

# 基于改进 CeiT 的 GTAW 焊接熔透状态识别方法

王颖, 高胜

(东北石油大学, 大庆, 163318)

**摘要:** 针对熔池信息与背景相似度高、噪声多、预测实时性差、识别精度低等问题, 提出了基于改进 CeiT 网络模型的 GTAW 焊接熔透状态识别方法. 首先通过 MobileNetV3 对 Image-to-Tokens 模块进行轻量化改进, 提升预测的实时性能; 其次设计 DGCA 模块改进 LeFF 模块来增强特征间的远程依赖关系、丰富类标记中所包含的分类信息; 最后将 LeFF 模块中的底层特征和高层语义特征进行融合, 提高模型对熔池特征的表示能力. 仿真结果表明, 与 MobileNetV3, ResNet50, ShuffleNetV2, DeiT 和 CeiT 模型相比, 所提出的模型获得了更高的准确率和较快的检测速度.

**创新点:** 建立了一种基于改进 CeiT 的 GTAW 焊接熔透状态识别模型, 与其他模型相比获得了更高的准确率和较快的检测速度.

**关键词:** 图像处理; 熔池; 熔透状态; CeiT 网络; 熔化极气体保护焊

**中图分类号:** TG 409      **文献标识码:** A      **doi:** 10.12073/j.hjxb.20230327002

## 0 序言

焊缝熔透状态是表征焊接质量、影响焊接产品使用性能最重要的指标之一, 然而由于焊接过程的瞬时性和复杂性, 使得利用正面熔池信息实时准确获得焊缝熔透状态成为焊接过程监测和质量控制的关键瓶颈<sup>[1-2]</sup>. 已有的关于焊缝熔透状态识别的研究主要采用机器学习的方法, 通过人为设计特征, 建立其与熔透状态之间的非线性映射关系; 然而人为设计的特征具有较强的主观性, 严重影响识别精度和速度. 随着深度学习的不断发展, 其通过自动提取所需的特征进行预测, 提高了识别的自动化程度和准确率, 有效解决了传统方法所面临的问题<sup>[3-4]</sup>. 因此, 近些年有许多学者将深度学习应用到熔池变化的在线监测中, 其中基于视觉信号的研究和应用最为广泛<sup>[5-8]</sup>. 文献 [9] 利用短时傅里叶变换获得电弧声的时频频谱图像, 建立卷积神经网络识

别焊缝的熔透状态; 文献 [10] 以感兴趣区域的温度场图像为输入, 基于深度残差网络建立了焊缝熔透预测模型; 文献 [11] 以采集的二维熔池图像作为输入, 建立了 CNN 熔透状态预测模型; 文献 [12] 以匙孔图像作为输入, 建立 LeNet-5 网络模型对熔透状态进行识别; 文献 [13] 以正面熔池图像作为输入, 搭建 CNN 模型预测烧穿, 并建立回归模型对穿透深度进行估计; 文献 [14] 利用 HDR 焊接摄像机获得熔池/匙孔图像, 使用图像增强方法提高训练数据集的多样性, 并建立 Resnet 网络对焊接状态进行识别; 文献 [15] 以熔池/匙孔图像序列作为底层输入, 利用 CNN 提取图像序列中每一帧图像的空域特征, 同时建立 BiLSTM 网络挖掘时域内的上下文信息, 最后利用获得的时空特征映射到具体的熔透状态上. 基于深度学习的方法大大提升了识别的准确率, 为了进一步提升预测的实时性, 文献 [16] 以 SSD 作为主干网络, 利用视觉传感信号作为输入, 将特征提取网络 VGG 替换为轻量网络 Mobilenet, 熔池的定位与识别速度得到了提升; 文献 [17] 基于 MobileNetV2 构建了迁移学习模型, 将 ImageNet 数据集进行预训练获得的参数迁移到自建的 GTAW 熔池图像数据预测中, 训练速度得到了大幅提升; 文献 [18] 利用从振荡焊接熔池表面反射的点

收稿日期: 2023-03-27

**基金项目:** 国家自然科学基金资助项目 (61702093); 国家重点研发计划项目 (2018YFE0196000); 黑龙江省自然科学基金 (F2018003); 黑龙江省博士后专项 (LBH-Q20077); 东北石油大学青年科学基金项目 (2020QNL-10).

结构激光图像进行焊缝熔深预测,在不对图像做任何预处理的情况下,利用提出的小批量梯度下降法进行训练,获得了较快的训练速度和较高的准确率。

Google 提出的 Vision Transformer<sup>[19-20]</sup>(ViT) 在图像识别领域也取得了突破性进展,其将图像分割成固定大小的像素块输入到编码器,编码器由多头自我注意力机制 (multi-head self-attention mechanism, MSA) 和多层感知模块 (multi-layer perceptron, MLP) 组成,利用最后一个编码器层的分类标记作为全局特征表示,获得最终的分类结果。ViT 在图像识别领域取得了与卷积网络相当的性能,但需要大量的训练数据或额外的监督。为了克服这个限制,提出了 CeiT 网络<sup>[21]</sup>,它集合了 CNN 在提取低级特征、增强局部性方面,以及 ViT 在建立远程依赖关系方面的优势,在数据量、准确率及收敛速度方面进行了均衡。但为了进一步满足焊接过程的实时性和准确性要求,对 CeiT 网络进行了改进。首先通过聚焦 (Focus) 模块、通道分组、通道混洗、多尺度特征融合构建了一种 FMCbneck 模块,使得模型

能够获得更多的熔池空间位置信息;其次将 FMCbneck 模块嵌入 MobileNetV3(Large) 网络中,并调整其结构代替 Image-to-Tokens 模块进行熔池图像浅层特征提取,在保证较高检测精度的条件下,极大的降低了模型参数量,提高了模型的检测速度;设计了 DGCA 模块并将其应用到 LeFF 中,有效增强了特征间的远程依赖关系、丰富了类标记中所包含的分类信息,并通过将 LeFF 模块中的底层特征和高层语义特征进行融合,提高了模型对熔池特征的表示能力、进一步提升了检测效果;最后与不同的网络模型进行对比,证明了所提模型在保持较低参数量的同时拥有着更高的精确度。

## 1 基于改进 CeiT 的熔透状态识别模型

### 1.1 CeiT 网络结构

CeiT 网络结构如图 1 所示,其在 ViT 的基础上设计了 Image-to-Tokens(I2T) 模块、LeFF 模块和 LCA 模块。

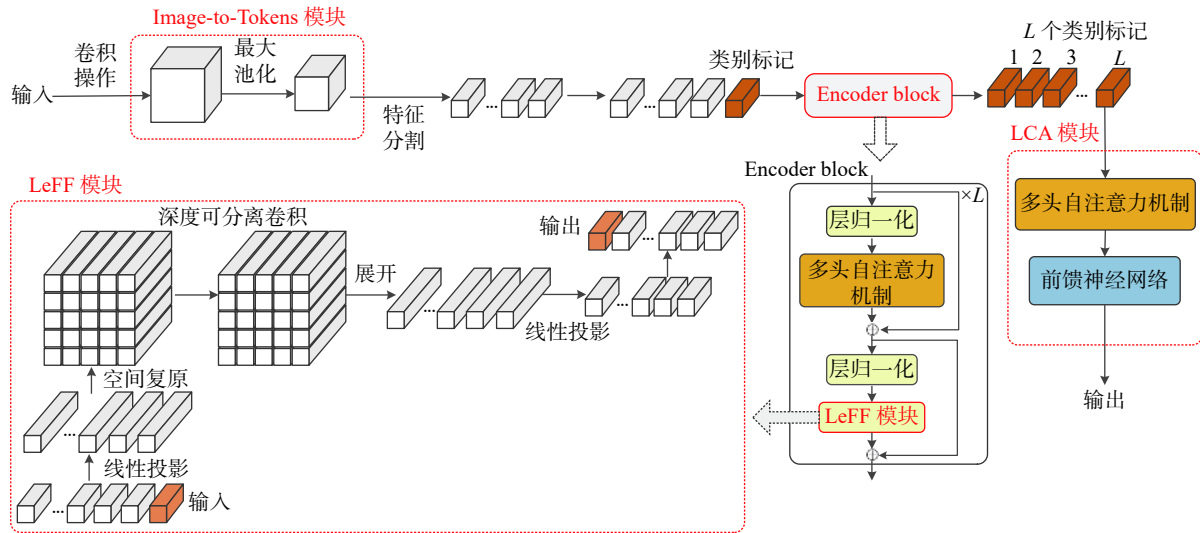


图 1 CeiT 网络结构

Fig. 1 CeiT network structure

I2T 模块由一个卷积层和一个最大池化层组成,用于提取熔池图像的底层特征,即

$$X' = \text{I2T}(X) = \text{MaxPool}(\text{BN}(\text{Conv}(X))) \quad (1)$$

式中:  $X$  为 I2T 模块的输入特征。

将提取的熔池底层特征  $X'$  分割 (Split) 成特征小块,并通过 flatten 操作将特征小块转换成一维向量,同时添加一个可学习的类别块,这个类别块用

于与所有的特征小块进行交互,最终从类别块中学习得到用于分类的特征。特征序列加入位置编码送入 Encoder block 中,重复堆叠  $L$  次。在 Encoder block 中除了包含 MSA 模块、Layer Normalization 模块 (Layer Norm)、Add(残差块) 外,还设计了局部增强前馈模块 (LeFF),目的是通过使用深度卷积增加相邻特征小块在空间维度上的相关性。抽取 Encoder

block 所得交互特征中的  $L$  个类别块送入 LCA 模块 (layer-wise class token attention), 目的是为了关注不同层的类标记信息, 将注意力放在不同层的特征块上, 最后获得输出信息。

### 1.2 Image-to-Tokens 模块轻量化改进

为充分利用熔池图像中的浅层特征、减少冗余操作、提高焊接检测的实时性与部署便捷性, 使用 MobileNetV3(Large) 作为 Image-to-Tokens 模块的特征提取网络. MobileNetV3<sup>[22]</sup> 网络的逆残差结构、深度可分离卷积、SE(squeeze and excitation) 通道注意力机制和 h-swish 激活函数, 使得模型能够很好地获得熔池图像中的重要特征信息, 并且避免训练时出现梯度消失及计算量过大的问题. 然而轻量化的网络结构也会带来检测精度上的损失, 为此对 MobileNetV3 的 bneck 结构进行改进, 改进的 bneck 模块结构 (FMCbneck) 如图 2 所示。

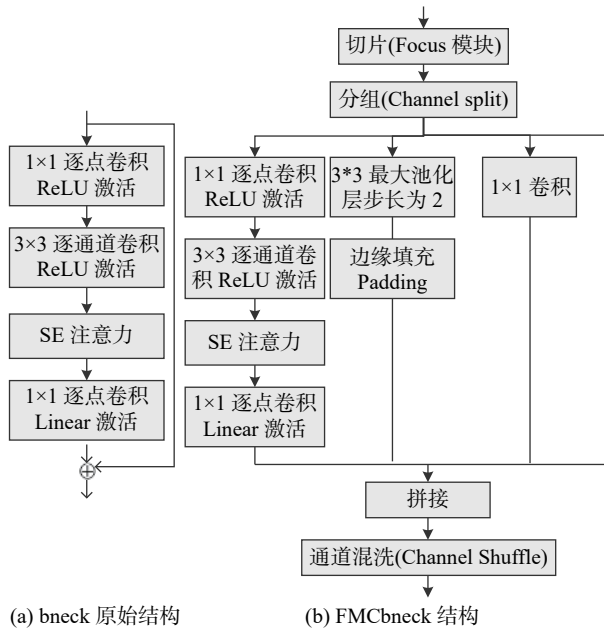


图 2 改进前后的 bneck 结构

Fig. 2 Improved bneck structure before and after. (a) bneck original structure; (b) FMCbneck structure

(1) 由于熔池图像像素低、携带的信息少, 并且在下采样过程中特征信息会部分丢失, 使得在处理过程中容易导致图像失真. 为解决这一问题, 在 bneck 输入端引入了聚焦模块 (Focus 模块), Focus 模块通过间隔采样对图片进行切片操作, 目的在于将高分辨率特征图拆分成多个低分辨率的特征图, 且没有信息丢失, Focus 切片采样原理如图 3 所示。

(2) 为了融合不同尺度的缺陷特征, 提升模型的泛化能力、抑制过拟合、扩大特征感受野, 并且不

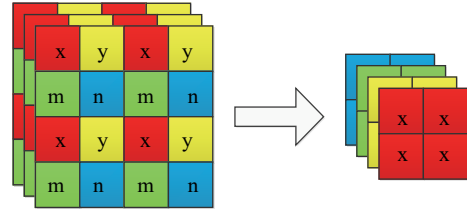


图 3 Focus 切片采样原理

Fig. 3 Focus slice sampling principle

会大幅度增加模型的复杂度, 在原 bneck 基础上增加 2 个分支, 分别为 Maxpooling 和  $1 \times 1$  的普通卷积。

(3) 为了进一步降低计算量, 增加特征信息间的交流和特征的表达能力, 对 Focus 切片后的特征通道进行了分组操作 (channel split), 并在多尺度特征融合后进行了通道混洗操作 (channel shuffle)。

除了对 MobileNetV3 的 bneck 结构进行了改进, 还对其中的 SE 模块以及  $5 \times 5$  卷积核的位置进行了调整. 通过消融试验发现 SE 模块在越靠后的位置对模型精度的提升越大,  $5 \times 5$  卷积核在网络的中后部对模型性能的提升作用更明显, 并且在步长为 1 时采用 FMCbneck 结构, 步长为 2 时采用原 bneck 结构, 模型效果更好. 所采用的 MobileNetV3 网络参数如表 1 所示. 表 1 中 exp 表示 FMCbneck 和 bneck 中第一层  $1 \times 1$  卷积升高的维度; #out 表示 FMCbneck 和 bneck 输出的通道数; SE 表示是否使用 SE 模块。

### 1.3 LeFF 模块改进

CeiT 网络的 LeFF 模块结构如图 4 所示, 输出的类标记将作为 LCA 模块的输入参与分类. 为了丰富类标记中所包含的分类信息、提高模型对熔池特征的表示能力、增强特征间的远程依赖关系, 改进后的 LeFF 结构如图 5 所示。

将 Depth-wise Convolution(DWConv) 操作替换成 DGCA 模块, 对特征图进行增强且保持较少的计算量. DGCA 模块结构如图 6 所示, 其在 DWConv 操作前后引入 Ghost Module、建立残差机制、添加 Coordinate Attention(CA) 注意力模块。

CA 注意力模块分别沿两个方向聚合特征, 一个方向捕捉长期依赖, 另一个方向保留精确的位置信息, 起到了聚焦显著目标区域、抑制飞溅、弧光等背景噪声的作用, 其结构如图 7 所示。

对输入尺寸为  $W \times H \times C$  的特征图 ( $W, H, C$  分别表示特征图的宽、高和通道数) 使用  $(1, W)$  和  $(H, 1)$  进行平均池化编码, 分别得到了水平方向感知

表 1 MobileNetV3 网络参数  
Table 1 MobileNetV3 network parameters

序号号	输入层	操作	步长	exp size	#out	SE	激活函数
1	$224^2 \times 3$	Conv2d	2	—	16	—	h-swish
2	$112^2 \times 16$	FMCbneck, $3 \times 3$	1	16	16	—	ReLU
3	$112^2 \times 16$	bneck, $3 \times 3$	2	64	24	—	ReLU
4	$56^2 \times 24$	FMCbneck, $3 \times 3$	1	72	24	—	ReLU
5	$56^2 \times 24$	bneck, $3 \times 3$	2	72	40	—	ReLU
6	$28^2 \times 40$	FMCbneck, $3 \times 3$	1	120	40	—	ReLU
7	$28^2 \times 40$	FMCbneck, $3 \times 3$	1	120	40	—	ReLU
8	$28^2 \times 40$	bneck, $3 \times 3$	2	240	80	—	h-swish
9	$14^2 \times 80$	FMCbneck, $5 \times 5$	1	200	80	√	h-swish
10	$14^2 \times 80$	FMCbneck, $5 \times 5$	1	184	80	√	h-swish
11	$14^2 \times 80$	FMCbneck, $5 \times 5$	1	184	80	√	h-swish
12	$14^2 \times 80$	FMCbneck, $5 \times 5$	1	480	112	√	h-swish

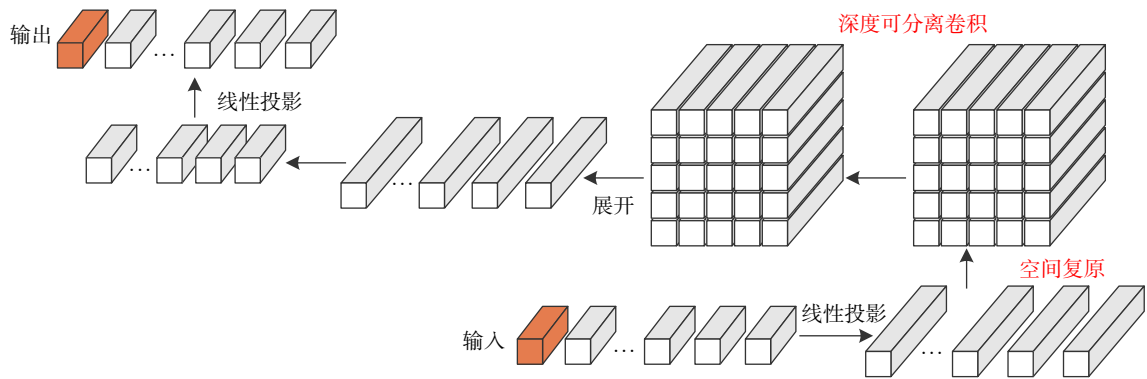


图 4 原始 LeFF 结构图

Fig. 4 Original LeFF structure diagram

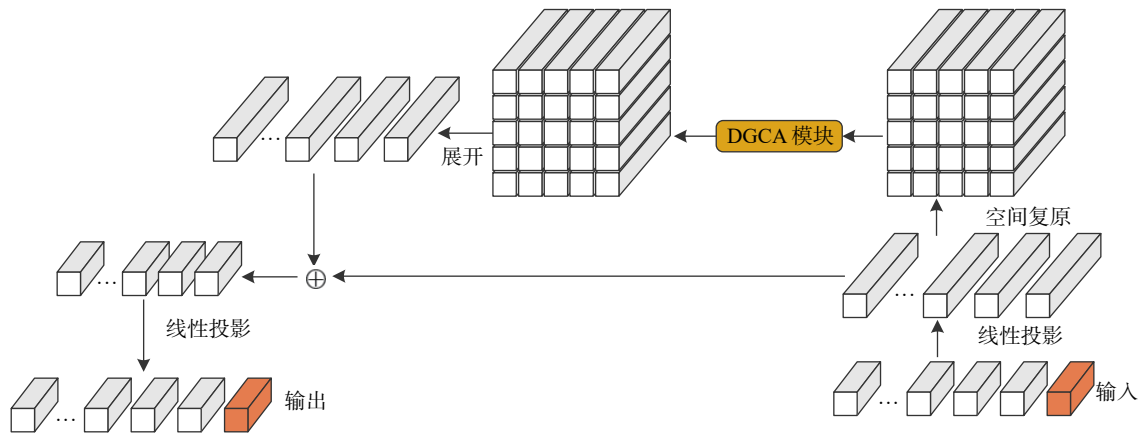


图 5 改进的 LeFF 结构图

Fig. 5 Improved LeFF structure diagram

特征图  $C \times H \times 1$  和垂直方向感知特征图  $C \times 1 \times W$ , 计算原理为

$$Z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (2)$$



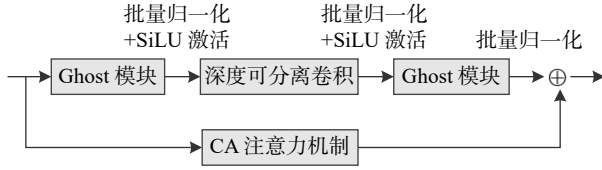


图 6 DGCA 模块

Fig. 6 DGCA Module

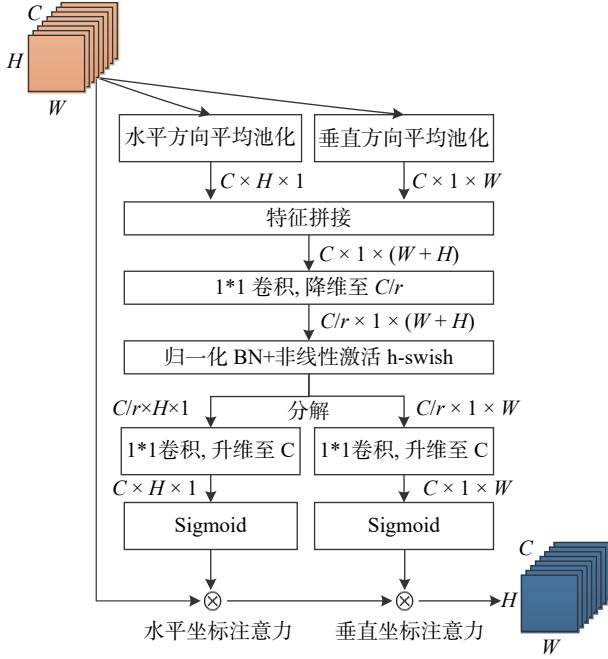


图 7 CA 注意力模块结构图

Fig. 7 CA attention module structure diagram

$$Z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (3)$$

式中： $Z_c^h(h)$ 表示高度为  $h$  的第  $c$  个通道的输出结果； $Z_c^w(w)$ 表示宽度为  $w$  的第  $c$  个通道的输出结果。

将垂直方向和水平方向感知特征图在空间维度进行拼接，得到特征图  $C \times 1 \times (W+H)$ 。利用  $1 \times 1$  卷积变换对其进行降维至  $C/r$ ， $r$  用于控制缩减率，再进行归一化和非线性变换操作。然后沿着空间维数将其分解为两个单独的张量  $C/r \times H \times 1$  和  $C/r \times 1 \times W$ ，再利用  $1 \times 1$  卷积变换分别对其升维至  $C$ ，并经过 sigmoid 激活函数将其作用于原特征图上，最终输出  $y_c(i, j)$  为

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (4)$$

式中： $g_c^h(i)$ 、 $g_c^w(j)$  分别为两个空间方向的权重。与此同时，考虑到每个特征图仅包含该层的信息，为了结合底层特征图的高分辨率以及高层特征的语义性能优势，将 LeFF 模块中空间复原前与复原后的特征图进行融合，以加强层与层之间的联系。

## 2 数据准备和试验设置

### 2.1 样本获取

采用的焊接方法为 GTAW 脉冲焊，保护气体为氩气，焊接材料为厚度 3.175 mm 的 304 不锈钢，摄像头采集频率为 1 000 Hz，曝光时间为 20  $\mu$ s，硬件设备如图 8 所示。将工业相机和数据采集卡采集到的熔池正面图片和相应的电流、电压信号传输到计算机，计算机对接收到的数据处理之后可通过调节电流电压控制器和伺服电机控制器控制焊接过程中的电流电压以及焊接速度，从而得到不同状态下的熔池数据。

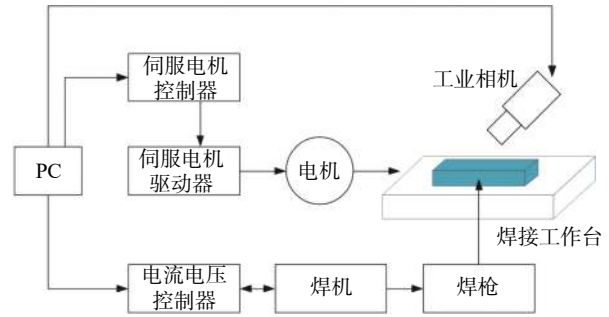


图 8 图像采集平台

Fig. 8 Image acquisition platform

焊接电流控制在 140 ~ 200 A，变化步长为 20 A；焊接速度控制在 20 ~ 40 cm/min，变化步长为 10 cm/min；通过调节焊接电流与焊接速度，获得未熔透（图 9(a)）、熔透（图 9(b)）、烧穿（图 9(c)）三种状态下的熔池图片，去掉起弧和收弧时不稳定状态下采集的图像，最终采集得到 1 475 张未熔透照片、7 567 张正常熔透照片和 725 张烧穿照片，其中未熔透样本标记为 0，正常熔透样本标记为 1，烧穿样本标记为 2。

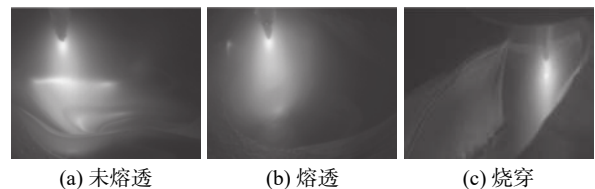


图 9 三种焊接熔池

Fig. 9 Three types of welding pools. (a) not melted through; (b) melt through; (c) burn through

### 2.2 图像预处理

熔池图像的预处理是通过提取 ROI、3 次样条插值和归一化进行的。视觉系统采集到的原始图像

尺寸为  $1\,280 \times 700$ , 直接对原始图像进行训练不仅干扰信息多而且计算量大, 会严重影响训练速度, 因此选择从原始图像中裁剪感兴趣的区域, 去掉无关信息. 由于在采集的图像中熔池的位置几乎保持在固定的区域, 因此选择 ROI 自动提取算法. 未熔透和熔透两类图像设定的左上角坐标为 (200, 50), 烧穿图像设定的左上角坐标为 (390, 150), 窗口大小为  $700 \times 350$ , 提取 ROI 的结果如图 10 所示.

采用 3 次样条插值方法将 ROI 图像调整为  $224 \times 224$ , 再将图像中的每个像素值除以 255 归一化为 [0-1]. 复制灰度图像矩阵, 使其变为  $224 \times 224 \times 3$ .

### 2.3 数据增强

在深度学习中, 样本数据集大小和数据质量会直接影响模型的泛化能力和鲁棒性. 在实际焊接过程中, 虽然数据量在不断增大, 但大多数数据为正常数据, 未熔透和烧穿的样本很少. 因此, 在保证图像自然特征表达情况下, 将对未熔透和烧穿的样本进行数据增强, 以弥补样本数量不均衡问题. 增强方法为改变图像的对比度和亮度、水平翻转、椒盐

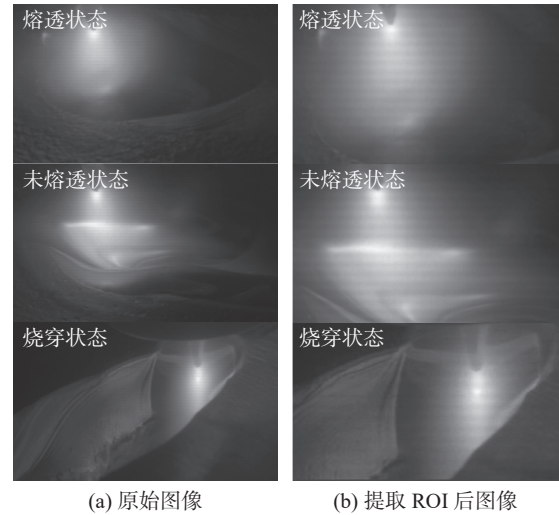


图 10 ROI 提取结果

Fig. 10 ROI extraction results. (a) original image; (b) image after ROI extraction

噪声、高斯噪声等, 其中烧穿的样本还采用了多种方式叠加的方法进行数据增强, 处理效果如图 11 所示. 最终未熔透样本扩充至 7 375 张、烧穿样本扩充至 6 579 张, 按照 8:1:1 的比例随机划分为训练集、验证集和测试集, 各数据集所含图片数量如表 2 所示.

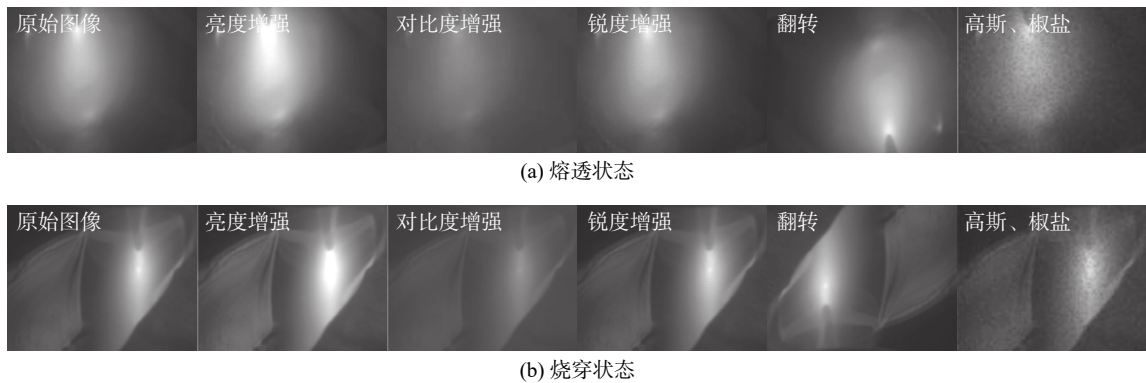


图 11 样本数据增强效果

Fig. 11 Sample data enhancement effect. (a) fusion state; (b) burn-through condition

表 2 扩充后各数据集数量 (张)

Table 2 Number of each data set after expansion

数据集	未熔透	熔透	烧穿	总和
训练集	5 899	6 053	5 263	17 215
验证集	738	757	658	2 153
测试集	738	757	658	2 153
总和	7 375	7 567	6 579	21 521

### 2.4 试验设置

CPU 型号为 Intel Core i9 13900K, GPU 为 16G

显存的 NVIDIA RTX 4080, 操作系统为 Ubuntu 16.04 LST 64 位, 深度学习框架为 Pytorch 1.8, TorchVision 0.8, CUDA 版本为 10.1. 批次大小设置为 128, 训练轮次为 100, 优化器使用 Adam (Adaptive momentum), 学习率衰减策略为

$$l_{t+1} = \begin{cases} b + \frac{rl_t}{R_2} (r < R_2) \\ \frac{r - R_2}{R_2 - R} (r \geq R_2) \end{cases} \quad (5)$$

式中:  $l_t$  为学习率;  $b$  为预热初始值;  $R_2$  为全局预热周期;  $R$  为训练总轮数;  $r$  为当前轮数.

## 2.5 损失函数

损失函数由两部分组成:分类损失函数和中心损失函数.

(1) 分类损失函数 $L_1$ . 采用交叉熵损失函数来计算预测值和真实值的误差, 计算式为

$$L_1 = \frac{1}{N} \sum_{i=1}^N \sum_{c=1}^M y_{ic} \log(p_{ic}) \quad (6)$$

式中:  $N$  为样本数;  $M$  为类别数目 ( $M=3$ );  $y_{ic}$  的取值为 0 或 1 (如果样本  $i$  的类别为  $c$  取 1, 否则取 0);  $p_{ic}$  为样本  $i$  属于类别  $c$  的预测概率.

(2) 中心损失函数 $L_2$ . 为了增大不同熔透类别间距离、减小同一类别内距离, 采用中心损失函数增强同类样本相似性, 计算式为

$$L_2 = \frac{1}{N} \sum_{i=1}^N (x_i - w_{y_i})^2 \quad (7)$$

式中:  $w_{y_i}$  为第  $y_i$  类训练样本特征的中心位置;  $x_i$  为样本  $i$  进入全连接层之前的特征.

(3) 总体损失函数  $L$  总体损失函数为

$$L = \lambda L_1 + (1 - \lambda) L_2 \quad (8)$$

式中:  $\lambda$  和  $(1 - \lambda)$  分别表示各自的权重,  $\lambda$  的初始值为 0.5, 在训练过程中不断对两个超参数进行更新.

## 2.6 评价指标

精确率 (Precision)、召回率 (Recall) 和 F1 分数 (F1-Score)、准确率 (Accuracy) 作为模型的评价指标, 模型训练时间和模型内存占用量作为模型计算复杂度的评价指标, 识别单个样本平均时间作为模型检测速度的评价指标.

$P_{\text{precision}}$  表示预测为正的样本实际为正样本的比例, 精确率越高则模型对负样本区分能力越强, 计算式为

$$P_{\text{precision}} = \frac{TP}{TP + FP} \times 100\% \quad (9)$$

$R_{\text{recall}}$  表示所有正样本中被预测为正样本的比例, 召回率越高则模型对正样本的区分能力越强, 计算式为

$$R_{\text{recall}} = \frac{TP}{TP + FN} \times 100\% \quad (10)$$

$F_{F1\text{-score}}$  是  $P_{\text{precision}}$  和  $R_{\text{recall}}$  的调和平均值, 计算式为

$$F_{F1\text{-score}} = \frac{2 \cdot P_{\text{precision}} \cdot R_{\text{recall}}}{P_{\text{precision}} + R_{\text{recall}}} \times 100\% \quad (11)$$

$A_{\text{accuracy}}$  表示预测正确的样本占总样本的比例,

计算式为

$$A_{\text{accuracy}} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (12)$$

式中:  $TP$  表示实际正类预测为正类的数量;  $TN$  表示实际负类预测为负类的数量;  $FP$  表示实际负类预测为正类的数量;  $FN$  表示实际正类预测为负类的数量.

## 3 试验结果与分析

### 3.1 数据增强对检测效果的影响

利用改进后的 CeiT 模型对数据增强效果进行对比试验, 表 3 为未使用数据增强的试验结果, 表 4 为使用数据增强的试验结果. 数据增强前未熔透照片 1 475 张、正常熔透照片 7 567 张、烧穿照片 725 张. 数据增强后未熔透样本 7 375 张、正常熔透样本 7 567 张、烧穿样本 6 579 张.

表 3 未数据增强试验结果

Table 3 No data enhancement test results

熔透类别	未熔透	正常熔透	烧穿
未熔透	82.5	12.5	5.0
正常熔透	3.0	96.0	1.0
烧穿	15.0	5.0	80

注: 12.5 表示未熔透状态的数据被预测为正常熔透的概率.

表 4 数据增强试验结果

Table 4 Data enhancement test results

熔透类别	未熔透	正常熔透	烧穿
未熔透	96.5	2.0	1.5
正常熔透	2.5	96.5	1.0
烧穿	1.0	2.0	97

通过试验结果可以看出使用数据增强三种类别的识别效果均得到了提升, 尤其是未熔透和烧穿. 数据样本不均衡时, 未熔透样本会有大部分误判为正常熔透, 烧穿样本误判为未熔透. 由此可见数据增强缓解了由于数据量不足造成的模型过拟合和泛化能力不足等问题.

### 3.2 学习率对检测效果的影响

学习率是深度学习中非常重要的超参数之一, 为了选择合适的学习率, 在控制其他条件相同的情况下, 分别设置学习率为 0.01, 0.001 和 0.000 1 进行分析, 图 12 为试验验证集结果. 比较图 12(a)

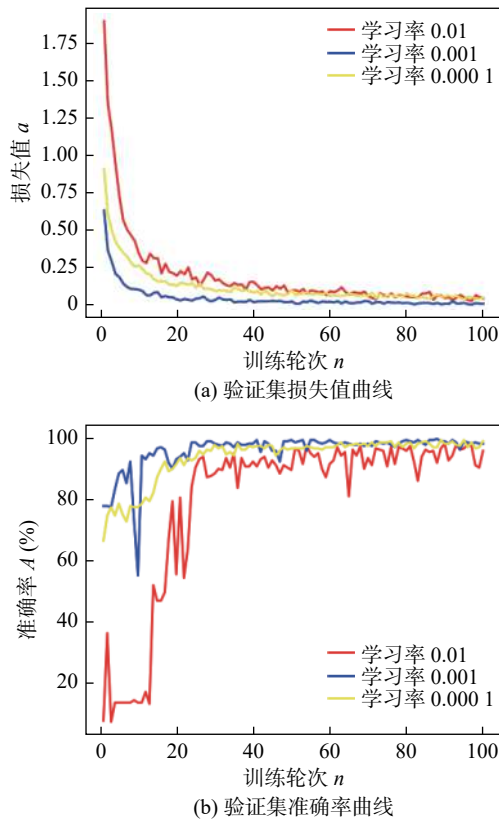


图 12 学习率对比结果

Fig. 12 Learning rate comparison results. (a) validation set loss value curve; (b) validation set accuracy curve

中曲线,学习率为 0.001 的损失值较低.对比图 12(b) 中曲线,训练初期学习率为 0.01 与 0.001 准确率波动较大,训练 40 轮时,学习率 0.000 1 的准确率随着训练轮数的增加变化稳定,而其他两种的准确率还有一定波动.综合对比验证集上损失值和准确率的变化情况,可以看出当学习率为 0.000 1 时,泛化能力最佳,可以获得较好的结果.

### 3.3 消融试验结果

为了证明对 CeiT 模型一系列改进在提升模型整体性能方面的有效性,在训练集上进行了消融试验.方案如表 5 所示,其中“I2T 模块改进”对应 1.2 节 Image-to-Tokens 模块轻量化改进,“LeFF 模块改进”对应 1.3 节;“M3”表示利用 MobileNetV3 进行底层熔池特征提取;“BNSA”表示 MobileNetV3 的 bneck 模块改进和结构调整;“DG”表示使用 DGCA 模块替换 DWConv 操作;“FI”表示将 LeFF 模块中空间复原前与复原后的特征图进行融合;“C0”表示未熔透样本对应的识别结果;“C1”表示正常熔透样本对应的识别结果;“C2”表示烧穿样本对应的识别结果;“—”表示在改进 CeiT 时不使用该策略;“II”表示在改进 CeiT 时使用该策略.

表 5 消融试验结果

Table 5 Ablation test results

方案	I2T模块改进		LeFF模块改进		精确率			召回率			F1分数			准确率			训练时间 t/h
	M3	BNSA	DG	FI	C0	C1	C2	C0	C1	C2	C0	C1	C2	C0	C1	C2	
1	—	—	—	—	85.1	86.9	85.3	80.6	81.3	80.0	84.6	87.3	84.5	87.5	87.9	86.8	4.07
2	II	—	—	—	81.2	83.7	81.9	74.7	80.7	79.2	81.1	83.8	80.9	82.1	84.3	83.5	3.23
3	II	II	—	—	91.2	92.3	92.0	87.8	88.8	88.6	92.5	93.3	92.9	90.3	93.8	91.4	3.58
4	—	—	II	—	90.6	91.6	91.4	87.4	89.0	86.7	90.6	92.4	91.0	90.8	92.4	91.3	4.94
5	—	—	II	II	91.3	92.5	92.1	87.7	89.5	87.6	91.7	91.6	92.2	91.7	92.6	92.4	5.16
6	II	II	II	II	94.6	95.4	94.9	91.0	92.7	90.7	94.3	95.5	94.3	95.9	96.0	95.8	3.76

由表 5 试验结果可知,方案 2 利用轻量化网络 MobileNetV3 进行熔池底层特征提取训练时间降低了 20.79%,准确率降低了 3.44%,说明轻量型网络能够有效降低模型的计算量,但会损失一定的检测精度;方案 3 在方案 2 的基础上对 MobileNetV3 的 bneck 模块进行了改进,并对其结构进行了调整,与方案 2 相比训练时间略有增加,但准确率提高了 10.24%,与方案 1 相比训练时间提升 12.04%,准确率提升了 6.62%,说明对 Image-to-Tokens 模块的轻量化改进有效的提高了模型的综合能力;与方

案 1 相比方案 4 中 Ghost Module、残差机制、CA 注意力模块的引入有效的提升了模型的识别精度,但训练时间增加了 21.38%;方案 5 在方案 4 的基础上融合了底层特征与高层语义特征,识别准确率得到了进一步提升.对比方案 1、方案 3 和方案 5,对 Image-to-Tokens 模块和 LeFF 模块的改进均提升了模型的准确率,但方案 3 有效降低了训练时间,方案 5 增加了训练时间.方案 6 为提出的改进后模型,即同时使用方案 3 和方案 5 进行改进,与方案 1 相比模型的识别准确率提升了 9.74%,训练



时间也缩小了 0.31 h.

组合改进策略的模型迭代曲线,如图 13 所示.可以看出,所提模型的损失值更小、收敛速度更快、在验证集上的识别准确率更高、泛化能力更强,进一步验证了对 CeiT 模型一系列改进的有效性.

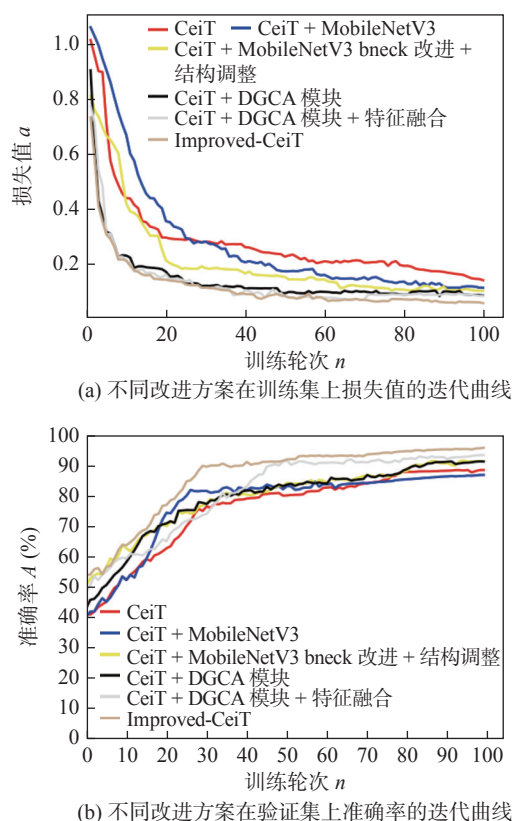


图 13 组合改进策略的模型迭代曲线

Fig. 13 Model iteration curves for combined improvement strategies. (a) Iteration curves of loss values on the training set for different improvement schemes; (b) Iteration curves of the accuracy of different improvement schemes on the validation set

### 3.4 不同检测模型对比试验

为了综合评估所提网络的有效性,将其与改进前的 CeiT 网络、DeiT 网络、MobileNetV3、ResNet

50 和 ShuffleNetV2 模型进行对比试验.利用测试集上获得的精确率、召回率、F1 分数和准确率评估识别效果,使用训练时间和模型内存占用量评估模型的复杂度,使用识别单个样本平均时间评估模型的检测速度.迭代过程中各模型在验证集上的准确率变化如图 14 所示.从图 14 中可以看出,所提模型与其他模型相比在验证集上有最高的识别准确率,而 DeiT 的表现最差.训练好的各个模型在测试集上的识别效果如表 6 所示.由表 6 数据可知,改进的 CeiT 模型和 ResNet50 均获得了较高的检测精度,但非轻量化的 ResNet50 在训练时间、模型内存占用量和识别单个样本平均使用时间要高出很多,无法满足焊接实时性要求.轻量化网络 ShuffleNetV2 和 MobileNetV3 训练时间相当,但 MobileNetV3 的检测精度比 ShuffleNetV2 低,说明在熔池数据集上 ShuffleNetV2 表现更好. DeiT 网络的检测精度最差,原因是其需要海量数据作为支撑,试验数据量远达不到要求,但其训练时间、模型内存占用量和识别单个样本平均时间比 CeiT 要少.改进的 CeiT 网络与原 CeiT 模型相比准确率提高了 8.66%,计算复杂度和检测速度都有了大幅提升.试

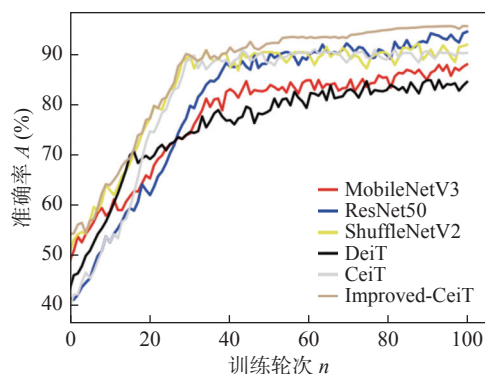


图 14 各对比模型在验证集上准确率的迭代曲线

Fig. 14 Iteration curves of the accuracy of each comparison model on the validation set

表 6 不同模型试验结果

Table 6 Test results of different models

模型	精确率	召回率	F1 分数	准确率	训练时间 $t/h$	模型内存占用量 $\delta/MB$	识别单个样本平均时间 $t'/s$
MobileNetV3	82.2	80.1	81.3	85.1	3.24	17.6	0.16
ResNet50	90.4	89.3	90.0	94.3	6.18	179	0.78
ShuffleNetV2	85.1	83.5	84.8	90.2	3.04	9.85	0.11
DeiT	75.5	73.3	74.2	80.3	3.27	35.8	0.35
tttCeiT	82.9	80.7	81.0	87.8	4.07	44.8	0.47
Improved-CeiT	92.7	90.2	91.6	95.4	3.76	27.7	0.24

验结果表明,所提模型对焊接熔透状态的识别取得了最好的效果。

## 4 结论

(1) 使用 MobileNetV3 作为 Image-to-Tokens 模块的特征提取网络,能够缩减网络模型的参数量和计算量,有效提升实时检测性能。

(2) 通过 Ghost Module、残差机制、注意力模块对 LeFF 模块的改进增强了特征间的远程依赖关系、提升了多尺度特征表达能力、丰富了局部和细节特征信息。

(3) 所提出的模型在试验数据集上与改进前的 CeiT 网络、DeiT 网络、MobileNetV3、ResNet 50 和 ShuffleNetV2 模型进行对比,结果表明所提模型获得了最高的准确率,且检测速度可以满足焊接实时性的要求,展现了较好的有效性和鲁棒性,为焊接熔透状态实时预测提供了新的方法。

## 参考文献

- [1] Wang Y M, Han J, Lu J, *et al.* TIG stainless steel molten pool contour detection and weld width prediction based on res-seg[J]. *Metals-Open Access Metallurgy Journal*, 2020, 10(11): 1495.
- [2] Jiao W, Wang Q, Cheng Y, *et al.* Prediction of weld penetration using dynamic weld pool arc images[J]. *Welding Journal*, 2020, 99(11): 295s – 302s.
- [3] Kim H, Nam K, Oh S, *et al.* Deep-learning-based real-time monitoring of full-penetration laser keyhole welding by using the synchronized coaxial observation method[J]. *Journal of Manufacturing Processes*, 2021, 68(8): 1018 – 1030.
- [4] Wu J Z, Shi J W, Gao Y F, *et al.* Penetration recognition in GTAW welding based on time and spectrum images of arc sound using deep learning method[J]. *Metals*, 2022, 12(9): 1549.
- [5] Ma G H, Yu L S, Yuan H T, *et al.* A vision-based method for lap weld defects monitoring of galvanized steel sheets using convolutional neural network[J]. *Journal of Manufacturing Processes*, 2021, 64: 130 – 139.
- [6] Chang B X, Huang J Y. Discrimination of molten pool penetration based on genetic algorithm optimization of BP neural network[J]. *Journal of Physics Conference Series*, 2020, 1437: 012110.
- [7] 陈宸,周方正,李成龙,等.融合空间和通道特征的等离子弧焊熔池熔透状态预测方法[J].*焊接学报*, 2023, 44(4): 30 – 38.
- Chen Chen, Zhou Fangzheng, Li Chenglong, *et al.* Prediction method of plasma arc welding molten pool melting state based on spatial and channel characteristics[J]. *Transactions of the China Welding Institution*, 2023, 44(4): 30 – 38.
- [8] 王颖,高胜,吴立明.基于胶囊网络的 TIG 熔透预测[J].*焊接*, 2023(4): 15 – 20,28.
- Wang Ying, Gao Sheng, Wu Liming. TIG penetration prediction based on capsule network[J]. *Welding & Joining*, 2023(4): 15 – 20,28.
- [9] Gao Y F, Wang Q S, Xiao J H, *et al.* Weld penetration identification with deep learning method based on auditory spectrum images of arc sounds[J]. *Welding in the World*, 2022, 66(12): 2509 – 2520.
- [10] Yu R W, He H Y, Han J, *et al.* Monitoring of back bead penetration based on temperature sensing and deep learning[J]. *Measurement*, 2022, 188. DOI:10.1016/J.MEASUREMENT.2021.110410.
- [11] 段明瑞.基于深度学习的乏燃料池不锈钢 GTAW 焊接质量在线监测[D].哈尔滨:哈尔滨工业大学,2021.
- Duan Mingrui. On-line monitoring of welding quality of stainless steel GTAW in spent fuel pool based on deep learning [D]. Harbin: Harbin Institute of Technology, 2021.
- [12] Liu S K, Wu D, Luo Z Y, *et al.* Measurement of pulsed laser welding penetration based on keyhole dynamics and deep learning approach[J]. *Measurement*, 2022, 199. DOI:10.1016/J.MEASUREMENT.2022.111579.
- [13] Nomura K, Fukushima K, Matsumura T, *et al.* Burn-through prediction and weld depth estimation by deep learning model monitoring the molten pool in gas metal arc welding with gap fluctuation[J]. *Journal of Manufacturing Processes*, 2020. DOI:10.1016/j.jmapro.2020.10.019.
- [14] Xia C Y, Pan Z X, Fei Z Y, *et al.* Vision based defects detection for Keyhole TIG welding using deep learning with visual explanation[J]. *Journal of Manufacturing Processes*, 2020, 56(8): 845 – 855.
- [15] 刘天元,鲍劲松,汪俊亮,等.融合时序信息的激光焊接熔透状态识别方法[J].*中国激光*, 2021, 48(6): 228 – 238.
- Liu Tianyuan, Bao Jinsong, Wang Junliang, *et al.* Laser welding penetration state recognition method fused with timing information[J]. *Chinese Journal of Lasers*, 2021, 48(6): 228 – 238.
- [16] 卢振洋,宫兆辉,闫志鸿,等.基于深度学习的 TIG 焊背部熔池检测和熔宽提取[J].*北京工业大学学报*, 2020, 46(9): 988 – 996.
- Lu Zhenyang, Gong Zhaohui, Yan Zhihong, *et al.* Deep learning based detection and width extraction of back molten pool in TIG welding[J]. *Journal of Beijing University of Technology*, 2020, 46(9): 988 – 996.
- [17] Wang Z M, Li L Y, Chen H Y, *et al.* Recognition of GTAW weld penetration based on the lightweight model and transfer learning[J]. *Welding in the World*, 2022, 67(1): 251 – 264.

- notch bar impact testing of metallic materials: ASTM E23 [S]. West Conshohocken, Pa: ASTM International, 2013.
- [19] Zhang J C, Di H S, Deng Y G, *et al.* Effect of martensite morphology and volume fraction on strain hardening and fracture behavior of martensite–ferrite dual phase steel[J]. *Materials Science and Engineering A*, 2015, 627: 230 – 240.
- [20] 赵洋洋, 林可欣, 王颖, 等. 基于位错模型的增材制造构件疲劳裂纹萌生行为 [J]. *焊接学报*, 2023, 44(7): 1 – 8.
- Zhao Yangyang, Lin Kexin, Wang Ying, *et al.* Fatigue crack initiation behavior of additive manufacturing components based on dislocation model[J]. *Transactions of the China Welding Institution*, 2023, 44(7): 1 – 8.
- [21] 田成川, 赵海, 田妮, 等. 长期服役对 P91 钢蒸汽管道接头疲劳裂纹扩展行为的影响 [J]. *材料与冶金学报*, 2023, 22(6): 588 – 594.
- Tian Chengchuan, Zhao Hai, Tian Ni, *et al.* Effect of long-term service on fatigue crack propagation behavior of P91 steel steam pipe joint[J]. *Journal of Materials and Metallurgy*, 2023, 22(6): 588 – 594.

---

**第一作者:**张超, 硕士; 主要从事低活化铁素体/马氏体钢、耐热钢搅拌摩擦焊连接机理、铝合金高能束焊接技术以及固相连接技术的相关研究工作; Email: 569405731@qq.com.

(编辑: 郑红)

---

[ 上接第 35 页 ]

- [18] Li C, Wang Q, Jiao W, *et al.* Deep learning-based detection of penetration from weld pool reflection images[J]. *Welding Journal*, 2020, 99(9): 239s – 245s.
- [19] Liu Z, Lin Y, Cao Y, *et al.* Swin transformer: Hierarchical vision transformer using shifted windows[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021: 10012 – 10022.
- [20] Liang S J, Yu M X, Lu W S, *et al.* A lightweight vision transformer with symmetric modules for vision tasks[J]. *Intelligent Data Analysis*, 2023, 27(6): 1741 – 1757.
- [21] Yuan K, Guo S, Liu Z, *et al.* Incorporating convolution designs into visual transformers[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021: 559 – 568.
- [22] Howard A, Sandler M, Chu G, *et al.* Searching for mobilenetv3[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 1314-1324.

---

**第一作者:**王颖, 硕士, 副教授; 主要研究方向为智能焊接、先进机器人工程、深度学习、人工智能等; Email: nepu\_wy@163.com.

(编辑: 郑红)