

基于焦点损失与门控优化的单阶段事件抽取改进模型

刘向杰, 梁志剑*

(中北大学 计算机科学与技术学院, 山西 太原 030051)

摘要: 事件抽取旨在从非结构化文本中提取结构化信息,但在处理重叠与嵌套复杂事件时,现有单阶段模型面临任务级误差传播、长尾类别识别瓶颈及特征表示冗余等挑战。为解决上述问题,本文提出一种基于结构化代价敏感学习与特征去冗余门控的改进模型。首先,在损失函数设计上,提出结构化焦点联合损失,通过引入聚焦参数动态调节难分样本权重,并设计任务依赖加权因子,构建梯度层级约束,从算法机制上阻断上游触发词检测误差向论元角色的传播。其次,针对特征融合中的噪声干扰,设计求和式门控事件融合层,利用Sigmoid门控信号生成加性干涉,替代传统线性拼接策略,在维持特征维度稳定的前提下实现内生性去噪。在中文金融及生物医学数据集上的实验显示,该模型在FewFC上的F1值较OneEE基准提升1.6个百分点,优于生成式模型Degree及大模型修正策略LC4EE;在Genia13数据集上,论元识别F1值达到80.2%,提升3.4个百分点。消融实验表明,移除结构化损失会导致F1下降1.8个百分点。本文提出的改进方法在保持并行推理高效性的同时,有效解决了逻辑约束缺失与特征噪声问题,提升了复杂场景下的结构化语义解析精度。但对于训练样本极少的极端长尾类别,模型的泛化能力仍有待通过少样本学习进一步增强。

关键词: 事件抽取; 多标签分类; 焦点损失; 结构化损失; 重叠事件; 嵌套事件

中图分类号: TP391.1

文献标识码: A

doi: 10.62756/jnuc.issn.1673-3193.2025.10.0008

引用格式: 刘向杰, 梁志剑. 基于焦点损失与门控优化的单阶段事件抽取改进模型[J]. 中北大学学报(自然科学版), 2026, 47(3): 395-405.

LIU Xiangjie, LIANG Zhijian. Improved single-stage event extraction model based on focal loss and gating optimization[J]. Journal of North University of China(Natural Science Edition), 2026, 47(3): 395-405.

Improved Single-Stage Event Extraction Model Based on Focal Loss and Gating Optimization

LIU Xiangjie, LIANG Zhijian*

(School of Computer Science and Technology, North University of China, Taiyuan 030051, China)

Abstract: Event extraction aims to extract structured information from unstructured text. However, when handling complex overlapping and nested events, existing single-stage models face challenges such as task cascading error propagation, long-tail class recognition bottlenecks, and feature representation redundancy. To address these issues, this paper proposed an improved model based on structured cost-sensitive learning and feature redundancy elimination gating. First, in the design of the loss function, the structured

收稿日期: 2025-10-13

基金项目: 山西省2024年度研究生教育创新计划项目(2024SZ23); 2024年山西省高等学校教学改革创新项目(J20240839)

作者简介: 刘向杰(2001—),男,硕士生,主要从事人工智能、自然语言处理的研究。

* 通信作者: 梁志剑(1978—),男,教授,博士,主要从事人工智能、自然语言处理的研究。E-mail: 116585916@qq.com。

focal joint loss was proposed. By introducing a focusing parameter, this mechanism dynamically adjusted the weights of hard samples. We also designed novel task-dependent weighting factors to construct hierarchical gradient constraints, which algorithmically prevented the propagation of upstream trigger detection errors to argument roles. Second, to tackle noise interference during feature fusion, we designed a summation-based gated event fusion layer. This layer utilized Sigmoid gating signals to generate additive interference, replacing traditional linear concatenation strategies to achieve endogenous redundancy elimination while maintaining stable feature dimensionality. Experimental results on the Chinese financial and biomedical datasets show that the $F1$ score of the proposed model on FewFC improves by 1.6 percentage points compared to the OneEE baseline, outperforming the generative model Degree and the LLM correction strategy LC4EE. On the Genia13 dataset, the argument identification $F1$ score reaches 80.2%, improvement of 3.4 percentage points. Ablation studies indicate that removing the structured loss leads to a 1.8 percentage points decrease in $F1$ score. The improved method proposed in this study effectively solves the problems of missing logical constraints and feature noise while retaining the high efficiency of parallel inference, and enhances the accuracy of structured semantic parsing in complex scenarios. However, for extreme long-tail classes with very few training samples, the generalization ability of the model remains to be further enhanced through few-shot learning.

Key words: event extraction; multi label classification; focal loss; structured loss; overlapping events; nested events

0 引言

随着信息技术的高速发展,在线平台产生了海量文本资源,为信息抽取(Information Extraction, IE)提供了广阔应用前景。作为IE的重要分支,事件抽取(Event Extraction, EE)因其能够从自然语言中自动提取结构化事件信息(如触发词及其论元角色)而备受关注^[1-2]。事件抽取旨在从纯文本中获取关于事件的结构化信息,该任务通常通过触发词抽取和论元抽取来完成^[3-5]。近年来,基于深度学习的方法在这一领域取得了显著进展,涵盖了从句子级到篇章级的多种范式^[6-7]。然而,尽管相关研究层出不穷,事件抽取仍被视为自然语言处理领域中一项重要且极具挑战性的任务^[8]。事件抽取在多个领域展现了显著价值,例如,在安全领域,Tanev等^[9]通过实时新闻事件抽取实现了全球危机监测;在智能交通领域,Sakaki等^[10]利用社交媒体提取实时驾驶信息,为驾驶员提供了交通拥堵和天气报告等关键事件;在金融领域,Sheng等^[11]探索了重叠事件抽取问题,Zheng等^[12]提出了一种从中文金融公告中进行端到端文档级事件抽取的新框架。

传统方法多侧重于平坦事件的提取,主要依赖于基于BIO方案的序列标注模型或多阶段流水线

框架。然而,在面对中文金融领域的重叠事件或生物医学领域的嵌套事件时,这些方法因标签歧义及任务间级联误差传播而面临性能瓶颈。单阶段框架OneEE^[13]通过定义跨度关系与角色关系,将复杂抽取任务转化为并行词对预测,实现了复杂事件结构的统一建模。

尽管单阶段框架在并行推理效率上取得了显著进展,但本文深入分析发现,其在处理长尾分布与高噪声数据时仍面临两个本质缺陷。首先,该框架忽略了触发词与论元间的逻辑依存关系。这种结构约束的缺失导致上游错误被级联放大,严重制约了长尾样本的识别性能。其次,简单的特征拼接引入了大量无关噪声。特征拼接引入的冗余信息不仅增加了模型拟合的难度,更削弱了其在重叠事件场景下的判别鲁棒性。与此同时,虽然Degree^[14]和Code4Struct^[15]等生成式方法在少样本场景下表现优异,但其自回归机制带来的高延迟与幻觉风险,严重制约了其在复杂工业场景下的应用。因此,如何在保持单阶段框架高效推理优势的同时,克服其逻辑缺失与噪声干扰的短板,成为当前亟待解决的关键问题。

针对上述挑战,本文致力于挖掘判别式单阶段框架的潜力,提出了一种基于结构化约束与内生去冗余的改进模型。为了阻断级联误差并解决样本失衡,本文构建了结构化焦点联合损失(SF-

Joint Loss), 赋予上游检测更高的梯度权重, 引入动态焦点参数, 建立层级化约束。针对特征交互中的噪声问题, 本文提出了求和式门控策略, 利用加性干涉替代线性拼接, 在同一维度空间内有效滤除无效噪声。通过上述结构优化与特征重构, 本文模型不仅保留了并行推理的高效性, 更成功实现了对生成式 SOTA 模型的性能超越。

1 相关工作

1.1 事件抽取

在自然语言处理(NLP)的研究中, 事件抽取在信息抽取领域中被公认为是最具挑战性的任务之一。如图 1 所示, 事件抽取通过触发词抽取和论元抽取, 涉及命名实体识别和关系抽取, 其性能在很大程度上依赖于这些基础子任务的准确性。根据 ACE2005 的事件抽取任务定义, 事件被描述为一种状态变化, 具体表现为在特定时间和地点涉及一个或多个参与者的动态事项。这一定义为事件抽取提供了理论基础, 同时也凸显了其在复杂文本分析中的重要性。传统事件提取方法主要分为序列标注方法和多阶段方法两大类, 各具特色但也存在显著局限。

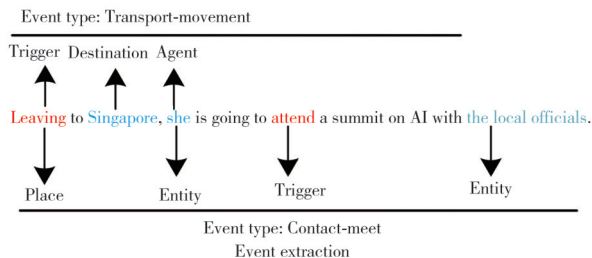


图 1 事件抽取示例图

Fig. 1 Example of event extraction

序列标注方法以 BIO 标记方案为基础, 常结合先进的预训练模型, 如 BERT-CRF, 利用序列标注技术提取触发词和论元^[16]。这种方法通过将事件抽取转化为序列预测任务, 依托上下文语义的连续性来识别事件组件。然而, 在面对重叠事件或嵌套事件时, 序列标注方法因标签歧义和上下文依赖性不足而表现不佳, 导致识别精度显著下降。此外, BERT-CRF 模型虽然在单一事件提取中表现出色, 但其对复杂结构的支持有限, 难以适应现实场景中多样化的事件模式。

多阶段方法则采用分步处理的策略, 通过先后执行触发词检测和论元抽取来完成事件提取任

务^[17]。例如 PLMEE 和 CasEE 分别利用预训练语言模型和多阶段优化, 通过逐步细化提高了抽取性能。动态多池卷积神经网络^[18]作为多阶段方法的代表, 采用两个动态多池层分别处理触发词和论元分类, 尽管 Yang 等^[19]和 Xu 等^[20]在一定程度上解决了论元重叠问题, 但触发词重叠的挑战仍未有效克服。这种方法因子任务之间的错误传播而受到限制, 尤其在处理嵌套事件时, 假设事件不重叠的简化模型设计使其难以满足复杂现实需求的准确性要求。此外, 多阶段方法的计算复杂度较高, 推理效率较低, 进一步限制了其在大规模数据集中的应用。

针对传统方法的不足, OneEE 提出了一个创新的单阶段框架, 定义了跨度关系和角色关系, 旨在克服重叠和嵌套事件提取中的难题。该框架依托强大的 BERT 编码器, 深度挖掘文本的语义表示, 并通过自适应事件融合层整合多源特征信息, 有效捕捉事件间的复杂依赖关系。

此外, 针对特定领域数据稀缺的问题, Liu 等^[21]提出了动态前缀微调技术以提升模板化抽取性能; Gao 等^[22]则提出了 Mask-then-fill 框架, 利用大语言模型的生成能力进行数据增强; Zhu 等^[23]提出的 LC4EE 利用 LLM 作为修正器改进了抽取结果。然而, 此类生成式或 LLM 辅助方法在大规模并行处理及重叠场景下仍面临推理延迟高、召回保守或幻觉干扰的挑战。

1.2 损失函数

损失函数的设计是提升事件抽取性能的核心环节, 其关键在于平衡样本难度与缓解类别分布不均。焦点损失(Focal Loss)由 Lin 等^[24]首次提出, 最初应用于目标检测领域, 通过引入聚焦参数 γ 和类别平衡参数 α , 以 $(1 - p_i)^\gamma$ 动态调整易样本和难样本的权重, 迫使模型聚焦于稀有触发词等难分样本, 在 NLP 长尾场景中展现出显著潜力。

近年来, 代价敏感学习在 NLP 领域的应用持续深化。Henning 等^[25]指出, 优化长尾分布的损失函数是提升模型鲁棒性的关键, 尤其在细粒度抽取任务中。在此背景下, Su 等^[26]提出了 MiLe Loss。该方法利用信息熵动态评估样本难度, 相比传统 Focal Loss 更有效修正了多标签分类偏差。此外, Ma 等^[27]针对样本稀缺问题, 在 DICE 框架中引入对比学习目标。该策略通过扩大类间特征距离, 显著改善了模糊边界下的抽取性能。

现有的改进策略主要针对扁平化标签空间或生成式架构,侧重于解决通用的分布不平衡,却忽视了事件抽取任务中触发词决定论元。由于缺乏对逻辑蕴含关系的显式建模,上游任务的预测偏差往往会无阻碍地传递至下游。针对这一问题,本文提出了结构化焦点联合损失。该方法继承了Focal Loss挖掘难样本的优势,并引入了结构化梯度约束。通过赋予上游触发词检测更高的优化权重,模型促使下游论元任务必须与触发词的判别状态保持一致。这种非对称的梯度流机制从根本上阻断了级联误差的隐性传播。

2 方法设计

2.1 问题形式化

事件抽取的目标是提取事件触发词及其论元。给定由 N 个词组成的输入句子 $X=\{x_1, x_2, \dots, x_N\}$ 和事件类型 $e_i \in E$,目标是预测词对 (x_i, x_j) 之间的跨度关系(Span)和角色关系(Role),其中 E 表示事件类型集合。预定义标签集合被划分为跨度标签集 S 与角色标签集 R ,如图2所示,对于图中输入的句子,模型需要捕捉特定词对间的依赖属性。通过预测触发词“acquired”(对应 x_i)与实体“Citic Securities”(对应 x_j)之间存在的跨度关联(S-T与S-A),进一步界定其角色关系标签为施事者(R-S)。通过上述直观的对应规则,将复杂的重叠与嵌套事件抽取任务统一转化为基于预定义标签 S 和 R 的词对关系分类

任务。

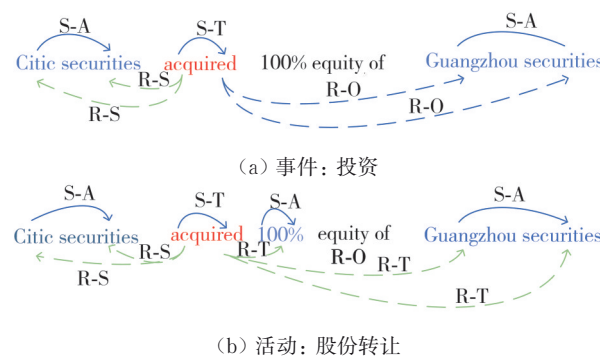


图2 跨度关系和角色关系图

Fig. 2 Span relationship and role relationship diagram

2.2 模型架构

如图3所示,总体框架分为三层:预训练编码层、求和门控事件融合层与多头词对标注器。该架构在保留单阶段处理高并行度优势的同时,通过层级协同实现精准语义与事件特征捕捉:PLM Encoder依托预训练语言模型生成含上下文信息的文本向量;求和式门控事件融合层(SGEF)以门控机制动态调节事件特征权重,完成多维度信息筛选融合;MHPT则通过多头部结构聚焦不同词对关系维度,既提升标注针对性又借头部互补缓解样本不平衡;同时,架构通过合理特征交互与任务拆分降低模型对特定样本模式的依赖,有效缓解过拟合,最终实现处理效率与任务性能稳定性的平衡。

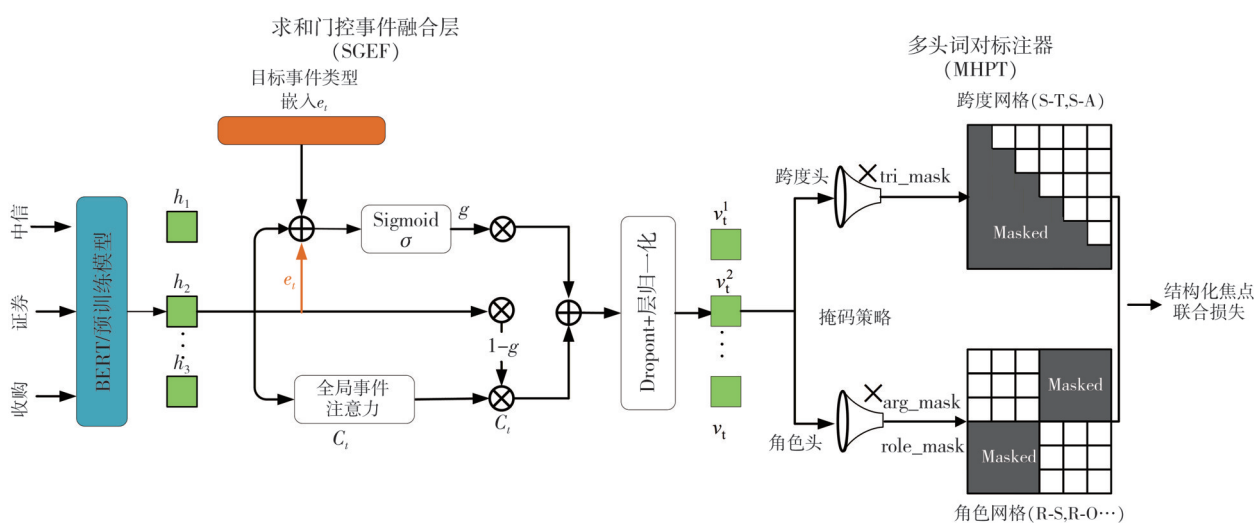


图3 模型架构图

Fig. 3 Model architecture diagram

求和门控事件融合层:针对传统拼接操作引入噪声的问题,本文设计了SGEF层(见图3前半

部分)。不同于传统的线性拼接策略,该层通过求和运算生成门控信号,实现了特征空间的深度干

涉与内生性去冗余。此外,如图3所示,在门控输出后显式引入了Dropout+LayerNorm模块,以进一步缓解过拟合风险并提升模型在噪声环境下的泛化能力。设PLM输出的上下文表示为 $H \in \mathbb{R}^{N \times d}$,预定义的事件类型嵌入矩阵为 $E_{\text{type}} \in \mathbb{R}^{K \times d}$ (其中 N 为句子长度, K 为事件类型数, d 为隐藏层维度)。

事件上下文交互:首先,利用多头注意力机制捕获上下文与特定事件类型的关联。对于第 t 个事件类型嵌入 $e_t \in \mathbb{R}^d$,计算其与上下文 H 的交互特征 C_t 。

$$C_t = \text{Attention}(Q = e_t, K = H, V = H), \quad (1)$$

式中: Q 、 K 、 V 分别为注意力机制中的查询向量(Query)、键向量(Key)和值向量(Value)。在计算中,以特定事件类型嵌入 e_t 作为查询向量去匹配上下文表示 H (即键向量与值向量)。这一步将全局上下文信息映射到特定事件的语义空间中。

求和式门控融合:不同于OneEE直接拼接 H 与 e_t ,提出求和门控机制。该机制利用门控信号 g_t 动态平衡原始上下文的保留与事件信息的注入。

$$g_t = \sigma(W_g H + W_e e_t + b_g), \quad (2)$$

$$V_t = g_t \odot H + (1 - g_t) \odot C_t, \quad (3)$$

式中: σ 为Sigmoid激活函数; W_g 、 W_e 、 b_g 为可学习参数。这里的求和操作不仅是降维,更是一种特征干涉。

OneEE的拼接操作仅简单扩展了特征空间,本文提出的求和策略具有明确的物理意义。从信息论角度,拼接操作保留了 H 中所有的上下文噪声,导致关键语义特征被无关维度稀释,削弱分类器的判别显著性,难以划定清晰的决策边界。这里的求和操作使事件特定特征 C_t 与原始上下文 H 在同一 d 维特征空间进行加性干涉。门控信号 g_t 在此充当了语义滤波器的角色:根据当前事件类型的拓扑结构,动态抑制与事件无关的背景噪声,仅允许有效语义通过。这种机制实现了特征空间的内在提纯,在维持表征紧凑性的前提下,增强了模型对重叠事件中细微差别的判别能力。

正则化与归一化:为防止过拟合,本文在输出前引入Dropout与LayerNorm。

$$V_t^{\text{final}} = \text{LayerNorm}(V_t + \text{Dropout}(V_t)). \quad (4)$$

式中: V_t^{final} 为针对第 t 个事件类型,经过求和门控融合后输出的特定事件特征表示; V_t 为经过求和门控机制初步融合后的中间变量。

多头词对标注器与掩码机制:在预测阶段,模型并行预测跨度与角色关系。为减少冗余计算与无效梯度,在MHPT层引入了精细化的掩码策略,如图3右侧Masking示意。具体包括:tri_mask用于过滤跨度网格的上三角区域;arg_mask与role_mask用于在角色网格中动态屏蔽无关的论元区域。这种掩码机制确保了损失函数仅关注有效的“触发词-论元”对,从而提升了训练效率。

2.3 任务依赖的结构化焦点损失

针对多标签分类中正负样本失衡以及任务间的级联依赖,本文提出了结构化焦点联合损失。

基础焦点损失:对于任意词对 (x_i, x_j) 的特定标签分类,设模型给出的预测概率为 p_{ij} ,对应的真实标签 $y_{ij} \in \{0, 1\}$ 。传统交叉熵容易被大量易分类的负样本主导梯度。引入聚焦参数 γ 和平衡参数 α 。

$$\mathcal{L}_{\text{focal}}(p_{ij}, y_{ij}) = -\alpha(1 - p_{ij})^\gamma \log(p_{ij}) \cdot y_{ij} - (1 - \alpha)p_{ij}^\gamma \log(1 - p_{ij}) \cdot (1 - y_{ij}). \quad (5)$$

当模型对负样本预测准确时,权重项 p_{ij}^γ 迅速趋近于0,使损失函数自动“忽略”这些简单样本,强迫梯度下降方向聚焦于稀有的触发词与难以区分的论元边界。

任务依赖加权:为了阻断误差传播,本文设计了具有层级约束的总损失函数。

$$\mathcal{L}_{\text{total}} = \lambda_{\text{tri}} \cdot \mathcal{L}_{\text{trigger}} + \lambda_{\text{arg}} \cdot \mathcal{L}_{\text{argument}} + \lambda_{\text{role}} \cdot \mathcal{L}_{\text{role}}, \quad (6)$$

式中: $\mathcal{L}_{\text{total}}$ 为模型的总联合损失; $\mathcal{L}_{\text{trigger}}$ 、 $\mathcal{L}_{\text{argument}}$ 、 $\mathcal{L}_{\text{role}}$ 分别为触发词检测、论元识别与角色分类三个独立子任务对应的焦点损失项; λ_{tri} 、 λ_{arg} 、 λ_{role} 分别为分配给这三个子任务的任务依赖加权因子。

在优化理论层面,本文提出的联合损失函数 $\mathcal{L}_{\text{total}}$ 本质上是多个焦点损失函数的非负线性组合。根据Lin等^[24]的研究,焦点损失函数在定义域内连续可微且下界有界,具备良好的收敛特性。根据多任务学习的凸优化性质,其线性加权组合依然保持了原有的梯度下降收敛属性。在此数学基础上,设计了触发词检测、论元识别与角色分类权重分配 $\lambda_{\text{tri}} > \lambda_{\text{arg}} > \lambda_{\text{role}}$ (本文设定为1.5:1.0:0.8)。这种递减的权重设计并非仅改变梯度的数值大小,而是构建了一种结构化梯度势能。在反向传播过程中,较大的 λ_{tri} 使得触发词检测任务产生的梯度在更新中占据主导地位。一旦上游触发词识别错误,损失函数产生的强烈惩罚信号将占据统治地位,进而削弱下游论元任务对整体优化方向的影响。这种机制实际上在

并行计算图中模拟了序列模型的逻辑依赖,有效防止了模型在触发词未确定的情况下过度拟合论元细节,从而解决了OneEE中因触发词误检导致的下游无效计算与幻觉问题。

3 实验设置

3.1 数据与预处理

本文在FewFC^[28](中文金融,重叠事件)与Genia2011^[29]/Genia2013(生物医学,嵌套事件)数据集上进行实验。统计分析表明,FewFC数据集呈现明显的样本长尾特征,其中占比不足10%的稀有类别事件贡献了大部分识别难点。为应对此分布特征,本文在数据预处理阶段通过实体边界对齐来确保小众类别的结构完整性。为提高样本代表性,本文将数据集按8:1:1的比例划分为训练集、验证集和测试集。在划分过程中采用分层抽样策略,确保各事件类型与论元角色的分布平衡。文本预处理包括中英文统一、符号正则化及实体边界对齐,各数据集的统计信息见表1。

表1 数据集统计信息

Tab.1 Dataset statistics

数据集	子集	Ovlp.	Nest.	Sent.	Events.
FewFc	train	1 560	—	7 185	10 277
	dev	205	—	899	1 281
	test	210	—	898	1 332
Genia11	train	954	1 628	8 730	6 401
	dev	121	199	1 091	824
	test	125	197	1 092	775
Genia13	train	347	784	4 000	2 743
	dev	44	100	500	352
	test	42	88	500	320

注:“Ovlp.”和“Nest.”分别表示包含重叠事件和嵌套事件的句子数量。

3.2 评价指标

针对触发词识别(Trigger Identification, TI)、触发词分类(Trigger Classification, TC)、论元识别(Argument Identification, AI)和论元分类(Argument Classification, AC)4项任务,分别计算精确率 P (Precision)、召回率 R (Recall)和 $F1$ 值^[30-31]。为增强统计稳健性,采用3次随机种子实验取均值±标准差,并以验证集宏平均AC-F1作为早停准则。

3.3 实验设置

本文模型基于PyTorch框架实现。为确保与基

准模型OneEE进行公平对比,保持了与其一致的基础训练配置:优化器选用AdamW,批次大小设为16。学习率采用分层策略:BERT编码器参数的学习率设为 2×10^{-5} ,而新增层的学习率设为 1×10^{-3} ,以加快收敛。针对改进模块,超参数配置如下:

1) 焦点损失函数:聚焦参数 $\gamma=2.0$ (标准值,用于调节难样本的聚焦程度),类别平衡参数 $\alpha=0.25$ 以修正正负样本比例平衡。

2) 结构化权重:损失函数的任务依赖权重设定为 $\lambda_{\text{tri}}:\lambda_{\text{arg}}:\lambda_{\text{role}}=1.5:1.0:0.8$,该设置赋予上游触发词检测更高的梯度权重,旨在抑制级联误差传播。

3) 正则化设置:SGEF层中Dropout率设为0.1,MHPT层前加LayerNorm,以防止过度拟合。训练在单卡RTX3090上完成,若验证集5轮无提升则触发早停。

为了验证本文改进模型的有效性,将实验结果与涵盖不同技术范式的代表性基线方法进行了对比。首先是传统的序列标注方法,包括利用BERT结合softmax或CRF层的BERT-softmax与BERT-CRF,以及通过扩展标签空间实现同步抽取的BERT-CRF-joint;其次是针对重叠与嵌套挑战的多阶段或问答式框架,涵盖了采用Suffixmax策略的PLMEE、基于机器阅读理解架构的MQAEE,以及通过类型检测与级联解码处理复杂重叠目标的CasEE;最后对比当前先进的单阶段与生成式模型,包括作为本文核心改进基准的并行化词对关系建模框架OneEE,以及基于提示学习的条件生成模型Degree和利用大语言模型作为外部修正器的LC4EE策略。所有对比实验均在验证集上完成了超参数优化,且事件嵌入均从零开始训练,旨在从不同技术演进方向展现本文改进机制在处理复杂中文金融与医疗文本时的优势。

实验设置:运行环境配置为Ubuntu20.4操作系统,硬件配备10vCPU,GPU采用单卡RTX3090;本文在FewFC和Genia11数据集的最大训练轮数设定为20轮,Genia13设定为30轮;训练过程中配合早停机制以防止过拟合;所有超参数均在验证集上完成优化;所有事件类型嵌入均从零开始训练。

4 实验结果

4.1 所有事件抽取结果

表2呈现了各方法在重叠事件抽取数据集

FewFC 上的实验结果,表 3 为相关方法在嵌套事件抽取数据集 Genia11 与 Genia13 上的性能指标。

从表 2 实验结果可明确观察到,改进方法的综合性能优于 OneEE 基准方法,在重叠事件场景下,本模型相较于 OneEE 在 AC 召回率提升了 1.6 百分点, $F1$ 分数提升了 1.6 百分点。对比生成式模型 Degree,本方法在 AC- $F1$ 指标上高出

2.8 百分点。与 LC4EE 模型相比,本方法依然保持了 0.4 百分点的优势。证明了在处理复杂中文金融重叠事件时,本模型通过结构化焦点联合损失对长尾样本的精准捕捉能力,优于通用的 LLM 修正策略,且避免了后者高昂的推理开销,有效缓解了长尾类别样本的“淹没效应”。其结构化建模能力相比生成式方法的软约束更具鲁棒性。

表 2 FewFC 上提取各种事件的结果
Tab. 2 Extract the results of various events on FewFC

方法分类	模型	TI			TC			AI			AC		
		P	R	$F1$	P	R	$F1$	P	R	$F1$	P	R	$F1$
FlatEE	BERT-softmax	89.8	79.0	84.0	80.2	61.8	69.8	74.6	62.8	68.2	72.5	60.2	65.8
	BERT-CRF	90.8	80.8	85.5	81.7	63.6	71.5	75.1	64.3	69.3	72.9	61.8	66.9
	BERT-CRF-joint	89.5	79.8	84.4	80.7	63.0	70.8	76.1	63.5	69.2	74.2	61.5	67.1
Ovlp. & Nest.EE	PLMEE	83.7	85.8	84.7	75.6	74.5	75.1	74.3	67.3	70.6	72.5	65.5	68.8
	MQAEE	89.1	85.5	87.4	79.7	76.1	77.8	70.3	68.3	69.3	68.2	66.5	67.3
Gen & Onestage	CasEE	89.4	87.7	88.6	77.9	78.5	78.2	72.8	73.1	72.9	71.3	71.5	73.4
	Degree	89.2	88.1	88.6	79.5	78.8	79.1	74.2	73.5	73.8	72.8	71.6	72.2
	OneEE	88.7	88.7	88.7	79.1	80.3	79.7	75.4	77.0	76.2	74.0	72.9	73.4
改进后	LC4EE	89.5	89.5	89.5	80.8	80.2	80.5	77.5	76.9	77.2	75.2	74.0	74.6
	SFG-Net	90.0	89.8	89.9	81.6	80.2	80.9	76.8	78.5	77.6	75.5	74.5	75.0

表 3 展示了各模型在生物医学数据集上的性能。改进后的模型在 Genia11 和 Genia13 上的 TC- $F1$ 分数平均优于 OneEE 约 2.1 百分点。在 AI(论元识别)指标上,本模型在 Genia13 数据集上达到了 80.2% 的高水平,相较于 OneEE 显著提升了 3.4 百分点。这一结果充分表明,针对生物医学文本中高度嵌套的论元识别痛点,本文优化的求和式门控事件融合层能够更有效地提取事件间的深层语义关联。此外,在 Genia13 数据集上,本模型的 AC- $F1$ 优于 LC4EE 约 0.7 百分点,表明在生物医学高度嵌套的场景下,基于 SGEF 的内生性特征去噪机制,比依赖外部知识的 LLM 修正更能精确解析细微的论元边界。

表 3 Genia11 和 Genia13 上提取所有事件的 $F1$ 值
Tab. 3 $F1$ scores for extracting all events on Genia11 and Genia13

模型	$F1/\%$							
	Genia11				Genia13			
	TI	TC	AI	AC	TI	TC	AI	AC
BERT-softmax	67.5	64.5	57.4	56.1	77.5	76.0	69.8	67.8
BERT-CRT	68.3	64.8	58.2	56.8	78.8	77.4	70.1	68.2
BERT-CRF-joint	67.1	64.2	60.3	58.1	77.8	75.6	71.8	68.1
PLMEE	67.3	65.5	60.6	59.3	79.5	78.3	72.2	70.6
CasEE	70.1	67.1	62.1	60.3	80.5	78.5	73.7	71.8
Degree	72.1	70.5	63.8	62.1	82.4	81.1	74.9	72.9
OneEE	71.5	69.4	65.7	62.8	81.7	80.6	76.8	72.6
LC4EE	72.5	70.9	68.0	63.8	82.9	81.8	79.2	73.5
改进后	73.0	71.5	69.1	64.4	83.5	82.7	80.2	74.2

综合上述重叠和嵌套数据集的实验结果,本模型在 AC- $F1$ 指标上表现稳健,特别是在 FewFC 数据集上,AC- $F1$ 分数较核心基准 OneEE 提升了 1.6 百分点。上述结果充分证明,改进方法在复杂重叠事件与嵌套事件抽取任务中具备超越 OneEE 的处理能力,展现出更高的有效性。

综合实验数据分析,改进后的单阶段框架在无误差传播的前提下,通过对正负样本损失贡献的平衡优化,提升了触发词分类和论元分类的精度。尤其是在面对稀有触发词和深度嵌套论元等难分类样本时,本模型表现出更高的识别灵敏度与解析性能,验证了其在复杂场景下作为单阶段架构改进方案的有效性与领先性。

4.2 消融实验结果

为了进一步探究结构化焦点联合损失与求和门控事件融合层对模型性能的具体贡献,本文在 FewFC 数据集上进行了消融实验。通过在改进模型的基础上分别移除关键组件,对比其 TI、TC、AI 和 AC 指标的变化。实验结果如表 4 所示。

表 4 FewFC 的消融研究
Tab. 4 Ablation studies using FewFC

模型	$F1/\%$			
	TI	TC	AI	AC
改进后	89.9	80.9	77.6	75.0
w/o SF-Joint Loss	89.2	79.7	77.0	73.2
w/o SGEF	88.9	79.9	76.2	74.1
w/o Dropout	89.4	80.3	77.0	74.5

表4的消融数据分析:

1) 结构化焦点联合损失的影响。当移除SF-Joint Loss恢复为标准交叉熵损失时, TC-F1和AC-F1分别下降了1.2个百分点和1.8个百分点, 验证了该损失函数在缓解事件抽取中“样本极度不平衡”问题上的核心价值。为了进一步揭示其生效机制, 结合图4进行深入分析。图4展示了聚焦参数 $\gamma=2$ 时的权重曲线与正负样本置信度密度的叠加关系。当预测置信度 p 较低时, 损失权重显著增大, 使模型聚焦于稀有触发词和边界模糊的论元; 同时, 由于负样本密度主要集中在低置信度区域, 焦点损失赋予其更低的权重占比, 从而有效缓解了负类淹没效应。这一可视化结果与消融实验中性能下降的结论相互印证, 充分说明了该模块通过优化梯度分布提升了模型对长尾样本的识别能力。

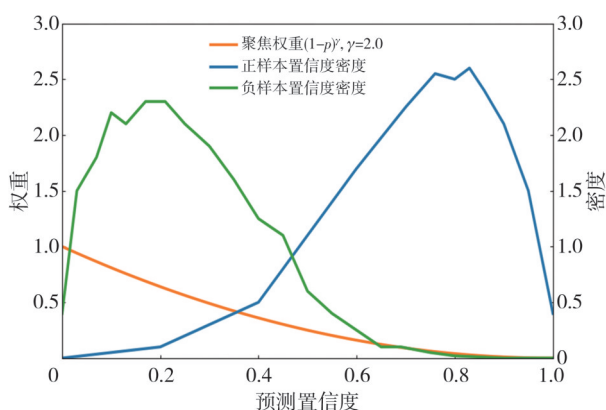


图4 聚焦损失权重与样本置信度分布对比

Fig. 4 Comparison of focal loss weight and sample confidence density distribution

2) 求和门控事件融合层的有效性。将改进的求和式门控机制替换为传统的特征拼接后, TI-F1出现了1.0个百分点的降幅。传统的拼接策略虽然保留了完整特征, 但同时也引入了大量冗余噪声, 导致单阶段模型在词对关系预测时产生语义干扰。而本文提出的SGEF通过求和运算实现了特征空间的深度交互与去冗余, 使得事件特征与上下文表示在相同子空间内进行干涉, 显著提升了触发词检测的准确性。

3) Dropout策略的增益。在自适应融合层中移除Dropout导致各指标平均下降约0.6个百分点, 说明在FewFC这种特定领域的事件抽取任务中, 引入适度的正则化能够有效抑制模型对特定触发词模式的过度拟合, 增强了模型在面对非典型表述时的泛化能力。

综上所述, 本文提出的各项改进机制均表现出正向增益。其中, 损失函数的优化对突破复杂重叠与嵌套场景下的分类瓶颈贡献最为显著。

5 讨论

5.1 实验结果分析

本文在FewFC数据集上开展了全面的实验验证, 旨在评估基于OneEE改进模型在触发词和论元抽取任务中的性能。结果表明, 模型在触发词识别、触发词分类、论元识别以及论元分类任务上均得到了提升。特别是, 如表2所示, 相比于单阶段核心基线模型OneEE, 改进模型在TC和AC任务中的F1分数分别提高了0.7个百分点和1.6个百分点, 这得益于焦点损失和门控优化的协同作用。实验采用了3次随机种子运行取均值的评估策略, 确保了结果的统计鲁棒性, 缓解了数据分布不均对性能评估的潜在偏差。值得注意的是, 模型在处理稀有触发词和嵌套论元时表现尤为突出, 这与类不平衡问题的缓解密切相关, 验证了所提方法的有效性。

5.2 优化策略有效性分析

本文提出的架构改进在复杂事件抽取中表现出了显著的优势, 其有效性主要归因于多种机制的协同作用。其中, SGEF层的去冗余机制至关重要。实验表明, 求和运算替代传统的线性拼接策略, 不仅降低了特征维度, 还通过门控信号实现了对无效上下文噪声的有效抑制。这种特征空间的重构使模型在处理重叠事件时, 能够更清晰地捕捉触发词与论元之间的逻辑关联。正则化策略增强了泛化能力。在门控融合层引入Dropout技术, 有效阻断了神经元之间的共适应关系, 降低了模型对特定中文金融或医疗文本模式的过拟合风险, 从而提升了在测试集上的鲁棒性。

为验证模型在抑制级联误差与特征冗余方面的优势, 本文选取FewFC数据集中的典型重叠事件样本进行案例分析。例句为: “公司A拟收购公司B的子公司C, 交易金额达5亿元”, 该句包含“投资”与“交易”两个重叠事件, 且共享论元“公司A”。基线模型OneEE虽然识别了投资事件, 但在交易事件中漏检了远距离论元“公司A”。这是因为并行处理机制无法有效应对多实体干扰, 导致关键特征淹没于上下文噪声中。相比之下, 本模型成功提取了所有

目标论元。其性能提升主要得益于两点:首先,SGEF 层的求和门控机制动态抑制了无关干扰,显著突出了“公司 A”的特征显著性;其次,结构化焦点损失赋予此类长距离难样本更高的权重,促使模型在反向传播中精准捕捉论元边界。

为验证多头词对标注器的有效性,本文深入分析了注意力权重分布情况。观察发现,MHPT 的不同头部呈现出显著的功能分化特征。Head-2 的高响应区域集中于矩阵对角线附近,体现了对局部邻域特征的强烈关注。这一特性与跨度边界识别任务高度契合。相比之下,Head-5 在长距离词对上表现出的激活强度,准确捕捉了触发词与远距离论元的语义关联。这证实了该头部主要负责角色关系的全局依赖建模。这种多头互补机制有效提升了模型在复杂重叠场景下的结构解析能力。

这一案例表明,本文提出的改进策略显著增强了模型在复杂重叠场景下的结构解析能力。

5.3 局限性分析

尽管改进模型在整体性能上取得了突破,但在处理极端长尾分布时仍存在瓶颈。通过对测试集预测结果的统计观察发现,对于训练集中出现频率极低的特定中文金融事件类别,模型的召回率提升幅度有限。这表明,虽然结构化焦点损失缓解了梯度消失问题,但固定的超参数配置(γ 和 α)可能无法完美适配所有类别的学习难度,模型在零样本或少样本场景下的泛化能力仍有待进一步增强。

6 结 论

针对单阶段事件抽取中重叠与嵌套结构的识别难题,本文提出了一种结合结构化焦点损失与门控优化的改进框架。该框架旨在解决特征冗余与级联误差传播问题。在理论层面,本文构建了任务依赖的梯度约束机制,通过层级化损失加权在上游触发词与下游论元间建立逻辑壁垒,有效阻断了误差传播。在特征层面,SGEF 模块利用门控信号实现特征空间的内生性去噪,降低了无关上下文干扰。实验结果表明,该模型在 FewFC 与 Genia 数据集上的表现优于 OneEE 基准。相比生成式及大模型辅助方法,本模型在维持高并行推理效率的同时,展现出更强的语义解析鲁棒性,证实了强结构化约束判别式模型在中文金

融、医疗等复杂场景中仍具备独特应用价值。

本文在解决样本不平衡与特征冗余方面取得进展,未来研究将聚焦于以下三个方向:1) 超参数的自适应优化。针对当前焦点损失参数需人工固定的局限,拟利用元学习或贝叶斯优化算法,实现对不同长尾分布样本难度的动态感知与参数自适应。2) 跨范式知识增强。鉴于大语言模型的泛化能力,计划研究“生成式-判别式”协同框架,通过知识蒸馏将通用语义推理能力注入轻量化模型,提升低资源场景下的抽取性能。3) 深层结构化建模。针对文档级或跨句挑战,拟引入图神经网络显式建模长距离论元依赖,进一步突破局部上下文窗口的性能瓶颈。

利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors declare no relevant conflict of interests.

参考文献:

- [1] LIU J, MIN L, HUANG X. An overview of event extraction and its applications[DB/OL]. (2021-11-05) [2025-10-13]. <https://arxiv.org/abs/2111.03212>.
- [2] 张斌龙, 雷海卫, 李成奇, 等. 基于关系标签语义与全局特征融合的专利实体关系抽取[J]. 中北大学学报(自然科学版), 2025, 46(5): 611-621.
ZHANG Binlong, LEI Haiwei, LI Chengqi, et al. Patent entity relation extraction based on the semantics of relation labels and global feature integration[J]. Journal of North University of China (Natural Science Edition), 2025, 46(5): 611-621. (in Chinese)
- [3] GAO J, YU C, WANG W, et al. Mask-then-fill: A flexible and effective data augmentation framework for event extraction[C]//Findings of the Association for Computational Linguistics: EMNLP 2022, 2022: 4537-4544.
- [4] LI Q, LI J, SHENG J, et al. A survey on deep learning event extraction: Approaches and applications[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(5): 6301-6321.
- [5] LIU X, HUANG H, SHI G, et al. Dynamic prefix-tuning for generative template-based event extraction [C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022: 5216-5228.
- [6] 万齐智, 万常选, 胡蓉, 等. 面向研究问题的深度学习事件抽取综述[J]. 自动化学报, 2024, 50(11): 2079-

2101.
WAN Qizhi, WAN Changxuan, HU Rong, et al. Event extraction based on deep learning: A survey of research issue[J]. *Acta Automatica Sinica*, 2024, 50(11): 2079-2101. (in Chinese)
- [7] 胡瑞娟, 周会娟, 刘海砚, 等. 基于深度学习的篇章级事件抽取研究综述[J]. *计算机工程与应用*, 2022, 58(24): 47-60.
HU Ruijuan, ZHOU Huijuan, LIU Haiyan, et al. Survey on document-level event extraction based on deep learning[J]. *Computer Engineering and Applications*, 2022, 58(24): 47-60. (in Chinese)
- [8] CHEN Y, XU L, LIU K, et al. Event extraction *via* dynamic multi-pooling convolutional neural networks [C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2015: 167-176.
- [9] TANEV H, PISKORSKI J, ATKINSON M. Real-time news event extraction for global crisis monitoring [C]//Natural Language and Information Systems. Berlin, Heidelberg, 2008: 207-218.
- [10] SAKAKI T, MATSUO Y, YANAGIHARA T, et al. Real-time event extraction for driving information from social sensors[C]//2012 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), 2012: 221-226.
- [11] SHENG J, GUO S, YU B, et al. CasEE: A joint learning framework with cascade decoding for overlapping event extraction[DB/OL]. (2021-07-04) [2025-10-13]. <https://arxiv.org/abs/2107.01583>.
- [12] ZHENG S, CAO W, XU W, et al. Doc2EDAG: An end-to-end document-level framework for Chinese financial event extraction [DB/OL]. (2019-09-23) [2025-10-13]. <https://arxiv.org/abs/1904.07535>.
- [13] CAO H, LI J, SU F, et al. OneEE: A one-stage framework for fast overlapping and nested event extraction [DB/OL]. (2022-09-06) [2025-10-13]. <https://arxiv.org/abs/2209.02693v1>.
- [14] HSU I H, HUANG K H, BOSCH E, et al. DEGREE: A data-efficient generation-based event extraction model[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2022: 1890-1908.
- [15] WANG X, LI S, JI H. Code4Struct: Code generation for few-shot event structure prediction[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023: 3640-3663.
- [16] NGUYEN T H, CHO K, GRISHMAN R. Joint event extraction *via* recurrent neural networks[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016: 300-309.
- [17] NGUYEN T, GRISHMAN R. Graph convolutional networks with argument-aware pooling for event detection[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, 32(1): 5900-5907.
- [18] CHEN Y, XU L, LIU K, et al. Event extraction *via* dynamic multi-pooling convolutional neural networks [C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2015: 167-176.
- [19] YANG S, FENG D, QIAO L, et al. Exploring pre-trained language models for event extraction and generation[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 5284-5294.
- [20] XU N, XIE H, ZHAO D. A novel joint framework for multiple Chinese events extraction [C]//Proceedings of the 19th Chinese National Conference on Computational Linguistics, 2020: 950-961.
- [21] LIU X, HUANG H, SHI G, et al. Dynamic prefix-tuning for generative template-based event extraction [C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022: 5216-5228.
- [22] GAO J, YU C, WANG W, et al. Mask-then-fill: A flexible and effective data augmentation framework for event extraction [C]//Findings of the Association for Computational Linguistics: EMNLP 2022, 2022: 4537-4544.
- [23] ZHU M, ZENG K, JIBINGWU J, et al. LC4EE: LLMs as good corrector for event extraction [C]//Findings of the Association for Computational Linguistics: ACL 2024, 2024: 12028-12038.
- [24] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]//2017 IEEE International Conference on Computer Vision (ICCV), 2017: 2999-3007.
- [25] HENNING S, BELUCH W, FRASER A, et al. A survey of methods for addressing class imbalance in deep-learning based natural language processing [C]//Proceedings of the 17th Conference of the European

- Chapter of the Association for Computational Linguistics, 2023: 523-540.
- [26] SU Z, LIN Z, BAIXUE B, et al. MiLe loss: A new loss for mitigating the bias of learning difficulties in generative language models[C]//Findings of the Association for Computational Linguistics: NAACL 2024, 2024: 250-262.
- [27] MA M D, TAYLOR A, WANG W, et al. DICE: Data-efficient clinical event extraction with generative models[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023: 15898-15917.
- [28] ZHOU Y, CHEN Y, ZHAO J, et al. What the role is *vs.* what plays the role: Semi-supervised event argument extraction *via* dual question answering[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(16): 14638-14646.
- [29] KIM J D, WANG Y, TAKAGI T, et al. Overview of genia event task in bionlp shared task 2011 [C]//Proceedings of BioNLP Shared Task 2011 Workshop, 2011: 7-15.
- [30] DU X, CARDIE C. Event extraction by answering (almost) natural questions [DB/OL]. (2021-02-04) [2025-10-13]. <https://arxiv.org/abs/2004.13625>.
- [31] SHENG J, GUO S, YU B, et al. CasEE: A joint learning framework with cascade decoding for overlapping event extraction[DB/OL]. (2021-07-04) [2025-10-13]. <https://arxiv.org/abs/2107.01583>.

(上接第 394 页)

- [21] DING X, ZHANG X, MA N, et al. RepVGG: Making VGG-style ConvNets great again [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 13728-13737.
- [22] JHA D, SMEDSRUD P H, RIEGLER M A, et al. Kvasir-SEG: A segmented polyp dataset [C]//Multi-Media Modeling. Cham: Springer, 2020: 451-462.
- [23] BERNAL J, SÁNCHEZ J, VILARIÑO F. Towards automatic polyp detection with a polyp appearance model [J]. Pattern Recognition, 2012, 45(9): 3166-3182.
- [24] BERNAL J, SÁNCHEZ F J, FERNÁNDEZ-ESPARRACH G, et al. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation *vs.* saliency maps from physicians [J]. Computerized Medical Imaging and Graphics, 2015, 43: 99-111.
- [25] SILVA J, HISTACE A, ROMAIN O, et al. Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer [J]. International Journal of Computer Assisted Radiology and Surgery, 2014, 9(2): 283-293.
- [26] WANG J, HUANG Q, TANG F, et al. Stepwise feature fusion: Local guides global [C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2022. Cham: Springer, 2022: 110-120.
- [27] JHA D, SMEDSRUD P H, RIEGLER M A, et al. ResUNet++: An advanced architecture for medical image segmentation [C]//2019 IEEE International Symposium on Multimedia (ISM), 2020: 225-2255.
- [28] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]//Computer Vision-ECCV 2018. Cham: Springer, 2018: 833-851.
- [29] HUANG C H, WU H Y, LIN Y L. HarDNet-MSEG: A simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 FPS[DB/OL]. (2021-01-20) [2025-09-25]. <https://arxiv.org/abs/2101.07172>.
- [30] QIN X, ZHANG Z, HUANG C, et al. U²-Net: Going deeper with nested U-structure for salient object detection [J]. Pattern Recognition, 2020, 106: 107404.