基于深层卷积随机配置网络的电熔镁炉工况识别方法研究

李帷韬 [1] 童倩倩 [1] 王殿辉 [2, 3] 吴高昌 [3]

摘 要 为解决电熔镁炉工况识别模型泛化能力和可解释性弱的缺陷, 提出一种基于深层卷积随机配置网络 (Deep convolutional stochastic configuration networks, DCSCN) 的可解释性电熔镁炉异常工况识别方法. 首先, 基于监督学习机制生成具有物理含义的高斯差分卷积核, 采用增量式方法构建深层卷积神经网络 (Deep convolutional neural network, DCNN), 确保识别误差逐级收敛, 避免反向传播算法迭代寻优卷积核参数的过程. 定义通道特征图独立系数获取电熔镁炉特征类激活映射图的可视化结果, 定义可解释性可信度评测指标, 自适应调节深层卷积随机配置网络层级, 对不可信样本进行再认知以获取最优工况识别结果. 实验结果表明, 所提方法较其他方法具有更优的识别精度和可解释性.

关键词 电熔镁炉, 深层卷积随机配置网络, 高斯差分卷积核, 类激活映射图, 可解释性

引用格式 李帷韬, 童倩倩, 王殿辉, 吴高昌. 基于深层卷积随机配置网络的电熔镁炉工况识别方法研究. 自动化学报, 2024, **50**(3): 527−543

DOI 10.16383/j.aas.c230272

Research on Fused Magnesium Furnace Working Condition Recognition Method Based on Deep Convolutional Stochastic Configuration Networks

LI Wei-Tao[1] TONG Qian-Qian[1] WANG Dian-Hui[2, 3] WU Gao-Chang[3]

Abstract In order to solve the defects of generalization ability and weak interpretability of fused magnesium furnace working condition recognition model, an interpretable fused magnesium furnace abnormal working condition recognition method based on deep convolutional stochastic configuration networks (DCSCN) is proposed in this paper. Firstly, based on the supervised learning mechanism to generate Gaussian differential convolution kernel with physical meaning, an incremental method is used to construct a deep convolutional neural network (DCNN) to ensure that the recognition error converges step by step, and to avoid the process that back propagation algorithm iteratively finds the optimal convolutional kernel parameters. This paper defines channel feature map independent coefficients to obtain visualization results of fused magnesium furnace feature class activation mapping map, defines interpretable credibility measure to adaptively adjust deep convolutional stochastic configuration network layers, and recognizes untrustworthy samples to obtain optimal working condition recognition results. The experimental results show that the proposed method in this paper has better recognition accuracy and interpretability than other methods.

Key words Fused magnesium furnace, deep convolutional stochastic configuration networks (DCSCN), Gaussian differential convolution kernel, class activation mapping map, interpretability

Citation Li Wei-Tao, Tong Qian-Qian, Wang Dian-Hui, Wu Gao-Chang. Research on fused magnesium furnace working condition recognition method based on deep convolutional stochastic configuration networks. *Acta Automatica Sinica*, 2024, **50**(3): 527−543

氧化镁作为电熔镁砂 (又称电熔镁) 的主要成分, 是一种碱性耐火原材料, 广泛应用于航空航天、核子熔炉、电子电器等领域. 作为全球最大的电熔镁生产国和供应国, 我国菱镁矿石普遍存在品位低、成份波动大、矿物组成复杂等特性, 需要采用特有的三相交流电极电熔镁炉进行熔炼. 电熔镁炉的冶炼过程是边进料边冶炼, 由机器将原料倒入电熔镁

收稿日期 2023-05-10 录用日期 2023-09-26
Manuscript received May 10, 2023; accepted September 26, 2023
国家重点研发计划 (2018AAA0100304), 国家自然科学基金 (62173120, 62103092), 安徽省自然科学基金 (2108085UD11), 111 引智项目 (BP0719039) 资助
Supported by National Key Research and Development Program of China (2018AAA0100304), National Natural Science Foundation of China (62173120, 62103092), Anhui Provincial Natural Science Foundation (2108085UD11), and 111 Project (BP0719039)
本文责任编委 段书凯
Recommended by Associate Editor DUAN Shu-Kai
1. 合肥工业大学电气与自动化工程学院 合肥 230009 2. 中国矿业大学人工智能研究院 徐州 221116 3. 东北大学流程工业综合自动化国家重点实验室 沈阳 110819
1. School of Electrical Engineering and Automation, Hefei University of Technology, Hefei 230009 2. Institute of Artificial Intelligence, China University of Mining and Technology, Xuzhou 221116 3. State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819

炉, 通过炉内高温电弧对原料进行加热生成氧化镁晶体[1−2]. 由于氧化镁的熔点高达 2 850 ℃, 整个生产过程电能消耗极高, 单台电熔镁炉的日均耗电量在 40 000 kWh 左右, 占生产成本的 60% 以上, 属于典型的重大耗能设备.

为了保证电熔镁的品质, 需要对生产过程进行监控, 防止冶炼过程中可能出现的异常工况, 包括欠烧、过热和异常排气等. 欠烧工况是由于原料中含有杂质和复杂矿物导致电熔镁炉中原料燃烧不充分而产生的异常工况, 此时炉壁局部被烧红、发亮; 过热工况时炉口火焰较亮, 可能会产生镁烟尘和氧化镁等不良物质; 异常排气工况时炉口会有高温熔体喷射, 此时电流变化剧烈. 出现这些异常工况时, 需要及时发现并处理, 否则会导致能耗浪费, 镁资源利用率和镁砂品位降低, 镁炉烧穿, 原料泄漏威胁操作人员的安全. 目前镁炉异常工况的诊断很大程度上还需要依靠人工决策. 由于炉口存在燃烧的火焰无法直接观测内部熔池, 操作人员需要克服炉壁周围冷却管道形成的不确定位置/浓度高亮度水雾干扰, 围绕炉壁感知随着原料间断加入形成的熔池高度变速率增长多模态工况, 凭借经验反复推敲比对评估当前工况, 利用电流控制系统使三相电极电流跟踪熔炼电流设定值[3−4]. 然而, 每个巡检人员负责多台熔炉, 受制于人员多角度多方位反复观测的经验、责任心和劳动强度等主观因素以及复杂烧制环境中高温、噪音、灰尘、水雾和炉壁固有白斑等客观因素的影响, 知识无法解释和积累传承, 容易导致漏检或误检而造成电熔镁炉烧穿等不可逆损失, 难以满足实时巡检的运维需求.

近年来, 随着人工智能技术的不断发展, 借助机器学习、深度学习对电熔镁炉工况进行异常诊断受到广泛关注. 这些方法通过对镁炉生产数据进行采集和处理, 提取各种特征参数, 利用算法对异常工况进行自动诊断. 例如文献 [5−6] 提出了一种基于贝叶斯网络的镁炉异常工况诊断方法, 引入迁移学习解决异常工况数据量少的问题. 文献 [7] 采用半监督学习对无标签电流数据进行自动标注, 对在半监督学习框架下构造的分类器进行训练. 文献 [8] 将卷积神经网络 (Convolutional neural network, CNN) 与长短期记忆网络 (Long short-term memory, LSTM) 相结合, 分别提取电熔镁炉的空间特征和时序特征, 以识别欠烧工况. 文献 [9] 采用深度卷积生成对抗网络进行数据增强, 利用卷积神经网络提取 RGB 图像和红外热成像的特征. 文献 [10] 使用卷积网络、多层双向长短期记忆网络 (Bi-directional long short-term memory, Bi-LSTM) 和堆叠自动编码器分别提取电熔镁炉工况的图像、声音、电流特征, 将不同特征进行融合后训练工况分类器. 文献 [11] 使用 YOLO 目标检测算法检测关键目标区域, 基于 AlexNet 模型进行镁炉工况分类.

深层卷积神经网络 (Deep convolutional neural network, DCNN) 是一种人工神经网络, DCNN 通常由多个卷积层、激活函数、池化层和全连接层组成, 形成深度的网络结构. 深度结构有利于网络学习更抽象和高级的特征, 使用多个卷积层和池化层用于逐层提取图像特征, 提高了网络的表达能力. 具体而言, 卷积层使用卷积核来捕获图像的局部特征, 而池化层则通过降采样操作减少特征图的维度, 保留重要的信息. 虽然深度学习在图像领域已被广泛使用, 但是依然存在一些亟待解决的难题. 深层神经网络的黑盒特性本质上是由于其内部神经元学习到的特征与人类所理解的语义概念之间存在不一致性所导致的. 深度学习的可解释性是指模型的决策过程可以被清晰地理解和解释, 是人们理解模型程度和对决策信任程度的重要指标. 目前, 深度学习模型的可解释性具有不同层面的呈现方式, 包括特征可视化[12]、分析可视化[13]、局部和全局可解释性[12]、具有解释性的网络结构设计[14] 等. 其中, 具有解释性的网络结构可以使用户理解模型, 不仅能观察模型的预测结果, 还能了解模型产生决策的原因, 在模型出错时可以自行修复模型[15]. 传统的网络构建方法通常是基于人类的先验知识和经验进行迭代试错, 寻找最佳的超参数, 时间消耗巨大. 近年来, 研究者提出了一些自动的网络结构搜索方法. 文献 [16] 通过构建计算图共享子图间的参数进行训练, 寻找最优神经网络架构. 文献 [17] 使用 RNN 生成网络的描述并采用强化学习训练 RNN. 文献 [18] 使用组稀疏性正则器自动确定网络每层节点个数. 然而, 这些方法搜索空间巨大、计算效率不高且需要提前确定网络层级. 另外, 采用反向传播梯度下降方法训练神经网络存在权重初始化、局部最小值以及学习性能对学习率设置敏感等问题[19], 同样制约了深度学习模型的性能. 因此, 亟待研究一种具有可解释性网络结构的快速有效自动构建方法.

为了解决神经网络训练时间长、易陷入局部最小、与学习能力密切相关的隐含层节点个数难以确定等问题, 随机学习算法应运而生. 该算法随机分配输入权重和偏移并通过最小二乘法计算输出权重. 与随机向量函数连接网络 (Random vector functional link networks, RVFL)[20] 相比, 随机配置网络 (Stochastic configuration networks, SCNs)[21] 基于增量式随机算法, 采用不等式约束神经元随机参

数的分配并自适应地选择随机参数的范围, 基于监督机制确保所构建随机学习器的万局逼近能力. 文献 [22] 提出二维随机配置网络 (Two dimension stochastic configuration network, 2DSCN), 可以直接处理二维数据, 较 SCNs 在图像数据建模方面的泛化性能有所提高. 文献 [23] 提出一种深度随机配置网络 (Deep stochastic configuration network, DeepSCN), 网络中每一层的隐含节点都与输出相连, 使网络可以学习到更丰富的特征表征. 为解决非稳态数据流的持续学习问题, 文献 [24] 提出一种深度堆叠随机配置网络 (Deep stack stochastic configuration network, DSSCN), 网络结构可以自主加深和减浅. 然而, 虽然 SCNs 构建了可学习的神经元网络, 但是对于图像特征的提取能力较弱, 无法很好地表征感兴趣的图像信息.

电熔镁炉运行环境的特殊性导致了识别模型的泛化能力问题. 高亮度水雾、炉壁白斑、强光等干扰因素对图像质量造成了极大影响, 使得已经训练良好的模型在测试集上的表现出现较大下降, 产生了泛化能力弱的现象. 解决这一问题需要寻找适用于电熔镁炉环境的处理方法, 以确保模型能够在多样化的工况下保持准确性. 因此, 本文引入具有物理含义的高斯差分卷积核, 有助于区分由水雾、白斑、强光等因素导致的图像纹理信息缺失, 使模型能够较为准确地区分镁炉图像的水雾和白斑. 并且本文使用随机配置方法构建深层卷积网络, 其网络结构在增量学习过程中确定, 每增加一个卷积核/卷积层, 网络误差逐渐收敛, 使得网络结构更加紧致, 冗余卷积核更少, 更加利于模型的泛化性能. 同时, 传统模型架构固定 (卷积层数和卷积次数固定), 设计缺乏明确意义, 卷积核没有物理含义, 会出现卷积核冗余导致网络结构复杂等现象, 这限制了模型在实际应用中的可靠性和可信度. 而本文采用的深层卷积随机配置网络 (Deep convolutional stochastic configuration networks, DCSCN) 架构, 从单层网络单个卷积开始构建, 直至满足预设误差限, 构建过程中网络误差逐渐收敛且无需反复训练, 使得网络更加可信. 这种网络结构紧致且具有高度的可信度, 使得本文模型在实际应用中更具有可信度.

综上所述, 本文针对电熔镁炉运行环境的特殊性导致识别模型的泛化能力差问题以及可解释性弱的缺陷, 借鉴 SCNs 的增量学习方法, 提出一种基于深层卷积随机配置网络的可解释性电熔镁炉异常工况识别方法, 主要工作包括:

1) 为了避免传统增量式网络的不足, 本文首次采用随机配置方法构建深层卷积神经网络, 从单层网络单个卷积核开始递增, 直至满足停止迭代条件, 避免了反向传播算法迭代寻优卷积核参数的过程. 具有物理含义的高斯差分卷积核参数通过与数据相关的参数选择策略自动配置, 确保识别误差逐级收敛. 给出了深层卷积神经网络全局收敛能力的证明, 以保证网络的收敛性.

2) 为了避免高精度的超分辨算法需要较多训练样本及计算量大, 本文采用双线性插值方法将 DCSCN 构建的特征图集合上采样至输入图像大小, 与原始图像进行叠加后, 重新输入至当前随机配置卷积层级条件下的 DCSCN, 以获得多模态工况隶属于不同类别的得分. 定义通道特征图独立系数, 获取不同通道特征独立得分, 将类别得分、通道特征独立得分与对应通道特征图进行线性组合, 得到类激活映射图, 叠加至原始图像可获得当前层级条件下的电熔镁炉特征可视化结果. 定义可解释性可信度评测指标, 自适应调节 DCSCN 的层级, 以获取最优工况识别结果.

3) 对 9 000 幅电熔镁炉工况数据进行实验验证, 结果表明, 本文方法在测试集上的识别精度为 92.31%, 较其他方法具有更高的准确率和可解释性.

1　可解释性电熔镁炉异常工况识别模型

本文提出的基于深层卷积随机配置网络的可解释性电熔镁炉异常工况识别模型, 采用三层结构实现信息的耦合传递, 包括训练层、反馈层和测试层, 模型结构如图 1 所示.

训练层包括数据增强、预处理和深层卷积随机配置网络模块. 首先, 将电熔镁炉训练图像数据进行数据增强和预处理, 以扩展训练集样本数量. 然后, 将训练集送入深层卷积随机配置网络, 避免反向传播迭代更新卷积核参数, 基于监督机制自适应选取新的卷积核参数, 从单层单个卷积核开始, 逐个逐层构建增量式卷积特征提取层, 确保识别误差收敛. 从第一层增量式卷积特征提取层开始 (图 1 中增量式卷积特征提取层 1 表示), 将训练集送入第一层增量式卷积特征提取层, 获取特征图 (图 1 中第一个黄色框表示), 经由全连接层进行电熔镁炉工况分类, 若识别误差满足预设条件, 则停止训练; 否则继续增加单层卷积核个数 (图 1 中第二个黄色框表示), 直至单层最大卷积核个数 C_1, 若此时识别误差仍不满足预设条件, 在单层增量式卷积特征提取层基础上, 继续增加第二层增量式卷积特征提取层 (图 1 中增量式卷积特征提取层 2 表示). 重复上述过程直至设置的最大卷积层数 (图 1 中增量式卷积特征提取层 $L_{\max}$ 表示), 停止训练.

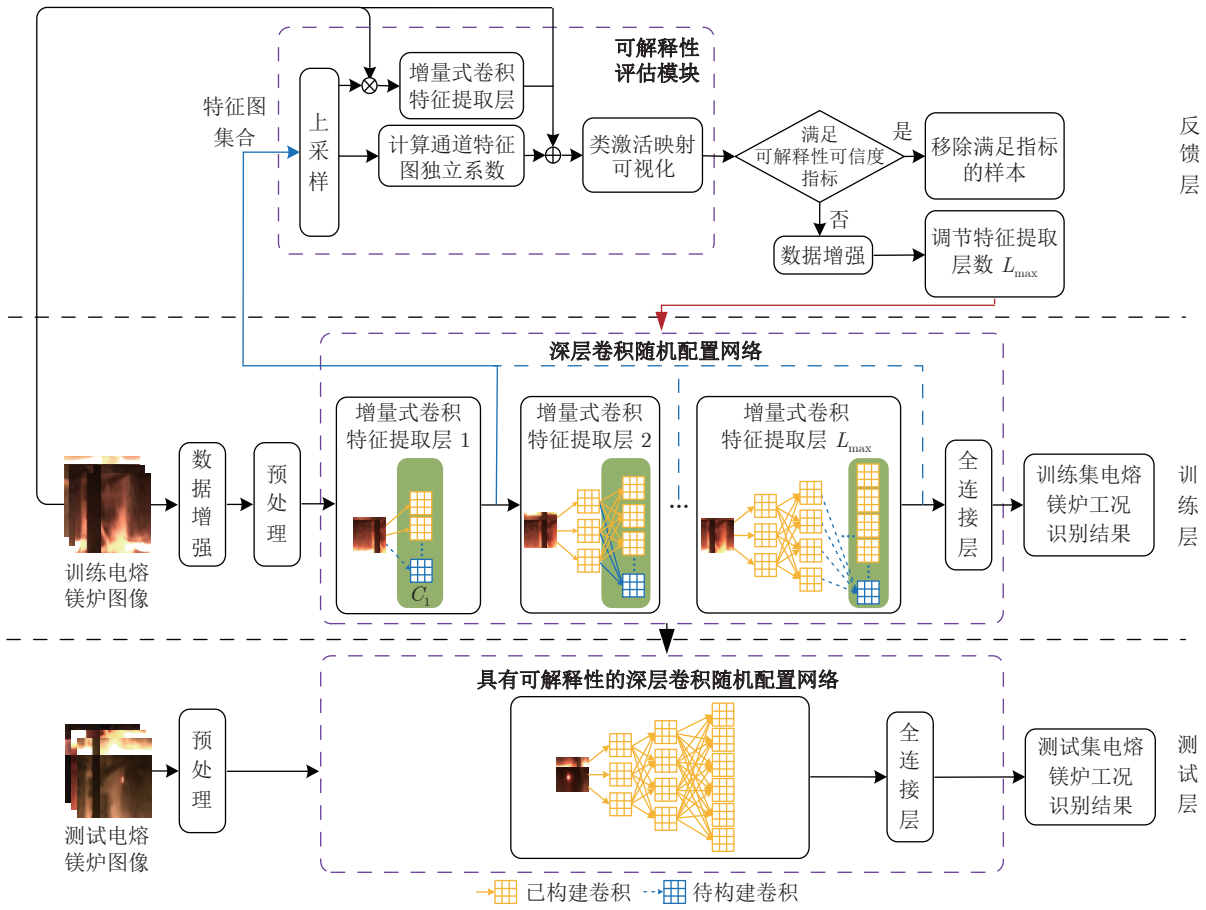


图 1 基于深层卷积随机配置网络的可解释电熔镁炉工况识别模型结构图

Fig. 1 Structure of interpretable fused magnesium furnace working condition recognition model based on deep convolutional stochastic configuration networks

反馈层包括可解释性评估模块和基于可解释性可信度指标的 DCSCN 卷积层调节模块. 电熔镁炉训练完成后, 若此时已构建 $L_{max}-1$ 层 DCSCN, 将深层卷积随机配置网络提取到的特征图集合输入至反馈层 (图 1 中蓝色箭头) 并上采样至原始输入图像大小, 与原始图像叠加后再次输入增量式卷积网络获取类别得分, 定义通道特征图独立系数, 计算每个通道特征独立得分, 最后将类别得分、通道特征独立得分与特征图线性组合, 获取特征图的类激活映射图. 定义可解释性可信度评测指标, 针对训练集电熔镁炉工况不确定识别结果, 判别是否满足可解释性可信度指标. 若满足可信度指标阈值, 则从训练集中移除, 否则将压缩后的训练集数据增强至原始训练集大小, 调节深层卷积随机配置网络的层数 (图 1 中红色箭头), 在固定前一层深层卷积随机配置网络参数条件下, 生成新的增量式卷积特征提取层, 使得 DCSCN 调整为 L_{max} 层, 对不满足可信度指标阈值的训练样本进行工况可信度再识别.

测试层使用构建的具有可解释性深层卷积随机配置网络, 获取测试样本的电熔镁炉工况最优识别结果.

2 基于深层卷积随机配置网络的可解释性电熔镁炉异常工况识别方法

2.1 数据增强和预处理

电熔镁炉生产过程中, 异常工况相对较少, 因此存在图像样本不平衡问题, 会导致训练的学习模型出现过拟合现象. 为了解决该问题, 本文采用非生成式方法对图像进行数据增强, 例如水平翻转、调整对比度和亮度、增加噪声等, 以提高数据的多样性, 减轻过拟合问题, 使得学习模型能够更好地适应各种场景下的镁炉图像变化. 其中, 水平翻转是一种常用的数据增强方法, 它通过将图像水平翻转来生成新的图像, 从而增加了数据的多样性. 这种操作可以模拟现实中不同视角的观察, 帮助模型更好地泛化到不同情况. 另一方面, 增加高斯噪声也是一种有效的数据增强方法. 在真实世界中, 图

像往往会受到各种干扰, 例如光照变化和传感器噪声. 通过在图像中添加高斯噪声, 能够让模型更好地适应这些现实干扰, 从而提高其在实际应用中的稳健性和准确性.

水平翻转可以描述为

$$I'(x, y) = I(w - x - 1, y) \tag{1}$$

其中, $I(x, y)$ 为原始图像在坐标 (x, y) 处的像素值, $I'(x, y)$ 为水平翻转后图像在 (x, y) 处的像素值, w 表示原始图像宽度.

调整对比度和亮度可以描述为

$$I'(x, y) = 1.5 \times (I(x, y) - 0.5) + 0.5 \tag{2}$$

为了避免出现像素值越界, 将原始图像中的每个数值减去 0.5, 乘以对比度增强系数 1.5, 最后再加上 0.5, 以同时增强图像的亮度和对比度.

添加高斯噪声可以描述为

$$I'(x, y) = I(x, y) + \mathrm{N}(0, \eta^2) \tag{3}$$

其中, $\mathrm{N}(0, \eta^2)$ 表示均值为 0、标准差为 η 的高斯噪声, 通过改变 η 的值可以控制噪声的强度.

采集的图像中可能包含一些与电熔镁炉无关的信息, 为了减少这些信息的影响, 将图像中心裁剪至 $1\,080 \times 1\,080$ 并缩放至 256×256, 将输入值归一化到 $[-1, 1]$ 范围内, 以方便后续的图像处理.

2.2　深层卷积随机配置策略

为了解决卷积神经网络在结构设计、超参数调整的问题, 借鉴随机配置网络[25−30], 如图 2 所示, 本文提出一种高效构建卷积神经网络的方法, 即深层卷积随机配置网络. 该策略从初始单层单个卷积核开始, 通过增量学习的方式产生新的随机卷积核来构建卷积神经网络, 克服了传统方法中手动设计网络结构和超参数调整的繁琐性. 在卷积核参数的选取过程中采用监督学习机制, 确保了深度学习模型的全局近似能力.

2.2.1　随机配置卷积核生成策略

随机配置卷积核生成策略自适应确定卷积核参数范围并在范围内随机产生新的卷积核参数, 包括权重和偏置. 深层神经网络从零卷积核开始构建, 使用监督学习机制产生具有可解释性的新卷积核, 确保网络性能的提升. 最后采用最小二乘法更新网络输出权重, 当网络误差小于预设误差时停止卷积核的生成.

令 $F = [f_1, f_2, \cdots, f_m] : \mathbf{R}^{d_1 \times d_2 \times d_3} \to \mathbf{R}^m$ 为一组实值函数, d_1, d_2, d_3 分别为空间上三个维度的大小, 其 L_2 范数定义为

$$\|F\| = \left(\sum_{q=1}^{m} \int_D |f_q(x)|^2 \, \mathrm{d}x \right)^{\frac{1}{2}} < \infty \tag{4}$$

实值函数 F 与实值函数 $G = [g_1, g_2, \cdots, g_m] : \mathbf{R}^{d_1 \times d_2 \times d_3} \to \mathbf{R}^m$ 的内积可表示为

$$\langle F, G \rangle = \sum_{q=1}^{m} \langle f_q, g_q \rangle = \sum_{q=1}^{m} \int_D f_q(x) g_q(x) \mathrm{d}x \tag{5}$$

设输入矩阵 I, 卷积核 K 的大小为 $\rho \times n$, 经过互相关操作后, 输出矩阵 S 中 (i, j) 位置的元素为

$$S(i, j) = (I * K)(i, j) = \sum_{\rho} \sum_{n} I(i + \rho, j + n) K(\rho, n) \tag{6}$$

其中, $*$ 为互相关操作符, $I(i + \rho, j + n)K(\rho, n)$ 表示 I 中 $(i + \rho, j + n)$ 位置的元素与卷积核 K 中 (ρ, n)

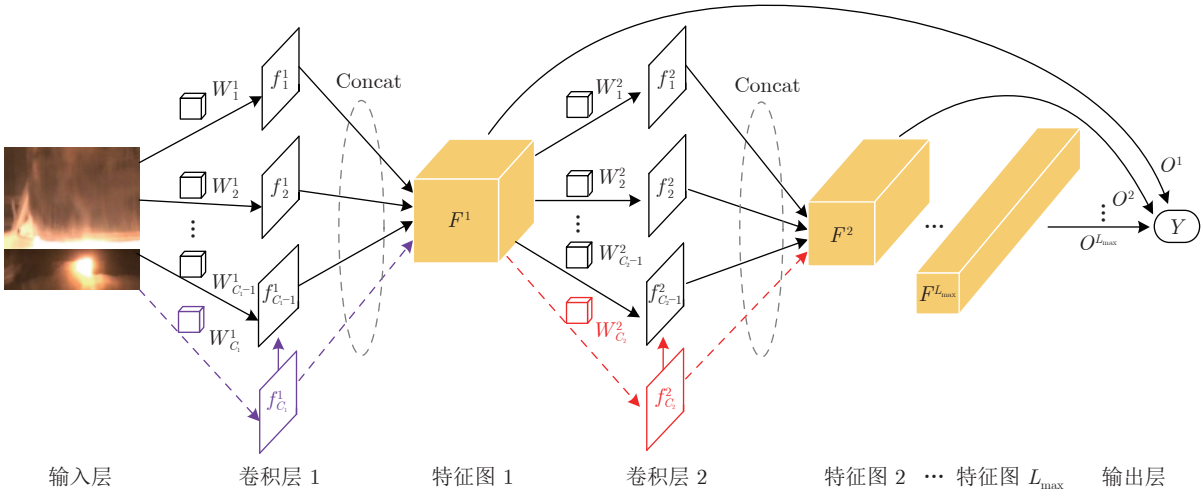


图 2　深层卷积随机配置网络结构图

Fig. 2　Deep convolutional stochastic configuration networks structure diagram

位置的元素相乘.

给定目标实值函数 $F:\mathbf{R}^{d_1\times d_2\times d_3}\to\mathbf{R}^m$, 假设一个 DCSCN 由 $L_{\max}$ 层卷积构成, 每层卷积的卷积核个数分别为 $C_1,\cdots,C_l,\cdots,C_{L_{\max}}$, $l\in[1,L_{\max}]$, 卷积核的大小为 $k\times k$, 则 DCSCN 可表示为

$$F^l(x)=\sum_{l=1}^{L_{\max}}\sum_{C=1}^{C_l}O_C^lA_C^l(W_C^l,b_C^l,I^l)\tag{7}$$

其中, $F^l(x)$ 为第 l 层的目标实值函数, W_C^l 和 b_C^l 为第 l 层卷积中第 C 个卷积核的参数, I^l 为第 l 层的输入, $W_C^l=\begin{bmatrix}W_{C,[1,1]}^l & \cdots & W_{C,[1,k]}^l\\ \vdots & \ddots & \vdots\\ W_{C,[k,1]}^l & \cdots & W_{C,[k,k]}^l\end{bmatrix}$ 为第 l 层卷积核的权重, A_C^l 为卷积函数, 包括卷积操作、激活函数和下采样操作, O_C^l 为输出权重, 误差 $e^l=F-F^l=[e_1^l,e_2^l,\cdots,e_m^l]$.

假设由 Γ 组成的函数空间 $span(\Gamma)$ 在 L_2 空间上稠密, $\forall A\in\Gamma$, $0<\|A\|<b$, 其中 $b\in\mathbf{R}^+$. 给定 $p>0$ 和非负收缩序列 $\{u_C\}$, $\{u_C\}$ 满足 $\lim_{C\to+\infty}u_C=0$ 且 $\sum_{C=1}^{\infty}u_C=\infty$, 随机生成第 l 层第 C 个卷积核的参数, 若满足

$$\left\langle e_{C,q}^l,A_C^l\right\rangle^2\geq pu_Cb^2\|e_{C-1,q}^l\|^2\tag{8}$$

其中, $e_{C,q}^l$, A_C^l为第 C个卷积核在第 q类的误差值, $q=1,2,\cdots,m$, 则 $\lim_{l\to+\infty}\|F-F^l\|=0$ 成立, 否则, 重新生成第 l 层第 C 个卷积核. 当第 l 层卷积中卷积核的个数增加至 C_l 时, 若误差 e^l 仍大于预设定误差值, 则新增加第 $l+1$ 层的第 1 个卷积核, 此时若满足

$$\left\langle e_{C_l,q}^l,A_1^{l+1}\right\rangle^2\geq pu_{C_l}b^2\|e_{C_l-1,q}^l\|^2\tag{9}$$

则 $\lim_{l\to+\infty}\|F-F^l\|=0$ 成立, 否则, 重新生成第 $l+1$ 层的第 1 个卷积核参数, 直至满足卷积核增加的终止条件.

因此, 一个 DCSCN 的构建问题可描述如下: 给定训练图像数据 $X=\{x_1,x_2,\cdots,x_N\}$, 其中, $x_i\in\mathbf{R}^{d_1\times d_2\times d_3}$, 对应的输出 $Y=\{y_1,y_2,\cdots,y_N\}$, 其中, $y_i\in\mathbf{R}^m$ 为图像类别标签. 记第 l 层卷积输入数据 I^l 的第 t 个通道为 I_t^l, $t=1,2,\cdots,C_{l-1}$, 则第 l 层卷积的第 C 个卷积核输出的特征图 M_C^l 可表示为

$$M_C^l=g\left(\sum_{t=1}^{C_{l-1}}W_C^l*I_t^l+b_C^l\right)\tag{10}$$

其中, $g(\cdot)$ 为激活函数, M_C^l 的维度为 $H\times W$. M_C^l 经过最大池化后可获得下采样特征图 A_C^l

$$A_C^l=\max_{m=1,\cdots,k-1}\max_{n=1,\cdots,k-1}M_{C,\Psi,\Omega}^l\tag{11}$$

其中, $\Psi\in[h,h+m]$, $h=1,2,\cdots,H$ 表示特征图的长度范围, $\Omega\in[w,w+n]$, $w=1,2,\cdots,W$ 表示特征图的宽度范围.

基于式 (11) 可以获取第 l 层卷积的特征图集合 $A^l=\{A_1^l,A_2^l,\cdots,A_C^l\}$, 将 A_C^l 与卷积网络的输出权重 O_C^l 相乘, 可以得到 l 层卷积的输出 F^l

$$F^l=\sum_{1}^{l}\sum_{C=1}^{C_l}O_C^lA_C^l\tag{12}$$

通过全局最小二乘法更新 O_C^l

$$O_C^l=[O_{C,1}^l,O_{C,2}^l,\cdots,O_{C,m}^l]=\arg\min_{O}\left\|Y-\sum_{l}\sum_{C=1}^{C_l}O_C^lA_C^l\right\|\tag{13}$$

第 l 层卷积中第 C 个卷积核的误差 e_C^l 为

$$e_C^l=[e_{C,1}^l,e_{C,2}^l,\cdots,e_{C,m}^l]\tag{14}$$

若 e_C^l 的 L_2 范数 $\|e_C^l\|$ 大于预设定误差值, 则新生成参数为 W_{C+1}^l 和 b_{C+1}^l 的卷积核, 直到满足停止迭代条件.

2.2.2 卷积核参数选取策略

深度学习模型中卷积核的构建会直接影响学习模型与输入数据之间的关联性, 具有良好性能的深度卷积网络, 其卷积核应呈高斯分布[31]. 炉壁和炉口高亮区域与水雾高亮区域具有截然不同的纹理特性, 边界存在显著的亮度变化. 高斯差分卷积核通过对图像进行高斯平滑和差分操作, 可以强化图像中的纹理信息, 以区分由于水雾遮挡导致减弱或丢失纹理信息的图像区域, 因此这里被选取作为 DCSCN 的卷积核, 其定义如下

$$f(x,y)=\frac{1}{2\pi}\left(\mathrm{e}^{-\frac{x^2+y^2}{2\xi^2}}-\frac{1}{r}\mathrm{e}^{-\frac{x^2+y^2}{2r^2\xi^2}}\right)\tag{15}$$

其中, ξ 表示标准差, 控制卷积核的宽度范围; r 表示尺度倍数, 较小的 r 可以检测细微边缘和细节信息, 较大的 r 可以检测粗糙边缘和轮廓结构. 不同的 ξ 和 r 组合可以获取不同的镁炉工况信息.

不同于传统基于梯度下降算法迭代更新卷积核参数的方法, 本文采用随机选取策略以生成不同的卷积核. 卷积核的权重 W_C^l 服从如式 (15) 所示的高斯差分分布, 偏置 b_C^l 在如下式所示的均匀分布中选取

$$b_C^l\sim\mathrm{U}(0,1)\tag{16}$$

基于式 (15) 和式 (16) 生成 $T_{\max}$ 个候选高斯差

分卷积核, 采用下式评估每个候选高斯卷积核的收敛性得分 $\sigma_{C,q}^{l}$

$$\sigma_{C,q}^{l}=\frac{\left((e_{C,q}^{l})^{\mathrm{T}}A_{C,q}^{l}\right)^{2}}{(A_{C,q}^{l})^{\mathrm{T}}A_{C,q}^{l}}-(\xi+r)u_{C}(e_{C,q}^{l})^{\mathrm{T}}e_{C,q}^{l}\quad(17)$$

其中, $A_{C,q}^{l}\in\{A_{C,1}^{l},A_{C,2}^{l},\cdots,A_{C,m}^{l}\}$, 保留满足下式的候选高斯卷积核

$$\sum_{q=1}^{m}\sigma_{C,q}^{l}>0\quad(18)$$

选取收敛性得分 $\sum_{q=1}^{m}\sigma_{C,q}^{l}$ 最高的候选高斯卷积核作为第 l 层中第 C 个卷积核. 若 $T_{\max}$ 个候选高斯卷积核中无满足要求的卷积核, 则基于式 (15) 重新选取 ξ 和 r 以生成新的高斯差分卷积核, 直至找到第 l 层中第 C 个满足式 (18) 条件的卷积核.

2.2.3　深层卷积随机配置网络收敛性证明

假设 DCSCN 中第 l 层卷积的卷积核个数为 C_l, 则易得第 l 层卷积满足

$$\|e_{C_l}^{l}\|^{2}\leq(\xi+r)u_{C_l}\|e_{C_l-1}^{l}\|^{2}\quad(19)$$

第 $l+1$ 层卷积中生成的第 1 个卷积核的输出误差 e_{1}^{l+1} 为

$$\|e_{1}^{l+1}\|^{2}=\|e_{C_l}^{l}-O_{1}^{l+1}A_{1}^{l+1}\|^{2}\quad(20)$$

其中, $O_{1}^{l+1}=[O_{1,1}^{l},O_{1,2}^{l},\cdots,O_{1,m}^{l}]$. 记 $\overline{O}_{1}^{l+1}=[\overline{O}_{1,1}^{l},\overline{O}_{1,2}^{l},\cdots,\overline{O}_{1,m}^{l}]$ 为最优输出权重, 则 $\overline{O}_{1,q}^{l+1}$ 可表示为

$$\overline{O}_{1,q}^{l+1}=\frac{\langle e_{C_l,q}^{l},A_{1}^{l+1}\rangle}{\|A_{1}^{l+1}\|^{2}}\quad(21)$$

由式 (18) ~ (21) 可将 $\|e_{1}^{l+1}\|^{2}$ 变换为

$$\|e_{1}^{l+1}\|^{2}\leq\|e_{C_l}^{l}-\overline{O}_{1}^{l+1}A_{1}^{l+1}\|^{2}=$$

$$\|e_{C_l}^{l}\|^{2}-\sum_{q=1}^{m}\overline{O}_{1,q}^{l+1}\leq\|e_{C_l}^{l}\|^{2}\quad(22)$$

因此

$$\|e_{C_{l+1}}^{l+1}\|^{2}\leq\|e_{1}^{l+1}\|^{2}\leq|e_{C_l}^{l}\|^{2}\leq$$

$$(\xi+r)u_{1}^{(\sum_{l}C_l)-1}\|e_{1}^{1}\|^{2}\quad(23)$$

由此, 可以证明 DCSCN 的输出误差 e_{1}^{l+1} 单调递减, 网络具有收敛性.

2.3　可解释性评估

电熔镁炉异常工况识别需要确保识别结果的可靠性和可解释性, 以便于操作人员理解深度学习模型的可信决策过程. 深层卷积随机配置网络在确保全局收敛获取可靠识别结果的同时, 基于监督学习生成的高斯分布卷积核具有与输入数据强关联的物理含义, 确保了模型的机理可解释性. 类激活映射可视化作为一种视觉可解释性验证, 表征了数据的重要性分布, 其采用最后一层卷积的梯度信息作为通道特征图的权重值, 与特征图加权求和得到类别重要性, 最终叠加至原图得到模型的特征数据可解释性. 由于 DCSCN 的构建过程中不包含梯度信息, 因此, 本文提出了一种基于特征图独立性加权的类激活映射方法. 首先将特征图集合上采样至原始图像大小, 与原始图像叠加后, 再次送入增量式卷积网络获取类别得分; 定义通道特征独立系数, 计算每个通道特征图独立得分, 将类别得分、通道特征图独立得分与通道特征图进行线性组合, 可以得到特征类激活图, 再叠加至原图即可得到图 3 所示的特征可视化结果.

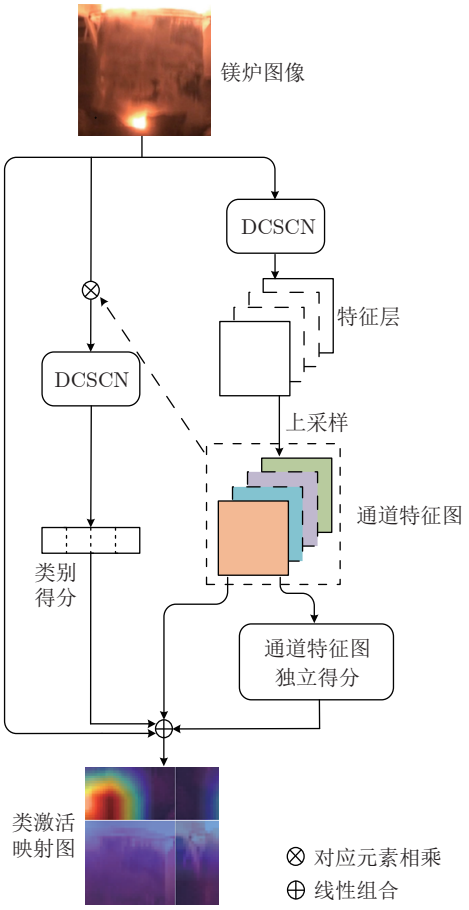


图 3　基于特征图独立性得分的类激活映射示意图

Fig. 3　Schematic diagram of the class activation mapping based on feature map independence scores

采用双线性插值将第 l 层卷积的特征图集合 A^{l} 的每个通道特征上采样至原始输入大小并归一化至 [0, 1]. A^{l} 的第 ϱ 个通道特征图 A_{ϱ}^{l} 与原始图像 x_i 叠加后可记为 $\bar{A}_{\varrho}^{l}$

$$\bar{A}_{\varrho}^{l} = A_{\varrho}^{l} \otimes x_i \tag{24}$$

其中, $\otimes$ 为点乘操作. 将 $\bar{A}_{\varrho}^{l}$ 输入 DCSCN 模型 (记为 Δ), 经由 Softmax 可输出类别得分 S_q^l

$$S_q^l = \text{Softmax}(\Delta(\bar{A}_{\varrho}^{l})) \tag{25}$$

DCSCN 模型针对 x_i 的类别预测结果为 y^q, $q = 1, \cdots, m$, 则第 ϱ 个通道特征图的类别得分为 $S_{\varrho,q}^{l}$. 通道特征独立得分的定义如下

$$FC_{\varrho}^{l} = \frac{\|A^l\|_* - \|\Xi_{\varrho}^{l} \odot A^l\|_*}{\|A^l\|_*} \tag{26}$$

其中, FC_{ϱ}^{l} 为每个通道特征图的独立得分, $A^l \in \mathbf{R}^{H\times W\times C_l}$, $A_{\varrho}^{l} \in \mathbf{R}^{H\times W}$ 表示第 l 层卷积的第 ϱ ($\varrho \in [1, C_l]$) 个特征图通道, $\|\cdot\|_*$ 表示核范数, $\odot$ 表示哈德马积, Ξ_{ϱ}^{l} 表示第 ϱ 行为零且其他行为 1 的卷积核掩码矩阵. 将掩码矩阵 Ξ_{ϱ}^{l} 与特征图 A^l 进行哈德马积运算, 表示第 ϱ 行被删除后的特征图. 特征图的秩可以表示特征图的线性独立程度, 并且核范数是对矩阵秩进行很好的凸近似的方法. 因此, $\|A^l\|_* - \|\Xi_{\varrho}^{l} \odot A^l\|_*$ 表示被删除的行对矩阵的线性程度的影响, 最后将核范数之差进行归一化, 从而得到通道特征独立系数.

将 A^l 中的全体特征图分别乘以 $S_{\varrho,q}^{l}$ 和 FC_{ϱ}^{l} 后再进行求和, 即将每个通道特征图赋予不同类别得分和不同通道独立得分, 即可得到 l 层 DCSCN 模型条件下样本 x_i 的类激活映射图

$$\mathcal{L}^q = \text{ReLU}\left(\sum_{\varrho} FC_{\varrho}^{l} S_{\varrho,q}^{l} A_{\varrho}^{l}\right) \tag{27}$$

基于 $\mathcal{L}^q$ 可定义可解释性可信度指标, 以量测预测目标与真实目标之间的偏差, 判断可解释性结果是否与真实结果一致. 可解释性可信度指标定义为当前可解释性结果高亮区域 d 与真实标注区域 $\bar{d}$ 之间的交并比 IoU

$$\text{IoU} = \frac{d \cap \bar{d}}{d \cup \bar{d}} \tag{28}$$

其中, $d \cap \bar{d}$ 表示 d 和 $\bar{d}$ 相交的面积大小, $d \cup \bar{d}$ 表示 d 和 $\bar{d}$ 相并的面积大小. 若 IoU 大于预设定阈值 $\mathcal{I}$, 则从训练样本集中移除该样本, 否则将移除压缩后的训练集数据增强至原始训练集大小, 增加 DCSCN 的特征提取层数, 固定前一层网络参数, 在新的特征提取层中从零开始生成新的卷积核, 对不满足可解释性可信度指标阈值的训练样本进行工况可信度再识别, 直至满足停止迭代要求.

3 实验分析

3.1 数据描述及处理

为了验证所提方法的有效性, 本文选取了源于辽宁省某电熔镁炉厂的生产视频. 对视频进行拆帧处理, 获取分辨率为 1 080 × 1 920 的图像, 经过图像数据增强后, 共获得 9 000 张图像样本. 图 4 ~ 7 分别为正常工况、欠烧工况、过热工况和异常排气工况数据增强后的部分结果. 通过增加对比度和亮度, 图像的明暗部分更加突出, 有利于模型提取细节特征. 此外, 还对图像进行了镜像处理, 模拟了在真实场景中可能出现的情况. 本文在图像上增加标准差为 0.3 的高斯噪声, 模拟真实情况下的噪声强度. 电熔镁炉图像的工况由专家进行类别标注. 随机选取 80% 的数据(共 7 200 张图像) 作为训练集, 其中 4 种工况样本各占四分之一, 剩余 20% 的数据 (共 1 800 张样本) 作为测试集. 所有实验均在同一平台上进行, 硬件包括 Intel i9-10900K 处理器、16 GB 内存以及 RTX 3060 显卡, 编程语言为 python 3.9.

图 4 正常工况图像数据增强后的结果

Fig. 4 Results of normal conditions image data enhancement

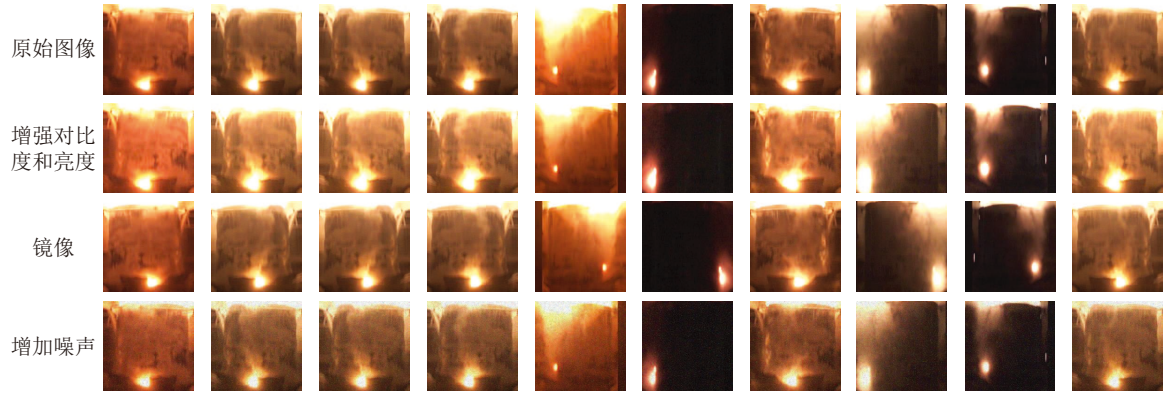


图 5　欠烧工况图像数据增强后的结果

Fig. 5　Results after image data enhancement for underburning conditions

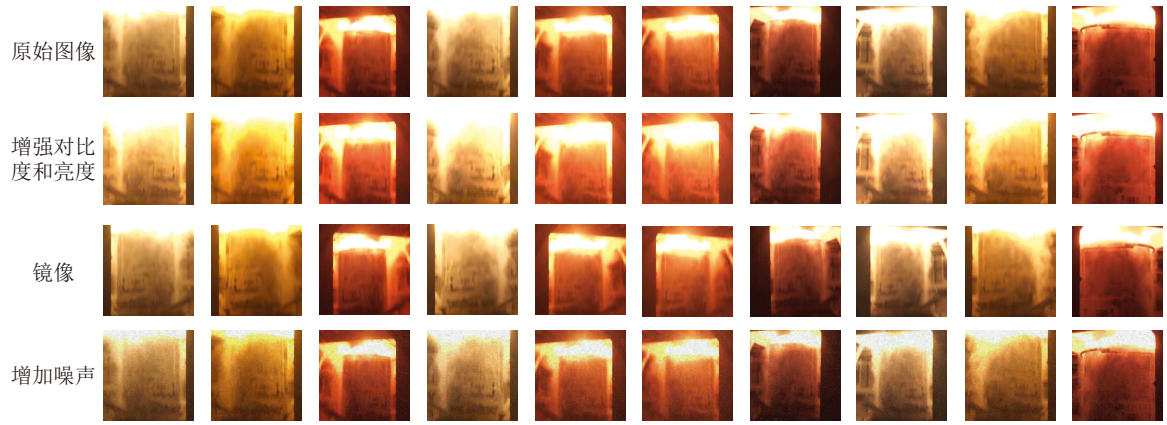


图 6　过热工况图像数据增强后的结果

Fig. 6　Results after image data enhancement for superheated operating conditions

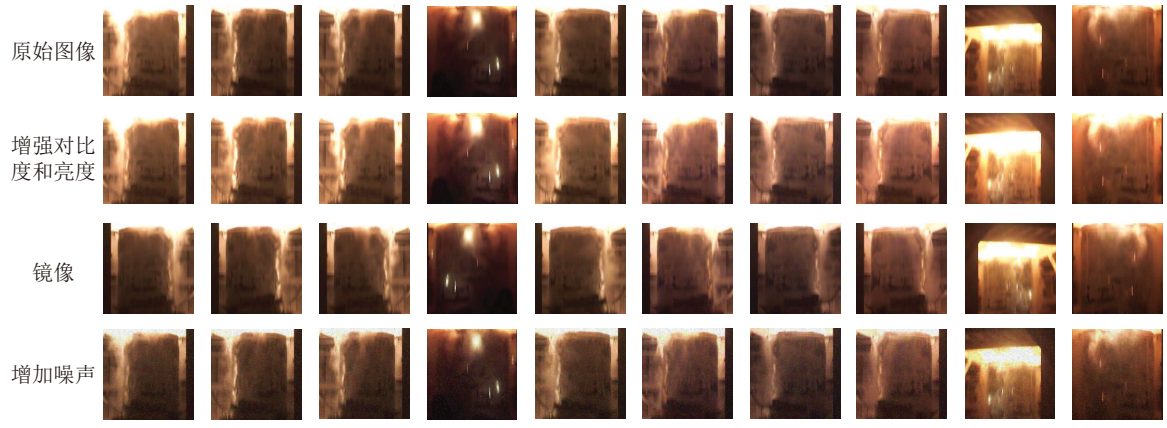


图 7　异常排气工况图像数据增强后的结果

Fig. 7　Results after image data enhancement for abnormal exhaust conditions

3.2　性能评价指标

漏诊率 R_m 定义为实际是正例但被预测为负例的样本数量与全体样本数量的比例, 可以衡量模型对正样本的识别能力, 计算方法如下

$$R_m = \frac{FN}{N} \tag{29}$$

其中, FN 为实际是正例但被预测为负例的样本数量, N 表示全体样本数量.

误诊率 R_f 定义为实际是负例但被预测为正例

的样本数量与全体样本数量的比例, 可以衡量模型对负样本的识别能力, 计算方法如下

$$R_f = \frac{FP}{N} \tag{30}$$

其中, FP 表示实际是负例但被预测为正例的样本数量.

精度 R_a 为模型正确分类样本数量与全体样本数量的比例, 计算方法如下

$$R_a = \frac{TP + TN}{N} \tag{31}$$

其中, TP 表示实际是正例且被正确预测为正例的样本数量, TN 表示实际是负例且被正确预测为负例的样本数量.

PA 可以衡量模型所占内存大小, 即卷积层参数量 CA 与全连接层参数量 FA 之和, 计算方法如下

$$PA = \sum_{l=1}^{L_{\max}} (CA_l + FA_l) = \sum_{l=1}^{L_{\max}} (k \times k \times C_{l-1} + 1) \times C_l + (m_{in} + 1) \times m \tag{32}$$

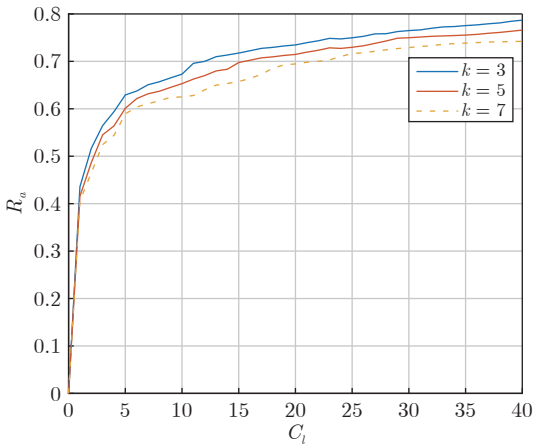
其中, k 表示高斯差分卷积核大小, C_{l-1} 为第 $l-1$ 层卷积核个数, C_l 为第 l 层卷积核个数, m_{in} 为全连接层输入神经元个数, m 为网络输出的维数.

3.3 实验结果与分析

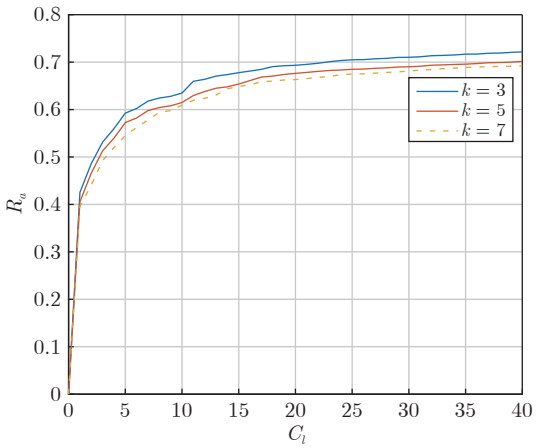
3.3.1 深层卷积随机配置网络实验结果

深层卷积随机配置网络的参数设置为期望误差限 $\bar{e} = 0.01$, 候选高斯差分卷积核个数 $T_{\max} = 100$, 标准差 $\xi \in [0.5, 5]$, 尺度倍数 $r \in [0.8, 1.5]$, 高斯差分卷积核大小 $k \in \{3, 5, 7\}$. 卷积网络中最大卷积层数 $L_{\max} = 10$, 每层卷积的最大卷积核个数 $C_l = 40$, 采用 sigmoid 激活函数, 池化层选取最大池化核, 池化核 $k_p = 2$, 非负收缩序列 $u_C = 1/C$, $C \in [1, C_{\max}]$. 所有实验结果均基于 50 次独立实验的平均.

图 8 给出了某次采样实验中, 当网络卷积层数 $l = 1$ 时, 不同大小卷积核 k 条件下, 电熔镁炉工况训练样本和测试样本识别精度曲线. 可以看出, 在训练和测试过程中, 随着具有物理含义的不同大小随机配置卷积核逐一生成, DCSCN 模型均呈现快速收敛性, 在 40 次卷积操作后, 模型精度趋于平缓. 由于单层卷积的性能有限, DCSCN 在给定的卷积核个数内未达到预设的误差水平, 因此需要多层卷积网络以提取抽象特征, 达到理想的分类性能. 此



(a) 训练集
(a) Training set



(b) 测试集
(b) Testing set

图 8 不同卷积核大小条件下的识别精度曲线

Fig. 8 Recognition accuracy curves under different convolutional kernel sizes

外, 不同大小的卷积核虽然可以获取相似的模型性能, 但同时也带来了更多的计算量. 两个 3×3 的卷积核与一个 5×5 的卷积核具有相同的感受野, 但两个 3×3 的卷积核可以进行两次非线性变换, 具有更好的非线性变换能力且计算量更小. 因此, 大小为 3 的卷积核可以更好地平衡模型性能与复杂度.

为了验证本文每个具有物理含义的卷积核生成的必要性, 采用强化学习方法对经由生成的随机配置卷积核提取的特征图集合进行选取, 与选取前的特征图集合进行性能对比. 构建 $l = 1$ 和 $l = 3$ 的深层卷积随机配置网络, 卷积核大小 k 为 3.

强化学习模型中, 为了降低深层卷积随机配置网络过拟合的风险, 平衡模型精度与较低相关性特征图提取之间的权重, 定义联合奖励函数 $R(s_{t,l}, s_{t+1,l})$ 如下

$$R(s_{t,l},\, s_{t+1,l}) = \alpha \times \left(1 - \sum_{l=1}^{L_{\max}} LFI_t^l\right) + \beta \times R_a \quad (33)$$

其中, α 和 β 分别表示特征图独立性和精度的权重系数, t 表示时刻, $s_{t,l}$ 和 $s_{t+1,l}$ 分别为 t 时刻和 $t+1$ 时刻智能体的状态, α 数值高表示强化学习模型侧重相互独立的特征图作为输出, β 数值高表示侧重提升精度的特征图作为输出. $\sum_{l=1}^{L_{\max}} LFI_t^l$ 表示每层卷积输出特征图集合 A^l 的独立性之和, 特征图独立性指标可以评测每个卷积层特征图集合的独立程度, 当某个特征图高度依赖于其他通道的特征图时, 意味着其包含的信息很大程度上已经被编码在其他特征图中, 其定义如下

$$LFI^l(\{A_\varrho^l\}_{\varrho=1}^{C_l}) = \sum_{\varrho=1}^{C_l}(\|A^l\|_* - \|\Xi_\varrho^l \odot A^l\|_*) \quad (34)$$

强化学习模型的参数设置为: 评估 Q 网络和目标 Q 网络均为包含 300 个神经元的两层全连接网络, 折扣因子 γ 为 0.9, 经验回放池大小为 2 000, 学习率为 0.002, 训练轮数为 400 个 epoch. 使用 ε-贪心搜索策略[32] 更新评估 Q 网络, 在前 100 轮实验中, ε 固定为 1.0, 100 轮后 ε 固定为 0.1.

图 9 为 $l=1$ 和 $l=3$ 时 DCSCN 中高斯差分卷积核提取的特征图集合被送入强化学习模块后的平均奖励曲线. 从图中可以看出, 在训练 300 轮后, 单层 DCSCN 和三层 DCSCN 的平均奖励值均稳定在 0.85 附近, 表明此时强化学习模块收敛, 选取了可以较好平衡模型精度与较低相关性的特征图.

训练样本集和测试样本集在是否采用强化学习方法条件下的漏诊率、误诊率和精度如表 1 所示. 可以看出, 本文方法在训练样本集和测试样本集的所有性能指标上, 均较采用强化学习方法选取特征图的策略有着更优的表现, 深层网络的性能优于浅层网络, 模型的泛化性能得到提升. 具体而言, 使用强化学习方法对特征图进行选取后, 训练样本集和测试样本集在单层 DCSCN 条件下和三层 DCSCN 条件下相较于本文方法, 精度分别下降 2.48%、2.67%、3.07%、2.49%, 漏诊率分别上升 1.47%、2.05%、0.56%、1.33%, 误诊率分别上升 0.99%、0.62%、2.51%、1.16%. 由此表明, 本文 DCSCN 所生成的不同标准差和不同尺度倍数的高斯差分卷积核具有特定的物理含义, 可以提取电熔镁炉图像截然不同细节和轮廓结构的特征图, 对工况的识别结果均可提供独特的贡献度, 因此, 特征图的选取导致 DCSCN 模型性能下降.

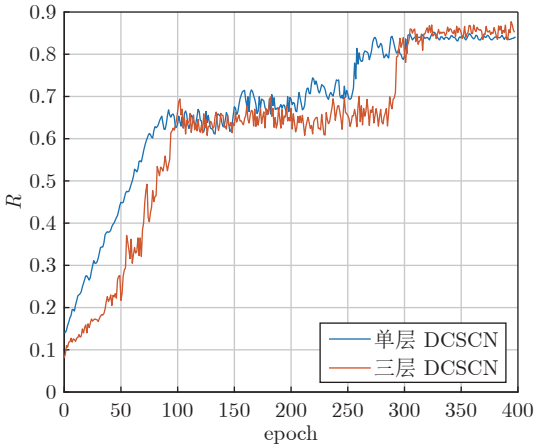


图 9　强化学习训练过程的平均奖励曲线

Fig. 9　Average reward curves for training process of reinforcement learning methods

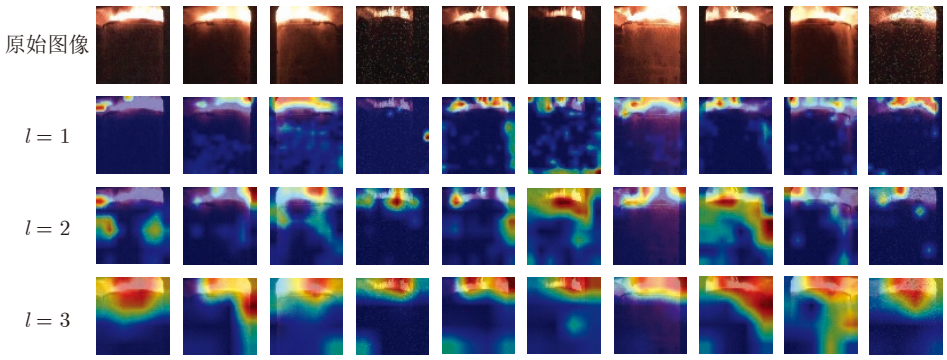
3.3.2　可解释性评估实验结果

深层卷积随机配置网络中不同卷积层的类激活映射图如图 10 所示, 图 10(a) ~ 10(d) 分别对应 4 种不同工况, 图中高亮的部分表示激活区域, 即该卷积层所关注的区域. 可以看出, $l=1$ 所对应的类激活映射图提供了较为详细的关键目标信息, 但是高亮区域较为分散且杂乱, 说明噪声污染对识别结果的影响比较严重. 随着卷积层数的增加, 噪声逐渐被抑制, 模型关注的关键目标更加连贯和准确. 在 $l=3$ 所对应的类激活映射图中, 较好地提取了关键特征, 定位了不同工况所需关注的目标区域, 即正常工况下的炉口火焰位置、欠烧工况下的炉壁烧红和炉口位置、过热工况下的炉口火焰位置、异常排气工况下的炉口溶液溢出位置.

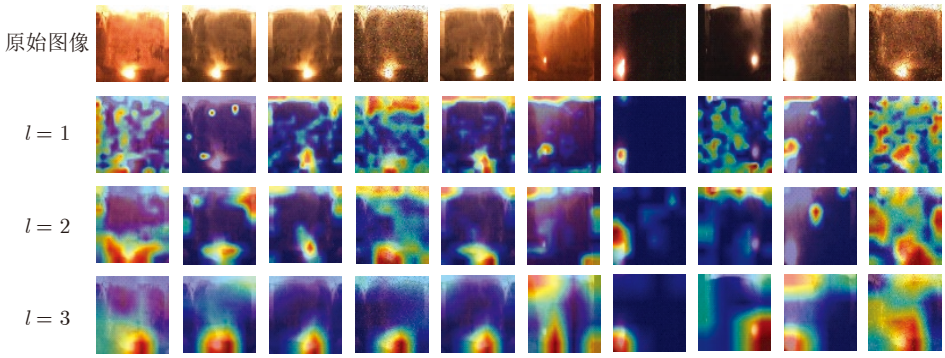
表 1　基于强化学习的漏诊率、误诊率和精度对比 (%)

Table 1　Comparison of missed diagnosis rate, misdiagnosis rate and accuracy based on reinforcement learning (%)

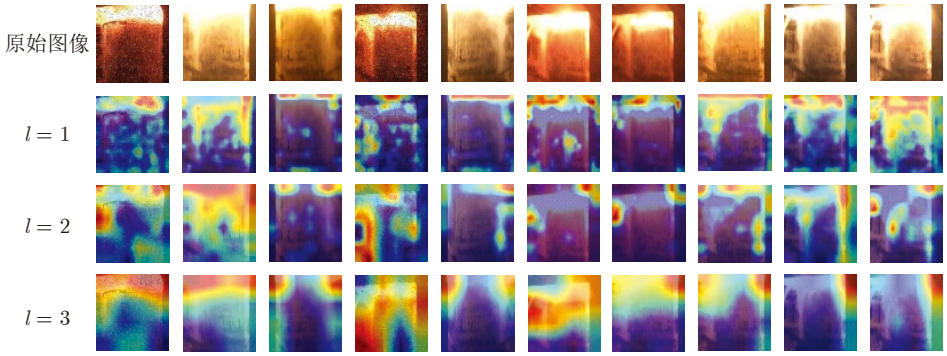
模型		训练集			测试集		
		漏诊率	误诊率	精度	漏诊率	误诊率	精度
单层	本文方法	7.61 ± 0.189	9.15 ± 0.331	83.24 ± 0.195	9.95 ± 0.216	10.30 ± 0.231	79.75 ± 0.108
	强化学习	9.08 ± 0.082	10.14 ± 0.354	80.76 ± 0.228	10.51 ± 0.172	12.81 ± 0.390	76.68 ± 0.305
三层	本文方法	5.31 ± 0.239	1.96 ± 0.165	92.73 ± 0.166	5.24 ± 0.245	2.45 ± 0.203	92.31 ± 0.283
	强化学习	7.36 ± 0.361	2.58 ± 0.313	90.06 ± 0.313	6.57 ± 0.361	3.61 ± 0.313	89.82 ± 0.329



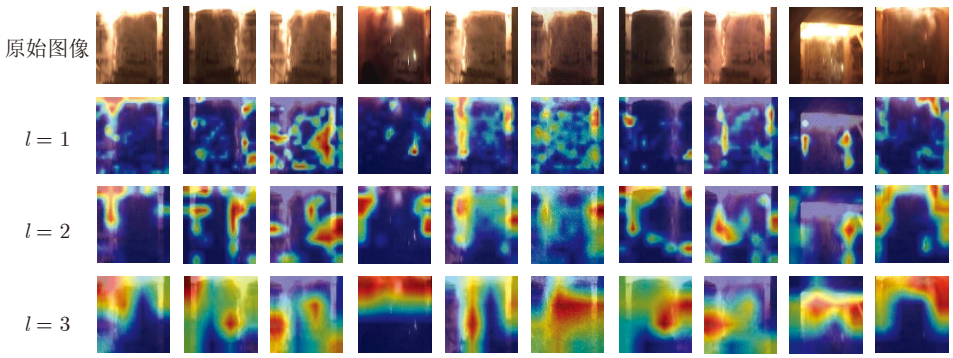
(a) 正常工况
(a) Normal conditions



(b) 欠烧工况
(b) Underburning conditions



(c) 过热工况
(c) Superheated operating conditions



(d) 异常排气工况
(d) Abnormal exhaust conditions

图 10 不同卷积层类激活映射图

Fig. 10 Different convolutional layer class activation mapping maps

为了更好地表明本文可解释性评估策略的准确性, 采用深度强化学习方法选取特征图进行可解释性评测, 与本文方法的实验对比结果如图 11 所示. 其中, 原始图像右侧第一行为使用强化学习方法选取特征图的可视化结果, 第二行为本文方法的可视化结果. 从图中可以看出, 本文方法相较于强化学习方法, 其高亮区域更加丰富连贯. 具体而言, 强化学习方法可解释性分析的高亮区域较小, 表明模型仅关注图像中的小部分细节和轮廓信息, 可能会影响模型的鲁棒性和泛化性能. 本文方法可以更加准确地定位与工况具有强关联性的感兴趣高亮区域, 具有物理含义的卷积核使得本文方法更加透明和易于理解.

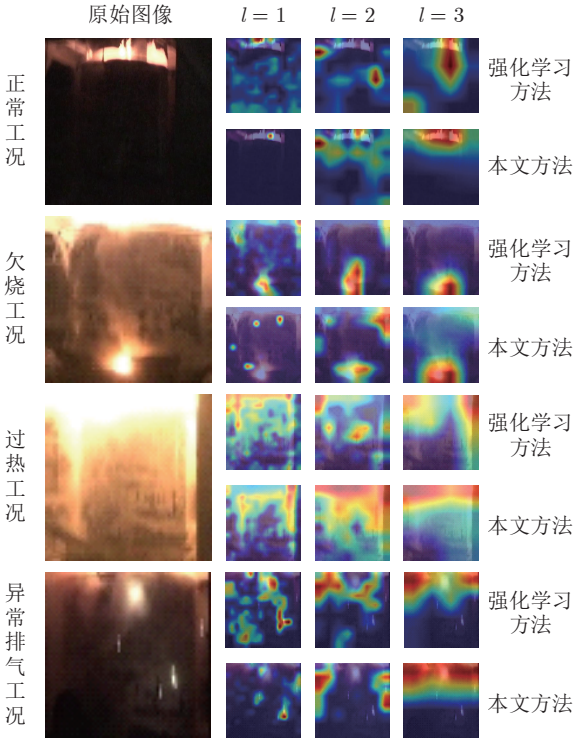


图 11　本文方法与基于强化学习的类激活映射图对比

Fig. 11　Comparison of the method proposed in this paper with the class activation mapping maps based on reinforcement learning

图 12 给出了某次采样实验中不同卷积层条件下, 本文方法与基于强化学习方法的可信识别样本比例变化曲线, 这里可解释性可信度指标阈值 $\mathcal{I}$ 设为 0.5, 当训练样本的 IoU 大于 $\mathcal{I}$ 时, 认为该样本的类别预测结果可信, 否则认为类别预测结果不可信. 从图中可以看出, 本文方法较采用强化学习选取特征图的策略均有着更优的表现. 当 $l=1$ 时, 本文方法中约有 87% 的训练样本满足可解释性可信度指标阈值, 将不满足的训练样本数据增强至原始数据

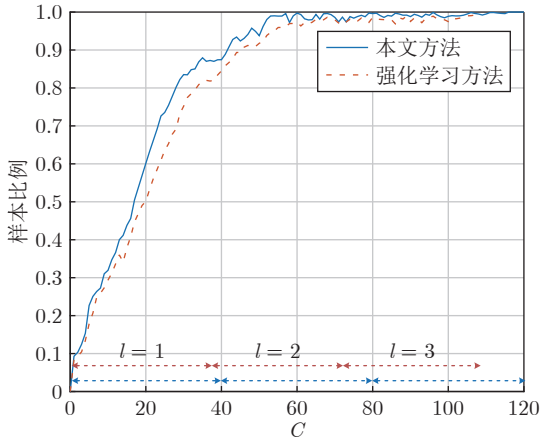


图 12　本文方法与基于强化学习的可信识别样本比例变化曲线

Fig. 12　The proportion change curves of trusted recognition samples based on reinforcement learning and the method proposed in this paper

大小, 增加新的卷积特征提取层, 当卷积层数增加至 3 时, 全体训练样本均满足可解释性可信度指标阈值, 网络层数停止增加, 表明深层卷积随机配置网络可以提取不同工况的有效特征, 关注的特征区域与真实标注区域重合. 强化学习方法移除了部分可提供独特贡献度的高斯差分卷积核, 降低了样本关注特征区域的表征能力, IoU 下降导致可信识别样本比例较低.

3.4　消融及对比实验

为了验证本文所提出的各个模块的有效性, 将本文方法与三个变体进行对比, 分别是未加入可解释性模块的方法、未加入高斯卷积核以及可解释性模块的方法、未加入高斯卷积核的方法. 实验中采用三层 DCSCN 网络, 实验结果见表 2.

由表 2 可知, 本文方法相比于去除可解释性模块、去除高斯卷积核、同时去除可解释性模块和高斯卷积核的方法, 训练集精度和测试集精度分别提升了 0.81%、1.53%、3.00%、1.19%、1.77%、3.17%. 因此, 本文所提出的可解释性模块不仅使模型更加透明和可信, 同时也提升了模型的性能. 另外, 使用具有物理含义的高斯差分卷积核构建的深层卷积网络, 使模型能够更好地适应电熔镁炉复杂情况的识别, 确保精度的不断提升. 使用类激活映射获取特征的可视化结果并定义可信度指标, 使工况识别结果更加准确及可信.

为了验证高斯噪声对数据的具体影响, 引入不同标准差 η 的高斯噪声并对其影响进行实验验证. 具体而言, 在三层 DCSCN 网络基础上, 实验中分

别设置了三个不同的标准差值, 即, $\eta = 0.3$ (本文方法所采用的标准差), $\eta = 0.6$, $\eta = 0.9$, 实验结果见表 3.

从表 3 可以看出, 随着高斯噪声标准差 η 的增大, 训练集和测试集性能逐渐下降, $\eta = 0.6$ 模型和 $\eta = 0.9$ 模型相较于本文方法, 训练集精度分别下降 1.86% 和 3.33%, 测试集精度分别下降 2.02% 和 6.77%, 结果表明, 噪声逐渐增加会使网络的构建更加困难, 容易导致过拟合或欠拟合问题. 具体原因分析如下, 高斯噪声、裁剪和缩放处理后, 图像会损失一部分细节特征, 适量噪声可以使模型对输入中的小变化更健壮, 当噪声水平过大时, 噪声会掩盖图像中的有用特征, 使模型难以从数据中学习有效的特征信息, 从而导致过拟合或欠拟合问题.

为了验证本文方法的有效性, 本文将其与 SCN[21]、块增量 BSC[33]、2DSCN[22]、DeepSCN[23]、CNN[34]、贝叶斯网络[6] 以及 CNN+LSTM[8] 的工况识别模型进行性能对比. 其中, SCN、块增量 BSC 和 2DSCN 的隐含层数均为 1, 隐含层节点数设置为 2 000, 块宽 $\triangle k = 5$. DeepSCN 的隐含层数为 4, 每个隐含层节点数均设置为 500, 激活函数为 sigmoid, 隐含层节点参数范围 $\lambda \in \{1, 3, 5, 7, 9, 10, 25, 50, 100\}$, 收缩序列 $r \in \{0.9, 0.99, 0.999, 0.999\,9, 0.999\,99, 0.999\,999\}$. 贝叶斯网络中使用 BN 参数迁移学习方法[6] 学习目标域 BN 模型参数. CNN 的网络结构包含 3 个卷积层, 3 个 sigmoid 激活层, 3 个池化层, 1 个全连接层, 训练轮数为 100 个 epoch.

图 13 给出了 5 种网络模型的训练样本识别精度曲线. 由图 13 可以看出, 其他 4 种模型的训练样本识别精度曲线随着卷积核个数/隐含层节点数的递增均呈现收敛趋势, SCN、块增量 BSC、2DSCN 和 DeepSCN 在 2 000 个隐含层节点处分别趋于 0.77、0.79、0.79、0.84 且上升缓慢, 而本文方法的识别精度变化更为剧烈, 在 120 次卷积时, 识别精度就可以达到 0.92.

表 4 给出了本文方法与其他模型的测试样本漏诊率、误诊率和精度对比. 可以看出, 本文方法相较于其他 7 种模型, 在漏诊率和误诊率上保持较低水平, 分别为 5.24% 和 2.45%. 精度相较于 SCN、块增量 BSC、2DSCN、DeepSCN、CNN、贝叶斯网络、CNN+LSTM 分别提升了 16.17%、15.46%、14.32%、8.67%、4.59%、2.39%、2.74%. 上述结果表明 SCN、块增量 BSC、2DSCN、DeepSCN 中的隐含层节点对于直接输入的图像数据特征提取能力不足, 本文方法采用增量式方法构建深层卷积随机配置网络, 基于监督学习机制随机配置具有物理含义的高斯差分卷积核参数, 有效提取电熔镁炉不同工况的特征, 确保识别误差逐级收敛. 本文方法构建类激活映射图获取电熔镁炉特征的可视化结果, 使得深度学习模型内部机理更加清晰, 定义可解释性可信度评测指标, 自适应调节网络层级对不可信识别结果的样本进行再认知, 以获取最优工况识别结果.

表 5 给出了本文方法与其他模型某次采样实验中的参数量、训练和测试时间对比. 在参数量方面, 本文方法中的卷积操作具有参数共享的优点, 加之深层网络的多次卷积和下采样操作, 使得输入全连接层的特征图尺寸减小, 参数量较 SCN、块增量 BSC、2DSCN 和 DeepSCN 减少了 1 至 2 个数量级, 复杂度大幅降低, 避免了模型的过拟合风险. 贝叶斯网络的参数量较少且训练时间较短, 但是推断

表 2 消融实验结果 (%)

Table 2 Results of ablation experiments (%)

模型	训练集			测试集		
	漏诊率	误诊率	精度	漏诊率	误诊率	精度
本文方法	5.31 ± 0.239	1.96 ± 0.165	92.73 ± 0.166	5.24 ± 0.245	2.45 ± 0.203	92.31 ± 0.283
未加入可解释性模块	5.57 ± 0.232	2.51 ± 0.223	91.92 ± 0.278	7.29 ± 0.173	1.59 ± 0.181	91.12 ± 0.347
未加入高斯卷积核	4.29 ± 0.274	4.51 ± 0.391	91.20 ± 0.264	3.45 ± 0.255	2.50 ± 0.329	90.54 ± 0.231
未加入可解释性模块以及高斯卷积核	6.02 ± 0.183	4.25 ± 0.231	89.73 ± 0.325	4.13 ± 0.242	6.73 ± 0.228	89.14 ± 0.179

表 3 不同高斯噪声的实验结果 (%)

Table 3 Experimental results with different Gaussian noises (%)

模型	训练集			测试集		
	漏诊率	误诊率	精度	漏诊率	误诊率	精度
本文方法 ($\eta = 0.3$)	5.31 ± 0.239	1.96 ± 0.165	92.73 ± 0.166	5.24 ± 0.245	2.45 ± 0.203	92.31 ± 0.283
$\eta = 0.6$ 模型	6.92 ± 0.232	2.21 ± 0.223	90.87 ± 0.206	7.19 ± 0.173	2.52 ± 0.181	90.29 ± 0.347
$\eta = 0.9$ 模型	8.31 ± 0.423	2.29 ± 0.248	89.40 ± 0.297	7.45 ± 0.382	7.01 ± 0.274	85.54 ± 0.288

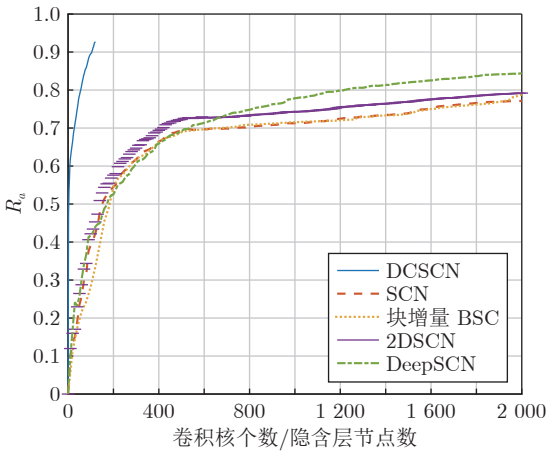


图 13　不同网络模型的训练样本识别精度曲线

Fig. 13　Recognition accuracy curves of training samples for different network models

表 4　不同模型的测试样本漏诊率、误诊率和精度对比 (%)

Table 4　Comparison of missed diagnosis rate, misdiagnosis rate and accuracy of test samples with different models (%)

模型	漏诊率	误诊率	精度
SCN	14.21 ± 0.228	14.21 ± 0.228	76.14 ± 0.215
块增量 BSC	12.58 ± 0.285	10.57 ± 0.153	76.85 ± 0.233
2DSCN	6.49 ± 0.263	15.52 ± 0.303	77.99 ± 0.353
DeepSCN	9.04 ± 0.285	7.32 ± 0.075	83.64 ± 0.209
CNN	6.82 ± 0.376	5.46 ± 0.167	87.72 ± 0.231
贝叶斯网络[6]	5.36 ± 0.268	4.72 ± 0.252	89.92 ± 0.256
CNN+LSTM[8]	6.91 ± 0.201	3.52 ± 0.184	89.57 ± 0.337
本文方法	5.24 ± 0.245	2.45 ± 0.203	92.31 ± 0.283

表 5　不同识别模型的综合性能对比

Table 5　Comprehensive performance comparison of different recognition models

模型	参数量 (MB)	训练时间 (s)	测试时间 (s)
SCN	500.038	10278.834	0.011
块增量 BSC	500.038	8341.094	0.011
2DSCN	1000.038	12352.771	0.013
DeepSCN	127.899	15411.081	0.013
CNN	0.664	20714.322	0.014
贝叶斯网络[6]	0.046	26.258	0.022
CNN+LSTM[8]	4.127	20159.642	0.015
本文方法	12.854	18218.021	0.014

过程时间较其他方法更长. 此外, 由于 DCSCN 每个特征图集合均与输出层连接, 因此, 虽然参数量较 CNN 和 CNN + LSTM 有所提高, 但漏诊率、误诊率和精度均为更优.

在训练时间方面, 本文方法、DeepSCN、CNN 和 CNN+LSTM 的多层模型训练时间均多于 SCN、块增量 BSC 和 2DSCN 的单层模型训练时间, 测试时间具有相同的数量级. 本文方法采用三通道的彩色图像作为输入, 加之卷积操作替代隐含层节点, 因此训练时间较采用灰度图像作为输入的 SCN、块增量 BSC、2DSCN 和 DeepSCN 更长. 不同于 CNN 和 CNN+LSTM 采用反向传播梯度下降法进行训练, 本文方法则基于监督学习机制增量式生成卷积核, 避免了权重初始化、局部最小值以及学习率敏感等问题.

3.5　跨数据集验证实验

为了验证本文所提出的深层卷积随机配置网络的泛化性和鲁棒性, 本文选择一种太阳能电池板公共数据集[35], 该数据集提取自单晶和多晶光伏模块图像. 该数据集包含 2 624 张图像, 图像分辨率为 300 × 300 像素, 涵盖了 4 种不同的太阳能电池板故障类型. 为保证实验的一致性, 首先对数据集进行统一的数据增强和预处理, 得到总计 10 496 张图像. 随机选取 80% 的图像作为训练集, 剩余 20% 的图像作为测试集. 实验参数设置与消融和比较实验保持一致. 表 6 给出了本文方法在太阳能电池板数据集上的漏诊率、误诊率和精度结果. 根据表中的数据, 可以清楚地看出在单层 DCSCN 和三层 DCSCN 条件下, 本文方法在测试集上表现出了较好的性能. 本文方法相较于未加入可解释性模块方法分别实现了 1.64% 和 1.39% 的精度提升, 并且进一步证明了本文方法在不同数据集上均具有良好的泛化性和鲁棒性.

表 6　太阳能电池板数据集实验结果对比 (%)

Table 6　Comparison of experimental results for solar panel dataset (%)

	模型	漏诊率	误诊率	精度
单层	本文方法	7.31 ± 0.187	7.86 ± 0.259	84.83 ± 0.245
	未加入可解释性模块	9.87 ± 0.252	6.94 ± 0.243	83.19 ± 0.279
三层	本文方法	3.45 ± 0.213	3.51 ± 0.169	93.04 ± 0.323
	未加入可解释性模块	4.13 ± 0.192	4.22 ± 0.257	91.65 ± 0.236

4　结论

针对现有电熔镁炉异常工况识别方法泛化能力差、可解释性弱等问题, 本文提出异常工况可解释性识别模型, 创新点如下:

1) 基于监督学习机制采用增量式随机配置策略, 生成具有物理含义的高斯差分卷积核构建深层

卷积网络, 确保识别误差逐级收敛, 具有网络结构透明和可解释性的特点.

2) 采用类激活映射方法对模型进行可解释性分析, 标识需关注的镁炉特征区域, 定义可解释性可信度评测指标, 自适应调节网络层级对不可信样本进行再认知, 以获取最优工况识别结果.

3) 本文方法的电熔镁炉异常工况漏诊率为 5.24%, 误诊率为 2.45%, 精度为 92.31%, 较其他识别方法更优.

未来将采用块增量技术, 进一步提升深层卷积随机配置网络的建模速度和精度.

References

1 Lu Shao-Wen, Wen Yi-Xin. Semi-supervised classification of semi-molten working condition of fused magnesium furnace based on image and current features. *Acta Automatica Sinica*, 2021, **47**(4): 891−902
(卢绍文, 温乙鑫. 基于图像与电流特征的电熔镁炉欠烧工况半监督分类方法. 自动化学报, 2021, **47**(4): 891−902)

2 Liu Qiang, Kong De-Zhi, Lang Zi-Qiang. Multi-level dynamic principal component analysis for abnormality diagnosis of fused magnesia furnaces. *Acta Automatica Sinica*, 2021, **47**(11): 2570−2577
(刘强, 孔德志, 郎自强. 基于多级动态主元分析的电熔镁炉异常工况诊断. 自动化学报, 2021, **47**(11): 2570−2577)

3 Wu Z W, Wu Y J, Chai T Y, Sun J. Data-driven abnormal condition identification and self-healing control system for fused magnesium furnace. *IEEE Transactions on Industrial Electronics*, 2015, **62**(3): 1703−1715

4 Wu Zhi-Wei. Embedded Intelligent Control System for Fused Magnesium Furnace [Ph.D. dissertation], Northeastern University, China, 2015.
(吴志伟. 嵌入式电熔镁炉智能控制系统研究 [博士学位论文], 东北大学, 中国, 2015.)

5 Li Hui, Wang Fu-Li, Li Hong-Ru. Abnormal condition identification and self-healing control scheme for the electro-fused magnesia smelting process. *Acta Automatica Sinica*, 2020, **46**(7): 1411−1419
(李荟, 王福利, 李鸿儒. 电熔镁炉熔炼过程异常工况识别及自愈控制方法. 自动化学报, 2020, **46**(7): 1411−1419)

6 Yan Hao, Wang Fu-Li, Sun Yu-Feng, He Da-Kuo. Abnormal condition identification based on Bayesian network parameter transfer learning for the electro-fused magnesia. *Acta Automatica Sinica*, 2021, **47**(1): 197−208
(闫浩, 王福利, 孙钰沣, 何大阔. 基于贝叶斯网络参数迁移学习的电熔镁炉异常工况识别. 自动化学报, 2021, **47**(1): 197−208)

7 Lu S W, Wen Y X. Semi-supervised condition monitoring and visualization of fused magnesium furnace. *IEEE Transactions on Automation Science and Engineering*, 2022, **19**(4): 3471−3482

8 Wu Gao-Chang, Liu Qiang, Chai Tian-You, Qin S. Joe. Abnormal condition diagnosis through deep learning of image sequences for fused magnesium furnaces. *Acta Automatica Sinica*, 2019, **45**(8): 1475−1485
(吴高昌, 刘强, 柴天佑, 秦泗钊. 基于时序图像深度学习的电熔镁炉异常工况诊断. 自动化学报, 2019, **45**(8): 1475−1485)

9 Lu S W, Gao H R. Deep learning based fusion of RGB and infrared images for the detection of abnormal condition of fused magnesium furnace. In: Proceedings of the IEEE 15th International Conference on Control and Automation (ICCA). Edinburgh, UK: IEEE, 2019. 987−993

10 Bu K Q, Liu Y, Wang F L. Operating performance assessment based on multi-source heterogeneous information with deep learning for smelting process of electro-fused magnesium furnace. *ISA Transactions*, 2022, **128**: 357−371

11 Lu Shao-Wen, Wang Ke-Dong, Wu Zhi-Wei, Li Peng-Qi, Guo Zhang. Online detection of semi-molten of fused magnesium furnace based on deep convolutional neural network. *Control and Decision*, 2019, **34**(7): 1537−1544
(卢绍文, 王克栋, 吴志伟, 李鹏琦, 郭章. 基于深度卷积网络的电熔镁炉欠烧工况在线识别. 控制与决策, 2019, **34**(7): 1537−1544)

12 Zeiler M D, Fergus R. Visualizing and understanding convolutional networks. In: Proceedings of the 13th European Conference on Computer Vision (ECCV). Zurich, Switzerland: Springer, 2014. 818−833

13 Samek W, Binder A, Montavon G, Lapuschkin S, Müller K R. Evaluating the visualization of what a deep neural network has learned. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, **28**(11): 2660−2673

14 Larsen A B L, Sonderby S K, Larochelle H, Winther O. Autoencoding beyond pixels using a learned similarity metric. In: Proceedings of the 33rd International Conference on Machine Learning. New York, USA: PMLR, 2016. 1558−1566

15 Zhang Q S, Wu Y N, Zhu S C. Interpretable convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 8827−8836

16 Pham H, Guan M, Zoph B, Le Q, Dean J. Efficient neural architecture search via parameters sharing. In: Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: PMLR, 2018. 4095−4104

17 Ding Z X, Chen Y R, Li N N, Zhao D B, Sun Z Q, Chen C L P. BNAS: Efficient neural architecture search using broad scalable architecture. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, **33**(9): 5004−5018

18 Alvarez J M, Salzmann M. Learning the number of neurons in deep networks. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc., 2016. 2270−2278

19 Kawaguchi K. Deep learning without poor local minima. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc., 2016. 586−594

20 Pao Y H, Park G H, Sobajic D J. Learning and generalization characteristics of the random vector functional-link net. *Neurocomputing*, 1994, **6**(2): 163−180

21 Wang D H, Li M. Stochastic configuration networks: Fundamentals and algorithms. *IEEE Transactions on Cybernetics*, 2017, **47**(10): 3466−3479

22 Li M, Wang D H. 2-D stochastic configuration networks for image data analytics. *IEEE Transactions on Cybernetics*, 2021, **51**(1): 359−372

23 Wang D H, Li M. Deep stochastic configuration networks with universal approximation property. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN). Rio de Janeiro, Brazil: IEEE, 2018. 1−8

24 Pratama M, Wang D H. Deep stacked stochastic configuration networks for lifelong learning of non-stationary data streams. *Information Sciences*, 2019, **495**: 150−174

25 Li W T, Zhang Q, Wang D H, Sun W, Li Q Y. Stochastic configuration networks for self-blast state recognition of glass insulators with adaptive depth and multi-scale representation. *Information Sciences*, 2022, **604**: 61−79

26 Li W T, Deng Y L, Ding M S, Wang D H, Sun W, Li Q Y. Industrial data classification using stochastic configuration networks with self-attention learning features. *Neural Computing*

and Applications, 2022, **34**: 22047−22069

27 Peng S Y, Ding L J, Li W T, Sun W, Li Q Y. Research on intelligent recognition method for self-blast state of glass insulator based on mixed data augmentation. *High Voltage*, 2023, **8**(4): 668−681

28 Zhang Q, Li W T, Li H, Wang J P. Self-blast state detection of glass insulators based on stochastic configuration networks and a feedback transfer learning mechanism. *Information Sciences*, 2020, **522**: 259−274

29 Li W T, Tao H, Li H, Chen K Q, Wang J P. Greengage grading using stochastic configuration networks and a semi-supervised feedback mechanism. *Information Sciences*, 2019, **488**: 1−12

30 Li W T, Chen K Q, Wang D H. Industrial image classification using a randomized neural-net ensemble and feedback mechanism. *Neurocomputing*, 2016, **173**: 708−714

31 He Y, Ding Y H, Liu P, Zhu L C, Zhang H W, Yang Y. Learning filter pruning criteria for deep convolutional neural networks acceleration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020. 2006−2015

32 Sutton R S, Barto A G. *Reinforcement Learning: An Introduction*. Cambridge: MIT Press, 1998.

33 Dai W, Li D P, Zhou P, Chai T Y. Stochastic configuration networks with block increments for data modeling in process industries. *Information Sciences*, 2019, **484**: 367−386

34 LeCun Y. LeNet-5, convolutional neural networks [Online], available: http://yann.lecun.com/exdb/lenet, January 11, 2024

35 Deitsch S, Christlein V, Berger S, Buerhop-Lutz C, Maier A, Gallwitz F, et al. Automatic classification of defective photovoltaic module cells in electroluminescence images. *Solar Energy*, 2019, **185**: 455−468

李帷韬 合肥工业大学电气与自动化工程学院副教授. 主要研究方向为深度学习, 图像处理和智能认知.
E-mail: wtli@hfut.edu.cn
(**LI Wei-Tao** Associate professor at the School of Electrical Engineering and Automation, Hefei University of Technology. His research interest covers deep learning, image processing, and intelligent cognition.)

童倩倩 合肥工业大学电气与自动化工程学院硕士研究生. 主要研究方向为智能认知.
E-mail: 2021110400@mail.hfut.edu.cn
(**TONG Qian-Qian** Master student at the School of Electrical Engineering and Automation, Hefei University of Technology. Her main research interest is intelligent cognition.)

王殿辉 中国矿业大学人工智能研究院教授. 主要研究方向为工业大数据建模与分析, 随机配置学习理论及工业应用. 本文通信作者.
E-mail: dh.wang@deepscn.com
(**WANG Dian-Hui** Professor at the Institute of Artificial Intelligence, China University of Mining and Technology. His research interest covers industrial big data modeling and analysis, stochastic configuration learning theory and industrial applications. Corresponding author of this paper.)

吴高昌 东北大学流程工业综合自动化国家重点实验室副教授. 主要研究方向为智能计算成像, 深度学习和异常工况智能感知与预测.
E-mail: wugc@mail.neu.edu.cn
(**WU Gao-Chang** Associate professor at the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University. His research interest covers intelligent computational imaging, deep learning, and intelligent sensing and prediction of abnormal working conditions.)