# 基于人工经验网络架构为初始化的 NAS 算法

包振山, 秘博闻, 张文博 (北京工业大学信息学部, 北京 100124)

摘 要:为了解决神经架构搜索(neural architecture search, NAS)算力要求高、搜索耗时长等缺陷,结合深度神经网络的人工设计经验,提出基于人工经验网络架构初始化的 NAS 算法. 该算法对搜索空间进行了重新设计,选取 VGG-11 作为初始架构,有效减少了由参数的随机初始化带来的无效搜索. 基于上述设计方案,在图像分类经典数据集 Cifar-10 上进行了实验验证,经过仅 12 h 的搜索便获得 VGG-Lite 架构,其错误率低至 2.63%,参数量为 1.48 M. 比现阶段性能最佳的人工设计结构 DenseNet-BC 错误率低0.83%,参数量减少至 DenseNet-BC 的 1/17. 结果表明,该方法可以搜索到优秀的网络架构并显著提高搜索效率,对 NAS 算法的普及有着重要的意义.

关键词:神经架构搜索(NAS);自动化深度学习;人工网络架构设计;深度学习;强化学习;图像分类
 中图分类号:TP183
 文献标志码:A
 文章编号:0254-0037(2021)08-0854-09
 doi; 10.11936/bjutxb2021010004

# NAS Algorithm Based on Manual Experience Network Architecture Initialization

BAO Zhenshan, MI Bowen, ZHANG Wenbo

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

**Abstract**: To solve the shortcomings of neural architecture search (NAS), such as high computing power requirements and long search time, combined with the manual design experience of deep neural networks, an NAS algorithm based on manual experience network architecture initialization was proposed. The algorithm redesigned the search space and selected VGG-11 as the initial architecture, which effectively reduced the invalid search caused by the random initialization of the parameters. Based on the above design scheme, experimental verification was carried out on the classic image classification dataset Cifar-10. The VGG-Lite structure was obtained by searching for 12 hours, and the error rate of this model was 2. 63%. The model VGG-Lite was 0. 83% more accurate than DenseNet-BC, the best-performing artificial design structure at this stage. The number of parameters of this architecture was 1. 48 M, which was about 1/17 of the DenseNet-BC number of parameters. Results show that this method can search for excellent network architectures and significantly improve the search efficiency, which is of great significance to the popularization of NAS algorithms.

Key words: neural architecture search (NAS); Auto-DL; manual architecture design; deep learning; reinforcement learning; image classification

收稿日期: 2021-01-04

基金项目:国家重点研发计划资助项目(2017YFC0803300);北京市教育委员会资助项目(KM201910005027)

作者简介:包振山(1965—),男,副教授,主要从事计算机体系架构、深度学习模型优化方面的研究, E-mail: baozhenshan@ bjut. edu. cn

通信作者:张文博(1980—),女,讲师,主要从事异构计算、深度学习模型优化方面的研究, E-mail: zhangwenbo@ bjut. edu. en

神经架构搜索(neural architecture search, NAS)<sup>[1]</sup>是一类使用计算机自动搜索神经网络架构 的算法.如今,NAS 已实现了在计算机视觉 (computer vision,CV)领域内对于通用数据集(如 ImageNet、COCO等)的高精度架构搜索<sup>[2-5]</sup>.然而对 于解决实际问题,却受到搜索耗时长、算力要求高等 因素的制约,难以进行大规模应用.该问题已经成 为 NAS 领域内的热门话题<sup>[6-7]</sup>.经调研发现,NAS 算法效率低下的主要原因来自于搜索初始化时的随 机过程.该过程引起大量无效搜索,这些无效的网 络结构导致了搜索效率的降低.为了减少初始化过 程中的不稳定因素,本文提出一种优化方法:将基于 人工经验的网络架构作为 NAS 搜索过程的起点,以 降低初始化过程中的不稳定性.

基于人工经验的网络架构的优势在于面对领域 内的经典问题已有成熟的解决方案<sup>[8-9]</sup>.其可以取 代 NAS 中随机的初始化过程,从而提高搜索效率. 本工作将基于人工设计的经典网络架构 VGG-11<sup>[10]</sup> 作为自动化搜索的网络结构的起点.

NAS 通常选取目标网络的准确率(在平均准确 率较高时的错误率)以及搜索算法的耗时作为算法 性能的评价指标.本文通过对目标数据集 Cifar-10<sup>[11]</sup>的搜索,使用单张 GTX 1080Ti 作为计算单元, 仅用时 12 h,便搜索得到 VGG-Lite 架构,错误率仅 为 2.63%.该方法显著地提高了算法的搜索效率, 这对 NAS 算法的普及有着重要的意义.

# 1 算法优化方法

# 1.1 NAS 算法

神经架构搜索起源于 2016 年 Zoph 等<sup>[1]</sup>使用强 化学习<sup>[12]</sup>算法进行自动化搜寻网络架构的工作,是 自动化深度学习(auto-deep learning, Auto-DL)<sup>[13]</sup>领 域内一个研究分支. 2018 年 Elasken 等<sup>[14]</sup>,依据当 时已存在的 NAS 工作,将 NAS 算法分为 3 个组成部 分:搜索空间(search space)、搜索策略(search strategy)以及性能评价策略(performance estimation strategy),如图 1 所示. 搜索空间是人为定义的包含 大量候选架构的集合. 搜索策略是在搜索空间上进 行采样并组合候选网络架构的方法. 常用的算法包 括强化学习算法(reinforcement learning, RL)<sup>[12]</sup>、进 化算法(evolution algorithm, EA)<sup>[15]</sup>、基于梯度的算 法(gradient)等. 性能评价策略负责对候选网络进 行性能评估. 候选网络架构被送入性能评估阶段进 行训练以及测试,得到候选网络的性能指标. 这些 性能指标作为搜索策略的输入信号,对算法参数进行调整,再进行新一次的生成.整个循环执行多次 直到算法收敛,算法的输出结果为此过程中性能最 优秀的架构.所以针对 NAS 算法的研究通过对其 中的一项或者多项进行优化,从而提升算法的整体 性能.



Fig. 1 General procedure of NAS

本文从 NAS 算法 3 个部分中搜索空间入手,通 过对搜索空间的有效限定,从而减少无效网络的生成,以达到提升搜索效率的目的.

大量文献[1-7,14,16]表明主流 NAS 算法在搜 索初期大多使用随机参数的初始化来进行网络架构 的生成.这种方式排除了搜索过程中人为因素的影 响,但是也导致搜索策略需要花费较长的时间才能 搜索到性能优异的网络架构.这对于拥有大量计算 资源的企业或者研究机构来说并不构成影响,但对 于大多数无法掌握大量计算资源的研究者或使用者 来说,网络的自动化搜索往往会因此受限.

经过近 10 年的研究,图像分类问题作为深度学 习领域内最经典的问题之一,已在人工网络设计领 域取得了许多优秀的成果<sup>[11,17-22]</sup>.这些成果体现在 网络结构的设计中.优秀的人工网络设计可以改善 NAS 搜索初期存在大量无效架构搜索的问题.所以 本工作将两者结合,让人工设计的架构作为 NAS 搜 索算法的初始状态,既对搜索进行了加速,又能确保 产生性能优异的架构.

# 1.2 分类网络架构的人工设计经验

自 2012 年 AlexNet<sup>[17]</sup> 在 ILSVRC (ImagineNet Large-Scale Visual Recognition Challenge)图像分类 任务中取得突破性成功之后,人工设计的深度神经 网络在图像分类领域内取得了巨大的进步<sup>[89]</sup>.历年的 ILSVRC 获奖网络结构也是深度学习研究者对 网络架构总结归纳的对象.历年获奖网络结构的具体信息如表1所示.

表1中VGG架构作为一个代表性的架构,受到 了广泛的关注.一方面VGG架构总结了AlexNet以 表 1 ILSVRC 2012—2017 年分类任务获奖网络架构 Table 1 ILSVRC 2012—2017 classified task award winning architecture

	8		
年份	架构	top5 错误率/%	结构类型
2012	AlexNet <sup>[17]</sup>	16.42	链式结构
2013	ZFNet <sup>[18]</sup>	13. 51	链式结构
2014	VGG <sup>[10]</sup>	6.80	链式结构
2014	GoogLeNet <sup>[19]</sup>	6.67	细胞结构
2015	$\operatorname{ResNet}^{[20]}$	3. 57	细胞结构
2016	${\rm ResNeXt}^{[21]}$	3.03	细胞结构
2017	SENet <sup>[22]</sup>	2.25	细胞结构

及 ZFNet 的网络的经验,通过对网络深度的加深构

建出了性能最佳的纯链式网络架构,其网络结构如图2所示.另一方面 VGG 架构中小尺寸滤波器连续堆叠,以及使用2×2最大池化层倍缩特征图的方法,为后续 ResNet 等优秀结构所借鉴沿用. VGG 架构中所包含的人工设计经验在分类任务中具有很强借鉴意义,因此本工作选用 VGG 架构作为 NAS 算法的初始架构.

### 1.3 目标数据集

实验所用数据集为 Cifar-10 和 Cifar-100. Cifar-10 和 Cifar-100 均为带标签的数据集,是 80 Million Tiny Image 科学数据集的 2 个子集,由深度学习科 学家 Alex Krizhevsky、Vinod Nair 和 Geoffrey Hinton 收集创建.







如图 3 所示, Cifar-10 数据集由 60 000 张图片 组成,60 000 张图片分为 10 类,每类包含 6 000 张图 片.对于每个类而言,5 000 张作为训练集数据, 1 000 张作为测试集数据,每张图片的大小为32 × 32 像素,图片为 RGB 三通道彩色图片.类与类之间数 据相互不重叠.

Cifar-100 数据集与 Cifar-10 不同之处在于 Cifar-100 拥有 100 类,每类包含 600 张图片,500 张 为训练图像和 100 张为测试图像.

Cifar-10 作为图像分类问题常用科学数据集,在 该数据集上训练的准确率常被用作基于人工设计网 络性能的一个重要评价标准.与此同时,大部分基 于 NAS 方法搜索出来的网络结构也以 Cifar-10 作为 主要的评判标准.所以在 Cifar-10 数据集上既可以 将搜索出来的网络架构与人工设计的深度神经网络 进行性能对比,又可以与 NAS 方法搜索出来的架构 进行比较.同时,本工作使用 Cifar-100 数据集对于 搜索算法得到的最优的网络架构进行训练,来评估 此网络的泛化情况.

### 1.4 搜索空间的建立

使用 VGG-11 架构作为网络的起始搜索架构, 设定在一个定长的链式结构下对搜索空间内的网络 结构进行搜索,有助于进一步缩小网络空间的范围.

遵循 VGG-11 的链式结构模型,网络中前一层 的输出作为后一层的输入,并且数据流向遵循从小 序号结点方向流向大序号结点的单一方向.整个网 络结构表示为所有节点组合的合集,即

$$\mathbf{A} = L_n L_{n-1} L_{n-2} \cdots L_1 \tag{1}$$

式中:A 代表候选架构;L 代表每层中的参数;角标 n 代表对应的层数.



图 3 Cifar-10 的图片分类以及图片示例<sup>[11]</sup> Fig. 3 Categories of data sets and picture instances of Cifar-10<sup>[11]</sup>

将目前高度流行的深度神经网络计算模块作为 搜索的候选集合,这些操作包括卷积、池化、激活以 及零操作(ZeroOp).这其中卷积操作包括 $\{1 \times 1 \\$  conv、3 × 3 conv、3 × 3 separable conv、5 × 5 separable conv、ZeroOp $\}$ ,对于卷积操作的层,同时搜索该层滤 波器的数量,数量的范围是 $\{8,16,32,64,128,256,$ 512 $\}$ ,池化操作包括 $\{2 \times 2 \text{ average pooling}, 3 \times 3 \\$  average pooling、3 × 3 max pooling、5 × 5 max pooling、 ZeroOp $\}$ ,激活操作包括 $\{\text{relu},\text{sigmoid},\text{ZeroOp}\}$ ,如 果本层零操作被搜索到,则表示此层将被跳过,零操 作也存在于其他 3 种搜索空间中.遵循 VGG-11 的 网络结构,对各层进行网络超参数的搜索.

图 4 为网络层数 n = 6 时,由控制器所生成的网络结构.控制器生成的网络架构由字符串表示,其内容包括从图像输入到分类方法之间的参数.控制器对于每层生成 2 个参数:一个为动作参数,一个为数量参数.如果此层的搜索结果为卷积操作,则对数量参数继续进行搜索;如果此层搜索到的结果为非卷积参数,则默认此层的数量参数为0.由控制器

生成的字符串被编码成网络结构,生成的网络结构 进入性能评估阶段进行训练以及验证.



#### 1.5 搜索算法

本工作在搜索算法方面使用强化学习的方法, 该算法的流程如图 5 所示.强化学习算法中的控制 器由长短时记忆(long short term memory, LSTM)<sup>[23]</sup> 网络构成,该网络结构负责生成网络模型字符串序 列.控制器根据概率生成候选网络架构模型,将其





放入性能评估阶段得到该候选网络架构的准确度 R,R 作为强化学习中的奖励信号对 LSTM 控制器进 行更新,如此不断迭代直至收敛,最终找到合适的网 络结构.

对于控制器内部的参数  $\theta_{e}$  更新方法,本工作使 用模型公式

$$J(\theta_{\rm c}) = E_{P(a_{1;T};\theta_{\rm c})}[R]$$
(2)

控制器每次对网络进行生成可看成一个动作, 该动作可被记为 *a*<sub>1:*T*</sub>,其中 *T* 是要预测的超参数的 数量. 当模型收敛时其在验证集上的精度是 *R*. 使 用 *R* 来作为强化学习的奖励信号,通过调整参数 *θ*<sub>e</sub> 来最大化 *R* 的期望.

由于 R 不可导,本文使用 Williams 等<sup>[24]</sup>提出的 REINFORCE 方法,来进行参数  $\theta_{e}$  的更新,有

$$\nabla_{\theta_{c}} J(\theta_{c}) = \sum_{t=1}^{T} E_{P(a_{1:T};\theta_{c})} \left[ \nabla_{\theta_{c}} \log P(a_{t} | a_{(t-1):1};\theta_{c}) R \right]$$
(3)

# 2 实验配置及结果分析

#### 2.1 实验概要

本实验环境部署在一台仅含单张 GTX-1080Ti 的服务器上,算法所使用的库以及对应版本分别为 pytorch 1.0.0 版本以及 torchvision 0.3.0 版本. 实 验分为4个部分,分别为数据预处理、候选网络初 筛、网络生成阶段以及网络精调阶段.

# 2.2 数据预处理

数据预处理是一种提高网络训练准确率的方法.通过对于训练数据进行例如旋转、缩放、遮挡、 裁剪等操作,减少训练中过拟合出现的概率,从而提 高网络训练的准确率.

数据预处理的过程如图 6 所示,具体过程如下: 首先,将训练集中全部图片等比例缩放至 8 × 8 像素 到 64 × 64 像素之间的某一尺寸,得到的数据集被命 名为预裁数据集.其次,若预裁数据集中图片的尺寸 大于 32 × 32 像素,对每张图片随机整块裁取32 × 32 像素的区域;若预裁数据集中的图片小于 32 × 32 像 素,对于图像的边沿部分用图像的平均 RGB 颜色进 行填充,经过此流程得到的数据集称之为裁剪数据 集.最终,裁剪数据集中随机选取半数图像进行水平 翻转并代替原有图像,得到进行训练的最终数据集.





在数据预处理过程中,图像缩放的尺寸是影响测 试准确率的重要因素.为保证试验中最优缩放尺寸 具有广泛代表性,本文选用了3种基于人工设计的网 络结构(AlexNet、VGG-11、ResNet-18)同时进行实验, 测试结果如图7所示.由结果可知,对图像进行等比 例缩小(小于32×32像素)以及大幅度的等比例放大 (大于40×40像素)会降低网络训练的准确率.对于 训练图像略微等比例放大有助于提升训练过程的准 确率.同时,3种网络架构均在图像尺寸等于40×40 像素时得到了最高的准确率,所以,本实验选择40× 40像素作为训练数据的缩放尺寸.

#### 2.3 候选网络初筛

在网络搜索过程中存在性能较差的网络架构, 本工作设定初筛参数来提升网络搜索阶段的效率.





根据 Tan 等<sup>[4]</sup>提出的方法,如果候选网络架构在前20个训练周期内没有达到60%的准确率,则其训练将被停止.如图8所示,就结果而言83%的架构通过了网络架构初筛的过程.



图 8 初筛阶段网络合格率

Fig. 8 Network qualification rate at the preliminary screening stage

#### 2.4 网络生成阶段

在经过了网络初筛阶段之后,用10个通过了初 筛的候选架构(命名为C1~C10)进行实验,寻找候 选架构合适的训练轮次数.如图9所示,经实验发 现在经历了100轮次训练之后,网络的准确率进入 到了缓慢的爬坡期,候选网络的相对排名几乎不再 变化.于是对于通过初筛阶段的候选网络进行100 轮次的训练,得到候选网络架构的相对准确率排名.







#### 2.5 网络精调阶段

根据生成阶段网络训练的相对排名,本工作将 10 个在生成阶段中准确率最高的架构进行网络精 调.根据 Zhou 等<sup>[7]</sup>所提出的训练方法,生成阶段以

#### 及精调阶段的网络训练配置细节如表2所示.

表 2 网络训练参数			
Table 2         Network training parameters			
配置参数	生成阶段	精调阶段	
优化器	SGD	SGD	
学习率	0. 1	0.1(0~60 epoch), 0.02(61~120 epoch), 4×10 <sup>-3</sup> (121~160 epoch), 8×10 <sup>-4</sup> (161~200 epoch)	
动量方法	0.9	0. 9	
权值衰减	$5 \times 10^{-4}$	$5 \times 10^{-4}$	
批大小	64	128	
训练轮次	100	200~收敛	

# 2.6 实验结果及分析

根据实验配置以及四部分的实验流程,经过12h 搜索得到性能最佳的网络结构 VGG-Lite,现将此 网络结构的性能与现阶段图像分类问题中性能优秀 的网络结构进行对比.

表3为 Cifar-10数据集下各网络架构的性能指标.选取2种基于人工设计的网络架构,分别为 VGG 类型准确率最高的 VGG-19架构以及现阶段基于人工设计准确率最高的 DenseNet-BC 架构进行对比.对于使用 NAS 算法搜索出来的网络架构,本工作选择了3种使用强化学习搜索策略的架构

表 3 Cifar-10 数据集训练结果 Table 3 Comparison on Cifar-10 fully trained

	•		v	
网纹加劫	错误率/	耗时/	参数	搜索
网络朱钩	%	$\text{GPUD}^{\textcircled{1}}$	量/M	方法
VGG-19 <sup>[10]</sup>	6.80		39.0	$Manual^{\textcircled{D}}$
$DenseNet\text{-}BC^{[25]}$	3.46		25.6	Manual
NAS <sup>[1]</sup>	5.50	22 400	3.9	$\mathrm{RL}^{(3)}$
NASNet-A <sup>[26]</sup>	2.65	1 800	3.3	RL
AmoebaNet-B <sup>[27]</sup>	2.55	3 150	3.2	EA <sup>④</sup>
ENAS <sup>[28]</sup>	2.89	0.5	4.6	RL
DARTS <sup>[29]</sup>	2.76	1	3.3	$\operatorname{Gradient}^{5}$
VGG-Lite	2.63	0.5	1.5	RL

①GPUD 为常用作 NAS 的耗时单位. ②Manual 为基于 人工设计. ③RL 即 reinforcement learning,为强化学习算法.
④EA 即 evolutionary algorithm,为进化算法. ⑤Gradient 为基 于梯度. (NAS、NASNet-A 以及 ENAS),一种使用进化算法的架构 AmoebaNet-B,以及近期经常被引用的 DARTS 架构来做比较.

在与基于人工设计的网络架构对比下,相较于 VGG-19,验证集的错误率从 6.80% 下降到 2.51%, 同时网络的参数量减少了 37.52 M,这意味着搜索 算法取得了成功,网络架构从错误率到参数量上都 呈现出明显的下降.与现阶段基于人工设计准确率 最高的 DenseNet-BC 相比,VGG-Lite 网络架构在网 络参数以及错误率上也具有优势,其原因在于 VGG-Lite 是为 Cifar-10 数据集定制的网络架构,而综合性 能最佳的人工设计网络在某一特定数据集上可能无 法取得最优的效果.

相比同样使用强化学习算法的 NAS 和 NASNet-A,搜索得到的 VGG-Lite 架构大幅缩短了搜索耗时, 并且错误率更低.其中,相对于 ENAS 在错误率和 参数量更低.究其原因,本工作在搜索空间上做出 了限制,降低了性能较差的候选架构生成的可能性, 加速了搜索过程.对于使用梯度算法的 DARTS 架 构,VGG-Lite 在 3 项评价标准中均处于领先地位. 相比于使用进化算法的 AmoebaNet-B,虽然在错误 率方 面 略 高,但是 VGG-Lite 搜索耗时约为 AmoebaNet-B 的 1/6 000.本方法无疑能让研究者更 快地得到所需要的网络架构.

在表4中,对于 Cifar-100 数据集,本工作选择 了同系 VGG 架构以及在此数据集上有着高准确率 的网络架构进行比较.与同类型的 VGG-11 以及 VGG-19 对比, VGG-Lite 在参数量上已经大幅降低, Top1 和 Top5 的错误率均低于 VGG-11 以及 VGG-19.与其他在此数据集上有着出色表现的网络架构 相比,相对于 ResNet 系列架构,在精度上有1个百

3	表 4	Cifar-10	0 数	如据集训练	结果	
Table 4	Com	parison	on	Cifar-100	fullv	traine

网络架构	参数量/ M	Topl 错 误率/%	Top5 错 误率/%
VGG-11 <sup>[10]</sup>	28.50	31.36	11.85
VGG-19 <sup>[10]</sup>	39.00	27.77	8.84
$MobilenetV2^{[30]}$	2.36	31.92	9.02
$ResNet101^{[20]}$	42.70	22. 22	5.61
$ResNet152^{[20]}$	58.30	22. 31	5.81
NasNet <sup>[26]</sup>	5.20	22.71	5.91
VGG-Lite	1.48	24.45	6.96

分点的差距,但它在参数量上具有更明显的优势. 以上结果说明,VGG-Lite 架构在与 Cifar-10 同源的 Cifar-100 数据集上进行完全训练也能具备较好的可 泛化性.

#### 2.7 架构分析

本工作最终网络架构(如图 10 所示)整体上与 初始架构 VGG-11 相似,不同点在于,原本 VGG-11 网络中存在的连续 3 × 3 卷积核堆叠的结构,被"卷 积+激活+池化"交替出现的结构所取代.由 VGG 网络的相关文献可知,设计这种连续 3 × 3 卷积堆叠 的目的在于,用更少的计算量来模拟 5 × 5 卷积、7 × 7 卷积甚至更大的卷积核.这些连续堆叠 3 × 3 卷积 的消失表明,Cifar-10 数据集可以通过尺寸较小的卷 积核就能完成图像特征的提取工作.与 ImageNet 数 据集 227 × 227 像素的图像相比,Cifar-10 数据集 32 × 32 像素的图像分辨率较小,此时不使用大尺寸



Fig. 10 Final Architecture of VGG-Lite

卷积核也能很好地提取到特征. 在全连接层之前使 用平均池化有效减小了特征图的大小,降低了全 连接层的工作,使得网络有更少的参数量而不损 失准确率. 通过上述分析不难发现,虽然待处理的 任务(本文中主要讨论图像分类任务)看似完全相 同,但对于不同数据集上搜索得到的最优架构区 别却很大. 上述的分析有助于更深刻地认识到网 络架构与具体应用场景之间的关联,这些特点对 于深度神经网络的可解释性研究具有重要意义. 然而上述研究目前仍处于定性研究阶段,未来还 需要从定量描述的角度对待处理任务和场景特征 (数据集)进行测评,这样才能揭示深度卷积网络, 实现真正高效的应用.

# 3 结论

 1)针对具体应用的 NAS 搜索中,使用基于人 工经验的网络作为算法的初始化可以在保证相对低 的错误率的同时大幅提升搜索效率.

2)本文方法开启了对 NAS 优化的一种思路, 该思路可以在不同的数据集以及网络架构中拓展, 进而实现对 NAS 方法的优化.

#### 参考文献:

- [1] ZOPH B, LE Q V. Neural architecture search with reinforcement learning[J]. ArXiv Preprint ArXiv, 2016: 1611. 01578.
- [2] 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.
- [3] CAI H, GAN C, WANG T, et al. Once-for-all: train one network and specialize it for efficient deployment [J]. ArXiv Preprint ArXiv, 2019: 1908. 09791.
- [4] TAN M, CHEN B, PANG R, et al. Mnasnet: platformaware neural architecture search for mobile [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 2820-2828.
- [5] PENG H, DU H, YU H, et al. Cream of the crop: distilling prioritized paths for one-shot neural architecture search[J]. ArXiv Preprint ArXiv, 2020: 2010. 15821.
- [6] REN P, XIAO Y, CHANG X, et al. A comprehensive survey of neural architecture search: challenges and solutions[J]. ArXiv Preprint ArXiv, 2020: 2006. 02903.
- [7] ZHOU D, ZHOU X, ZHANG 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.

- [8] 尹宝才,王文通,王立春. 深度学习研究综述[J]. 北京工业大学学报,2015,41(1):48-59.
  YIN B C, WANG W T WANG L C. Review of deep learning[J]. Journal of Beijing University of Technology, 2015,41(1):48-59. (in Chinese)
- [9] 成科扬,王宁,师文喜,等. 深度学习可解释性研究进展[J]. 计算机研究与发展,2020,57(6):1208-1217.
  CHENG K Y, WANG N, SHI W X, et al. Research advances in the interpretability of deep learning [J].
  Journal of Computer Research and Development, 2020, 57 (6): 1208-1217. (in Chinese)
- [10] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. ArXiv Preprint ArXiv, 2014: 1409. 1556.
- [11] CIFAR-10 and CIFAR-100 datasets (toronto. edu)[DB/ OL]. [2020-12-01]. http://www.cs.toronto.edu/~ kriz/cifar.html.
- [12] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning[J]. ArXiv Preprint ArXiv, 2013: 1312. 5602.
- [13] HE X, ZHAO K, CHU X. AutoML: a survey of the state-of-the-art [J]. Knowledge-Based Systems, 2019, 212: 106622.
- [14] ELSKEN T, METZEN J H, HUTTER F. Neural architecture search: a survey [J]. J Mach Learn Res, 2019, 20(55): 1-21.
- [15] SAMPSON J R. Adaptation in natural and artificial systems[J]. SIAM Review, 1976, 18(3):529-530.
- [16] 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. (in Chinese)
- [17] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60 (6): 84-90.
- [18] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [C] // European Conference on Computer Vision. Berlin: Springer, 2014: 818-833.

- [19] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C] // International Conference on Machine Learning. New York; ACM, 2015; 448-456.
- [20] HE K, ZHANG X, REN S, 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.
- [21] XIE S, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 1492-1500.
- [22] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
- [23] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [24] WILLIAMS R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning [J]. Machine Learning, 1992, 8(3/4): 229-256.
- [25] 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: 4700-4708.

- [26] ZOPH B, VASUDEVAN V, SHLENS J, et al. Learning transferable architectures for scalable image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8697-8710.
- [27] REAL E, AGGARWAL A, HUANG Y, et al. Regularized evolution for image classifier architecture search [C] // Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2019: 4780-4789.
- [28] PHAM H, GUAN M, ZOPH B, et al. Efficient neural architecture search via parameters sharing [C] // International Conference on Machine Learning. New York: ACM, 2018: 4095-4104.
- [29] LIU H, SIMONYAN K, YANG Y. DARTS: differentiable architecture search [C] // International Conference on Learning Representations. New Orleans: ICLR, 2019.
- [30] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 4510-4520.

(责任编辑 杨开英)