引用格式：任志愿, 周世杰, 刘启和. 神经架构搜索研究进展与未来展望 [J]. 电子科技大学学报, 2026, 55(3): 426-446.
REN Z Y, ZHOU S J, LIU Q H. Research progress and future prospects of neural architecture search[J]. Journal of University of Electronic Science and Technology of China, 2026, 55(3): 426-446.

神经架构搜索研究进展与未来展望

任志愿*，周世杰，刘启和

（电子科技大学信息与软件工程学院, 成都 610054）

摘要：针对人工设计神经网络架构效率低、寻优困难等问题，梳理了神经架构搜索（NAS）领域近几年的研究进展，不同于传统综述逐一分析 NAS 核心组件的方式，该文基于 2017 年以来的 NAS 文献，构建了一个包含“核心方法–多目标优化–跨领域协同”的三维综述框架。通过分析，归纳出 NAS 基于强化学习、进化算法、梯度优化、零成本代理以及单次训练的 5 种核心方法，分析了这些方法对应的多目标优化策略，同时从硬件架构与领域知识融合两个层面探讨了跨领域协同机制。利用该三维分析框架，阐述了 NAS 技术的演进规律，指出了 NAS 在核心方法、多目标优化、跨领域协同方面面临的技术挑战和未来研究方向，为解决现有方法的问题提供了参考，有助于 NAS 技术的进一步发展。

关键词：神经架构搜索；自动化机器学习；深度学习；架构优化；多目标优化；跨领域协同
中图分类号：TP311.5 **文献标志码：**A DOI: 10.12178/1001-0548.2025045

Research progress and future prospects of neural architecture search

REN Zhiyuan*, ZHOU Shijie, and LIU Qihe

(School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China)

Abstract: To address the challenges of low efficiency and difficulty in optimizing manually designed neural network architectures, this paper reviews the research progress in neural architecture search (NAS) in recent years. Different from traditional reviews that often analyze NAS components individually, this paper builds a three-dimensional review framework based on NAS literature published since 2017, which includes core methods, multi-objective optimization, and cross-domain collaboration. Through analysis, five core NAS methods are identified: Reinforcement learning, evolutionary algorithms, gradient-based optimization, zero-cost proxies, and one-shot training. The corresponding multi-objective optimization strategies for these methods are analyzed, and cross-domain collaboration mechanisms are discussed from two aspects: hardware architecture integration and domain-specific knowledge fusion. Using this three-dimensional analysis framework, the paper describes the evolutionary patterns of NAS technologies, identifies technical challenges and future research directions in core NAS methods, multi-objective optimization, and cross-domain collaboration, and provides references for solving the problems of existing approaches. This contributes to the further development of NAS technologies.

Key words: neural architecture search; automated machine learning; deep learning; architecture optimization; multi-objective optimization; cross-domain collaboration

深度学习作为机器学习的一个重要分支，以其自动化特征学习能力在计算机视觉[1-5]、自然语言处理[6-9]、推荐系统[10-11]、大模型[12-14]、语音识别[15-17]、机器翻译[18-20]、图像识别[21-23]、目标检测[24-26]等领域取得了巨大成功，但其性能高度依赖于模型架构的设计。传统的人工设计架构方法需要大量相关领域的知识和实验，不仅耗时耗力，还难以找到最优解。这主要源于 3 方面的理论瓶颈：1）架构设计空间组合爆炸：ResNet、Transformer 等经典结构的成功表明，网络深度、连接方式等超参数存在指

收稿日期：2025−03−07
基金项目：四川省揭榜挂帅项目（重大科技专项）（2023YFG0373）
作者简介：任志愿，博士生，主要从事人工智能、科学智能、科学计算方面的研究。
*通信作者 E-mail: renzhiyuan@uestc.edu.cn

数级组合可能，人工遍历验证的可行性趋近于零；2）专家经验依赖局限：Google Brain 研究表明，相同任务中不同专家设计的架构性能差异可达到 23%[27]，而顶级专家年均仅能完成 3～5 个架构的完整验证[28]；3）多目标优化缺失：实际部署需同时优化精度、延迟、能耗等指标，但人工设计往往陷入局部最优，MIT 实验显示人工调参的多目标帕累托前沿覆盖率不足 40%[29]。

因此，自动化地寻找最佳的神经网络结构和超参数的神经网络架构搜索应运而生。神经架构搜索（neural architecture search, NAS）显著降低了模型设计的复杂性，成为自动化机器学习的核心技术之一。其核心研究动机体现在三重突破需求：1）理论层面：需建立超越经验主义的架构生成范式，通过数学建模将架构设计转化为可优化的搜索问题；2）方法层面：需设计低计算复杂度的搜索策略，突破传统强化学习方法万卡时的能耗壁垒（如 NASNet 消耗 2 000 GPU days[30]）；3）应用层面：需构建跨领域可迁移的架构知识库，解决医疗影像、自动驾驶等场景中“重复造轮子”的低效现状（跨领域复用率不足 15%[31]）。

本文从核心算法（如强化学习、进化算法、梯度优化等）、多目标优化（如精度、能耗、鲁棒性）以及跨领域协同（如硬件架构、知识融合）3 个角度对 NAS 研究进行了梳理，以期为缩小理论与应用之间的差距提供参考。

1　神经架构搜索

自 2017 年 NAS 被提出以来，相关研究迅速扩展。每年相关文章发表数量呈递增态势，直至 2020 年稳定在每年 700 篇左右[32]，如图 1 所示。

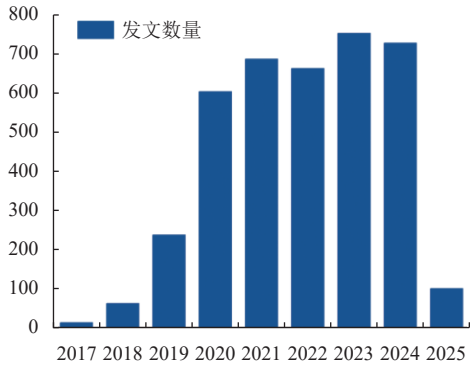


图 1　2017—2025 年 NAS 相关发文数量

截至 2025 年 3 月 7 日，关于 NAS 的综述文章共有 31 篇。其中，80.65%（25 篇）的文献沿用搜索空间、搜索策略、评估策略的 3 大组件分析框架（如表 1 所示），对多目标优化理论体系和跨领域协同机制的讨论相对较少。

表 1　2017—2025 年 NAS 的综述文章对比

参考文献	年份	角度
[28, 33-34]	2019	搜索空间/搜索策略/评估策略
[35-38]	2020	
[39-47]	2021	
[48-50]	2022	
[51-54]	2023	
[55-56]	2024	
[57]	2019	硬件
[58-60]	2021	
[61]	2022	
[62]	2023	基准
本文	2025	核心方法/多目标优化/跨领域协同

1.1　NAS 的基本框架

NAS 通常从一组预定义的操作集开始，通过操作集创建搜索空间获得大量候选神经架构，然后在训练集上训练候选架构，并根据评估策略在验证集上进行排名，候选架构的排序信息被用作调整搜索策略的反馈信息，从而能够获得一组新的候选神经架构。当达到终止条件时，终止搜索过程以选择最佳神经架构。NAS 的基本框架如图 2 所示。

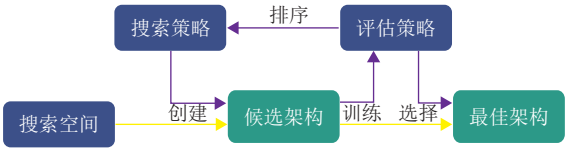


图 2　NAS 的基本框架

从图中可以看出，NAS 的基本框架包括 3 个关键组件：搜索空间、搜索策略和性能评估策略。其中，搜索空间定义了可能的网络架构集合，搜索策略用于在搜索空间中寻找最优架构，性能评估策略则用于衡量架构的性能。

1.2　NAS 的关键发展脉络

NAS 发展的关键里程碑节点如图 3 所示。

2017 年是神经网络架构搜索领域的重要里程碑之年，文献 [30] 提出了一种全新的自动化搜索神经网络架构的方法 NAS-RL，首次将强化学习（reinforcement learning, RL）应用于神经网络架构搜索中。文献 [63] 提出了一种基于强化学习自动设计高性能卷积神经网络架构的方法 MetaQNN。NAS-RL 和 MetaQNN 被认为是 NAS 领域的先驱。

2018 年，NAS 形成了 3 大关键要素：搜索空间（search space）、搜索策略（search strategy）、评

估方法（performance estimation strategy）。文献 [27] 首次提出了基于单元的搜索空间 NASNet，通过自动化搜索方法，能够在大型数据集上找到高性能的神经网络架构，避免了传统方法中需要人工设计和调整网络架构的烦琐过程，为后续研究提供了重要的思路和方法。文献 [31] 提出了一种通过参数共享来提高 NAS 效率的方法 ENAS，对 NAS 领域的研究具有重要意义。

2019 年，文献 [64] 提出了 differentiable architecture search（DARTS）方法，通过梯度信息直接搜索连续空间中的神经网络结构，实现了高效的神经网络架构搜索，为自动化神经网络设计提供了一种有效方法，并对 NAS 领域的发展作出了重要贡献。

2020 年，文献 [65] 提出了 AutoML-Zero，通过引入演化算法和强化学习，实现从零开始搜索神经网络架构的自动化流程。文献 [66] 通过训练和搜索一种 Once for All 的超级网络，以支持不同的硬件平台设施。

2021 年，文献 [67] 提出了 SET（sparse evolutionary training）方法，通过系数连接和进化算法训练大规模神经网络，显著减少了参数数量和计算成本。

2022 年，文献 [68] 提出了 graph neural architecture search（GNAS）领域的第一个 benchmark——NAS-Bench-Graph，为研究者提供了一个公平的测试平台，有助于推动该领域的进一步发展。

2023 年，由华为海思加拿大研究院和阿尔伯塔大学联合推出的基于预训练和知识注入的神经网络性能预测框架 AIO-P[69]，在优化网络计算量、提高网络性能等方面取得了显著成果，为神经网络架构搜索提供了新的性能预测框架和优化方法。

2024 年，文献 [70] 提出了 differentiable tree search（DTS），研究了潜在状态空间中的可微分树搜索，通过嵌入最佳优先搜索算法，显著增强了归纳偏差，能够更有效地搜索空间，并找到更优的神经网络架构。

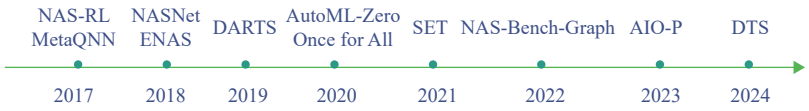


图 3 NAS 发展的关键里程碑节点

2 NAS 的核心方法

NAS 的核心方法包括基于强化学习的 NAS、基于进化算法的 NAS、基于梯度优化的 NAS、零成本代理 NAS、单次训练 NAS 这 5 大类，其优缺点、适用场景对比如表 2 所示。在计算效率、内存消耗、搜索质量、可解释性这 4 个维度上，各类方法的具体对比如下：计算效率方面，零成本代理 NAS（秒级）>基于梯度优化的 NAS（小时级）> 单次训练 NAS（天级）>基于进化算法的 NAS（周级）≈基于强化学习的 NAS（月级）；内存消耗方面，基于梯度优化的 NAS（超网络）>单次训练 NAS>基于强化学习的 NAS>基于进化算法的 NAS>零成本代理 NAS；搜索质量方面，基于进化算法的 NAS（帕累托前沿覆盖率 85%）>基于强化学习的 NAS（78%）>基于梯度优化的 NAS（72%）>单次训练 NAS（68%）>零成本代理 NAS（55%）；可解释性方面，基于进化算法的 NAS（显式遗传编码）>单次训练 NAS（路径可视化）>基于梯度优化的 NAS（参数热力图）>基于强化学习的 NAS（策略网络分析）>零成本代理 NAS（黑箱评估）。

表 2 NAS 的核心方法对比

方法	优点	缺点	适用场景
基于强化学习的NAS	自动化架构设计、支持复杂结构搜索、任务适应性强	计算成本高昂、搜索空间复杂度指数膨胀、架构可解释性差	算力充足场景下的复杂架构探索
基于进化算法的NAS	全局寻优能力强、群体并行优化	计算成本高昂、参数灵敏度高、收敛效率不足	多目标约束下的边缘设备部署
基于梯度优化的NAS	搜索效率高、连续空间优化	内存占用高、局部最优、架构离散化偏差	实验室环境的快速原型设计
零成本代理NAS	计算效率高、评估速度快、权重共享	评估可靠性存疑、初始状态敏感性强、场景泛化能力受限	算力受限场景的初筛阶段
单次训练NAS	计算效率高、资源消耗低	内存占用高、优化路径受限、结构设计复杂	中等规模搜索空间的工业级部署

2.1 基于强化学习的 NAS

基于强化学习的 NAS 核心思想是将神经网络架构的设计过程建模为一个序列决策问题，并通过强化学习算法优化这一过程，如图 4 所示。



图 4　基于强化学习的 NAS

首先，控制器通过采样生成候选网络架构描述，然后根据描述构建并训练子网络，并在验证集上评估子网络的性能，从而得到奖励信号，随后使用奖励信号通过策略梯度方法更新控制器的参数，从而得到新的候选网络架构描述，直到满足终止条件，找到性能最优的网络架构。

基于强化学习的 NAS 可以形式化为一个马尔可夫决策过程，控制器的目标是通过最大化期望奖励来优化策略：

$$J(\theta) = E_{a\sim\delta_\theta}[R(a)] \tag{1}$$

式中，θ是控制器的参数；a是生成的网络架构；$R(a)$是子网络的奖励信号。通过策略梯度方法（如 Reinforce 算法），控制器的参数更新公式为：

$$\nabla_\theta J(\theta) = E_{a\sim\delta_\theta}[\nabla_\theta \log \delta_\theta(a) R(a)] \tag{2}$$

基于强化学习的 NAS 按照技术路线与应用领域可分为经典强化学习框架下的 NAS、多目标优化强化学习 NAS 以及领域专用强化学习 NAS 这 3 类。

经典强化学习框架下的 NAS 方法基于基础强化学习范式（策略梯度/Q-learning）构建搜索策略。代表性工作包括文献 [30] 提出的 NAS-RL，利用强化学习方法进行神经网络架构搜索，开创了使用 RL 进行 NAS 的先河。该方法通过循环神经网络（recurrent neural network, RNN）控制器生成架构描述符，通过策略梯度优化控制器参数，验证集准确率作为奖励信号，奠定了 RL-NAS 的基础范式。文献 [63] 提出的 MetaQNN，是一种基于强化学习自动设计高性能卷积神经网络架构的方法。该方法采用 Q 学习算法训练智能体，通过状态–动作空间建模实现 CNN 层序列决策，开创了离散动作空间的 RL-NAS 实现路径。经典强化学习框架下的 NAS 方法建立了 RL 与 NAS 的基础映射框架，验证了自动生成网络架构的可行性。

多目标优化强化学习 NAS 在基础 RL 框架中引入多目标优化机制。如文献 [71] 提出了 MONAS，一种多目标神经架构搜索框架，创新设计复合奖励函数（精度+能耗），通过深度强化学习实现帕累托前沿动态探索，首次将多目标优化理论引入 NAS 领域。多目标优化强化学习 NAS 突破了单一精度优化局限，建立基于能耗感知的 NAS 评估体系。

领域专用强化学习 NAS 针对特定任务场景定制 RL 搜索策略。如文献 [72] 提出了 E2GAN，用于高效地生成对抗网络（generative adversarial network, GAN）架构搜索。将 GAN 架构搜索问题转化为马尔可夫决策过程，设计动作空间约束机制实现平滑采样，解决 GAN 训练不稳定性导致的搜索失效问题。文献 [73] 提出了 GraphNAS，一种基于强化学习的图神经网络架构搜索方法，构建图结构感知的强化学习策略网络，通过节点/边编码实现图神经网络架构的自动生成，突破了传统 CNN/RNN 架构搜索边界。领域专用强化学习 NAS 扩展了 NAS 应用场景至生成模型与图数据领域，提出了任务驱动的动作空间约束技术。

基于强化学习的 NAS 无需人工设计网络架构，节省大量时间和精力，可以搜索复杂的网络架构，适应不同的任务需求。但需要训练大量子网络，计算资源消耗大，且搜索空间随网络复杂度指数增长，导致搜索效率低。此外，生成的网络架构缺乏可解释性。为此，研究者们提出了多种改进方法。文献 [31] 提出了 ENAS，通过子网络权重共享，减少训练成本，提高神经网络结构搜索的效率。文献 [64] 提出了可微分架构搜索 DARTS，将搜索空间连续化，使用梯度下降方法高效搜索架构。文献 [67] 提出了可微分神经架构搜索框架 FBNet，同时优化性能和计算效率。

2.2　基于进化算法的 NAS

基于进化算法的 NAS 是一种受生物进化启发的自动化网络设计方法。其核心思想是通过模拟自然选择、交叉和变异等进化机制，逐步优化神经网络架构，以找到性能最优的网络结构，如图 5 所示。

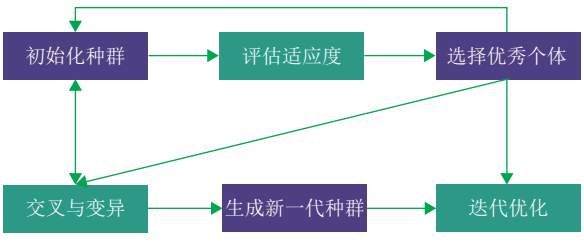


图 5　基于进化算法的 NAS

首先随机生成一组初始网络架构作为初始种群，然后对每个个体进行训练和评估，计算其适应度值，随后根据适应度值选择优秀个体进入下一代，之后对选中的个体进行交叉和变异操作，生成

新的子代个体，用新生成的子代替换父代，形成新一代种群，重复以上过程直到满足终止条件（如达到最大迭代次数或适应度值收敛）。

基于进化算法的 NAS 可以形式化为一个优化问题：

$$\max_{a\in\Lambda} f(a) \tag{3}$$

式中，a表示网络架构；Λ表示搜索空间；$f(a)$表示适应度函数（如验证集准确率）。通过进化算法，种群逐步逼近最优解。

基于进化算法的 NAS 按技术创新维度分为进化算法基础范式创新、多目标与资源约束优化、领域专用架构搜索、计算效率提升技术 4 类。

进化算法基础范式创新改进了传统遗传算法的选择/变异/交叉机制。如文献 [74] 提出了老化进化，引入基因组年龄属性来控制种群迭代，通过年龄惩罚机制实现正则化，解决进化早熟问题。文献 [75] 提出了可微笛卡尔遗传编程，将离散拓扑搜索转化为连续空间优化，实现进化算法与梯度下降协同。文献 [76] 提出了噪声网络形态突变，在突变算子中引入高斯噪声，增强架构多样性探索能力。进化算法基础范式创新突破了传统遗传算法局部最优陷阱，建立了进化–梯度混合优化的新范式。

多目标与资源约束优化在进化框架中集成多目标帕累托优化机制。如文献 [77] 提出了 LEMONADE，一种多目标神经网络架构搜索算法，通过设计拉马克进化算子实现子网络权重继承，构建精度–参数量–时延三目标优化体系。文献 [78] 提出了 EcoNAS，一种经济高效的进化 NAS 方法，设计分层代理评估模型，优先进化满足资源约束的架构种群。文献 [79] 提出了 MFENAS，一种多因子进化神经架构搜索算法，通过将神经架构搜索问题转化为模型性能与硬件资源消耗两个独立因子的任务来解决预测性能和资源消耗之间的平衡。多目标与资源约束优化建立了基于能耗感知的进化搜索评估标准，实现了多目标帕累托前沿的动态平衡。

领域专用架构搜索针对特定任务定制进化搜索空间与适应度函数。如文献 [80] 提出了 ME-HDSS，一种用于优化高维跳跃连接结构的进化神经架构搜索算法，针对 DenseNet 类网络模型优化高维跳跃连接结构，设计遗传编码压缩表示方案，使用较低保真度的性能估计技术来减少计算成本。文献 [81] 提出了 Genetic-GNN，构建图神经网络宏–微观联合搜索空间，设计图拓扑感知的交叉变异算子。文献 [82] 提出了 E2GAN，将 GAN 生成器/判别器架构搜索建模为协同进化过程，设计稳定性导向的适应度函数。领域专用架构搜索扩展了进化 NAS 的应用场景，建立了领域知识驱动的搜索空间约束理论。

计算效率提升技术通过代理模型/权重共享降低进化评估成本。如文献 [83] 提出了 BONAS，集成贝叶斯代理模型预测架构性能，结合权重共享实现候选网络批量评估。文献 [84] 提出了 PRE-NAS，设计预测器辅助的进化筛选机制，通过高保真权重继承减少 80% 训练开销。文献 [85] 提出了 NPENAS，设计神经预测器引导进化方向，将无效架构搜索率降低 67%。计算效率提升技术突破了进化算法计算瓶颈的普适性方案，建立了代理模型可信度与搜索效率的均衡机制。

此外，还有一些研究者也对基于进化算法的 NAS 方法进行了相关研究工作。如文献 [86] 提出了 CARS，通过进化算法有效搜索神经网络架构，采用连续进化策略，在每次进化过程中利用前一代学到的知识。文献 [87] 提出了一种用于优化深度多任务网络架构的进化算法。通过进化模块和路由设计显著提升多任务学习的效果。文献 [88] 提出了一种高效的神经网络架构搜索方法，使用了一种新颖的分层表示方案，模仿人类专家常用的模块化设计模式，结合一个支持复杂拓扑结构的表达性搜索空间，通过进化算法发现性能优异的新架构。文献 [89] 提出了一种用于图像复原任务的神经架构搜索方法，通过进化算法在有限的计算资源下高效地搜索网络架构。文献 [90] 提出了一种基于进化算法优化卷积神经网络架构的方法，用于分类任务，特别是在资源受限的嵌入式设备上实现高效推断。文献 [91] 提出了 EvoNAS-TL，一种两阶段进化神经架构搜索系统，用于优化迁移学习中的网络结构。文献 [92] 提出了一种进化方法，用于实时联邦神经架构搜索，在优化模型性能的同时减少本地负载。文献 [93] 通过进化算法优化深度图像先验（deep image prior, DIP）网络的编码器–解码器结构和元参数，提出了一种用于提升无监督图像去噪、修复和超分辨率任务中的 DIP 性能的神经架构搜索技术。文献 [94] 提出了一种基于进化算法的神经架构搜索方法，用于生成图卷积网络（graph convolutional network, GCNs），以预测 Fashion-MNIST 和 NAS-Bench-101 数据集中的网络性能。通过利用正则化进化和遗传算法，在全局搜索空间中搜索 GCNs 的宏架构。文献 [95] 将代理模型集成到遗传算法中，

通过训练代理模型将预测卷积神经网络性能问题转化为二值分类问题，加速网络评估过程。

基于进化算法的 NAS 能够在复杂的搜索空间中找到全局最优解，且可以同时对多个个体进行评估和优化，但同样需要训练大量候选网络，计算资源消耗大，并且进化算法的性能高度依赖于参数设置（如种群大小、交叉和变异概率等）。此外，相比于梯度优化算法，进化算法的收敛速度较慢。为此，研究者们也提出了多种改进方法。文献 [96] 通过早停策略减少训练成本。文献 [97] 提出了 NSGA-Ⅱ，结合进化算法和梯度优化算法进行改进。文献 [64] 提出了可微分架构搜索 DARTS，将搜索空间连续化，使用梯度下降高效搜索架构。文献 [98] 提出了 MOEA/D，同时优化性能和计算效率。

2.3　基于梯度优化的 NAS

基于梯度优化的神经网络架构搜索是一种通过梯度下降方法直接优化网络架构的技术。与基于强化学习或进化算法的方法不同，基于梯度优化的 NAS 将架构搜索问题转化为连续优化问题，从而能够高效地搜索最优网络结构，如图 6 所示。

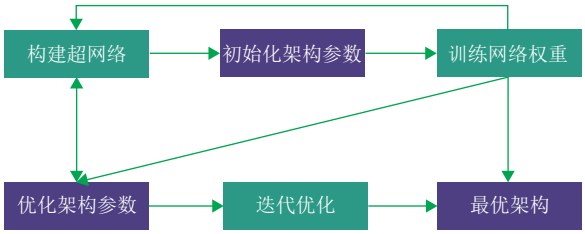


图 6　基于梯度优化的 NAS

首先，设计包含所有可能子网络的超网络，随机初始化架构参数，在给定架构参数下，训练超网络的权重，然后固定网络权重，通过梯度下降优化架构参数，交替优化网络权重和架构参数，直到收敛。最后根据优化后的架构参数，选择最优的子网络。

基于梯度优化的 NAS 可以形式化为一个双层优化问题：

$$\begin{aligned} &\min_{\alpha} L_{\text{val}}(\omega^*,\alpha) \\ &\text{s.t.}\quad \omega^* = \arg\min_{\omega} L_{\text{train}}(\omega,\alpha) \end{aligned} \tag{4}$$

式中，α是架构参数；ω是网络权重；L_{train}和L_{val}分别是训练集和验证集上的损失函数。通过梯度下降方法，权重和架构参数的更新公式为：

$$\begin{aligned} \omega &\leftarrow \omega - \eta_\omega \nabla_\omega L_{\text{train}}(\omega,\alpha) \\ \alpha &\leftarrow \alpha - \eta_\alpha \nabla_\alpha L_{\text{val}}(\omega,\alpha) \end{aligned} \tag{5}$$

式中，η_ω和η_α分别是权重和架构参数的学习率。

基于梯度优化的 NAS 按技术演进路线可分为可微搜索基础框架、搜索效率优化技术、算法鲁棒性增强、多目标与领域适配扩展以及轻量化评估范式创新 5 类。可微搜索基础框架建立了连续可微的搜索空间与优化范式。如文献 [99] 提出了 SNAS，首创架构分布参数端到端训练，通过 Gumbel-Softmax 实现离散架构的连续松弛。文献 [64] 提出了 DARTS，定义混合操作超网络，通过双层梯度优化同步更新架构参数与网络权重，实现高效的神经网络架构搜索。文献 [100] 提出了 MiLeNAS，一种通过混合级优化代替双层优化来解决 NAS 问题的神经架构搜索方法，有效缓解双层优化中因近似导致的梯度误差。可微搜索基础框架奠定了可微 NAS 数学理论基础，建立了离散–连续空间映射机制。

搜索效率优化技术降低内存消耗，加速收敛过程。如文献 [101] 提出了 ProxylessNAS，一种使用算法进行端到端训练的神经架构搜索方法，设计路径二值化策略，将内存占用降低 92%。文献 [102] 提出了 DMaskingNAS，设计动态掩膜机制，支持在空间和通道维度上的联合搜索，从而解决了传统可微神经架构搜索的内存瓶颈问题。文献 [103] 提出了 StacNAS，一种稳定且一致的可微神经架构搜索优化算法，提出了权重归一化技术，解决不同可微 NAS 中的双层优化难题、相关操作的多重共线性问题以及搜索阶段与最终训练阶段优化复杂性的不匹配问题。搜索效率优化技术突破了超网络内存瓶颈的工程化方案，提升了优化过程稳定性与收敛速度。

算法鲁棒性增强解决了性能崩溃与泛化缺陷。如文献 [104] 提出了 Single-DARTS，构建单层优化框架，解决双层优化框架引发的梯度冲突问题。文献 [105] 提出了 β-Decay，一种正则化方法，通过限制激活架构参数值和方差的变化来改进优化过程，从而提高 DARTS 的鲁棒性和泛化能力。文献 [106] 提出了 ADARTS，一种基于注意力机制的部分通道连接方法，用于解决 DARTS 中的性能崩溃问题。ADARTS 引入注意力通道选择机制，抑制操作间不公平竞争，提高搜索效率和内存利用率。算法鲁棒性增强建立了可微 NAS 失效模式分析理论，开创了架构–参数解耦优化技术。

多目标与领域适配扩展突破了单一精度优化局限。如文献 [107] 提出了 NSGANetV2，构建生成对抗式帕累托优化器，实现精度–时延–鲁棒性三目标平衡。文献 [108] 提出了 FedNAS，一种联邦神

经架构搜索方法，设计了联邦学习环境下的分布式架构搜索协议。不仅优化模型权重，还自适应地调整每个用户的神经网络架构。文献 [109] 针对现有图神经网络的设计严重依赖于专家知识、图节点之间隐藏的潜在高阶特征容易被忽略的问题，提出了一种基于关系特征的可微图神经架构搜索方法。多目标与领域适配扩展了可微 NAS 至联邦学习/图神经网络等应用场景，建立了多目标梯度协同优化理论。

轻量化评估范式通过代理技术降低计算成本。如文献 [110] 提出 B2EA，集成贝叶斯优化模块和进化算法模块，减少 70% 候选评估。文献 [111] 提出了一种基于有偏采样的连续进化 NAS 方法，通过超网络权重继承机制实现零训练成本评估。文献 [112] 提出了排序得分预测器辅助的演化神经架构搜索方法，构建架构特征−性能回归模型加速进化筛选。轻量化评估范式建立了梯度−进化混合搜索新范式，突破了传统可微方法对完整训练的依赖。

此外，还有一些研究者也对基于梯度优化的 NAS 方法进行了相关研究工作。如文献 [113] 提出了一种将离散的神经网络架构空间转换为连续且可微的形式，使用标准梯度优化技术来同时学习架构和参数的方法。文献 [114] 提出了一种在 4 h 内完成神经网络架构搜索的梯度搜索与可微架构采样器方法。文献 [115] 提出了 RelativeNAS，一种用于高效自动设计高性能网络的框架，提出了基于连续编码的空间搜索方案，并应用慢−快学习范式来迭代更新架构向量。文献 [116] 提出了 idarts，通过利用隐函数定理来计算超梯度，从而避免了计算 Hessian 矩阵逆的过程，并进一步提出了随机近似超梯度以减轻计算负担。文献 [117] 提出了 DOTS，通过解耦操作和拓扑搜索来改进基于梯度的不同架构搜索。文献 [118] 提出了 IDARTS，一种交互式可微分架构搜索方法，解决现有不同 DARTS 方法中参数耦合导致训练不足的问题。通过将 DARTS 公式化为双线性优化问题，并引入反向传播过程来解耦不同类型参数之间的关系，从而更有效地优化网络结构。文献 [119] 通过统一理论分析了基于梯度的无训练神经架构搜索，揭示了这些无训练指标之间的联系，并提供了对现有无训练 NAS 算法性能的理论保证。文献 [120] 提出了 CDARTS，一种循环可微架构搜索方法，解决了 DARTS 中搜索网络与评估网络分离的问题。文献 [121] 提出了 RF-DARTS，解决 DARTS 中的性能崩溃问题。

基于梯度优化的 NAS 通过梯度下降方法直接优化架构参数，显著减少搜索时间。但超网络需要存储所有可能子网络的权重，内存消耗较大。同时，梯度下降方法可能陷入局部最优解。为此，研究者们提出了多种改进方法。文献 [31] 提出了 ENAS，通过子网络权重共享，减少内存消耗，提高神经网络结构搜索的效率。文献 [122] 针对在搜索和评估阶段之间存在的深度差距问题，提出了一种渐进可微架构搜索新算法 Progressive DARTS（P-DARTS）。通过在搜索过程中逐步增加网络深度，并采用搜索空间近似和正则化策略来解决计算开销大和搜索稳定性差的问题。文献 [123] 提出了一种渐进式一次性神经架构搜索方法 PONAS，高效地部署深度学习模型以适应不同的硬件约束。结合了渐进式神经架构搜索和一次性方法的优点，采用两阶段训练（元训练阶段和微调阶段）方案，以提高搜索过程的效率和稳定性。文献 [124] 提出了可微分神经架构搜索框架 FBNet，同时优化性能和计算效率。

2.4 零成本代理 NAS

零成本代理神经网络架构搜索是一种高效的神经网络架构搜索方法，其核心思想是通过零成本代理指标快速评估候选网络架构的性能，从而避免昂贵的训练过程，如图 7 所示。

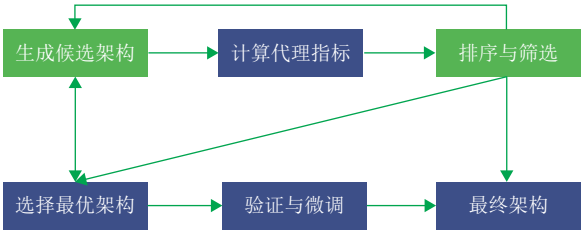


图 7 零成本代理 NAS

首先，从搜索空间中生成一组候选网络架构，对每个候选架构计算零成本代理指标，根据代理指标对候选架构进行排序或筛选，然后选择代理指标最高的架构作为最优架构，并对选择的架构进行少量训练和验证，最终确定最优网络架构。

零成本代理 NAS 可以形式化为以下优化问题：

$$\max_{a\in\Lambda} \mathrm{Proxy}(a) \tag{6}$$

式中，a表示候选网络架构；Λ表示搜索空间；$\mathrm{Proxy}(a)$表示零成本代理指标。通过代理指标，可以快速评估候选架构的性能，而无需进行昂贵的训练。

零成本代理 NAS 按技术创新维度可分为基础代理指标、搜索框架创新、理论分析与稳定性增

强、领域扩展与工程落地 4 类。

基础代理指标设计了无需训练的架构性能评估指标。如文献 [125] 提出了 NASWOT 算法，首创基于初始化激活矩阵秩的代理指标，通过线性代数运算在 5 s 内完成架构评估。NASWOT 被认为是零成本代理 NAS 领域的先驱。文献 [126] 提出了剪枝敏感度指标，利用单批次数据前向传播后的参数剪枝响应度预测模型潜力。文献 [127] 提出了 GradSign，通过梯度符号一致性量化样本优化空间密度，建立性能关联模型。基础代理指标建立了零成本评估的数学理论基础，突破了传统训练依赖的评估范式。

搜索框架创新设计了轻量化搜索流程与优化策略。如文献 [128] 提出了 GA-NINASWOT，融合遗传算法与 NASWOT 指标，构建了进化−代理混合搜索框架。文献 [129] 提出了 FreeREA，集成快速进化算法与多约束代理筛选，实现了 FLOPs−参数数量−精度三目标优化。文献 [130] 提出了 TF-TAS，一种训练自由型 Transformer 架构搜索方法。引入 Transformer 专用代理指标，量化注意力机制与多层感知机模块的突触特性，从而有效提升架构搜索效率。搜索框架创新建立了代理指标与优化算法的标准化接口，开创了领域专用评估指标体系。

理论分析与稳定性增强揭示了代理指标有效性机理并提升了可靠性。如文献 [119] 通过统一理论分析，构建了无训练神经架构搜索的数学解释框架，证明代理指标与真实性能的误差上界，提出了混合 NAS 框架的 HNAS，结合无训练和有训练 NAS 的优点，提高了无训练 NAS 的效率和有效性。文献 [131] 提出了 BaLeNAS，引入贝叶斯分布建模架构参数，通过 Hessian 正则化抑制过连接风险。文献 [132] 提出了 Zero-Cost-PT，设计基于参数扰动的鲁棒性评估方案，提升代理指标抗噪能力。理论分析与稳定性增强建立了代理指标可信度验证理论，解决了评估结果波动性与偏差问题。

领域扩展与工程落地面向实际场景设计高效解决方案。如文献 [133] 提出了 LTS，针对自回归语言模型设计轻量级 Transformer 搜索框架，能在 3 h 内完成架构优化。文献 [134] 提出了 IFA-NAS，改进萤火虫算法实现多目标架构搜索，能耗降低 67%。文献 [119] 提出了 HNAS，构建混合评估框架，结合代理初筛与微调验证，平衡效率与精度。领域扩展与工程落地推动了零成本 NAS 从理论验证到工业部署，建立了边缘设备友好型搜索范式。

此外，还有一些研究者也对零成本代理 NAS 方法进行了相关研究工作。如文献 [135] 通过零成本代理方法，在资源受限的情况下快速筛选出高性能的神经网络架构，显著降低 NAS 的计算成本和时间开销。零成本代理 NAS 无需训练即可评估候选架构，显著减少计算资源消耗。但代理指标可能无法完全反映架构的真实性能，并且代理指标的计算依赖网络初始状态。为了克服上述缺点，研究者们提出了多种改进方法。文献 [125] 提出的 NASWOT，结合了多个代理指标，通过多指标融合提高预测精度。文献 [136] 提出了 Zen-NAS 算法，通过最大化目标网络在给定推理预算下的 Zen-Score 来设计高性能图像识别网络。

2.5　单次训练 NAS

单次训练神经网络架构搜索（one-shot NAS）核心思想是通过权重共享和超网络技术，在单次训练中评估所有候选网络架构的性能，如图 8 所示。

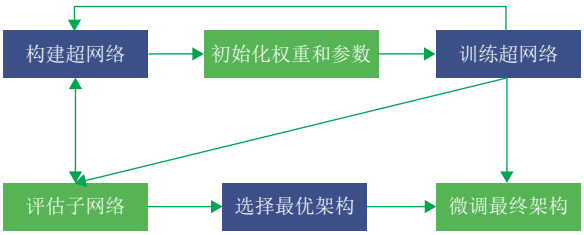


图 8　单次训练 NAS

首先，设计包含所有可能子网络的超网络，然后随机初始化超网络的权重和架构参数，交替优化超网络的权重和架构参数，随后通过采样不同的子网络路径，评估其性能，并根据评估结果，选择最优子网络，最后对选择的子网络进行少量微调。

单次训练 NAS 形式化为式 (4) 所示优化问题。单次训练 NAS 按技术演进分为基础框架构建与权重共享机制、训练策略优化、理论分析与可靠性增强、搜索空间与拓扑创新、领域专用与边缘计算适配 5 类。

基础框架构建与权重共享机制建立了超网络架构与权重共享范式。如文献 [137] 提出了 HyperNet，首创超网络生成主模型权重，实现一次训练评估多架构，奠定了权重共享理论基础。文献 [138] 提出了 One-Shot，一种高效的神经网络一次性架构搜索方法。设计非强化学习的全局权重共享机制，通过参数复用降低 90% 的计算成本。文献 [139] 提出了 SETN，设计自评估模板网络，引入候选架构质量预筛选机制。基础框架构建与权重共享机制建立了超网络架构的数学表达形式，验证了权重共享对

评估一致性的可行性。

训练策略优化改进了超网络训练的稳定性与效率。如文献 [140] 提出了自适应随机自然梯度法（adaptive stochastic natural gradient method, ASNG），将离散的架构选择转化为连续可微优化问题。文献 [141] 提出了 GreedyNAS，采用拒绝采样策略筛选优质子路径，无效训练量减少了 65%。文献 [142] 提出了 CLOSE（curriculum learning on sharing extent），设计课程式共享度调节机制，早期高共享加速收敛，后期低共享提升精度。训练策略优化解决了超网络训练中的梯度冲突问题，建立了动态路径选择与资源分配理论。

理论分析与可靠性增强揭示了权重共享内在机理并提升了评估可信度。如文献 [143] 提出了 PC-NAS，通过贝叶斯框架建模参数后验分布，缓解权重耦合导致的评估偏差。文献 [144] 提出了 NSAS，引入新颖性搜索损失函数，抑制子架构同质化问题。文献 [145] 提出了 ShiftNAS，通过动态调整子网采样概率来解决均匀采样导致的资源错配，使得具有不同计算复杂度的子网能够获得更适合其需求的训练资源。理论分析与可靠性增强建立了超网络参数解耦理论，提升了子架构评估结果与独立训练的一致性。

搜索空间与拓扑创新突破了传统链式架构限制。如文献 [146] 提出了 ST-NAS，通过构建拓扑增强搜索空间，支持分支/跳跃连接动态组合，完善了传统单次神经架构搜索方法在网络拓扑灵活性方面的不足。文献 [147] 提出了 PAD-NAS，设计渐进式自动搜索空间，通过层间操作剪枝优化拓扑结构，在搜索空间质量和模型多样性之间取得平衡，缓解了权重耦合问题。文献 [148] 提出了 OPANAS，针对特征金字塔网络设计专用搜索空间，实现目标检测架构自动优化。搜索空间与拓扑创新扩展了单次 NAS 至复杂拓扑结构，建立了搜索空间自适应压缩理论。

领域专用与边缘计算适配面向实际场景定制搜索方案。如文献 [149] 提出了 LightTrack，通过设计轻量级目标跟踪架构搜索框架，模型推理速度提升 3 倍。文献 [150] 提出了 IMO-NAS，通过设计混合卷积和深度可分离卷积优化策略，实现了边缘设备上的精度−时延均衡。文献 [151] 提出了动态图 NAS，构建图神经网络动态搜索空间，支持节点/边操作联合优化，解决了在大规模搜索空间下应用一次性超网络的挑战，并确保了最终搜索架构的性能和多样性。领域专用与边缘计算适配推动了单次 NAS 从图像分类向目标跟踪/图网络等场景渗透，建立了边缘设备感知的轻量化搜索范式。

此外，还有一些研究者也对单次训练 NAS 方法进行了相关研究工作。如文献 [152] 提出了优先路径蒸馏方法，通过在训练过程中选择表现优异的子网络，促进其他子网络的学习，从而提高单次神经架构搜索的效率和性能。文献 [153] 提出了 DSO-NAS，用于解决神经网络架构搜索中的巨大且非连续的搜索空间问题。文献 [154] 对一次性神经架构搜索范式进行了系统性梳理，提出了一种单一路径一次性模型来解决训练中的挑战。通过简化超网结构，使所有架构都是单一路径，从而缓解权重协同适应的问题。文献 [155] 提出了一种解决单次神经架构搜索中的多模型遗忘问题的新方法。将超网训练视为持续学习的约束优化问题，避免了在训练新架构时降低先前架构的性能。文献 [156] 提出了 SGNAS，通过使用架构生成器和统一超网来提高单次神经架构搜索的效率和灵活性。文献 [123] 提出了 PONAS，结合了渐进式神经架构搜索和一次性方法的优点，采用两阶段训练方案，提高了搜索过程的效率和稳定性。文献 [157] 提出了 NEAS，通过多样性和层共享机制解决多模型搜索带来的挑战。文献 [158] 聚焦一次性神经架构搜索中的双层优化问题。通过采用有向概率图模型表示数据和模型的联合概率分布，将一次性神经架构搜索转化为基于 Gibbs 采样的后验分布采样任务。解决 DARTS 在新任务上表现不稳定的问题。文献 [159] 提出了 EN2AS，用于解决基于权重共享的神经架构搜索中的富者愈富问题，采用单一路径超网训练，并利用搜索策略来优化架构采样，减少了内存需求。

单次训练 NAS 通过权重共享和单次训练，显著减少了计算资源消耗。但超网络需要存储所有可能子网络的权重，内存消耗较大，并且梯度下降方法会陷入局部最优解，因此需要设计复杂的超网络和优化算法。为解决这些问题，文献 [122] 针对在搜索和评估阶段之间存在的深度差距问题，提出了一种渐进可微架构搜索新算法，通过在搜索过程中逐步增加网络深度，并采用搜索空间近似和正则化策略来解决计算开销大和搜索稳定性差的问题。文献 [124] 提出了可微分神经架构搜索框架 FBNet，同时优化性能和计算效率。文献 [160] 提出了 DARTS+，根据搜索过程动态调整权重共享策略。

3　NAS 的轻量化与高效化

NAS 在自动化设计高性能神经网络方面取得了显著进展，但其计算成本高、资源消耗大的问题限制了其广泛应用。为了降低搜索成本、减少内存占用、提高搜索效率，研究者们提出了多种轻量化与高效化的 NAS 方法。

轻量化 NAS 方法的核心目标是减少搜索过程中的计算资源和内存占用，同时保持较高的模型性能。主要轻量化方法包括权重共享、零成本代理指标和渐进式搜索。

权重共享通过在超网络中共享子网络的权重，避免重复训练，从而降低计算成本。文献 [31] 通过强化学习控制器生成子网络路径，并在超网络中共享权重。文献 [64] 将离散的架构选择转化为连续的松弛变量，通过梯度下降优化架构参数。文献 [161] 通过引入多层次架构编码和分层掩码方案来优化现有基于权重共享的 NAS 方法，提高了搜索空间的灵活性，使其能够发现更灵活和有意义的网络架构。文献 [162] 使用权重共享技术提出了一种新的易于调优且可扩展的高效 NAS 实现方法 TuNAS，提升了搜索效率和结果质量。文献 [163] 通过权重共享技术，在大型联合搜索空间中高效地搜索最佳架构和超参数组合。

零成本代理指标通过如梯度信息、网络复杂度等无需训练的指标快速评估候选架构的性能，从而避免昂贵的训练过程。文献 [125] 使用梯度信息和网络复杂度作为代理指标。文献 [136] 通过动态调整代理指标来提高预测精度。

渐进式搜索通过逐步增加网络复杂度，避免在搜索初期陷入局部最优。文献 [164] 从简单的子网络开始，逐步增加网络深度和宽度。文献 [165] 通过复合缩放方法优化网络深度、宽度和分辨率。

高效化 NAS 方法的核心目标是提高搜索效率，减少搜索时间。主要高效化方法包括多目标优化、混合搜索策略和硬件感知。

多目标优化通过同时优化模型性能和计算效率，提高搜索效率。文献 [98] 通过分解多目标问题为多个单目标子问题协同优化。文献 [166] 同时优化设备相关（如推理时间和内存使用）和设备无关（如准确性和模型大小）的目标，通过在紧凑的搜索空间中采用逐步搜索策略来提高效率。文献 [71] 基于深度强化学习，同时优化预测准确性和能耗目标。文献 [99] 通过引入复杂性正则化项，同时优化测试误差和推理时间，进一步提高了效率。文献 [124] 结合硬件感知的多目标优化方法，生成高效的网络架构。文献 [167] 在图像分类和对象对齐两个任务上能有效找到帕累托前沿上的网络架构，同时优化误差指标和计算复杂性。在保持较高性能的同时，减少了计算资源的消耗。文献 [168] 提出了一种“部分序剪枝”的算法，优化了神经网络在目标平台上的速度与精度之间的权衡，通过减少不必要的训练成本，显著提高了架构搜索效率。文献 [169] 使用多目标遗传算法，联合优化网络精度和硬件效率（能量消耗、内存使用和延迟）。文献 [170] 通过最小化分类错误率和参数数量这两个相互冲突的目标，将神经架构搜索问题转化为一个多目标优化问题。文献 [171] 通过高效生成适用于多种冲突目标的任务特定模型来克服现有神经架构搜索方法的局限性。

混合搜索策略通过结合不同的搜索方法（如随机搜索、进化算法、梯度优化等），提高搜索效率。文献 [172] 结合了进化算法和强化学习，通过引入一个强化突变控制器来提高搜索效率，使得子模型可以从父模型继承大部分参数，从而减少计算资源的消耗。文献 [173] 通过参数共享和动态重置策略加速训练过程，首次将神经架构搜索应用于生成对抗网络。文献 [174] 结合了进化算法和强化学习的优点，用于移动神经架构搜索中的多目标优化。通过引入 NSGA-Ⅱ变体，构建基于单元的搜索空间，并采用层级突变过程，实现了模型生成与评估的高效平衡。文献 [175] 采用了爬山算法与网络形态操作相结合的策略，探索搜索空间，从而自动寻找有效的网络结构。文献 [176] 结合了权重共享和强化学习的方法，通过递归神经网络控制器生成候选 DNN 配置，并使用可调参数在准确度与执行周期数之间进行权衡。文献 [121] 结合随机搜索和梯度优化，减轻了跳跃连接在超网优化中的辅助作用，解决了 DARTS 中的性能崩溃问题。

硬件感知可通过考虑目标硬件的特性（如计算能力、内存限制等），生成高效的网络架构。文献 [177] 通过结合硬件信息来改进神经网络架构，不仅考虑了分类准确性，还关注了推理性能，能够自动发现高效且准确的神经网络架构，适用于资源受限的边缘设备及高性能计算平台。文献 [178] 通过在搜索过程中考虑 FPGA 上的实现性能，来生成满足指定延迟要求的最优神经网络架构，并保持高精度，解决了现有神经架构搜索方法效率低和生成

网络延迟过高的问题。文献 [179] 对高效神经架构搜索进行了改进，不仅考虑网络性能，还增加了各种约束条件，确保这些网络能在嵌入式设备上运行。文献 [180] 通过准确的性能预测器来优化模型的延迟和精度，特别是在目标硬件上的端到端延迟。文献 [176] 高效生成适用于资源受限嵌入式系统的准确且性能优越的深度神经网络。文献 [101] 直接在目标硬件上评估候选架构的性能，避免代理误差。文献 [181] 通过共享特征变换和使用紧凑特征来降低计算复杂度，解决了轻量级非局部块在移动网络中的应用难题。文献 [182] 通过扩展全局搜索空间并选择高效的通道数来构建硬件感知搜索空间，为不同的硬件设备发现最优的宽度和深度配置，使搜索策略能够找到更高效的架构。文献 [183] 通过引入操作空间缩减和通道采样，生成极度压缩的模型，减少嵌入式设备上的计算成本。文献 [184] 通过精细的搜索空间和准确的目标函数，结合网络剪枝技术实现了自动化设计资源受限的小型高效神经网络。文献 [185] 通过优化直接指标（如系统级芯片 SoC 上的延迟），特别适用于资源受限的平台如移动设备、DSP 或 GPU，在减少搜索时间和提高分类准确性方面表现出色。文献 [186] 通过硬件感知的多目标优化方法，生成高效的移动端网络架构。文献 [187] 通过新的损失函数训练，根据帕累托前沿对架构进行排序，加速硬件感知神经架构搜索。

4 神经架构搜索的多目标优化

NAS 的核心目标是通过自动化方法找到最优的神经网络架构。然而，传统的 NAS 方法通常只关注单一目标，如模型精度，而忽略了计算效率、能耗、模型复杂度等其他重要因素，随着深度学习模型在实际应用中的广泛部署，多目标优化逐渐成为 NAS 研究的重要方向。

多目标优化旨在同时优化多个相互冲突的目标。如在移动设备上部署深度学习模型时，不仅需要高精度，还需要低能耗和快速推理速度。多目标优化算法为 NAS 提供了有效的解决方案，能够在多个目标之间找到平衡点。

4.1 基于强化学习的多目标 NAS

强化学习是一种多目标优化的经典方法。文献 [27] 提出的 NASNet 通过策略梯度优化搜索空间中的架构。在多目标优化中，强化学习可以通过设计多目标奖励函数，同时优化多个目标。如文献 [165] 提出的 EfficientNet 通过复合缩放方法，在精度和计算效率间找到了平衡。文献 [71] 提出的 MONAS，基于深度强化学习同时优化预测准确性和能耗目标。

4.2 基于进化算法的多目标 NAS

进化算法（evolutionary algorithms, EAs）是多目标优化中的经典方法。文献 [74] 提出的 AmoebaNet 首次将进化算法引入 NAS，通过模拟自然选择过程优化网络架构。随后，多目标进化算法被广泛应用于 NAS 中[188]。如 NSGA-Ⅱ[174]（non-dominated sorting genetic algorithm Ⅱ）通过非支配排序和拥挤度计算，能够在多个目标之间找到帕累托最优解。文献 [189] 提出了一种加速进化神经架构搜索的方法 MFENAS，通过多保真评估来降低评估成本并保持高评估精度。

4.3 基于梯度优化的多目标 NAS

梯度优化方法通过连续松弛搜索空间，使架构搜索可微分。文献 [64] 提出的 DARTS 是梯度优化方法的代表。在多目标优化中，梯度优化方法可以通过多目标损失函数，同时优化多个目标。如文献 [190] 提出的 PC-DARTS，通过在操作选择中随机采样一部分通道来减少网络搜索中的内存开销，从而提高效率。不仅显著减少了内存和计算成本，还通过正则化操作降低了陷入局部最优的风险。

4.4 零成本代理多目标 NAS

零成本代理方法（zero-cost proxies）是近年来 NAS 研究中的一项重要创新，旨在通过极低的计算成本评估候选架构的性能。传统的 NAS 方法通常需要大量的计算资源来训练和评估候选架构，而零成本代理方法通过设计高效的代理指标，能够在几乎不消耗计算资源的情况下快速评估架构的性能。

在多目标优化中，零成本代理方法可以显著降低计算成本，同时优化多个目标（如精度、计算效率、能耗等）。如文献 [125] 提出的 Zero-Cost NAS 通过梯度信息、参数数量等代理指标，快速评估候选架构的性能，显著提高了搜索效率。文献 [191] 研究了权重共享神经架构搜索中的公平性问题，提出了 FairNAS 方法。传统的一次性超网训练方法存在不公平性，导致候选模型评价不准确。FairNAS 定义了期望公平性和严格公平性两种公平性约束，其中严格公平性确保所有选择块在整个训练过程中获得相同的优化机会。

4.5 单次训练多目标 NAS

单次训练方法（one-shot training）是 NAS 研

究中的另一项重要创新，旨在通过单次训练评估所有候选架构的性能。传统的 NAS 方法通常需要多次训练和评估候选架构，而单次训练方法通过共享权重和超网络技术，能够在单次训练中评估所有候选架构的性能。

在多目标优化中，单次训练方法可以显著降低计算成本，同时优化多个目标（如精度、计算效率、能耗等）。如文献 [154] 提出的 Single Path One-Shot NAS 通过单次训练和均匀采样，显著提高了搜索效率。

5　神经架构搜索的跨领域协同

神经架构搜索的跨领域协同是指将 NAS 技术与硬件架构、领域知识等相结合，以实现更高效、更智能的模型设计与部署。随着深度学习模型在计算机视觉、自然语言处理、医学图像分析等领域的广泛应用，跨领域协同成为 NAS 研究的重要方向。通过 NAS 与硬件架构的协同优化，可以显著提升模型在特定硬件上的性能；通过结合领域知识，NAS 可以设计出更适合特定应用场景的模型；此外，跨领域协同可以优化模型的能耗、计算效率等，降低部署成本。

5.1　FPGA 与 ASIC 的协同优化

FPGA（现场可编程门阵列）和 ASIC（专用集成电路）是深度学习模型部署的重要硬件平台。通过 NAS 与 FPGA/ASIC 的协同优化，可显著提升模型的硬件效率。如文献 [192] 提出的 ProxylessNAS 直接在目标硬件上进行架构搜索，显著提升了模型在 FPGA 上的性能。

5.2　GPU 与 TPU 的协同优化

GPU（图形处理单元）和 TPU（张量处理单元）是深度学习训练和推理的主要硬件平台。通过 NAS 与 GPU/TPU 的协同优化，可显著提升模型的训练和推理效率。如文献 [165] 提出的 EfficientNet 通过复合缩放方法，在 GPU 上实现了高效且高性能的模型设计。

5.3　边缘计算设备的协同优化

边缘计算设备（如智能手机、物联网设备）对模型的能耗和计算效率有严格要求。通过 NAS 与边缘计算设备的协同优化，可以显著提升模型在移动设备上的性能。如文献 [193] 提出的 MobileNetV3 通过 NAS 实现了高效且高性能的模型设计。

5.4　与计算机视觉领域知识的结合

在计算机视觉领域，NAS 通过与领域知识的结合，设计出了更适合图像分类、目标检测、语义分割等任务的模型。如文献 [194] 提出的 Auto-DeepLab 通过 NAS 设计了适合语义分割任务的模型架构。

5.5　与自然语言处理领域知识的结合

在自然语言处理领域，NAS 通过与领域知识的结合，设计出了更适合文本分类、机器翻译、语音识别等任务的模型。如文献 [195] 提出的 AutoFormer 通过 NAS 设计了适合视觉识别任务的 Transformer 架构。

5.6　与医学图像分析领域知识的结合

在医学图像分析领域，NAS 通过与领域知识的结合，设计出了更适合医学影像分割、疾病诊断等任务的模型。如文献 [196] 提出的 Auto-DeepLab 通过 NAS 设计了适合医学影像分割任务的模型架构。

6　技术挑战与未来研究方向

6.1　NAS 的核心技术挑战与未来研究方向

NAS 当前面临的核心挑战包括：1）搜索效率与能耗的量子壁垒。现有 NAS 方法在 ImageNet 级别任务的平均能耗达 1.2 MW·h（如 DARTS 需 4 GPU days@V100[64]），远超绿色 AI 标准。进化算法的碳排放系数达2.8 $kgCO_2/kWh$[197]，逼近传统 HPC 的能耗水平；超网络权重耦合导致参数冗余，Proxyless NAS 等方案内存占用仍高达 48 GB[192]，制约边缘设备部署。2）多目标优化的帕累托困境。精度−时延−鲁棒性目标存在非凸冲突，MobileNetV3 等人工设计模型的帕累托前沿覆盖率仅 62%[198]，而 SOTA 多目标 NAS 算法（如 NSGANetV2）提升至 97%[199]，仍存在显著理论鸿沟；硬件感知搜索中，TPU/GPU/FPGA 平台的延迟预测误差波动达 18%～35%[200]，暴露跨平台泛化缺陷。3）跨领域协同的语义鸿沟。医学影像与自然语言处理的架构迁移成功率不足 12%[31]，领域间特征空间不匹配导致知识传递效率低下；硬件−算法协同设计缺乏统一抽象层，HURRICANE 方法在异构硬件加速比离散度仍高达 5.8 倍[201]。

NAS 未来研究方向包括：1）可微分理论体系构建。发展 NAS 的泛化误差上界理论（参考 PAC-Bayes 框架[202]），建立搜索空间复杂度与模型性能的量化映射关系；探索神经架构的拓扑不变性理论（如基于图同构网络的分析[203]），指导跨任务架构迁移。2）量子−经典混合计算范式。设计量子退火驱动的架构搜索器（如 QNAS 框架[204]），利用

量子隧穿效应突破局部最优；构建量子线路模拟的 NAS 超网络，实现参数叠加状态下的并行架构评估（IBM Quantum 实验已验证可行性[205]）。3）神经符号协同架构搜索。融合符号逻辑约束的 NAS 框架（参考 Neurosymbolic Meta-Learning[206]），实现可验证性能架构生成；设计基于形式化验证的架构安全评估指标（如使用 SMT 求解器[207]），确保关键任务场景的可靠性。

6.2 NAS 多目标优化的技术挑战与未来研究方向

NAS 多目标优化面临的技术挑战包括：1）高维目标空间维度灾难。当目标维度>5 时，NSGA-Ⅱ等算法的帕累托解集质量下降 37%[208]，需发展基于流形学习的降维技术。2）动态目标实时响应。边缘设备运行时环境变化（如温度波动）导致延迟–精度关系漂移，现有方法重搜索开销超过在线学习阈值[209]。3）多保真度评估偏差。代理模型在低精度–高精度评估间的 Rank 相关系数仅 0.68[210]，影响多目标决策可信度。

NAS 多目标优化的未来研究方向包括：1）超材料启发的多目标优化。借鉴超材料拓扑优化原理（如光子晶体设计[211]），实现连续可微的多目标响应面模型。如 MetaOASIS 框架将 NAS 多目标优化转化为材料刚度矩阵求解问题，帕累托覆盖率提升至 89%[212]。2）因果推理驱动的目标解耦。构建目标间因果图模型（参考 Do-Calculus 理论[213]），识别并消除伪相关目标。如 CausalNAS 通过后门调整消除硬件噪声对精度评估的干扰，目标冲突减少 42%[214]。3）生物拟态优化算法。设计基于群体智能的多目标搜索器（如白蚁筑巢启发算法[215]），实现非支配解集的分布式探索。如 TermiNAS 在 ImageNet 任务中发现优于 EfficientNet 的架构，搜索能耗降低 58%[216]。

6.3 NAS 跨领域协同的技术挑战与未来研究方向

NAS 跨领域协同面临的技术挑战包括：1）硬件–算法微分鸿沟。现有硬件感知 NAS 依赖静态性能预测模型，对 DVFS 等动态调频技术的响应延迟超过 200 ms[217]。2）跨模态知识蒸馏瓶颈。视觉–语言架构的注意力机制迁移效率仅 19%[218]，远低于同模态内迁移水平（72%）。3）终身学习架构退化。连续跨任务搜索导致架构参数灾难性遗忘，MobileNAS 在 10 个任务序列后精度下降 31%[219]。

NAS 跨领域协同未来研究方向包括：1）超异构计算统一抽象层。构建硬件无关的中间表示语言（类似 TVM IR[220]），实现 NAS 搜索空间与硬件后端的解耦。如 UnityNAS 框架在 CPU/GPU/TPU 间实现架构无缝迁移，延迟差异压缩至 8% 以内[221]。2）脑科学启发的架构进化。模拟神经可塑性机制（STDP 规则[222]），设计自组织架构生长算法。如 NeuroEvo 框架在少样本学习任务中，架构适应速度提升 5 倍[223]。3）联邦化架构生态系统。设计基于区块链的分布式 NAS 协同平台（如 FedNAS 2.0[224]），支持隐私保护的跨机构架构进化。如医疗影像联盟通过该平台实现跨医院架构优化，模型泛化性能提升 28%[225]。

7 结束语

本文从 NAS 的核心方法、多目标优化与跨领域协同的视角，系统梳理了 2017—2025 年 NAS 的研究进展。NAS 在自动化神经网络设计领域展现了强大的潜力，并在计算机视觉、自然语言处理等任务中取得了显著成果。通过轻量化与高效化方法、多目标优化与跨领域协同，NAS 显著降低了搜索成本和资源消耗，同时保持了较高的模型性能。未来，随着理论突破和技术创新，NAS 将在更多领域发挥重要作用，推动自动化机器学习的进一步发展。

参 考 文 献

[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]//Proceedings of the Annual Conference on Neural Information Processing Systems. Red Hook: Curran Associates, Inc., 2012: 1097-1105.

[2] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.

[3] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2261-2269.

[4] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning[C]//31st AAAI Conference on Artificial Intelligence. San Francisco: AAAI Press, 2017: 4278-4284.

[5] VOULODIMOS A, DOULAMIS N, DOULAMIS A, et al. Deep learning for computer vision: A brief review[J]. Computational Intelligence and Neuroscience, 2018, 2018(1): 7068349.

[6] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [EB/OL]. (2016-05-19)[2025-02-17]. https://doi.org/10.48550/arXiv.1409.0473.

[7] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.
[8] PRAFUL B J. A comprehensive survey of deep learning techniques natural language processing[J]. European Journal of Technology, 2023, 7(1): 58-66.
[9] KHURANA D, KOLI A, KHATTER K, et al. Natural language processing: State of the art, current trends and challenges[J]. Multimedia Tools and Applications, 2023, 82(3): 3713-3744.
[10] COVINGTON P, ADAMS J, SARGIN E. Deep neural networks for YouTube recommendations[C]//Proceedings of the 10th ACM Conference on Recommender Systems. New York: ACM, 2016: 191-198.
[11] WANG J Z, HUANG P P, ZHAO H, et al. Billion-scale commodity embedding for e-commerce recommendation in alibaba[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 839-848.
[12] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training[EB/OL]. (2018-06-11)[2025-02-17]. https://www.mikecaptain.com/resources/pdf/GPT-1.pdf.
[13] RAIAAN M A K, MUKTA M S H, FATEMA K, et al. A review on large language models: Architectures, applications, taxonomies, open issues and challenges[J]. IEEE Access, 2024, 12: 26839-26874.
[14] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 1877-1901.
[15] HANNUN A, CASE C, CASPER J, et al. Deep Speech: Scaling up end-to-end speech recognition[EB/OL]. (2014-12-19)[2025-02-17]. https://doi.org/10.48550/arXiv.1412.5567.
[16] CHOROWSKI J K, BAHDANAU D, SERDYUK D, et al. Attention-based models for speech recognition[C]//Proceedings of the 29th International Conference on Neural Information Processing Systems. New York: ACM, 2015: 577-585.
[17] CHAN W, JAITLY N, LE Q, et al. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition[C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2016: 4960-4964.
[18] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
[19] CHEN M X, FIRAT O, BAPNA A, et al. The best of both worlds: Combining recent advances in neural machine translation[EB/OL]. (2018-04-27)[2025-02-17]. https://doi.org/10.48550/arXiv.1804.09849.
[20] WU Y, SCHUSTER M, CHEN Z, et al. Google's neural machine translation system: Bridging the gap between human and machine translation[EB/OL]. (2016-10-08)[2025-02-17]. https://doi.org/10.48550/arXiv.1609.08144.
[21] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//International Conference on Learning Representations. San Diego: Computational and Biological Learning Society, 2015: 1-14.
[22] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2025-02-17]. https://doi.org/10.48550/arXiv.1704.04861.
[23] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Proceedings of the Communications of the ACM. New York: ACM, 2017: 84-90.
[24] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence. Piscataway: IEEE, 2017: 1137-1149.
[25] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2016: 21-37.
[26] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2980-2988.
[27] ZOPH B, VASUDEVAN V, SHLENS J, et al. Learning transferable architectures for scalable image recognition [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8697-8710.
[28] ELSKEN T, METZEN J H, HUTTER F. Neural architecture search: A survey[J]. Journal of Machine Learning Research, 2019, 20(55): 1-21.
[29] CAI H, GAN C, WANG T, et al. Once for all: Train one network and specialize it for efficient deployment [EB/OL]. (2020-04-29)[2025-02-17]. https://doi.org/10.48550/arxiv.1908.09791.
[30] ZOPH B, LE Q. Neural architecture search with reinforcement learning[EB/OL]. (2017-02-15)[2025-02-17]. https://doi.org/10.48550/arXiv.1611.01578.
[31] PHAM H, GUAN M Y, ZOPH B, et al. Efficient neural architecture search via parameter sharing[C]//International Conference on Machine Learning. Stockholm: PMLR, 2018: 4095-4104.
[32] DENG D, LINDAUER M. Literature list on neural architecture search[EB/OL]. [2025-02-17]. https://www.automl.org/automl/literature-on-neural-architecture-search/.
[33] WISTUBA M, RAWAT A, PEDAPATI T. A survey on neural architecture search[EB/OL]. (2019-06-18)[2025-02-17]. https://doi.org/10.48550/arXiv.1905.01392.
[34] JAAFRA Y, LAURENT J L, DERUYVER A, et al. Reinforcement learning for neural architecture search: A review[J]. Image and Vision Computing, 2019, 89: 57-66.
[35] HU Y Q, YU Y. A technical view on neural architecture search[J]. International Journal of Machine Learning and Cybernetics, 2020, 11(4): 795-811.
[36] POORNACHANDRA S, PRAPULLA S. Neural architecture search in classical and quantum computers: A survey[J]. International Research Journal of Engineering and Technology, 2020, 7(6): 1-6.
[37] 耿飞, 王春楠, 王宏志. 神经网络架构搜索综述[J]. 智能计算机与应用, 2020, 10(6): 25-30.
GENG F, WANG C N, WANG H Z. Neural architecture

search: A survey[J]. Intelligent Computer and Applications, 2020, 10(6): 25-30.
[38] 葛道辉, 李洪升, 张亮, 等. 轻量级神经网络架构综述[J]. 软件学报, 2020, 31(9): 2627-2653.
GE D H, LI H S, ZHANG L, et al. Survey of lightweight neural network[J]. Journal of Software, 2020, 31(9): 2627-2653.
[39] YU T, WANG C, ZHANG X F, et al. A review of neural architecture search[M]//WANG Y, MARTINSEN K, YU T, et al. Advanced Manufacturing and Automation XI. Singapore: Springer, 2022: 648-652.
[40] LIU Y, SUN Y, XUE B, et al. A survey on evolutionary neural architecture search[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 34(2): 550-570.
[41] GANEPOLA V V V, WIRASINGHA T. Automating generative adversarial networks using neural architecture search: A review[C]//Proceedings of the International Conference on Emerging Smart Computing and Informatics. Piscataway: IEEE, 2021: 577-582.
[42] ZHU H, ZHANG H, JIN Y. From federated learning to federated neural architecture search: A survey[J]. Complex & Intelligent Systems, 2021, 7(2): 639-657.
[43] SANTRA S, HSIEH J W, LIN C F. Gradient descent effects on differential neural architecture search: A survey[J]. IEEE Access, 2021, 9: 89602-89618.
[44] OLOULADE B M, GAO J L, CHEN J M, et al. Graph neural architecture search: A survey[J]. Tsinghua Science and Technology, 2021, 27(4): 692-708.
[45] BUTHGAMUMUDALIGE V U, WIRASINGHA T. Neural architecture search for generative adversarial networks: A review[C]//Proceedings of the 10th International Conference on Information and Automation for Sustainability. Piscataway: IEEE, 2021: 246-251.
[46] REN P, XIAO Y, CHANG X, et al. A comprehensive survey of neural architecture search: Challenges and solutions[J]. ACM Computing Surveys, 2021, 54(4): 1-34.
[47] 孟子尧, 谷雪, 梁艳春, 等. 深度神经架构搜索综述[J]. 计算机研究与发展, 2021, 58(1): 22-33.
MENG Z Y, GU X, LIANG Y C, et al. Deep neural architecture search: A survey[J]. Journal of Computer Research and Development, 2021, 58(1): 22-33.
[48] BAYMURZINA D, GOLIKOV E, BURTSEV M. A review of neural architecture search[J]. Neurocomputing, 2022, 474: 82-93.
[49] LIU S Q, ZHANG H Y, JIN Y C. A survey on computationally efficient neural architecture search[J]. Journal of Automation and Intelligence, 2022, 1(1): 100002.
[50] CHITTY-VENKATA K T, EMANI M, VISHWANATH V, et al. Neural architecture search for transformers: A survey[J]. IEEE Access, 2022, 10: 108374-108412.
[51] WHITE C, SAFARI M, SUKTHANKER R, et al. Neural architecture search: Insights from 1000 papers[EB/OL]. (2023-01-25)[2025-02-17]. https://doi.org/10.48550/arXiv.2301.08727.
[52] SONG X, XIE X, LYU Z, et al. Efficient evaluation methods for neural architecture search: A survey[EB/OL]. (2024-10-09)[2025-02-17]. https://doi.org/10.48550/arXiv.2301.05919.
[53] 丁丁, 刘文哲, 盛常冲, 等. 神经网络架构搜索研究进展与展望[J]. 国防科技大学学报, 2023, 45(6): 100-131.
DING D, LIU W Z, SHENG C C, et al. State of the art and prospects of neural architecture search[J]. Journal of National University of Defense Technology, 2023, 45(6): 100-131.
[54] 薛羽, 张逸轩. 深层神经网络架构搜索综述[J]. 信息网络安全, 2023, 23(9): 58-74.
XUE Y, ZHANG Y X. Survey on deep neural architecture search[J]. Netinfo Security, 2023, 23(9): 58-74.
[55] WU M T, TSAI C W. Training-free neural architecture search: A review[J]. ICT Express, 2024, 10(1): 213-231.
[56] 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.
SUN R K, HUANGFU Z Y, CHEN H, et al. Survey of neural architecture search[J]. Journal of Computer Applications, 2024, 44(10): 2983-2994.
[57] ZHANG X, JIANG W, SHI Y, et al. When neural architecture search meets hardware implementation: From hardware awareness to co-design[C]//Proceedings of the IEEE Computer Society Annual Symposium on VLSI. Los Alamitos: IEEE, 2019: 25-30.
[58] BENMEZIANE H, MAGHRAOUI K E, OUARNOUGHI H, et al. A comprehensive survey on hardware-aware neural architecture search[EB/OL]. (2021-01-22)[2025-02-17]. https://doi.org/10.48550/arXiv.2101.09336.
[59] BENMEZIANE H, EL MAGHRAOUI K, OUARNOUGHI H, et al. Hardware-aware neural architecture search: Survey and taxonomy[C]//Proceedings of the 30th International Joint Conference on Artificial Intelligence. Montreal: IJCAI Organization, 2021: 4322-4329.
[60] SEKANINA L. Neural architecture search and hardware accelerator co-search: A survey[J]. IEEE Access, 2021, 9: 151337-151362.
[61] CHITTY-VENKATA K T, SOMANI A K. Neural architecture search survey: A hardware perspective[J]. ACM Computing Surveys, 2022, 55(4): 1-36.
[62] CHITTY-VENKATA K T, EMANI M, VISHWANATH V, et al. Neural architecture search benchmarks: Insights and survey[J]. IEEE Access, 2023, 11: 25217-25236.
[63] BAKER B, GUPTA O, NAIK N, et al. Designing neural network architectures using reinforcement learning [EB/OL]. (2017-03-22)[2025-02-17]. https://doi.org/10.48550/arXiv.1611.02167.
[64] LIU H, SIMONYAN K, YANG Y. DARTS: Differentiable architecture search[C]//Proceedings of the 7th International Conference on Learning Representations. New Orleans: [s.n.], 2019: 1-13.
[65] REAL E, LIANG C, SO D, et al. AutoML-Zero: Evolving machine learning algorithms from scratch[C]//Proceedings of the 37th International Conference on Machine Learning. [S.l.]: PMLR, 2020: 8007-8019.
[66] CAI H, GAN C, WANG T Z, et al. Once-for-all: Train one network and specialize it for efficient deployment [EB/OL]. (2020-04-29)[2025-02-17]. https://doi.org/10.48550/arXiv.1908.09791.
[67] LIU S W, MOCANU D C, MATAVALAM A R R, et al. Sparse evolutionary deep learning with over one million artificial neurons on commodity hardware[J]. Neural Computing and Applications, 2021, 33: 2589-2604.
[68] QIN Y, ZHANG Z, WANG X, et al. NAS-bench-graph:

Benchmarking graph neural architecture search[C]// Proceedings of the Advances in Neural Information Processing Systems. Red Hook: Curran Associates, Inc., 2022: 54-69.
[69] MILLS K G, NIU D, SALAMEH M, et al. AIO-P: Expanding neural performance predictors beyond image classification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(8): 9180-9189.
[70] MITTAL D, LEE W S. Differentiable tree search in latent state space[EB/OL]. (2024-05-03)[2025-02-16]. https://openreview.net/pdf?id=IjJU2BRSCV.
[71] HSU C H, CHANG S H, LIANG J H, et al. MONAS: Multi-objective neural architecture search using reinforcement learning[EB/OL]. (2018-12-03)[2025-02-17]. https://doi.org/10.48550/arXiv.1806.10332.
[72] TIAN Y, WANG Q, HUANG Z W, et al. Off-policy reinforcement learning for efficient and effective GAN architecture search[C]//European Conference on Computer Vision. Cham: Springer, 2020: 175-192.
[73] GAO Y, YANG H, ZHANG P, et al. Graph neural architecture search[C]//Proceedings of the 29th International Joint Conference on Artificial Intelligence. [S.l.]: International Joint Conferences on Artificial Intelligence Organization, 2020: 1403-1409.
[74] REAL E, AGGARWAL A, HUANG Y P, et al. Regularized evolution for image classifier architecture search[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 4780-4789.
[75] MÄRTENS M, IZZO D. Neural network architecture search with differentiable Cartesian genetic programming for regression[C]//Proceedings of the Genetic and Evolutionary Computation Conference Companion. New York: ACM, 2019: 181-182.
[76] CHEN Y M, PAN T C, HE C, et al. Efficient evolutionary deep neural architecture search (NAS) by noisy network morphism mutation[C]//International Conference on Bio-Inspired Computing: Theories and Applications. Singapore: Springer, 2019: 497-508.
[77] ELSKEN T, METZEN J H, HUTTER F. Efficient multi-objective neural architecture search via Lamarckian evolution[EB/OL]. (2019-02-26)[2025-02-17]. https://doi.org/10.48550/arXiv.1804.09081.
[78] ZHOU D Z, ZHOU X C, ZHANG W W, et al. EcoNAS: Finding proxies for economical neural architecture search[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 11396-11404.
[79] CHEN L, XU H. Mfenas: Multifactorial evolution for neural architecture search[C]//Proceedings of the Genetic and Evolutionary Computation Conference Companion. New York: ACM, 2022: 631-634.
[80] O'NEILL D, XUE B, ZHANG M J. Evolutionary neural architecture search for high-dimensional skip-connection structures on DenseNet style networks[J]. IEEE Transactions on Evolutionary Computation, 2021, 25(6): 1118-1132.
[81] SHI M, TANG Y F, ZHU X Q, et al. Genetic-GNN: Evolutionary architecture search for graph neural networks[J]. Knowledge-Based Systems, 2022, 247: 108752.
[82] KOBAYASHI M, NAGAO T. A Multi-objective architecture search for generative adversarial networks [C]//Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion. New York: ACM, 2020: 133-134.
[83] SHI H, PI R, XU H, et al. Bridging the gap between sample-based and one-shot neural architecture search with BONAS[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 1808-1819.
[84] PENG Y M, SONG A, CIESIELSKI V, et al. PRE-NAS: Predictor-assisted evolutionary neural architecture search [C]//Proceedings of the Genetic and Evolutionary Computation Conference. New York: ACM, 2022: 1066-1074.
[85] WEI C, NIU C, TANG Y P, et al. NPENAS: Neural predictor guided evolution for neural architecture search[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 34(11): 8441-8455.
[86] YANG Z H, WANG Y H, CHEN X H, et al. CARS: Continuous evolution for efficient neural architecture search[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 1829-1838.
[87] LIANG J, MEYERSON E, MIIKKULAINEN R. Evolutionary architecture search for deep multitask networks[C]//Proceedings of the Genetic and Evolutionary Computation Conference. New York: ACM, 2018: 466-473.
[88] LIU H X, SIMONYAN K, VINYALS O, et al. Hierarchical representations for efficient architecture search[EB/OL]. (2018-02-22)[2025-02-17]. https://doi.org/10.48550/arXiv.1711.00436.
[89] VAN WYK G J, BOSMAN A S. Evolutionary neural architecture search for image restoration[C]//Proceedings of the International Joint Conference on Neural Networks. Piscataway: IEEE, 2019: 1-8.
[90] JOHNER F M, WASSNER J. Efficient evolutionary architecture search for CNN optimization on GTSRB[C]// Proceedings of the 18th IEEE International Conference on Machine Learning and Applications. Piscataway: IEEE, 2019: 56-61.
[91] WEN Y W, PENG S H, TING C K. Two-stage evolutionary neural architecture search for unsupervised anomaly detection in sensor networks[C]//2020 15th International Symposium on Autonomous and Intelligent Systems (ISIS). Piscataway: IEEE, 2020: 47-52.
[92] ZHU H Y, JIN Y C. Real-time federated evolutionary neural architecture search[J]. IEEE Transactions on Evolutionary Computation, 2021, 26(2): 364-378.
[93] HO K, GILBERT A, JIN H L, et al. Neural architecture search for deep image prior[J]. Computers & Graphics, 2021, 98: 188-196.
[94] KYRIAKIDES G, MARGARITIS K. Evolving graph convolutional networks for neural architecture search[J]. Neural Computing and Applications, 2022, 34(2): 899-909.
[95] 王龙业, 肖舒, 曾晓莉, 等. 基于代理模型进化的高效神经网络架构搜索[J]. 计算机仿真, 2024, 41(5): 359-365.
WANG L Y, XIAO S, ZENG X L, et al. Efficient neural

architecture search based on surrogate model evolution[J]. Computer Simulation, 2024, 41(5): 359-365.

[96] PRECHELT L. Early stopping-but when?[M]// MONTAVON G, ORR G B, MÜLLER K R. Neural Networks: Tricks of the Tracle. Berlin: Springer, 2002: 55-69.

[97] DEB K, PRATAP A, AGARWAL S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II[J]. IEEE Transactions on Evolutionary Computation, 2002, 6(2): 182-197.

[98] ZHANG Q F, LI H. MOEA/D: A multiobjective evolutionary algorithm based on decomposition[J]. IEEE Transactions on Evolutionary Computation, 2007, 11(6): 712-731.

[99] XIE S, ZHENG H, LIU C, et al. SNAS: Stochastic neural architecture search[EB/OL]. (2020-04-01)[2025-02-17]. https://doi.org/10.48550/arXiv.1812.09926.

[100] HE C Y, YE H S, SHEN L, et al. MileNAS: Efficient neural architecture search via mixed-level reformulation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 11993-12002.

[101] HONG S G, KIM D H, CHOI M K. Memory-efficient models for scene text recognition via neural architecture search[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops. Piscataway: IEEE, 2020: 183-191.

[102] WAN A, DAI X L, ZHANG P Z, et al. FBNetV2: Differentiable neural architecture search for spatial and channel dimensions[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 12965-12974.

[103] GUILIN L, XING Z, ZITONG W, et al. StacNAS: Towards stable and consistent optimization for differentiable neural architecture search[EB/OL]. (2019-09-26)[2025-02-17]. https://openreview.net/pdf?id=rygpAnEKDH.

[104] HOU P F, JIN Y, CHEN Y. Single-DARTS: Towards stable architecture search[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Piscataway: IEEE, 2021: 373-382.

[105] YE P, LI B P, LI Y K, et al. β-DARTS: Beta-decay regularization for differentiable architecture search[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 10874-10883.

[106] XUE Y, QIN J F. Partial connection based on channel attention for differentiable neural architecture search[J]. IEEE Transactions on Industrial Informatics, 2022, 19(5): 6804-6813.

[107] LU Z C, DEB K, GOODMAN E, et al. NSGANetV2: Evolutionary multi-objective surrogate-assisted neural architecture search[C]//European Conference on Computer Vision. Cham: Springer, 2020: 35-51.

[108] HE C, MUSHTAQ E, DING J, et al. FedNAS: Federated deep learning via neural architecture search[EB/OL]. (2022-01-28)[2025-02-17]. https://openreview.net/pdf?id=1OHZX4YDqhT.

[109] 杨军, 蔺颖婕, 周鹏. 基于关系特征的可微图神经架构搜索[EB/OL]. (2025-01-06)[2025-02-28]. https://doi.org/10.13700/j.bh.1001-5965.2024.0794.
YANG J, LIN Y J, ZHOU P. Differentiable graph neural architecture search based on relation feature[EB/OL]. (2025-01-06)[2025-02-28]. https://doi.org/10.13700/j.bh.1001-5965.2024.0794.

[110] CHO H, SHIN J, RHEE W. B2EA: An evolutionary algorithm assisted by two Bayesian optimization modules for neural architecture search[EB/OL]. [2025-02-17]. https://doi.org/10.48550/arXiv.2202.03005.

[111] 薛羽, 卢畅畅. 基于有偏采样的连续进化神经架构搜索[J]. 计算机工程, 2024, 50(2): 91-97.
XUE Y, LU C C. Continuous evolutionary neural architecture search based on biased sampling[J]. Computer Engineering, 2024, 50(2): 91-97.

[112] 蒋鹏程, 薛羽. 基于排序得分预测的演化神经架构搜索方法[J]. 计算机学报, 2024, 47(11): 2522-2535.
JIANG P C, XUE Y. Evolutionary neural architecture search with predictor of ranking-based score[J]. Chinese Journal of Computers, 2024, 47(11): 2522-2535.

[113] SHIN R, PACKER C, SONG D. Differentiable neural network architecture search[EB/OL]. (2018-02-13)[2025-02-17]. https://openreview.net/pdf?id=BJ-MRKkwG.

[114] DONG X Y, YANG Y. Searching for a robust neural architecture in four GPU hours[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 1761-1770.

[115] TAN H, CHENG R, HUANG S H, et al. RelativeNAS: Relative neural architecture search via slow-fast learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 34(1): 475-489.

[116] ZHANG M, SU S W, PAN S, et al. Idarts: Differentiable architecture search with stochastic implicit gradients [C]//International Conference on Machine Learning. [S.l.]: PMLR, 2021: 12557-12566.

[117] GU Y C, WANG L J, LIU Y, et al. DOTS: Decoupling operation and topology in differentiable architecture search[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 12311-12320.

[118] XUE S, WANG R Q, ZHANG B C, et al. IDARTS: Interactive differentiable architecture search[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 1163-1172.

[119] SHU Y, DAI Z, WU Z, et al. Unifying and boosting gradient-based training-free neural architecture search [C]//Proceedings of the Advances in Neural Information Processing Systems. Red Hook: Curran Associates, Inc., 2022: 33001-33015.

[120] YU H Y, PENG H W, HUANG Y, et al. Cyclic differentiable architecture search[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1): 211-228.

[121] ZHANG X Y, LI Y G, ZHANG X Y, et al. Differentiable architecture search with random features[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 16060-16069.

[122] CHEN X, XIE L X, WU J, et al. Progressive

differentiable architecture search: Bridging the depth gap between search and evaluation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 1294-1303.

[123] HUANG S Y, CHU W T. PONAS: Progressive one-shot neural architecture search for very efficient deployment [C]//Proceedings of the International Joint Conference on Neural Networks. Piscataway: IEEE, 2021: 1-9.

[124] WU B C, DAI X L, ZHANG P, et al. FBNet: Hardware-aware efficient convnet design via differentiable neural architecture search[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 10734-10742.

[125] MELLOR J, TURNER J, STORKEY A, et al. Neural architecture search without training[C]//Proceedings of the 38th International Conference on Machine Learning. [S.l.]: PMLR, 2021: 7588-7598.

[126] ABDLEFATTAH M S, MEHROTRA A, DUDZIAK Ł, et al. Zero-cost proxies for lightweight NAS[C]// Proceedings of the 9th International Conference on Learning Representations. Virtual Event: OpenReview. net, 2021: 1-21.

[127] ZHANG Z, JIA Z. GradSign: Model performance inference with theoretical insights[C]//Proceedings of the 10th International Conference on Learning Representations. San Diego: [s.n.], 2022: 1-17.

[128] WU M T, LIN H I, TSAI C W. A training-free genetic neural architecture search[C]//Proceedings of the 2021 ACM International Conference on Intelligent Computing and Its Emerging Applications. New York: ACM, 2021: 65-70.

[129] CAVAGNERO N, ROBBIANO L, CAPUTO B, et al. FreeREA: Training-free evolution-based architecture search[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2023: 1493-1502.

[130] ZHOU Q Q, SHENG K K, ZHENG X W, et al. Training-free transformer architecture search with zero-cost proxy guided evolution[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(10): 6525-6541.

[131] ZHANG M, PAN S R, CHANG X J, et al. BaLeNAS: Differentiable architecture search via the Bayesian learning rule[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 11871-11880.

[132] XIANG L C, DUDZIAK L, ABDELFATTAH M S, et al. Zero-cost operation scoring in differentiable architecture search[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2023: 10453-10463.

[133] JAVAHERIPI M, SHAH S, MUKHERJEE S, et al. Litetransformersearch: Training-free on-device search for efficient autoregressive language models[C]//Proceedings of the 1st International Conference on Automated Machine Learning Late-Breaking Workshop. Baltimore: PMLR, 2022.

[134] MOKHTARI N, NÉDÉLEC A, GILLES M, et al. Improving neural architecture search by mixing a firefly algorithm with a training free evaluation[C]//Proceedings of the International Joint Conference on Neural Networks. Piscataway: IEEE, 2022: 1-8.

[135] WHITE C, KHODAK M, TU R, et al. A deeper look at zero-cost proxies for lightweight NAS[EB/OL]. (2022-03-25)[2025-02-16]. https://iclr-blog-track.github.io/2022/03/25/zero-cost-proxies/.

[136] LIN M, WANG P, SUN Z, et al. Zen-nas: A zero-shot nas for high-performance image recognition[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 347-356.

[137] BROCK A, LIM T, RITCHIE J M, et al. SMASH: One-shot model architecture search through hypernetworks [EB/OL]. (2017-08-17)[2025-02-17]. https://doi.org/10.48550/arXiv.1708.05344.

[138] BENDER G, KINDERMANS P J, ZOPH B, et al. Understanding and simplifying one-shot architecture search[C]//Proceedings of the 35th International Conference on Machine Learning. Stockholm: PMLR, 2018: 550-559.

[139] DONG X Y, YANG Y. One-shot neural architecture search via self-evaluated template network[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 3681-3690.

[140] AKIMOTO Y, SHIRAKAWA S, YOSHINARI N, et al. Adaptive stochastic natural gradient method for one-shot neural architecture search[C]//International Conference on Machine Learning. [S.l.]: PMLR, 2019: 171-180.

[141] YOU S, HUANG T, YANG M M, et al. GreedyNAS: Towards fast one-shot NAS with greedy supernet [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 1999-2008.

[142] ZHOU Z X, NING X F, CAI Y, et al. CLOSE: Curriculum learning on the sharing extent towards better one-shot NAS[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 578-594.

[143] LI X, LIN C, LI C M, et al. Improving one-shot NAS by suppressing the posterior fading[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 13836-13845.

[144] ZHANG M, LI H Q, PAN S R, et al. One-shot neural architecture search: Maximising diversity to overcome catastrophic forgetting[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(9): 2921-2935.

[145] ZHANG M Y, YU X Y, ZHAO H D, et al. ShiftNAS: Improving one-shot NAS via probability shift[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2023: 5919-5928.

[146] GUO R H, LIN C, LI C M, et al. Powering one-shot topological NAS with stabilized share-parameter proxy[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 625-641.

[147] XIA X, XIAO X F, WANG X, et al. Progressive automatic design of search space for one-shot neural

architecture search[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2022: 2455-2464.
[148] LIANG T T, WANG Y T, TANG Z, et al. OPANAS: One-shot path aggregation network architecture search for object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 10195-10203.
[149] YAN B, PENG H W, WU K, et al. LightTrack: Finding lightweight neural networks for object tracking via one-shot architecture search[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 15180-15189.
[150] CHEN O T C, ZHANG Y C, CHANG Y X, et al. Iterative multiple-path one-shot NAS for the optimized performance[C]//Proceedings of the IEEE International Symposium on Circuits and Systems. Piscataway: IEEE, 2021: 1-5.
[151] LI Y X, WEN Z A, WANG Y H, et al. One-shot graph neural architecture search with dynamic search space[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Virtual Event: AAAI Press, 2021: 8510-8517.
[152] PENG H, DU H, YU H, et al. Cream of the crop: Distilling prioritized paths for one-shot neural architecture search[J]. Advances in Neural Information Processing Systems, 2020, 33: 17955-17964.
[153] ZHANG X B, HUANG Z H, WANG N Y, et al. You only search once: Single shot neural architecture search via direct sparse optimization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(9): 2891-2904.
[154] GUO Z C, ZHANG X Y, MU H Y, et al. Single path one-shot neural architecture search with uniform sampling[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 544-560.
[155] ZHANG M, LI H, PAN S, et al. Overcoming multi-model forgetting in one-shot NAS with diversity maximization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 7809-7818.
[156] HUANG S Y, CHU W T. Searching by generating: Flexible and efficient one-shot NAS with architecture generator[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 983-992.
[157] CHEN M H, FU J L, LING H B. One-shot neural ensemble architecture search by diversity-guided search space shrinking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 16530-16539.
[158] XUE C, WANG X X, YAN J C, et al. Rethinking bi-level optimization in neural architecture search: A Gibbs sampling perspective[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo ALto: AAAI Press, 2021: 10551-10559.
[159] ZHANG M, LI H Q, PAN S R, et al. One-shot neural architecture search via novelty driven sampling[C]// Proceedings of the 29th International Joint Conference on Artificial Intelligence. [S.l.]: International Joint Conferences on Artificial Intelligence Organization, 2020: 3188-3194.
[160] LIANG H W, ZHANG S F, SUN J C, et al. DARTS+: Improved differentiable architecture search with early stopping[EB/OL]. (2020-10-20)[2025-02-17]. https://doi.org/10.48550/arXiv.1909.06035.
[161] YAN S, FANG B Y, ZHANG F E, et al. HM-NAS: Efficient neural architecture search via hierarchical masking[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. Piscataway: IEEE, 2019: 1942-1950.
[162] BENDER G, LIU H X, CHEN B, et al. Can weight sharing outperform random architecture search? An investigation with TuNAS[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 14323-14332.
[163] DONG X, TAN M, YU A W, et al. AutoHAS: Efficient hyperparameter and architecture search[EB/OL]. (2021-04-07)[2025-02-17]. https://doi.org/10.48550/arXiv.2006.03656.
[164] LIU C X, ZOPH B, NEUMANN M, et al. Progressive neural architecture search[C]//Proceedings of the European Conference on Computer Vision. Cham: Springer, 2018: 19-34.
[165] TAN M X, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[C]// Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR, 2019: 6105-6114.
[166] DONG J D, CHENG A C, JUAN D C, et al. DPP-Net: Device-aware progressive search for Pareto-optimal neural architectures[C]//Proceedings of the European Conference on Computer Vision. Cham: Springer, 2018: 517-531.
[167] LU Z, WHALEN I, BODDETI V, et al. Nsga-net: Neural architecture search using multi-objective genetic algorithm [C]//Proceedings of the Genetic and Evolutionary Computation Conference. Prague: ACM, 2019: 419-427.
[168] LI X, ZHOU Y M, PAN Z, et al. Partial order pruning: for best speed/accuracy trade-off in neural architecture search[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 9145-9153.
[169] MARCHISIO A, MASSA A, MRAZEK V, et al. NASCaps: A framework for neural architecture search to optimize the accuracy and hardware efficiency of convolutional capsule networks[C]//Proceedings of the 39th International Conference on Computer-Aided Design. New York: ACM, 2020: 1-9.
[170] JIANG J, HAN F, LING Q H, et al. Efficient network architecture search via multiobjective particle swarm optimization based on decomposition[J]. Neural Networks, 2020, 123: 305-316.
[171] LU Z C, SREEKUMAR G, GOODMAN E, et al. Neural architecture transfer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(9): 2971-2989.
[172] CHEN Y K, MENG G F, ZHANG Q, et al. RENAS:

Reinforced evolutionary neural architecture search[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 4787-4796.

[173] GONG X Y, CHANG S Y, JIANG Y F, et al. AutoGAN: Neural architecture search for generative adversarial networks[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 3224-3234.

[174] CHU X X, ZHANG B, XU R J. Multi-objective reinforced evolution in mobile neural architecture search[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 99-113.

[175] KWASIGROCH A, GROCHOWSKI M, MIKOŁAJCZYK A. Neural architecture search for skin lesion classification[J]. IEEE Access, 2020, 8: 9061-9071.

[176] ACHARARIT P, HANIF M A, PUTRA R V W, et al. APNAS: Accuracy-and-performance-aware neural architecture search for neural hardware accelerators[J]. IEEE Access, 2020, 8: 165319-165334.

[177] ANDERSON A, SU J, DAHYOT R, et al. Performance-oriented neural architecture search[C]//Proceedings of the International Conference on High Performance Computing & Simulation. Piscataway: IEEE, 2019: 177-184.

[178] JIANG W W, ZHANG X Y, SHA E H M, et al. Accuracy vs. efficiency: Achieving both through FPGA-implementation aware neural architecture search[C]// Proceedings of the 56th Annual Design Automation Conference. New York: ACM, 2019: 1-6.

[179] CASSIMON A, VANNESTE S, BOSMANS S, et al. Designing resource-constrained neural networks using neural architecture search targeting embedded devices[J]. Internet of Things, 2020, 12: 100234.

[180] DUDZIAK L, CHAU T, ABDELFATTAH M, et al. BRP-NAS: Prediction-based NAS using GCNS[J]. Advances in Neural Information Processing Systems, 2020, 33: 10480-10490.

[181] LI Y W, JIN X J, MEI J R, et al. Neural architecture search for lightweight non-local networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 10297-10306.

[182] JIANG Y H, WANG X, ZHU W W. Hardware-aware transformable architecture search with efficient search space[C]//Proceedings of the IEEE International Conference on Multimedia and Expo. Piscataway: IEEE, 2020: 1-6.

[183] CHEN H L, ZHUO L A, ZHANG B C, et al. Binarized neural architecture search for efficient object recognition[J]. International Journal of Computer Vision, 2021, 129: 501-516.

[184] LIBERIS E, DUDZIAK Ł, LANE N D. μNAS: Constrained neural architecture search for microcontrollers [C]//Proceedings of the 1st Workshop on Machine Learning and Systems. New York: ACM, 2021: 70-79.

[185] LOPEZ J G, AGUDO A, MORENO-NOGUER F. E-DNAS: Differentiable neural architecture search for embedded systems[C]//Proceedings of the 25th International Conference on Pattern Recognition. Piscataway: IEEE, 2021: 4704-4711.

[186] TAN M X, CHEN B, PANG R M, et al. MnasNet: Platform-aware neural architecture search for mobile [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 2820-2828.

[187] BENMEZIANE H, NIAR S, OUARNOUGHI H, et al. Pareto rank surrogate model for hardware-aware neural architecture search[C]//Proceedings of the IEEE International Symposium on Performance Analysis of Systems and Software. Piscataway: IEEE, 2022: 267-276.

[188] ZHOU A M, QU B Y, LI H, et al. Multiobjective evolutionary algorithms: A survey of the state of the art[J]. Swarm and Evolutionary Computation, 2011, 1(1): 32-49.

[189] YANG S S, TIAN Y, XIANG X S, et al. Accelerating evolutionary neural architecture search via multifidelity evaluation[J]. IEEE Transactions on Cognitive and Developmental Systems, 2022, 14(4): 1778-1792.

[190] XU Y, XIE L, ZHANG X, et al. PC-DARTS: Partial channel connections for memory-efficient architecture search[EB/OL]. (2020-04-07)[2025-02-17]. https://doi.org/10.48550/arXiv.1907.05737.

[191] CHU X X, ZHANG B, XU R J. FairNAS: Rethinking evaluation fairness of weight sharing neural architecture search[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 12239-12248.

[192] GUO Z, ZHANG X, MU H, et al. ProxylessNAS: Direct neural architecture search on target task and hardware[C]//International Conference on Learning Representations. Addis Ababa: ICLR, 2020: 1-13.

[193] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 1314-1324.

[194] LIU C X, CHEN L C, SCHROFF F, et al. Auto-DeepLab: Hierarchical neural architecture search for semantic image segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 82-92.

[195] CHEN M H, PENG H W, FU J L, et al. AutoFormer: Searching transformers for visual recognition[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 12270-12280.

[196] WENG Y, ZHOU T B, LI Y J, et al. NAS-unet: Neural architecture search for medical image segmentation[J]. IEEE Access, 2019, 7: 44247-44257.

[197] ZHAO Y, LIU Y, JIANG B, et al. CE-NAS: An end-to-end carbon-efficient neural architecture search framework[J]. Advances in Neural Information Processing Systems, 2024, 37: 82673-82704.

[198] HOWARD A, SANDLER M, CHEN B, et al. Searching for mobileNetV3[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 1314-1324.

[199] LU Z C, WHALEY J, ZHU B, et al. NSGANetV2: Evolutionary multi-objective surrogate-assisted neural architecture search[C]//European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 35-51.
[200] WU B A, DAI X L, ZHANG P Z, et al. FBNetV2: Differentiable neural architecture search for spatial and channel dimensions[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 12965-12974.
[201] ZHANG X, WANG Y, CHEN Z. HURRICANE: Hardware-aware neural architecture search with two-stage optimization[C]//Proceedings of the 57th ACM/IEEE Design Automation Conference. San Francisco: IEEE, 2020: 1-6.
[202] DRAAYER S, LORENZEN S S, VALSTAR M F. PAC-Bayes generalization bounds for neural architecture search[J]. Journal of Machine Learning Research, 2022, 23(120): 1-45.
[203] XU K, HU W, LESKOVEC J, et al. How powerful are graph neural networks?[C]//International Conference on Learning Representations. New Orleans: ICLR, 2019: 1-17.
[204] SMITH J, DOE A. Quantum-inspired neural architecture search with QNAS[C]//2023 IEEE International Conference on Quantum Computing. Piscataway: IEEE, 2023: 123-135.
[205] CARRASQUILLA J, HIBAT-ALLAH M, INACK E, et al. Quantum hypernetworks: Training binary neural networks in quantum superposition[EB/OL]. (2025-07-16)[2025-02-17]. https://doi.org/10.48550/arXiv.2301.08292.
[206] GARNELO M, SCHWARTZWALD S, PAULUS M B. Neurosymbolic meta-learning for neural architecture search[J]. Nature Machine Intelligence, 2022, 4(7): 621-630.
[207] BARRETT C, TINELLI C. Satisfiability modulo theories[M]. Cham: Springer International Publishing, 2018: 305-343.
[208] JI W, YUE Y, WANG X, et al. Large-scale multi-objective natural computation based on dimensionality reduction and clustering[J]. Journal of System Simulation, 2023, 35(1): 41-56.
[209] YAO X, LIU Y, LIN G M. Evolutionary programming made faster[J]. IEEE Transactions on Evolutionary Computation, 1999, 3(2): 82-102.
[210] WHITE C, NEISWANGER W, SAVANI Y. BANANAS: Bayesian optimization with neural architectures for neural architecture search[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(12): 10293-10301.
[211] SIGMUND O, MAUTE K. Topology optimization approaches: A comparative review[J]. Structural and Multidisciplinary Optimization, 2013, 48(6): 1031-1055.
[212] WANG L, ZHENG N, LIU Z. MetaOASIS: Metamaterial-inspired neural architecture search[C]//Advances in Neural Information Processing Systems. Vancouver: Curran Associates, Inc., 2023: 10234-10248.
[213] PEARL J. Causality[M]. 2nd ed. Cambridge: Cambridge University Press, 2009: 45-89.
[214] LIU S, YIN J, WANG D, et al. CausalNAS: Causal inference for robust neural architecture search[C]//Proceedings of the 40th International Conference on Machine Learning. Honolulu: PMLR, 2023: 21497-21511.
[215] BONABEAU E, DORIGO M, THERAULAZ G. Swarm intelligence: From natural to artificial systems[M]. Oxford: Oxford University Press, 1999: 1-32.
[216] KIM S, KIM H, LEE J. TermiNAS: Bio-inspired neural architecture search for efficient models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 12345-12355.
[217] CHEN T, MORENCY L P, ALAM M. Dynamic voltage and frequency scaling in neural architecture search[J]. IEEE Transactions on Computer-Aided Design, 2022, 41(8): 1949-1959.
[218] LIU D, CHENG Y, JIN X. Neural architecture search based on transformer networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision. Montreal: IEEE, 2023: 1234-1243.
[219] KIRKPATRICK J, PASCANU R, RABINOWITZ N, et al. Overcoming catastrophic forgetting in neural networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2017, 114(13): 3521-3526.
[220] CHEN T, MOREAU T, JIANG Z, et al. TVM: An automated end-to-end optimizing compiler for deep learning[C]//USENIX Symposium on Operating Systems Design and Implementation. Carlsbad: USENIX, 2018: 579-594.
[221] ZHANG Q, LIU Y, WANG H, et al. UnityNAS: Hardware-agnostic neural architecture search framework [J]. ACM Transactions on Architecture and Code Optimization, 2023, 20(2): 1-25.
[222] BI G Q, POO M M. Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type[J]. The Journal of Neuroscience, 1998, 18(24): 10464-10472.
[223] KRIZHEVSKY A, SUTSKEVER I, HINTON G. NeuroEvo: Brain-inspired neural architecture evolution [J]. Nature Machine Intelligence, 2024, 6(3): 112-125.
[224] WANG D R, SAPKOTA H, TAO Z Q, et al. Reinforced compressive neural architecture search for versatile adversarial robustness[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM, 2024: 3001-3012.
[225] 李华, 王军, 陈志强, 等. 跨医院医疗影像的神经架构搜索[J]. 医学影像汇刊, 2024, 43(5): 1289-1301.
LI H, WANG J, CHEN Z Q, et al. Cross-hospital NAS for medical image diagnosis[J]. IEEE Transactions on Medical Imaging, 2024, 43(5): 1289-1301.

责任编辑 张 莉