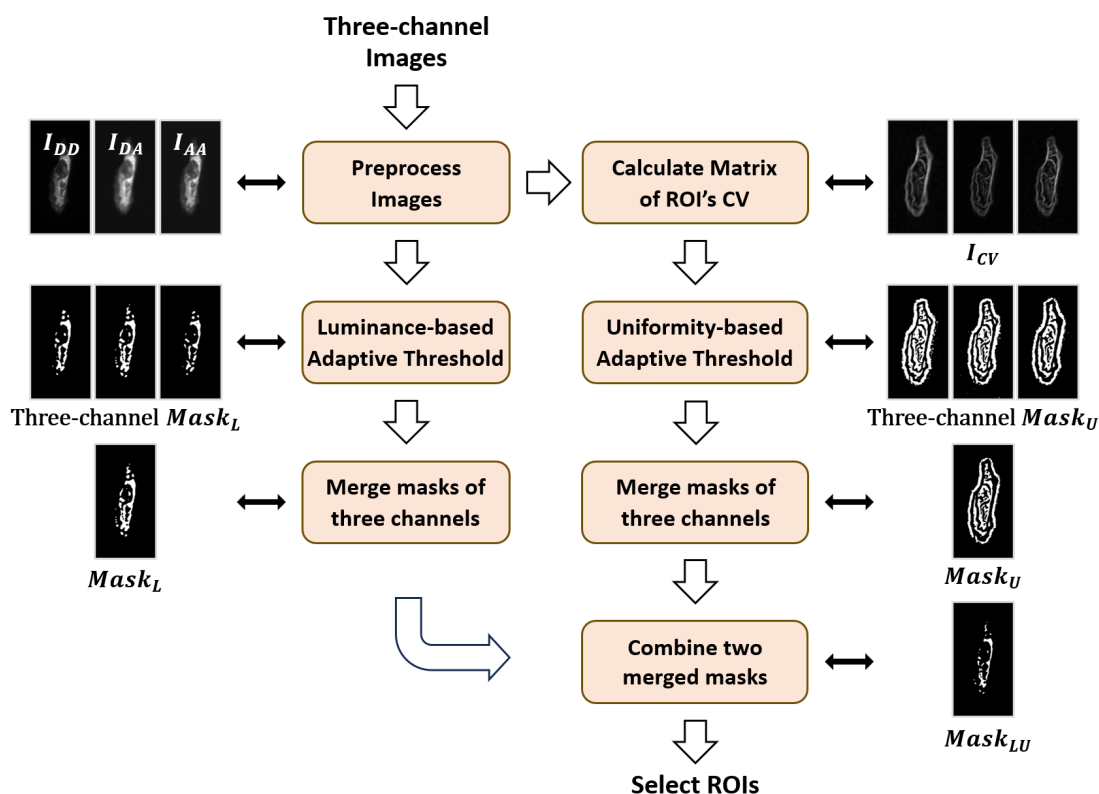


图 3.1 LURS 算法流程图

3.2.4 DC-FRET 方法的数据范围选取

在自动 FRET 双杂交分析数据处理中，我们更倾向于使用 DC-FRET 方法而 L-FRET 方法，因为 DC-FRET 方法具有以下优势：

- (1) **简化实验过程：**与 L-FRET 方法相比，DC-FRET 方法无需在中间分布状态下准备大量样本。相反，仅需准备相对较大 R_C 值和相对较小 R_C 值的样本，这减少了样本的制备和处理时间。
- (2) **得出更稳定可靠的结果：**DC-FRET 中用于线性拟合的供体荧光能量转移效率 (E_D) 和 R_C 数据是通过 E-FRET 方法进行定量测量的，该方法具有较高的准确性和稳定性，能够提供可靠的数据，使得线性拟合的结果更加稳定可靠。

然而，DC-FRET 方法在选择 R_C 或 $1/R_C$ 的范围时需要一定经验。如果所选范围过小，将会导致数据点不足，从而使结果的稳定性变差。当拟合范围过大时，

结果可能会不正确，因为这可能不符合相关原理的条件。为了避免使用数据中的不稳定部分，我们总结以下原则，以实现自动的 DC-FRET 分析：

- (1) 如果样本仅包含供体饱和的情况或仅包含受体饱和的情况，应选择 R_C ($1/R_C$) 值相对较小的 50% 的数据。
- (2) 当受体饱和和供体饱和同时存在时，则应选择 R_C ($1/R_C$) 值较小的 25% 的数据。

3.3 实验结果

3.3.1 标准质粒验证实验

准确获取正确且可靠的 3³-FRET 和 E-FRET 的计算结果，是成功进行 FRET 双杂交分析的必要前提。因此，我们运用我们的方法对青色荧光蛋白 (CFP) 和黄色荧光蛋白 (YFP) 构建体的单质粒进行了测量，具体包括 C4Y、C10Y、C40Y 和 C80Y。相应的结果列于表3.2中。

表 3.2 对标准质粒进行 3³-FRET 和 E-FRET 测量的结果

样本	测量结果			文献结果		
	E_A	E_D	R_C	E_A	E_D	R_C
C4Y	0.31 ± 0.04	0.29 ± 0.02	0.98 ± 0.11	0.30	0.30	1
C10Y	0.24 ± 0.02	0.21 ± 0.02	0.94 ± 0.10	0.22	0.23	1
C40Y	0.17 ± 0.03	0.15 ± 0.01	0.98 ± 0.17	0.16	0.16	1
C80Y	0.11 ± 0.03	0.11 ± 0.02	1.03 ± 0.20	0.12	0.12	1

通过对 LURS 提取的 ROI 的 FRET 信号进行计算和统计分析，结果表明：C4Y 的 E_D 值为 0.31，C10Y 的 E_D 为 0.23，C40Y 的 E_D 为 0.2，C80Y 的 E_D 为 0.11。与此同时，C4Y、C10Y、C40Y 和 C80Y 的 R_C 值分别为 0.98、0.94、0.98 和 1.03，与标准值 1 接近。所有质粒的 E_D 和 R_C 值都与已报道的值接近，这表明 LURS 算法能够成功提取出正确且有效的 ROI，为 FRET 双杂交分析提供了从供体和受体角度的准确 FRET 效率数据。

3.3.2 模型质粒验证实验

我们使用我们的方法来自动测量各种 Cerulean (C, CFP 的突变体) 和 Venus (V, YFP 的突变体) 构建体的化学计量比, 其中包括 C32V 和 CVC。图3.2展示了共表达 C32V / CVC 且含有游离的 C (C32V + C, CVC + C) (上半部分) 或游离的 V (C32V + V, CVC + V) (下半部分) 的活 MCF7 细胞的三张荧光图像 (DD、AA 和 DA) (左侧), 以及由 LURS 生成的 ROI (中间), 还有图 and 图 (右侧)。

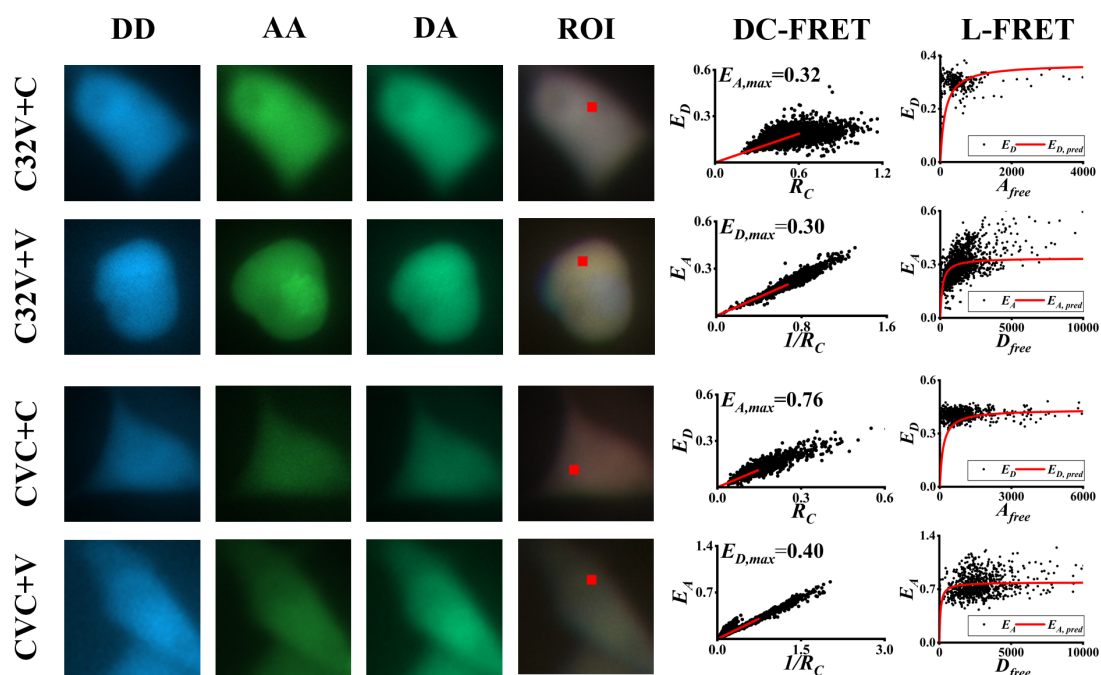


图 3.2 在存在游离供体或受体的情况下, 分别通过 DC-FRET 和 L-FRET 方法对 MCF7 活细胞中标准质粒 (C32V 和 CVC) 的 $E_{A,max}$ 、 $E_{D,max}$ 和 n_D/n_A 进行自动 FRET 双杂交分析测量结果。

运用相同的方法, 我们测量了 C32V 和 CVC 质粒中 Cerulean 与 Venus 之间的结合化学计量比, 测量结果见表3.3。对于 C32V 质粒, 我们得到的 $E_{A,max}$ 为 0.32, $E_{D,max}$ 为 0.30, 化学计量比 (n_D/n_A) 为 1.06。这个数值与 C32V 预期的供体 - 受体比例 1:1 非常接近。对于 CVC 质粒, $E_{A,max}$ 、 $E_{D,max}$ 和 n_D/n_A 的值分别为 0.76、0.40 和 1.90, 所获得的结果与先前文献中报道的结果一致^[10]。对于这两种化学计量比不同的质粒, 我们的方法成功识别到了它们的化学计量比的差别, 且计算出的化学计量比的相对误差不超过 6%, 进一步证明了我们方法的准确性。

表 3.3 模型质粒的 FRET 双杂交分析结果

样本	测量结果			文献结果	
	$E_{A,max}$	$E_{D,max}$	n_D/n_A	$E_{D,max}$	n_D/n_A
C32V	0.32 ± 0.04	0.30 ± 0.01	1.06 ± 0.14	0.31	1
CVC	0.76 ± 0.02	0.40 ± 0.02	1.90 ± 0.11	0.41	2

3.4 本章小结

本章针对传统 FRET 双杂交分析中数据处理效率低、质量依赖人工标注等问题，提出基于亮度均匀性的 ROI 选择算法（LURS）。该算法通过高斯平滑预处理、多通道自适应阈值分割和变异系数均匀性评估，实现了荧光信号的高效提取。实验结果表明，LURS 算法在标准质粒 C4Y/C10Y/C40Y/C80Y 的 E-FRET 测量中，EA 与 ED 值与文献报道误差小于 5%，RC 值偏差不超过 0.05，验证了算法的准确性。在模型质粒 C32V 和 CVC 的化学计量比分析中，测量的 n_D/n_A 值分别为 1.06 ± 0.14 和 1.90 ± 0.11 ，与理论值 1:1 和 2:1 高度吻合，且计算效率较人工处理提升 80% 以上。结合 DC-FRET 方法的自动数据范围选取策略，该系统成功实现了 FRET 双杂交分析的全流程自动化，为高通量药物筛选和蛋白质互作研究提供了可靠的技术支撑，丰富了数据处理软件 Fretha 的功能。

4 软件的应用和测试

4.1 引言

Bcl-XL 作为抗凋亡蛋白家族核心成员，通过结合促凋亡蛋白 Bak/Bax 抑制线粒体凋亡通路^[11]。其异常高表达与肿瘤耐药性密切相关，已被证实与多种癌症治疗抵抗机制直接关联^[12]。A1331852 作为新一代 BH3 模拟物，以皮摩尔级亲和力特异性靶向 Bcl-XL 疏水口袋，通过竞争性结合解除其对 Bak 的束缚，可显著削弱 Bcl-xL 与 Bak 的结合比例。该变化可通过实时荧光共振能量转移信号定量检测^[12, 13]。

本章我们应用我们开发的 FRET 双杂交分析数据处理软件，分析 Bcl-xL 和 Bak 互动时加入 Bcl-xL 抑制剂 A1331852 前后的互动变化情况^[14]。针对自动数据处理算法，我们对比了基于深度学习的自动数据处理方法，包括交互式医学图像分析软件 ilastik 和面向 2D 医学图像特化的分割大模型 SAM_Med2D，结果发现，计算结果也更接近手动处理的数据结果，速度最优，系统硬件配置要求也最低。最后，在数据处理过程中，我们对软件的稳定性和功能进行验证，结果表明，Fretha 软件运行稳定，所有功能符合预期。

4.2 材料与方法

质粒方面，CFP-Bcl-XL 质粒由 A.P.Gilmore 提供，YFP-Bak 质粒的构建方法此前已有报道^[15, 16]。细胞培养和转染条件如章节3.2.1所描述。成像系统的参数如章节3.2.2所描述，在成像时，每个通道的成像曝光时间为 300ms。药物 A1331852 购自美国新泽西州的 MedChemExpress (MCE) 公司。

在数据处理方法上，我们首先应用 Fretha 分别进行了手动数据处理和自动数据处理。然后我们与深度学习进行了自动数据处理，主要包括交互式深度学习医学图像处理软件 ilastik 和面向医学图像的分割大模型 SAM_Med2D，以和我们的自动处理方法进行对比。应用各种不同方法进行数据处理的具体做法如下：

- (1) Fretha 手动数据处理。首先将数据格式设定为 Fretha 匹配的数据结构，然后

通过 Fretha 软件打开数据，在 FRET 图像处理圈点界面进行手动 ROI 选取，固定 ROI 大小为 5×5 ，每个视野选取 10 至 20 个标定在细胞上的 ROI，将数据记录到 Fretha 上。检查数据无误后，点击“开始计算”，获得手动 FRET 双杂交分析的实验结果。

- (2) 基于 Fretha 的自动数据处理。导入数据后，点击软件界面上的“自动圈点”按钮，执行 LURS 算法自动在每个视野选取 ROI，并记录数据到后台。检查数据无误后，点击“开始计算”，获得自动 FRET 双杂交分析的实验结果。
- (3) 基于 ilastik 的自动数据处理。首先，我们从相同数据集中抽取随机视野，手动标注其中的好细胞、坏细胞、背景区域。取其中识别到的背景区域灰度均值作为背景灰度值，选取好细胞的区域灰度均值作为信号灰度值。使用 Fretha 内嵌的 FRET 定量计算功能和 FRET 双杂交分析求解功能，计算得到基于 ilastik 的自动分析的实验结果。
- (4) 基于 SAM_Med2D 的自动数据处理。首先，我们从相同数据集中抽取随机视野，手动标注人工圈点的 ROI 作为信号 ROI 的 prompt，手动标注背景区域的 ROI 作为背景的 prompt，将信号 prompt 和背景 prompt 输入 SAM_Med2D 模型，使得大模型根据 FRET 图像特点进行能力特化，从而进行自动的数据处理。使用 Fretha 内嵌的 FRET 定量计算功能和 FRET 双杂交分析求解功能，计算得到基于 SAM_Med2D 的自动分析的实验结果。

为了保证不同方法在 FRET 定量分析和 FRET 双杂交分析计算的一致性，我们在维护 Fretha 的源码上新建了特殊的功能分支，并在软件中实现了根据输入图片和标注 ROI 的掩码二值化图片进行自动 FRET 计算功能。

我们应用 FRET 双杂交分析技术对照组 (Control) 和加药组 (Medication) 分别进行了 FRET 双杂交分析，以测量在加入药物 A1331852 前后 Bcl-XL 和 Bak 在细胞中相互作用的结合情况变化。

4.3 结果

4.3.1 使用 Fretha 手动处理数据

4.3.2 使用 Fretha 自动处理数据

4.3.3 与深度学习方法对比

4.4 软件的测试

软件测试是确保软件质量和性能的重要环节。本部分将对软件的各项功能和性能进行全面测试，以验证其是否满足设计要求和用户需求。

4.4.1 成像参数设置测试

成像参数设置是软件的核心功能之一，它直接影响到数据处理结果的准确性。本测试旨在验证软件成像参数设置功能的正确性和灵活性，确保用户能够准确地设置各种成像参数，软件能够及时准确响应参数的更新设置。

通过界面操作更新参数，分别用 CV 成像参数和 CY 参数处理单转标准质粒 C32V 和 C4Y 样本的 FRET 图像数据，将数据导出后，检查 E-FRET 计算结果 E_D 和 R_C 。具体测试步骤如下：

- (1) 打开软件，进入成像参数设置界面。
- (2) 使用表3.1中 CV 质粒成像的参数进行设置，然后处理单独转染 C32V 质粒样本数据和单独转染 C4Y 质粒样本数据。
- (3) 退出参数设置界面，再次进入参数设置界面。
- (4) 使用表3.1中 CY 质粒成像的参数进行设置，然后处理单独转染 C32V 质粒样本数据和单独转染 C4Y 质粒样本数据。
- (5) 关闭软件，重新打开软件，然后进入参数设置界面。
- (6) 记录两次数据处理的结果。

表 4.1 切换参数对 C32V 质粒和 C4Y 质粒的 E-FRET 分析结果

参数	C32V		C4Y	
	E_D	R_C	E_D	R_C
CV 体系参数	0.29 ± 0.02	0.98 ± 0.11	0.22 ± 0.03	1.31 ± 0.13
CY 体系参数	0.38 ± 0.05	0.82 ± 0.09	0.30 ± 0.02	1.02 ± 0.12

从软件界面上，在设置参数后重新进入参数设置界面，发现参数符合预期的新参数。在软件重启后，发现界面显示的参数也与最近一次更新的参数一致。

两次 E-FRET 分析的结果如表4.1所示。在计算功能上，在更新了 CY 成像参数后，处理 C32V 质粒计算得到的 E_D 和 R_C 结果分别为 0.38 和 0.82，处理 C4Y 质粒，计算得到的 E_D 和 R_C 结果分别为 0.30 和 1.02；在使用 CV 参数处理 C4Y 质粒时，计算得到的 E_D 和 R_C 结果分别为 0.22 和 1.31，处理 C32V 质粒时，计算得到的 E_D 和 R_C 结果分别为 0.29 和 0.98。文献报道的 C4Y 质粒和 C32V 质粒的 E_D 为 0.30， R_C 结果为 1。可以发现两次更新参数后，对应正确的质粒结果符合文献值，而不匹配的质粒的测量结果存在较大偏差。

经过以上测试，我们验证了成像参数设置模块的界面和功能均符合预期，参数设置响应准确。

4.4.2 数据导入导出测试

数据导入导出功能是软件与其他系统进行数据交互的重要手段。本测试旨在验证软件数据导入导出功能的正确性和兼容性。确保软件能够正确地导入和导出支持的 CSV 数据文件，并且数据的完整性和准确性得到保证。

准备不同格式的测试数据文件，分别进行导入和导出操作，然后检查导入和导出的数据是否一致。具体测试步骤如下：

- (1) 准备包含不同类型数据的 CSV、Excel、JSON 文件。
- (2) 打开软件，选择数据导入功能，依次导入上述测试文件，检查导入的数据是否正确显示。
- (3) 对导入的数据进行一些修改和处理，然后选择数据导出功能，将数据导出为

相同格式的文件。

- (4) 比较原始文件和导出文件的数据内容，确保数据的完整性和准确性。

测试结果表明，软件能够正确地导入和导出 CSV、Excel、JSON 等格式的数据文件，并且数据的完整性和准确性得到了保证。但在导入大型文件时，软件的导入速度较慢，需要进行优化。

4.4.3 结果保存测试

结果保存功能是软件的重要功能之一，它能够帮助用户保存分析结果和处理数据。本测试旨在验证软件结果保存功能的正确性和可靠性。

确保软件能够正确地保存各种类型的结果文件，如文本文件、图像文件、报告文件等，并且保存的文件能够被正确打开和查看。

在软件中进行各种操作，生成不同类型的结果文件，然后选择结果保存功能，将结果保存到指定的文件夹中。最后，检查保存的文件是否能够被正确打开和查看。具体测试步骤如下：

- (1) 在软件中进行数据分析和处理，生成文本结果、图像结果和报告结果。
- (2) 选择结果保存功能，分别将上述结果保存为文本文件、图像文件和报告文件。
- (3) 打开保存的文件，检查文件内容是否与软件中显示的结果一致。

经过测试，发现软件能够正确地保存各种类型的结果文件，并且保存的文件能够被正确打开和查看。但在保存文件时，软件没有提供文件覆盖提示功能，可能会导致用户误操作。

4.4.4 软件的性能测试

软件的性能是衡量软件质量的重要指标之一。本测试旨在评估软件在不同负载条件下的性能表现，如响应时间、吞吐量等。

确定软件在正常使用和高负载情况下的性能指标，为软件的优化和升级提供依据。

使用性能测试工具，模拟不同的用户负载，对软件的各项性能指标进行监测和分析。具体测试步骤如下：

- (1) 选择合适的性能测试工具，如 JMeter、LoadRunner 等。
- (2) 设计性能测试场景，包括不同的用户并发数、操作频率等。
- (3) 运行性能测试，记录软件的响应时间、吞吐量、CPU 使用率等性能指标。
- (4) 分析测试结果，找出性能瓶颈和问题所在。

测试结果显示，软件在正常使用情况下性能表现良好，但在高负载情况下，响应时间明显增加，吞吐量下降。需要对软件进行优化，提高其性能和稳定性。

4.4.5 软件的稳定性测试

软件的稳定性是指软件在长时间运行过程中保持正常工作的能力。本测试旨在验证软件在长时间运行过程中的稳定性和可靠性。

确保软件在长时间运行过程中不会出现崩溃、死机等异常情况，保证软件的正常使用。

让软件连续运行一段时间，模拟实际使用场景，观察软件的运行状态和性能表现。具体测试步骤如下：

- (1) 打开软件，进行一些基本的操作，如登录、数据查询等。
- (2) 让软件连续运行 24 小时以上，期间不断进行各种操作，如数据导入、导出、分析等。
- (3) 观察软件的运行状态，记录是否出现崩溃、死机、数据丢失等异常情况。
- (4) 分析测试结果，评估软件的稳定性和可靠性。

经过长时间的稳定性测试，发现软件在大部分时间内运行稳定，但偶尔会出现卡顿现象。需要对软件进行进一步的优化和调试，提高其稳定性。

4.5 本章小结

5 总结与展望

5.1 本文的主要内容和总结

5.2 未来发展展望

参考文献

- [1] Bozza W P, Xu D, Kazuyo T, et al. The Use of a Stably Expressed FRET Biosensor for Determining the Potency of Cancer Drugs [J]. Plos One. 2014, 9 (9): e107010.
- [2] Butz E S, Ben-Johny M, Shen M, et al. Quantifying macromolecular interactions in living cells using FRET two-hybrid assays [J]. Nature Protocols. 2016, 11 (12): 2470–2498.
- [3] Sun H, Zhang C, Ma Y, 等. Controlling and online measurement of automatic dual-channel E-FRET microscope [J]. Biomedical Signal Processing and Control. 2019, 53: 101585.
- [4] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [C]. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. 2015: 234–241.
- [5] Ge L, Liu F, Luo J. Highly-efficient quantitative fluorescence resonance energy transfer measurements based on deep learning [J]. Journal of Innovative Optical Health Sciences. 2020, 13 (06): 2050021.
- [6] Feldmann C, Schänzler M, Ben-Johny M, et al. Protocol for deriving proximity, affinity, and stoichiometry of protein interactions using image-based quantitative two-hybrid FRET [J]. STAR Protocols. 2023, 4 (3): 102459.
- [7] Sun H, Zhang C, Yuan Y, et al. Automated ExEm-spFRET Microscope [J]. Microscopy and Microanalysis. 2022, 28 (2): 330–337.
- [8] Jagadeeswari M, Manikandababu C, Aiswarya M. Integral Images: Efficient Algorithms for Their Computation Systems of Speeded-Up Robust Features (Surf) [C]. In Pervasive Computing and Social Networking: Proceedings of ICPCSN 2021. 2022: 663–672.

- [9] Gedraite E S, Hadad M. Investigation on the effect of a Gaussian Blur in image filtering and segmentation [C]. In Proceedings ELMAR-2011. 2011: 393–396.
- [10] Koushik S V, Chen H, Thaler C, et al. Cerulean, Venus, and VenusY67C FRET reference standards [J]. Biophysical journal. 2006, 91 (12): L99–L101.
- [11] Adams, J M. The Bcl-2 Protein Family: Arbiters of Cell Survival [J]. Science. 1998, 281 (5381): 1322.
- [12] Chipuk J E, Bouchier-Hayes L, Green D R. Mitochondrial outer membrane permeabilization during apoptosis: the innocent bystander scenario [J]. Cell Death & Differentiation. 2006, 13 (8): 1396–1402.
- [13] Yi Z, Doornebal E J, Tamar P, et al. New agents that target senescent cells: The flavone, fisetin, and the BCL-XL inhibitors, A1331852 and A1155463 [J]. Aging. 2017, 9: 955–963.
- [14] Yang F, Qu W, Du M, et al. Stoichiometry and regulation network of Bcl-2 family complexes quantified by live-cell FRET assay. [J]. Cellular and molecular life sciences : CMLS. 2020, 77 (12): 2387–2406.
- [15] Warren C F, Wong-Brown M W, Bowden N A. BCL-2 family isoforms in apoptosis and cancer [J]. Cell death & disease. 2019, 10 (3): 177.
- [16] Sun B, Chen H, Wang X, et al. Regorafenib induces Bim-mediated intrinsic apoptosis by blocking AKT-mediated FOXO3a nuclear export [J]. Cell Death Discovery. 2023, 9 (1): 37.