

## 基于CNN-Transformer双流网络的烧结火焰燃烧状态分类

梁秀满 安金铭 曹晓华 曾凯 王福斌 刘贺飞

### Classification of combustion state of sintering flame based on CNN-Transformer dual-stream network

LIANG Xiuman, AN Jinming, CAO Xiaohua, ZENG Kai, WANG Fubin, LIU Hefei

引用本文:

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. DOI: 10.5768/JAO202344.0502003

LIANG Xiuman, AN Jinming, CAO Xiaohua, et al. Classification of combustion state of sintering flame based on CNN-Transformer dual-stream network[J]. Journal of Applied Optics, 2023, 44(5): 1030–1036. DOI: 10.5768/JAO202344.0502003

在线阅读 View online: <https://doi.org/10.5768/JAO202344.0502003>

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 基于卷积神经网络的焊缝缺陷图像分类研究

Research on weld defect image classification based on convolutional neural network

应用光学. 2020, 41(3): 531–537 <https://doi.org/10.5768/JAO202041.0302007>

#### 基于时空双流卷积神经网络的红外行为识别

Infrared behavior recognition based on spatio-temporal two-stream convolutional neural networks

应用光学. 2018, 39(5): 743–750 <https://doi.org/10.5768/JAO201839.0506002>

#### 基于卷积神经网络的蛋胚活性精准检测方法研究

Research on accurate detection method of egg embryo activity based on convolutional neural network

应用光学. 2021, 42(2): 268–275 <https://doi.org/10.5768/JAO202142.0202003>

#### 基于小波变换与卷积神经网络的图像去噪算法

Image denoising algorithm based on wavelet transform and convolutional neural network

应用光学. 2020, 41(2): 288–295 <https://doi.org/10.5768/JAO202041.0202001>

#### 基于全卷积神经网络的图像去雾算法

Image defogging algorithm combined with full convolution neural network

应用光学. 2019, 40(4): 596–602 <https://doi.org/10.5768/JAO201940.0402003>

#### 基于深度神经网络的太阳能电池组件缺陷检测算法研究

Research on detection algorithm of solar cell component defects based on deep neural network

应用光学. 2020, 41(2): 327–336 <https://doi.org/10.5768/JAO202041.0202006>



关注微信公众号, 获得更多资讯信息

文章编号: 1002-2082 (2023) 05-1030-07

# 基于 CNN-Transformer 双流网络的烧结火焰燃烧状态分类

梁秀满, 安金铭, 曹晓华, 曾 凯, 王福斌, 刘贺飞

(华北理工大学 电气工程学院, 河北 唐山 063210)

**摘 要:** 烧结火焰图像中具有细粒度的局部火焰状态特征信息和复杂多变的全局火焰状态特征信息, 而传统的卷积神经网络往往对局部特征更加敏感, 难以提取火焰状态的全局特征信息, 从而制约烧结火焰特征的表达能力, 导致烧结火焰状态分类识别精度低。针对此类问题, 提出一种基于 CNN-Transformer 的双流网络特征融合分类方法, 该方法包含 CNN (convolutional neural networks) 流和 Transformer 流两个模块, 首先将 CNN 块和 Transformer 块并行设计, CNN 流提取烧结火焰 RGB 图像中的局部特征信息, Transformer 流提取烧结火焰 GRAY 图像的全局特征信息; 然后, 将双流网络分别提取的烧结火焰状态局部特征信息和全局特征信息进行级联交互特征融合; 最后, 利用 softmax 分类器实现烧结火焰状态的分类。实验结果表明, 火焰分类准确率可达 96.20%, 与传统卷积神经网络相比提升 6%~8% 的识别准确率。

**关键词:** 卷积神经网络; Transformer; 特征融合; 烧结火焰; 图像分类

中图分类号: TN911.73; TP391.41

文献标志码: A

DOI: 10.5768/JAO202344.0502003

## Classification of combustion state of sintering flame based on CNN-Transformer dual-stream network

LIANG Xiuman, AN Jinming, CAO Xiaohua, ZENG Kai, WANG Fubin, LIU Hefei

(College of Electrical Engineering, North China University of Science and Technology,  
Tangshan 063210, China)

**Abstract:** The sintering flame image has fine-grained local flame state feature information and complex global flame state feature information. However, the traditional convolutional neural network is often more sensitive to local features, and it is difficult to extract the global feature information of the flame state, which restricts the expression ability of sintering flame features, resulting in low accuracy in the classification and recognition of the sintering flame state. In response to such problems, a dual-stream network feature fusion classification method based on CNN-Transformer was proposed, which includes two modules: convolutional neural networks (CNN) flow and Transformer flow. Firstly, the CNN block and the Transformer block were designed in parallel. The CNN stream extracts the local feature information of the RGB image of the sintering flame, and the Transformer stream extracts the global feature information of the GRAY image of the sintering flame. Then, the local feature information and the global feature information of the sintering flame state extracted by the dual-stream network was fused using the cascade interactive feature fusion method. Finally, the softmax classifier was used to achieve the classification of sintering flame states. The experimental results show that the flame classification accuracy can reach 96.20%, which is 6%~8% higher than that of the traditional

收稿日期: 2022-09-14; 修回日期: 2022-11-10

基金项目: 教育部产学研协同育人项目 (202002133014); 河北省自然科学基金高端钢铁联合研究基金 (F2019209323)

作者简介: 梁秀满 (1973—), 女, 副教授, 硕士生导师, 主要从事检测技术与智能装置研究。E-mail: 13230857556@126.com

通信作者: 曾凯 (1990—), 男, 讲师, 博士研究生, 主要从事机器视觉研究。E-mail: kevinzengkai@126.com

convolutional neural network.

**Key words:** convolutional neural network; Transformer; feature fusion; sintering flame; image classification

## 引言

优质的铁矿石不仅能节约能源、降低生产成本,而且能减少对环境的污染,烧结作为炼铁过程中的一个重要环节,工艺过程十分复杂,包括焦炭燃料破碎、矿物配料、加水制粒、烧结及筛分等多个环节<sup>[1]</sup>。烧结终点是评价烧结成品品质的重要依据,是与烧结矿的产量、成本和能源消耗相关的重要工艺参数。烧结终点的判断依据,主要分为两类:1) 分析烧结工艺影响烧结终点的相关参数,包括烧结机台车速度、点火温度、煤气压力、煤气流量、负压、终点温度等,由于工艺的复杂性、计算的偏差,因此通用性不强;2) 基于机器视觉的烧结火焰图像状态分类,通过图像处理分析料层与烧结终点的位置关系,该方法具有节省看火工人力资源、自动化程度高等优点。烧结断面火焰图像的燃烧状态能有效判断烧结终点的位置,从而提高优质铁矿的产出率<sup>[2]</sup>。因此,准确识别烧结火焰的不同燃烧状态对优化钢铁产业及提高能源利用率都具有重要意义。

卷积神经网络在图像分类领域占据着重要的地位,以强大的图像特征提取能力被广泛应用于各种图像分类任务,取得了十分显著的效果。近几年,深度学习已逐步应用在钢铁产业,包括:应用图像分类技术的烧结看火、应用目标检测技术的板坯表面质量检测、应用图像分割技术的带式运输机皮带跑偏检测以及铁水包扒渣识别等<sup>[3]</sup>。2012年, AlexNet 网络<sup>[4]</sup>的出现标志着卷积神经网络逐步走向深度学习。ResNet 网络<sup>[5]</sup>解决了卷积神经网络随着卷积层的加深出现的梯度消失、梯度爆炸等网络退化问题,但由于参数过多导致模型训练速度不理想。MobileNet 网络<sup>[6]</sup>中的深度可分离网络能够在不影响识别准确率的情况下提升模型运行速度。卷积神经网络善于提取局部信息,然而,烧结火焰图像的大面积火焰区域依赖于全局特征信息,采用卷积神经网络进行图像分类,使得火焰识别任务的效率低,准确率难以提升。近年来,Transformer 在自然语言处理取得了重大的研究进展,其在图像分类也得到了广泛的应用。ViT 网络<sup>[7]</sup>将 Transformer 应用在图像分类任务上,但无法建模一张图片的局部信息。

通过上述分析,本研究结合 CNN 网络善于提取局部特征和 Transformer 更加聚焦于全局特征信息的特点,构建了一种基于 CNN-Transformer 双流网络分类模型,以解决烧结火焰利用单流网络结构提取特征信息不全面的问题,从而提高烧结火焰图像分类识别精度。

## 1 模型架构设计

卷积神经网络利用卷积核对图像的像素进行卷积运算,然后得到特征映射矩阵,提取相应位置的局部特征,虽然可以通过扩大感受野捕捉全局特征,但这可能需要更密集但具有破坏性的池化操作,例如 RseNet-18<sup>[5]</sup>由于中间池化层缺乏细粒度的全局上下文建模,难以获取烧结火焰全局图像信息,从而影响分类识别精度。Transformer 将图像分成固定大小的 patch,然后通过线性变换得到 patch embedding,最后使用注意力机制,编码图像特征矩阵的任意两个位置反映了复杂的空间变换和长距离特征依赖性,构成了全局表示,但受限于计算复杂度,Transformer 需要减小输入分辨率、增大下采样步长,这造成切分 patch 阶段损失图像细节信息,忽略了局部特征细节,降低了背景和前景之间的可辨别性。因此,本文为了提高火焰识别的准确率,在传统卷积神经网络引入 Transformer 结构,构建 Transformer 全局特征流,与 CNN 流局部特征流进行特征融合建模双流网络。

根据烧结断面火焰图像的特点, RGB(red-green-blue) 彩色火焰图像具有更加细粒度的局部特征信息,利用 CNN 流提取彩色图像中更多颜色通道的细粒度局部信息,能够增强与全局信息的交互性。烧结火焰状态分类同时依赖于火焰的全局表观特征,利用 Transorformer 流提取灰度火焰图像中的全局特征信息,在满足全局表观特征信息的需求下,加快了模型运行速度。然后,利用级联交互特征融合模块,将 CNN 流提取的局部特征与 Transformer 流提取的全局特征进行融合,得到具有全局性且细粒度的烧结火焰特征信息。整体网络模型如图 1 所示,该网络模型包括:基于 ResNet-18 的 CNN 模块、基于 ViT 的 Transformer 模块、特征融合和 softmax 损失函数。

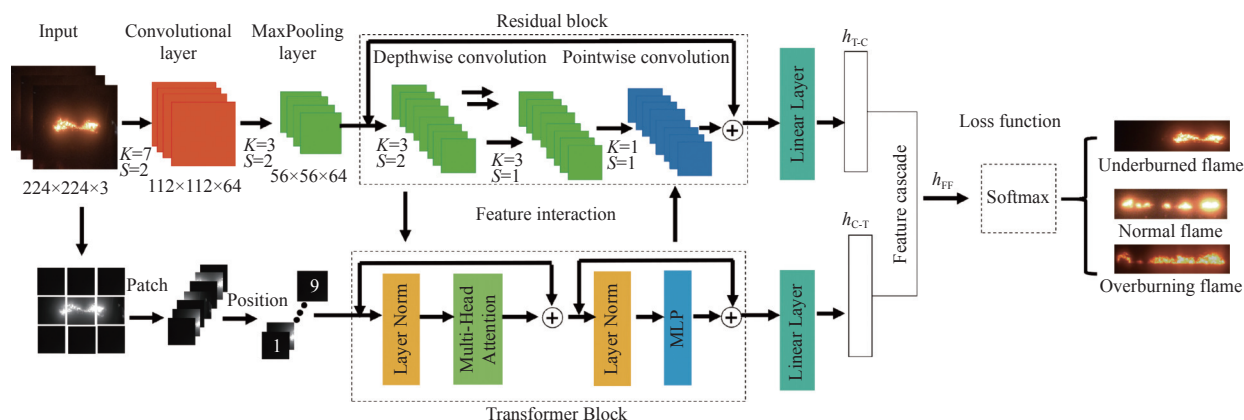


图1 CNN-Transformer 双流网络模型

Fig. 1 CNN-Transformer two-stream network model

### 1.1 CNN 流

充分利用卷积神经网络局部感知的特点,可以有效地提取烧结断面火焰空间局部信息,因此双流网络基于卷积神经网络设计了CNN流。综合考虑烧结断面火焰图像的小数据集和网络模型训练效率的问题,选取ResNet-18作为CNN流的骨干网络结构,并改变网络残差结构为深度可分离卷积结构。

深度可分离卷积包含两种卷积结构:深度卷积(DW)和逐点卷积(PW),如图2所示。首先,深度卷积对输入特征图的每个通道利用 $3 \times 3$ 卷积层独立进行卷积运算,得到相同数量的特征;然后,将输出的特征图经过 $1 \times 1$ 卷积核进行通道数的加权组合。采用深度可分离卷积能够使网络在不影响准确率的前提下明显降低网络计算量,获得更高效的轻量化网络<sup>[8]</sup>。

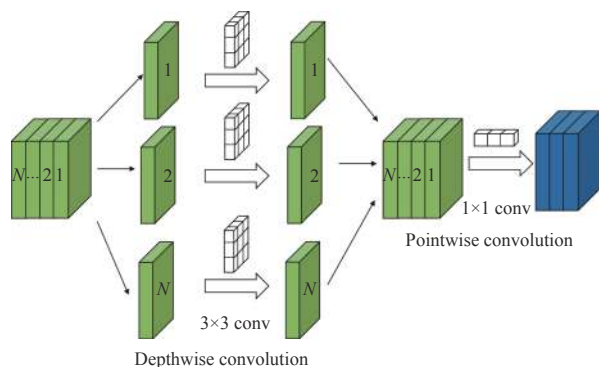


图2 深度可分离卷积结构

Fig. 2 Structure diagram of depth separable convolution

### 1.2 Transformer 流

虽然CNN流利用深度可分离卷积的ResNet-18网络提取了局部特征信息,但未能根据烧结火

焰图像大面积火焰区域很好地提取全局特征信息。因此,为了更好地保留烧结断面火焰图像的空间信息,引入Transformer流对烧结火焰图像特征提取进行补充。

Transformer结构对图像进行全局self-attention,得到较好的全局信息特征,适用于有大面积火焰区域的烧结图像,由于烧结火焰图像以位置信息、火焰区面积等特征进行火焰状态的分类,因此将RGB彩色图像转化为灰度图输入Transformer流中,减少参数量提升模型训练速度,同时不损失全局特征信息。Transformer是以MSA(multi-head attention)多头注意力为基础的模块,注意力根据输入特征的重要程度进行权重匹配。

该结构为多个单头注意力并行,但由于自注意力计算方式的特点,架构推理速度被严重限制。EdgeNeXt<sup>[9]</sup>的分割深度转置注意编码器(SDTA)中自注意力计算模块使以往查询矩阵 $Q(HW \times C)$ 、键矩阵 $K(C \times HW)$ 之间的空间维度计算转变为通道维度上的注意力计算,仅对查询矩阵进行转置,得到 $C \times C$ 矩阵,可以更加有效地编码全局上下文。计算公式如式(1)所示,对图像特征矩阵进行线性变换。

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{Q^T K}{\sqrt{d_k}}\right)V$$

$$\text{MSA}(X) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^o \quad (1)$$

$$\text{head}_i = \text{SA}_c(XW_i^Q, XW_i^K, XW_i^V)$$

式中: $Q$ 为查询矩阵; $K$ 为键矩阵; $V$ 为值矩阵; $\text{SA}_c$ 为通道self-attention计算; $W^o$ 为特征融合矩阵。

### 1.3 双流特征融合

CNN-Transformer双流网络将卷积结构与Transformer结构的优点有效结合,提高烧结火焰图像的分类识别精度。ZADEH A提出的TFN(tensor fusion



network)<sup>[10]</sup> 特征融合方式将提取的特征进行维度扩增, 然后对不同特征求笛卡尔积, 不仅计算了两个特征空间的相关性, 而且保留了各自的特定特征信息。但该方法增加了特征维度, 导致模型参数量增大, 模型容易过拟合。低秩多模态融合 (LMF, low-rank multimodal fusion)<sup>[11]</sup> 利用低秩权重张量和输入张量的并行分解来计算基于张量的融合, 解决了 TFN 维度增加的问题, 降低了时间复杂度。但 CNN 流和 Transformer 流的特征维度不同, 上述特征融合方式无法将全局特征与局部特征精准地相互嵌入, 使得识别精度不高。

为了更好地融合 CNN 流提取的局部特征与 Transformer 流提取的全局特征, 加入特征耦合单元 (FCU)<sup>[12]</sup> 和级联的融合策略。FCU 为一种交互性的方式将两路网络获取的特征融合, 逐步融合特征图和 patch embedding。这种交互特征融合方式的优点是特征信息交互共享, 但两路网络信息交互过程中也会不可避免地损失较多信息。为了解决这个问题, 本文提出了一种新的基于级联特征耦合单元 (cascading feature coupling units, C-FCU) 的特征融合模块, 可以获取更全局且具体的特征表达。

所提出的特征融合模块如图 3 所示。由于 CNN 和 Transformer 的特征维度不一致, 因此利用  $1 \times 1$  卷积和上下采样来实现维度和通道的对齐。当馈送到 Transformer 分支时, 特征图首先需要通过  $1 \times 1$  卷积来对齐 patch embedding 的通道数, 然后使用下采样模块完成空间维度对齐。FCU 下采样仅采用平均池化来压缩信息, 受到文献 [13] 启发, 本文将最大池化与平均池化结合, 最大池化能够编码目标的显著性信息, 可以更加精准地将 CNN 流的分辨性较强的特征信息交互到 Transformer 流。当从 Transformer branch 反馈到 CNN 分支时, patch embedding 需要双线性插值上采样以对齐空间尺度, 然后通过  $1 \times 1$  卷积将通道维度与 CNN 特征图的维度对齐, 最后将两路网络输出的特征进行级联操作, 即得到了级联交互特征融合图。整体运算过程如下:

$$\begin{cases} h_{C-T} = \sigma_1(k_{1 \times 1} * (\text{AvgPool}(h_C) + \text{MaxPool}(h_C))) + h_T \\ h_{T-C} = \sigma_2(\text{Interpolate}(\sigma_1(k_{1 \times 1} * (\text{AvgPool}(h_C) + \text{MaxPool}(h_C))) + h_T) * k_{1 \times 1}) + h_C \\ h_{FF} = [h_{C-T}, h_{T-C}] \end{cases} \quad (2)$$

式中:  $h_C$  表示 CNN 流局部特征图;  $h_T$  表示 Transformer 流全局特征图;  $h_{C-T}$  表示 CNN 流交互到 Transformer 流的特征图;  $h_{T-C}$  表示 Transformer 流交互到 CNN 流的特征图;  $h_{FF}$  表示融合后的特征图;  $k_{n \times n}$

表示  $n \times n$  卷积核;  $\sigma_1$  表示 GELU 激活函数;  $\sigma_2$  表示 ReLU 激活函数; AvgPool 表示平均池化运算; MaxPool 表示最大池化运算, Interpolate 表示双线性插值;  $[*]$  表示级联 (cascading) 操作。

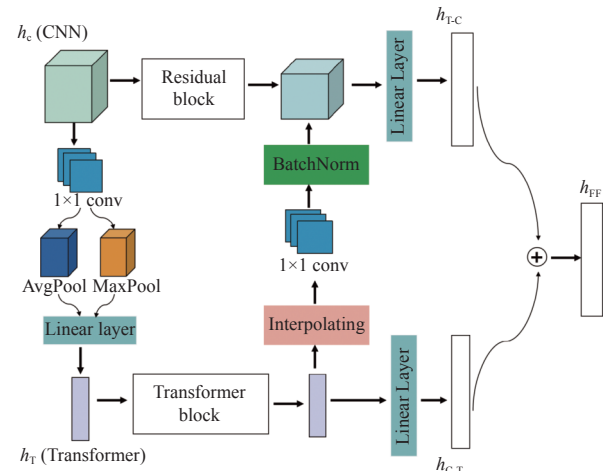


图 3 级联特征耦合单元

Fig. 3 Cascading feature coupling units

最后, 将融合后的特征输入到 softmax 损失函数层中, 预测 3 种火焰燃烧状态的概率, 进而实现烧结火焰图像正常燃烧、过烧以及欠烧状态的分类。

## 2 构建数据集

### 2.1 火焰样本图像采集

本研究试验自建火焰数据集, 采集的图像来源于唐钢 320 m<sup>2</sup> 烧结机的烧结机尾断面火焰图像, 数据集中的类别可分为正常烧结、欠烧和过烧 3 种状态, 如图 4 所示, 正常燃烧火焰能量的剖面分布成正态分布, 火焰面积大, 烧结过程中矿粉利用率高, 炼铁成品率高; 而欠烧图像和过烧图像能量剖面分布成非正太分布, 火焰面积均减少, 浪费燃料并且不符合生产要求。

### 2.2 图像数据预处理

火焰图像数据预处理是为了提高网络的泛化性能, 抑制网络模型过拟合, 并提高网络模型训练速度, 本研究实验采用 3 种方法:

1) 对获得的烧结结尾断面火焰图像进行数据扩增, 包括随机旋转、水平和垂直翻转、调整图像的亮度饱和度、对比度等方式;

2) 采集到的图像受烧结环境中烟气、粉尘的影响, 导致火焰图像模糊、噪声退化, 选取部分原始数据集中的烧结火焰图像进行图像复原并划入数据集中;

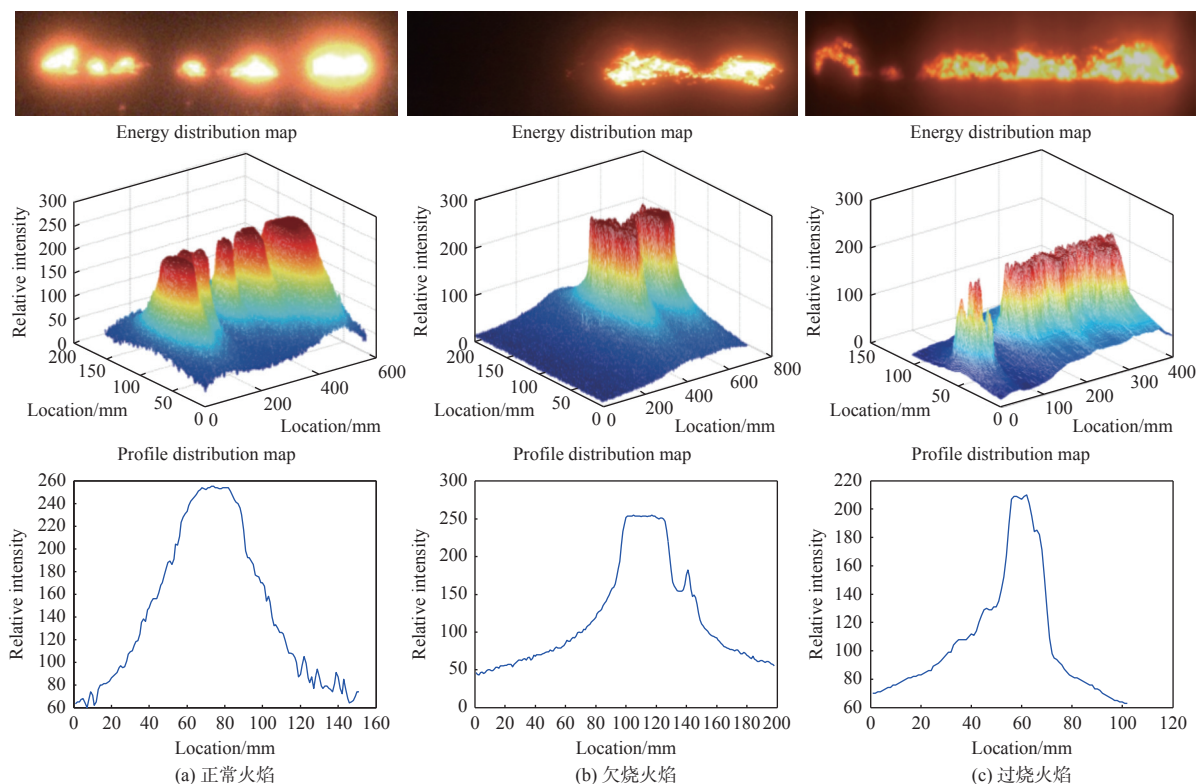


图 4 烧结断面 3 种火焰状态图像

Fig. 4 Three flame state images of sintered section

3) 烧结断面火焰图像有大面积的背景区域, 火焰图像背景区域具有大量的冗余信息, 其特征对断面火焰图像的识别分类影响不大, 为了减少网络背景区域的特征学习所消耗的时间, 对部分数据集中的图像进行火焰目标区域分割与背景区域的分割。

图像预处理后的数据集数量为原来的 10 倍, 共 13 116 张图像, 包括正常火焰图像 6 447 张, 过烧图像 4 305 张, 欠烧图像 2 364 张。

### 3 实验结果与分析

#### 3.1 实验设置

本次实验环境硬件训练平台为: 32 G Intel(R) Xeon(R)E7CPU, 12 G 显存 NVIDIA-RTX3060TiGPU; 软件平台为: 操作系统 Ubuntu20.04, 深度学习框架 PyTorch1.8.0 实现网络模型训练, 编程语言 Anaconda3, 并行计算架构 CUDA11 和 GPU 加速库 CUDNN8.2.1 等。

#### 3.2 模型参数对 CNN-Transformer 双流网络模型的影响分析

为了比较不同训练参数在训练过程中对 CNN-Transformer 双流网络模型的影响, 本实验分别选

用了随机梯度下降(SGD)优化器和自适应矩估计(Adam)优化器, 设置学习率为 0.01、0.001 和 0.000 1, 批大小选取 8、16、32 进行对比试验。学习率控制了基于损失梯度调整神经网络权值的速度。如果学习率较小, 梯度下降就会很慢; 如果学习率较大, 则梯度下降步长超过最优值。经过多次试验调参后, 最后选择随机梯度下降法(SGD), 当学习率为 0.001、动量参数为 0.937、批大小为 32、权重衰减率为 0.000 5 时, 测试集准确率可达 96.20%, 实验效果对比如表 1 所示。

表 1 CNN-Transformer 网络模型在不同参数下学习效果对比

Table 1 Comparison of learning effects of CNN-Transformer network models under different parameters

Optimizers	Learning rate	Accuracy rate/%	
		Training set	Validation set
SGD	0.01	99.56	95.07
	0.001	99.85	96.20
	0.000 1	99.37	94.36
Adam	0.01	98.67	85.69
	0.001	99.23	92.35
	0.000 1	98.26	78.96

### 3.3 消融实验

针对如何将局部特征和全局表示精确地相互嵌入的问题,本研究实验在烧结断面火焰图像数据集上对比了不同的特征融合方法对网络的影响,包括 TFN、LMF、FCU 和本文提出的级联特征耦合单元(C-FCU)的特征融合方式,如表 2 所示。实验结果表明,FCU 特征融合对比 TFN、LMF 特征融合方法火焰识别准确率有较为明显的提升,这是由于 FCU 可实现 CNN 流和 Transformer 流的精准特征交互。然而,由于特征在交互过程中导致部分信息丢失,因此 C-FCU 特征融合方法对 CNN-Transformer 双流网络分类火焰状态准确率效果较好,网络模型准确率可达 96.20%,相较于 FCU 方法提升了 1.06 个百分点。

表 2 不同特征融合方法结果对比

Table 2 Comparison of results of different feature fusion methods

Methods	Accuracy/%
TFN	92.53
LMF	93.02
FCU	95.14
C-FCU	96.20

此外为进一步验证所提双流网络中 CNN 流和 Transformer 流的有效性,在烧结断面火焰图像数据集上进行了详细的消融实验。表 3 为所提方法的消融实验结果,其中 CNN 流为 ResNet18 框架基础上将残差结构替换为深度可分离卷积结构,Transformer 流的骨干网络为 ViT,自注意力计算由空间转换到通道维度上。实验结果表明,本文所提的 CNN-Transformer 算法的分类精度要高于 CNN 流和 Transformer 流,在火焰图像数据集上的准确率分别提高 3.69% 和 2.2%,这表明 CNN-Transformer 算法能够有效地提取并融合烧结断面火焰像中的局部信息和全局信息,使得特征表达更加具有鉴别力,进一步提高火焰分类的精度。

表 3 CNN-Transformer 算法的消融实验

Table 3 Ablation experiment of CNN-Transformer algorithm

Models	Accuracy/%			Average accuracy/%
	Normal flame	Overburning flame	Underburned flame	
CNN	93.25	92.72	91.56	92.51
Transformer	95.00	93.45	93.55	94.00
CNN-Transformer	97.00	95.38	96.22	96.20

### 3.4 不同算法模型的对比分析

为了验证本文算法的有效性和先进性,在烧结火焰图像数据集上,将本文算法与现有的主流图像分类方法进行比较,如表 4 所示。需要说明的是,所有算法均采用相同的实验条件。从对比实验结果可以看出,本文提出的双流网络模型的性能均优于经典的卷积神经网络 ResNet18 和 InceptionV3;相较于轻量化卷积神经网络模型 MobileNet-V2,本文模型分类精度提高 8.85 个百分点;对比 Transformer 结构的 ViT 网络模型,只有其近 23% 的模型复杂度,本文模型分类精度提高 5.71 个百分点。这表明双流网络模型火焰分类效果优于单流网络模型。同时对比主流的特征融合网络,包括 Conformer<sup>[12]</sup>和 CMT<sup>[16]</sup>,同样融合了基于 CNN 和 Transformer 两个分支提取的特征,Conformer 网络模型采用 FCU 特征交互方式进行融合,CMT 网络模型将 CNN 层插入 Transformer 结构以提取特征。实验结果表明,相较于 Conformer 和 CMT 模型,本文模型参数量减少几乎近一半,分类准确率分别提升了 1.85 和 1.2 个百分点,模型训练时间减少 12.6 min 和 15.04 min,网络运行速度为 25.78 f/s,能够满足快速识别烧结火焰状态的要求。这表明深度可分离卷积和通道维度的注意力使得网络模型更加轻量化,同时采用 C-FCU 特征融合方法取得了与目前主流方法更优的分类准确率。

表 4 CNN-Transformer 双流网络模型与其他模型学习效果对比

Table 4 Comparison of learning effects between CNN-Transformer dual-stream network model and other models

Models	Params/M	FLOPs/G	Accuracy/%	Training time/min	Speed/fs <sup>-1</sup>
InceptionV3 <sup>[14]</sup>	23.21	6.02	90.50	85.72	20.79
ResNet18 <sup>[9]</sup>	11.65	3.82	91.00	67.95	22.53
ViT <sup>[7]</sup>	55.50	77.91	90.49	157.65	12.24
MobileNet-V2 <sup>[15]</sup>	3.51	0.58	87.35	34.15	30.58
Conformer <sup>[12]</sup>	23.53	5.23	94.35	73.85	23.86
CMT <sup>[16]</sup>	25.10	4.02	95.00	76.29	20.56
Ours	12.74	3.54	96.20	61.25	25.78

## 4 结论

传统的 CNN 算法使用局部感受野来进行特征捕获,无法直接对烧结断面火焰图像全局环境进



行建模。因此,本文提出面向烧结机尾断面火焰图像的 CNN-Transformer 双流网络。该算法包含两个组成部分: CNN 流和 Transformer 流。CNN 流将深度可分离卷积的 ResNet-18 模型提取的特征作为火焰场景的局部表达; Transformer 流基于自注意力获取上下文信息来提取火焰图像的全局表达。CNN-Transformer 网络模型有效融合了这两种特征,相较于一些现有的方法对火焰图像进行分类,分类精度取得了令人满意的结果。本研究用图像识别在线判断烧结终点,对减轻工人劳动强度、提高系统自动化程度、实现钢铁工业智能制造等具有重大工业应用价值,强化深度学习在钢铁工业的实际应用。

#### 参考文献:

- [1] 王福斌,刘贺飞,何江红,等. 烧结运行工艺参数分析及烧结行为模型构建[J]. 烧结球团, 2020, 45(6): 29-34.  
WANG Fubin, LIU Hefei, HE Jianghong, et al. Analysis of sintering operation process parameters and construction of sintering behavior model[J]. Sintering Pellet, 2020, 45(6): 29-34.
- [2] 王福斌,刘贺飞,王蕊,等. 烧结断面火焰图像多核 Boosting 显著性检测[J]. 计算机辅助设计与图形学学报, 2021, 33(9): 1466-1474.  
WANG Fubin, LIU Hefei, WANG Rui, et al. Multi-core Boosting saliency detection of flame images of sintered sections[J]. Journal of Computer Aided Design and Graphics, 2021, 33(9): 1466-1474.
- [3] 李江昀,杨志方,郑俊锋,等. 深度学习技术在钢铁工业中的应用[J]. 钢铁, 2021, 56(9): 43-49.  
LI Jinagyun, YANG Zhifang, ZHENG Junfeng, et al. Application of deep learning technology in iron and steel industry[J]. Iron and Steel, 2021, 56(9): 43-49.
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [5] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016.
- [6] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17) [2022-09-13]. <https://arxiv.org/abs/1704.04861>.
- [7] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. [2022-09-13]. <https://www.xueshufan.com/publication/3119786062>.
- [8] 张乾闯,郭晨霞,杨瑞峰,等. 基于轻量级网络的光纤环图像超分辨率重建[J]. 应用光学, 2022, 43(5): 913-920.  
ZHANG Qianchuang, GUO Chenxia, YANG Ruifeng, et al. Super resolution reconstruction of optical fiber ring image based on lightweight network[J]. Applied Optics, 2022, 43(5): 913-920.
- [9] MUHAMMAD M, ABDELRAHMAN S, HISHAM C, et al. EdgeNeXt: efficiently amalgamated cnn-transformer architecture for mobile vision applications[EB/OL]. (2022-06-21) [2022-09-13]. <https://arxiv.org/abs/2206.10589v3>.
- [10] ZADEH A, CHEN M, PORIA S, et al. Tensor fusion network for multimodal sentiment analysis[EB/OL]. (2017-07-23) [2022-09-13]. <https://arxiv.org/pdf/1707.07250.pdf>.
- [11] LIU Z, SHEN Y, LAKSHMINARASIMHAN V B, et al. Efficient low-rank multimodal fusion with modality-specific factors[EB/OL]. (2018-05-31) [2022-09-13]. <https://arxiv.org/abs/1806.00064>.
- [12] PENG Z, HUANG W, GU S, et al. Conformer: local features coupling global representations for visual recognition[EB/OL]. (2021-05-09) [2022-09-13]. <https://arxiv.org/abs/2105.03889>.
- [13] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[EB/OL]. (2018-07-17) [2022-09-13]. <https://arxiv.org/abs/1807.06521>.
- [14] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 2818-2826.
- [15] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. New York: IEEE, 2018: 4510-4520.
- [16] GUO J, HAN K, WU H, et al. CMT: convolutional neural networks meet vision transformers[EB/OL]. (2021-07-13) [2022-09-13]. <https://arxiv.org/abs/2107.06263>.