

文章编号: 1671-7449(2024)03-0315-08

基于特征增强的汽车发动机积碳程度识别模型

张永玲, 黄倩, 陈友兴*, 陈香, 张航佳

(中北大学信息与通信工程学院, 山西太原 030051)

摘要: 汽车发动机积碳的长期累积容易加速汽车老化, 及时检测并清理可以有效延长汽车使用寿命。对此提出了一种基于视觉图像的积碳程度识别方法, 首先针对积碳图像数据量小且类别分布不均的问题对数据进行预处理, 其次基于积碳图像特征分布范围广及细粒度特点设计了一个特征重采样模块, 从空间和通道两个方向加强特征表达, 最后设计了一个轻量化的积碳程度识别模型方便检测部署。结果表明, 与其他方法相比, 该方法具有较高的预测速度, 为179帧/s, 且测试精度为84.5%, 满足实际需求。

关键词: 积碳程度识别; 小样本学习; 细粒度图像; 特征重采样; 轻量化模型

中图分类号: TP391 **文献标识码:** A **doi:** 10.3969/j.issn.1671-7449.2024030

引用格式: 张永玲, 黄倩, 陈友兴, 等. 基于特征增强的汽车发动机积碳程度识别模型[J]. 测试技术学报, 2024, 38(3): 315-322.

ZHANG Yongling, HUANG Qian, CHEN Youxing, et al. Recognition model of automobile engine carbon deposits degree based on feature enhancement[J]. Journal of Test and Measurement Technology, 2024, 38(3): 315-322.

Recognition Model of Automobile Engine Carbon Deposits Degree Based on Feature Enhancement

ZHANG Yongling, HUANG Qian, CHEN Youxing*, CHEN Xiang, ZHANG Hangjia

(School of Information and Communication Engineering, North University of China, Taiyuan 030051, China)

Abstract: Long-term accumulation of carbon deposits in automobile engines can easily accelerate the aging of automobiles. Timely detection and cleaning can effectively prolong the service life of automobiles. In this paper, a carbon deposits degree recognition method based on a visual image is proposed to automatically recognize the degree of carbon deposits and provide guidance for carbon deposits cleanup. Firstly, to address the issue of small data volume and uneven category distribution in carbon deposits images dataset, data preprocessing is conducted. Secondly, aimed at the wide range of feature distribution and fine-grained characteristics of carbon images, a feature resampling module is designed to improve the feature expression from both spatial and channel directions. Finally, a lightweight carbon deposits degree recognition model is designed to facilitate the detection of deployed. The experimental results demonstrate that compared to other methods, the method proposed in this thesis achieves the highest inference speed of 179 frames/s with a testing accuracy of 84.5%, meeting the needs of the industry.

Key words: recognition of the degree of carbon deposits; few-shot learning; fine-grained images; feature resampling; lightweight model

收稿日期: 2023-09-25

基金项目: 山西省回国留学人员科研资助项目(2022-145)

作者简介: 张永玲(1983-), 女, 讲师, 主要从事光电检测、信号与信息处理等研究。E-mail: 5yzyl@nuc.edu.cn。

通信作者: 陈友兴(1978-), 男, 教授, 主要从事自动检测、信号与信息处理、成像与重建技术研究。E-mail: chenyouxing@nuc.edu.cn。

0 引言

汽车发动机积碳的长期累积会对汽车产生较大危害,如堵塞喷油嘴的针阀、阀孔,影响电子喷射系统精密部件的工作性能;致使进气阀关闭不严,导致发动机怠速不稳、油耗增大并伴随尾气排放恶化;附着在活塞顶和气缸盖等部位形成坚硬的氧化物,引起发动机爆震;等等。因此,及时检测并清理汽车发动机积碳非常重要。

针对汽车发动机积碳程度检测的方法可以分为基于物理量的检测和基于视觉的检测。Roe等^[1]、穆大芸等^[2]、杨万里等^[3]、Bhaskar等^[4]利用传感器设备获取汽车特定部位的状态信息量,通过分析油道温度、进气歧管压差、积碳厚度、积碳表面光学信号等物理量并结合相应阈值判断汽车是否存在积碳,通过提前干涉处理预防故障的发生。但基于物理量的检测方式较为繁琐,需要综合多种信息且检测结果不够直观,在实际应用中存在很大的局限性。

基于视觉的方法主要利用图像采集装置进行反馈,相较于物理量的检测方法更直观、易推广。晏娟等^[5]通过氧传感器获取反馈电压值判断是否存在积碳,再利用微型摄像头观测发动机内积碳厚度,通过成像显示屏反馈成像观察积碳程度。这种检测方法便捷易操作,但在程度识别时存在以下问题:无法做到自动识别,大大降低检测效率;主观意识影响较大,导致结果存在随机性;采集到的积碳图像特征复杂多变,给程度识别带来很大困难。针对上述问题,肖剑平等^[6]采集不同部位的积碳图像提取图像特征,之后与预设特征通过预辨识模型进行相似对比,从而得到部位污损程度,预设辨识模型为基于大量的样本图片经过机器学习训练得到的神经网络模型。采用基于机器学习的方法对积碳图像进行判别可以实现积碳程度的自动识别,但目前的检测精度较低,不能达到高效精准识别积碳程度的需求。

由于基于传统机器学习的模型参数较多且层数受限不能解决较为复杂的问题,而基于深度学习的方法通过更深层的网络结构可以学习到隐藏在图像深层的特征,增加分类特点,提高图像分类准度。且目前基于深度学习的网络模型广泛应用于工程、医疗、农业、能源等领域,如路面病害检测任务^[7]、结肠癌病变识别^[8]、跨作物叶片疾病检测^[9]、光伏板积灰程度识别^[10]等。

本文以发动机燃烧室内部缸壁部位的图像为研究对象,基于改进的深度神经网络架构实现积碳程度的精准识别,采用ResNet18^[11]作为骨干网络,在其基础上添加特征重采样模块提高分类精度,模块分别从空间和通道出发提取多尺度和多频率特征,并在模型中添加轻量化卷积降低计算开销,提高模型推理速度,为后续发动机积碳的清理工作提供建议。

1 模型方法

1.1 空间特征提取模块

基于程度识别的分类任务与一般的语义级图像分类任务不同,在整个图像区域没有具象的物体或类别特点供网络学习,从肉眼观察到的多是一些特征像素位置及像素值的差异,这与细粒度图像分类任务类间差异小、类内差异大的特点较为相似,但程度分类图像的特征分布范围较广,标准卷积在特征提取时无法从比自身卷积核更大的特征视野中提取图像特征,这就导致图像特征空间的紧密性降低,因此,从特征提取的角度出发,采用感受野更大的膨胀卷积提取图像特征。然而,基于膨胀卷积的模型规模大且效率低,导致模型预测低效,基于此引入了极高效的空间金字塔(Extremely Efficient Spatial Pyramid, EESP)模块^[12],如图1所示。

EESP模块采用分组卷积将输入特征从高维空间映射到低维空间,并采用深度可分离膨胀卷积(Depthwise Separable Dilated Convolution, DSDC)代替昂贵的膨胀卷积,提高膨胀卷积的计算效率。其中每个DSDC的膨胀系数为 r ,卷积核大小为 $k \times k$,其对应的感受野大小为 $[(k-1)r+1]^2$ 。采用DSDC后的感受野为标准卷积($k=3, r=1$)的 $[(k-1)r+1]^2/9$ 倍。为了弥补膨胀卷积存在栅格效应的弊端,采用层次特征融合(Hierarchical Feature Fusion, HFF)的方式填补特征像素间的空洞,HFF的输出函数 X_n 定义为

$$X_n(x) = N(D_n(x), R_n(x)), \quad (1)$$

式中: x 为DSDC前组卷积的输出; n 为并行DSDC的序数; N 为聚合点; $D_n(x)$ 为第 n 个DSDC的输出; $R_n(x)$ 的定义如下

$$R_n(x) = \begin{cases} 0, & n=1, \\ D_{n-1}(x) + R_{n-1}(x), & \text{otherwise.} \end{cases} \quad (2)$$

将HFF的输出特征在通道维度进行拼接恢复特征维数空间,采用组卷积融合不同分辨率的特

征图降低计算开销。EESP 模块还加入了如残差网络的捷径连接改善梯度流。

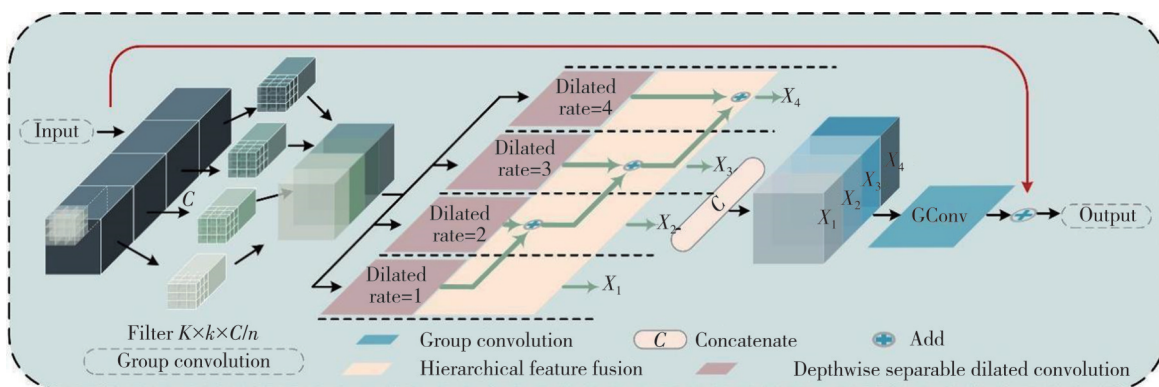


图 1 EESP 模块

Fig. 1 EESP module

1.2 通道特征提取模块

EESP 模块在逐层卷积的过程中会产生特征结构损失, 因此引入了多频谱通道注意力(Multi-Spectral Channel Attention, MSCA)模块^[13], 以此获得更好的通道表达。将通道表示视为压缩问题, SENet^[14]采用全局平均池化(Global Average Pooling, GAP)压缩通道, 卷积块注意模块(Convolutional Block Attention Module, CBAM)^[15]和基于风格的再校准模块(Style-based Recalibration Module, SRM)^[16]在 SENet 的基础上增加了全局最大池化和全局标准差池化, MSCA 模块则从频域的角度考虑通道的压缩问题, 证明了 SENet 只从 0 频获取了特征信息, 并将 GAP 扩展到多个频率。

图 2 所示为 MSCA 模块为特征通道加权的过

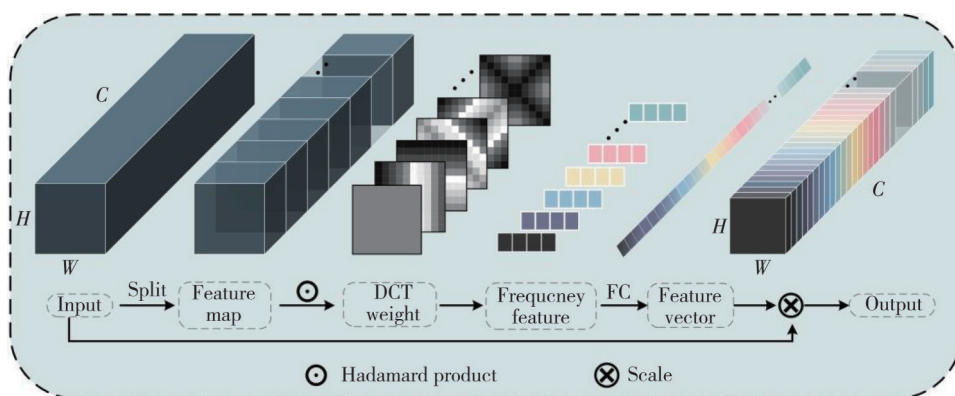


图 2 MSCA 模块

Fig. 2 MSCA module

频率特征通过全连接(Fully Connected, FC)拼接为 1 维向量, 即

$$F = FC([F^0, F^1, \dots, F^{m-1}]), \quad (5)$$

则整个 MSCA 模块可以表示为

$$MSCA = sigmoid(scale(F)), \quad (6)$$

程, 首先通过对特征映射做离散余弦变换(Discrete Cosine Transform, DCT)获取不同频率下的权值, 然后选取 m 个性能最高的频率分量, 将输入特征划分为 m 等份, 每个特征量的通道数为 C/m , 为每个特征量分配对应的 DCT 权重做哈达玛积(Hadamard Product)得到频率特征, 每个频率特征可以表示为

$$F^i = 2DDCT^{u_i, v_i}(Y^i) = \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} Y_{:,h,w}^i B_{u_i, v_i}^{h,w}, \quad (3)$$

式中: Y^i 为特征分量, $i \in \{0, 1, \dots, m-1\}$; u_i, v_i 为 Y^i 对应二维频率分量的索引; H, W 为输入特征的高和宽, $h \in \{0, 1, \dots, H-1\}$, $w \in \{0, 1, \dots, W-1\}$; 二维 DCT 的函数 $B_{u_i, v_i}^{h,w}$ 为

$$B_{u_i, v_i}^{h,w} = \cos\left(\frac{\pi u_i}{H} \left(h + \frac{1}{2}\right)\right) \cos\left(\frac{\pi v_i}{W} \left(w + \frac{1}{2}\right)\right). \quad (4)$$

式中: $scale$ 将 1 维向量变换到输入特征维度, $sigmoid$ 为非线性激活函数。

1.3 特征重采样模块

为了更好地捕获特征信息提高识别精度, 本

文从空间和通道两个维度获取细粒度特征,设计了多尺度 & 频率特征提取模块(Multi-scale & Frequency Feature Extraction Module, MFM)完成特征的重采样,如图3所示。输入特征分别经过

EESP模块和MSCA模块,EESP从不同尺度获取图像特征,既能分辨细节特征,也保证了宏观维度上的差异性,MSCA则从多个频率获取通道特征的全局平均池化,提高特征传递的丰富度。

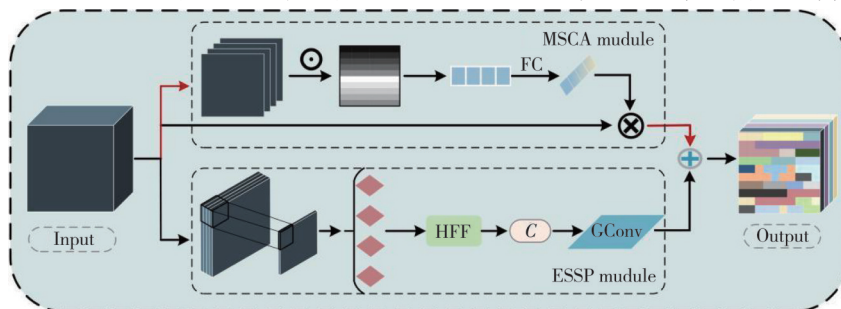


图3 MFM模块

Fig. 3 MFM module

特征重采样模块的结构设计受到残差结构通过恒等映射放大细小差异的启发,将MSCA模块作为EESP模块中的捷径连接融合特征,其目的是突出重要特征,减少冗余特征,使网络在经过MFM模块时能最大化获取特征差异。模块的设计允许在卷积神经网络的不同空间级别上进行特征重采样,在不改变特征图大小的情况下丰富内核,同时保证了网络的轻量化。

如图4所示为网络的整体设计结构。为了提高模型的推理速度并保障边缘设备的处理效率,在结构中添加了深度可分离卷积和分组卷积两种轻量化卷积(DConv、GConv)及Dropout减少冗余特征,提高全连接层的分类速度。

结构设计的必要性在2.3.2节的结构对比实验中进行验证。

1.4 模型结构

2 实验

积碳图像在采集过程中受环境因素的干扰容易产生噪声,因此,在骨干网络的选择上采用浅层网络,在保证网络学习到有效的特征元素的同时避免过拟合的产生。骨干网络结构采用Resnet18,利用标准的3×3卷积提取图像的主体特征,MFM模块嵌入不同尺度特征的卷积块后加强特征重用。

2.1 数据集介绍及预处理

数据采用内窥镜设备进行采集,具体为拆下喷油嘴后,将内窥镜管线伸入喷油孔,进入燃烧室从不同角度全方位拍摄。如图5所示为积碳图像部分样本图片,以缸壁面上的碳附着情况将积碳污染程度划分为轻度、中度和重度3种。

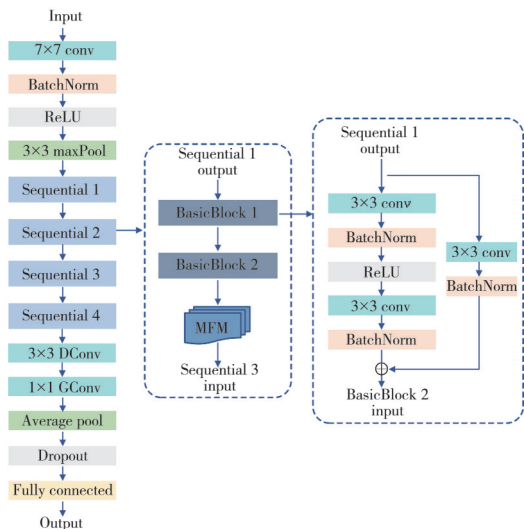


图4 网络整体设计结构

Fig. 4 Overall network design structure

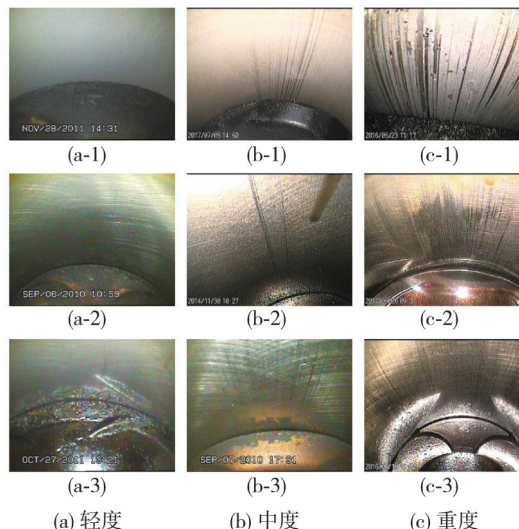


图5 样本示例图

Fig. 5 Sample examples

表 1 为数据分布情况,首先对数据量随机抽样做均衡化处理,其次对抽样数据按 8:2 的比例划分训练集和测试集。由于本文实验数据量较少,因此对原始图像做数据扩充。数据扩充采用旋转、裁剪等多种数据处理相结合的方式,从而提高与原始数据的差异性,有效防止过拟合

的产生。本文只针对训练数据进行扩充,对测试数据不做任何处理,保证数据与实际应用中采集到的图片特征相匹配。将均衡抽样后剩余的数据补充为测试集数据,通过更多的测试数据测试模型泛化性能,使本文实验结果更具实际应用价值。

表 1 数据集分布

Tab. 1 Distribution of datasets

标签	类别	原始数据		均衡抽样	训练集	测试集	训练集扩充后	测试集补充后
		数目	占比/%					
0	轻度	1 450	42	550	440	110	8 800	900
1	中度	1 450	42	550	440	110	8 800	900
2	重度	550	16	550	440	110	8 800	110
	总计	3 450	100	1 650	1 320	330	26 400	1 910

2.2 实验参数设置及评价指标

实验的批处理大小为 32,训练轮次为 100,优化器采用 SGD(Stochastic Gradient Descent)优化器,学习率调整使用余弦衰减策略,初始学习率为 0.001,损失函数为交叉熵损失函数。图像处理单元(Graphics Processing Unit, GPU)规格为 NVIDIA GeForce GTX 1080 Ti*3,采用并行训练的方式完成实验。

本文使用 Accuracy、Precision、Recall、Specificity、F1-score、FLOPs、Parameters、Inference Speed 等作为实验的评价指标,Accuracy、Precision、Recall、Specificity、F1-score 为常用的图像分类评价指标,FLOPs 为模型所需的计算量大小,Parameters 为模型的参数量大小,Inference Speed 为衡量模型预测速度的指标,单位为帧/s,表示每秒可预测的图片数量。

2.3 实验结果分析及对比实验

2.3.1 模型对比实验

本文选取了若干个网络模型进行对比实验,包括在卷积神经网络中应用较为广泛的模型,如 VGGNet16^[17]、ResNet50^[11]、DenseNet169^[18]、EfficientNet-B4^[19],及应用在视觉领域的 Transformer 模型 ViT^[20]和基于特征融合的 FFVT 模型^[21]。并对 EESP 模块所属的 ESPNetV2 模型作对比。图 6 从 FLOPs、Parameters、Inference Speed 结合准确率以及 Specificity、F1-score 展示了不同模型的性能表现。本文模型在图中用红色圆圈标出。

从图 6(a)和图 6(b)中可以看出基于 Transformer 的模型结构识别准确率相对较高,FFVT 和 ViT 因其不同于传统卷积神经网络的架构,对

数据噪声表现出较强的鲁棒性,所以在积碳程度判别中有较好的识别能力。其中,FFVT 针对细粒度图像聚合的来自低级、中级和高级 token 的局部信息进行分类,在本数据集上表现优于 ViT,证明这种特征融合的思想适用于积碳程度识别任务。VGGNet16 的识别准确率也相对较高,其采用 3×3 的小卷积核能较好地提取局部特征,对程度分类的帮助较大。但上述 3 种模型的计算量和参数量都很大,模型训练效率低,在实际应用中不易部署。从 ResNet50、DenseNet169 的结果看到更深层的网络并没有得到更好的结果,相反更深的网络由于图像噪声因素的干扰易造成网络过拟合,识别效果反而不佳。因此,深层次的网络结构并不适用于积碳程度的识别。EfficientNet-B4 展示了其较为高效的模型架构,但由于移动翻转瓶颈卷积块中采用 1×1 的点卷积降低了局部特征间的关联性,导致模型识别准确率略显不足。轻量化模型 ESPNetV2 的计算量和参数量虽然大大降低,但测试精度偏低不能满足实际应用需求,同时,ESPNetV2 的实验结果证明一味追求大的感受野并不会显著提升本模型性能。

图 6(c)为不同模型的预测推理速度,从图中看到模型的推理速度虽然会受到模型复杂度的影响,但与其并不成比例。本文设计的模型在推理速度中取得了最好的成绩,并在测试中达到了 84.5% 的准确率,基本达到 Transformer 模型的测试精度,且推理速度快,能更好地部署在边缘设备中。

图 6(d)从 Specificity、F1-score 两个指标评价模型性能,Precision 和 Recall 的分数由 F1-score 综合反映,没有单独列出。本文提出的方法与 VGGNet16 和 ResNet50 特异度分数处在第一梯

队,表示多次堆叠的卷积结构在本文的任务中具有较好的类间分离特性。 $F1$ -score 指标仅略低于

Transformer模型,在剩余的卷积神经网络中处于最高位,与Accuracy表现一致。

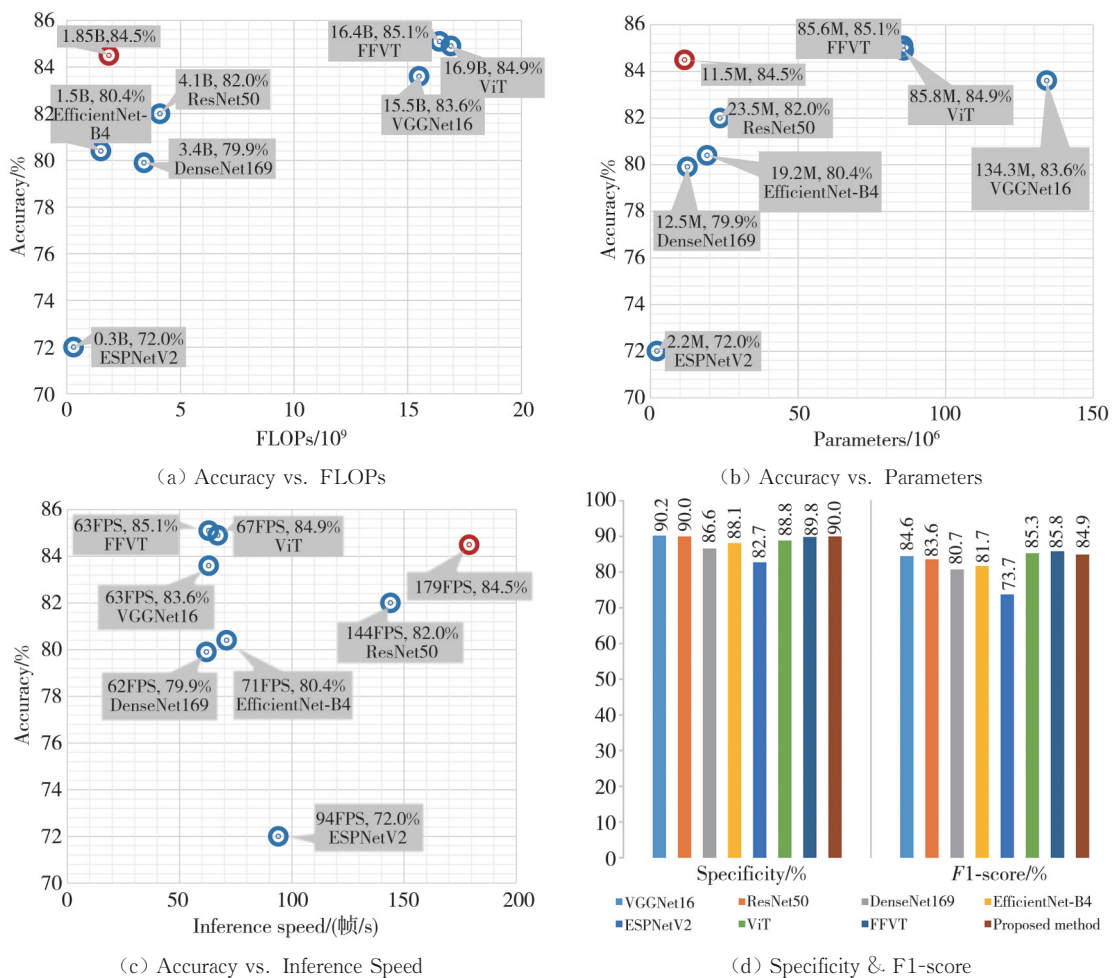


图6 不同模型的性能对比

Fig. 6 Performance comparison of different models

由于汽车发动机积碳程度识别任务的特殊性,在实际应用中需要将模型更好地部署在边缘计算设备中,即在保证实时性的同时不会降低太多的精度。综合上述对比实验的结果,本文设计的网络与该任务适配度最高。

2.3.2 结构对比

本文从骨干网络的选择、特征重采样模块的添加位置、轻量化卷积添加的必要性出发设计了3种

网络结构: A模型采用ResNet34作为骨干网络,ResNet34相较ResNet18每个卷积序列的卷积块数量均增加,但卷积块的结构不变; B模型将特征重采样模块增加在卷积序列中的每个卷积块后,而图4结构是在每个卷积序列后即2个卷积块后添加; C模型将全连接层前的两层轻量化卷积去除,其余结构不变。上述模型均在图4结构的基础上进行修改,保证单一变量原则,实验结果如表2所示。

表2 结构对比的实验结果

Tab. 2 Experimental results of structural comparison

结构	Accuracy/%	Precision/%	Recall/%	Specificity/%	F1-score/%	FLOPs/($\times 10^9$)	Parameters/($\times 10^6$)	Inference speed
A	83.0	85.1	83.0	89.7	84.0%	3.71	21.6	133
B	80.4	83.7	80.4	89.0	82.0%	1.88	11.7	121
C	82.9	84.6	83.0	89.0	83.8%	1.85	11.3	152
本文方法	84.4	85.3	84.5	90.0	84.9%	1.85	11.5	179

表2中A模型的准确率为83.0%,根据图6骨干网络为ResNet18的模型准确率为84.5%,证

明更深层次的深度模型在小样本的积碳程度判别任务中并不适用,并且结构复杂的模型不利于在

边缘设备中部署。B模型的准确率为80.4%，相较图6中基于ResNet18的模型性能大大降低。B模型的复杂结构获取了更加丰富的特征信息，但过于复杂的模型会导致模型产生过拟合从而降低模型分类性能。C模型的准确率和推理速度分别为82.9%和152帧/s，相较本文未去除轻量化卷积的方法均有所降低，证明轻量化卷积在降低模型冗余参数和提高推理速度中有着重要作用。

2.4 消融实验

为了验证单一模块对本文结构性能的影响，替换图4结构中的MFM模块进行实验对比。实验对比除了单独添加MSCA和EESP模块外，还设置了一组对比实验，即添加带步长的EESP模块(SEESP)。如图7所示为SEESP模块的结构图，与MFM模块结构类似，但SEESP是通过2个堆叠的卷积层(左侧旁支)弥补空间结构信息，与本文设计中的MSCA作用相同。表3给出了不同

模块添加在Conv2_x、Conv3_x、Conv4_x、Conv5_x层后的实验结果。

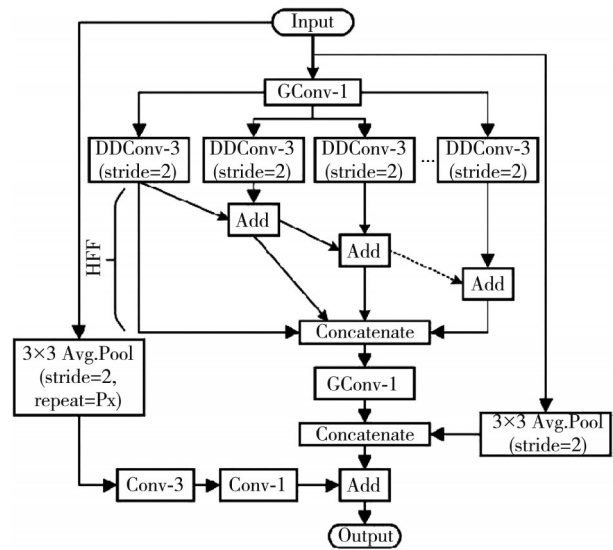


图7 SEESP模块

Fig. 7 SEESP module

表3 不同特征采样模块对模型训练指标的影响

Tab. 3 Effect of different feature sampling modules on model training metrics

结构	Accuracy/%	Precision/%	Recall/%	Specificity/%	F1-score/%
model-w/-EESP	83.6	84.1	83.5	88.4	83.8
model-w/-MSCA	83.5	84.6	83.5	88.8	84.0
model-w/-SEESP	80.3	81.1	80.3	85.8	80.7
本文方法	84.5	85.3	84.5	90.0	84.9

由表3实验结果可知，单独添加空间特征提取模块的模型准确率为83.6%，单独添加通道注意力模块的模型准确率为83.5%，相较于原始模型分类性能均有提高。

空间特征提取模块通过膨胀卷积提取多尺度的空间特征，可有效应对积碳图像特征分布范围广的问题。通道注意力模块则获取了不同频率下的通道信息特征，增强特征图的空间结构特征，针对积碳图像的细粒度特点有效提高了类别可分离性。SEESP模块虽然与本文结构类似，但在积碳程度识别任务中表现不佳，其分类准确率仅为80.3%，这主要是由于带步长的卷积在特征提取时造成了信息遗漏，并且单一的卷积提取的空间结构信息不够丰富。

3 结论

本文针对积碳图像程度自动识别任务精度低的问题提出了一个基于深度神经网络的分类模型。模型基于积碳图像特征分布范围广及细粒度

的特点和轻量易部署的需求进行设计。首先通过并行不同扩张率的DSDC增大特征提取的感受野，捕获多尺度的特征信息，在扩大关联特征范围的同时加强对细节信息的把握，DSDC保证了模型的轻量化。其次采用MSCA模块从多个频率对通道特征图进行加权，以此获得更丰富的空间结构信息。最后在全连接层前添加轻量化卷积提取重要特征，减少冗余特征提高模型的推理速度。实验结果表明本文结构在保证识别精度的同时拥有更小的模型体量，能较好地与实际工程中开展应用。

参考文献:

[1] ROE A P, THOMPSON S A, PETROU A T, et al. System for determining piston damage based on carbon deposit growth; US20180030890 [P]. 2018-02-01.

[2] 穆大芸, 吕宪勇, 宋国梁, 等. 一种喷气阀积碳的检测方法、装置及电子设备; CN110005524B [P]. 2019-09-20.

- [3] 杨万里, 刘岩, 洪安新, 等. 一种发动机气缸盖、发动机、车辆及沉积物处理方法: CN113818970B[P]. 2022-12-23.
- [4] PRABHAKAR B, CONTI R. Non-contact process for engine deposit layer measurement; US11650043 [P]. 2023-05-16.
- [5] 晏娟, 邵传刚, 王利锋, 等. 一种燃料发动机内积碳厚度检测装置: CN209131594U[P]. 2019-07-19.
- [6] 肖剑平, 季明, 兰龙, 等. 发动机检测方法、系统、计算机设备和存储介质: CN113284131A[P]. 2021-08-20.
- [7] APEAGYEI A, ADEMOLAKE T E, ADOM-ASAMOA M. Evaluation of deep learning models for classification of asphalt pavement distresses [J]. *International Journal of Pavement Engineering*, 2023, 24(1): 2180641.
- [8] 梁美彦, 张倩楠, 任竹云, 等. 基于注意力机制的结肠癌病理学图像识别研究 [J]. *测试技术学报*, 2022, 36(2): 93-100.
LIANG Meiyuan, ZHANG Qiannan, REN Zhuyun, et al. Research on identification of colon pathology image based on attention mechanism [J]. *Journal of Test and Measurement Technology*, 2022, 36(2): 93-100. (in Chinese)
- [9] SHARMA V, TRIPATHI A K, MITTAL H. DLMC-Net: deeper lightweight multi-class classification model for plant leaf disease detection [J]. *Ecological Informatics*, 2023, 75: 102025.
- [10] FAN S, WANG Y, CAO S, et al. A deep residual neural network identification method for uneven dust accumulation on photovoltaic(PV) panels [J]. *Energy*, 2022, 239: 122302.
- [11] 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.
- [12] MEHTA S, RASTEGARI M, SHAPIRO L, et al. Espnetv2: a light-weight, power efficient, and general purpose convolutional neural network [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 9182-9192.
- [13] QIN Z, ZHANG P, WU F, et al. FcaNet: frequency channel attention networks [C]// *Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual*, 2021: 763-772.
- [14] 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.
- [15] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]// *Proceedings of the European Conference on Computer Vision (ECCV)*. Cham: Springer, 2018: 3-19.
- [16] LEE H J, KIM H E, NAM H. SRM: a style-based recalibration module for convolutional neural networks [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual*, 2019: 1854-1862.
- [17] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. *Computer Science*, 2014. DOI: 10.48550/arXiv.1409.1556.
- [18] 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*, 2017: 2261-2269.
- [19] TAN M, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks [EB/OL]. <http://doi.org/10.48550/arxiv.1905.11946>.
- [20] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale [EB/OL]. <https://doi.org/10.48550/arxiv.2010.11929>.
- [21] WANG J, YU X, GAO Y. Feature fusion vision transformer for fine-grained visual categorization [EB/OL]. <http://arxiv.org/abs/2017.02341v3>.