DOI: 10.19666/j.rlfd.202602026

基于改进 ConvNeXt 的垃圾焚烧炉火焰燃烧状态识别

赵 征，杜科江
（华北电力大学自动化系，河北 保定 071003）

［摘 要］【目的】针对炉排式垃圾焚烧炉膛中典型火焰燃烧状态具有多样性且人工识别效率低的问题，提出了一种基于 ConvNeXt 网络改进的火焰燃烧状态识别模型 MSDRNeXt，旨在实现垃圾炉膛燃烧工况的智能判别。【方法】首先，为了在炉膛亮度波动、扬尘遮挡等环境下识别有效火焰燃烧区域，提出了多尺度特征融合(multi scale feature fusion, MSFF)模块，通过引入卷积块注意力模块(convolutional block attention module, CBAM)机制，精炼网络浅层火焰特征图的纹理特征，并补充至深层语义特征图，提升了模型在复杂环境下识别火焰的鲁棒性。其次，基于炉膛左右侧配风的结构特性设计了双区域差分池化模块，使模型能高效识别偏烧、出生料、堆料这类左右燃烧具有明显差异的工况。【结果】实验结果表明，MSDRNeXt 在垃圾焚烧火焰数据集上的整体准确率为 91.36%，推理速度达到 277.85 帧/s，同时模型参数量较基线模型减少了 53.43%。【结论】所提出的改进网络 MSDRNeXt 在垃圾焚烧炉膛场景中能高效、准确识别火焰燃烧状态，可为垃圾焚烧的实时监测提供技术支持，从而优化后续控制策略。

［关 键 词］垃圾发电；火焰图像；注意力机制；多尺度特征融合；差分池化

［引用本文格式］赵征，杜科江. 基于改进 ConvNeXt 的垃圾焚烧炉火焰燃烧状态识别[J]. 热力发电，年，卷(期): 起始页码-终止页码. ZHAO Zheng, DU Kejiang. Recognition of flame combustion state in garbage incinerators based on improved ConvNeXt[J]. Thermal Power Generation, 年，卷(期): 起始页码-终止页码.

Recognition of flame combustion state in garbage incinerators based on improved ConvNeXt

ZHAO Zheng, DU Kejiang
(Department of Automation, North China Electric Power University, Baoding 071003, China)

Abstract: [Objective] As municipal solid waste generation increases, waste incineration has become a mainstream treatment method due to its advantages in reduction, harmlessness, and resource recovery. To address the challenges posed by the diversity of flame combustion states and the low efficiency of manual identification in grate-fired municipal solid waste incineration furnaces, this study proposes an improved flame combustion state recognition model, termed MSDRNeXt, based on the ConvNeXt architecture. The aim is to achieve intelligent identification of combustion conditions within the furnace, thereby reducing reliance on manual judgment and improving operational stability and safety in industrial applications.[Methods] To accurately extract effective flame regions under complex conditions such as significant brightness fluctuations and dust occlusion, a Multi-Scale Feature Fusion (MSFF) module is first proposed. This module integrates multi-scale feature information and incorporates the Convolutional Block Attention Module (CBAM) to enhance fine-grained texture

收 稿 日 期：2026-02-10 修 回 日 期：2026-04-17 接 受 日 期：2026-04-22 网络首发日期：
基 金 项 目：国家自然科学基金项目(52276007)；深圳市科学计划项目 (KCXFZ20201221173402007)
Supported by：National Natural Science Foundation of China(52276007)；Shenzhen Special Sustainable Development Science and Technology Project(KCXFZ20201221173402007)
第一作者简介：赵征(1978），女，副教授，博士，主要研究方向为热工过程建模与控制，电力视觉等, zheng_zhao@ncepu.edu.cn。

features in shallow feature maps while suppressing irrelevant noise. The refined shallow features are further fused into deeper semantic feature maps, enabling a better balance between local texture representation and global semantic understanding. In addition, data augmentation and normalization strategies are applied during training to further improve the model's robustness to environmental variations. This design improves the robustness and generalization performance of the model in complex environments. Furthermore, considering the inherent asymmetry of air distribution on the left and right sides of the furnace, a Dual-Region Differential Pooling (DRDP) module is designed. This module partitions the feature space into left and right regions and performs differential pooling to explicitly model spatial asymmetry. As a result, the model can effectively capture subtle inter-regional differences, which are critical for distinguishing combustion states with significant lateral variations, such as biased combustion, raw material emergence, and material accumulation. The integration of MSFF and DRDP enhances the discriminative capability of the model while maintaining computational efficiency.[Results] Experimental results on a dedicated waste incineration flame dataset demonstrate that the proposed MSDRNeXt model achieves an overall classification accuracy of 91.36%, outperforming the baseline ConvNeXt model and several comparative methods. In addition, the model achieves an inference speed of 277.85 frames per second (FPS), which satisfies the requirements for real-time monitoring in industrial scenarios. Meanwhile, the number of parameters is reduced by 53.43% compared with the baseline model, indicating a lightweight architecture with lower computational cost and memory consumption.[Conclusion] The proposed MSDRNeXt model can efficiently and accurately identify flame combustion states in municipal solid waste incineration furnaces. It provides a reliable solution for real-time monitoring of the incineration process and supports the optimization of subsequent control strategies.

Keywords: waste-to-energy; flame images; attention mechanism; multi-scale feature fusion; differential pooling

随着我国城市化进程的加速和居民消费水平的持续提升，城市固体废弃物（municipal solid waste, MSW）的产生速度日益增长[1]，如何高效、安全、环保地处理垃圾已成为全球范围内亟待解决的重大问题。目前，垃圾焚烧发电是实现固废“无害化、减量化、资源化”处理的主要途径[2]，由于 MSW 具有成分复杂、含水率高、热值波动大的燃烧特性，为了维持焚烧炉内垃圾的稳定燃烧，需要实时监测炉膛内的燃烧过程，运行人员根据炉内火焰燃烧状态调节给料量、风配比、炉排运行速度等参数，确保炉膛温度、压力等参数的稳定性，从而降低如二噁英、NO_x 等污染物排放[3]，提高垃圾发电效率。

目前，运行人员根据炉膛内部工业相机捕捉到的火焰图像进行燃烧状态判断，同时，为了提高判断准确性，仍然需要运行人员不定期地前往炉膛侧面看火孔进一步观测，并依靠工作经验主动优化焚烧控制策略。然而，这种需要依靠人工经验的诊断方式存在较强的主观性，并且还伴随着响应滞后、标准不统一等问题[4-5]，难以满足垃圾焚烧发电厂的智能控制需求。此外，在炉排式垃圾焚烧炉膛这一特定场景中，一方面，受炉膛内的强光干扰，火焰与炉膛背景在颜色空间上的区分度较低，导致火焰纹理模糊不清，难以有效实现关键提取的提取；另一方面，若炉排上垃圾堆料不均匀或者炉膛两侧配风不当，可能导致垃圾燃烧不均匀，从而产生偏烧现象。

为了减少对人工经验的依赖，同时准确高效识别火焰燃烧状态，通过引入计算机视觉技术对炉膛火焰图像进行智能判别。在图像识别领域，早期研究主要采用传统的机器学习方法。这类方法通过人工设计视觉特征，实现对目标的分类识别。例如，葛红等[6]基于图像处理方法，通过机器学习提取火焰亮度、温度、闪烁频率等特征参数，实现火焰稳定性的定量检测和评估。杨绪兵等[7]提出了一种基于矩阵分块的半监督学习算法 Semi-MHKS，提升了模型对林火的识别能力。宋昱等[8]基于 Log-Gabor 小波提取火焰图像纹理特征，再通过支持向量机分类，实现了锅炉燃烧火焰状态的识别。虽然上述机器方法具有一定的识别效果，但不具备自动提取高层特征的能力，特别是面对垃圾焚烧炉膛内部扬尘密集、强光干扰等复杂环境时，十分依赖人工设计特征，泛化能力较差。

近几年，深度学习技术在火焰图像识别领域迅速发展并成为主流方法[9]。赵东卓[10]提出了一种结合灰度反转的卷积神经网络，实现了对火焰图像燃烧状态的准确诊断，且具有较强的鲁棒性。白家宇等[11]通过在颈部网络层中引入高效注意力机制(efficient channel attention，ECA)，从而增强对输出特征的表达能力，提升了对目标的检测精度。文献[12]构建了一种轻量级的双向特征金字塔网络，用于识别隧道火灾火焰。 文献[13]提出了一种颈部采用 FPN+FAN 的改进 YOLOV5 结构，实现了双向路

径的多尺度融合，对不同的火焰识别具有良好鲁棒性。文献[14]设计了一种基于多尺度特征提取的卷积神经网络进行火情识别，并融入特征自适应权重模块，增强了多场景下火焰关键特征的识别能力。文献[15]设计了一种结合 ECA 注意力机制以及 DSConv 卷积技术的火焰检测算法，在火焰烟雾探测中具有较高的准确度。

上述深度学习方法在一般火焰识别任务中表现良好，然而对于垃圾焚烧炉膛类场景具有一定局限性。因此，针对上述情况，本文提出了一种契合炉排式垃圾焚烧炉膛的火焰识别模型，实现了火焰燃烧状态的高效、准确判断，有效弥补了传统人工诊断方式的不足，同时可为垃圾发电厂的燃烧监测提供技术支持。主要贡献如下：

1）提出了基于改进 ConvNeXt 的炉排式垃圾焚烧火焰识别模型 MSDRNeXt，实现了火焰燃烧状态的高效识别。

2）设计了一种跨层级的多尺度特征融合(multi scale feature fusion，MSFF)模块，增强了模型对火焰细节的表达能力。

3）基于炉排式垃圾焚烧炉膛内部结构先验，设计了双区域差分池化(dual-region difference pooling, DRDP)模块，从而提升模型整体识别准确率。

1 火焰识别模型

垃圾焚烧过程主要通过安装在炉膛后墙看火孔处的电荷耦合器件(CCD)工业相机进行实时监测，垃圾焚烧炉膛结构如图 1 所示。

由图 1 可见，该焚烧炉炉排由具有一定倾角的多级往复炉排组成，其中，干燥炉排对垃圾进行干燥并析出水分；随后送至气化炉排进行剧烈燃烧；垃圾在燃烬炉排大部分燃尽并形成炉渣，随后送入出渣口推出炉膛。同时，炉排底部的一次风在干燥段辅助垃圾水分析出，在气化段为垃圾焚烧提供充足氧气；二次风在炉膛侧墙喷入，与未燃烧的气体充分反应，减少污染物的排放。

针对炉排式垃圾焚烧炉膛这一场景，本文将结合炉膛内部的物理结构对火焰识别模型进行设计。

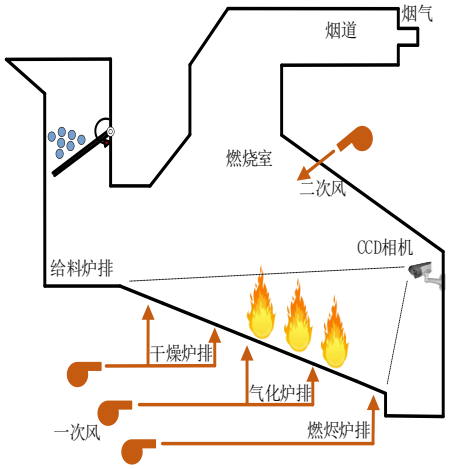


图 1 炉排式垃圾焚烧炉膛示意
Fig.1 Diagram of grate type waste incinerator

1.1 MSDRNeXt 整体架构

ConvNeXt[17]是一种基于 ResNet 架构并融入 Swin Transformer 设计理念的纯卷积神经网络。ConvNeXt 系列包含 Tiny、Small、Base、Large 4 种规格，鉴于本文的数据集规模有限，实验选用轻量级的 ConvNeXt_Tiny 作为基线分类模型，其网络结构包含 4 个核心网络层；然而，深层阶段（阶段 4）的输出特征图分辨率过低，虽然富含高层语义信息，但会丢失大量重要的纹理与边缘信息；此外，其过深的网络层级大幅增加了模型的参数计算量，易导致过拟合现象。因此，本文对主干网络进行精简优化，移除了原始架构中的阶段 4 模块，在保留火焰全局空间分布、宏观燃烧形态等语义信息的同时，实现了网络的轻量化改进。图 2 为 MSDRNeXt 模型结构。

由图 2 可见，原始火焰图像首先调整为 224×224×3 尺寸，并输入至 MSDRNeXt 的 Stem 初始网络层。接着，送入主干网络 Stage1、Stage2、Stage3 做进一步的特征提取。随后，通过 MSFF 模块进行特征融合，使火焰细小边缘、火焰亮度等浅层信息能够补充至高层语义特征中，从而更准确地识别各类燃烧状态。其次，在特征图向量化阶段，利用 DRDP 模块替代原网络的全局平均池化(global average pooling, GAP)，有效量化火焰偏烧程度，从而提高模型对偏烧、堆料、出生料类别图像的识别能力。经过上述过程，网络最终会输出一个 11 维的向量，并通过 Softmax 函数激活后转换为对应类别的预测概率。

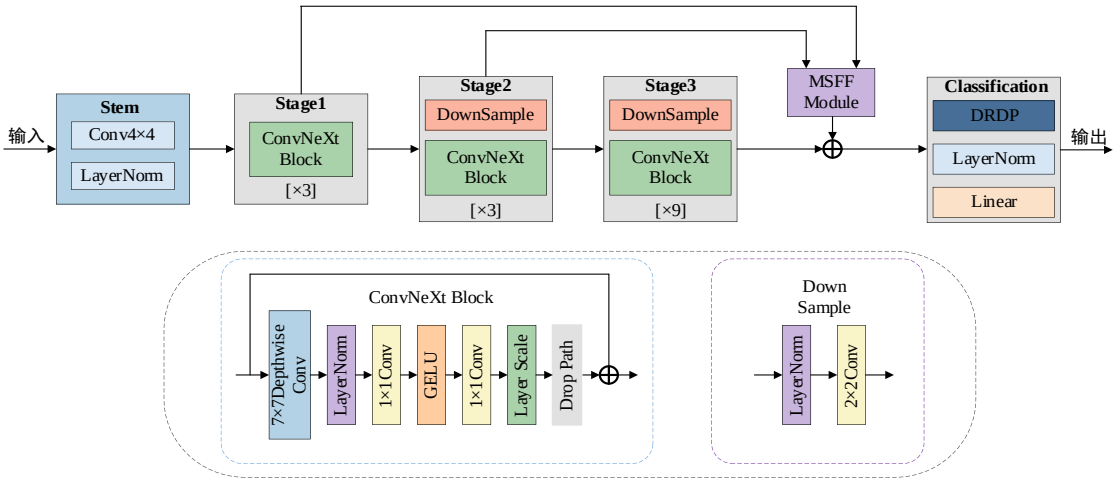


图 2 MSDRNeXt 模型结构

Fig. 2 MSDRNeXt model architecture

1.2 MSFF 模块

原始 ConvNeXt 网络经过多次下采样后，其特征图的空间分辨率会被极大压缩，导致部分关键纹理与边缘等信息在逐级压缩中丢失，模型在区分“扬尘”“堆料-生料”等高度依赖纹理细节特征的燃烧状态时效果不佳，为此设计多尺度特征融合 MSFF 模块（图 3）。模型提取网络浅层特征图 F_1 和 F_2 的特征，这些特征经过拼接融合后，通过卷积块注意力模块(convolutional block attention module, CBAM)在通道与空间维度上进行加权筛选，再补充至深层特征图 F_3 中，增强模型对火焰细微燃烧差异的表达能力。MSFF 模块的具体实现过程如下：

定义骨干网络 3 个阶段的特征输出分别为 $\{F_1, F_2, F_3\}$，其中 F_3 为主干特征输出。在进行多尺度特征融合前，首先，通过自适应平均池化分别将浅层特征 F_1 和 F_2 的空间分辨率下采样至与 F_3 空间尺寸对齐；其次，通过 1×1 卷积层对浅层特征的通道维度进行压缩。该过程描述为：

$$F_1'=\mathrm{Conv}(\mathrm{AvgPool}(F_1)) \tag{1}$$

$$F_2'=\mathrm{Conv}(\mathrm{AvgPool}(F_2)) \tag{2}$$

式中：Avgpool 表示自适应平均池化；Conv 表示卷积操作；F_1'和F_2'分别表示经过空间尺度对齐后的特征。

随后，将处理后的浅层特征沿通道维度进行拼接，并利用 1×1 卷积进行融合，将通道数调整为与 F_3 一致。该过程描述为：

$$F_\mathrm{f}=\mathrm{Conv}(\mathrm{Concat}[F_1', F_2']) \tag{3}$$

式中：Concat 表示通道维度拼接；F_f 表示融合特征。

为了筛选与火焰纹理与亮度变化等有关的特征，同时抑制噪声干扰，引入了卷积块注意力模块(convolutional block attention module, CBAM)对融合特征 F_f 在通道和空间维度进行精炼。

其中，通道注意力模块的压缩比设置为 4，采用较小的压缩比是为了保留更多的通道维度细节，减少火焰特征信息丢失；空间注意力模块采用 7×7 的卷积核，能够赋予模型更大的空间感受野。如图 3 中虚线框所示，首先，F_f 经过通道注意力模块生成通道权重向量$\boldsymbol{\omega}_\mathrm{c}$，并对特征输入进行加权计算，得到中间特征图$F_\mathrm{c}$。该过程表示为：

$$F_\mathrm{c}=\boldsymbol{\omega}_\mathrm{c}\square F_\mathrm{f} \tag{4}$$

接着，中间特征图经过空间注意力模块，生成空间权重注意力图，二者同样进行加权计算，得到空间注意力加权后的特征图，最后，经过完整 CBAM 注意力机制精炼后的浅层特征补充至网络深层特征 F_3，实现多尺度的特征融合。该过程表示为：

$$F_3'=F_3+\boldsymbol{\omega}_\mathrm{s}\square F_\mathrm{c} \tag{5}$$

式中：$\square$表示 Hadamard 积；$\boldsymbol{\omega}_\mathrm{s}$表示空间注意力模块生成的空间权重矩阵；$F_\mathrm{c}$表示经过通道注意力加权后的中间特征图；$F_3'$表示 MSFF 模块最终输出的多尺度融合特征图。

通过上述机制，MSFF 模块在保留原深层语义特征的前提下，有效地融合了多尺度的浅层细节特征，提升了模型在识别复杂纹理火焰图像时的鲁棒性。

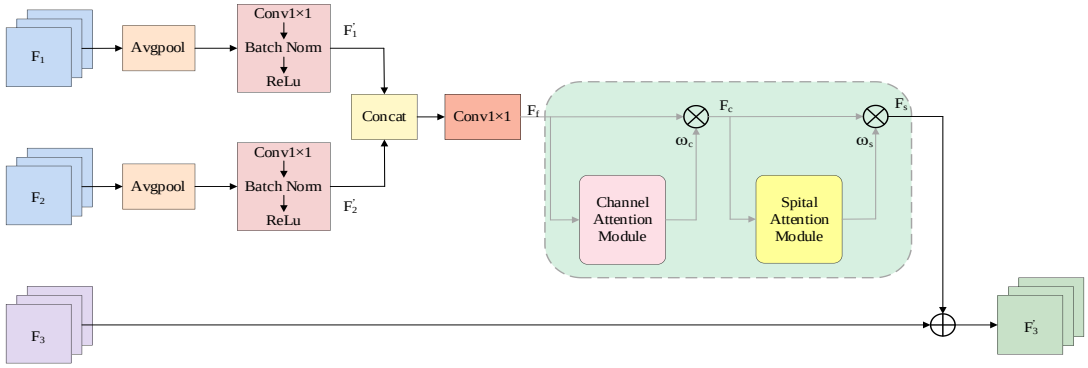


图 3 多尺度特征融合结构

Fig. 3 Structure diagram of cross-stage feature fusion

1.3 双分区差分池化

MSFF 模块虽然有效增强了模型的特征表达能力，但仅依靠多尺度融合后的特征信息仍难以敏感捕捉火焰在炉膛空间分布上的差异。在特征图转换为向量的过程，传统全局平均池化导致特征空间信息被均质化，无法有效表征炉膛左右两侧的燃烧差异。为此，本文设计了 DRDP 双分区差分池化模块，该模块基于炉排式垃圾焚烧炉膛的内部结构先验，首先，将特征图沿宽度方向划分为左右 2 个子区域，分别进行池化以提取区域特征；然后，进行特征差分计算，量化火焰左右区域的燃烧差异；最后，利用轻量化的通道注意力机制 ECA[18]自适应地生成各特征通道的权重系数，提升了模型对火焰空间分布差异的表征能力。图 4 为双分区差分池化结构。

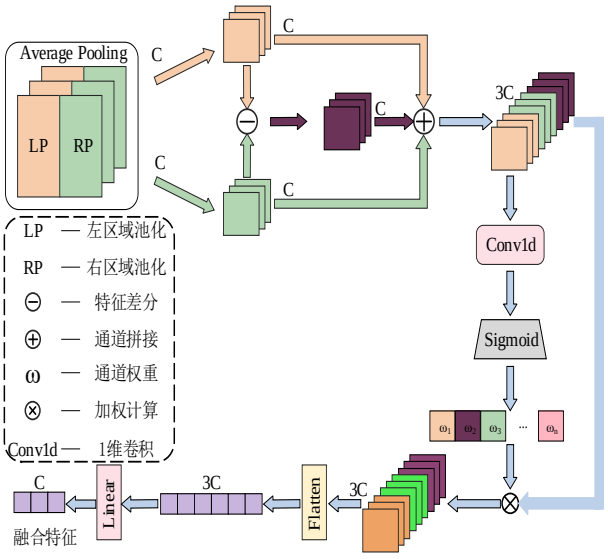


图 4 双分区差分池化结构

Fig. 4 Structure diagram of the dual-region difference pooling

将经过 MSFF 模块输出的多尺度融合特征图 $F_3' \in \mathbb{R}^{C\times H\times W}$ 沿宽度方向均分为左侧区域 R_l 和右侧区域 R_r，R_l、$R_r \in \mathbb{R}^{C\times H\times W/2}$；随后，分别对 2 个子区域平均池化并计算其特征差分。此过程表示为：

$$\Delta \boldsymbol{F}=\mathrm{LP}(\boldsymbol{R}_l)\ominus\mathrm{RP}(\boldsymbol{R}_r) \tag{6}$$

式中：$\Delta \boldsymbol{F} \in \mathbb{R}^{C\times 1\times 1}$ 表示特征差分张量。

其次，通过将左、右侧区域平均池化得到的特征张量 $\boldsymbol{F}_l$ 和 $\boldsymbol{F}_r$ 与其特征差分张量 $\Delta \boldsymbol{F}$ 沿通道维度拼接，得到一个携带区域燃烧强度信息的联合特征张量。此过程可以表示为：

$$\boldsymbol{F}_c=\mathrm{Concat}(\boldsymbol{F}_l,\Delta \boldsymbol{F},\boldsymbol{F}_r) \tag{7}$$

式中：Concat 表示拼接操作；$\boldsymbol{F}_c \in \mathrm{R}^{3C\times 1\times 1}$ 表示拼接后的融合特征。

为进一步对融合特征进行精炼，引入一种轻量级通道注意力机制，对 $\boldsymbol{F}_c$ 进行卷积计算。其采用尺寸自适应的一维卷积核，大小为 k，并通过 Sigmoid 激活函数生成通道权重。此过程可以表示为：

$$k=\left|\frac{\log_2(C)}{\gamma}+\frac{b}{\gamma}\right|_{\mathrm{odd}} \tag{8}$$

$$\omega=\sigma(\mathrm{Conv1d}_{\boldsymbol{k}}(\boldsymbol{F}_c)) \tag{9}$$

式中：C 表示图像通道；γ 表示比例参数；b 表示偏置参数，采用最优经验配置 γ=2，b=1；σ表示 Sigmoid 激活函数；$\boldsymbol{\omega}$表示生成的通道权重。

将通道权重向量与融合特征 $\boldsymbol{F}_c$ 逐元素相乘，实现通道级加权，得到加权融合特征 $\boldsymbol{F}_c'$ 此过程可以表示为：

$$\boldsymbol{F}_c'=\boldsymbol{F}_3+\boldsymbol{\omega}\Delta \boldsymbol{F}_c \tag{10}$$

最后，通过全连接层将通道维度从 $3C$ 映射回原始通道数 C，最终输出至分类头进行决策。

这种基于特征差分和 ECA 的双区域池化改进机制，得到的特征图携带很强的燃烧差异信息，从而有效提升了模型对左右侧燃烧具有明显差异工况的识别能力。

2 实验设置

2.1 数据集来源

本文以某垃圾焚烧电厂的炉排式垃圾焚烧炉为研究对象，火焰图像数据集通过对炉膛内连续 3 日的运行视频进行抽帧构建。在数据采集过程中，镜头首先接收外界光信号，再由 CCD 图像传感器转换为电信号；随后，这些电信号在主控芯片中进行图像信号处理与数字信号处理，转换为可按照 H.265 标准进行压缩编码的数据并通过 PHY 芯片封装为网络视频码流，经 RJ45 接口实现数据传输，最终值控现场通过接入局域网获取实时视频流。视频最高分辨率 1 980×1 080，标准采样帧率为 25 FPS。所采集各类样本数量分布见表 1，共 5 280 张图片，其中训练集与测试集样本数量比值为 5:1。

表 1 各类别样本数量
Tab.1 Sample count for each category

序号	类别	训练集	测试集	样本数量
1	右偏烧	470	94	564
2	右出生料	415	83	498
3	右堆料	375	75	450
4	堆料-生料	265	53	318
5	左偏烧	440	88	528
6	左出生料	430	86	516
7	左堆料	390	78	468
8	扬尘	430	86	516
9	正常	535	107	642
10	火线高	300	60	360
11	面积大	350	70	420

2.2 数据集划分

图 5 展示了一种典型偏烧工况。对于单张火焰图像，可将图像横向划分为如图 5 所示 4 个区域，其中I、II、III区域分别对应炉排的部分燃烬段、燃烧段、干燥段[19]。图 5 中的红色线段为“燃烬火线”，即固化物燃烧结束变为灰烬的位置，可以定量反映当前垃圾在炉排上的燃烧状况。由图 5 可见，其燃烬火线左右两侧分布区域不同，表现出偏烧现象。

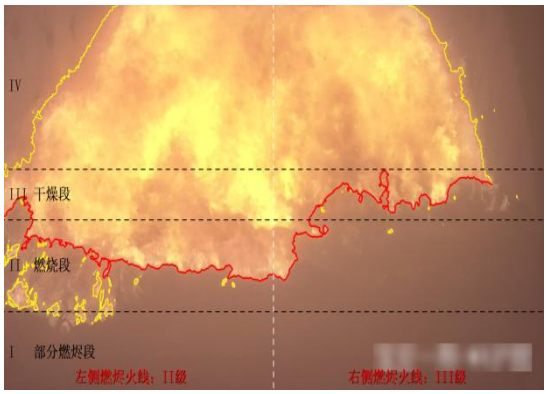


图 5 典型偏烧工况
Fig.5 Typical uneven combustion conditions

火焰图像按燃烧状态划分为正常、左偏烧、右偏烧、左堆料、右堆料、面积大、堆料-生料、右出生料、左出生料、火线高、扬尘 11 类，图 6 为 11 类火焰燃烧状态（镜像）。这些数据集的类别划分能够反映垃圾焚烧过程中的大部分典型工况，具有一定的代表性。

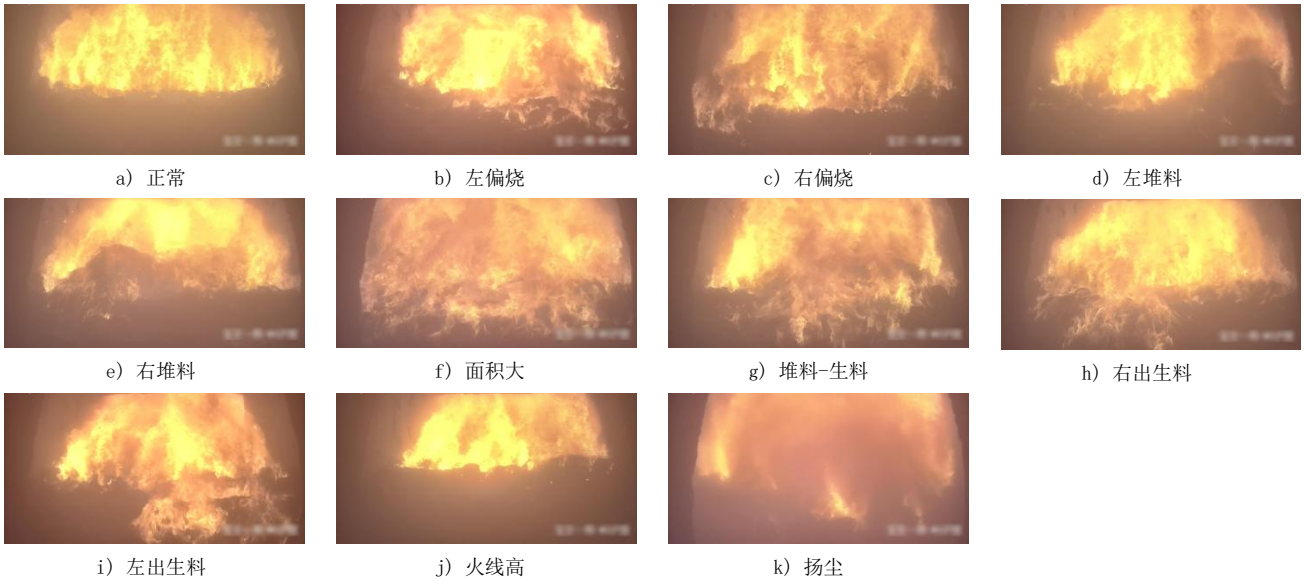


图 6 11 类火焰燃烧状态
Fig.6 11 classes of flame conditions

以火焰图像中轴线为基准，划分为左右两区域。分别统计两侧燃烬火线像素点在各区域的分布情况，并确定其像素点分布最多的区域，即燃烬火线对应的等级，其中等级小的一侧为偏烧方向。具体判定规则如下。

1）扬尘：在标注前，首先计算火焰图像数据集的平均梯度（average gradient, AG），并将 AG< 0.04 的图像结合专家经验判定为扬尘。

2）正常：两侧燃烬火线等级均为II级。

3）火线高：两侧燃烬火线等级均为III级或IV级。

4）面积大：两侧燃烬火线等级均为I级。

5）出生料：一侧燃烬火线等级为I级，另一侧为II或III级。

6）偏烧：一侧燃烬火线等级为II级，另一侧为III级。

7）堆料：一侧燃烬火线等级为IV级，另一侧为II或III级，而且火线下方具有明显料层堆积，如图 6d)、图 6e)所示。

8）堆料-生料：一侧燃烬火线等级为I级，另一侧为IV级。

对于左右侧燃烧不均类别，即“出生料”“偏烧”“堆料”“堆料-生料”4 种状态，可通过计算其偏烧率做进一步的定量验证，表 2 统计了上述各类所有数据集的偏烧率平均值，其计算公式为：

$$S=\frac{\sum_{1}^{n/2}y_i-\sum_{1+n/2}^{n}y_i}{\sum_{1}^{n}y_i}\times 100\% \tag{11}$$

式中：S 为偏烧率；n 为图像宽度；y_i 为燃烬火线在 i 处的纵坐标值。S 值越大，表明左右两侧的燃烧不均衡程度越大。

表 2 部分类别平均偏烧率

Tab.2 Average biased combustion rate for selected classes

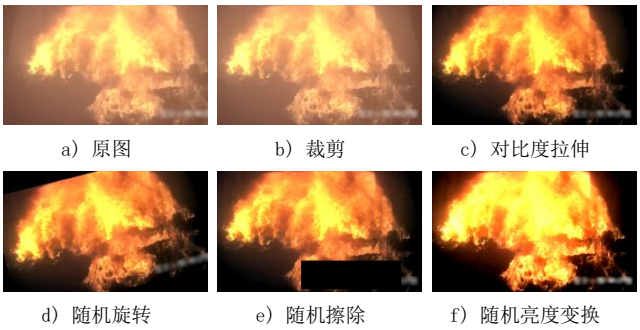
燃烧类别	正常	偏烧	出生料	堆料	堆料-生料
平均偏烧率/%	7.96	18.35	22.94	20.93	23.57

由表 2 可知，正常状态的偏烧率在 10%的范围内，而其余类别偏烧率均超过 15%，且各类别结果与其两侧燃烬火线等级差值成正相关的关系，可有效验证左右燃烧不均衡类别的定义。

2.3 预处理

在垃圾焚烧炉膛中采集的原始图像往往包含炉膛壁、扬尘、火焰强光等干扰信息。因此，需要对火焰数据集进行预处理，旨在去除图像中的噪声干扰与无关信息，同时提升模型应对焚烧炉内复杂工况时的鲁棒性。

首先，裁剪图像中炉膛壁的非火焰区域；其次，基于 RGB 通道对火焰图像进行对比度拉伸（contrast stretching），以凸显火焰燃烧区域；同时，为了防止模型过拟合，对数据集进行了额外的几何、色彩增强，包括随机旋转、随机擦除、随机亮度变换。通过上述预处理操作以提升模型的泛化能力，所得结果如图 7 所示。



a）原图　b）裁剪　c）对比度拉伸

d）随机旋转　e）随机擦除　f）随机亮度变换

图 7 图像预处理

Fig.7 Image preprocessing

2.4 评价指标

为了评估改进模型对复杂火焰图像的识别性能，本文选取精确率（precision）P、召回率（recall）R、整体准确率（overall accuracy）OA、加权平均 F_1 分数（weighted F_1-score）$\overline{F1}$作为改进模型的评价指标。

精确率表示预测某一类别样本时，实际属于该类别的样本所占的比例；召回率表示某一类别的实际样本中，被模型成功识别出来的比例；

$$P=\mathrm{TP}/(\mathrm{TP}+\mathrm{FP}) \tag{12}$$

$$R=\mathrm{TP}/(\mathrm{TP}+\mathrm{FN}) \tag{13}$$

式中：TP 表示指正样本被模型正确分类为正样本的数量；FP 表示负样本被模型错误分类为正样本的数量；FN 表示正样本被模型错误分类为负样本的数量。

总体准确率表示在所有测试样本中，被模型正确分类的样本所占的比例；加权平均 F_1 分数是多分类问题中评估模型整体性能的重要指标，它是所有类别 F_1 分数的加权平均值。

$$\mathrm{OA}=\sum_i^C n_i/N \tag{14}$$

$$F_1=2\mathrm{PR}/(P+R) \tag{15}$$

$$\overline{F1}=\sum_i^C n_i \cdot F1_i/N \tag{16}$$

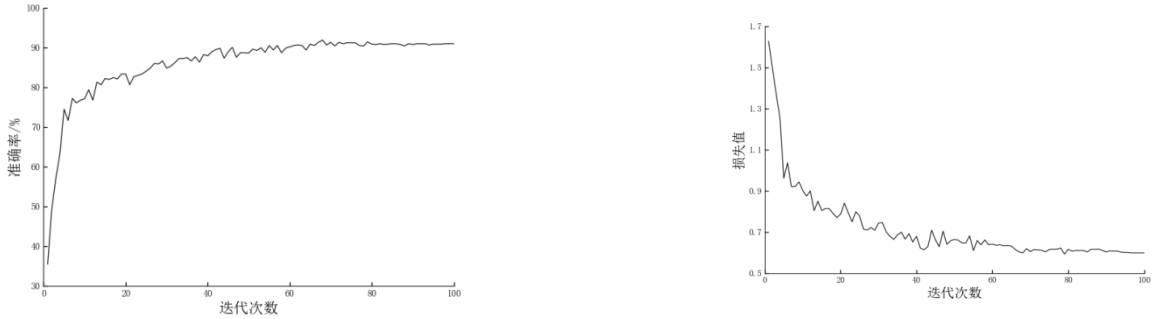
式中：C 为样本类别数；N 为样本总数；n_i 为第 i 个类别的真实样本数量；F_{1i} 为第 i 个类别的 F_1 分数。

同时，为全面衡量模型的效率与实用性，引入参数量、浮点运算数和每秒处理帧数作为模型复杂度与计算效率的评价指标。其中，参数量用于衡量模型对存储空间及显存的占用情况，反映模型的空间复杂度；浮点运算数用于评估模型推理所需的计算资源，反映模型的计算复杂度；FPS 则用于衡量模型处理图像的速度，反映其在实际部署中的实时性。通过对上述指标的综合分析，可以更加全面地反映模型的识别性能，从而验证其可行性。

2.5 实验环境与参数配置

本实验硬件和部分参数如下：硬件方面，CPU 采用 Intel Core i7-13700H，主频率为 2.40GHz，配备 16GB 运行内存。GPU 型号为 NVIDIA Geforce RTX 4060，显存为 8GB，CUDA 版本为为 12.6。实验平台为 Windows11，实验代码基于 Python3.9.23 版本运行。

在模型训练阶段，输入图像统一调整为 224×224 像素，训练批次大小为 16，总共训练 100 轮次。使用 Adam W 优化器[20]，初始学习率设置为 0.001，权重衰减系数为 0.05。在损失函数设计上，针对样本数量不平衡的问题，本实验在训练阶段采用重加权策略，根据各类别样本数量的倒数频率重新分配损失权重，并将其用于交叉熵损失计算，从而使模型更加关注少数类别的学习。同时，采用了标签平滑技术，平滑因子为 0.05，以防止模型过拟合。此外，使用了余弦退火衰减(cosine annealing decay)动态调整学习率。融合上述改进模块后模型训练过程的准确率与损失值曲线如图 8 所示。



a)准确率 b)损失值

图 8 训练结果

Fig.8 Training results

3 结果与分析

3.1 对比实验

为了验证本文提出模型 MSDRNeXt 在垃圾焚烧火焰燃烧状态识别任务中的优越性，本文以 ConvNeXt_Tiny 作为基线模型，选取经典深度学习模型 ResNet50、VGGNet，基于 Vision Transformer 架构的模型如 DeiT、Swin，以及近两年前沿的视觉分类模型 DINOv3、FastViT 进行对比分析。所有对比实验均在相同的软硬件环境下进行，且初始化参数与训练策略保持一致。各模型在测试集上的具体评价指标结果见表 3。

由表 3 可见，早期经典深度学习模型的识别性能相对较低，而近年来的前沿视觉模型整体表现较为优异，但是其普遍引入了多头自注意力机制、动态卷积、稀疏窗口计算等，复杂拓扑结构限制了模型架构灵活度。

相比之下，本文提出 MSDRNeXt 以架构简洁、模块化程度高的 ConvNeXt 网络为基础，该结构更易于结合具体任务进行针对性改进，通过引入双改进模块，在分类准确率和加权平均 F_1 分数上均取得了最佳结果。同时，在保持轻量化的前提下，其对燃烧状态的识别精度仍显著优于其他对比模型。

此外，MSDRNeXt 对单张图像的推理速度达到 278 帧/s，而当前主流垃圾焚烧电厂工业摄像头的采样率通常为 25 帧/s 或 30 帧/s，因此，满足垃圾焚烧炉燃烧状态实时监测的工程部署需求，证实了本文提出模型在火焰状态识别中具有优越性。

表 3 不同模型性能对比
Tab.3 Performance comparison of different model

模型	OA/%	$\overline{F1}$/%	计算量	参数量	速率/(帧·s^{-1})
GoogLeNet[21]	86.95	86.59	2.07G	5.98M	270
VGGNet[22]	85.91	85.72	15.47G	134.31M	158
ResNet50[23]	85.23	84.84	4.13G.	23.53M	242
MobileNet[24]	84.89	84.79	326M	2.22M	323
EfficientNet[25]	87.27	87.10	3.79G	20.19M	93
DeiT-S[26]	88.13	88.09	4.25G	21.59M	318
Swin-T[27]	86.59	86.62	3.91G	18.95M	85
MobileViT_V2[28]	87.16	87.02	1.81G	4.37M	170
DINOv2-S[29]	89.42	89.53	5.51G	21..51M	105
DINOv3-S[30]	89.70	89.85	4.33G	21.56M	139
FastViT[31]	88.66	88.84	0.70G	4.00M	162
ConvNeXt-T	88.30	88.11	4.45G	27.81M	256
MSDRNeXt	91.36	91.32	3.72G	12.95M	278

3.2 消融实验

3.2.1 评价指标消融实验

为了验证本文提出的多尺度融合模块 MSFF 与双区域差分池化模块 DRDP 的有效性，在相同的数据集和参数设置下设计了消融实验，详细结果汇总见表 4。实验 1 设置使用基线模型；实验 2 在实验 1 基础上截取 Stage4；实验 3、实验 4、实验 5 在实验 2 的基础上，并分别引入不同模块。其中，实验 3 与实验 4 分别引入 MSFF 模块、DRDP 模块；实验 5 同时引入 MSFF 模块和 DRDP 模块。

根据表 4 的结果，基线模型的整体准确率和加权 F_1 分数分别为 88.30%、88.11%；在计算复杂度方面，其计算量和参数量分别为 4.45G 浮点运算量和 27.81M。而 MSDRNeXt 的 OA 和$\overline{F1}$相较于基线模型分别提升了 3.06%、3.21%；计算量和参数量分别减少了 16.40%和 53.43%，证明了本文改进模型的有效性。

表 4 消融实验结果对比
Tab.4 Comparison of Ablation Study Results

实验	MSFF	DRDP	Stage4	OA/%	$\overline{F1}$/%	计算量	参数量
1	×	×	✓	88.30	88.11	4.45G	27.81M
2	×	×	×	88.37	88.16	3.70G	**12.34M**
3	✓	×	×	90.34	90.30	3.72G	12.51M
4	×	✓	×	90.11	90.06	**3.70G**	12.78M
5	✓	✓	×	**91.36**	**91.32**	3.72G	12.95M

注：×为不添加模块；✓ 为添加模块。

为进一步明确 MSFF 和 DRDP 模块的单独贡献，同时计算各燃烧状态的精确率和召回率，实验结果见表 5。

表 5 各类燃烧状态识别对比
Tab.5 Comparison of Recognition Performance for Various Combustion States

燃烧类别	实验 2		实验 3		实验 4		实验 5	
	P/%	*R*/%	*P*/%	*R*/%	*P*/%	*R*/%	*P*/%	*R*/%
右偏烧	87.06	78.72	86.21	79.78	86.67	82.98	**88.76**	**84.04**
右出生料	92.50	89.16	96.15	90.37	**96.25**	**92.77**	96.15	90.37
右堆料	92.75	85.33	98.46	85.33	**98.59**	**93.33**	98.57	92.00
堆料-生料	71.23	**98.11**	77.61	**98.11**	78.46	96.23	**82.54**	**98.11**
左偏烧	97.37	84.09	97.37	84.09	**97.44**	**86.36**	**97.44**	**86.36**
左出生料	88.30	96.51	85.71	**97.67**	85.71	**97.67**	**88.42**	**97.67**
左堆料	71.84	**94.87**	74.75	**94.87**	74.75	**94.87**	**77.89**	**94.87**
扬尘	97.06	**100**	**100**	**100**	**100**	**100**	**100**	**100**
正常	89.05	96.06	89.86	97.63	90.30	95.27	**91.91**	**98.42**
火线高	94.12	53.33	95.00	63.33	**97.14**	56.66	95.23	**66.67**
面积大	**100**	87.14	98.48	92.86	98.44	90.00	98.52	**95.71**

分析表 5 中实验 2 和实验 3 结果可以发现，引入 MSFF 模块，模型识别“正常”“火线高”“面积大”类别的准确率明显上升，验证了该模块提取细节纹理特征的能力；其次，分析实验 2 和实验 4 结果，引入 DRDP 模块，模型识别“堆料”“偏烧”“出生料”类别的准确率明显上升，弥补了模型区分火焰水平方向燃烧差异的短板；同时结合两模块，根据实验 5 的结果，MSDRNeXt 在各类别上的识别准确率相较于实验 2 基线模型均有提升，验证了各模块对模型识别能力带来的提升。

3.2.2 纹理特征消融实验

火焰纹理本质上是火焰丰富的细小边缘，可计算其平均梯度(average gradient, AG)来量化纹理活跃程度，该值越大，说明特征图包含的边缘信息越丰富，代表模型对火焰纹理的表征能力越强；反之则说明火焰受炉膛内扬尘、烟雾、堆料等干扰遮挡

越严重，模型提取到的有效纹理信息越少。由于火焰纹理属于浅层网络所提取到的特征，因此，在相同数据集与环境下，设计计算特征图 F_1 平均梯度的消融实验，结果见表 6。

表 6 中，由对比实验 2 与其他实验的结果可知，基线模型融入 MSFF 模块后，在模型训练过程中，对火焰的纹理表征能力更强。此外，对于“扬尘”状态，由于受灰尘遮挡严重，导致火焰区域模糊不清，可提取到的边缘纹理就相对最少；对于“偏烧”“生料”“面积大”状态，其燃烧程度较为剧烈，因此其边缘梯度值偏大，从而直观量化了不同燃烧状态的纹理差异。

表 6 平均梯度结果对比

Tab.6 Average Gradient Comparison

实验	MSFF	DRDP	Stage4	右偏烧	右出生料	右堆料	堆料-生料	左偏烧	左出生料	左堆料	扬尘	正常	火线高	面积大
1	×	×	✓	1.3831	1.4530	1.3437	1.4707	1.3736	1.4259	1.3524	0.8976	1.2480	1.2339	1.5136
2	×	×	×	1.4785	1.5306	1.4489	1.5447	1.4643	1.5059	1.4513	0.9498	1.3543	1.3662	1.5826
3	✓	×	×	1.9840	2.1042	1.9516	2.1297	1.9674	2.0574	1.9492	1.0437	1.7604	1.7642	2.1612
4	×	✓	×	1.6894	1.7562	1.6666	1.7752	1.6798	1.7283	1.6606	1.0356	1.5463	1.5495	1.7817
5	✓	✓	×	1.9966	2.1200	1.9752	2.1558	1.9842	2.0793	1.9656	1.0568	1.7698	1.7815	2.1550

3.3 特征可视化

为进一步从视觉层面验证模型是否从火焰图像中捕捉有效的燃烧特征，本文通过热力图[32]对基线模型与 MSDRNeXt 的特征提取过程进行了可视化分析。热力图以颜色空间分布反映特征关注权重，区域越接近红色表示模型关注度越高，越接近蓝色则表示模型关注度越低，具体结果如图 9 所示。

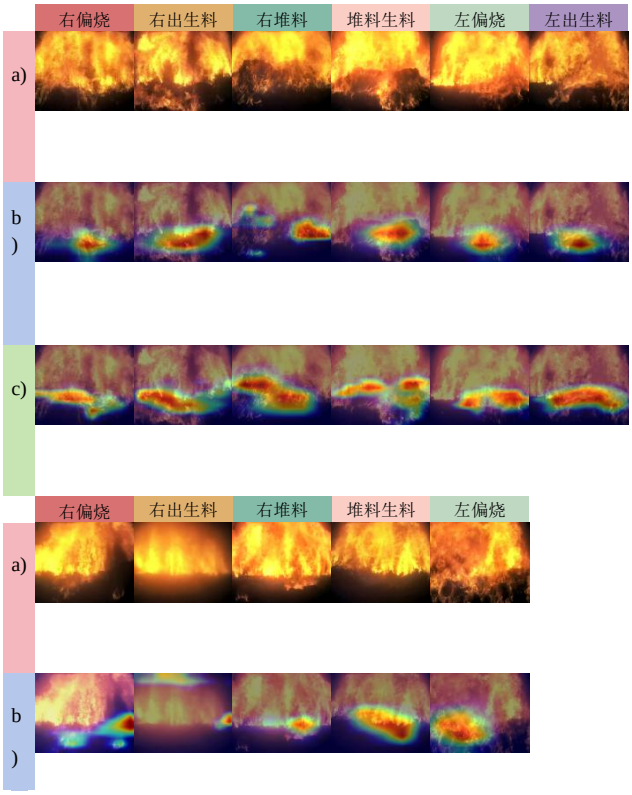


图 9 热力图对比

Fig.9 Heatmap comparison

根据图 9 结果可以发现，基线模型的关注区域较小且具有明显偏差，往往局限于关注小范围燃烧区域；相比之下，MSDRNeXt 的关注面积更大且更精准，对火焰纹理差异、局部燃烧差异、料层堆积部分、燃烬火线位置变化等物理特征地捕获上具有更强的表征能力。这种对核心特征的捕捉能力，使得模型能够有效区分复杂、相似度高的燃烧类别，体现了本文所提出模型在复杂识别任务中具有更优的泛化性能。

3.4 预测结果

为了检验模型在实际火焰图像预测中的具体表现，图 10 展示了基线模型与 MSDRNeXt 对各类别的具体预测概率。由图 10 可见，基线模型预测偏差更大，存在误判的现象，且预测置信度较低；相比之下，MSDRNeXt 的预测更精准，置信度维持在较高水平，充分验证了 MSDRNeXt 的优越性。

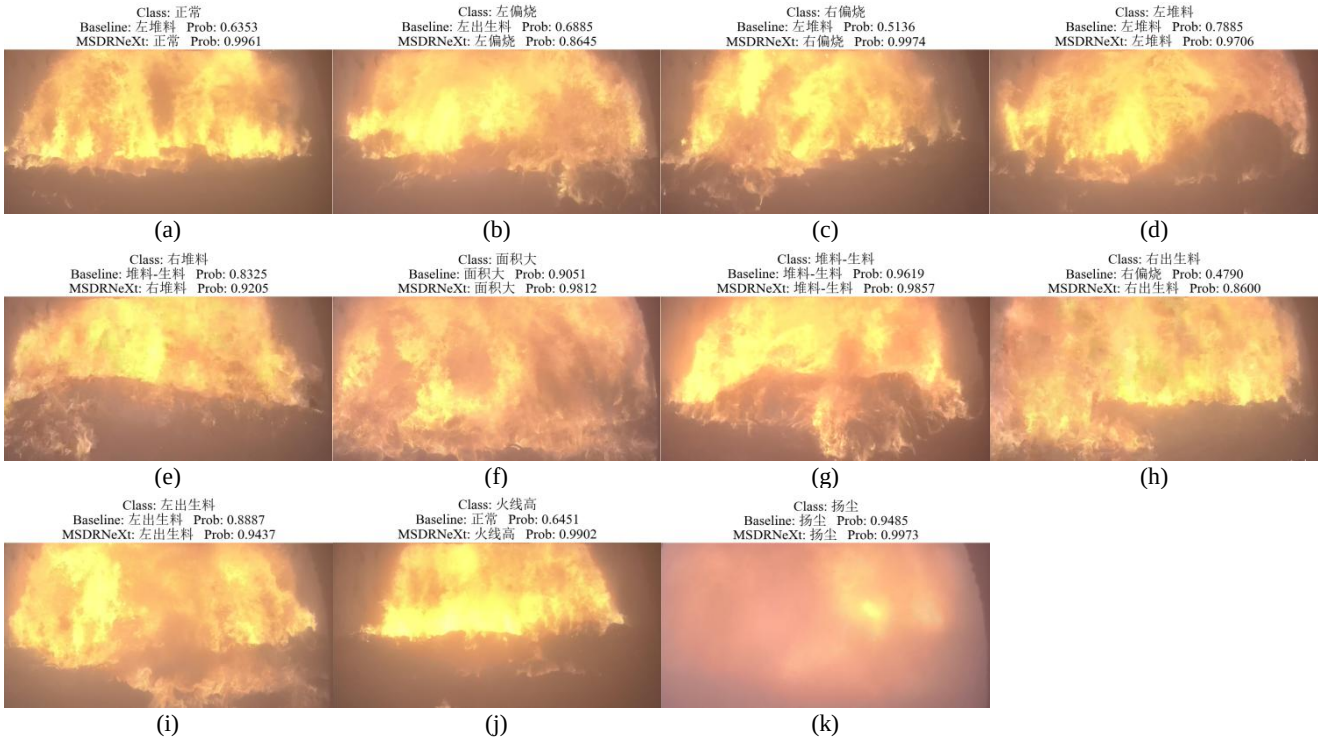


图 10 11 类火焰燃烧状态预测结果对比

Fig. 10 Prediction results comparison of 11 flame combustion state

3.5 分析与讨论

本文基线模型在融入 MSFF 模块与 DRDP 模块后，取得了更高的识别准确率。Liu[12]、Jia[13]、SHI[14]等人的研究表明多尺度特征提取机制能够有效提升模型在光照波动、烟尘遮挡场景下的火焰识别能力，本文提出的 MSFF 模块在多尺度提取的基础上更侧重于跨层级信息的融合，不仅提升了上述复杂场景下模型的识别能力，而且使模型能够更加关注如亮度差异、局部火焰面积等贡献于状态判断的关键特征。

此外，同为垃圾焚烧炉火焰状态识别任务，赵东卓[10]所提出的方法需提取火焰区域面积、像素灰度等统计特征，再结合卷积神经网络进行状态分类。针对炉排左右侧燃烧差异的类别，其通过裁剪图像燃烧不对称的区域来凸显局部差异，不仅效率低，且在火线移动或镜头发生偏移时鲁棒性较差。本文提出的DRDP模块则可有效量化炉排左右两侧燃烧差异，对于偏烧、堆料、出生料类别的识别效果明显提升，同时在炉膛复杂环境下具有较强的鲁棒性。

4 结论与展望

1）针对垃圾焚烧炉炉膛燃烧过程中火焰纹理模糊、受背景干扰强等问题，同时根据炉膛内部左右侧配风的结构特性，提出了一种基于 ConvNeXt 网络的改进模型 MSDRNeXt；通过裁剪无关信息区域、对比度拉伸等操作，有效避免了炉膛内强光和烟雾等背景噪声干扰并突出了火焰区域，实现了垃圾焚烧火焰燃烧状态的准确识别。

2）通过 MSFF 模块提取多尺度的浅层网络特征，再经过注意力机制精炼补充至深层网络层中，增强了模型对火焰纹理的捕捉能力。其次，在特征聚合阶段设计 DRDP 替代 GAP 模块，通过计算特征图左右子区域的特征差分，量化火焰两侧的燃烧差异，并结合 ECA 注意力机制关注表征燃烧强度的通道信息，进一步提升了的整体识别准确率，尤其在堆料、偏烧、出生料等左右燃烧不均衡类别上提升幅度最大。

3）对比实验结果表明，本文模型在垃圾焚烧火焰数据集的状态识别任务中，整体准确率与加权平均 F_1 分数均最高，分别达到了 91.36%和 91.32%，相较基线模型分别提升了 3.06%和 3.21%；此外，模型参数量减少 53.43%，表明了 MSDRNeXt 具有更轻量化的结构以及更强的识别性能，可为垃圾发电的燃烧监测提供技术支持。

4）尽管本文提出的方法在垃圾焚烧火焰分类任务中表现优越，但仍存在一定局限性。首先，本文方法主要针对炉排式焚烧炉这一特定场景，在其他燃烧工艺或场景下模型的泛化能力不足；其次，受限于 3 日的数据采集周期，如炉膛内发生爆燃、闪烁、颜色异常等低频工况无法被完全囊括，导致异常工况的数据集不够完整，一定程度上限制了模型的泛化能力；最后，模型尚未投入现场进行在线测试，在实际运行环境下的稳定性有待进一步验证。后续工作将在不同垃圾发电厂开展部署验证，检验模型的泛化能力。同时，将继续探索该领域自适应、小样本学习等前沿方法，提升模型在新型炉膛结构或数据稀缺场景下的适应能力。

[参 考 文 献]

[1] 李钦钦, 冉根柱, 张鹏. 全国生活垃圾焚烧飞灰处置现状及趋势分析[J]. 中国资源综合利用, 2023, 41(2): 99-102.
LI Qinqin, RAN Genzhu, ZHANG Peng. Analysis on the current situation and trend of domestic waste incineration fly ash disposal in China[J]. China Resources Comprehensive Utilization, 2023, 41(2): 99-102.

[2] 宋建华, 陈亮广, 柯仕忠, 等. 生活垃圾焚烧飞灰处置技术与资源化利用途径综述[J]. 中国资源综合利用, 2024, 42(9): 113-117.
SONG Jianhua, CHEN Liangguang, KE Shizhong, et al. Review of disposal technology and resource utilization of domestic waste incineration fly ash[J]. China Resources Comprehensive Utilization, 2024, 42(9): 113-117.

[3] 汤健, 王天峥, 孙明俊, 等. 城市固废焚烧过程 AI 安全赋能智慧化发展与展望[J/OL]. 控制工程, 1-28[2026-04-03]. DOI:10.14107/j.cnki.kzgc.20250600.
TANG Jian, WANG Tianzheng, SUN Mingjun, et al. Development and prospect of AI security-empowered intelligentization in municipal solid waste incineration process[J/OL]. Control Engineering of China, 1-28[2026-04-03]. DOI:10.14107/j.cnki.kzgc.20250600.

[4] 汤健, 王天峥, 夏恒, 等. 城市固废焚烧智能算法测试与验证模块化半实物平台[J]. 自动化学报, 2024, 50(12): 2432-2461.
TANG Jian, WANG Tianzheng, XIA Heng, et al. Modular hardware-in-the-loop platform for testing and verification of intelligent algorithms in municipal solid waste incineration[J]. Acta Automatica Sinica, 2024, 50(12): 2432-2461.

[5] 张强. 基于计算机视觉的垃圾焚烧特征的提取分析[J]. 计算机与数字工程, 2020, 48(2): 488-491.
ZHANG Qiang. Extraction and analysis of waste incineration characteristics based on computer vision[J]. Computer & Digital Engineering, 2020, 48(2): 488-491.

[6] 葛红, 徐伟程, 闫勇, 等. 基于数字图像处理的生物质混燃火焰稳定性检测方法[J]. 热力发电, 2019, 48(9): 64-70.
GE Hong, XU Weicheng, YAN Yong, et al. Flame stability detection method for biomass co-firing based on digital image processing[J]. Thermal Power Generation, 2019, 48(9): 64-70.

[7] 杨绪兵, 葛彦齐, 张福全, 等. 基于矩阵模式的林火图像半监督学习算法[J]. 图学学报, 2019, 40(5): 835-842.
YANG Xubing, GE Yanqi, ZHANG Fuquan, et al. Forest fire image semi-supervised learning algorithm based on matrix patterns[J]. Journal of Graphics, 2019, 40(5): 835-842.

[8] 宋昱, 吴一全. Log-Gabor 小波和分数阶多项式 KPCA 的火焰图像状态识别[J]. 中国图象图形学报, 2014, 19(12): 1785-1793.
SONG Yu, WU Yiquan. Flame image state recognition based on Log-Gabor wavelet and fractional order polynomial KPCA[J]. Journal of Image and Graphics, 2014, 19(12): 1785-1793.

[9] 张驰, 郭媛, 黎明. 人工神经网络模型发展及应用综述[J]. 计算机工程与应用, 2021, 57(11): 57-69.
ZHANG Chi, GUO Yuan, LI Ming. Review of development and application of artificial neural network models[J]. Computer Engineering and Applications, 2021, 57(11): 57-69.

[10] 赵东卓. 基于机器视觉的垃圾焚烧炉燃烧状态诊断方法研究[D].保定: 华北电力大学, 2025.
ZHAO Dongzhuo. Research on combustion state diagnosis method of waste incinerator based on machine vision[D]. Baoding: North China Electric Power University, 2025.

[11] 白家宇, 沙玲, 魏丹, 等. 基于改进 YOLOv5 的风机叶片表面缺陷检测方法[J]. 热力发电, 2025, 54(9): 95-103.
BAI Jiayu, SHA Ling, WEI Dan, et al. A surface defect detection method for wind turbine blades based on improved YOLOv5[J]. Thermal Power Generation, 2025, 54(9): 95-103.

[12] LIU Z M, ZHANG L, MA H, et al. Research on lightweight tunnel cable fire recognition algorithm based on multi-scale features[J]. Scientific Reports, 2025, 15(1): 25841.

[13] JIA C R. Robust flame detection based on YOLOv5 and bidirectional feature pyramid[C]//2023 IEEE International Conference on Image Processing and Computer Applications. New York: IEEE Press, 2023: 219-225.

[14] SHI X, SHI Z Y, WANG Z, et al. Study on wildfire detection method for transmission corridor based on multi-scale adaptive feature fusion[C]//2025 IEEE 8th Information Technology and Mechatronics Engineering Conference. New York: IEEE Press, 2025: 86-91.

[15] ZOU X. Flame image recognition detection based on improved YOLOv7[C]//2023 IEEE 6th International Conference on Automation, Electronics and Electrical Engineering. New York: IEEE Press, 2023: 956-961.

[16] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]//European Conference on Computer Vision. Cham: Springer, 2018: 3-19.

[17] LIU Z, MAO H, WU C Y, et al. A ConvNet for the 2020s[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2022: 11966-11976.

[18] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2020: 11531-11539.

[19] 卢叶. 基于图像处理的焚烧炉燃烧状态监测方法研究[D].保定: 华北电力大学, 2024.
LU Ye. Research on monitoring method of combustion state of incinerator based on image processing[D]. Baoding: North China Electric Power University, 2024.
[20] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization[C]//The 7th International Conference on Learning Representations. New York: OpenReview.net, 2019: 1-18.
[21] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2015: 1-9.
[22] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//The 3rd International Conference on Learning Representations. San Diego: ICLR, 2015.
[23] 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. New York: IEEE Press, 2016: 770-778.
[24] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17) [2025-01-15]. https://arxiv.org/abs/1704.04861.
[25] TAN M X, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[C]//The 36th International Conference on Machine Learning. New York: PMLR, 2019: 6105-6114.
[26] TOUVRON H, CORD M, DOUZE M, et al. Training data-efficient image transformers & distillation through attention[C]//The 38th International Conference on Machine Learning. New York: PMLR, 2021: 10347-10357.
[27] LIU Z, HU H, LIN Y, et al. Swin Transformer V2: scaling up capacity and resolution[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2022: 12009-12019.
[28] MEHTA S, RASTEGARI M. Separable self-attention for mobile vision transformers[J]. Transactions on Machine Learning Research, 2023.
[29] OQUAB M, DARCET T, MOUTAKANNI T, et al. DINOv2: Learning robust visual features without supervision[EB/OL]. (2023-04-14)[2026-04-21]. https://arxiv.org/abs/2304.07193.
[30] SIMÉONI O, VO H V, SEITZER M, et al. DINOv3[EB/OL]. (2025-08-13)[2026-04-21]. https://arxiv.org/abs/2508.10104.
[31] VASU P K A, GABRIEL J, ZHU J, et al. FastViT: a fast hybrid vision transformer using structural reparameterization[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2023: 5785-5795.
[32] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[J]. International Journal of Computer Vision, 2020, 128(2): 336-359.

（责任编辑　杜亚勤）