

基于CvT模型的高原环境下燃气燃烧火焰图像识别方法研究

周悦¹, 刘敏², 王敬桑¹, 毕德贵^{1*}

¹上海理工大学, 环境与建筑学院, 上海

²淮南师范学院, 电子工程学院, 安徽 淮南

收稿日期: 2025年3月21日; 录用日期: 2025年4月14日; 发布日期: 2025年4月21日

摘要

针对高原低压环境下燃气锅炉燃烧状态监测难题, 本文提出基于CvT模型的火焰图像智能识别方法。在青海西宁的一个1.4 MW燃气锅炉试验台上进行实验, 采集典型工况燃气火焰图像, 提取、分析其火焰图像特征, 在此基础上, 提出CvT-13混合模型, 融合CNN局部特征提取与Transformer全局时序建模优势对火焰状态进行识别, 实现模型特征与图像特征参数的协同分析。结果表明, CvT-13模型可实现高原低压环境下火焰状态的多维度表征, 在背景、稳定及不稳定燃烧状态的分类测试中达到99.51%平均准确率, 其中熄火不稳定状态召回率达99.67%, 背景识别精确度100%。本文为高原低氧环境下的燃烧稳定性实时诊断提供了高精度、强鲁棒性的解决方案。

关键词

高原环境, 燃气燃烧, 火焰图像, 火焰识别, CvT模型

Research on Flame Image Recognition Method for Gas Combustion in Plateau Environment Based on CvT Model

Yue Zhou¹, Min Liu², Jingshen Wang¹, Degui Bi^{1*}

¹School of Environment and Architecture, University of Shanghai for Science and Technology, Shanghai

²School of Electronic Engineering, Huainan Normal University, Huainan Anhui

Received: Mar. 21st, 2025; accepted: Apr. 14th, 2025; published: Apr. 21st, 2025

Abstract

To address the challenges of combustion state monitoring in gas-fired boilers under plateau low-

*通讯作者。

文章引用: 周悦, 刘敏, 王敬桑, 毕德贵. 基于 CvT 模型的高原环境下燃气燃烧火焰图像识别方法研究[J]. 建模与仿真, 2025, 14(4): 607-616. DOI: 10.12677/mos.2025.144313

pressure environments, this study proposes an intelligent flame image recognition method based on a Convolutional vision Transformer (CvT) model. Experimental investigations were conducted on a 1.4 MW gas-fired boiler testbed in Xining, Qinghai Province, where characteristic flame images under typical operating conditions were collected and analyzed. Building upon extracted flame image features, we developed a CvT-13 hybrid model that integrates the advantages of CNN-based local feature extraction and Transformer-based global temporal modeling for flame state identification, enabling synergistic analysis between model characteristics and image feature parameters. Experimental results demonstrate that the CvT-13 model achieves multi-dimensional characterization of flame states in low-pressure plateau environments, attaining 99.51% average accuracy in classifying background, stable combustion, and unstable combustion states. Particularly notable performance includes 99.67% recall rate for flameout-unstable states and 100% precision in background recognition. This research provides a high-precision and robust solution for real-time diagnosis of combustion stability in hypoxic plateau environments.

Keywords

Plateau Environment, Gas Combustion, Flame Images, Flame Recognition, CvT Model

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

天然气因其高热低污的特性大大地推动了燃气锅炉普及,但燃烧失稳易引发未燃气体富集爆炸风险,尤其是在高原地区,低压低氧环境显著影响了燃烧过程[1],这对燃烧状态实时监测与稳定性诊断技术提出更高挑战[2]。研究表明,低压条件下,氧气浓度降低会引发燃烧不完全,增加烟气中 CO 浓度,对高原地区燃气锅炉的燃烧效率和烟气传热性能造成负面影响[3]。有学者针对青海省燃气热水锅炉的能效进行了分析,发现这些锅炉相较于平原热效率更低,无法全负荷运行[4]。现有燃烧诊断技术依赖标准大气参数,其火焰识别模型及热力学计算,在低压环境中适应性不足,难以准确捕捉燃烧状态异变[5]。因此开发适应高原环境的智能监测方法,是突破天然气高效清洁利用瓶颈、保障能源安全的关键需求。

传统燃气锅炉普遍采用接触式传感器来监测火焰稳定性[6],但其易损性、延迟及低灵敏度问题促使非接触式[7]数字图像处理技术快速发展,通过燃烧模态预识别实现高效监测与控制。在图像识别领域,研究者们已发展出各种算法以辨识火焰的特征和状态,卷积神经网络[8] (CNN)基于局部连接架构与参数共享机制优化图像特征提取,能够识别并聚焦火焰图像中的局部细节信息;在自然语言处理(NLP)方面,Transformer 模型[9]已实现显著的研究突破, Vision Transformer (ViT)将 Transformer 架构引入图像分类领域,其模型性能依赖大规模训练数据与高算力支撑,被用于从灰度火焰图像中提取全局特征信息,该模型优化了运行速度[10];卷积视觉变换器(CvT)结合了 CNN 流以捕获彩色图像内各颜色通道的精细局部信息,有助于提升模型与全局信息的互动性;并且 CvT 模型能够自适应地学习火焰图像中的关键特征,减少人工设计特征的工作量[11],模型通过训练自动调整参数,适应不同场景和条件下的火焰图像分类需求。本文融合 CNN 的局部特征提取优势和 Transformer 对全局信息加权的能力[12],创建一种基于 CvT-13 架构的融合 CNN-Transformer 混合模型用于高原火焰识别研究。

本文针对高原低压环境下燃气锅炉燃烧状态识别难题,通过实验的方法,采集大量高原环境下燃气锅炉燃烧火焰图像,对采集火焰图像进行预处理,并提取火焰动静态特征,建立典型工况下火焰图像数

据库,在此基础上,通过构建基于 CvT-13 架构的 CNN-Transformer 混合模型对燃烧过程火焰状态进行识别。本文探讨了一种提高复杂条件下火焰图像燃烧状态分类精度和效率的识别方法,为高原特殊环境下的燃烧过程智能化监控提供参考。

2. 实验系统及方法

2.1. 实验装置

本文的实验系统是青海省西宁市(海拔 2212 米,大气压实测为 76.6 kPa)的一台 1.4 MW 燃气锅炉,实验系统示意图如图 1。实验系统主要包括燃气燃烧系统和图像采集系统两部分,燃烧系统采用低氮扩散式燃烧器、卧式燃气锅炉,空气和燃料在燃烧器前端管道混合之后被喷射进入炉膛进行燃烧过程。图像采集系统由 CCD 相机、计算机及搭载的图像分析、识别算法等组成,相机沿燃烧器轴线方向布置,采集到的火焰图像传输至计算机进行处理。本研究采用 Matlab 程序对采集图像进行预处理及提取特征值,采用了适用于燃气燃烧湍流火焰的中值滤波法[13]及多阈值分割法[14]进行图像预处理。同时,提取火焰总面积、平均灰度、连通度、矩满度、光流速度大小均值和面积比这 6 种动静态图像特征参数。利用这些图像特征参数在 CvT 图像分类模型中对高原燃气锅炉炉膛燃烧图像进行分类训练及测试。

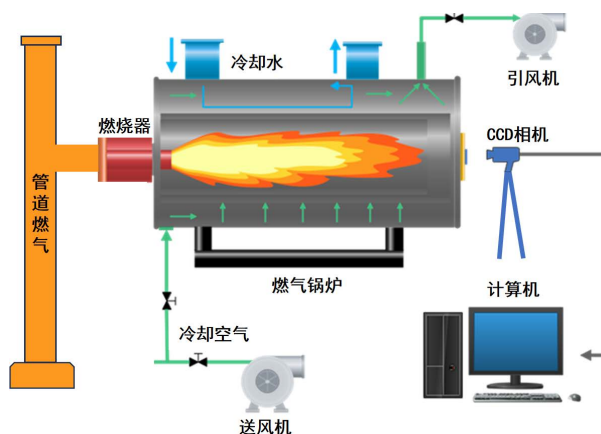


Figure 1. Schematic diagram of the 1.4 MW gas-fired boiler test platform in plateau environment

图 1. 高原 1.4 MW 燃气锅炉实验台示意图

2.2. CvT 模型及参数设置

本文采用 CvT-13 架构模型进行炉膛火焰图像燃烧状态分类。CvT-13 架构模型主要由三个阶段(Stage 1~Stage 3)和一个 MLP Head 组成,其核心设计思想是通过层级式卷积特征嵌入与局部-全局注意力协同机制实现火焰图像的多尺度特征提取,结构如图 2 所示。每个 Stage 包含卷积图像嵌入模块(Conv. Embedding Layer)和卷积变换器模块(Convolutional Transformer Block),其中后者由卷积投影(Conv.Proj)、多头自注意力(MHSA)和多层感知机(MLP)构成。

1) 卷积特征嵌入的渐进式学习

每个 Stage 的卷积嵌入模块通过重叠卷积核(Overlap Convolution)实现图像的空间下采样与通道扩展。以 Stage 1 为例,输入图像($224 \times 224 \times 3$)经 7×7 卷积核(步长 $S = 4$, 填充 $P = 3$)处理后,输出特征图尺寸为 $56 \times 56 \times 64$ 。该过程通过式(1)实现局部纹理特征提取:

$$Z_l = \text{Conv2D}(X_{in}, K = 7, C_{out} = 64, S = 4) \quad (1)$$

其中 X_{in} 为输入张量, C_{out} 为嵌入后的特征图。相较于传统 ViT 的块嵌入(Patch Embedding), 卷积操作通过参数共享保留火焰边缘的连续性特征, 避免硬切割导致的燃烧振荡信息丢失。

2) 卷积-注意力融合机制

在卷积变换器模块中, Conv.Proj 层采用 3×3 深度可分离卷积(Depthwise Separable Conv), 对输入特征进行通道间独立滤波, 计算式如(2):

$$Q/K/V = DWConv(Z_i, K=3, G=C_{in}) \quad (2)$$

其中 $Q/K/V$ 为 Query、Key、Value 矩阵, G 为分组卷积的组数, 该设计使 MHSA 的注意力计算聚焦于火焰局部区域内的响应关联性, 有效抑制高原低气压环境下的通道间干扰噪声。

3) 多阶段特征抽象增强

各 Stage 通过动态调整注意力头数(H)与嵌入维度(D)实现特征表达的渐进优化:

Stage1 ($H=1, D=64$): 侧重燃烧核心区的辐射强度分布, MHSA 计算窗口覆盖 56×56 像素区域, 捕获火焰根部脉动频率特征;

Stage2 ($H=2, D=192$): 通过 2 个注意力头学习火焰发光区域的时序特征;

Stage3 ($H=4, D=384$): 采用 4 头注意力机制建立火焰整体形态, 例如高度/摆动幅度与燃烧稳定性的全局映射关系。

最终, Stage3 输出的 $14 \times 14 \times 384$ 特征图经全局平均池化(GAP)压缩为 384 维向量, 通过 MLP Head 中的线性层(含 Dropout = 0.3)映射至燃烧状态类别概率空间, 完成不稳定燃烧、稳定燃烧的分类任务。

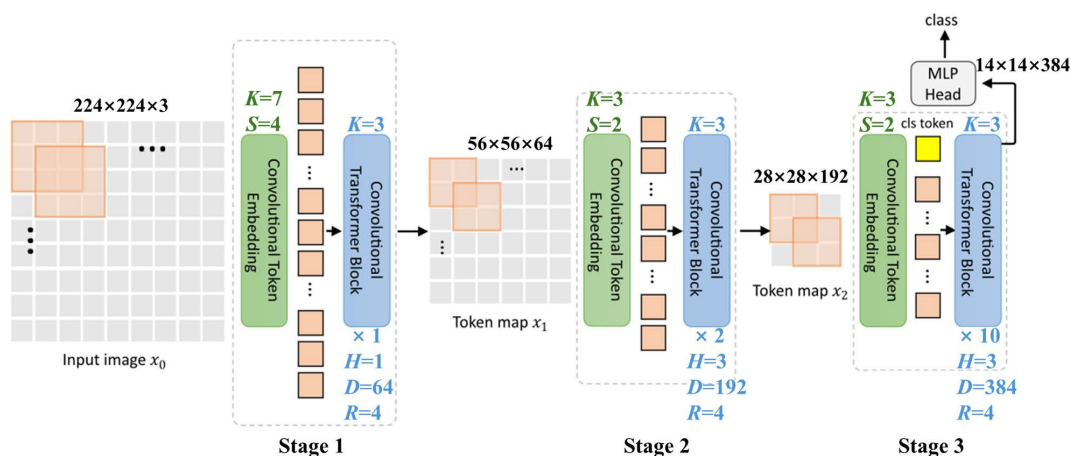


Figure 2. Structure of the CvT-13 model architecture [15]

图 2. CvT-13 模型结构[15]

本文的实验环境包括硬件训练平台: 16G 12th Gen Intel(R) Core(TM) i7-12700F。软件平台包括: 操作系统 Windows 11 专业版 23H2, 使用 Anaconda3 管理 Python 环境, 并采用 PyTorch 1.8.0 作为深度学习框架实现 CvT-13 网络模型的训练。训练参数设置如下, 初始学习率为 0.001, 在第 20 个 epoch 减少学习率, 学习率衰减比率为 0.1, 使用 AdamW 作为优化器, 批次大小为 32 个样本, 训练总周期数为 30, 权重衰减设置为 $5e-4$ 。

经过图像预处理操作后火焰图像数据集总共有 7190 张图像, 其中背景类有 300 张、点火不稳定类有 995 张、稳定类有 5745 张、熄火不稳定类有 150 张, 将预先标定好的已知类别火焰图像打散, 按照 4:1 的比例将所有图像数据分为训练集和测试集之后进行训练和测试, 具体的数据集划分如表 1 所示。

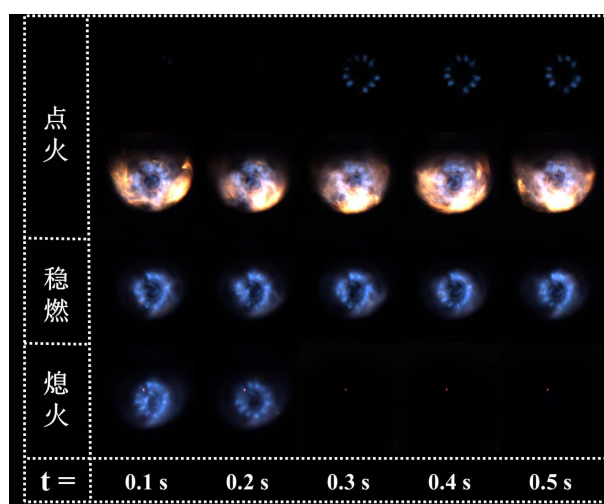
Table 1. Training and testing datasets of gas flame images**表 1.** 燃气火焰图像训练及测试数据集

数据集	燃烧状态	图像数	数据集图像总数
训练集	背景	240	5752
	点火不稳定	796	
	稳定	4596	
	熄火不稳定	120	
测试集	背景	60	1438
	点火不稳定	199	
	稳定	1149	
	熄火不稳定	30	

3. 燃气结果分析与讨论

3.1. 高原环境下燃气火焰图像特征分析

本文选取了高原地区典型燃烧工况 85% 负荷下燃气燃烧情况进行分析。如图 3 所示，展示了 85% 负荷下燃气火焰点火 - 稳燃 - 熄火连续变化过程。点火初期，火焰集中于燃烧器喷口附近，呈点状排列，颜色为淡蓝色，亮度较低，形态清晰可见。进入不稳定燃烧阶段，火焰前端(观火孔侧)逐渐扩展至炉内大部分区域，颜色转为橘红色，表明末端空气不足导致不完全燃烧，此时橘黄色火焰遮挡画面，但喷口附近仍维持蓝色火焰，反映该区域空气供给充足。稳定燃烧阶段，残留燃气消耗完毕，火焰以蓝色为主，形态与亮度趋于均匀，燃烧状态平稳。熄火过程因锅炉保护机制触发燃烧器停机，火焰迅速熄灭(仅持续 1~3 帧，0.1~0.3 s)。该过程揭示了实际燃烧中空气分布不均对火焰稳定性的影响，以及点火阶段动态不稳定性与稳定阶段的差异特征。

**Figure 3.** Ignition-Stable combustion-flameout process images at 85% load**图 3.** 85% 负荷点火 - 稳燃 - 熄火过程图像

对上述工况下燃烧火焰采集大量火焰图像，并提取其主要特征值进行分析，结果如图 4 所示。

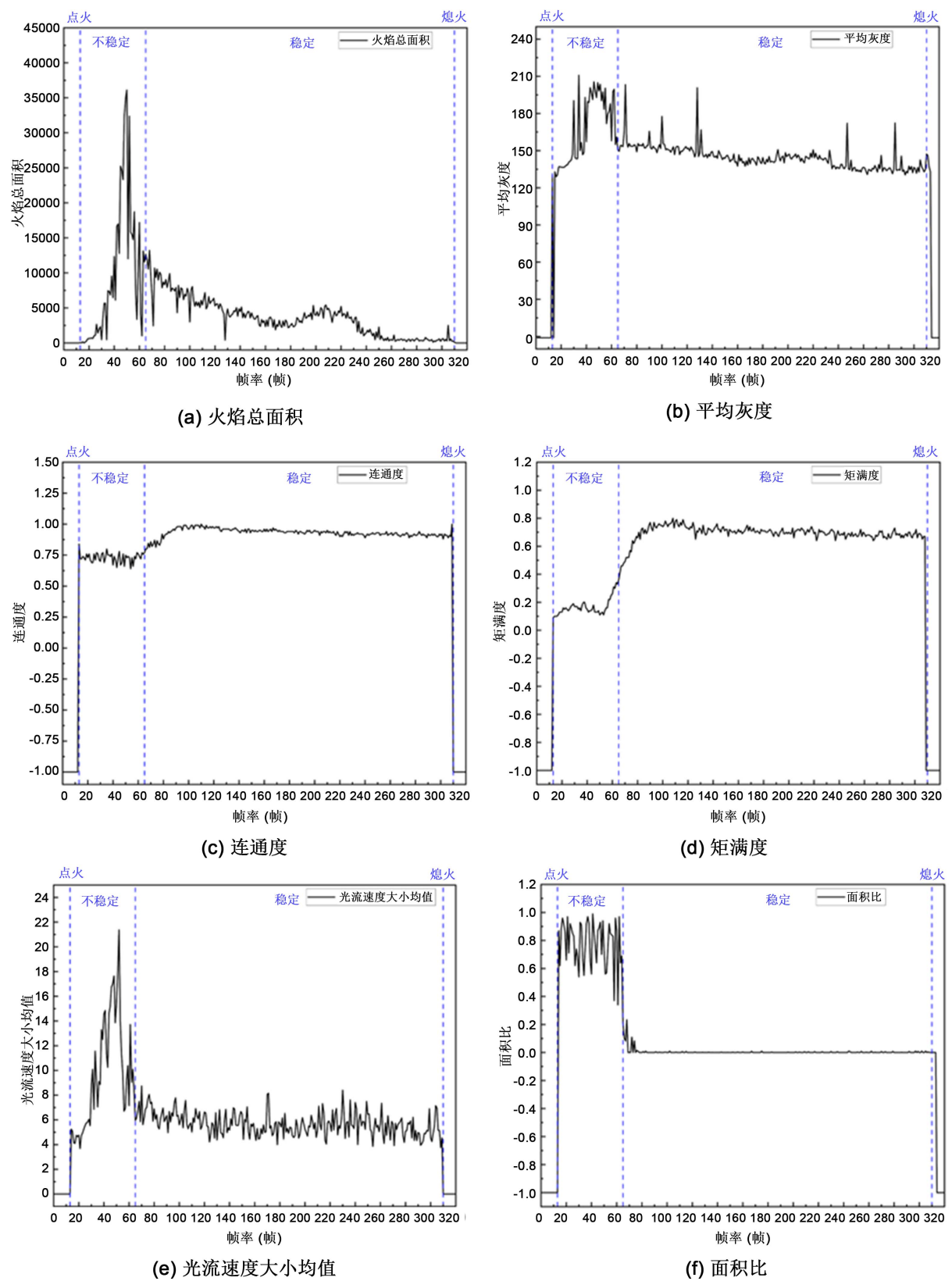


Figure 4. Trends of image feature parameters during ignition-stable combustion-flameout process at 85% load
图 4. 85%负荷点火 - 稳燃 - 熄火过程图像特征参数变化规律

85%负荷下燃气锅炉火焰图像特征参数随时间变化趋势显示,在点火阶段的不稳定燃烧过程中,火焰总面积、平均灰度和速度大小均值均在短时间内出现大幅度的数值波动,呈现先升高后降低的变化过程,这与观察到的炉膛火焰在点火阶段的不稳定燃烧图像变化情况一致。连通度在 0.75 附近波动,矩满度则在 0.1 附近波动,面积比则在 0.6~0.9 波动,这种现象的原因在于拍摄时燃烧器喷口在初始点火阶段最先出现点状火焰,火焰分散程度高,局部燃烧不均匀,每一个喷口都存在火焰,这对面积比、连通度矩满度产生了一定影响,反映了实际炉膛火焰监测过程中的复杂性和难度。

稳定燃烧阶段,火焰总面积、平均灰度和光流速度大小均值随着负荷趋于稳定而逐渐变得平稳,但仍有一定幅度波动,这主要是由炉膛燃烧时的工况波动引起的,表明这些参数对稳定燃烧过程中运行情况具备有效监测能力。而在不稳定阶段这些特征均出现了对于稳定燃烧阶段明显的差异性和大幅度的波动,这很好地地区分了火焰稳定燃烧和不稳定燃烧状态,说明本文提取的图像特征参数能够很好地反应高原燃气锅炉炉膛火焰状态。

3.2. 基于 CvT-13 模型的火焰图像多层次特征提取与可视化分析

本文进一步基于 CvT-13 模型通过采用卷积图像嵌入模块和卷积变换器模型可以学习到燃气锅炉炉膛火焰图像的火焰轮廓、亮度模式和纹理信息等关键特征,这些特征构成了火焰状态的视觉指标。结果如图 5 所示,从图中可以看出,通过可视化 Stage1 的特征层输出尺寸为 $56 \times 56 \times 64$ 特征图像,可以洞察模型是如何捕捉到火焰图像的基础层面信息的:特征层图像捕捉火焰轮廓的扩散范围直接对应图像特征参数中火焰总面积的动态变化;不同通道对颜色敏感的卷积核,例如偏向红色或蓝色的激活区域,能够区分灰度差异,可以与图像特征参数中的平均灰度作联合分析;离散火焰区域的边界分割(如图中分散的亮斑)直接反映连通度和面积比的计算逻辑。不同通道的特征映射层展示了 CvT-13 模型在早期阶段提取的低层次特征,这为后续层做出更精细的燃烧分析提供了基础。

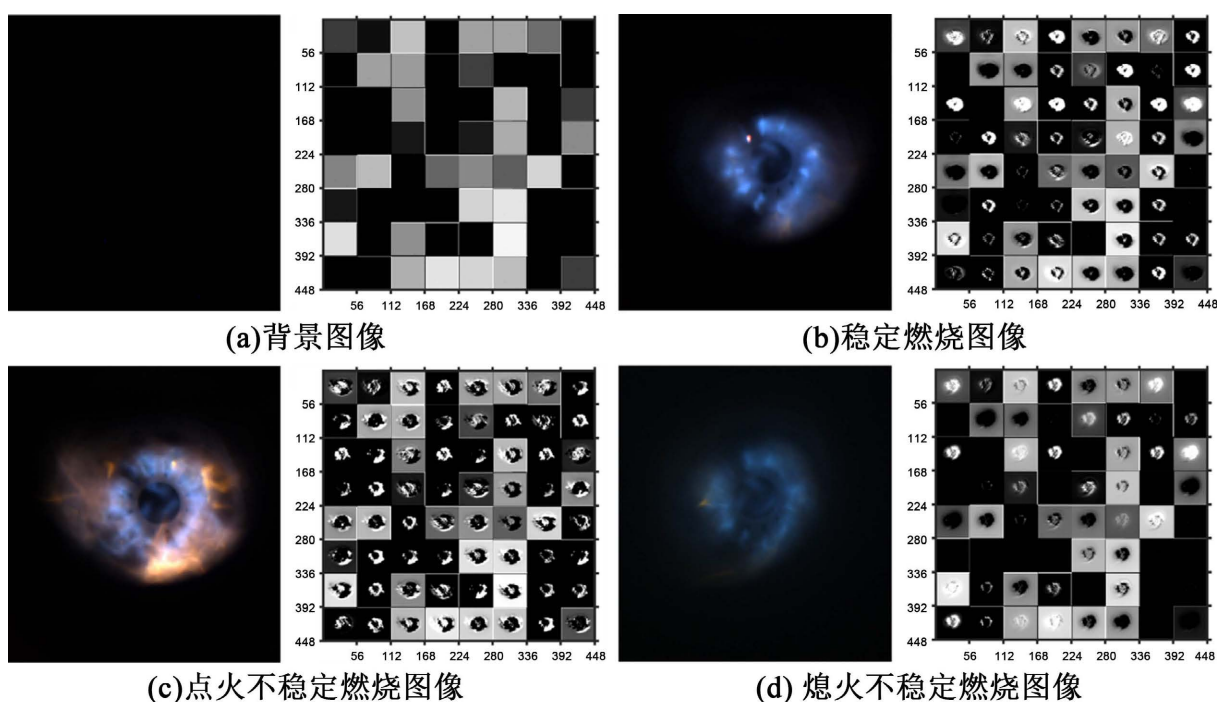


Figure 5. Raw flame images and visualized Stage 1 output feature layers under different combustion states

图 5. 不同燃烧状态下的火焰原始图像和可视化 Stage 1 输出特征层图像

还可以看到，CvT-13 模型深层部分的卷积变换器模块能够识别和表达更为复杂的火焰特性，诸如火焰结构、稳定性以及火焰锋面形态等深层特征：模型通过深层纹理特征解释灰度的局部不均匀性；通过结构特征解释连通度波动；全局注意力机制整合火焰锋面移动与结构稳定性，间接支持了光流速度分析。这些高级特征的提取表明 CvT-13 模型结合传统的 CNN 方法具有更强大的性能，它不仅捕获了局部特征，而且可以实现关联分析，将图像特征参数的实验数据与模型特征形成闭环验证，两者共同验证燃烧状态的动态演变，加强了模型在捕捉火焰复杂行为方面的能力。

3.3. CvT-13 模型在火焰状态识别中的高精度分类

经过训练，在测试集上验证了 CvT-13 模型的识别效果，结果如表 2 和图 6 所示。CvT-13 模型的召回率在所有类别上都超过了 98%，特别是对于稳定和熄火不稳定状态，召回率达到了 99.74%和 99.67%，这意味着模型能够非常有效地识别出大部分正类样本，有非常小的遗漏率。精确度也非常高，特别是对于背景状态，精确度达到了完美的 100%。然而其他模型的精确度略低，在 98.99%至 99.67%之间。这样的精确度表明很少有被错误分类到这些类别的样本。F1 分数结合了精确度和召回率，提供了一个关于模型精度和健壮性的单一度量。CvT-13 模型在所有类别的 F1 分数都非常接近 1，这表明了召回率和精确度之间非常好的平衡。最后，CvT-13 模型平均准确率为 99.51%，这是一个出色的成绩，表示 CvT-13 模型能够非常准确地为绝大多数样本分类。

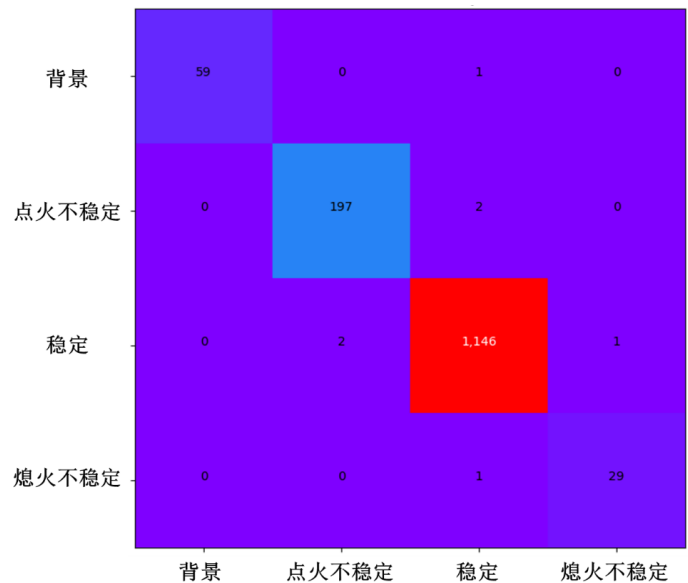


Figure 6. Figure 6 confusion matrix results of the CvT-13 model on the test set
图 6. CvT-13 模型测试集混淆矩阵结果

基于 CvT-13 模型的炉膛火焰图像燃烧状态识别效果较好。模型在所有类别(即“背景”、“点火不稳定”、“稳定”和“熄火不稳定”)上都达到了比较高的性能，准确率为 99.51%。这表明 CvT-13 模型能够非常可靠和稳定地识别火焰状态，这对 CvT-13 模型的在高原低气压条件下的燃气锅炉火焰稳定性检测与识别具有重要意义，对 CvT-13 模型在实际工业应用十分关键。

3.4. 不同模块对 CvT-13 模型贡献的消融实验验证

上述识别结果表明，CvT-13 模型在高原火焰状态识别任务中展现出显著的分类性能优势。为深入解

析模型架构中各组件的协同作用机制，本研究进一步通过消融实验对卷积特征提取、全局注意力及特征融合模块进行系统性验证。

Table 2. Flame state recognition results of the CvT-13 model

表 2. CvT-13 模型火焰状态识别结果

类别	召回率	精确度	F1 分数	准确率
背景	98.33%	100.00%	0.9916	99.51 %
点火不稳定	98.99%	98.99%	0.9899	
稳定	99.74%	99.65%	0.9969	
熄火不稳定	99.67%	99.67%	0.9667	

通过构建消融实验验证模型架构合理性，实验结果见表 3，当移除卷积特征提取模块时，测试准确率下降 12.7%，证明局部特征捕获对低压燃烧诊断的重要性；而禁用 Transformer 的全局注意力机制导致不稳定状态召回率降低 9.2%，表明时序建模能力对燃烧动态过程识别起关键作用。

Table 3. Comparison of ablation study results

表 3. 消融实验结果对比

模型变体	召回率	精确度	F1 分数
CvT-13 完整模型	99.67	99.51	0.9967
移除卷积模块	88.23 ↓	86.81 ↓	0.8642 ↓
禁用 Transformer 注意力	90.47 ↓	93.45 ↓	0.9241 ↓
无特征融合机制	92.15 ↓	95.12 ↓	0.9356 ↓

4. 结果与展望

本研究基于 CvT-13 架构的 CNN-Transformer 混合模型，利用图像特征参数和 CvT 图像分类模型对高原炉膛燃气燃烧过程进行图像特征分析和火焰稳定性状态识别，得出以下结论：

1) 提取的火焰动静态特征参数在 76.6kPa 低压条件下呈现显著响应特性，不稳定阶段出现了对于稳定燃烧阶段明显的差异性和大幅度的波动，很好地描述点火阶段的火焰不稳定状态，说明这些图像特征参数在对高原燃气锅炉炉膛火焰稳定性监测方面具有良好的效果。

2) CvT-13 模型成功融合 CNN 局部特征提取与 Transformer 全局建模优势，以图像特征参数为量化指标，模型可视化特征为解释性工具，两者共同验证燃烧状态的动态演变，实现高原低压环境下火焰状态的多维度表征。

3) 模型在测试集达到 99.51% 的平均准确率，实现了混合架构对高原低压燃烧状态的高精度识别，在高原 1.4 MW 燃气锅炉中验证了其较好的鲁棒性。

实验结果揭示模型在高原燃烧诊断中的独特优势：① 多尺度特征融合机制可同步捕捉火焰微结构变化与动态过程；② 注意力权重分布与燃烧不稳定性理论高度契合，验证了物理引导的机器学习有效性。但研究仍存在以下局限：当前数据集仅覆盖一种 76.6 kPa 固定气压条件，未来需构建包含连续海拔梯度 (2000~5000 m) 的扩展数据集；模型参数量(21.7 M)对边缘设备部署构成挑战，后续将通过神经架构搜索进行轻量化改进。

基金项目

本研究由国家重点研发计划资助(2021YFF0600601)。

参考文献

- [1] 杜勇博, 张井坤, 笪耀东, 等. 高原锅炉燃料燃烧和烟气特性的研究与进展[J]. 工业锅炉, 2020(5): 1-6.
- [2] Sai, Q., Zhao, J., Bi, D., Qin, B. and Meng, L. (2024) Improved Yolov8 for Gas-Flame State Recognition under Low-Pressure Conditions. *Sensors*, **24**, Article 6383. <https://doi.org/10.3390/s24196383>
- [3] 李晓红, 李通征, 陶瑞鑫. 基于 Fluent 的青海高原燃气锅炉燃烧特性数值研究[J]. 节能, 2019, 38(10): 65-69.
- [4] 于吉明, 笪耀东, 张井坤, 等. 高原 WNS 型燃气热水锅炉能效测试与数据分析[J]. 中国特种设备安全, 2022, 38(12): 48-51+62.
- [5] Zhang, J., Ji, W., Yuan, Y., Sun, Q. and Tang, F. (2025) Effect of Low Atmospheric Pressure on Combustion Characteristics of Polypropylene: An Experimental Investigation. *Applied Thermal Engineering*, **265**, Article 125545. <https://doi.org/10.1016/j.applthermaleng.2025.125545>
- [6] 舒文. 基于颜色厚度的预混碳氢火焰当量比软测量模型的研究[D]: [硕士学位论文]. 贵阳: 贵州大学, 2023.
- [7] 蔡李靖, 蔡伟伟, 施展, 等. 燃烧诊断中的光学信号采集方法[J]. 信号处理, 2023, 39(12): 2091-2114.
- [8] 刘鸿达, 孙旭辉, 李沂滨, 等. 基于卷积神经网络的图像分类深度学习模型综述[J]. 计算机工程与应用, 2025(3): 1-29.
- [9] 杜丽霞, 梁晓萍. 基于 Swin Transformer 和 PJF 颜色空间的火灾检测算法[J]. 现代电子技术, 2025, 48(5): 147-152.
- [10] 梁秀满, 安金铭, 曹晓华, 等. 基于 CNN-Transformer 双流网络的烧结火焰燃烧状态分类[J]. 应用光学, 2023, 44(5): 1030-1036.
- [11] 陈浩霖, 高尚兵, 相林, 等. FIRE-DET: 一种高效的火焰检测模型[J]. 南京信息工程大学学报, 2023, 15(1): 76-84.
- [12] 胡瀚淳. 融合卷积神经网络和 Transformer 架构的遥感影像自监督学习方法[D]: [硕士学位论文]. 武汉: 华中师范大学, 2024.
- [13] 刘光宇, 曹禹, 王帅, 等. 基于自适应中值滤波的图像去噪技术研究[J]. 安徽电子信息职业技术学院学报, 2022, 21(5): 1-6.
- [14] 吴一全, 宋昱, 周怀春. 基于灰度熵多阈值分割和 SVM 的火焰图像状态识别[J]. 中国电机工程学报, 2013, 33(20): 66-73+13.
- [15] Wu, H., Xiao, B., Codella, N., Liu, M., Dai, X., Yuan, L., *et al.* (2021) CvT: Introducing Convolutions to Vision Transformers. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 22-31. <https://doi.org/10.1109/iccv48922.2021.00009>