

引用格式: CAO Bin, ZHENG Enrang, SHEN Junge. Neural network attention architecture search for optical remote sensing image scene classification[J]. Remote Sensing Technology and Application, 2023, 38(4): 913-923. [曹斌, 郑恩让, 沈钧戈. 基于神经网络注意力架构搜索的光学遥感图像场景分类[J]. 遥感技术与应用, 2023, 38(4): 913-923.]
DOI: 10.11873/j.issn.1004-0323.2023.4.0913

基于神经网络注意力架构搜索的光学遥感图像场景分类

曹斌¹, 郑恩让¹, 沈钧戈²

(1. 陕西科技大学 电气与控制工程学院, 陕西 西安 710021;
2. 西北工业大学 无人系统技术研究院, 陕西 西安 710072)

摘要: 针对光学遥感图像场景分类存在类别变化、样本数量变化, 场景图像中背景与重要物体变换大、尺度变化多的问题, 提出基于神经网络注意力架构搜索的光学遥感图像场景分类方法, 由算法自适应在神经网络中搜索卷积、池化、注意力等操作, 构建能完成光学遥感图像场景分类任务的神经网络。为保证搜索神经网络过程稳定性, 提出两段式贪婪策略网络搜索方法, 分阶段丢弃无用操作, 减少搜索算法负担、提高搜索速度。最后为了关注各物体与场景关联信息, 提出自上而下的网络连接策略, 充分复用各阶段多尺度特征图的语义。实验结果证明: 该方法相较于手工设计的经典深度学习方法具有更好的性能。在AID、NWPU、PATTERNET 3个遥感图像标准数据集上总体精度均超过经典方法。在AID数据集上准确率达到94.04%; 在PATTERNET数据集上准确率达到99.62%; 在NWPU数据集上达到95.49%。

关键词: 遥感; 场景分类; 神经网络架构搜索; 贪婪算法; 网络连接策略

中图分类号: TP751; TP18 **文献标志码:** A **文章编号:** 1004-0323(2023)04-0913-11

1 引言

近年来, 随着遥感观测技术不断升级, 越来越多高分辨率遥感图像数据被获取^[1-2], 如何高效和充分利用这些数据成为学术界的研究重点。遥感场景分类^[3]任务作为分析遥感数据的基础性工作存在数据数量和数据类别不断增多的情况, 传统深度学习方法对这种情况有两种解决方案: 第一种根据新数据集特点重新构建网络, 然后重新训练神经网络。这种方法耗费大量时间而且需要很多的人力来调整网络结构。第二种使用迁移学习的方法^[4], 以自然图像分类的网络作为基础网络, 在导入自然图像分类任务上预训练完善的权重后, 输入遥感图

像进行再训练。这种解决方案能节约训练时间, 但由于经典网络和预训练权重并不是专门对应遥感图像问题设置, 所以得到的网络往往并不是新数据集的最优网络。实验使用神经网络架构搜索方法^[5], 通过算法从搜索空间中自适应取出各种网络操作, 算法自动根据网络构建策略构成适用于遥感图像场景分类问题的网络架构, 这样既节约人力算力, 又得到适用于目标场景分类问题的神经网络。

在遥感场景分类图像问题上, 往往存在图像场景复杂的问题, 而且各类场景中具有代表性的不同尺度物体。如何利用复杂的场景信息和复杂的多尺度物体成为解决遥感场景分类任务的重点研究方向。Xu等^[6]根据复杂场景和多尺度物体提出—

收稿日期: 2021-04-15; 修订日期: 2021-06-24

基金项目: 国家自然科学基金项目(61603233), 河南省水下重点实验室开放基金项目(D5204200587)。

作者简介: 曹斌(1997—), 男, 河南新郑人, 硕士研究生, 主要从事深度学习、计算机视觉、遥感图像分析研究。E-mail: caobnas@163.com

通讯作者: 郑恩让(1962—), 男, 陕西凤翔人, 教授, 主要从事智能信息处理研究。E-mail: zhenger@sust.edu.cn

种软注意力机制和反馈注意力机制的结合。Wang 等^[7]提出一种循环注意力架构重点在高层语义特征中通过注意力机制提取全局信息。实验将采用神经网络架构搜索的方式自适应构建注意力操作在网络模块中的位置,通过算法充分利用注意力机制的特点,充分发掘各个层中语义信息。根据遥感分类任务,提出通道注意力机制的卷积搜索空间,使得网络可以通过通道注意力机制重点关注通道语义信息。Ye 等^[8]通过改进网络连接策略,提取网络不同阶段语义信息进行融合,利用不同尺度物体和背景的语义特征完成分类。实验根据神经网络架构搜索任务,提出网络连接方式融合不同阶段特征信息。

综上,本实验提出基于神经网络注意力架构搜索的光学遥感图像场景分类方法具有以下优点:①为使深度学习算法自动搜索适合遥感场景分类问题的网络,提出基于神经网络注意力架构搜索的一种自动深度学习范式来处理光学遥感图像场景分类问题,节省人力和计算机算力;②针对遥感场景特征图在不同通道中稀疏排列的特点,提出具有通道注意力操作的搜索空间;③为增强架构搜索稳定性并提升搜索算法对注意力操作的适应性,提出两段式贪婪策略网络搜索方法;④针对各特征层中同源异构的场景语义信息,提出一种自上而下能融合 3 个阶段特征图的网络连接策略。

2 神经网络架构搜索基础

近年来,随着自动深度学习技术的发展,3 种自动深度学习方法成为学界研究重点:自动特征工程^[9]、元学习^[10]、神经网络架构搜索^[11]。其中神经网络架构搜索在图像分类和目标检测任务上取得了优秀的效果。神经网络架构搜索主要思想是通过智能算法从搜索空间取出构成神经网络的基础操作,使用一定的网络搜索策略取出操作,然后根据网络连接策略构成能完成特定任务的神经网络。这种智能的方法省去了人工从头构建网络和开发人员不断尝试调整网络参数的时间成本,极大地解放了科研人员在基础构建网络问题上的精力。

神经网络架构搜索算法主要有 3 部分技术流程(图 1):搜索空间(Search space)、搜索阶段(Search phase)和评估阶段(Evaluate phase),搜索空间主要定义算法或智能体使用的操作(operation),这些操作构成一个组成神经网络的基础组成超图(Super Graph)。搜索阶段算法或智能体搭配和使用搜索

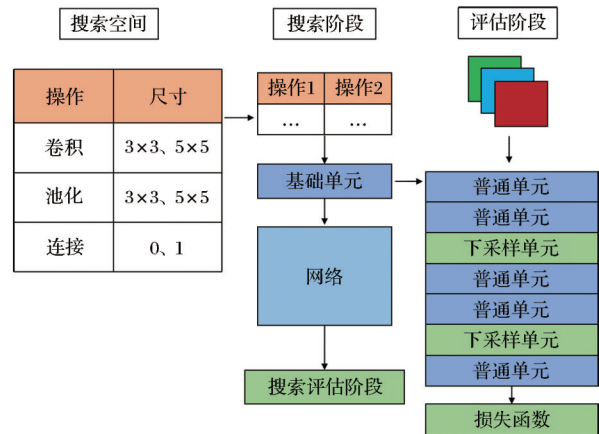


图 1 神经网络架构搜索技术流程

Fig.1 Neural network architecture search technical process

空间的操作进行搜索最终构成基础单元,并根据网络连接策略连接基础单元成神经网络。在评估阶段使用各种搜索得到的基础单元完整使用神经网络在特定任务上,例如遥感图像场景分类问题在评估阶段输入遥感图像数据进行训练,在评估阶段结束时得到搜索获得神经网络架构的性能和准确率评价。

神经网络架构搜索算法根据搜索策略的不同分为 3 种:进化算法^[12]、强化学习^[13]、基于梯度的松弛化搜索^[14]。进化算法主要通过编码网络操作和结构信息生成种群,通过遗传算法、进化算法等方式变异编码信息中的各种操作成为新种群,然后根据新种群解码文本信息复现网络并从头训练网络,依次循环变异直到获得满意网络。强化学习通过构建一个智能体生成神经网络信息,智能体从搜索空间中取出操作构成网络,使用评估策略对生成神经网络性能进行评估,然后智能体根据评估结果进行调整,最终获得效果不错的神经网络。基于梯度的松弛化搜索方法通过松弛搜索空间中的操作成一个超图,每个操作是超图中的一个边,对每条边赋予权重,训练每条边的权重并根据最终训练权重排序决定组成网络的所有操作。

由于进化算法和强化学习在重构神经网络和训练神经网络过程中,需要很大的计算机算力,目前神经网络架构搜索发展方向主要是基于梯度松弛化搜索方式,从梯度角度搜索网络的算力消耗最少只需要几块 GPU。Xu 等^[15]通过修改梯度松弛化搜索方法的网络通道连接策略,对每阶段网络连接的通道进行随机采样,只取一半的通道通过网络操作进行运算,这样能减小一半以上计算开销,提升搜索速度和效率。Liang 等^[16]使用早停机制,在搜

索过程中根据搜索方法提前停止搜索,解决了搜索时间过长导致网络跳跃连接过多,使搜索获得网络陷入局部最优而且网络特征提取能力差的问题。

3 神经网络注意力架构搜索方法

3.1 注意力搜索空间设置

实验根据遥感图像场景信息特点,为使搜索得到的网络具有综合多尺度物体和背景的能力,如表1所示,在搜索空间中引入通道注意力模块^[17]。使算法能针对遥感图像类间和类内的差异提取重点语义信息,并且算法能智能地在卷积、池化操作的基础上自适应地在隐藏层合适的位置加入合适的注意力操作。

实验在神经网络架构搜索空间内引入通道注意力机制,如图2所示,通道注意力机制主要从通道

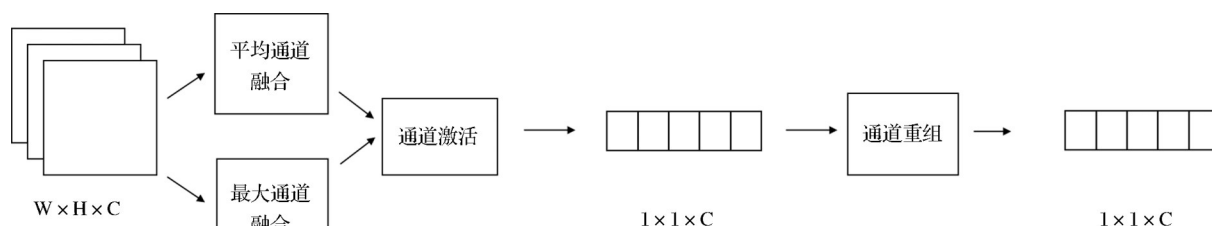


图2 通道注意力机制

Fig.2 Channel attention mechanism

对于搜索空间采用松弛化操作,将搜索空间所有操作松弛为超图中两节点之间的边,超图中的节点代表两操作输入特征图和输出特征图。本方法设置多个节点构成两种构成网络的单元:基础单元(normal cell)和下采样单元(reduction cell)。在搜索阶段主要是为了确定两种组成网络基础单元的各操作,利用梯度的反向传播训练每条边上各操作的权重,最终根据网络搜索方法确定每条边上最重要的权重,固定基础单元和下采样单元的所有操作。

基础单元和下采样超图松弛化构建计算公式如下所示:

$$\bar{o}_{ij}(x) = \sum_{o \in O} \frac{\exp(\alpha_o^{ij})}{\sum_{o' \in O} \exp(\alpha_o^{ij})} o(x) \quad (1)$$

其中:将搜索空间(如表1所示)所有操作表示为 $O = \{o_1, o_2, o_3, o_4, \dots\}$,利用softmax函数松弛化所有操作 o 在 i 节点到 j 节点的权重 α_o^{ij} ,最终获得两节点间对特征图 x 的操作 $\bar{o}_{ij}(x)$ 。所有节点间操作从松弛化设置为超图到构建网络过程如图3所示。

表1 注意力搜索空间操作及尺寸

Table 1 Operation and size of attention search space

操作	尺寸
空洞卷积	$3 \times 3, 5 \times 5$
深度可分离卷积	$3 \times 3, 5 \times 5$
平均池化	$3 \times 3, 5 \times 5$
最大池化	$3 \times 3, 5 \times 5$
通道注意力操作	1
无操作	0
跳跃操作	1

维度进行全局平均池化和全局最大池化,提取通道数为 C 、长和宽为 W 和 H 的特征图各通道全局语义,然后对提取出的特征进行非线性激活,最终通过全连接层对特征图进行特征重组。通道注意力机制能挖掘各特征图在通道维度上的重要信息,增强对图像背景语义信息和各场景物体语义信息的关注。

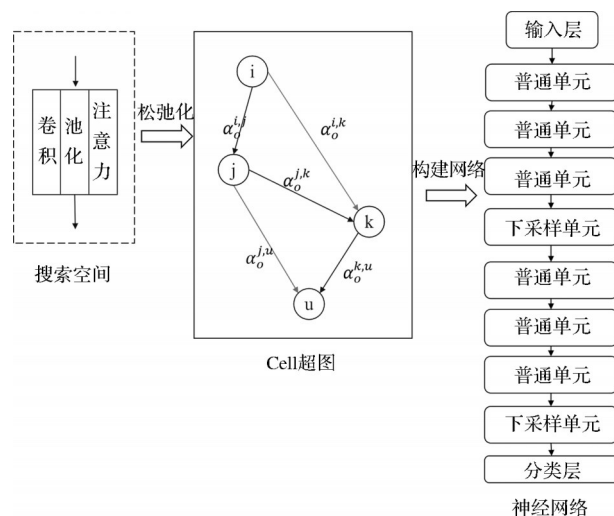


图3 松弛化操作超图和构成神经网络

Fig.3 Relaxation operation hypergraph and construct neural network

3.2 自上而下式网络连接策略

实验在神经网络架构搜索网络级联连接策略的基础上,提出自上而下式网络连接策略,针对遥感图像场景分类问题,融合浅层特征中小物体的语

义信息,中层特征中大中型物体和局部背景语义信息和深层特征全局场景语义信息。重点关注遥感图像中场景与物体之间与对应类别的关系。具体实现为:从网络级增加网络特征融合操作,融合网络对浅、中、深三阶段特征语义。

在网络构建最终评估、测试阶段,采用本方法三阶段网络连接策略结合普通单元和下采样单元构成网络。以下采样单元分割网络为3个阶段,分别提取浅层、中层和深层特征语义。网络连接策略如图4所示。

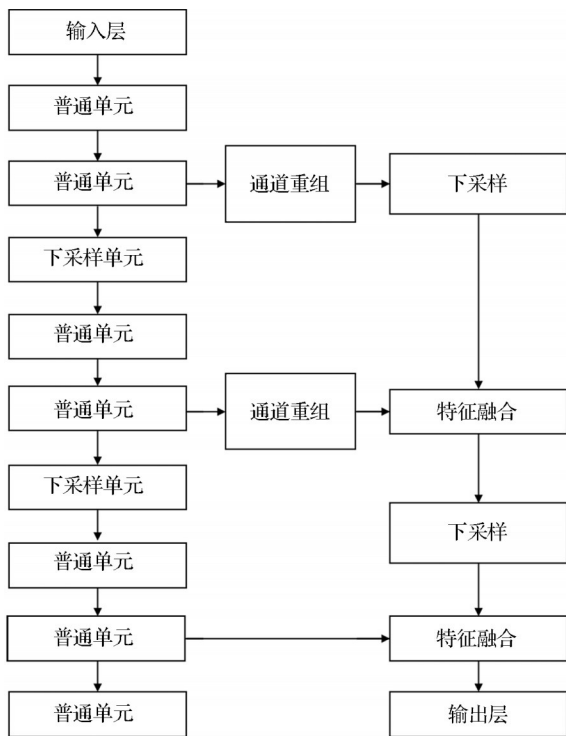


图4 自上而下式网络连接策略

Fig.4 Top-to-down network connection strategy

实验采用自上而下的特征融合策略,对浅层特征图进行下采样,然后将下采样获得的特征图与调整过通道的中层特征进行融合,之后对中层特征图和深层特征图重复以上操作,最终根据3层融合后的结果进行分类,自上而下网络架构设置如表2所示。

自上而下的三阶段结构分别使用通道重组(Channel shuffle)、特征融合(Merge)、 3×3 卷积下

表2 自上而下网络架构设置

Table 2 Top down network architecture setting

特征层	单元	操作
浅层特征	浅层普通单元	通道重组、下采样
中层特征	中层普通单元	通道重组、特征融合、下采样
深层特征	深层普通单元	特征融合

采样(Down Sample)3种操作。

对浅层特征图使用通道重构后进行下采样操作,然后与通道重构后的中层特征图进行特征融合,浅层中层特征图融合的计算公式为式(2):

$$y_{M,L} = y_M * k_M + \text{DownSample}(y_L) * k_L \quad (2)$$

其中: $y_{M,L}$ 表示网络浅中层融合特征图; y_M 表示中层特征图; $*$ 表示二维卷积操作; k_M 表示对中层特征图使用 1×1 大小的通道调整卷积; $\text{DownSample}(y_L)$ 表示 3×3 卷积步长为2的下采样操作; y_L 表示浅层特征图; k_L 表示对浅层特征图使用 1×1 大小的通道调整卷积。

然后对浅中层特征图使用下采样操作后与通道重构的浅层特征图进行特征融合,融合操作的计算公式为式(3):

$$y_{out} = y_H + \text{DownSample}(y_{M,L}) \quad (3)$$

其中: y_{out} 表示网络深中浅层融合后网络输出特征图; y_H 表示深层特征图; $\text{DownSample}(y_{M,L})$ 表示 3×3 卷积步长为2的下采样操作; $y_{M,L}$ 表示浅中层融合特征图。

3.3 两段式贪婪策略搜索方法

传统神经网络架构搜索方法随着搜索周期增加,网络算法会更倾向于选择跳跃连接和无连接。为解决上述问题本实验提出两段式贪婪策略搜索算法,在搜索过程中分阶段增加网络层数,并在各阶段结束时使用贪婪策略保留搜索空间中最合适场景分类任务的前几种操作。两段式贪婪策略搜索方法如图5所示。

本方法在搜索阶段过程中,将搜索过程分为两个阶段:操作筛选阶段和单元固定阶段。在操作筛选阶段为了避免超图操作过多导致的参数量巨大,使用较少的普通单元和下采样单元构成层数较少的神经网络,对网络参数进行优化。训练阶段的参数有两种:网络权重 w 和操作权重 α ,这样搜索网络就成了一个二阶优化问题。搜索阶段训练环节如式(4)所示:

$$\begin{cases} \min_w L_{train}(\alpha^*(w), w) \\ \alpha^*(w) = \arg \min_{\alpha} L_{val}(w, \alpha) \end{cases} \quad (4)$$

其中: L_{train} 为训练环节交叉熵损失函数, $\alpha^*(w)$ 为在网络权重为 w 时的固定架构权重 α 。搜索阶段验证环节为式(5):

$$\begin{cases} \min_w L_{val}(w^*(\alpha), \alpha) \\ w^*(\alpha) = \arg \min_w L_{train}(w, \alpha) \end{cases} \quad (5)$$

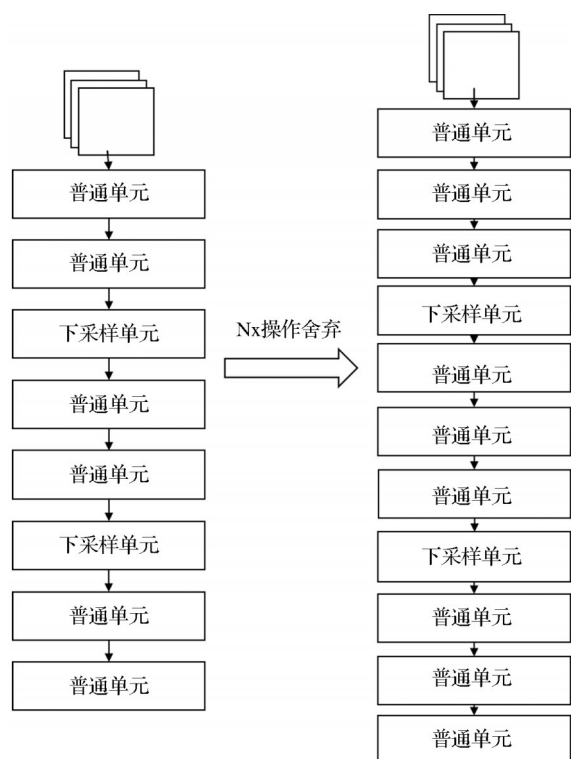


图5 两段式网络搜索策略

Fig.5 Two stage network search strategy

其中: L_{val} 为验证环节交叉熵损失函数; $w^*(\alpha)$ 为在架构权重为 α 时的固定网络权重 w 。

在操作筛选阶段先在固定架构权重 α_1 ,通过

L_{train} 训练阶段损失函数反向传播优化网络权重 w 。然后再固定网络权重 w ,通过 L_{val} 验证阶段损失函数反向传播优化架构权重 α_1 。依此循环直到第一阶段搜索结束,本方法根据架构权重 α_1 的数值大小,使用贪婪策略取出最优的几个操作,设定为第二搜索阶段新的搜索空间和架构权重 α_2 。

在单元固定阶段使用和第一阶段同样的设置,循环搜索到第二阶段结束后,取出各节点间最优的架构权重 α_2 对应的操作,以此固定全局超图的所有操作,得到最终的基础单元和下采样单元。

4 实验结果与分析

本节中,在3个标准公开数据集设置实验以验证本方法的有效性。首先介绍了实验数据集的设置;其次说明实验评估指标和超参数设置;最后与最新的一些方法进行比较,讨论本方法的优越性。

4.1 数据集

实验在3个标准公开数据集进行实验,如表3所示,标注数据集分别是:Aerial Image Data Set(AID)^[18]、NWPU-RESISC45(NWPU)^[19]、PatternNet data set (PatternNet)^[20]。数据集训练比例设置与其他对比方法一致。

表3 标准数据集信息

Table 3 Standard dataset information

数据集	每类图片数/张	类别数/种	全部图片数/张	图片大小	数据集年份
AID	220~420	30	10 000	600×600	2017年
NWPU	700	45	31 500	256×256	2016年
PatternNet	800	38	30 400	256×256	2017年

根据神经网络架构搜索特点在搜索和评估阶段对数据集使用不同比例设置,如表4所示。

表4 3个标准数据集上的实验设置

Table 4 Experimental setup on three standard datasets

数据集	搜索阶段	评估阶段
AID	80%:20%	50%:50%
NWPU	80%:20%	60%:40%
PatternNet	80%:20%	50%:50%

4.2 评价指标

实验使用两种评价指标验证算法和模型精度,总体分类精度和混淆矩阵。

(1)总体分类精度(Overall Accuracy, OA):全部分类正确的样本数占整体样本数的比例,计算方法如式(6)所示:

$$OA(\%) = \frac{1}{N} \sum_{i=1}^N a_i \quad (6)$$

其中: N 为数据集类别总数; a_i 为第 i 类数据的分类精度。

(2)混淆矩阵(Confusion Matrix, CM):在矩阵中表示预测正确和错误的类别数目或准确率,行和列分别代表预测类别和真实类别,对应元素反映了网络对物体的正确检测和错误检测的数量或准确率。

4.3 超参数讨论

(1)搜索设置:在搜索阶段使用两个优化器优化架构参数和网络权重,使用自适应矩估计算法(Adam)^[21]优化网络架构参数,学习率设置为0.000 5,动量项设置中一阶动量项设置为0.5、二阶动量项设置为0.999,权重衰减系数为0.000 3。使用动量优

化的随机梯度下降(SGD)的参数更新方法优化网络权重,学习率按照余弦退火学习率设置学习率从最高 0.025 衰减到最低 0.001,余弦退火的 1/4 周期设置为每阶段总训练轮次,权重衰减系数为 0.004。本实验均在 python3.6.10 上进行,硬件平台为 Intel CoreI7-8700CPU, GPU 为两张显存为 11 GB 的 NVIDIA GeForce GTX 2080Ti,内存为 16 GB。

(2)评估设置:在评估阶段使用动量优化的随机梯度下降优化网络参数,余弦退火学习率按余弦规律设置学习率从最高 0.025 衰减到最低 0.005,余弦退火的 1/4 周期设置为整个训练轮次。动量项设置为 0.5。

(3)单元层数:神经网络架构搜索通过优化基础单元超图最终固定操作成一个网络基础单元,在网络层的 1/3、2/3 处设置下采样单元,剩余位置使用基础单元构成。不同的网络单元设置对网络的分类效果会产生较大的影响。实验通过设置消融实验,在 AID 数据集上按训练集:验证集 4:6 设置,验证单元层数对网络准确率的影响。实验结果如表 5 所示,在 10 层时网络分类效果最好,达到 92.83% 总体精度,在单元层数增加和减少时网络分类效果开始变差。

(4)两段式贪婪策略消融实验:实验设置两段式贪婪策略超参数,一阶段搜索轮次 25 轮,网络单元层数为 7,网络通道数为 16,一阶段结束丢弃 3 个操作。二阶段搜索轮次 35 轮,网络单元层数为 10,网络通道数为 20,二阶段结束保留各节点间架构权

表 5 不同单元层数网络分类总体精度

Table 5 Overall accuracy of network classification with different cell layers

网络层数	网络总体精度/%
9	91.80
10	92.83
12	91.98
15	91.30
20	90.80

重最高的两个操作。本研究设置消融实验,在 AID 数据集上按训练集:测试集 4:6 设置,在搜索阶段搜索网络基础单元。使用网络基础单元并在评估阶段按训练集:测试集 4:6 设置对网络进行评估,验证两段式贪婪策略对网络准确率的影响。

搜索基础单元与下采样单元如图 6 所示,图中各操作如表 6 所示,搜索结果的总体精度如表 7 所示,由此看出,本实验与 3 种搜索策略进行对比,单

表 6 单元图操作对照表

Table 6 Cell operation comparison table

名称	操作
none	无连接
skip_connect	跳跃连接
max_pool_3×3	3×3 最大池化
avg_pool_3×3	3×3 平均池化
sep_conv_3×3	3×3 空洞卷积
sep_conv_5×5	5×5 空洞卷积
dil_conv_3×3	3×3 深度可分离卷积
dil_conv_5×5	5×5 深度可分离卷积

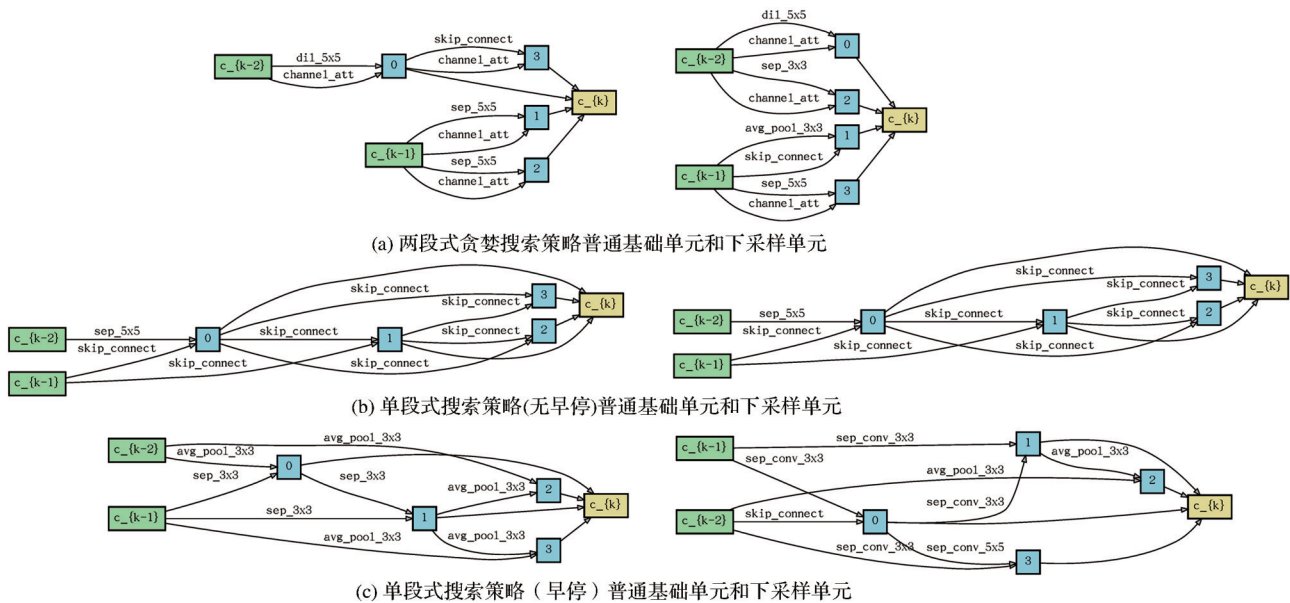


图 6 AID 数据集搜索单元结果

Fig.6 AID dataset search cell results

段式搜索策略结果如图6(b)所示,在搜索末期会大量选择跳跃操作放弃使用注意力操作,这样严重影响分类准确率,使神经网络中有大量无用操作。单段式搜索策略使用早停后结果如图6(c)所示,获得了较好的分类效率91.8%,但是网络中仍有很多无用的跳跃操作。本方法搜索出的基础单元如图6(a)所示,构成网络取得最优分类效果92.83%。本实验会让搜索单元中注意力机制与跳跃连接同时使用,这样既保证两特征层之间单独使用注意力机制又减少跳跃连接,通过实验结果可以证明搜索策略的优异性。

表7 网络搜索策略总体精度

Table 7 Overall accuracy of network search strategy

网络搜索策略	总体精度/%
单段式搜索策略(早停)	91.80
单段式搜索策略(无早停)	90.50
两段式贪婪策略	92.83

4.4 标准数据集分类性能对比

讨论本实验方法和深度学习经典方法以及神经网络架构搜索最新方法之间的性能对比,评价标准为OA和CM。

(1) Aerial Image Data Set(AID)

如表8所示,一些使用迁移学习的经典神经网络和经典神经网络架构搜索方法,在AID数据集上总体精度。在训练集比例50%时进行对比,本实验方法取得了最高的总体精度94.04%。与使用迁移学习的效果最高的手工设计神经网络Resnet-50对比,总体精度提高2.06%。与经典神经网络架构搜索算法DARTS对比,总体精度提升0.84%。同时表8中给出神经网络架构搜索算法与经典神经网络

表8 在AID数据集上总体精度

Table 8 Overall accuracy on AID data set

方法	50% 训练集 比例 OA/%	(搜索时间)训练时间 /Gpu-days
VGG-16(pretraining) ^[22]	91.58	0.6
Resnet-50(pretraining) ^[23]	91.98	0.9
GoogLeNet(pretraining) ^[24]	89.69	0.7
DARTS ^[13] (Early stop)	93.20	(2.1)0.7
Our Method	94.04	(1.3)0.6

在训练时间和搜索时间上的对比,可以看出本算法在搜索速度为1.3 GPU-days超过经典算法DARTS。在评估阶段网络训练时间为0.6 GPU-days,训练速度接近VGG-16,这意味着本方法对各

种全新数据集的适应能力强,构建网络速度快,同时构建的网络有较快的收敛速度。

通过混淆矩阵分析在AID数据集上实验结果如图7所示,分类准确率小于90%的样本类别分别是center、industrial、park、resort、school、square,在这几类图片中产生了错误分类情况,是因为个别图片有很强的类间相似性,通常会降低准确率。在这种情况下,本实验仍取得了最高的总体精度。

(2) PatternNet Data Set(PatternNet)

如表9所示,本实验前面已经在30类别数据集AID上证明了本算法的优越性。在PatternNet数据集,将场景分类类别提升到38类,并且图片总数提升到30 400张,增加少量数据类别和各类别图片数。与训练集比例50%进行对比,本文方法取得了最高的总体精度99.62%。与使用迁移学习总体精度最高的VGG-16神经网络对比,总体精度提高1.31%。与经典神经网络架构搜索算法DARTS对比,总体精度提升1.37%。

在PatternNet数据集上单元搜索结果如图8所示,可以看出本研究搜索策略充分利用了搜索空间中的各种操作,最终得到效果较好的网络基础单元。在PatternNet数据集上实验获得混淆矩阵如图9所示,在该数据集上每一类分类效果都在99%以上,只有harbor类别取得97%的准确率,其中有几张图片错分为ferry terminal分类。证明本方法可以很好地分辨该数据集不同类别的数据,并且本实验方法在各种方法对比中取得了最高的总体精度。

(3) NWPU-RESISC45(NWPU)

如表10所示,在NWPU数据集上增加场景类别到45类,并进一步提高各类图片数量。通过实验结果可以看出本方法在应对数据类别和数量不断增加的情形下,对遥感图像场景分类任务都能很好地适应。与最好效果的使用迁移学习手工设计神经网络Resnet-50神经网络对比,总体精度提升3.86%。与神经网络架构搜索算法DARTS对比,总体精度提升2.39%,均有较大的提升。

如图10所示,在NWPU数据集上搜索获得的基础单元也充分使用搜索空间中的各种操作,在测试阶段同样获得了最高的总体分类精度。

本实验方法在类别最高的NWPU数据集上实验得到混淆矩阵如图11所示,分类准确率小于90%的只有church和commercial area类别,church中个别图像被分类成parking lot类别,Commercial类别

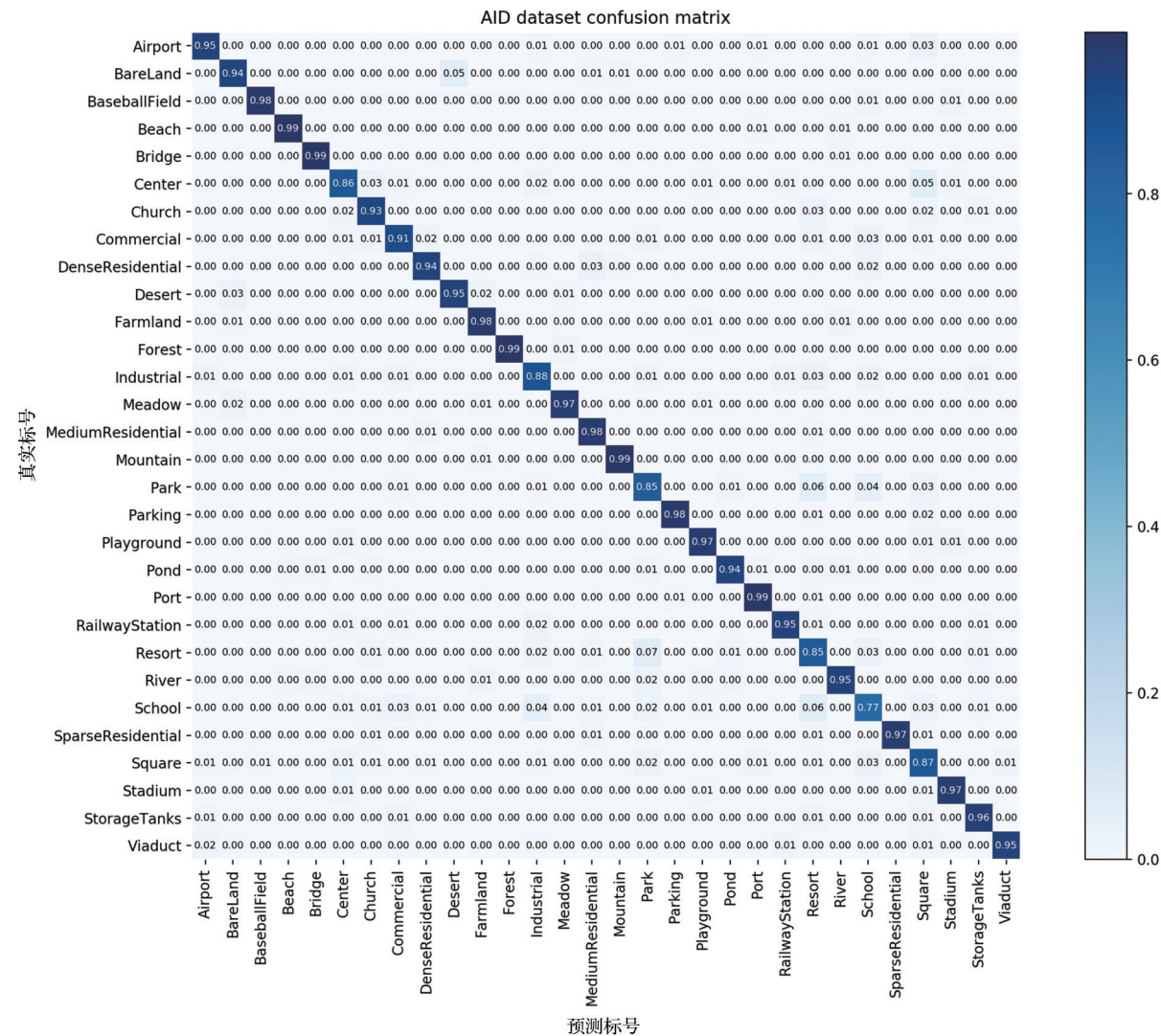


图 7 AID数据集 50%训练样本混淆矩阵

Fig.7 confusion matrix of 50% training samples in AID dataset

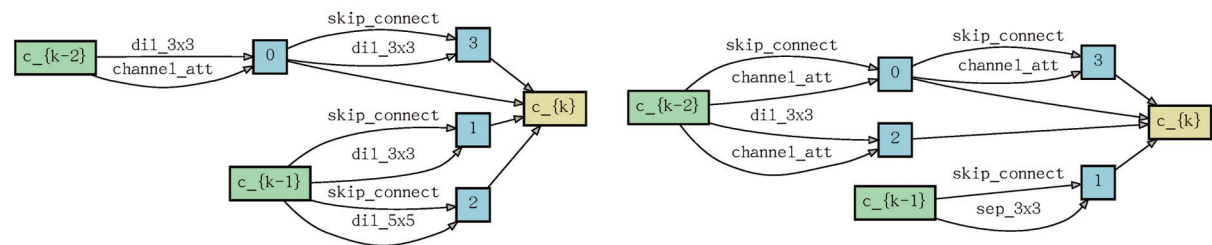


图 8 PatternNet数据集搜索基础单元和下采样单元结果

Fig.8 Results of searching basic cell and down-sampling cell in PatternNet dataset

表 9 在 PatternNet数据集上总体精度

Table 9 Overall accuracy on PatternNet dataset

方法	训练集比例 50%OA/%
VGG-16(pretraining)	98.31
Resnet-50(pretraining)	98.23
GoogLeNet(pretraining)	97.56
DARTS(Early stop)	98.25
Our Method	99.62

表 10 在 NWPU数据集上总体精度

Table 10 Overall accuracy on NWPU data set

方法	训练集比例 60%OA/%
VGG-16(pretraining)	91.32
Resnet-50(pretraining)	91.63
GoogLeNet(pretraining)	89.42
DARTS(Early stop)	93.04
Our Method	95.49

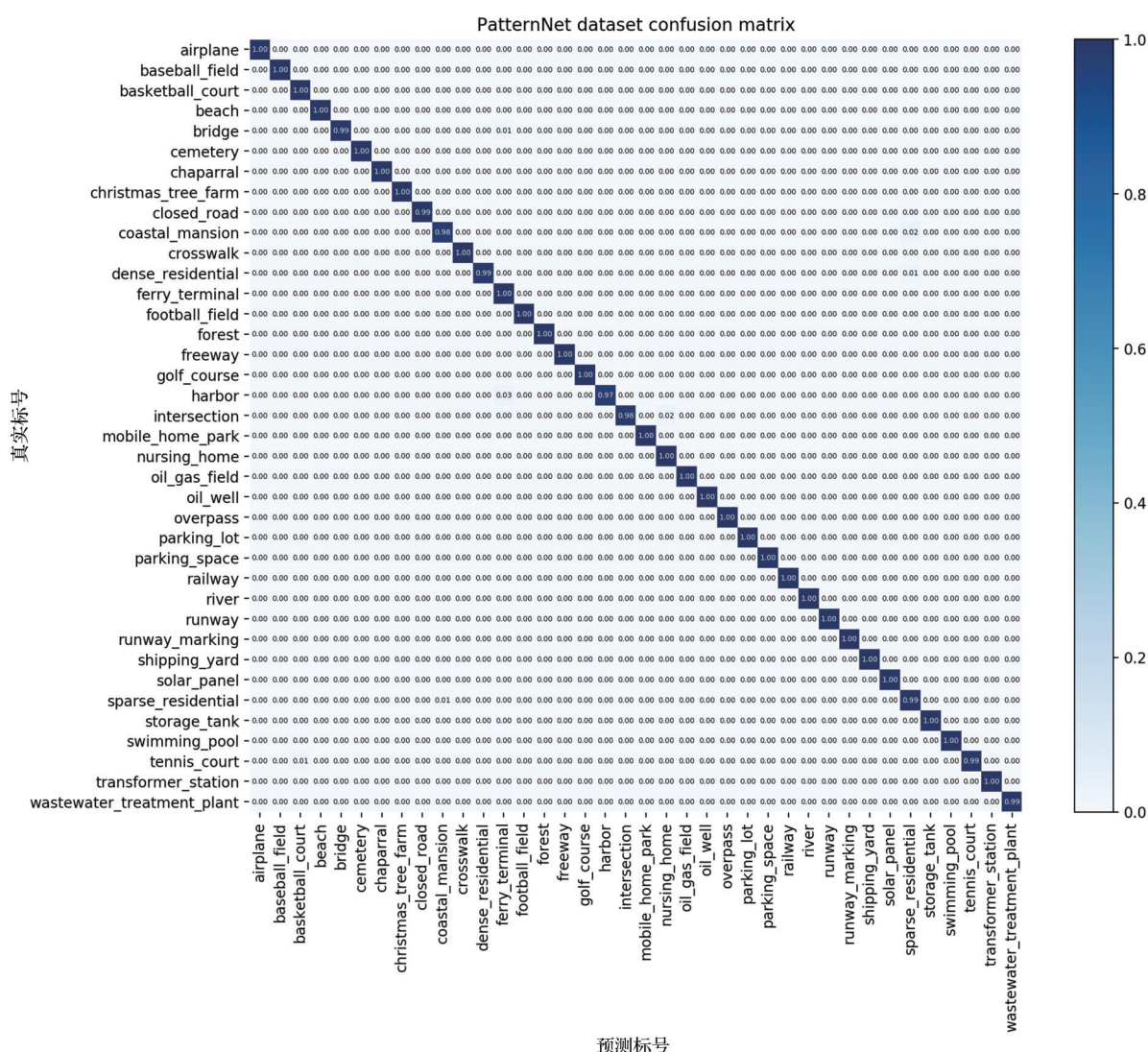


图 9 PatternNet数据集 50% 训练样本混淆矩阵

Fig.9 Confusion matrix of 50% training samples in PatternNet dataset

容易被分类为 church,这是因为 church 中包含 parking lot 类别和 commercial 类别的局部语义信息。但本实验方法在其余类别都取得了很高的分类精度,特别对比其他方法取得了最高的总体分类精度。

5 结 论

实验提出一种适合遥感图像场景分类问题的

神经网络注意力架构搜索方法,能较好应对遥感图像数据类别和数据量不断变化的问题,降低研究人员手工设计网络的时间成本和试错成本。同时,提出网络连接策略关注遥感场景图像特有的复杂背景和多尺度场景物体问题。实验结果表明,本方法能够稳定地构建适合遥感图像场景分类任务的注意力神经网络。同时,本方法在遥感图像场景分类

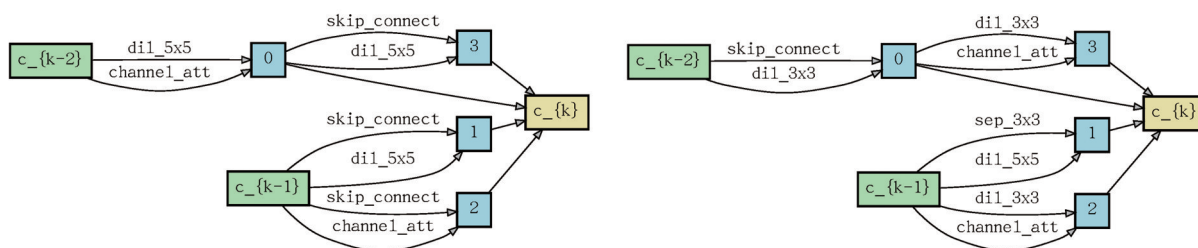


图 10 NWPU 数据集搜索基础单元和下采样单元结果

Fig.10 Results of searching basic cell and down-sampling cell in NWPU data set

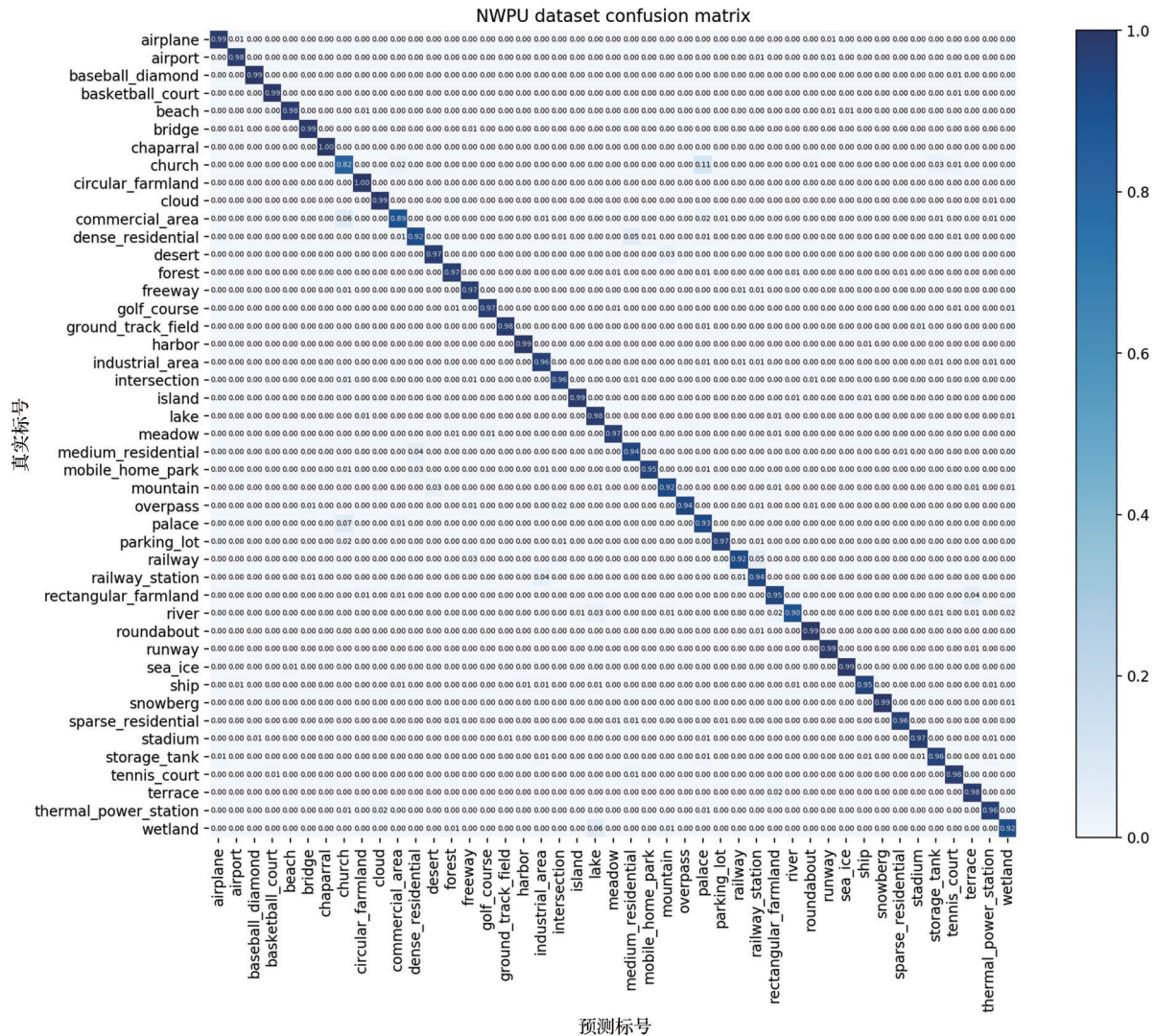


图11 NWPU数据集60%训练样本混淆矩阵

Fig.11 Confusion matrix of 60% training samples in NWPU dataset

任务上能获得与人工设计网络具有竞争力的效果。未来将从网络特征提取能力和网络特征复用能力出发,进一步提升网络对遥感图像特征的挖掘能力。

参考文献 (References):

- [1] TONG Qingxi, ZHANG Bing, ZHANG Lifu. Advance in hyperspectral remote sensing in China [J]. Acta Remotica Sinica, 2016, 20(5): 689-707. [童庆禧, 张兵, 张立福. 中国高光谱遥感的前沿进展[J]. 遥感学报, 2016, 20(5): 689-707.]
- [2] GONG Peng, LI Xia, XU Bing. Some problems in the theory and application of high resolution image interpretation [J]. Acta Remotica Sinica, 2006, 10(1): 1-5. [宫鹏, 黎夏, 徐冰. 高分辨率影像解译理论与应用方法中的一些研究问题[J]. 遥感学报, 2006, 10(1): 1-5.]
- [3] CHENG G, YANG C, YAO X, *et al.* When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs [J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(5): 2811-2821.
- [4] HAN X, ZHONG Y, CAO L, *et al.* Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification [J]. Remote Sensing, 2017, 9(8): 848. DOI: 10.3390/rs9080848
- [5] ELSKEN T, METZEN J H, HUTTER F. Neural architecture search: A survey [J]. arXiv preprint arXiv:1808.05377, 2018.
- [6] XU R, TAO Y, LU Z, *et al.* Attention-mechanism-containing neural networks for high-resolution remote sensing image classification [J]. Remote Sensing, 2018, 10(10): 1602. DOI: 10.3390/rs10101602
- [7] WANG Q, LIU S, CHANUSSOT J, *et al.* Scene classification with recurrent attention of VHR remote sensing images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 57(2): 1155-1167. DOI: 10.1109/tgrs.2018.2864987
- [8] YE L, WANG L, SUN Y, *et al.* Parallel multi-stage features fusion of deep convolutional neural networks for aerial scene classification [J]. Remote Sensing Letters, 2018, 9(3): 294-303. DOI: 10.1080/2150704x.2017.1415477
- [9] KAUL A, MAHESHWARY S, PUDI V. Autolearn—auto-

- mated feature generation and selection[C]// 2017 IEEE International Conference on Data Mining (ICDM). IEEE, 2017: 217-226.
- [10] VILALTA R, DRISSI Y. A perspective view and survey of meta-learning[J]. Artificial Intelligence Review, 2002, 18(2): 77-95.
- [11] ELSKEN T, METZEN J H, HUTTER F. Neural architecture search: A survey[J]. arXiv preprint arXiv:1808.05377, 2018.
- [12] REAL E, AGGARWAL A, HUANG Y, *et al.* Aging evolution for image classifier architecture search[C]// AAAI Conference on Artificial Intelligence, 2019.
- [13] ZOPH B, LE Q V. Neural architecture search with reinforcement learning[J]. arXiv preprint arXiv:1611.01578, 2016.
- [14] LIU H, SIMONYAN K, YANG Y. Darts: Differentiable architecture search[J]. arXiv preprint arXiv:1806.09055, 2018.
- [15] XU Y, XIE L, ZHANG X, *et al.* PC-DARTS: Partial channel connections for memory-efficient architecture search[J]. arXiv preprint arXiv:1907.05737, 2019.
- [16] LIANG H, ZHANG S, SUN J, *et al.* Darts+: Improved differentiable architecture search with early stopping[J]. arXiv preprint arXiv:1909.06035, 2019.
- [17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141. DOI:10.1109/cvpr.2018.00745
- [18] XIA G S, HU J, HU F, *et al.* AID: A benchmark data set for performance evaluation of aerial scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(7): 3965-3981. DOI:10.1109/tgrs.2017.2685945
- [19] CHENG G, HAN J, LU X. Remote sensing image scene classification: Benchmark and state of the art[J]. Proceedings of the IEEE, 2017, 105(10): 1865-1883.
- [20] ZHOU W, NEWSAM S, LI C, *et al.* PatternNet: a benchmark dataset for performance evaluation of remote sensing image retrieval[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 145: 197-209. DOI:10.1016/j.isprsjprs.2018.01.004
- [21] KINGMA D P, BA J. Adam: a method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.
- [22] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [23] 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, 2016: 770-778.
- [24] SZEGEDY C, LIU W, JIA Y, *et al.* Going deeper with convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.

Neural Network Attention Architecture Search for Optical Remote Sensing Image Scene Classification

CAO Bin¹, ZHENG Enrang¹, SHEN Junge²

(1. School of Electrical and Control Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China;

2. Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: With majority problems in image scene of optical remote, changing category in classification, variational size in sample, diverse changing of scale between backgrounds and essential objectives, for instance, new Classification Algorithm for scene classification of optical remote sensing image base on attention architecture search of neural network is proposed in this paper. This algorithm can search convolution, pooling, attention and other operations in the neural network, adaptively; and complete the construction task of scene classification for optical remote sensing images in neural network. Two-stage greedy algorithms network search is mentioned in order to ensure the stability of neural search network. This method abandons useless operations in stage which can reduce algorithm burden and improve speed of search. Furthermore, a top-bottom connection strategy of network, which can fully reuse the semantics of multi-scale feature maps in each stage, is proposed to merge information between each object and scene. The experimental results proved that the method proposed in this paper has better performance than the classical deep learning method designed by hand. Overall, the accuracy of this method in all three remote sensing image-standard data sets (AID, NWPU and PatterNet) is exceeding the classic method. The accuracy rate of AID data set, PatterNet data set, and NWPU data set are 94.04%, 99.62%, and 95.49%, respectively.

Key words: Remote sensing; Scene classification; Neural network architecture search; Greedy algorithms; Network connection strategy