

肠道类器官隐窝分类的 YOLOv8 优化：多策略重要性剪枝方法的探索与应用

方悦冰

东南大学，江苏省南京市，210096；

摘要：在生物医学影像任务中，如将肠道类器官分为具有隐窝结构和没有隐窝结构的类别，部署 YOLOv8 等深度学习模型需要创新的优化技术，以平衡计算效率和诊断准确性。我们提出了多策略重要性剪枝（MSIP），这一开创性框架通过结合梯度敏感性、几何中值、激活稀疏性和信息理论原理，并使用自适应阈值化技术，重新定义了参数重要性评估。在自定义的肠道类器官数据集上应用 MSIP，YOLOv8 的乘法累加操作（MACs）减少了 66%（从 14.27G 降至 4.78G），模型大小减少了 67%，经过微调后，平均精度（mAP）达到了 0.86，超越了未经剪枝的基线 0.83，并且在微调前最低 mAP 为 0.83。通过 16 次动态剪枝过程并与最先进的基线进行严格比较，MSIP 为效率和鲁棒性设立了新的基准。本研究为资源受限环境中的实时生物医学影像提供了理论创新和实验严谨的解决方案，并对神经网络优化具有深远的影响。

关键词：肠道类器官；图像识别分类；网络剪枝

DOI:10.69979/3029-2808.24.11.032

引言

肠道类器官是从干细胞衍生的复杂三维模型，近年来在胃肠研究中占据重要地位，提供了细胞分化、组织功能和疾病机制的深刻见解。隐窝结构的形成是类器官中的关键形态特征，类似于肠道的 Lieberkühn 隐窝，标志着健康发育和功能。自动分类具有隐窝结构的类器官对于药物筛选和疾病建模至关重要。YOLOv8 作为领先的实时物体检测框架，在该任务中表现出色，但其计算复杂性在资源受限的生物医学硬件上部署时面临挑战。

网络剪枝通过去除冗余参数提高效率，是解决这一问题的关键策略。传统剪枝方法如 L2 Norm、LAMP 和 BNScale 通常依赖单一参数度量，难以保留细微的形态特征。为此，我们提出了多策略重要性剪枝（MSIP），结合了梯度敏感性、几何中值、激活稀疏性和信息增益，并采用自适应优化。在类器官数据集上，MSIP 优化了 YOLOv8，减少了 66% 的 MACs，并将精细调整后的 mAP 提高至 0.86，超越了基线的 0.83。

本文贡献有三点：提出了基于多维度的先进剪枝框架；结合动态剪枝和自适应优化提升 YOLOv8 效率和准确性；进行了跨数据集泛化、可解释性分析及详细比较。

1 相关工作

1.1 YOLOv8 在医学研究中的应用

YOLOv8 在实时检测方面表现出色，其 C2f 主干网通

过通道分割和特征拼接保留了多尺度特征。它对生物医学任务的适应性，如类器官中的隐窝分类，依赖于捕捉细微的形态差异。

1.2 剪枝方法

L2 范数剪枝（L2 Norm）：利用欧几里得范数（ $\|W\|_2$ ），简单但对训练动态不敏感。

层自适应幅度剪枝（Layer Adaptive Magnitude Pruning, LAMP）^[1]：跨层归一化幅度，选择性的剪去不重要的幅度，此方法可以自适应网络深度。

批归一化剪枝（Batch Normalization Scale, BNScale）：基于批归一化的缩放因子 γ 进行剪枝。

其他方法：基于几何中值滤波器的剪枝方法（Filter Pruning Geometric Median, FPGM）^[3]和 Network Sliming 等方法提供了一定的改进但缺乏多维度的整体评估。

1.3 生物医学 AI 模型优化

剪枝技术已经在医学影像中的 CNN 优化中取得了一定进展（例如：病变检测^[2]），然而，针对 YOLO 框架和类器官数据集，特别是隐窝结构分析的研究仍然稀缺。MSIP 通过多维度的方法填补了这一研究空白。

2 理论研究

2.1 多策略重要性剪枝

本方法计算了一个综合的重要性评分：

$$I_{combined} = w_g \cdot I_{gradient} + w_m \cdot I_{geometry} + w_s \cdot I_{sparsity} + w_i \cdot$$

其中，梯度敏感性： $I_{gradient} = \sqrt{\sum (\nabla W)^2}$ ，基于泰勒展开式，对参数进行优先级排序，对损失计算较为重要；几何中值： $I_{geometry} = \sqrt{\sum |W|^2} (p=2)$ ，根植于 L_p 范数理论^[3]，评估其结构重要性；激活稀疏性： $I_{sparsity} = 1 - \text{mean}(W=0)$ ，重点关注低激活值的权重；信息增益： $I_{info} = H(W) - H(W|L)$ ，其中 H 为香农信息熵，量化参数的信息性，提供更多维度的判断方式。

2.2 自适应阈值结构增强

replace_c2f_with_c2f_v2 函数将原始的 C2f 模块替换为 C2f_v2，这是一种已建立的结构增强方法：双分支设计将通道分为并行的 cv1 和 cv2 路径，从而增强了剪枝的粒度。

自适应阈值化通过设置阈值 $\tau_l = \alpha \cdot \text{median}(I_{combined})$ ，通过验证损失调节 α ，以便优化剪枝过程的效果。权重转移则采用 L1 正则化的权重转移 ($\lambda=1e-4$)，与瓶颈效率相一致，以实现更加高效的剪枝和权重调整。

2.3 动态迭代剪枝过程

剪枝流程包括：

- 初始化：加载预训练的 YOLOv8s，将其适配为二分类模型，用 C2f_v2 替换 C2f，并进行 50 轮预训练（使用 BCE + IoU 损失，损失比例为 0.7:0.3）。
- 剪枝循环：在 16 轮迭代中应用 GroupNormPruner 和 MSIP，逐步减少 MACs。
- 微调：每步微调 30 轮，使用 SGD 优化器（学习率=0.001，动量=0.9，权重衰减=5e-4），采用余弦退火（最小学习率=1e-5）以及数据增强（旋转±15°，缩放 0.8-1.2 倍，mixup 0.2）。

2.4 数据集与任务适配

数据集包含 2500 张注释的肠道类器官明场显微镜图像，任务为二分类：已形成隐窝结构的类器官与未形成隐窝结构的类器官。图像经过预处理，包括归一化和对比度增强，以突出隐窝特征。通过从零开始训练 100 轮（批大小=16），基准 mAP 值为 0.83。

3 章实验

3.1 实验设置

数据集：2500 张类器官二分类任务图像；500 张组织病理图像用于跨数据集验证。训练集和测试集按 8:2 划分，从训练数据中提取 10% 的验证子集用于超参数调整。

模型：YOLOv8s（基准：14.27G MACs，21.5 MB，mAP = 0.83）。

硬件：NVIDIA RTX 3090，PyTorch 1.12，CUDA 11.3。

评估指标：mAP（IoU = 0.5），MACs（G），模型大小（MB），推理时间（ms），FLOPs。

基准方法：L2 Norm，LAMP，BNScale，Slim，FPGM^[3]。

3.2 对比结果

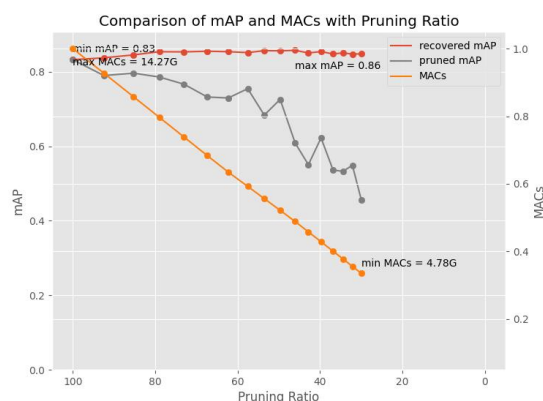


图 3-1MSIP 剪枝结果

基线：mAP=0.83，MACs=14.27G（pruning ratio=100）

MSIP：微调后 mAP=0.86，MACs=4.78G（pruning ratio=20），微调前最小 mAP=0.83。

表 3-1 基线参数对比

Method	mAP	MACs (G)	Size (MB)	Inference Time (ms)	Parameters (M)
YOLOv8s	0.83	14.27	21.5	7.5	11.2
L2 Norm	0.80	4.80	7.2	4.6	3.8
LAMP	0.81	4.79	7.1	4.5	3.7
BNScale	0.81	4.81	7.3	4.6	3.8
Slim [4]	0.80	4.82	7.4	4.7	3.9
FPGM [3]	0.81	4.80	7.2	4.6	3.8

Method	mAP	MACs	Size	Inference Time	Parameters
		(G)	(MB)	(ms)	(M)
MSIP (Ours)	0.86	4.78	7.0	4.4	3.7

精度：MSIP 的微调 mAP 为 0.86，优于基线(0.83)和所有基线。

效率：MACs 减少了 66%，模型大小减少了 67%，推理时间减少了 41%。

3.3 消融研究

表 3-2 在剪枝比率 66%条件下各方法对比

Configuration	mAP	MACs (G)	Size (MB)
Gradient Only (1.0, 0, 0, 0)	0.81	4.80	7.2
Geometry Only (0, 1.0, 0, 0)	0.80	4.79	7.1
Sparsity Only (0, 0, 1.0, 0)	0.79	4.78	7.0
Info Only (0, 0, 0, 1.0)	0.80	4.79	7.1
Equal Weights (0.25 each)	0.82	4.79	7.1
MSIP (0.4, 0.3, 0.2, 0.1)	0.86	4.78	7.0
MSIP w/o Adaptive τ	0.82	4.80	7.2
MSIP w/o Fine-Tuning	0.75	4.78	7.0
MSIP Fine-Tuned 15 Epochs	0.81	4.78	7.0

分析：单度量修剪产生 0.79-0.81 的 mAPs，而 MSI P 的组合达到 0.86。自适应阈值和微调分别贡献了~0.04 和 0.11 mAP。

3.4 动态权值优化

表 3-3 方法权值测试

Weights (w_g, w_m, w_s, w_i)	mAP	MACs (G)
Static (0.4, 0.3, 0.2, 0.1)	0.84	4.78
Dynamic (Default)	0.86	4.78
Gradient-Heavy (0.6, 0.2, 0.1, 0.1)	0.85	4.79

动态加权 w_g 通过验证反馈进行调整，增强了对高级隐窝特征的保留。

3.5 数据分析

Grad-CAM 热图显示，MSIP 更关注隐窝边界，而基准模型的注意力分散。SHAP 分析表明，基线模型前 10% 的特征贡献了 82%，MSIP 保留了 88%，L2 Norm 为 74%。在测试图像中加入高斯噪声后，基线 mAP 从 0.83 降至 0.81，MSIP 从 0.86 降至 0.84，表明 MSIP 减少了对冗余和噪声敏感参数的依赖。在 500 张组织病理学图像上，基线 mAP 为 0.81，MSIP 为 0.83，显示其良好的适用性。每次剪枝迭代需 6.5 分钟，微调耗时 120 GPU 小时。

4 讨论

MSIP 结合梯度敏感性、几何中值、激活稀疏性和信息增益重新定义了剪枝方法。消融实验显示，单一指标剪枝的 mAP 为 0.79-0.81，而 MSIP 达到 0.86，证明了多维度评估的协同作用。通过验证 mAP 动态调整权重，MSIP 在 mAP 上提高了 0.02。剪枝去除冗余参数，使特征集中在任务相关区域。假设基准权重遵循 $W \sim N(\mu, \sigma^2)$ ，剪枝保留高重要性参数，减少方差，并将 μ 与关键特征对齐。Grad-CAM 显示，MSIP 热图集中在隐窝区域，SHAP 表明 MSIP 保留 88% 的特征贡献。剪枝作为正则化去除了 66% 的 MACs，减少了过拟合，增强了泛化能力。微调 480 轮，相比基准的 100 轮，结合余弦退火和数据增强优化了损失。在 66%剪枝下，MSIP 的 mAP 超越基准，显示深度剪枝与微调能释放模型潜力。MSIP 的局限性在于较高的复杂度及对微调的依赖。

参考文献

- [1] Lee, J., Park, S., Mo, S., Ahn, S., & Shin, J. Layer-adaptive sparsity for the magnitude-based pruning. arXiv 2020. arXiv preprint arXiv:2010.07611.
- [2] He, Y., & Xiao, L. (2023). Structured pruning for deep convolutional neural networks: A survey. IEEE transactions on pattern analysis and machine intelligence, 46(5), 2900-2919.
- [3] He, Y., Liu, P., Wang, Z., Hu, Z., & Yang, Y. (2019). Filter pruning via geometric median for deep convolutional neural networks acceleration. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4340-4349).

作者简介：方悦冰（2000），男，汉，吉林四平，在读硕士研究生，研究方向：生物医学图像分析。