

DOI:10.19651/j.cnki.emt.2519011

ArgusFusion: 基于 MLP 的轻量化 高效的氩花分割网络*

李豆 王静宇 任国印 褚佳兴

(内蒙古科技大学数智产业学院(网络安全学院) 包头 014010)

摘要: 钢包底吹氩是炼钢生产的关键环节,其中钢液的裸露面积(氩花)是评估底吹效果的重要依据。为实现氩花的量化分析,引入了图像分割技术。然而,现有的分割网络普遍存在参数量大,计算机资源要求高,分割精度不足等问题,无法满足工业生产中实时性和高效性的要求。为此,本文提出了一种创新性的氩花分割网络 ArgusFusion。该网络采用 U 形结构,在特征提取和重建阶段,结合卷积模块与创新性的注意力机制(Glo-MLP attention)实现高效信息交换。瓶颈层引入改进的多分支混合模块(MACA-Mixer)以增强特征表达能力。此外,跳跃连接中引入自适应层级特征融合架构(AHFF)优化边缘分割。实验结果表明,ArgusFusion 在工业氩花数据集上以 0.51 M 参数量,1.38 GFLOPs 计算量,实现 88.90%IoU 精度,具备高分割精度和低资源消耗优势,完全满足工业实时性要求。

关键词: 图像分割;氩花;实时性;高效性

中图分类号: TP399;TN911 **文献标识码:** A **国家标准学科分类代码:** 520.6040

ArgusFusion: A lightweight and efficient argon flower segmentation network based on MLP

Li Dou Wang Jingyu Ren Guoyin Chu Jiaying

(School of Digital and Intelligent Industry (School of Cyber Science and Technology), Inner Mongolia University of Science and Technology, Baotou 014010, China)

Abstract: Bottom argon blowing in the ladle is a critical step in steelmaking, where the exposed surface area of molten steel (argon flower) serves as an important indicator for evaluating the blowing efficiency. To achieve quantitative analysis of the argon flower, image segmentation techniques are employed. However, existing segmentation networks generally suffer from large parameter sizes, high computational resource requirements, and limited segmentation accuracy, making them unable to meet the real-time and efficiency demands of industrial production. This paper proposes an innovative argon flower segmentation network named ArgusFusion. The network adopts a U-shaped architecture and integrates convolutional modules with a novel Global Multi-Layer Perceptron-based Attention (Glo-MLP attention) mechanism during feature extraction and reconstruction stages to facilitate efficient information exchange. In the bottleneck layer, an improved Multi-Scale Channel Attention Mixer (MACA-Mixer) is introduced to enhance feature representation. Additionally, a Adaptive Hierarchical Feature Fusion (AHFF) is incorporated into the skip connections to optimize boundary segmentation. Experimental results on an industrial argon flower dataset demonstrate that ArgusFusion achieves an IoU of 88.90% with only 0.51 M parameters and 1.38 GFLOPs, showcasing high segmentation accuracy and low computational cost, fully meeting the requirements of real-time industrial applications.

Keywords: image segmentation; argon flower; real-time performance; efficiency

0 引言

在炉外精炼过程中,向钢包底部吹入氩气是确保钢水

质量的关键性步骤。这一步骤中需要人工监督,操作人员仅凭直观的视觉观察氩花的面积(钢水的裸露面积)并计算氩花占比来调整氩气的吹入流量^[1]。然而,这种单靠人工

收稿日期:2025-06-04

* 基金项目:国家自然科学基金(62466045)、2022年第三批自治区重点研发和成果转化计划项目(社会公益)(2022YFSH0044)、内蒙古自然科学基金(2024QN06012)项目资助

经验的方法不仅容易出现精度上的偏差,而且也不能实现实时调节;这种方法无法达到计算机控制的精准和速度。随着钢铁行业对产品质量的日益重视,传统的依赖人工经验的手动调节方式已逐渐显示出其不足之处,因此迫切需要引入更为智能和精确的控制技术来满足现代生产的需求^[2]。

为了解决钢包底部吹氩操作的实时性差和人工调节吹氩气流量不准确的问题,引入图像采集技术对钢包底吹氩的操作场景进行实时监控成为了一种可行的解决方案。通过图像分割技术,系统可以自动识别钢渣和钢水区域,并计算出氩花的面积^[3]。基于此,系统及时调整氩气流量,从而提高流量控制的精确性并减少人为干预的误差^[4-6]。因此,深入研究钢包底吹氩图像分割技术对钢铁智能化发展具有重要意义。

在图像分割领域中,卷积神经网络(convolutional neural network, CNN)面临空间信息丢失的问题。全卷积神经网络(fully convolutional networks, FCN)通过使用卷积层替换全连接层并采用反卷积操作进行上采样。随后, Ronneberger 等^[7]提出了 U-Net 网络,其编码-解码结构和跳跃连接设计,使得更多空间细节信息得以保留,成为炼钢图像分割的关键网络。在 U-Net 的基础上,后续的变体模型如 UNet++^[8]、Attention UNet^[9]、3D UNet 等^[10],通过引入更多的跳跃连接、注意力机制和 3D 卷积等方法,使其在炼钢环境下的适应性和分割效果更加优越。U-Net 及其变体在炼钢或其他领域中被广泛应用^[11-12]。现在 U-Net 的变体仍在不断发展来适应更高精度或者其他炼钢的任务需求。2024 年研究人员提出了 Mamba-UNet^[13],它的设计结合了 U-Net 和 Swin-UNet 网络结构^[14],并且引入了 Visual Mamba 块来增强长距离依赖关系的建模能力,进一步提升了在炼钢和氩花图像分割任务中的适用性。

为了克服 U-Net 以及其他卷积神经网络在炼钢分割任务中处理长程依赖和全局特征建模的局限性,研究人员提出了基于 Transformer 的网络。然而,传统 Transformer 直接应用于视觉任务时,因缺乏空间信息建模和对大规模图像处理的效率问题而面临挑战。为此,2020 年 Dosovitskiy 等^[15]提出视觉转换器(vision transformer, ViT)将图像切分为小块,并将其作为输入序列引入 Transformer 结构,首次展示了 Transformer 在视觉任务中的潜力。在炼钢行业中,图像分割的准确性直接影响到钢铁表面缺陷的识别和生产效率的提升^[16]。因此,研究人员越来越关注 Transformer 在这一领域的应用,期望通过其强大的特征建模能力来提升炼钢分割任务的效果。然而,ViT 对局部炼钢图像特征捕捉能力较弱,且对大数据的依赖较大。为此,后续提出了多种变体:Swin Transformer 通过引入局部窗口和跨窗口移动机制^[17],增强了对钢铁表面细微缺陷和整体结构的建模能力;SegFormer 结合了多尺度特征提取与高效自注意力机制^[18],为精确识别不同尺寸的

炉料缺陷提供了可能;TransUNet 融合了 U-Net 的编码-解码结构与 Transformer 的全局建模能力,提升了对炼钢图像中细节和全局信息的捕捉能力;这些网络结构虽然在一定程度上提升了分割的性能,但是牺牲了其参数量和增加了计算的复杂程度,大大增加能耗和计算机成本,难以应用在一些实时性要求比较高和资源有限的工业场景。

为了解决以上问题,更好地满足工业生产对高效,精确分割的需求,本文设计出一种适用于工业环境的轻量且高效的分割网络 ArgusFusion,以更好地应对吹氩图像分割的独特挑战。

1 ArgusFusion

1.1 网络总体结构

ArgusFusion 是一个对称的 U 形网络,主要由全局多层感知机注意力模块(global multi-layer perceptron-based attention, Glo-MLP Attention)、卷积模块、自适应层级特征融合架构(adaptive hierarchical feature fusion, AHFF)以及瓶颈区的多分支混合模块(multi-scale channel attention mixer, MACA-Mixer)组成,网络的结构如图 1 所示。输入的氩花图像首先经过 3 个卷积模块,以提取关键信息。每经过一个卷积模块,网络会进行下采样,从而进一步细化特征表达;特征图的尺寸减半,通道数则加倍。随后,图像通过两个 Glo-MLP Attention 模块,以实现全局信息交换。在编码器的末端,使用 MACA-Mixer 进行多尺度特征聚合,进一步增强特征的表达能力。解码器部分首先通过两个 Glo-MLP Attention 模块,增强其泛化能力。接着,通过多层卷积逐步进行上采样,并结合特征融合模块 AHFF,将编码器中的特征逐步恢复。最终,网络输出高精度的分割预测结果。

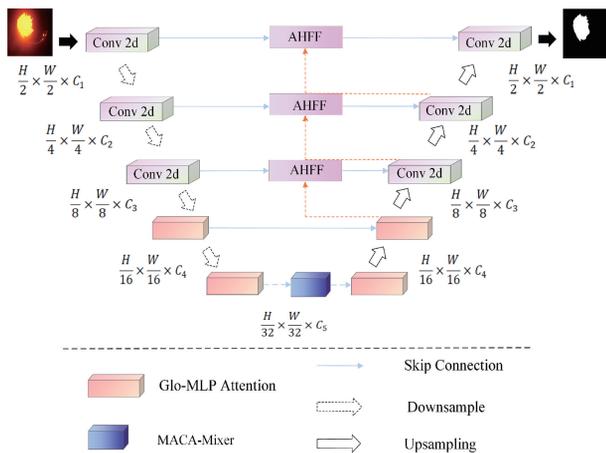


图 1 ArgusFusion 总体架构

Fig. 1 Overall architecture of ArgusFusion

1.2 自适应层级特征融合架构 AHFF

借鉴医学图像处理中多尺度特征融合的先验理念,本

研究创新性地提出了自适应层级特征融合架构^[19]。该模块通过结合输入特征与引导特征来增强氩花边界信息,从而显著提升了网络在解析图像时对关键特征的高层次表达能力。AHFF 的核心在于扩散过程,通过求解一个偏微分方程来平滑特征图如式(1)所示。

$$\frac{\partial F}{\partial t} = D \cdot \nabla^2 F \quad (1)$$

式中: F 表示特征图, D 是扩散系数, ∇^2 是拉普拉斯算子。这个方程表示特征图的信息会根据语义差异进行扩散和平滑。

由于需要提供语义信息, G 是较深层网络生成的特征图来引入到扩散函数中,所以采用了 ∇G 的差分图来作为语义引导图反映整个像素空间中的每个像素的变化情况。具体来说如果两个像素之间的特征差异较大,那么在这些位置上的扩散过程会受到抑制;相反,如果特征差异较小,扩散过程则会促进像素特征的平滑,如式(2)所示。

$$D = h(|\nabla G|^2) \quad (2)$$

其中, $h(|\nabla G|^2)$ 模拟扩散过程的非线性特征来控制不同区域的扩散速度变化, h 是用于投影的卷积函数。基于上述式(1)和(2),特征图会根据扩散过程进行更新,具体迭代式(3)所示。

$$\hat{F}_p^{t+1} = \sum_{p' \in \delta p} h(|G_{p'} - G_p|^2) \cdot (F_{p'}^t - F_p^t) \quad (3)$$

式中: p 表示特征图中的像素索引, δp 是以像素 p 为中心的 2×2 的二维领域, G_p 和 $G_{p'}$ 分别是像素点 p 及其领域中像素点的语义特征。 $(|G_{p'} - G_p|^2)$ 代表更精细化的特征。 $(F_{p'}^t - F_p^t)$ 是邻域像素 p' 和中心像素 p 在第 t 次迭代时的特征差异。 \hat{F}_p^{t+1} 表示像素 p 在下一迭代更新的特征值。这代表经过扩散过程之后,像素 p 的特征值如何被领域的像素特征所影响。只考虑扩散过程预测的边界不够准确,因此加入了 $(|G_{p'} - G_p|^2)$ 生成更加精细化的特征。最终的特征图更新公式为:

$$F_p^{t+1} = \lambda \cdot F_p^t + \nu \cdot \hat{F}_p^{t+1} \quad (4)$$

其中, λ 和 ν 是权重系数, F_p^{t+1} 是像素 p 的新特征值,由两部分加权求和得到。第一部分保留了当前像素的部分原始特征 F_p^t ,防止在迭代的过程中丢失原始信息;第二部分根据领域像素的特征差异对像素 p 进行扩散更新。这个扩散过程在图像的边界区域会受到限制,以保留边界特征,在其他区域则进行平滑。

为了加强特征信息的获取,模块通过显式和隐式差分核来更新特征图的过程,如式(5)所示。该式结合了卷积操作和差分信息,用于增强特征图中边界信息的表达。

$$F_p^{t+1} = \sum_{p' \in \delta p} \omega_{p'} \cdot |\alpha_{p'} G_{p'} - G_p|^2 \cdot (\beta_{p'} F_{p'}^t - F_p^t) \quad (5)$$

式中: $\alpha_{p'}$ 和 $\beta_{p'}$ 是控制差分操作的参数,这些参数控制了特征图中不同位置的语义差异和特征差异如何影响扩散。 $\omega_{p'}$

是卷积核中的权重,其中中心值为 -1 ,其他位置的值通过训练学习得到。这个负值的中心是为了强调差分操作,确保中心像素和领域的对比能够更好地捕捉到变化。如图 2 所示为 AHFF 的详细结构,先对原始特征进行运算生成差分特征,然后引入深层特征的语义信息,将原始差分特征细化为增强特征。

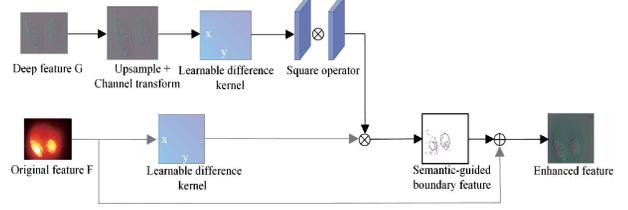


图 2 AHFF 模块结构

Fig. 2 Structural of the AHFF module

1.3 Glo-MLP Attention

Glo-MLP Attention 模块是一种创新性的注意力模块,其设计初衷是为了更好地捕捉氩花图像中特征的多样性和细节,从而提升网络对复杂任务的适应能力,结构如图 3 所示。该模块将输入图像划分为多个非重叠的图像片段(补丁),然后将这些片段投影到所需的隐藏维度 C 来进行特征提取。假设输入特征图大小为 $H \times W$,补丁大小为 $P \times P$,步幅为 S ,则可以用式(6)来计算补丁的数量 $num_patches$:

$$num_patches = \left(\frac{H-P}{S} + 1\right) \times \left(\frac{W-P}{S} + 1\right) \quad (6)$$

之后通过卷积层将每个补丁的像素值投影到新的特征空间,生成嵌入特征。再将每个嵌入特征展平得到 X_i 以便后续处理。接着对每个特征进行归一化后通过 Glo-MLP 块得到 Y_i ,如式(8)所示。再将原始嵌入特征 X_i 与 MLP 输出后的特征 Y_i 相加得到最终输出 $X_{out,i}$,如式(9)所示。

$$X_{i,norm} = (LayerNorm(X_i)) \quad i = 1, \dots, s \quad (7)$$

其中, $LayerNorm$ 是对单个样本的所有特征进行归一化操作, X_i 为原始嵌入特征。

$$Y_i = Glo-MLP(X_{i,norm}) \quad (8)$$

$$X_{out,i} = Y_i + X_i \quad (9)$$

之后通过全局平均池化和最大池化,学习到各个通道的重要性权重,进而增强关键信道的特征表达能力,如式(10)所示。

$$X_{out} = \sigma(Glo-MLP(AvgPool(X_{out,i}))) + Glo-MLP(MaxPool(X_{out,i})) \quad (10)$$

式中: σ 是激活函数, $AvgPool$ 为全局平均池化操作, $MaxPool$ 为最大池化操作。

接着识别重要区域,进一步提升这些区域的特征表达,如式(11)所示。

$$Y_{out} = \sigma(Concat(AvgPool(X_{out}), MaxPool(Y_{out}))) \quad (11)$$

最终将 X_{out} 和 Y_{out} 相乘之后得到了注意力增强的特征

图 Y :

$$Y = X_{out} \cdot Y_{out} \quad (12)$$

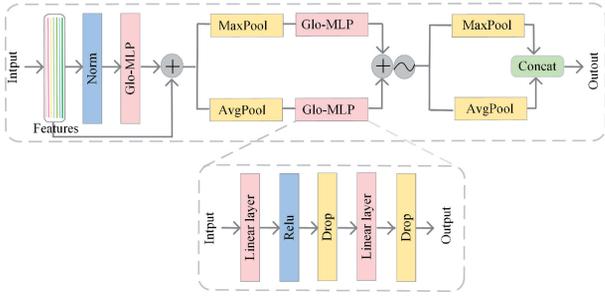


图 3 Glo-MLP Attention 结构

Fig. 3 Structural of Glo-MLP Attention

1.4 MACA-Mixer 模块

为了高效且轻量地建模图像中的多尺度上下文信息与通道注意力特征,提出了一种创新的低计算开销的多分支混合模块 MACA-Mixe^[20],如图 4 所示。该设计通过二值化和多尺度卷积结合,在降低计算机复杂度的同时保留了特征表达能力。该模块采用三支并行结构处理输入特征图 $X \in \mathbb{R}^{B \times C \times H \times W}$,每个分支首先通过通道注意力模块(channel attention mechanism, CAM)模块增强特征通道的依赖性:

$$X^{(i)} = CAM(X), i = 1, 2, 3 \quad (13)$$

然后使用二值函数 $Ba()$ 对特征进行量化,其中 $Ba()$ 的反向传播采用分段多项式梯度近似确保在二进制训练中梯度有效传递,避免因符号函数不可导导致的优化问题。量化后的特征分别经过不同膨胀率(1,3,5)的二值化 3×3 分组卷积层进行多尺度特征提取,之后将卷积层的输出通过激活函数增强非线性表达能力,最后将三个分组卷积层的输出特征经批量归一化层 bn 融合相加,得到最终的特征表示 Y 。这种设计通过二值化大幅降低了计算复杂度,同时利用多尺度卷积和通道注意力机制保持了模型的表征能力,实现了计算效率与特征表达能力的平衡。

$$Ba(X^{(i)}, a, b) = \begin{cases} +1, \frac{X^{(i)} - b}{a} \geq 0 \\ -1, \text{其他} \end{cases} \quad (14)$$

$$\frac{\partial Ba}{\partial X^{(i)}} = \begin{cases} 2 + 2(\frac{X^{(i)} - b}{a}), b - a \leq X^{(i)} < b \\ 2 - 2(\frac{X^{(i)} - b}{a}), b \leq X^{(i)} < b + a \\ 0, \text{其他} \end{cases} \quad (15)$$

其中, a 和 b 为可学习的尺度因子和偏置,通过训练动态调整二值化阈值。

$$X_{Bi}^{(i)} = PRRELU(Bi_Conv_{3 \times 3}^{1,3,5}(Ba(X^{(i)}, a, b) + X^{(i)})) \quad (16)$$

式中: $PRRELU$ 为激活函数, $Bi_Conv_{3 \times 3}^{1,3,5}$ 为不同膨胀率的二值化 3×3 分组卷积。

$$Y = bn(X_{Bi}^{(1)} + X_{Bi}^{(2)} + X_{Bi}^{(3)}) \quad (17)$$

CAM 是一种创新的神经网络模块,旨在增强特征图中每个通道的重要性。该模块通过全局平均池化提取通道特征,并利用一个独立的多层感知机(MLP)计算每个通道的权重,最后通过逐元素相乘将这些权重应用于原始特征图。

输入特征图 $F \in \mathbb{R}^{B \times C \times H \times W}$,其中 B 是批量大小, C 为输入通道数, H 和 W 是特征图的高度和宽度,接着通过全局平均池化得到每个通道的平均值:

$$avg_out = AvgPool2d(F) \quad (18)$$

式中: $AvgPool2d$ 为全局平均池化。接着将池化后的特征图展平为形状 (B, C) ,得到 $avg_out_flattened$,以便输入到 MLP。该 MLP 由两个全连接层 $Linear$ 组成,整体流程可表示为:

$$z = Relu(Linear(avg_out_flattened)) \quad (19)$$

其中, $Relu$ 为激活函数。然后通过第二层全连接层 $Linear$ 和激活函数计算通道权重:

$$weight = Sigmoid(Linear(z)) \quad (20)$$

式中: $Sigmoid$ 为激活函数。最后计算得到的权重 $weight$ 被重塑为形状 $(B, C, 1, 1)$,并通过逐元素相乘将其应用于输入特征图,得到输出的特征图:

$$output = F \times weight \quad (21)$$

这种结构使得模型减少参数数量的同时有效强化重要通道的信息,并抑制不重要通道,从而提高整体性能。

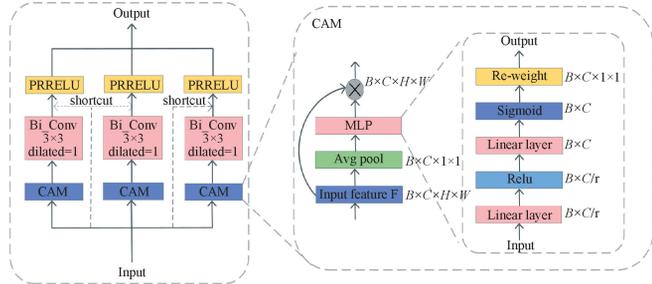


图 4 MACA-Mixer 模块结构

Fig. 4 Structural of the MACA-Mixer module

2 实验设置

2.1 数据集

本文所用数据集采集于某吹氩工业现场,全过程严格遵循实际操作规范。该现场配备了高精度的监控摄像头,能够在不同吹氩条件下连续稳定地捕获图像。采集过程中,摄像头被固定在吹氩作业区域上方,以确保能够全面覆盖整个操作区域。所采集的数据集由 803 张图像组成(分辨率均为 1280×720),这些图像是在不同的生产周期和工作时段内获取的,涵盖各种图像以反映吹氩过程的多变性和复杂性。

在像素层面,氩花区域作为关键目标,通常仅占原始图像总像素的约 $5\% \sim 50\%$,其余多为高温背景或钢包内壁

等结构,呈现出明显的前景稀疏。在样本层面,为进一步分析氩花图像在数据集中的分布情况,将原始数据集中的 803 张图像根据氩花面积及其形态划分为 4 类:面积偏小、面积正常、面积偏大和异常情况,如图 5 所示。统计结果表明,面积正常的氩花图像占比约为 36%,面积偏小约占 38%、面积偏大约占 6%、异常情况(包括烟雾干扰、孔堵塞及强光干扰)约为 20%。其中,面积偏小样本所占比例最高,主要源于实际生产过程中常出现流量不足、喷嘴堵塞或局部阻力增大等工况,致使氩气难以充分扩散,从而形成小面积氩花。这种类别分布的不均衡性在一定程度上反映了真实工况中的典型特征,尤其是面积偏小与正常状态在生产实际中更为常见。

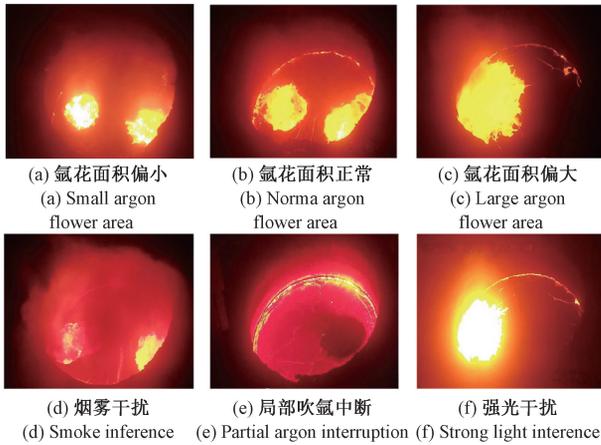


图 5 氩花图像

Fig. 5 Argon flower images

由于氩花图像的核心特征主要体现在形态和边缘轮廓上,本文采用了几何增强策略对数据集进行扩充,以提升模型的鲁棒性。具体增强方式主要包括图像的旋转(90° , 180°)和水平、垂直翻转方法。扩充后数据总量提升至 1 232 张,样本的占比调整为:面积正常为 32%,面积偏小约为 31%,面积偏大为 12%,异常情况为 25%。尽管各类仍存在一定不均衡,但整体分布已更加合理,有助于提升模型对边缘类别的学习能力。数据集按 8:2 的比例划分为训练集和验证集。随后,为统一模型输入尺寸,所有图像均被缩放至 $512 \text{ pixel} \times 512 \text{ pixel}$ 。图 6 展示了扩充后的部分样本图像。

2.2 实验细节

本文所有实验都是在单个 Nvidia GeForce RTX 3070 8 G GPU 上进行的,总共训练了 50 个 epoch。为了缓解样本中部分图片中前景背景(面积)不平衡带来的消极影响,本文选用 Dice Loss 损失函数,同时网络结合 Adam 优化器进行训练,学习率设置为 0.001,动量为 0.9,批量大小设置为 4。

2.3 评价指标

对于评价指标,本文使用交并比(intersection over

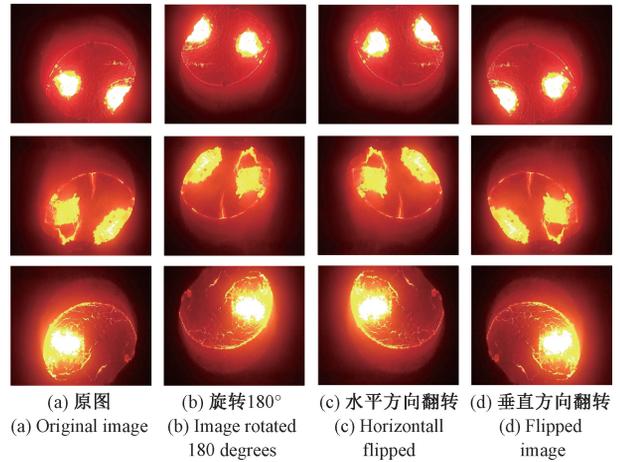


图 6 部分数据集扩充结果

Fig. 6 Some results of dataset augmentation

union, IoU), F1 分数(F1 Score, F1), Recall 来衡量模型分割的准确性;使用计算复杂度,模型参数量和推理速度来衡量模型的轻量化效果。IoU 和 F1 的计算公式如下:

$$IoU = \frac{TP}{TP + FP + FN} \quad (22)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (23)$$

$$Recall = \frac{TP}{TP + FN} \quad (24)$$

其中, TP 是模型正确地将正类样本预测为正类的数量, FP 是模型错误地将负类样本预测为正类的数量, FN 是模型错误地将正类样本预测为负类数量。

此外,计算复杂度,模型参数量(params size, Params)和推理速度(inference speed, Speed)的计算公式如下:

$$GFLOPs = W \times H \times K \times K \times C_{in} \times C_{out} \quad (25)$$

$$Params = C_{in} \times C_{out} \times K \times K \quad (26)$$

式中: W, H 分别表示特征图的高和宽, C_{in} 和 C_{out} 分别表示输入和输出特征图的通道数, K 为卷积核的大小。

$$Speed = t_1 + t_2 + t_3 \quad (27)$$

式中: t_1 是模型部署到 GPU 所用的时间, t_2 表示图像处理 123 张图片所用的时间, t_3 表示输出评价结果所需的时间。

图 7 展示了网络在训练过程中的 Loss 和 IoU 指标在训练集和验证集上的变化趋势。图中对原始损失曲线进行了简单移动平均(SMA)和平滑移动平均(EMA)的处理。从图 7 中可以看出,模型的训练损失和验证损失在早期迅速下降,表明网络在前几个 epoch 中能够快速学习到有效特征。在训练进行到大约 20 个 epoch 之后,损失曲线逐渐趋于平稳,表明网络的优化进入了后期的收敛状态。此外,图中下方的 IoU 曲线则显示了分割性能的提升趋势,训练 IoU 和验证 IoU 随着 epoch 增加逐步上升,并在后期趋于稳定。

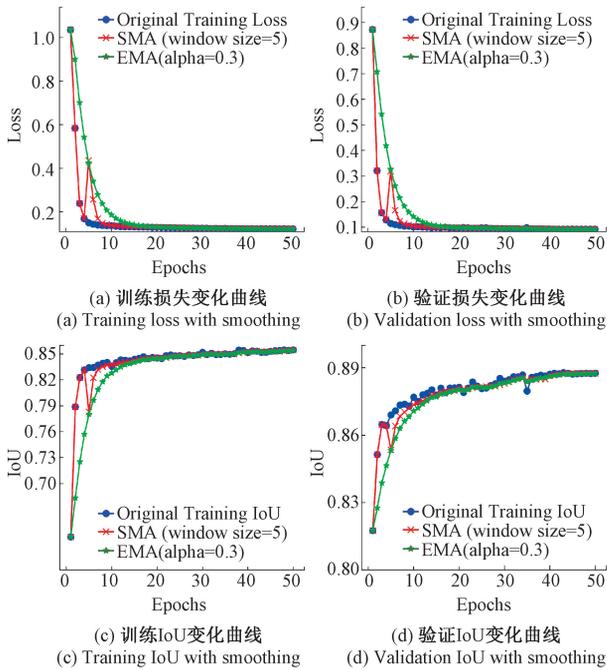


图 7 训练和验证损失与精度趋势

Fig. 7 Trends of training and validation loss and accuracy

3 实验设置

3.1 性能比较

在工业现场采集的氩花数据集中,将 ArgusFusion 与几种比较经典和先进的模型进行了比较,结果如表 1 所示。其中使用 IoU, F1 和 Recall 来衡量分割的质量,使用 Params 和 GFLOPs 来衡量模型的参数量和计算复杂度,最后使用 Speed 来衡量推理速度,Speed 表示在 GPU 上处理 123 张图片所需要的时间,包括模型部署到 GPU,图像处理以及输出评价结果的时间。

表 1 不同网络的性能比较

Table 1 Performance comparison of different networks

Networks	Params /M	GFLOPs	IoU	F1	Speed /s	Recall
U-Net(vgg)	24.00	45.23	86.24	92.61	8.71	91.08
U-Net(resnet50)	43.00	72.42	85.83	92.31	8.50	91.23
DeepLabv3+(mobilenet)	5.81	53.02	82.60	90.47	8.47	88.91
DeepLabv3+(xception)	54.71	54.00	74.79	90.38	8.30	91.74
BiSeNet	34.70	14.55	74.09	84.89	2.49	74.76
VM-UNet	22.03	4.11	72.70	75.48	10.46	75.20
SegFormer	3.71	13.53	87.01	93.04	11.59	92.00
UNeXt	1.47	0.57	88.10	93.64	3.56	93.53
UNeXt-S	0.32	0.10	87.45	93.33	3.17	93.16
ArgusFusion	0.51	1.38	88.90	93.88	3.58	94.69

从表 1 中的实验结果可以看出: ArgusFusion 在多个关键指标上表现卓越,尤其在 IoU, F1, Recall 上均达到了最高水平。相比于 U-Net, DeepLabV3, BiSeNet, VM-UNet, 本网络大大降低了参数量和计算量,提高了推理速度,保证了模型的轻量化。可以看出 ArgusFusion 在保持相对较小的模型规模的同时,展现出了更高的准确率和更出色的边界细节捕捉能力,使其成为工业环境分割任务的理想选择。

本网络在氩花图像中选取了几张具有挑战性的图像,不同模型的分割效果如图 8 所示。其中方框突出了本网络比其他网络表现更好的区域。可以看出, VM-UNet 和 BiSeNet 在氩花样本数据集中没有表现出分割优势,而 DeepLabV3(mobilenet)、U-Net(vgg)、UNeXt-S、UNeXt 没有正确处理上下文语义信息,导致位置信息丢失。SegFormer, UNeXt 可以得到近似正确的分割轮廓,但不能正确处理氩花的边缘信息。相比之下, ArgusFusion 在具有挑战性的数据集上取得了很好的分割效果。

3.2 ArgusFusion 的消融实验

本研究进行了消融实验,以了解 ArgusFusion 中每个模块的影响,如表 2 所示。该模型在编码器和解码器引入二维卷积层(2D convolutional layer, Conv2d)来提取特征信息。随后加入 Glo-MLP Attention 模块来加强全局特征信息的捕捉能力。通过消融实验结果可以看出,两个模块的结合极大地提升了模型的综合表现。然而,氩花数据集中的边缘形态较为复杂,为此网络结构引入了 AHFF。此外 MACA-Mixer 通过全局的空间信息聚合,增强了网络对背景和前景区域的区分能力。

3.3 Glo-MLP Attention 模块的消融实验

为验证 Glo-MLP Attention 模块在本任务中的有效性,本文设计了多组消融实验,如表 3 所示,将该模块与当前主流的注意力机制进行了系统性对比。实验结果表明, Glo-MLP Attention 在精度与计算效率之间实现了更优的平衡。其中,该模块在 F1 和 IoU 指标上分别达到 93.51% 和 88.25%,在所有对比方法中性能最为突出。同时,其计算复杂度仅为 0.72 GFLOPs,尽管参数量相对略高,但整体仍属于轻量化设计,具备良好的性能-效率权衡能力。相比之下,卷积块注意模块(convolutional block attention module, CBAM),高效多尺度注意力机制(efficient multi-scale attention, EMA),简单且无参数的注意力模块(simple attention module, SimAM),压缩和激励注意力模块(squeeze-and-excitation attention, SE)的计算复杂度均为 0.94 GFLOPs,存在更高的计算开销,且在关键性能指标(F1 和 IoU)上略有下降,说明这些模块未能充分提升特征表达效果,反而带来一定的计算冗余。因此,在精度要求与资源受限的工业视觉场景下, Glo-MLP Attention 是一种更加适用的注意力设计。

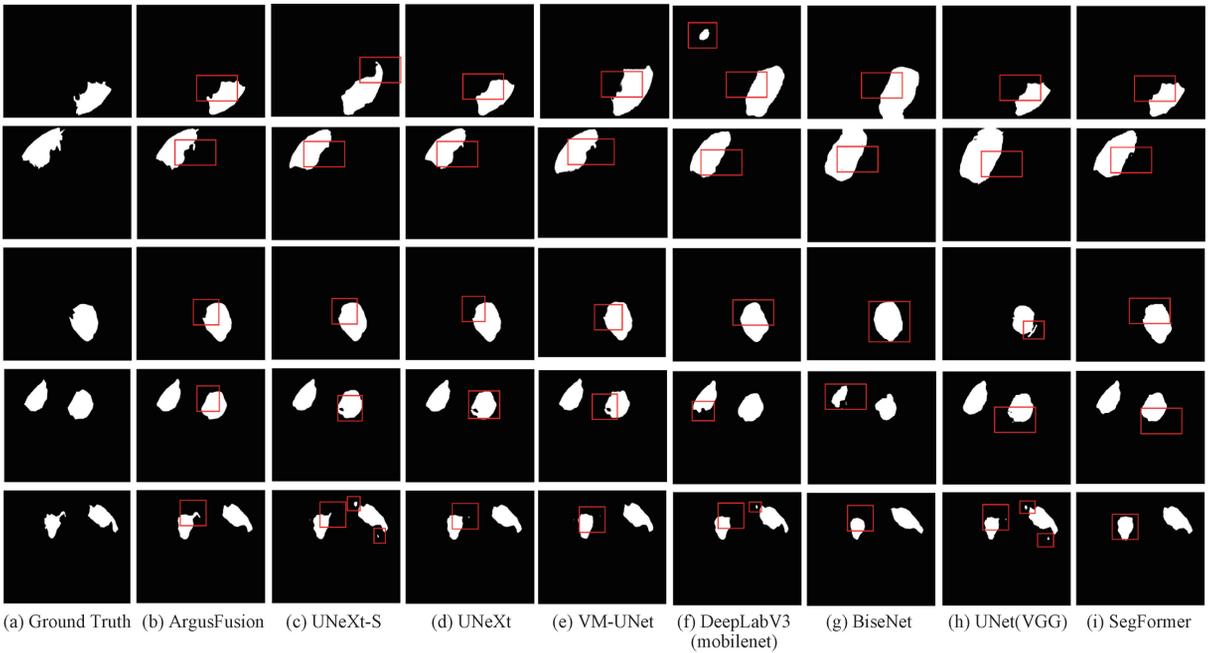


图 8 不同网络的分割结果比较

Fig. 8 Comparison of segmentation results of different networks

表 2 ArgusFusion 的消融实验

Table 2 Ablation experiment of the ArgusFusion

Networks	Params/M	GFLOPs	F1	IoU
Conv2d (Base)	0.12	0.56	92.60	86.85
+ Glo-MLP Attention	0.30	0.72	93.51	88.25
+ AHFF	0.42	1.20	93.72	88.50
+ MACA-Mixer (ArgusFusion)	0.51	1.38	93.88	88.90

表 3 Glo-MLP Attention 的消融实验

Table 3 Ablation experiment of Glo-MLP Attention

Networks	Params/M	GFLOPs	F1	IoU
+CBMA	0.24	0.94	93.35	88.19
+EMA	0.20	0.94	93.50	88.18
+SimAM	0.19	0.93	93.35	88.10
+SE	0.20	0.94	93.45	88.10
+Glo-MLP Attention	0.30	0.72	93.51	88.25

3.4 跳跃连接的分析实验

为进一步验证跳跃连接结构中不同特征融合方式对模型性能的影响,本文分别评估直接连接(skip connection)、CBAM、RepGhost 结构以及所提出的 AHFF 模块在跳跃连接中的应用效果,实验结果如表 4 所示。可以看出,尽管其他跳跃连接方式在参数量和计算复杂度方面表现更为轻量,但其在分割精度方面仍与 AHFF 模块存在一定差距。其中,CBAM 和 RepGhost 虽然在 F1 和 IoU 上略优于

直接连接方案,但其融合能力依然不足,无法充分挖掘浅层与深层特征之间的互补信息。而 AHFF 模块通过引入多尺度上下文信息并融合不同尺度特征,有效提升了特征表达的完整性,体现出更优的跨层特征整合能力。图 9 展示了有无 AHFF 模块的分割效果对比,验证了其在复杂结构边界和细节区域的分割表现的优势。从图 9 中可以直观地观察到,未使用 AHFF 模块的结果存在结构缺失,目标边界模糊不清,而引入 AHFF 后,模型能够更准确地还原目标的整体轮廓与尖锐边缘,进一步说明其在提高分割精度方面发挥了关键作用。

表 4 跳跃连接的分析实验

Table 4 Analysis experiment of skip connection

Networks	Params/M	GFLOPs	F1	IoU
+Skip Connection	0.30	0.72	93.51	88.25
+RepGhost	0.31	0.74	93.64	88.35
+CBMA	0.31	0.72	93.69	88.35
+AHFF	0.42	1.20	93.72	88.50

3.5 MACA-Mixer 模块的消融实验

针对不同的空间池化模块和 CAM 模块的组合方式进行了性能评估。实验选取了 Conv2d, Glo-MLP Attention 和 AHFF 模块作为基础架构,分别引入 4 种不同的池化策略进行对比:金字塔池化模块(pyramid pooling module, PPM)、通道金字塔池化模块(channel attention pyramid pooling module, CAPPMP)、空洞空间卷积金字塔池化模块

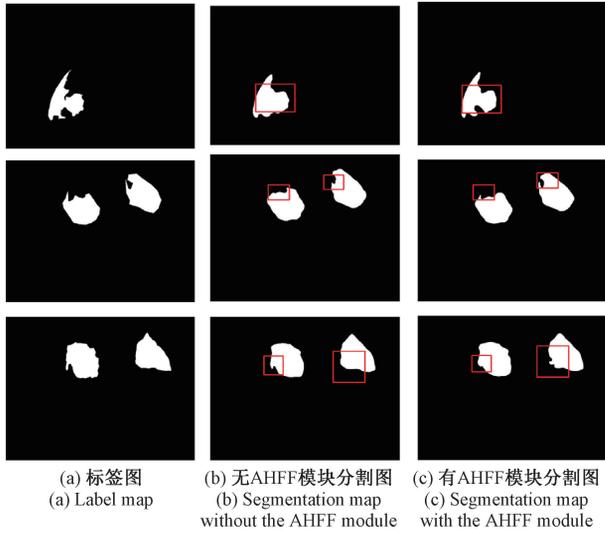


图 9 有无 AHFF 模块分割图对比结果

Fig. 9 Comparison results of segmentation maps with and without the AHFF module

(atrous spatial pyramid pooling, ASPP)和通道空间卷积金字塔池化模块(channel attention spatial pyramid pooling, CASPP)。其中, CAPPMM 与 CASPP 分别是在 PPM 与 ASPP 基础上引入 CAM 模块的改进版本。表 5 实验结果表明, MACA-Mixer 模块在参数量和计算开销都是最优的情况下,同时融合多尺度特征方面表现更出色。同时可以观察到,在相同结构的基础上,集成 CAM 模块的 CAPPMM 和 CASPP 相比原始的 PPM 和 ASPP 均取得了更高的分割精度,说明引入 CAM 能够有效增强模型对关键区域的感知能力,从而提升整体分割性能。该结果验证了 CAM 模块在多尺度融合中的有效性与通用性。

表 5 MACA-Mixer 模块消融实验

Table 5 Ablation experiment of the MACA-Mixer module

Networks	Params/M	GFLOPs	F1	IoU
+PPM	1.61	1.98	93.86	88.70
+CAPPMM	1.70	2.01	93.84	88.80
+ASPP	1.10	1.50	93.81	88.65
+CASPP	1.14	1.70	93.88	88.79
+MACA-Mixer	0.51	1.38	93.88	88.90

4 结 论

工业生产中使用图像分割技术时,对实时性和准确性有着较高要求。针对这一要求,本文提出了基于 U 形网络框架的轻量化结构 ArgusFusion。该网络能够在普通 CPU 上快速运行,有效解决了现有主流网络参数量大和资源限制的问题。ArgusFusion 通过结合卷积, Glo-MLP Attention 模块, AHFF 模块, MACA-Mixer 模块,成功实

现了轻量化的多尺度信息融合,显著提升氩花分割的性能。结果表明, ArgusFusion 具有较高的分割精度,同时保持了较低的模型参数量和计算复杂度,显著提升了模型效率,使其能够更好地适应资源受限的工业环境,在实际应用中展现出卓越的实用价值。

参考文献

- [1] XIN Z CH, ZHANG J S, PENG K X, et al. Modeling of LF refining process: A review[J]. Journal of Iron and Steel Research International, 2024, 31 (2): 289-317.
- [2] CHU K Y, CHEN H H, LAI P H, et al. The effects of bottom blowing gas flow rate distribution during the steelmaking converter process on mixing efficiency[J]. Metallurgical and Materials Transactions B, 2016, 47 (2): 948-962.
- [3] WANG H, WANG M, LIU Q, et al. Research progress on intelligent control and decision-making models for the ladle furnace refining process [J]. Chinese Journal of Engineering, 2024, 46 (10): 1739-1752.
- [4] LIU CH, TANG L X, ZHAO CH CH. A novel dynamic operation optimization method based on multiobjective deep reinforcement learning for steelmaking process[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 35 (3): 3325-3339.
- [5] SHI Q, TANG J, CHU M SH. Key issues and progress of industrial big data-based intelligent blast furnace ironmaking technology [J]. International Journal of Minerals, Metallurgy and Materials, 2023, 30(9): 1651-1666.
- [6] XU W, LIU J J, LI J M, et al. A novel hybrid intelligent model for molten iron temperature forecasting based on machine learning [J]. Aims Math, 2024, 9: 1227-1247.
- [7] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.
- [8] TIE J, WU W B, ZHENG L, et al. Improving walnut images segmentation using modified UNet3 + algorithm[J]. Agriculture, 2024, 14(1): 149.
- [9] ZHANG Y, CHEN J, MA X X, et al. Interactive medical image annotation using improved attention U-net with compound geodesic distance [J]. Expert Systems with Applications, 2024, 237: 121282.
- [10] 包广斌, 罗瞳, 彭璐, 等. 3DuA-Net: 融合 3D 卷积和注

- 意力的雷达回波外推预报[J]. 电子测量技术, 2024, 47(15): 153-160.
- BAO G B, LUO T, PENG L, et al. 3DuA-Net: Fusion of 3D convolution and attention for radar echo extrapolation forecasting[J]. Electronic Measurement Technology, 2024, 47(15): 153-160.
- [11] LI Y, LI Y J, LIU J CH, et al. Research on segmentation of steel surface defect images based on improved Res-UNet network [J]. Journal of Electronics & Information Technology, 2022, 44(5): 1513-1520.
- [12] 刘焯莹, 李超生, 李恭新, 等. CS-SegNet: 新型的脑卒中病灶分割网络[J]. 电子测量技术, 2025, 48(11): 1-11.
- LIU Y Y, LI CH SH, LI G X, et al. CS-SegNet: Novel segmentation network for cerebral stroke lesions [J]. Electronic Measurement Technology, 2025, 48(11): 1-11.
- [13] WU R K, LIU Y H, LIANG P CH, et al. Hvmunet: High-order vision mamba unet for medical image segmentation[J]. Neurocomputing, 2025, 624: 129447.
- [14] YAO J Y, JIN SH G. Multi-category segmentation of Sentinel-2 images based on the Swin UNet method[J]. Remote Sensing, 2022, 14(14): 3382.
- [15] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [J]. ArXiv preprint arXiv: 2010.11929, 2020.
- [16] VASAN V, SRIDHARAN N V, VAITHIYANATHAN S, et al. Detection and classification of surface defects on hot-rolled steel using vision transformers [J]. Heliyon, 2024, 10(19): e38498.
- [17] HUANG J H, FANG Y Y, WU Y ZH, et al. Swin transformer for fast MRI [J]. Neurocomputing, 2022, 493: 281-304.
- [18] XIE E Z, WANG W H, YU ZH D, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers [J]. Advances in Neural Information Processing Systems, 2021, 34: 12077-12090.
- [19] YOU X, DING M, ZHANG M H, et al. Pnpnet: Pull-and-push networks for volumetric segmentation with boundary confusion [J]. ArXiv preprint arXiv: 2312.08323, 2023.
- [20] GAO T, ZHANG Y, ZHANG ZH Y, et al. BHViT: Binarized hybrid vision transformer [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2025: 3563-3572.

作者简介

李豆, 硕士研究生, 主要研究方向为吹氩图像分割, 计算机视觉。

E-mail: 2665410335@qq.com

王静宇 (通信作者), 博士, 教授, 硕士生导师, 主要研究方向为大数据及安全、区块链及安全等。

E-mail: 13734728816@126.com

任国印, 博士, 副教授, 硕士生导师, 主要研究方向为模式识别与智能图像处理等。

褚佳兴, 硕士研究生, 主要研究方向为图像分割。