

可微分神经网络架构搜索综述

逯鹏^{1,2,3}, 李科研^{1,2}, 张宏坡^{3,4}, 陈立伟¹, 武家辉^{1,2}, 刘帅兵^{1,2}

(1. 郑州大学 电气与信息工程学院, 河南 郑州 450001; 2. 机器人感知与控制河南省工程实验室, 河南 郑州 450001; 3. 互联网医疗与健康服务河南省协同创新中心, 河南 郑州 450052; 4. 郑州大学 网络管理中心, 河南 郑州 450001)

摘要:神经网络架构搜索(neural architecture search, NAS)是深度学习领域的交叉研究,旨在自动化地设计神经网络结构。NAS需要反复训练和评估大量候选网络,计算代价高。可微分神经网络架构搜索(differentiable neural architecture search, DNAS)将离散架构搜索问题转化为可微的连续优化问题,降低了计算代价。首先从搜索空间、搜索策略和性能评估策略3个方面构建了可微分神经网络架构搜索算法框架;其次,分析、对比和总结了其参数化操作的性能估计偏差、架构过拟合与搜索稳定性问题以及优化搜索空间与提升效率问题的改进策略;然后,比较分析了典型DNAS算法在图像分类数据集上的错误率、参数量以及搜索消耗时间与实验硬件条件;最后指出DNAS在边缘设备部署、医学信号分析和跨模态匹配等复杂场景中的应用潜力,并提出面向多目标优化、任务驱动搜索空间设计及跨任务迁移复用的未来研究方向。

关键词:神经网络架构搜索;深度学习;可微分神经网络架构搜索;连续优化;性能估计

中图分类号: TP181 **文献标志码:** A **doi:** 10.13705/j.issn.1671-6833.2026.02.007

深度学习模型性能高度依赖其网络架构与参数权重的协同优化,模型的训练会面临样本数量和质量、计算和内存代价、过拟合和欠拟合^[1]、梯度消失或组合爆炸问题^[2]。对复杂模型的网络架构参数和对应的超参数进行优化是解决上述问题的途径之一。然而通过人工反复调整或常规的随机搜索与网格搜索效率非常低, NAS是自动寻优网络架构的新途径。

NAS通过将网络结构设计转化为优化问题,利用算法自动寻优,生成针对特定任务的高性能网络架构。NAS通常通过强化学习与进化算法等黑盒优化方法,搜索耗时和计算成本高是其瓶颈问题,尤其是在边缘设备等资源受限的环境中^[3]。

可微分神经网络架构搜索(DNAS)以可微分神经网络架构搜索算法(differentiable architecture search, DARTS)^[4]为代表,通过连续松弛策略重构搜索空间,将离散加架构选择转化为可微分优化问题,使搜索过程可通过梯度下降优化,搜索效率上取得重要突破。

以当前DARTS面临的问题与挑战为切入点,首

先从DARTS算法框架角度分析其面临的核心问题及原因;其次,针对其参数化操作的性能估计偏差、架构过拟合与搜索稳定性问题、优化搜索空间与提升效率问题的改进策略进行系统性地分析、阐述和归纳,并对典型DNAS算法进行对比;最后,提出DNAS在更复杂场景中的应用潜力及发展方向。

1 可微分神经网络架构搜索算法

DARTS的搜索空间被称为超网^[5],基于NAS-Net^[6]的模块化结构构建,由重复堆叠的正常单元(Normal Cell)和还原单元(Reduction Cell)组成,分别用于保持特征图尺寸和实现下采样,从而在提升搜索效率的同时压缩搜索空间。每个单元内部为一个包含N个节点的有向无环图(Directed Acyclic Graph, DAG),如图1所示,DAG中节点的具体计算方法定义为

$$x^{(j)} = \sum_{i < j} o^{(i,j)}(x^{(i)})。 \quad (1)$$

式中: $x^{(i)}$ 和 $x^{(j)}$ 分别代表第*i*和第*j*个节点对应的潜在特征表示; $o^{(i,j)}$ 代表第*i*个节点到第*j*个节点采

收稿日期:2026-02-10;修订日期:2026-04-03

基金项目:国家自然科学基金资助项目(62373330);河南省高等学校重点科研项目(25A520022)

作者简介:逯鹏(1974—),男,河南郑州人,郑州大学教授,博士,主要从事人工智能与大数据分析、康复机器人和视觉计算,E-mail: lupeng@zzu.edu.cn。

用的操作。

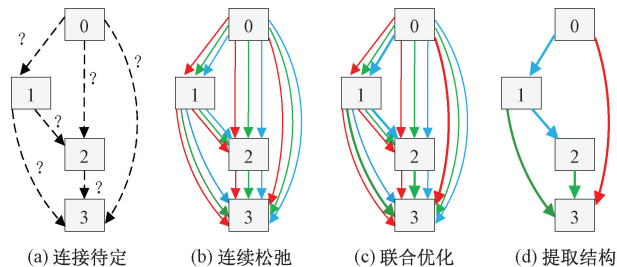


图1 DARTS 算法流程图

Figure 1 DARTS algorithm flowchart

为将离散的搜索空间变成连续松弛化, DARTS 采用了基于 softmax 函数的连续松弛化策略:

$$\bar{o}^{(i,j)}(x) = \sum_{o \in \mathcal{O}} \frac{\exp(\alpha_o^{(i,j)})}{\sum_{o' \in \mathcal{O}} \exp(\alpha_{o'}^{(i,j)})} o(x). \quad (2)$$

式中: $\bar{o}^{(i,j)}$ 为操作混合权重; $o(x)$ 代表操作类型; $\alpha_o^{(i,j)}$ 为第 i 个节点到第 j 个节点之间操作的权重, 代表之后需要搜索的架构参数。

将搜索空间连续化后, DARTS 将架构搜索转化为优化架构参数 α 与网络权重 w 的双层优化问题: 在训练集上优化权重, 在验证集上优化架构参数, 以提升模型泛化能力。该过程可形式化为一个双层优化问题, 用以下公式表示:

$$\begin{cases} \min_{\alpha} \mathcal{L}_{\text{val}}(w^*(\alpha), \alpha); \\ \text{s. t. } w^*(\alpha) = \operatorname{argmin}_w \mathcal{L}_{\text{train}}(w, \alpha). \end{cases} \quad (3)$$

式中: \mathcal{L}_{val} 和 $\mathcal{L}_{\text{train}}$ 分别表示验证和训练损失。DARTS 训练的目的就是从所有边中找到贡献最大的边, 并将其保留, 即以下公式:

$$o^{(i,j)} = \operatorname{argmax}_{o \in \mathcal{O}} \alpha_o^{(i,j)}. \quad (4)$$

内层优化网络权重需要消耗大量资源, 因此在实际求解中提出了一个近似方案:

$$\nabla_{\alpha} \mathcal{L}_{\text{val}}(w^*(\alpha), \alpha) \approx \nabla_{\alpha} \mathcal{L}_{\text{val}}(w - \xi \nabla_w \mathcal{L}_{\text{train}}(w, \alpha), \alpha). \quad (5)$$

式中: w 为当前网络权重; ξ 为迭代优化过程的学习率; $w^*(\alpha)$ 为最优网络权重, 在训练集上对网络权重执行一次梯度下降得到, 无需通过训练直到收敛来解决公式(3)的内层优化问题。将链式法则用于公式(5)得到:

$$\nabla_{\alpha} \mathcal{L}_{\text{val}}(w', \alpha) - \xi \nabla_{\alpha, w}^2 \mathcal{L}_{\text{train}}(w, \alpha) \nabla_w \mathcal{L}_{\text{val}}(w', \alpha). \quad (6)$$

为降低二阶求导大量的计算资源问题, 利用有限差分近似方法, 公式(6)的第二项可近似为

$$\frac{\nabla_{\alpha, w}^2 \mathcal{L}_{\text{train}}(w, \alpha) \nabla_w \mathcal{L}_{\text{val}}(w', \alpha) \approx \nabla_{\alpha} \mathcal{L}_{\text{train}}(w^+, \alpha) - \nabla_{\alpha} \mathcal{L}_{\text{train}}(w^-, \alpha)}{2\epsilon}. \quad (7)$$

将多个二次梯度转化为多个一次梯度, 只需对网络权重 w 进行两次前向传递, 对架构参数 α 进行

两次反向传递, 复杂度从 $O(|\alpha||w|)$ 优化到 $O(|\alpha| + |w|)$ 。当 $\xi=0$ 时, 公式(6)的二阶导数将会消失, 此时计算速度会大大加快, 但模型梯度计算准确率会下降。将 $\xi=0$ 的情况称为一阶近似, $\xi>0$ 的情况称为二阶近似。

DARTS 采用 One-shot 评估方法, 通过构建包含所有候选操作的超级网络, 在其中搜索子网络架构。子网络共享超网权重实现高效评估, 避免重复训练。如图2所示, 搜索阶段使用8个Cell构建浅层模型, 评估阶段扩展至20个Cell。

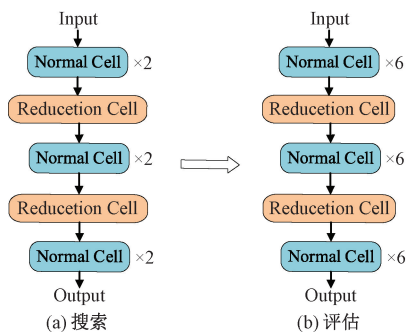


图2 搜索评估图

Figure 2 search evaluation map

2 参数化操作的性能估计偏差

DARTS 利用连续松弛和梯度近似优化架构参数, 提升搜索效率的同时也导致性能估计偏差, 表现为梯度方向偏离最优解、架构参数对真实精度的反映不准确、搜索与评估阶段存在失配等问题。围绕降低梯度近似偏差、准确评估操作重要性和缩小优化差距等方面已发展出多种改进策略以提升搜索质量与稳定性。

2.1 降低梯度近似偏差

梯度近似偏差会导致次优解, 现有方法分为两类, 优化目标改进和优化过程调整。在优化目标改进方面, MiLeNAS^[7] 通过构建混合级优化公式(8)来平衡训练和验证目标损失, 减轻梯度偏差:

$$\min_{\alpha, w} [\mathcal{L}_{\text{tr}}(w^*(\alpha), \alpha) + \lambda \mathcal{L}_{\text{val}}(w^*(\alpha), \alpha)]. \quad (8)$$

式中: λ 为调节系数, 用于平衡训练损失和验证损失的重要性。Fair DARTS^[8] 则采用 sigmoid 替代 softmax, 并引入零一损失 L_{0-1} 将架构权重的 sigmoid 值推向 0 或 1 来最小化离散化间隙。 L_{0-1} 公式如下:

$$L_{0-1} = -\frac{1}{N} \sum_i (\sigma(\alpha_i) - 0.5)^2. \quad (9)$$

式中: σ 代表 sigmoid 函数; α_i 为架构参数。

在优化过程调整方面, 温度退火方法 TA-DARTS^[9] 通过动态调整 softmax 温度来优化离散化过

程。DE-DARTS^[10]设计动态力注意网络(dynamic attention network, DAN),根据输入数据自适应调整架构权重。DE-DARTS算法如图3所示,动态注意力网络在DAG的每个边缘上实现,将DAN的输出乘以注意力权重 σ ,然后再将其添加到架构参数 α 中。

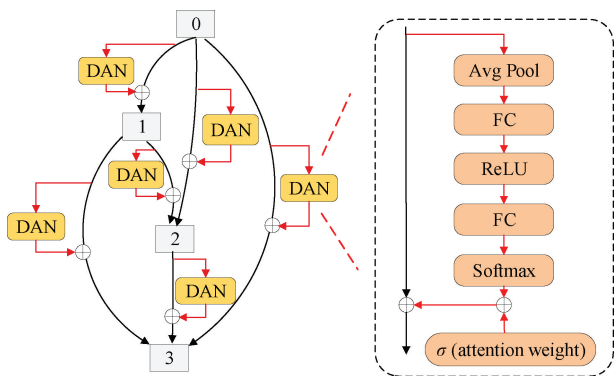


图3 DE-DARTS 算法示意图

Figure 3 Schematic diagram of DE-DARTS algorithm

MR-DARTS^[11]和DARTS-CGW^[12]算法在优化过程中调整梯度,前者采用动量递归估计器提升二阶导数精度,后者则通过通道重要性加权抑制噪声梯度。然而,目标改进类方法需额外调参,过程调整策略在大规模任务中的通用性仍待验证。

2.2 准确评估架构参数重要性

针对架构参数在操作重要性评估中的局限性,已有多种基于扰动的评估机制被提出,用以提升架构选择的可靠性与稳定性。DARTS-PT^[13]通过测量操作对模型精度的扰动幅度来评估其贡献;DARTS-PT-CORE^[14]在此基础上引入协作正则化以减小评估噪声;DARTS-EAST^[15]则结合网络拓扑信息,根据屏蔽边对性能的影响自适应调整离散顺序并优化结构选择。尽管以上方法提升了评估精度,但由于依赖单次扰动度量,仍易受搜索初期梯度不稳定性的干扰,进而带来离散结果的波动。

IS-DARTS^[16]和Shapley-NAS^[17]通过统计操作贡献来评估重要性。IS-DARTS在离散化阶段引入了多轮采样与扰动评估机制,对每条候选边在屏蔽操作前后验证集精度变化进行精确测量,以稳定方式量化操作贡献并确定最优操作。同样的,Shapley-NAS建议评估操作对验证准确性的直接影响,根据所有可能的操作子集与没有给定操作的对应子集的验证精度差异,直接评估操作对任务性能的边际贡献。

OStr-DARTS^[18]设计了一种基于操作强度的新评价指标,直接测量候选操作对损失函数的实际影响,避免了统计操作贡献方法所需的多次采样计算。

EoiNAS^[19]计算验证精度和训练时期的比率作为判断操作重要性的附加条件,以此作为指标来充分挖掘操作重要性并指导模型搜索。

与上述方法不同,AGNAS^[20]利用注意机制来引导神经架构搜索,如图4所示,将注意力模块插入到候选操作的输出特征图中,然后使用注意力权重来识别具有明显优势的最佳操作。SWD-NAS^[21]则进一步发现通道维度对权重评估的影响,提出通过上层注意力先筛选重要通道,再用下层注意力计算架构权重,构建层次化的重要性评估策略。

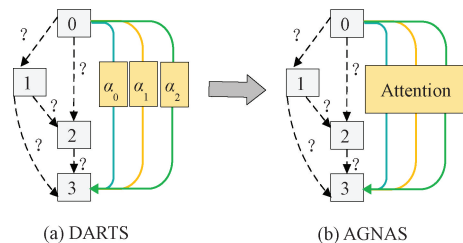


图4 AGNAS 算法示意图

Figure 4 Schematic diagram of AGNAS algorithm

2.3 缩小优化差距

架构深度与初始通道数在搜索与评估阶段的差异导致显著性能鸿沟。渐进式搜索方法通过分阶段逼近目标架构以缩小优化差距,如图5所示,P-DARTS^[22]逐步增加网络深度并剪枝冗余操作,在提升精度的同时降低搜索成本。MIG-DARTS^[23]提出渐进式通道桥接策略,将搜索过程划分为3个阶段逐步提升初始通道数,使搜索网络逐步逼近评估网络的通道规模。

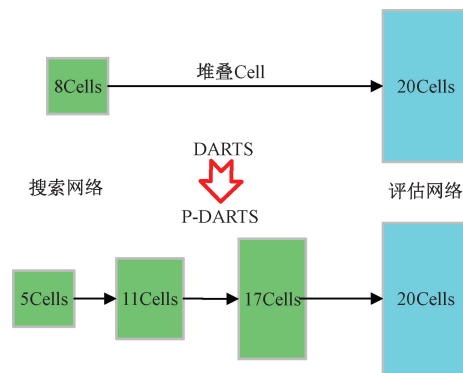


图5 P-DARTS 算法示意图

Figure 5 Schematic diagram of P-DARTS algorithm

渐进式搜索方法缩小了阶段间架构规模的差距,但操作间的参数竞争与训练动态差异仍会导致优化不稳定。EC-DARTS^[24]首次将这一问题归结为“操作差距”与“结构差距”,为缩小此差距,如图6所示,根据不同的操作类别进行解耦,优化操作权重,缩小操作差距;此外,通过诱导结构过渡来弥合搜索和再训练阶段模型结构之间的结构差距。

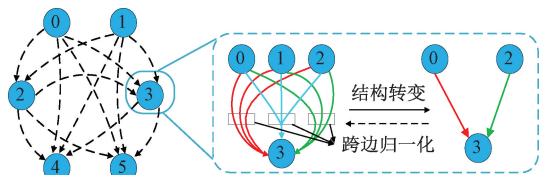


图 6 EC-DARTS 算法示意图

Figure 6 Schematic diagram of EC-DARTS algorithm

CDARTS^[25]进一步融合二者思想,构建搜索-评估网络的循环蒸馏闭环,以特征蒸馏与架构重构实现规模与结构的双重自适应,标志着缩小优化差距从分阶段调整迈向协同进化。

2.4 改进算法比较

表 1 总结三种方法所解决的核心问题,解决思路以及优缺点。在复杂搜索空间,小数据集任务采用降低梯度近似偏差方法,可有效缓解局部最优问题。对于高精度需求任务,存在非参数化操作场景则需要准确评估架构参数重要性,提升选择准确性。最后在跨阶段架构差异显著的任务,需要提高模型的通用性,则需要采用缩小优化差距方法。

3 架构过拟合与搜索稳定性问题

DARTS 在验证集上可能找到表现优异的架构,但由于超网共享权重带来的偏置,该架构在测试集上往往泛化能力不足。同时,DARTS 对超参数与初始条件敏感,易导致搜索结果不稳定,某些操作可能在训练早期迅速主导,抑制了架构多样性。

3.1 正则化

β -DARTS^[26]设计 Beta-Decay 正则化,用于动态调整架构参数的幅度和方差,在提升架构泛化能力与搜索稳定性方面展现出良好效果。正则化方法在控制架构参数的学习过程中不仅有助于防止过拟合,还能提高模型的泛化能力和稳定性。沿着这一思路,ANP-DARTS^[27]采用结构范数正则化器,使架构参数在搜索初期保持均衡,后期趋于稀疏,有利于精确选择操作。SmoothDARTS^[28]则采用基于扰动的正则化策略,通过添加小幅噪声评估稳定性以引导优化,但有可能陷入局部最优。为此,DARTS-

AER^[29]设计架构熵正则化鼓励结构多样性,避免搜索陷入局部最优; ℓ -DARTS^[30]通过 Softmax-Margin 调节权重分布,防止操作选择失衡; Λ -DARTS^[31]则设计两项梯度对齐正则,缓解操作偏置带来的性能退化问题。

3.2 早停策略

DARTS 倾向于过度表示已发现架构中的跳跃连接,跳跃连接是非参数化操作,学习数据表示的能力有限,会导致全局性能崩溃。针对这一挑战,P-DARTS 通过引入操作级 dropout 来规范搜索空间,对跳跃连接施加更高概率的丢弃,并手动限制其数量,从而规避其过度选择问题。NoisyDARTS^[32]则在架构梯度中添加高斯噪声,并通过验证集损失波动触发早停机制,防止训练后期因噪声扰动引发不稳定。RobustDARTS^[33]进一步基于 Hessian 最大特征值设定早停准则,防止模型陷入尖锐极小值带来的结构坍塌。尽管该方法具备理论依据,但其静态阈值限制可能错过潜在优良结构。为了采取更简单、直接的角度限制无参数化操作,OLES^[34]设计了操作级别的早停策略,算法直接监视超网中的每个操作,以确定其是否倾向于过度拟合训练数据,并在发生过度拟合时停止训练该操作。OLES 算法首次将早停策略细化至操作级别,但实现高效的早停机制需要精确的数据监控和分析能力。

3.3 动态调整策略

DARTS 框架中架构参数的动态调整策略旨在平衡搜索过程中的探索与利用,以缓解过拟合与性能崩溃问题。DrNAS^[35]通过将架构混合权重建模为狄利克雷分布随机变量,将搜索过程转化为分布学习问题,借助路径导数与梯度优化实现架构参数的端到端更新,从而提升了架构的多样性与泛化性能。SGAS^[36]采用分阶段贪婪策略,每轮仅对一条边进行操作选择或删除,使搜索路径更稳定,减少冗余探索,并提升最终架构的可控性。但其对初始操作集依赖较强,可能陷入局部最优。为克服局部最优问题,EG-NAS^[37]融合进化算法与梯度信息,通过进化策略提升全局搜索能力,同时利用梯度下降高

表 1 不同改进算法对比

Table 1 Comparison of different improved algorithms

方法类型	核心问题	解决思路	优点	缺点
降低梯度近似偏差	梯度估计不精确	优化目标改进或优化策略调整	提升搜索稳定性;缓解局部最优	计算复杂度高;超参数敏感
准确评估架构参数重要性	架构参数重要性度量偏差	基于扰动、新指标或注意力机制重新评估操作贡献	提升选择准确性;增强可解释性	计算开销大;实现复杂
缩小优化差距	搜索与评估阶段性能差异	渐进式搜索、均衡优化或蒸馏反馈缩小阶段差异	提升泛化能力;降低计算成本	通用性受限

效更新网络参数,提升了搜索质量与效率。

3.4 不同策略对比

表 2 对比了 DARTS 在架构过拟合与搜索稳定性问题上的三类改进策略:正则化、早停策略与动态调整策略。正则化可提升搜索稳定性但限制探索,早停策略提升效率却可能提前终止较优路径,动态调整策略增强探索但可能不稳定。RobustDARTS 提出初期采用强正则化稳定训练,结合早停监控,后期增强动态调整以提升探索。正则化、早停策略和动态调整策略的结合在 DNAS 中能够通过协同作用生成高效架构,前提是合理调优以缓解潜在冲突。

4 优化搜索空间与提升效率问题

4.1 优化搜索空间

DARTS 通过权重共享大大降低了搜索成本,然而这种方法也极大地减少了搜索空间,从而排除了潜在的有前途的架构。为此,D-DARTS^[38]采用单元级嵌套网络替代权重共享,引入替代搜索空间以提升多样性。U-DARTS^[39]扩展搜索空间,设计多样单元结构与连接方式,提升搜索公平性与效率。

不同于直接修改搜索空间的方式,如图 7 所示,DOTS^[40]解耦操作选择与拓扑构建,通过分别进行操作和拓扑搜索,提升收敛速度。与解耦操作与拓扑结构方法不同,ATL-DAS^[41]将拓扑参数引入到搜索空间中以显式地对网络拓扑进行建模,并使用两种采样策略来对独立的子网络进行采样以进行训练和评估。该算法专注于自动学习最优的拓扑结构,结合了自动拓扑学习和梯度优化技术,实现了对架构空间的高效探索。

4.2 提升搜索效率

DARTS 的搜索效率受限于超网结构的高内存占用与冗余计算开销。为此,PC-DARTS^[42]提出部分通道连接策略,在每一层中仅对部分特征通道进行操作选择,并配合边缘正则化以稳定性能,从而在大幅降低搜索资源的同时保持精度。然而,随机通道采样可能造成架构偏差。为提升稳定性,ADAR-

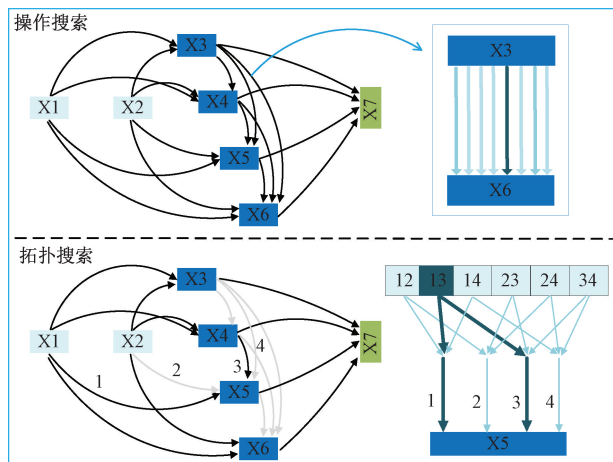


图 7 DOTS 算法示意图

Figure 7 Schematic diagram of DOTS algorithm

TS^[43]在 PC-DARTS 部分通道连接的基础上引入通道注意力机制优先选择高权重通道;EPC-DARTS^[44]进一步通过通道重要性建模捕获跨通道交互,优化资源分配。MCP-DARTS^[45]不但选择关键通道进行采样,还结合温度正则化增大操作区分度,实现效率与性能的双重优化。

LMD-DARTS^[46]通过稠密跳跃连接、参数共享与低秩压缩降低显存开销。为了更高效地利用计算资源,FP-DARTS^[47]设计了并行计算策略,通过并行化多个子任务的计算,显著提升了搜索速度,尤其适合需要高效处理大规模数据集的场景,但该方法对硬件资源要求较高。DLW-NAS^[48]面向边缘设备部署,通过构建轻量搜索空间并控制计算复杂度,在保证较小模型规模的同时,依然能够维持较高的精度和效率;ReCNAS^[49]则通过通道裁剪与动态规划提升资源适应性与架构稳定性。

4.3 搜索空间与效率协同分析

在 DNAS 中,通过解耦操作与拓扑结构、动态剪枝等优化搜索空间技术手段,可减少冗余信息并降低评估成本;而部分通道采样、并行计算等高效策略则在资源受限下扩展探索范围。Slim-DARTS^[50]结合空间重构与通道自适应调节,聚焦核心结构并加快搜索;SQNAS^[51]将全局优化拆解为节点级搜索,

表 2 不同改进策略对比

Table 2 Comparison of different improvement strategies

方法类型	核心思路	优点	缺点	协同场景	冲突场景
正则化	添加约束平衡参数学习	稳定搜索;抑制过拟合	计算开销较大;超参数敏感	与动态调整策略结合提升搜索多样性	正则过强抑制波动,限制搜索灵活性
早停策略	动态终止训练防止性能崩溃	提高效率,节省资源	阈值难设可能错失最优解	与正则化结合稳定后及时终止搜索	与动态调整策略不同步,影响后期优化
动态调整策略	实时调整参数增强搜索灵活性	增强探索能力;避免局部最优	优化复杂;结果易波动	与正则化协同,与早停策略结合加速搜索	干扰正则平滑,易导致训练不稳定

降低搜索空间复杂度,同时结合部分通道采样提升搜索效率。两者验证了优化搜索空间与提升效率协同作用时,一方面减少冗余并聚焦核心特征,缩短每轮搜索时间;另一方面,由于计算负担的减轻,可相应增加搜索迭代次数,对已精简的搜索空间实施更充分的探索。

5 典型 DNAS 算法性能比较

本节基于分类错误率、参数量与搜索耗时三项核心指标,对典型 DNAS 算法进行了比较分析。分

类错误率反映模型性能,参数量衡量架构复杂度与表达能力,而搜索耗时体现搜索效率。然而由于各算法在数据处理、训练设置与硬件条件等方面存在差异,横向比较具有一定挑战性。表 3 汇总了文中提及的 DNAS 算法在 CIFAR-10 和 CIFAR-100 数据集上的实验结果,表中显示多数算法在优化 DARTS 核心问题的同时取得了较好性能。其中多数算法的相应文献中也提供了硬件配置,为研究人员进行公平性评估与方法复现提供了参考依据。

表 3 DNAS 算法性能对比
Table 3 Comparison of DNAS algorithm performance

算法名称	解决的核心问题	测试错误率/%		参数量/ M	搜索消耗/ (GPU days)	实验条件
		CIFAR-10	CIFAR-100			
DARTS ^[4]		3.00	17.76	3.3	1.5	GTX 1080Ti
MiLeNAS ^[7]		2.51	—	3.87	0.30	Tesla V100
Fair DARTS ^[8]		2.54	17.61	2.80	0.40	Tesla V100
TA-DARTS ^[9]		2.60	18.33	4.10	4.00	RTX A4000
DE-DARTS ^[10]		2.46	—	4.10	1.00	RTX 2080Ti
MR-DARTS ^[11]		2.52	17.48	2.50	0.40	RTX 3090
DARTS-CGW ^[12]		2.43	—	4.10	0.40	Tesla P100
DARTS-PT ^[13]		2.61	—	3.00	0.80	GTX 1080Ti
DARTS-PT-CORE ^[14]		2.50	16.73	3.60	0.80	Tesla V100
DARTS-EAST ^[15]	参数化操作的 性能估计偏差	2.43	16.77	3.30	0.59	Tesla V100
IS-DARTS ^[16]		2.56	—	4.25	0.42	RTX 3090
Shapley-NAS ^[17]		2.47	—	3.40	0.30	—
OStr-DARTS ^[18]		2.42	15.78	3.50	0.40	—
EoiNAS ^[19]		2.50	17.30	3.40	0.60	GTX 1080Ti
AGNAS ^[20]		2.60	—	3.60	0.40	GTX 1080Ti
SWD-NAS ^[21]		2.51	17.08	3.17	0.13	RTX 3090
P-DARTS ^[22]		2.50	17.20	3.40	0.30	Tesla P100
MIG-DARTS ^[23]		2.44	16.38	3.00	0.10	RTX 3090
EC-DARTS ^[24]		2.39	16.13	3.20	0.30	TITAN V
CDARTS ^[25]		2.48	15.69	3.90	0.30	Tesla V100
β -DARTS ^[26]		2.53	16.24	3.75	0.40	—
ANP-DARTS ^[27]		2.40	15.89	3.56	0.60	Tesla V100
SmoothDARTS ^[28]		2.61	—	3.30	1.30	GTX 1080Ti
DARTS-AER ^[29]		2.60	—	3.39	4.00	Tesla V100
ℓ -DARTS ^[30]		2.47	16.38	4.10	0.06	—
Λ -DARTS ^[31]	架构过拟合与 搜索稳定性问题	2.43	16.15	3.60	—	—
NoisyDARTS ^[32]		2.65	—	3.10	0.40	Tesla V100
RobustDARTS ^[33]		2.95	18.01	—	1.60	—
OLEs ^[34]		2.30	16.30	3.4	0.40	—
DrNAS ^[35]		2.46	—	4.10	0.60	GTX 1080Ti
SGAS ^[36]		2.66	—	3.70	0.25	GTX 1080Ti
EG-NAS ^[37]		2.53	16.22	3.20	0.10	RTX 3090

表3(续)

算法名称	解决的核心问题	测试错误率/%		参数量/ M	搜索消耗/ (GPU days)	实验条件
		CIFAR-10	CIFAR-100			
D-DARTS ^[38]		2.98	18.90	1.70	0.50	RTX 3090
U-DARTS ^[39]		2.59	16.32	3.30	4.00	RTX 2080Ti
DOTS ^[40]		2.49	16.48	3.50	0.26	Tesla V100
ATL-DAS ^[41]		2.54	16.95	3.90	2.70	TITAN Xp
PC-DARTS ^[42]		2.57	—	3.60	0.10	GTX 1080Ti
ADARTS ^[43]		2.46	17.06	2.90	0.20	—
EPC-DARTS ^[44]	优化搜索空间与 提升效率问题	2.40	15.98	3.20	0.20	RTX 3090Ti
MCP-DARTS ^[45]		2.65	—	3.60	0.09	—
LMD-DARTS ^[46]		2.58	—	4.23	0.12	Tesla V100
FP-DARTS ^[47]		2.50	16.50	3.40	0.08	GTX 1080Ti
DLW-NAS ^[48]		2.73	17.12	2.30	0.45	GTX 1080Ti
ReCNAS ^[49]		2.38	—	5.10	0.30	RTX 2080Ti
Slim-DARTS ^[50]		2.39	16.78	1.70	0.17	—
SQNAS ^[51]		2.70	—	3.78	0.25	GTX 1080Ti

6 结论

DNAS 作为自动化模型设计的重要方法, 凭借其高效性与灵活性, 在图像分类等任务中已取得显著成果。本文以 DARTS 面临的问题为切入点, 从参数化操作性能估计偏差、架构过拟合与搜索稳定性以及搜索空间优化与效率提升等方面, 对典型 DNAS 算法进行了系统分析、对比与总结, 以帮助读者快速了解该领域的研究现状与发展脉络。未来, DNAS 有望逐步拓展至多种新兴应用场景, 并在实际部署中持续增强适应能力。例如, 在自动驾驶和边缘设备部署等资源受限环境中, 结合延迟与能耗约束的多目标优化 DNAS 方法有助于提升模型的部署适应性; 在医学信号分析领域, DNAS 可应用于心电图与脉搏波^[52]建模, 增强模型的结构紧凑性与判别能力; 在语音识别、跨模态匹配等任务中, DNAS 可通过引入任务先验与结构模块, 增强模型的泛化能力与鲁棒性。未来研究可进一步围绕多目标优化、任务驱动搜索空间设计以及跨任务迁移复用等方向深入开展。

参考文献:

[1] Dos Santos C F G, Papa J P. Avoiding overfitting: a survey on regularization methods for convolutional neural networks[J]. *ACM Computing Surveys*, 2022, 54(10s): 1–25.

[2] Hanin B. Which neural net architectures give rise to exploding and vanishing gradients? [C]//*Proceedings of the 32nd International Conference on Neural Information Processing Systems*. New York: ACM, 2018: 580–589.

[3] Sun Renke, Huangfu Zhiyu, Chen Hu, et al. Survey of neural architecture search[J]. *Journal of Computer Applications*, 2024, 44(10): 2983–2994. [孙仁科, 皇甫志宇, 陈虎, 等. 神经架构搜索综述[J]. *计算机应用*, 2024, 44(10): 2983–2994.]

[4] Liu Hanxiao, Simonyan K, Yang Yiming. DARTS: differentiable architecture search [PP/OL]. V2. arXiv (2019–04–23) [2025–10–09]. <https://doi.org/10.48550/arXiv.1806.09055>.

[5] Heuillet A, Nasser A, Arioui H, et al. Efficient automation of neural network design: a survey on differentiable neural architecture search[J]. *ACM Computing Surveys*, 2024, 56(11): 1–36.

[6] Zoph B, Vasudevan V, Shlens J, et al. Learning transferable architectures for scalable image recognition[C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2018: 8697–8710.

[7] He Chaoyang, Ye Haishan, Shen Li, et al. MiLeNAS: efficient neural architecture search via mixed-level reformulation[C]//*Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 2020: 11990–11999.

[8] Chu Xiangxiang, Zhou Tianbao, Zhang Bo, et al. Fair DARTS: eliminating unfair advantages in differentiable architecture search[C]//*Computer vision – ECCV 2020*. Cham: Springer International Publishing, 2020: 465–480.

[9] Shin J, Park K, Kang D K. TA-DARTS: temperature annealing of discrete operator distribution for effective differential architecture search[J]. *Applied Sciences*, 2023, 13(18): 10138.

- [10] Mun J, Ha S, Lee J. DE-DARTS: neural architecture search with dynamic exploration[J]. *ICT Express*, 2023, 9(3): 379–384.
- [11] Ma Benteng, Zhang Yanning, Xia Yong. Momentum recursive DARTS [J]. *Pattern Recognition*, 2024, 156: 110710.
- [12] Liu Wenbo, Deng Tao, An Rui, et al. DARTS-CGW: research on differentiable neural architecture search algorithm based on coarse gradient weighting [C]//*Pattern recognition and computer vision*. Singapore: Springer Nature Singapore, 2025: 31–44.
- [13] Wang Ruochen, Cheng Minhao, Chen Xiangning, et al. Rethinking architecture selection in differentiable NAS [PP/OL]. V1. arXiv (2021–08–10) [2025–10–09]. <https://doi.org/10.48550/arXiv.2108.04392>.
- [14] Xie Weisheng, Li Hui, Fang Xuwei, et al. DARTS-PT-CORE: collaborative and regularized perturbation-based architecture selection for differentiable NAS[J]. *Neurocomputing*, 2024, 580: 127522.
- [15] Fang Xuwei, Xie Weisheng, Li Hui, et al. DARTS-EAST: an edge-adaptive selection with topology first differentiable architecture selection method[J]. *Applied Intelligence*, 2025, 55(7): 526.
- [16] He Hongyi, Liu Longjun, Zhang Haonan, et al. IS-DARTS: stabilizing DARTS through precise measurement on candidate importance[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, 38(11): 12367–12375.
- [17] Xiao Han, Wang Ziwei, Zhu Zheng, et al. Shapley-NAS: discovering operation contribution for neural architecture search [C]//*Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 2022: 11882–11891.
- [18] Yang Le, Zheng Ziwei, Han Yizeng, et al. OStr-DARTS: differentiable neural architecture search based on operation strength[J]. *IEEE Transactions on Cybernetics*, 2024, 54(11): 6559–6572.
- [19] Zhou Yuan, Xie Xukai, Kung S Y. Exploiting operation importance for differentiable neural architecture search [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(11): 6235–6248.
- [20] Sun Zihao, Hu Yu, Lu Shun, et al. Agnas: attention-guided micro and macro-architecture search [C]//*International Conference on Machine Learning*. New York: ACM, 2022: 20777–20789.
- [21] Xue Yu, Han Xiaolong, Wang Zehong. Self-adaptive weight based on dual-attention for differentiable neural architecture search [J]. *IEEE Transactions on Industrial Informatics*, 2024, 20(4): 6394–6403.
- [22] Chen Xin, Xie Lingxi, Wu Jun, et al. Progressive DARTS: bridging the optimization gap for NAS in the wild [J]. *International Journal of Computer Vision*, 2021, 129(3): 638–655.
- [23] Hao Debei, Pei Songwei. MIG-DARTS: towards effective differentiable architecture search by gradually mitigating the initial-channel gap between search and evaluation [J]. *Neural Computing and Applications*, 2025, 37(8): 6085–6096.
- [24] Zhou Qinjin, Zheng Xiawu, Cao Liujuan, et al. EC-DARTS: inducing equalized and consistent optimization into DARTS [C]//*Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Piscataway: IEEE, 2021: 11966–11975.
- [25] Yu Hongyuan, Peng Houwen, Huang Yan, et al. Cyclic differentiable architecture search [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(1): 211–228.
- [26] Ye Peng, Li Baopu, Li Yikang, et al. β -DARTS: beta-decay regularization for differentiable architecture search [C]//*Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 2022: 10864–10873.
- [27] Zeng Xianhua, Wu Jie, Xia Yaoguang, et al. A differentiable neural architecture search algorithm with architecture norm regularization [J]. *Scientia Sinica (Informationis)*, 2024, 54(9): 2181–2199. [曾宪华, 吴杰, 夏耀光, 等. 一种结构范数正则化的可微神经结构搜索算法 [J]. *中国科学: 信息科学*, 2024, 54(9): 2181–2199.]
- [28] Chen Xiangning, Hsieh C J. Stabilizing differentiable architecture search via perturbation-based regularization [C]//*Proceedings of the 37th International Conference on Machine Learning*. New York: ACM, 2020: 1554–1565.
- [29] Jing Kun, Chen Luoyu, Xu Jungang. An architecture entropy regularizer for differentiable neural architecture search [J]. *Neural Networks*, 2023, 158: 111–120.
- [30] Hu Liwei, Wang Zidong, Li Han, et al. ℓ -DARTS: light-weight differentiable architecture search with robustness enhancement strategy [J]. *Knowledge-Based Systems*, 2024, 288: 111466.
- [31] Movahedi S, Adabinejad M, Imani A, et al. $\$$ λ $\$$ -DARTS: mitigating performance collapse by harmonizing operation selection among cells [PP/OL]. V2. arXiv (2023–03–01) [2025–10–26]. <https://doi.org/10.48550/arXiv.2210.07998>.
- [32] Chu Xiangxiang, Zhang Bo. Noisy differentiable architecture search [C]//*Proceedings of the British Machine Vi-*

- sion Conference 2021. British Machine Vision Association, 2021: 434.
- [33] Zela A, Elsken T, Saikia T, et al. Understanding and robustifying differentiable architecture search [PP/OL]. V2. arXiv (2020-01-28)[2025-08-11]. <https://doi.org/10.48550/arXiv.1909.09656>.
- [34] Huang Yihua, Ji Zipeng, Jiang Shen, et al. Operation-level early stopping for robustifying differentiable NAS [C] // Proceedings of the Advances in Neural Information Processing Systems 36. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2023: 70983-71007.
- [35] Chen Xiangning, Wang Ruochen, Cheng Minhao, et al. DrNAS: dirichlet neural architecture search [PP/OL]. V4. arXiv (2021-03-16)[2025-10-26]. <https://doi.org/10.48550/arXiv.2006.10355>.
- [36] Li Guohao, Qian Guocheng, Delgadillo I C, et al. SGAS: sequential greedy architecture search [C] // Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 1617-1627.
- [37] Cai Zicheng, Chen Lei, Liu Peng, et al. EG-NAS: neural architecture search with fast evolutionary exploration [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(10): 11159-11167.
- [38] Heuillet A, Tabia H, Arioui H, et al. D-DARTS: distributed differentiable architecture search [J]. Pattern Recognition Letters, 2023, 176: 42-48.
- [39] Huang Lan, Sun Shiqi, Zeng Jia, et al. U-DARTS: uniform-space differentiable architecture search [J]. Information Sciences, 2023, 628: 339-349.
- [40] Gu Yuchao, Wang Lijuan, Liu Yun, et al. DOTS: decoupling operation and topology in differentiable architecture search [C] // Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 12306-12315.
- [41] Liu Qinglin, Lin Jingbo, Lv Xiaoqian, et al. ATL-DAS: automatic topology learning for differentiable architecture search [J]. Displays, 2023, 80: 102541.
- [42] Xu Yuhui, Xie Lingxi, Zhang Xiaopeng, et al. PC-DARTS: partial channel connections for memory-efficient architecture search [C] // International Conference on Learning Representations. Addis Ababa: ICLR, 2019.
- [43] Xue Yu, Qin Jiafeng. Partial connection based on channel attention for differentiable neural architecture search [J]. IEEE Transactions on Industrial Informatics, 2023, 19(5): 6804-6813.
- [44] Cai Zicheng, Chen Lei, Liu Hailin. EPC-DARTS: efficient partial channel connection for differentiable architecture search [J]. Neural Networks, 2023, 166: 344-353.
- [45] Pan Jie, Zheng Xuechi, Zou Xiaoyu. Differentiable neural architecture search with channel performance measurement [J]. Control and Decision, 2024, 39(7): 2151-2160. [潘杰, 郑学驰, 邹筱瑜. 基于通道性能度量的神经网络结构搜索算法 [J]. 控制与决策, 2024, 39(7): 2151-2160.]
- [46] Li Zhongnian, Xu Yixin, Ying Peng, et al. LMD-DARTS: low-memory, densely connected, differentiable architecture search [J]. Electronics, 2024, 13(14): 2743.
- [47] Wang Wenna, Zhang Xiuwei, Cui Hengfei, et al. FP-DARTS: fast parallel differentiable neural architecture search for image classification [J]. Pattern Recognition, 2023, 136: 109193.
- [48] Li Shu, Mao Yuxu, Zhang Fuchang, et al. DLW-NAS: differentiable light-weight neural architecture search [J]. Cognitive Computation, 2023, 15(2): 429-439.
- [49] Peng Cheng, Li Yangyang, Shang Ronghua, et al. ReC-NAS: resource-constrained neural architecture search based on differentiable annealing and dynamic pruning [J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(2): 2805-2819.
- [50] Yin Shantong, Niu Ben, Wang Rui, et al. Spatial and channel level feature redundancy reduction for differentiable neural architecture search [J]. Neurocomputing, 2025, 630: 129713.
- [51] Biju G M, Pillai G N. Sequential node search for faster neural architecture search [J]. Knowledge-Based Systems, 2024, 300: 112145.
- [52] Lu Peng, Wang Hanzhang, Mao Xiaobo, et al. Pulse wave classification model based on convolutional autoencoder [J]. Journal of Zhengzhou University (Engineering Science), 2021, 42(5): 56-61. [逯鹏, 王汉章, 毛晓波, 等. 基于卷积自编码器网络的脉搏波分类模型 [J]. 郑州大学学报(工学版), 2021, 42(5): 56-61.]

Overview of Differentiable Neural Network Architecture Search

LU Peng^{1,2,3}, LI Keyan^{1,2}, ZHANG Hongpo^{3,4}, CHEN Liwei¹, WU Jiahui^{1,2}, LIU Shuaibing^{1,2}

(1. School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou 450001, China; 2. Robot Perception and Control Henan Engineering Laboratory, Zhengzhou 450001, China; 3. Henan Collaborative Innovation Center for Internet based Medical and Health Services, Zhengzhou 450052, China; 4. Network Management Center, Zhengzhou University, Zhengzhou 450001, China)

Abstract: Neural Architecture Search (NAS) was an interdisciplinary study in the field of deep learning, which aims to automate the design of neural network structures. NAS requires repeated training and evaluation of a large number of candidate networks, which is computationally expensive. Differentiable Neural Architecture Search (DNAS) transforms the discrete architecture search problem into a differentiable continuous optimization problem, which reduces the computational cost. Firstly, a search algorithm framework for differentiable network architecture was constructed from three aspects: search space, search strategy and performance evaluation strategy. Secondly, the performance estimation bias, architecture overfitting and search stability problems of parameterization operation, as well as the improvement strategies of optimizing search space and improving efficiency are analyzed, compared and summarized. Then, the error rate, parameter quantity, search time and experimental hardware conditions of typical DNAS algorithms on image classification datasets were compared and analyzed. Finally, it points out the application potential of DNAS in complex scenarios such as edge device deployment, medical signal analysis, and cross-modal matching, and proposes future research directions toward multi-objective optimization, task-driven search space design, and cross-task transfer and reuse.

Keywords: neural architecture search; deep learning; differentiable neural architecture search; Continuous optimization; performance estimation