



中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION



基于通道性能度量的神经网络结构搜索算法

潘杰, 郑学驰, 邹筱瑜

引用本文:

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.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.0257>

您可能感兴趣的其他文章

Articles you may be interested in

基于SAPSO算法的RBF神经网络设计

Design of RBF neural network based on SAPSO algorithm

控制与决策. 2021, 36(9): 2305–2312 <https://doi.org/10.13195/j.kzyjc.2020.0176>

基于反时限混沌郊狼优化算法的BP神经网络参数优化

Parameter optimization of BP neural network based on coyote optimization algorithm with inverse time chaotic

控制与决策. 2021, 36(10): 2339–2349 <https://doi.org/10.13195/j.kzyjc.2020.0408>

基于生物启发神经网络和DMPC的多机器人协同搜索算法

Multi-robot cooperative search algorithm based on bio-inspired neural network and DMPC

控制与决策. 2021, 36(11): 2699–2706 <https://doi.org/10.13195/j.kzyjc.2020.0959>

周围神经MicroCT图像中神经束轮廓获取算法的改进

An improved approach to obtain contours of fascicular groups from MicroCT images of peripheral nerve

控制与决策. 2021, 36(7): 1601–1610 <https://doi.org/10.13195/j.kzyjc.2019.1664>

基于强化学习的倒立摆分数阶梯度下降RBF控制

Reinforcement learning based fractional gradient descent RBF neural network control of inverted pendulum

控制与决策. 2021, 36(1): 125–134 <https://doi.org/10.13195/j.kzyjc.2019.0816>

基于通道性能度量的神经网络结构搜索算法

潘杰^{1†}, 郑学驰¹, 邹筱瑜²

(1. 中国矿业大学 信息与控制工程学院, 江苏 徐州 221116;

2. 中国矿业大学 机电工程学院, 江苏 徐州 221116)

摘要: 卷积神经网络的表征与预测能力往往依赖结构合理性,但其主流结构均由人工设计,存在设计难度高、算力要求强、时间开销大等问题. 如何让神经网络自主搜索合理结构并节约计算资源是当前的研究重点. 目前,基于部分通道连接的可微分结构搜索算法,以其高效的显存利用率在搜索速度和分类性能上表现良好. 然而,其针对通道的随机采样策略易造成重要信息丢失,当通道连接不足时性能明显下降. 为此,提出一种基于通道性能度量的神经网络结构搜索算法,利用注意力机制提取通道重要性系数,并以此对通道进行排序采样. 此外,考虑到预热阶段导致搜索不充分,产生较大离散化误差,在结构权重连续化的过程中设计温度正则化系数,提升权重差异. 实验表明,所提算法能够在节约计算资源的基础上搜索出更优的卷积神经网络结构.

关键词: 卷积神经网络; 可微分结构搜索; 通道注意力; 通道排序; 离散化误差; 温度正则化系数

中图分类号: TP183

文献标志码: A

DOI: 10.13195/j.kzyjc.2023.0257

引用格式: 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.

Differentiable neural architecture search with channel performance measurement

PAN Jie^{1†}, ZHENG Xue-chi¹, ZOU Xiao-yu²

(1. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China; 2. School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: The design of convolutional neural network architecture is critical for achieving high characterization and prediction performance. However, manual design can be time-consuming and computationally expensive. Automatic architecture search methods are thus becoming popular. Presently, partial channel connections for memory-efficient differentiable architecture search being efficient in terms of search speed, memory utilization, and classification performance. However, the random sampling strategy can cause the loss of important information. To address these issues, this paper proposes differentiable neural architecture search with channel performance measurement, which extracts importance coefficients of channels with attention mechanism and ranks the channels accordingly to select the most important ones. The proposed algorithm also includes a temperature regularization coefficient to improve the transformation from discrete to continuous architecture weight and reduce discretization errors. Experimental results show that the proposed algorithm achieves good performance while significantly reducing computational resources.

Keywords: convolutional neural network; differentiable architecture search; channel attention; channel sorting; discretization error; temperature regularization coefficient

0 引言

21 世纪以来,随着机器学习理论的不断深化与硬件算力的飞速提升,卷积神经网络(convolutional neural networks, CNN)在计算机视觉领域发挥着越

来越重要的作用^[1-2]. 目前,广泛应用的网络结构多由人工设计,如结构复杂的 VGGNet^[3]、ResNet^[4]、DenseNet^[5] 及轻量级的 MobileNet^[6] 与 ShuffleNet^[7] 等. 尽管这些经典结构具有一些变体可以应用于

收稿日期: 2023-03-08; 录用日期: 2023-06-16.

基金项目: 国家自然科学基金项目(62176258, 62273349, 52174152); 中央高校基本科研业务费专项资金项目(2021YCPY0111); 江苏省高校优势学科建设工程资助项目(PAPD).

责任编委: 张国山.

[†]通讯作者. E-mail: panjie1616@cumt.edu.cn.

不同背景下的视觉任务,但其依旧缺乏灵活性.近年来,CNN的研究逐渐注重对网络结构进行创新,而结构的设计依赖充分的专业知识与反复试验,且对计算硬件要求很高,一定程度上限制了神经网络的应用.神经网络结构搜索(neural architecture search, NAS)的出现,对于推动深度学习发展具有重要意义.

NAS旨在自动设计网络结构,由于卷积神经网络在深度学习中发挥着至关重要的作用,当前研究工作主要集中于卷积神经网络的结构搜索,通过特定搜索策略自主寻找最优网络结构^[8].2016年,Zoph等^[9]采用基于强化学习的NAS算法,在CIFAR10数据集上的精度超越人工设计的VGGNet和ResNet等经典网络结构,此后一系列基于进化学习^[10-12]、强化学习^[13-15]、梯度下降^[16-18]等搜索策略的NAS算法被提出.基于进化和强化学习的方法在研究早期发挥着重要作用,其原理是在元控制器指导下,从搜索空间中采样和评估结构.这些方法虽然取得了不错的性能,但计算开销格外庞大.如Zoph等^[9]所提算法在CIFAR10上完成搜索任务需要22 400 GPU天.大规模进化NAS算法^[12]尽管采用权重继承减少了计算开销,但在CIFAR10上完成搜索任务仍需要2 750 GPU天.

针对计算开销问题,Zoph等^[13]考虑采用可迁移搜索空间(NASNet search space),即通过小数据集搜索任务获得单元网络,堆叠单元网络并迁移到更复杂、规模更大的数据集,使搜索时间降低到2 000 GPU天.然而,这样的搜索效率依旧很难满足现实需求.采用梯度下降作为搜索策略的典型方法是可微分结构搜索(differentiable architecture search, DARTS)^[19],其网络模型以可微分参数化形式实现,将离散问题连续化,并将结构搜索融合到模型训练中,摆脱了对搜索结构反复验证的过程,将搜索成本大幅降低,仅需0.4 GPU天即可完成搜索任务.然而,其候选操作及输出需要储存在单元网络节点内,导致占用大量显存,一定程度上影响了算法效率与稳定性.为此,研究人员进一步提出了基于部分通道连接的可微分结构搜索算法(partial channel connections for memory-efficient architecture search, PC-DARTS)^[20],在所有输入通道发送至候选操作前随机采样部分通道用于计算以减少显存占用.

研究发现,PC-DARTS依旧存在如下关键问题:1)采样策略为随机选取部分通道,易使网络在训练过

程中损失有效信息,导致采样通道数降低到一定程度时,算法性能显著下降;2)为最大限度提升信息流动性,导致候选操作权重倾向于收敛到一个集中区间,并且算法引入预热阶段,降低了搜索阶段的迭代轮数,使算法在选取最优结构时产生更大的离散化误差.

鉴于此,本文提出基于通道性能度量的神经网络结构搜索算法(MCP-DARTS),针对PC-DARTS采用随机采样更少通道带来的信息损失,度量了各通道重要性,选取部分重要通道用于采样.此外,针对选取最优结构时产生更大的离散化误差问题,算法在结构搜索过程中使判断各候选操作优劣的权重分布变得更加具有区分性.

本文主要工作如下:1)提出通道排序机制,利用通道注意力在预热阶段获得的通道重要性系数,在结构搜索阶段开始前,衡量各通道重要性并排序,仅将重要性较高的通道用于搜索计算,以保留有效信息,缩短搜索时间,并减少过拟合;2)为提高搜索与评估阶段相关性,在归一化候选操作时引入温度正则化系数,扩大不同候选操作间的权重差距,减小离散化误差,提升算法的稳定性与精度;3)在CIFAR10和ImageNet数据集上进行了实验验证,与其他主流NAS算法相比,本文算法具有更高的搜索效率.

1 相关工作

1.1 可微分神经网络结构搜索

过大的计算开销使得NAS算法在面向现实需求时遇到了阻碍.为缓解该问题,最近的研究转向如何高效搜索网络结构.DARTS^[19]算法将离散搜索空间转化为可微搜索空间,在搜索阶段对超网络进行优化,保留最优子网络,在性能和效率上有了显著提高.尽管如此,研究人员依旧发现DARTS算法对于显存利用率不足,候选操作权重分布相对集中,导致搜索过程效率低、搜索结构相关性弱、搜索结果稳定性差等问题,为此出现了许多对DARTS进行优化的研究成果.SNAS^[21]通过直接优化NAS任务的目标函数,缓解了搜索阶段与评估阶段优化目标不一致的问题;P-DARTS^[22]利用逐步优化搜索阶段网络深度,缩小搜索网络与评估网络间的性能差距;SGAS^[23]采用贪心策略,通过选边指标逐步确定搜索结构来提升搜索与评估阶段相关性;RL-DARTS^[24]将参数的优化问题利用基于强化学习的元优化器解决,能够在更短的时间内找到最优模型,提高搜索效率;Fair-DARTS^[25]利用sigmoid函数取代softmax函数,将离

散搜索空间连续化,消除了候选操作竞争的不公平性,避免搜索出大量跳跃连接而导致网络性能崩溃,提升了搜索结构稳定性;MR-DARTS^[26]提出多路径搜索空间提高单元网络的特征表示能力,同时限制单元网络的连通性以加深其网络深度,弥补搜索与评估阶段的深度缺口,提高二者相关性;DLW-NAS^[27]构建了一个轻量级的搜索空间,用于降低模型复杂度,从而节约所需的计算资源,缓解模型在资源有限的设备上部署困难的问题;iDARTS^[28]引入节点归一化和去相关离散化策略,为算法提供更好的分类精度和稳定性;PC-DARTS^[20]通过对部分输入通道进行采样,减少了空间冗余,缓解了DARTS在显存方面的问题,提升了结构搜索效率。

1.2 注意力机制

注意力机制最初被使用在机器翻译领域,随后广泛应用于计算机视觉任务^[29],其与人类感知系统相似,选择性地聚焦于视觉图像的重要部分以更好捕捉视觉结构,可以看作是输入图像特征权重动态调整的过程^[30]。行人识别^[31]、目标跟踪^[32]、图像分类^[33]等视觉任务通过引入注意力机制都可以提高神经网络的特征提取能力,本文主要涉及的是通道注意力。如squeeze-and-excitation (SE)^[34]中的SE模块能够注意到通道之间的关系,模块通过全局平均池化特征来计算通道注意力系数,为每个通道赋予不同权重;gather-excite (GE)^[35]在SE模块中加入了深度卷积,能够更好地挖掘特征上下文信息;文献^[36]则提出卷积块注意模型,并指出通过SE模块得到的通道注意力特征仅是次优的,只有同时使用全局平均池化和全局最大池化并将特征合并,才能得到更精确的通道注意

力系数。

2 基于通道性能度量的神经网络结构搜索算法(MCP-DARTS)

为合理减少选取通道,以兼顾计算资源和算法稳定性,本文提出基于通道性能度量的神经网络结构搜索算法。本节从算法框架、单元网络构建、通道性能度量、搜索空间连续化策略与算法的总体流程等方面依次展开论述。

2.1 算法框架

本文算法网络是由normal cell 和 reduction cell 两种单元重复堆叠构成的 n 层神经网络,网络的1/3层和2/3层为reduction cell,其余为normal cell。通过两级优化策略分别优化结构权重和网络权重,输出训练后的最优神经网络结构,再将该结构扩大堆叠为 m 层单元网络,重新训练验证算法性能。算法框架由3部分构成:通道排序、选取部分通道连接与连续化搜索空间。首先,在各节点候选操作前加入通道排序机制,预热阶段学习各节点通道注意力系数,搜索阶段开始前按照注意力系数对通道进行排序。其次,选取注意力系数相对更大的部分通道用于结构搜索,剩余通道跳过该过程直接与输出合并。然后,在连续化搜索空间扩大各候选操作权重分布范围。算法框架如图1所示,最左侧为具有 n 个单元层的神经网络结构,其中单元层具体结构为其右侧具有4个节点的有向图,该有向图两两节点间的信息传输可进一步展开为由通道排序、通道采样、结构搜索等环节构成的系列流程。需要说明的是,算法的核心在于通道排序,通过全局池化与共享参数获得注意力系数,能够在采样通道部分子集时减少信息损失。

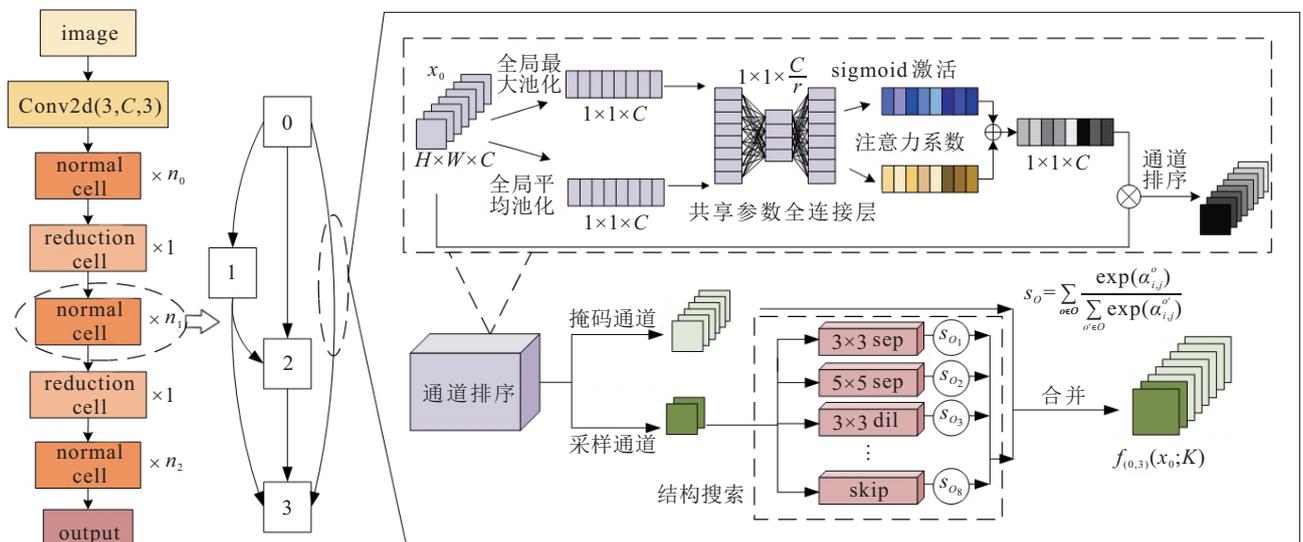


图1 算法框架

2.2 单元网络构建

考虑到算法在大数据集上的可迁移性以及良好搜索效率的维持,本文采用NASNet^[17]搜索空间构建最终的结构单元网络.该单元网络是由 C 个序列节点组成的有向无环图,每两个节点间存在着8种候选操作,分别为零操作(O_1)、 3×3 最大池化(O_2)、 3×3 平均池化(O_3)、跳跃连接(O_4)、 3×3 可分离卷积(O_5)、 5×5 可分离卷积(O_6)、 3×3 扩张可分离卷积(O_7)、 5×5 扩张可分离卷积(O_8).搜索得到的单元网络重复堆叠构成最终神经网络结构.将单元网络里所有候选操作被选择的可能性以参数化形式表示,利用softmax函数将搜索空间连续化表示,节点 x_i 向节点 x_j 传递的过程如下:

$$\bar{o}_{i,j}(x_i) = \sum_{o \in O} \frac{\exp(\alpha_{i,j}^o)}{\sum_{o' \in O} \exp(\alpha_{i,j}^{o'})} \cdot o(x_i). \quad (1)$$

其中: O 表示候选操作集合, $o(\cdot)$ 为卷积、池化等候选操作, $\alpha_{i,j}^o$ 表示节点 x_i 与节点 x_j 间候选操作权重.因此,结构搜索任务可以简化为学习一组连续变量 $\alpha = \{\alpha_{i,j}^o\}$,算法可以表示为如下利用梯度下降分别交替优化训练损失和验证损失的两级优化问题:

$$\min_{\alpha} L_{\text{val}}(\omega^*(\alpha), \alpha); \quad (2)$$

$$\text{s.t. } \omega^*(\alpha) = \arg \min_{\omega} L_{\text{train}}(\omega, \alpha). \quad (3)$$

其中: L_{val} 和 L_{train} 分别表示验证集损失和训练集损失; α 和 ω 分别表示结构权重和网络权重,通过梯度下降法对两者进行参数更新,有

$$\alpha_{k+1} = \alpha_k - \eta_2 \nabla_{\alpha} L_{\text{val}}(\omega, \alpha), \quad (4)$$

$$\omega_{k+1} = \omega_k - \eta_1 \nabla_{\omega} L_{\text{train}}(\omega, \alpha), \quad (5)$$

η_1 和 η_2 分别表示网络权重的学习率和结构权重的学习率.

2.3 通道性能度量

DARTS存在空间冗余性,因此PC-DARTS^[24]在通道选择比例 $K = 4$ 时可以获得良好性能,但设置 $K = 8$ 即选择1/8通道用于搜索任务,会导致有效信息丢失,算法性能明显下降.为缓解这一问题,本文在预热阶段度量通道性能,结构搜索时选择重要通道进行采样,对信息量不足通道进行掩码操作,从而避免有效信息丢失.

注意力系数的获取是衡量通道性能的关键,其原理在于利用全局信息有选择性地关注对当前任务有用的特征图通道,并根据重要程度赋予系数.首先利用全局平均池化和全局最大池化将各通道信息压缩

得到两个 $1 \times 1 \times C$ 的通道描述,然后将通道描述分别送入两层参数共享的全连接神经网络,再将输出结果相加,利用sigmoid函数激活获得介于 $0 \sim 1$ 的通道注意力系数 $M_c(F)$ 如下:

$$M_c(F) = \sigma(\text{MLP}(\text{Avgpool}(F)) + \text{MLP}(\text{Maxpool}(F))). \quad (6)$$

其中: F 表示各通道输入特征, σ 表示sigmoid函数,MLP表示全连接层,Avgpool和Maxpool分别表示全局平均池化和全局最大池化.利用获得的注意力系数从大到小对通道进行排序,选取前 $1/K$ 通道进行采样,剩下通道进行掩码操作,节点 x_i 向节点 x_j 传递的最终输出 $f_{(i,j)}(x_i; K)$ 为

$$f_{(i,j)}(x_i; K) = \sum_{o \in O} \frac{\exp(\alpha_{i,j}^o)}{\sum_{o' \in O} \exp(\alpha_{i,j}^{o'})} \cdot o(x) \cdot \frac{1}{K} + \left(1 - \frac{1}{K}\right) \cdot x_i, \quad (7)$$

其中 $\frac{1}{K}$ 和 $1 - \frac{1}{K}$ 代表被选择和被掩码的通道比例. K 值越大即为被选择的通道越少,所需计算开销越小,允许设置更大的batchsize提高运行效率,在每次迭代中采样更多数据,减少更新网络权重和结构权重带来的不确定性,提高结构搜索稳定性.此外,由于被掩码部分的通道可以起到缓解梯度消失的作用,选取更少部分的通道进行采样,对可微分结构搜索算法而言能够获得更好的跳跃连接正则化效果,提高搜索结构稳定性.

2.4 搜索空间连续化策略

可微分结构搜索算法在搜索结束后选取softmax最大 $\alpha_{i,j}^o$ 值对应的候选操作.搜索过程中,各节点间输入输出传递是通过对所有候选操作加权求和获得,候选操作结构权重倾向于收敛到集中区间以提升网络的信息流通性,导致搜索结构相关性弱.此外,为了解决算法训练中参数不确定问题,同时在更新结构权重之前度量通道性能,算法引入预热阶段.但通过实验发现预热阶段使最终各操作的结构权重差异集中在 $0.1 \sim 0.2$ 之间,如图2所示.图2中纵坐标表示结构单元上的每条边,横坐标表示不同的候选卷积操作.这导致判定操作优良的范围变得更窄,不同操作的最优值与次优值差异很小,最优操作易产生混淆,搜索出的网络与包含所有结构的原始网络存在着性能差异,造成一定离散化误差.

为使算法对于最优候选操作的选择更具分辨力,本文在式(7)softmax函数中加入温度正则化系数,节点 i 到节点 j 的转播过程变为

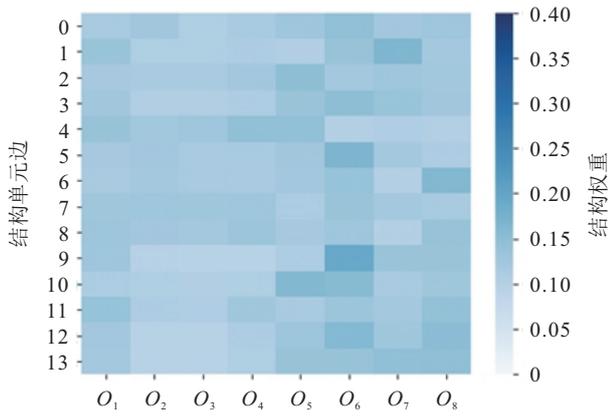


图2 PC-DARTS ($K = 8$)在CIFAR10数据集上的结构权重热力图

$$f_{(i,j)}(x_i; K) = \sum_{o \in O} \frac{\exp(\alpha_{i,j}^o / T)}{\sum_{o' \in O} \exp(\alpha_{i,j}^{o'} / T)} \cdot o(x) \cdot \frac{1}{K} + \left(1 - \frac{1}{K}\right) \cdot x_i, \tag{8}$$

其中 T 为介于 $0 \sim 1$ 的温度正则化系数. 引入 T 可以使不同的操作权重呈发散分布, 变得更加具有区分度, 从而扩大判定操作优良的范围, 进而能够达到减少离散化误差的目的, 缩小搜索阶段和评估阶段之间网络的性能差距.

2.5 算法流程

基于通道性能度量的NAS算法流程如下: 首先初始化算法参数, 在当前迭代轮数小于 K 时仅更新网络权重, 大于等于 K 时按照各边通道注意力系数对通道进行排序, 选取前 $1/K$ 进行采样以计算结构搜索任务, 通过二级优化对网络权重和结构权重进行梯度更新, 基于学习到的结构权重获得最终的网络结构.

算法1 MCP-DARTS.

输入: 迭代总轮数 M , 预热阶段迭代轮数 N , 训练集损失 L_{train} , 验证集损失 L_{val} , 网络权重学习率 η_1 , 结构权重学习率 η_2 , 通道选择比例 K ;

输出: 搜索得到的CNN结构.

初始化: 网络权重 ω 和结构权重 α , 注意力系数 M_c , 当前迭代轮数 t .

- 1) while $t < M$
- 2) for each (i, j)
- 3) if $t < N$
- 4) 随机选取 $1/K$ 通道
- 5) 根据式(6)计算注意力系数 M_c
- 6) else if $t = N$
- 7) 随机选取 $1/K$ 通道
- 8) 根据式(6)计算注意力系数 M_c 并排序

- 9) else if $t > N$
- 10) 选择前 $1/K$ 通道
- 11) 根据式(4)更新结构权重
- 12) end if
- 13) 根据式(5)更新网络权重
- 14) end for
- 15) $t = t + 1$
- 16) end while
- 17) return //通过学习到的结构权重获得最终网络结构.

3 实验

3.1 数据集

本文主要在 CIFAR10 和 ImageNet 这两个图像分类常用数据集上进行实验. CIFAR10 由 6 万张分辨率为 32×32 的图片组成, 包括 5 万张训练图片和 1 万张测试图片, CIFAR10 中的这些图片均匀分布在 10 个类中, 其中训练图片被均匀分成两个子集, 分别用于训练网络权重和结构权重. ImageNet 数据集包括 1000 个对象类别、130 万张训练图片和 5 万张验证图片, 图片大小为 224×224 .

3.2 实验设置

本文设置通道选择比例 K 为 8, $batchsize$ 为 512, 网络权重学习率初始值为 0.2, 学习率衰减采用余弦退火策略. 迭代总轮数 M 为 50, 预热阶段轮次 N 为 15. 搜索阶段网络层数为 8 层, 初始通道为 16, 损失函数为交叉熵损失函数. 网络权重优化采用 SGD 优化器, 权重衰减为 0.0003. 结构权重优化采用 Adam 优化器, 结构权重学习率固定为 0.0006, 权重衰减为 0.001.

在 CIFAR10 数据集上进行结构验证的参数设置如下: 网络层数为 20 层, 初始通道为 36, 迭代总轮次为 600, $batchsize$ 设置为 96, 网络权重初始学习率为 0.025, 学习率衰减策略为余弦退火策略. 对数据集使用 cutout 技术^[37]进行预处理, 正则化长度为 16.

在 ImageNet 数据集上进行结构验证的参数设置如下: 网络层数为 14 层, 初始通道为 48, 迭代总轮次为 250, $batchsize$ 为 128, 使用 SGD 优化器, 网络权重初始学习率为 0.1, 学习率每轮 epoch 衰减 0.97 倍.

3.3 与PC-DARTS算法的性能对比

本文分别进行 5 次独立实验, 与 PC-DARTS ($K = 4$) 和 PC-DARTS ($K = 8$) 进行性能对比, 结果如表 1 所示. 搜索得到的 normal cell 和 reduction cell 网络结构可视化结果分别如图 3(a)、(b) 所示. 通过对比可见, 在 CIFAR10 数据集上 PC-DARTS ($K = 8$) 得到的网

络结构准确率较低,本文算法相较于同样仅选取1/8通道进行采样的PC-DARTS算法,在仅增加0.01 GPU天基础上平均准确率从96.75%提升到97.38%,其中5次独立实验最优结构的准确率也有明显提升.相较于PC-DARTS($K = 4$),本文算法能节约29.6%搜索时间,平均准确率也高于97.33%,搜索得到的结果准确率最高达到97.44%,高于PC-DARTS算法的

97.43%.从表1结果可以看出,由于本文算法利用通道注意力机制在训练过程中衡量了各个通道的性能,在选取更少通道连接的情况下,减轻了结构搜索过程中产生的图像信息损失,并且在搜索阶段算法使用相同batchsize使节点间显存占用更少,提升了显存利用率.因此相较于PC-DARTS,MCP-DARTS在保证准确率不变的情况下,搜索效率更高.

表1 PC-DARTS取不同K值和本文算法在CIFAR10上的实验结果

操作	PC-DARTS($K = 4$)		PC-DARTS($K = 8$)		MCP-DARTS($K = 8$)	
	准确率/%	GPU/天	准确率/%	GPU/天	准确率/%	GPU/天
1	97.28	0.125	96.52	0.083	97.38	0.088
2	97.33	0.125	96.56	0.083	97.44	0.088
3	97.43	0.125	96.99	0.083	97.40	0.088
4	97.25	0.125	96.96	0.083	97.29	0.088
5	97.36	0.125	96.73	0.083	97.26	0.088
avg	97.33	0.125	96.75	0.083	97.35	0.088

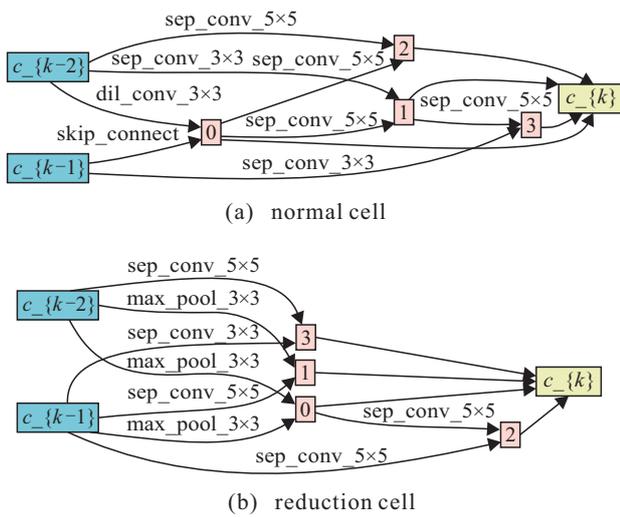


图3 CIFAR10上搜索得到的单元网络结构

3.4 温度系数正则化 T 对性能的影响

图4给出了不同温度正则化系数下候选操作结构权重在某条边上的操作.图4中横坐标为8个不同候选操作,纵坐标为该操作在一条边上所对应的结构权重.由图4可知 $T = 1$ 时权重分布曲线平滑,当 T 逐渐缩小时,结构权重分布曲线逐渐尖锐.图5(a)~(d)分别表示当 T 取1、0.6、0.2、0.05时结构权重在所有边上的热力图,图中纵坐标为单元网络中的所有边,横坐标为每条边的候选操作.当 $T = 1$ 时,每条边上的结构权重大部分集中在0.08~0.18之间;当 $T = 0.6$ 时,每条边上的结构权重集中区间较 $T = 1$ 时变化不大,但权重分布呈两极化趋势;当 $T = 0.2$ 时,每条边上的结构权重分布区别进一步扩大至0.05~0.3之间;当 $T = 0.05$ 时,每条边上的结构权重分布在

0.001~0.6之间.

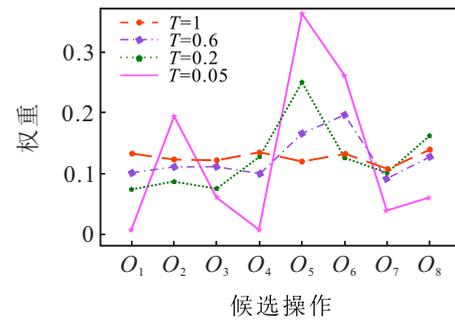


图4 不同温度正则化系数下候选操作结构权重在某条边上的分布

取不同的 T 值搜索所获得的结构验证性能对比如表2所示.可以看出:当 T 取1即不加温度正则化系数时,由于各候选操作权重分布范围集中,具有比较大的离散化误差,模型性能以及稳定性较差;当 T 取0.2时各边候选操作差异明显,对算法准确率的提升相较于其他 T 值达到了最好效果,算法稳定性也有一定的提升;而当 T 值过小时,神经网络信息流通性受到影响,使算法易陷入局部最优,对于结构搜索的探索性变弱,导致对算法稳定性也产生了一定影响,准确率波动较大.因此,选取合理的温度正则化系数,对于算法的准确率和稳定性都有着显著改善.

表2 温度正则化系数对精度的影响 %

T 值	准确率
1	97.20 ± 0.10
0.6	97.31 ± 0.08
0.2	97.35 ± 0.07
0.05	97.07 ± 0.22

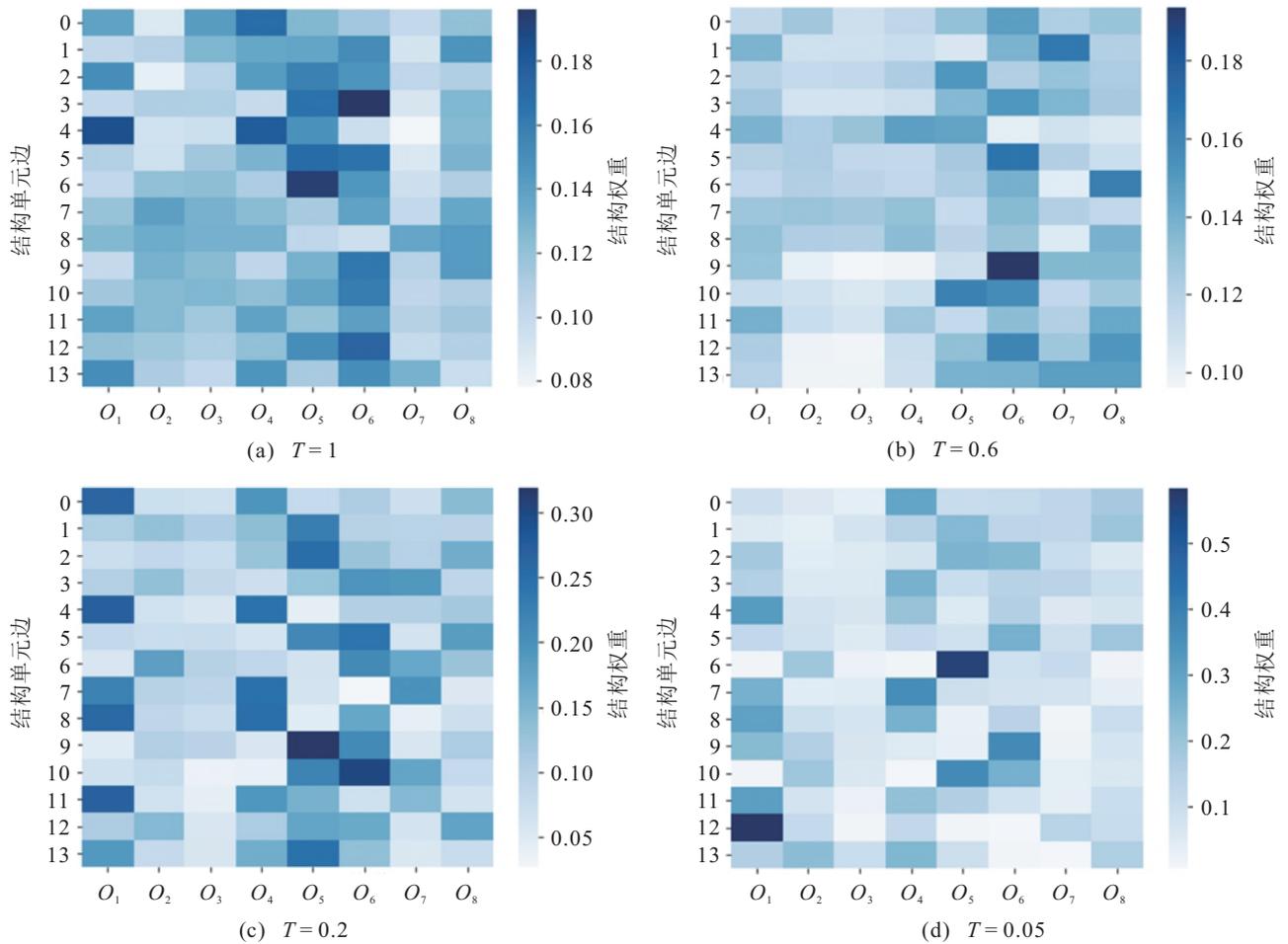


图5 取不同温度正则化系数下候选操作结构权重热力图

3.5 算法性能对比

为了验证效果,表3和表4对比了本文算法与人工设计CNN、基于强化学习的NAS算法、基于进化的NAS算法、基于梯度的NAS算法分别在CIFAR10和ImageNet数据集上的性能结果.在算法对比中使用搜索结构在两个数据集上的错误率、模型参数量和搜索所需的GPU时间作为衡量算法性能的指标.表3中:avg代表5次独立实验结果的平均值,best代表最优实验结果.为了公平比较,表4中PC-DARTS的错误率为在与本文所设置相同学习率和batchsize的基础上实验得到的结果.

对比早期人工设计的CNN结构,NAS算法能够考虑到专家手工设计网络过程中未考虑的因素,因此取得了更加优异的性能表现,在精度上有着较大提升,并且NAS算法搜索获得的模型参数量更少,更易于训练.

基于强化学习的NAS算法中,NASNet-A^[13]在精度上能够获得良好性能,但其在搜索过程中需要花费大量时间,难以应用到实际领域.BlockQNN^[15]的搜索时间有所减少,但其结构模型参数量过大,

难以训练.ENAS^[14]使用权重共享的方式进一步缩小了搜索时间,但精度较于NASNet-A有所下降.基于进化学习的NAS算法中,AmoebaNet^[10]精度性能表现良好,但需要庞大的计算开销.Hierarchical evolution^[11]在计算开销大的问题上有所缓解,但其结构模型参数量较大,不易于训练.

基于梯度的NAS算法以其搜索效率高和模型参数量少的特点相比其他设计方法表现出独特的优越性.本文算法是在其基础上做出的进一步改进,通过减小冗余空间架构,仅需0.09 GPU天就能搜索获得在CIFAR10上错误率最低为2.5%以及在ImageNet上top 1/top 5错误率为25.9%/8.4%的神经网络结构.对比目前该领域前沿的PC-DARTS、RL-DARTS^[24]与DLW-NAS^[27]等可微分搜索框架下的NAS算法,MCP-DARTS搜索得到的网络结构分类精度表现更佳,耗费的搜索时间更少,具有更优的搜索效率.对比目前分类性能较好的ProxylessNAS^[18]和MR-DARTS^[26],本文算法在搜索时间上具有明显的优势,并且结构模型参数量更低,搜索获得的网络更易于训练.

表3 CIFAR10上人工设计网络结构和各NAS算法结构的性能表现

设计方法	结构	错误率/%		参数量/MB	GPU/天
		top 1	top 5		
人工	VGG ^[3]	6.7	20.4	20.4	—
	DesNet-BC ($k = 40$) ^[5]	3.8	25.6	25.6	—
	ResNet (depth=101) ^[4]	6.4	25.3	25.3	—
基于强化	NASNet-A + cutout ^[13]	2.65	3.3	3.3	2 000
	BlockQNN ^[15]	3.54	39.8	39.8	96
	ENAS ^[14]	2.89	4.6	4.6	0.5
基于强化	Large Scale Evolution ^[12]	5.4	5.4	5.4	2 750
	AmoebaNet-Acutout ^[10]	3.12	3.1	3.1	3 150
	Hierarchical evolution ^[11]	3.6	15.7	15.7	300
基于梯度 + 强化	RL-DARTS + cutout ^[24]	2.74	3.2	3.2	0.18
基于梯度	ProxylessNAS + cutout ^[18]	2.08	5.7	5.7	4
	BayesNAS + cutout ^[17]	2.81 ± 0.04	3.4	3.4	0.2
	DARTS (first order) + cutout ^[19]	3.00 ± 0.14	3.3	3.3	1.5
	DARTS (second order) + cutout ^[19]	2.76 ± 0.09	3.3	3.3	4
	SNAS (moderate) + cutout ^[21]	2.85 ± 0.02	2.8	2.8	1.5
	P-DARTS + cutout ^[22]	2.5	3.4	3.4	0.3
	SGAS + cutout ^[23]	2.66 ± 0.24	3.7	3.7	0.25
	Fair-DARTS + cutout ^[25]	2.54 ± 0.05	3.32	3.32	0.4
	MR-DARTS + cutout ^[26]	2.49	4.3	4.3	0.4
	DLW-NAS + cutout ^[27]	2.74	2.3	2.3	0.45
	PC-DARTS (avg) ^[20]	2.67 ± 0.06	3.6	3.6	0.125
	PC-DARTS (best) ^[20]	2.57	3.6	3.6	0.125
	MCP-DARTS (avg)	2.65 ± 0.07	3.6	3.6	0.09
	MCP-DARTS (best)	2.56	3.4	3.4	0.09

表4 ImageNet上人工设计网络结构和各NAS算法结构的性能表现

设计方法	结构	错误率/%		参数量/MB	GPU/天
		top 1	top 5		
人工	Inception-V1 ^[12]	30.2	10.1	6.6	—
	MobileNet ^[6]	29.4	10.5	4.2	—
	ShuffleNet 2× (v1) ^[7]	26.4	10.2	5	—
基于强化	NASNet-A ^[13]	26.0	8.4	5.3	2 000
	NASNet-B ^[13]	27.2	8.7	5.3	2 000
	NASNet-C ^[13]	27.5	9.0	4.9	2 000
基于进化	AmoebaNet-A ^[10]	25.5	8.0	5.1	3 150
	AmoebaNet-B ^[10]	26.0	8.5	5.3	3 150
基于梯度	DARTS (second order) + cutout ^[19]	26.7	8.7	4.7	1
	ProxylessNAS + cutout ^[18]	25.0	7.5	7.1	8.3
	BayesNAS + cutout ^[17]	26.5	8.9	3.9	0.2
	SNAS (mild) ^[21]	27.3	9.2	4.3	1.5
	DLW-NAS + cutout ^[27]	26.1	8.1	3.8	0.45
	PC-DARTS ^[20]	26.0	8.5	5.3	0.125
	MCP-DARTS	25.9	8.4	5.0	0.09

4 结论

本文针对在选取极少部分通道随机采样时性能下降以及预热阶段导致产生更大的离散化误差等问题,提出基于通道性能度量的神经网络结构搜索算法(MCP-DARTS)。通过度量通道性能,解决了随机采样通道导致图像信息丢失严重的问题,并进一步提高了结构搜索效率。此外,本文还引入温度正则化系数,使得算法在搜索过程中能够做出更加具有分辨力的决策,提高搜索网络与验证网络间的相关性,减小确定最终网络时产生的离散化误差。在CIFAR10数据集上的实验结果表明,本文算法仅需0.09 GPU天就能搜索出分类错误率仅有2.56%的CNN结构,相比于主流算法具有明显的搜索效率优势。需要指出的是,当前基于梯度的结构搜索算法主要适用于小数据集。针对大数据集如ImageNet上采用的方法,仍然是在CIFAR10上搜索单元网络并进行结构堆叠。如何直接在大数据集上进行高效结构搜索是后续工作的研究重点。

参考文献(References)

- [1] 宋燕, 覃俞璋, 曾入. 基于注意力的共享参数胶囊网络[J]. 控制与决策, 2023, 38(6): 1577-1585.
(Song Y, Qin Y Z, Zeng R. Attention-based capsule network with shared parameters[J]. Control and Decision, 2023, 38(6): 1577-1585.)
- [2] Toshev A, Szegedy C. DeepPose: Human pose estimation via deep neural networks[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, 2014: 1653-1660.
- [3] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J/OL]. 2014, arXiv: 1409.1556.
- [4] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 770-778.
- [5] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 2261-2269.
- [6] Howard A G, Zhu M L, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[J/OL]. 2017, arXiv: 1704.04861.
- [7] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 6848-6856.
- [8] 刘建伟, 王新坦. 神经网络结构搜索方法综述[J]. 控制理论与应用, 2023, 40(1): 12-31.
(Liu J W, Wang X T. A review of neural network structure search method[J]. Control Theory & Applications, 2023, 40(1): 12-31.)
- [9] Zoph B, Le Q V. Neural architecture search with reinforcement learning[J/OL]. 2016, arXiv: 1611.01578.
- [10] 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.
- [11] Liu H X, Simonyan K, Vinyals O, et al. Hierarchical representations for efficient architecture search[J/OL]. 2017, arXiv: 1711.00436.
- [12] Real E, Moore S, Selle A, et al. Large-scale evolution of image classifiers[J/OL]. 2017, arXiv: 1703.01041.
- [13] Zoph B, Vasudevan V, Shlens J, et al. Learning transferable architectures for scalable image recognition[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 8697-8710.
- [14] Pham H, Guan M Y, Zoph B, et al. Efficient neural architecture search via parameter sharing[J/OL]. 2018, arXiv: 1802.03268.
- [15] Zhong Z, Yan J J, Wu W, et al. Practical block-wise neural network architecture generation[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 2423-2432.
- [16] Shin R, Packer C, Song D. Differentiable neural network architecture search[C]. International Conference on Learning Representations. Vancouver, 2018: 340-343.
- [17] Zhou H P, Yang M H, Wang J, et al. BayesNAS: A Bayesian approach for neural architecture search[J/OL]. 2019, arXiv: 1905.04919.
- [18] Cai H, Zhu L G, Han S. ProxylessNAS: Direct neural architecture search on target task and hardware[J/OL]. 2018, arXiv: 1812.00332.
- [19] Liu H X, Simonyan K, Yang Y M. DARTS: Differentiable architecture search[J/OL]. 2018, arXiv: 1806.09055.
- [20] Xu Y, Xie L, Zhang X, et al. Partial channel connections for memory-efficient differentiable architecture search[C]. International Conference on Learning Representations. Addis Ababa, 2020: 1-13.
- [21] Xie S R, Zheng H H, Liu C X, et al. SNAS: Stochastic neural architecture search[J/OL]. 2018, arXiv: 1812.09926.
- [22] Chen X, Xie L X, Wu J, et al. Progressive differentiable architecture search: Bridging the depth gap between search and evaluation[C]. 2019 IEEE/CVF International Conference on Computer Vision. Seoul, 2020:

- 1294-1303.
- [23] Li G H, Qian G C, Delgadillo I C, et al. SGAS: sequential greedy architecture search[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, 2020: 1617-1627.
- [24] Pang D, Le X Y, Guan X P. RL-DARTS: Differentiable neural architecture search via reinforcement-learning-based meta-optimizer[J]. Knowledge-Based Systems, 2021, 234: 107585.
- [25] Chu X X, Zhou T B, Zhang B, et al. Fair DARTS: Eliminating unfair advantages in differentiable architecture search[C]. European conference on computer vision. Glasgow, 2020: 465-480.
- [26] Gao F, Song B, Wang D, et al. MR-DARTS: Restricted connectivity differentiable architecture search in multi-path search space[J]. Neurocomputing, 2022, 482: 27-39.
- [27] Li S, Mao Y X, Zhang F C, et al. DLW-NAS: Differentiable light-weight neural architecture search[J]. Cognitive Computation, 2023, 15(2): 429-439.
- [28] Wang H Q, Yang R J, Huang D, et al. iDARTS: Improving DARTS by node normalization and decorrelation discretization[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(4): 1945-1957.
- [29] Corbetta M, Shulman G L. Control of goal-directed and stimulus-driven attention in the brain[J]. Nature Reviews Neuroscience, 2002, 3(3): 201-215.
- [30] 韩晓微, 张云泽, 谢英红, 等. 动态异构特征融合的水下图像增强算法[J]. 控制与决策, 2023, 38(6): 1560-1568.
(Han X W, Zhang Y Z, Xie Y H, et al. Underwater image enhancement algorithm based on dynamic heterogeneous feature fusion[J]. Control and Decision, 2023, 38(6): 1560-1568.)
- [31] 于明, 李学博, 郭迎春. 融合注意力机制的域泛化行人再识别[J]. 控制与决策, 2022, 37(7): 1721-1728.
(Yu M, Li X B, Guo Y C. Domain generalization person re-identification based on attention mechanism[J]. Control and Decision, 2022, 37(7): 1721-1728.)
- [32] 刘芳, 浦昭辉, 张帅超. 基于注意力特征融合的无人机多目标跟踪算法[J]. 控制与决策, 2023, 38(2): 345-353.
(Liu F, Pu Z H, Zhang S C. UAV multi-target tracking algorithm based on attention feature fusion[J]. Control and Decision, 2023, 38(2): 345-353.)
- [33] 石磊, 张晓涵, 洪晓鹏, 等. 多尺度梯度对抗样本生成网络[J]. 模式识别与人工智能, 2022, 35(6): 483-496.
(Shi L, Zhang X H, Hong X P, et al. Multi-scale gradient adversarial examples generation network[J]. Pattern Recognition and Artificial Intelligence, 2022, 35(6): 483-496.)
- [34] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 7132-7141.
- [35] Hu J, Shen L, Albanie S, et al. Gather-excite: Exploiting feature context in convolutional neural networks[J/OL]. 2018, arXiv: 1810.12348.
- [36] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[C]. Proceedings of the European conference on computer vision. New York, 2018: 3-19.
- [37] DeVries T, Taylor G W. Improved regularization of convolutional neural networks with cutout[J/OL]. 2017, arXiv: 1708.04552.

作者简介

潘杰(1986—), 男, 副教授, 博士, 从事深度学习、网络生理学等研究, E-mail: panjie1616@cumt.edu.cn;

郑学驰(1999—), 男, 硕士生, 从事神经网络结构搜索、计算机视觉等研究, E-mail: zhengxc@cumt.edu.cn;

邹筱瑜(1990—), 副教授, 博士, 从事深度学习、机器视觉与装备运维等研究, E-mail: zouxiaoyu@cumt.edu.cn.