



丁多,李春秀,金勇,等. CTSegFormer: 一种面向烟丝组分占比测定的语义分割网络模型[J]. 轻工学报, 2026, 41(5): 94-107.
Ding D, Li C X, Jin Y, et al. CTSegFormer: a semantic segmentation network for cut tobacco component proportion measurement[J]. Journal of Light Industry, 2026, 41(5): 94-107.
DOI: 10. 12187/2026. 05. 009

CTSegFormer: 一种面向烟丝组分占比测定的语义分割网络模型

丁多¹, 李春秀², 金勇¹, 辛义定³, 谭格⁴, 王艺斌⁵, 许文武⁵

1. 湖南中烟工业有限责任公司 技术中心, 湖南 长沙 410000;
2. 湖南中烟工业有限责任公司 常德卷烟厂, 湖南 常德 415000;
3. 湖南中烟工业有限责任公司 信息中心, 湖南 长沙 410000;
4. 湖南中烟工业有限责任公司 原料采购中心, 湖南 长沙 410000;
5. 南京焦耳科技有限责任公司, 江苏 南京 210000

摘要:【目的】提升对烟丝组分占比测定的效率和准确性。【方法】提出一种语义分割模型 CTSegFormer, 以 SegFormer 为基础骨架, 设计多类型卷积并行融合、增强的高效加性注意力机制及渐进式深度监督模块, 完成对叶丝、薄片丝、梗丝、膨胀丝 4 类烟丝及梗签的有效识别和分割, 在此基础上实现对烟丝组分占比的准确测定, 并构建烟丝组分测定数据集用于测试与验证。【结果】与现有语义分割模型相比, CTSegFormer 模型在分割和识别烟丝能力上的 mIoU 指标最高, 同时, 在测定多个不同组分配比样品时, 其占比预测相对误差不超过 6%。【结论】提出的 CTSegFormer 模型能较好地适用于烟丝组分占比测定任务, 对监控成品烟支品质有重要意义。

关键词: 烟丝; 组分占比; 语义分割; SegFormer

中图分类号: TS45 **文献标识码:** A **文章编号:** 2096-1553(2026)05-0094-14

0 引言

测定生产线上的烟丝组分, 评估其是否满足工艺标准, 对监控烟丝质量具有重要意义。传统烟丝组分占比测定^[1-2]主要通过人工对烟丝进行筛选和分类, 称重得到质量占比, 测定速度慢, 且由于人工提取的样本量有限, 结果准确性不高。传统图像分

析技术被用于提高测试的准确性和效率, 如刘栋等^[3]通过示踪物表征法, 研究了烟丝的掺配均匀性; 李斌等^[4]通过拟合模型预测了烟丝中的梗丝含量; 寇霄腾等^[5]基于图像处理技术实现了对梗丝掺配比例的定量测定。但这些方法主要依赖特殊的物理抽样和分析, 且只对某一类别烟丝适用, 这导致烟丝类别扩展时, 方法过于复杂。而利用计算机视觉

收稿日期: 2025-06-30; **修回日期:** 2026-02-01; **出版日期:** 2026-10-15

基金项目: 国家自然科学基金面上项目(62272231); 湖南中烟工业有限责任公司科技计划项目(KY2023CG0019、KY2022CG0013)

作者简介: 丁多(1979—), 男, 湖南省常德市人, 湖南中烟工业有限责任公司高级工程师, 主要研究方向为模式识别、烟草工艺与质检。E-mail: dingd1006@hngyobacco.com

通信作者: 李春秀(1973—), 女, 湖南省常德市人, 湖南中烟工业有限责任公司工程师, 主要研究方向为机器视觉、烟草工艺与质检。E-mail: ichx1113@hngyobacco.com

进行多类别烟丝识别的方法较之于这些方法更为简单。杨天卓等^[6]利用高光谱成像技术和机器学习方法预测了烟丝在线掺配比例。然而,高光谱采集设备成本高,且传统的回归模型学习能力有限。

随着机器视觉技术的不断发展,使用深度网络模型实现对烟丝组分占比的精准测定已成为主要发展趋势,深度语义分割网络模型^[7]能准确分割出烟丝区域,并正确识别每条烟丝类别。而Transformer架构^[8-9]在语义分割任务上的性能普遍优于其他架构模型,比如,SegFormer^[10]模型因其良好的上下文信息提取能力和局部信息融合能力,以及模型较为轻量化而备受关注。大量研究工作^[11-16]通过适配和改进 SegFormer,使其更好地应用于遥感图像目标分割、桥梁病害分割和自然场景文本检测等任务。然而,SegFormer 在细粒度烟丝组分的识别方面仍存在一定不足,例如重要空间语义信息丢失,无法有效捕捉烟丝图像中的边缘和细节信息等问题。

鉴于此,本文拟基于 SegFormer 语义分割模型,进一步设计由多特征并行融合(MFPF: Multi-Feature Parallel Fusion)模块、增强的高效加性注意力(EAA: Enhanced Efficient Additive Attention)模块和渐进式深度监督(PDS: Progressive Deep Supervision)模块组成构建的CTSegFormer模型,该模型通过在Segformer模型中引入多个功能模块,利用深度学习丰富的特征学习能力,结合注意力机制和有效学习信息对模型的监督作用,以期获得较高准确性的烟丝组分占比测定结果。

1 CTSegFormer 模型构建

1.1 符号说明

本文符号 $\mathbf{R}$ 表示实数集合, $*$ 表示 2 个实数矩阵对应的逐元素乘积; $\cdot$ 表示实数与实数相乘,或实数与实数向量(或实数矩阵)相乘; $\hat{\mathbf{X}}$ 表示对矩阵或向量 $\mathbf{X}$ 进行 L_1 归一化; K 表示卷积核的尺寸; S 表示卷积核窗口移动步长(stride); P 表示特征图 4 个边界补零(padding)行(或列)数; T 表示参数为 $W \in \mathbf{R}^{d \times d}$ 的线性变换; MLP 为多层感知机模块, Conv 表示卷积层。

1.2 模型架构

CTSegFormer 网络整体架构及对应模块如图 1 所示。由图 1 可知,CTSegFormer 由编码器和解码器组成,这种结构与 U-Net^[20]、SegFormer、SegNet^[21] 等类似,能有效地组合多个阶段不同尺度的特征,帮助模型理解更丰富的上下文语义信息,提高目标定位和分割精度。

编码器由 4 个 MFPF 模块和 4 个并行注意力通道(Parallel Attention Channels, PAC)模块组成,其中 PAC 模块由原始 SegFormer 中的 Transformer 模块和 EEAA 模块 2 个并行分支共同提取长距离上下文信息,每个 MFPF 模块后接 1 个 PAC 模块,构成一个整体,作为 CTSegFormer 结构中的一层。MFPF 模块的输出即为 PAC 模块的输入, PAC 模块对 MFPF 模块提取的局部特征进行注意力计算操作,将结果送入下一个整体模块进行处理,在此过程中 4 个整体模块分别将特征维度缩放到原始输入图像的 1/4, 1/8, 1/16 和 1/32。解码器的 PDS 模块先使用渐进的双线性插值对低分辨率特征进行上采样,再使用 MLP 融合特征信息,上采样后作为最终分割结果,同时使用多层次的监督损失进行训练,4 个 PAC 模块的输出均使用线性层上采样后计算 GHMLoss^[22]。

1.3 MFPF 模块

普通的卷积层受限于卷积核大小,其感受野区域有限,无法捕捉到更广泛的上下文信息,虽然增大卷积核可以缓解该问题,但会增加模型的参数数量,不利于小面积目标的特征提取。对于烟丝分割任务,常因出现小片烟丝,不宜过多使用大尺寸的卷积核。W. H. Wang 等^[23]在 PVT v2 中引入了重叠分块(Overlap Patch, OP)操作,在保留更多局部信息的同时,增强了模型对不同尺寸输入图像的适应能力。虽然 OP 在普通图像分割任务中表现优异,但烟丝语义分割更趋向于细粒度的图像分割,对纹理、形状、颜色、边缘信息等局部特征的提取能力要求更高,单一的重叠卷积难以充分捕捉和表征上述复杂信息。因此,为增强模型的特征提取能力并丰富烟丝局部特征表达,本文提出了 MFPF 模块。该模块采用并行结构,包括 OP 卷积分支、CDC 卷积分支和膨胀卷积分支。由图 1b)可知,OP 卷积分支和 CDC 卷积分支仅包含 1 层卷积层,而 DC 卷积分支

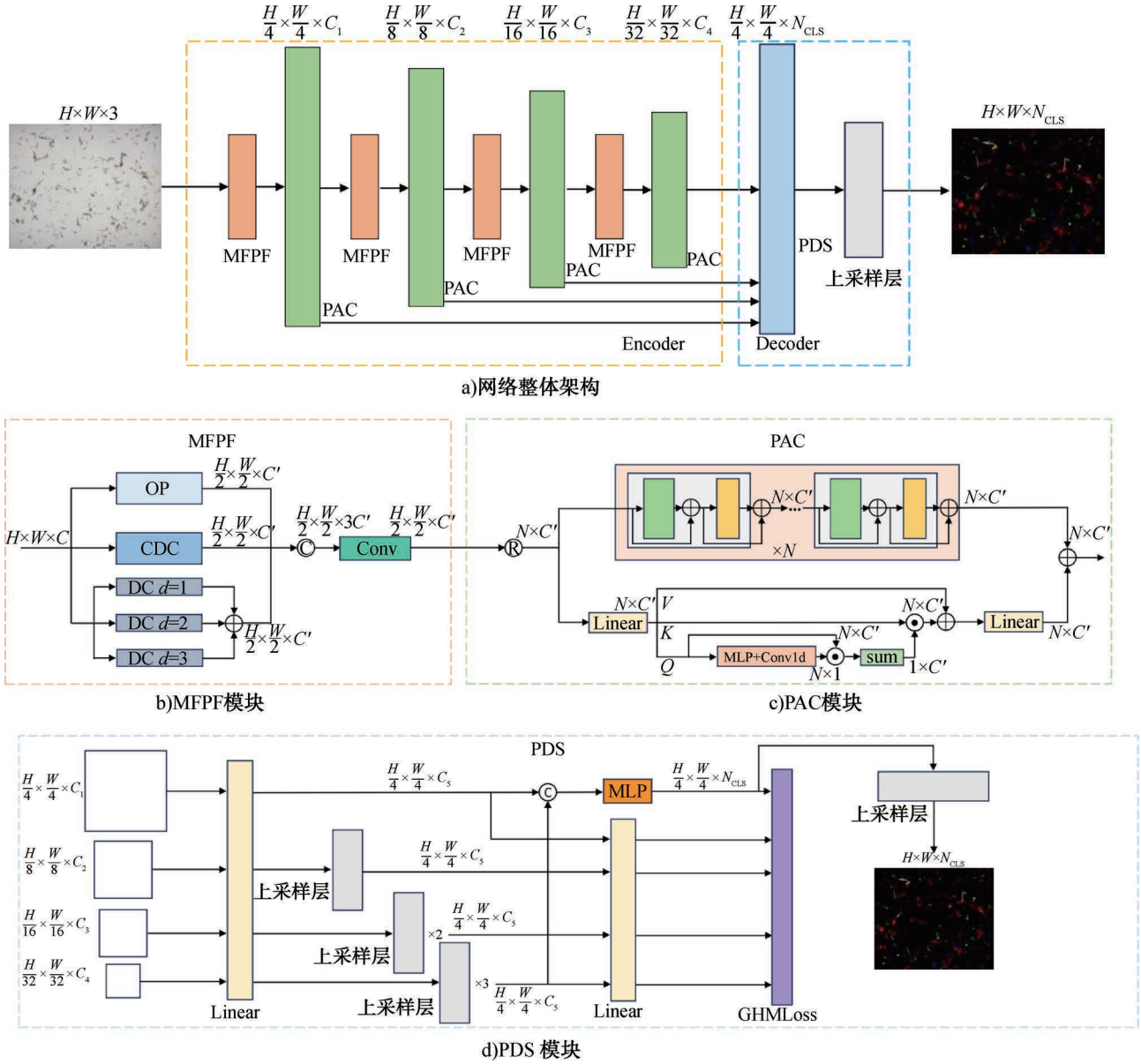


图 1 CTSegFormer 网络整体架构及对应模块

Fig. 1 Overall architecture of the CTSegFormer network and its corresponding modules

由 3 层并行的 DC 卷积组成,膨胀率分别设置为 1、2、3,这 3 条分支的输出与输入特征尺寸均保持相同。最后使用拼接操作对 3 条分支提取的特征进行融合,并使用 1 层卷积层来还原通道数,此外,增强模型对局部及边缘信息的抽取能力。膨胀卷积也能够提供中远距离的位置信息,帮助模型学习更丰富和层次化的特征表示。4 个阶段的 MFPP 模块的具体参数设置见表 1,遵循普遍的经验性设计,在第 1 阶段选用大尺寸卷积核提取特征,更大的感受野有助于在网络早期阶段捕获更多的空间信息,在第 2 阶段到第 4 阶段则是选择大小为 3 的卷积核,帮助模型捕获更深层次的精细特征。

该模块的整个特征提取过程表示如下:

$$x_{out} = \text{Conv}_{1 \times 1} \left(\text{Concat} \left(\text{OP}(x_{in}), \text{CDC}(x_{in}), \sum_{i=0}^2 \text{DC}_i(x_{in}) \right) \right)$$

式中, $x_{in} \in \mathbf{R}^{H \times W \times C}$ 是输入特征, $x_{out} \in \mathbf{R}^{\frac{H}{2} \times \frac{W}{2} \times C'}$ 是输出特征; $\text{Conv}_{1 \times 1}$ 是 1×1 的卷积核,用于融合 3 个卷积的信息; Concat 是通道的拼接操作, $\text{DC}_i (i = 0, 1, 2)$ 表示 3 个 DC 卷积分支,3 个分支相加,加权比例均设为 1。

1.4 PAC 模块

SegFormer 的 Efficient Self-attention 模块通过降低 K, V 特征维度的方式减少自注意机制的计算量,

对于局部特征和边缘信息精细的烟丝会导致测定精度下降。因此,本文提出了 PAC 模块,在仅增加少量模型参数和线性计算量的前提下,来增强 MFPP 模块的特征提取,弥补 Efficient Self-attention 模块性能损失,获取更加有效的特征表示。由图 1c)可知,MFPF 模块输出的特征被送入上下 2 个注意力分支,上方分支与 SegFormer 的 Transformer 模块采用了完全相同的结构,使用 Efficient Self-attention 和 Mix-FNN 代替原始的自注意力机制,在 CTSegFormer 模型的 4 个阶段中,Transformer 的层数分别为 3、4、3、6。下方分支是 EEAA 模块,每个阶段只有 1 层,最后通过加法融合 2 个注意力分支的结果。

本文提出的 EEAA 模块灵感来源于 SwiftFormer 的高效加性注意力 (Efficient Additive Attention, EAA) 模块,EAA 模块结构图如图 2 所示。由图 2 可知,与 MobileViT-v2^[24]提出的可分离自注意力类似,该模块同样使用逐元素乘积操作来计算注意力矩阵,但 EAA 模块移除了 Key 与 Value 间的无效交互,只关注 Query 与 Key 之间的有效交互,并使用额外的线性层学习 tokens 之间的联系。具体地,对于输入的编码特征 $x \in \mathbf{R}^{n \times d}$,使用线性变换得到查询 $Q \in \mathbf{R}^{n \times d}$ 和键 $K \in \mathbf{R}^{n \times d}$,其中 n 是 token 的长度, d 是编码向量的维度。EAA 模块的输出如下:

$$\hat{x} = Q + T(K * q)$$

式中, Q 是归一化的查询矩阵, T 是线性变换,使用 1 个线性变换 $W_T \in \mathbf{R}^{d \times d}$ 来学习 tokens 的隐藏表示。

可以看出,EAA 模块在最后一步中将查询矩阵 Q 与全局上下文矩阵的隐藏表示 $T(K * q)$ 直接相加,无显式地考虑值 V ,形式上缺少了 Key 与 Value 的关系刻画。

与 EAA 不同,EEAA 先使用 1 个 MLP 层加一维

表 1 MFPP 模块具体参数设置

Table 1 Detailed parameter settings of the MFPP module

卷积核 名称	阶段 1 K, S, P	阶段 2 K, S, P	阶段 3 K, S, P	阶段 4 K, S, P
OP	7, 4, 3	3, 2, 1	3, 2, 1	3, 2, 1
CDC	7, 4, 3	3, 2, 1	3, 2, 1	3, 2, 1
DC $d=1$	7, 4, 3	3, 2, 1	3, 2, 1	3, 2, 1
DC $d=2$	5, 4, 3	3, 2, 2	3, 2, 2	3, 2, 2
DC $d=3$	5, 4, 5	3, 2, 3	3, 2, 3	3, 2, 3

卷积层直接从查询 Q 中学习全局的注意力查询向量 $\alpha \in \mathbf{R}^n$;MLP 能够学习输入特征的非线性表示,有助于捕捉查询 Q 中更复杂的特征关系,生成更丰富的查询向量参数;一维卷积擅长捕捉局部特征,可以在查询 Q 上提取有用的局部模式,数学表达式如下:

$$\alpha' = \text{MLP}(Q) + \text{Conv}_{3 \times 1}(Q) / \sqrt{d} \quad (1)$$

式中,MLP 包括 1 个线性层 $\text{Linear}(d, 2d)$, 1 个 ReLU 激活函数,1 个线性层 $\text{Linear}(2d, 1)$, $\text{Conv}_{3 \times 3}$ 的卷积核大小为 3,步长为 1。

为了学习到更丰富的特征表示,EEAA 再使用 1 个线性层来获得 Key 与 Value 之间的交互,EEAA 模块的输出如下:

$$\hat{x} = \hat{V} + T(K * q)$$

式中, $\hat{V}$ 是归一化的矩阵 V 。

由上可知,EEAA 模块的整体计算复杂度依然是线性且未使用昂贵的矩阵乘法,以 3 个线性层获取了更丰富的注意力特征矩阵,为 PAC 模块提供了更多的空间和位置信息。

1.5 PDS 模块

由图 1d)可知,在原始的 SegFormer 解码器中,

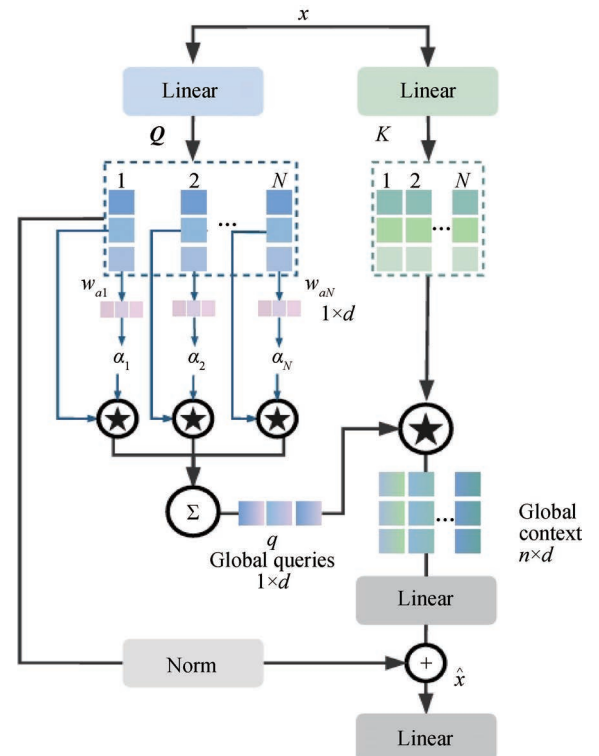


图 2 EAA 模块结构图

Fig. 2 Architecture of the EAA module

对于 4 个 Transformer 模块输出的特征图 $E_1 \in \mathbf{R}^{\frac{W}{4} \times \frac{H}{4} \times C_1}$, $E_2 \in \mathbf{R}^{\frac{W}{8} \times \frac{H}{4} \times C_2}$, $E_3 \in \mathbf{R}^{\frac{W}{16} \times \frac{H}{16} \times C_3}$, $E_4 \in \mathbf{R}^{\frac{W}{32} \times \frac{H}{32} \times C_4}$, 先用线性变换将它们的特征通道数全部变换为 C 来增强线性特征, 然后用双线性插值将特征图的尺寸全部恢复到原始输入图像的 1/4, 最后将这些特征在通道上进行拼接, 使用线性变换得到最终的分割结果, 整个过程的表达式如下所示:

$$\hat{E}_i = \text{Linear}(C_i, C)(E_i), \forall i \quad (2)$$

$$\hat{E}_i = \text{Upsample}\left(\frac{W}{4} \times \frac{H}{4}\right)(\hat{E}_i), \forall i \quad (3)$$

$$E = \text{Linear}(4C, C)(\text{Concat}(\hat{E}_i)), \forall i \quad (4)$$

$$M = \text{Linear}(C, N_{cls})(E) \quad (5)$$

式中, $\text{Linear}(C_i, C)$ 是提升通道数的线性层, $\text{Upsample}\left(\frac{W}{4} \times \frac{H}{4}\right)$ 是双线性插值上采样, $\text{Linear}(4C, C)$ 是降低通道数的线性层, M 是最终的预测分割图。

考虑到烟丝分割任务对边缘和细节信息的高度敏感特性, 低分辨率特征一次性插值到自身的 4 倍甚至 8 倍, 会带来位置和空间信息丢失。受 SETR^[25] 编码器设计理念的启发, 对低分辨率特征图 E_3 和 E_4 采用渐进式的双线性插值上采样, 对公式②线性变换后的 $\hat{E}_3$ 和 $\hat{E}_4$, 方法如下:

$$\hat{E}_3 = \text{Upsample}\left(\frac{W}{4} \times \frac{H}{4}\right)\left(\text{Upsample}\left(\frac{W}{8} \times \frac{H}{8}\right)(\hat{E}_3)\right)$$

$$\hat{E}_4 = \text{Upsample}\left(\frac{W}{4} \times \frac{H}{4}\right)\left(\begin{array}{l} \text{Upsample}\left(\frac{W}{8} \times \frac{H}{8}\right) \\ \left(\text{Upsample}\left(\frac{W}{16} \times \frac{H}{16}\right)(\hat{E}_4)\right) \end{array}\right)$$

式中, $\hat{E}_3$ 经过 2 次连续双线性插值恢复到 1/4 尺寸, $\hat{E}_4$ 经过 3 次连续双线性插值恢复到 1/4 尺寸。

另外, PDS 模块采用了深度监督损失来辅助训练, 对公式③中尺寸恢复到 $\left(\frac{W}{4} \times \frac{H}{4}\right)$ 的特征图 $\hat{E}_i(\forall i)$, 再使用卷积核大小为 1, 步长为 1 的卷积层, 将这些特征图的通道数降低到 N_{cls} , 与公式⑤中的分割结果 $M \in \mathbf{R}^{\frac{W}{4} \times \frac{H}{4} \times C}$ 共同送入损失函数模块, 将它们的损失求和作为最终损失 $Loss$, 该过程如下:

$$\hat{E}'_i = \text{Conv}(\hat{E}_i), \forall i$$

$$Loss = \text{GHMLoss}(M, \mathbf{y}_{\text{label}}) + \sum_{i=1}^4 \omega_i \cdot$$

$$\text{GHMLoss}(\hat{E}'_i, \mathbf{y}_{\text{label}})$$

式中, Conv 是卷积层, GHM (Gradient Harmonizing Mechanism) Loss 是 GHM 损失函数^[18], ω_i 是相应特征图损失的可学习权重, $\mathbf{y}_{\text{label}}$ 是烟丝图像的真实分割标签。

2 实验与分析

2.1 烟丝组分占比测定系统、数据集构建与实验设置

2.1.1 烟丝组分占比测定 将烟丝直接放在白底柔性振动台上, 打散烟丝, 使其均匀平铺, 尽量不发生粘连; 再用不反光玻璃压在烟丝上, 使其服帖; 最后使用手机摄像头拍摄烟丝图像, 通过手机上安装的自研 App 程序上传至服务器, 由服务器上的模型进行算法处理并将结果返回, App 则实时显示烟丝样品组分占比测定结果, 以及整丝率、碎丝率、梗签率等信息。实验设备及测定结果展示如图 3 所示。

2.1.2 烟丝组分占比数据集构建 数据集构建所用的实物烟丝主要为芙蓉王品牌烟丝 (大于 95%) 和引入了极少量的黄金叶品牌烟丝 (不到 5%), 上述烟丝的种类包括叶丝、薄片丝、梗丝、膨胀丝和梗签。所有图像的分辨率为 1400×1200 , 按照 7.0 : 1.5 : 1.5 的比例划分, 其中训练集 3357 张, 验证集和测试集分别为 720 张, 每张图像的 5 类别烟丝的质量占比均有差异, 包括接近成品烟支的占比, 以及均匀占比和自定义占比, 且尽可能包括了不同的光照条件, 也引入了不同主流手机品牌的摄像头的图像特性以及不同分辨率以保证图像多样化, 将此数据集命名为 CTobaccoDataset-v1。

2.1.3 实验设置 实验中的计算设备为一台含 Xeon(R) Gold 6226R 型号的 CPU, 内存 376 G, 4 块 RTX 3090 GPU 服务器, 运行的操作系统为 Ubuntu 18.04.4 LTS。模型训练的主要参数设置如下: 输入图片分辨率为 1400×1200 , 批大小为 4, 采用 GHM Loss 为损失函数, 采用 AdamW 为优化器, 动量因子为 0.9, 初始学习率为 1×10^{-4} , 最小学习率为 1×10^{-6} , 采用余弦学习率衰减策略, 迭代轮次为 200。

2.2 烟丝语义分割结果分析

笔者选择了代表性的基于深度学习的语义分割

方法作为对比方法,包括 FCN^[26]、U-Net、PSPNet^[27]、DeeplabV3 +、HRNet^[28]、SETR、SegFormer 和 SegNext^[29]。不同模型在 CTobaccoDataset-v1 数据集上的性能见表 2,对比指标有平均交并比(*mIoU*)、平均像素准确率(*mPA*)、叶丝率(PA_1)、薄片丝率(PA_2)、膨胀丝率(PA_3)、梗丝率(PA_4)及梗签率(PA_5),另外本文给出了每个模型的参数量,其中 *mIoU* 和 *mPA* 均不包含背景。

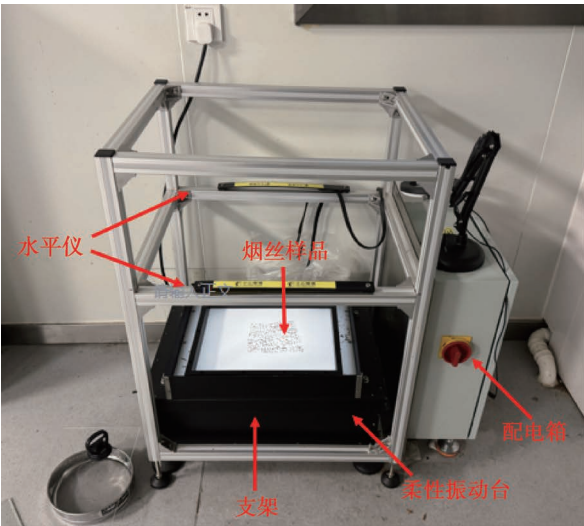
由表 2 可知,CTSegFormer 模型的综合性能最好,*mIoU* 和 *mPA* 分别达到了 76.1%和 87.6%,比性能排第二位的 SegNext-Base 模型分别提升了 3.3%和 2.2%,比性能排第三位的 SETR 模型分别提升了 3.3%和 2.5%。在梗丝的类别像素准确率上,CTSegFormer 比 SETR 提升了 5.1%;在叶丝率、膨胀丝率和梗签率的类别像素准确率上,也都提升了 2.5%以上。对于基于卷积神经网络的几个语义分割方法,PSPNet 和 DeeplabV3+模型的表现似乎没有达到预期,在 *mIoU* 指标上只高于 FCN 模型 1%左右,高于 U-Net 模型 5%左右,这是因为烟丝分割任务非常依赖于特征图的尺寸,PSPNet 和 DeeplabV3+均采用了 8 倍下采样策略,在上采样时也是从 1/8 大小直接恢复到原始尺寸,导致图像丢失了大量的细小烟丝特征,使模型的 *mIoU* 大幅下降,这一点从 FCN 模型的结果也可以得到证实,即使模型的参数量达到 134 M,远超对比模型,但因为同样采用了 8

倍下采样,性能表现较差。HRNet 模型在网络架构中则是采用了 1/4 大小到 1/32 大小的多尺度特征信息融合,虽仅有 29.8 M 的参数量,但性能只稍稍低于 SegFormer 模型。对于 SegNext 模型,选取了标准版本 SegNext-Base,虽然其在公开数据集上性能比原始 SegFormer 好一些,但本实验发现,将其直接用于烟丝的效果也不如 CTSegFormer, *mIoU* 降低了 3.3%。

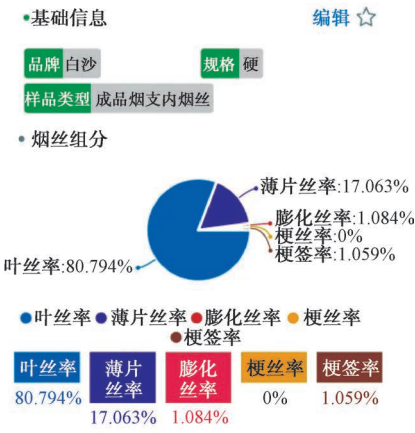
不同方法在 CTobaccoDataset-v1 数据集上的 *mIoU* 训练曲线如图 4 所示。由图 4 可知,CTSegFormer 在前 80 个周期内性能低于 SegFormer 和 SETR,后期提升较大。FCN、DeeplabV3 +(Xception)和 PSPNet 的 *mIoU* 训练曲线稳定性较差,存在较多震荡。不同模型的置信度区间如图 5 所示。由图 5 可知,这些模型在烟丝语义分割任务上的训练结果重复性较好,本文所建模型随训练次数增加,指标达到最大,进一步说明了本文模型的性能优势是稳定的。

2.3 可视化结果分析

不同模型分割烟丝的可视化结果如图 6 所示,本文从大分辨率的烟丝分割图像中剪裁出局部区域以方便观察。第 1 列是原始图像,包含了在不同光照条件拍摄到的烟丝样例图,第 2 列是从原始图像截取的局部区域,第 3 列是该局部区域的标签,第 4 列到第 7 列是用于比较的语义分割方法,包括



a)实验设备组成示意图



b)App返回的组分占比测定结果展示

图 3 实验设备及测定结果展示

Fig. 3 Experimental equipment and proportion determination results

表 2 不同模型在 CTobaccoDataset-v1 数据集上的性能

Table 2 Performance comparison of different methods on the CTobaccoDataset-v1 dataset

模型	<i>mIoU</i> /%	<i>mPA</i> /%	<i>PA</i> ₁ /%	<i>PA</i> ₂ /%	<i>PA</i> ₃ /%	<i>PA</i> ₄ /%	<i>PA</i> ₅ /%	参数量/M
FCN	58.5	72.4	73.7	68.9	74.6	74.0	70.8	134.0
PSPNet	59.1	74.5	76.1	58.1	73.4	78.0	86.8	46.8
DeeplabV3+(MobileNetV2)	59.6	74.9	78.7	71.5	69.1	73.2	81.9	5.8
DeeplabV3+(Xception)	62.3	76.6	80.6	72.9	72.4	74.1	83.0	56.2
U-Net	64.5	77.0	69.9	80.1	78.5	75.7	81.0	44.0
HRNet	67.4	81.4	81.3	81.3	78.7	81.7	84.0	29.8
SegFormer	70.4	83.9	83.3	82.6	79.2	84.2	80.2	27.4
SETR	72.8	85.1	84.4	86.8	81.2	82.9	90.2	92.6
SegNext-Base	72.8	85.4	85.3	84.6	84.4	84.9	87.8	27.6
CTSegFormer-Base	76.1	87.6	87.9	84.8	83.9	88.0	93.5	37.1

PSPNet、HRNet、SETR 和 CTSegFormer。由图 6 可知,CTSegFormer 相对其他分割方法表现更加出色,其中 PSPNet 由于下采样倍数过大,每一张图的分割结果都明显劣于其他方法,除了分割边界模糊的问题,也没有捕捉到烟丝中的细微结构,细节信息丢失严重,同时存在类别混淆和漏检的情况。HRNet 得益于其全程高分辨率的特征图,在分割细节上有了明显提升,但是漏检和分割不准确的问题依然存在,尤其在第 1 张和第 5 张图中能够观察到这一现象。SETR 的分割准确度进一步提高,但是在第 1 张图和第 6 张图仍然存在少量类别混淆的分割。CTSegFormer 通过其优秀的局部特征提取模块和注意力模块的改进,能够有效提取烟丝的细节和边缘信息,只存在极少量的类别混淆,并且没有出现漏检和烟丝细节结构信息丢失,因此 CTSegFormer 分割效果是最好的。

2.4 消融实验结果分析

MFPP、PAC 和 PDS 模块的消融实验结果对比见表 3。由表 3 可知,本文将 SegFormer 的编码器中每一层 OP 卷积模块替换为 MFPP 模块后,模型性能得到了明显提升,各项指标均提升了 2%,其中对梗签的像素级分类准确率更是提升了 11%。这说明 MFPP 模块受益于其较大的感受野和良好的细节处理能力,具备更强大的局部特征提取能力,能够更精细地处理烟丝细粒度的特征。膨胀卷积的 3 个分支取不同权值与等权值几乎没有明显差别,故最终取等权值。

为进一步了解 MFPP 中每一种卷积对模型性能

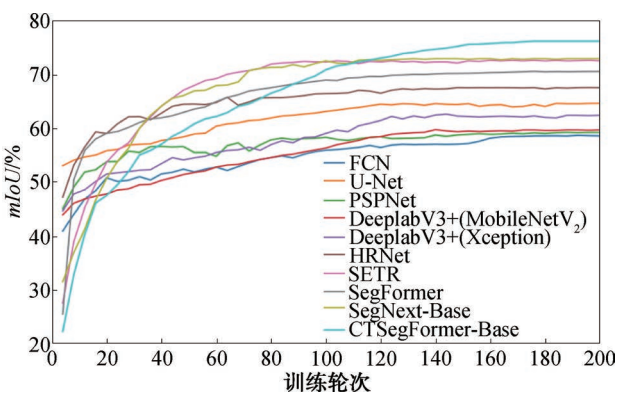


图 4 不同方法在 CTobaccoDataset-v1 数据集上的 *mIoU* 训练曲线

Fig. 4 *MIoU* training curves of different methods on the CTobaccoDataset-v1 dataset

的影响,设计实验将全部的 MFPP 模块退化为只包含单独的 OP 卷积、CDC 卷积或 DC 卷积,以及它们的不同组合,测定 *mIoU* 和 *mPA*,3 种卷积单独和组合的性能对比如图 7 所示。由图 7 可知,3 个卷积组合时性能最好,且包含 OP 卷积模块的组合比不包含 OP 卷积的组合性能更好一些。

PAC 模块性能分析:由表 3 可知,在模型参数量仅提升 0.9 M 的前提下,SegFormer+PAC 模型相较于 SegFormer 模型的 *mIoU* 和 *mPA* 分别获得了 1.4%和 3.2%的提升,*PA*₅ 同样得到了大幅度提升,这说明本文设计的 EEAA 模块能够弥补原 Transformer 模块的不足,可提供更多空间和位置信息,并充分感知烟丝目标位置,从而有效增强特征表达。不同改进模型的置信度区间如图 8 所示,每个模型测量 5 次。由图 8 可知,该模型及各改进模块的 *mIoU* 置信度区间的最大和最小值变动范围均

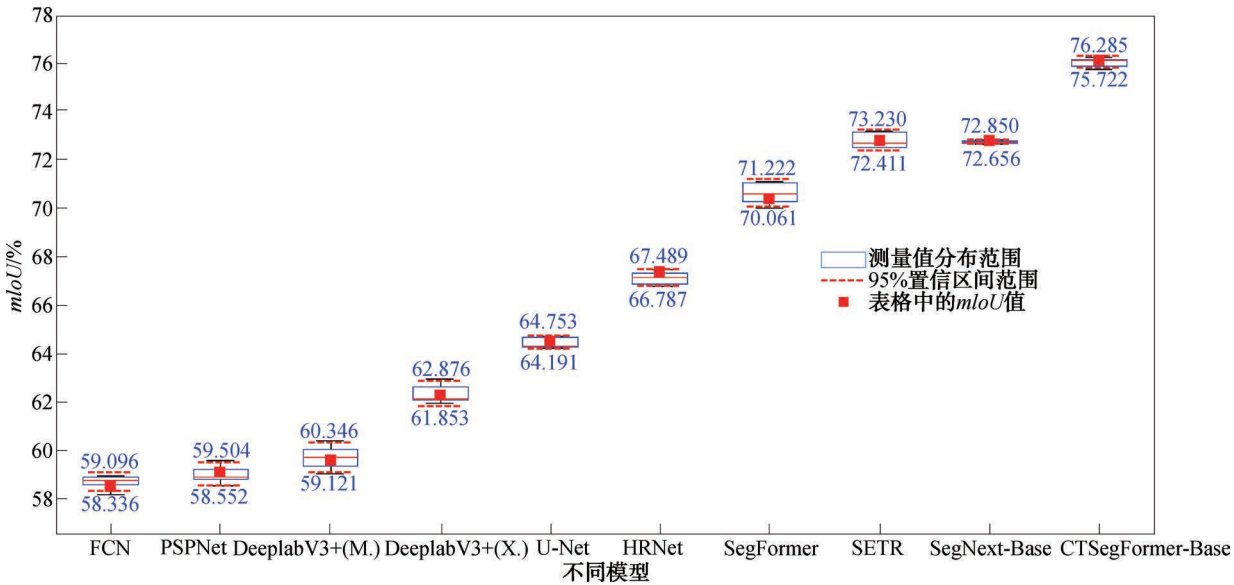


图 5 不同模型的置信度区间

Fig. 5 Confidence intervals of different models

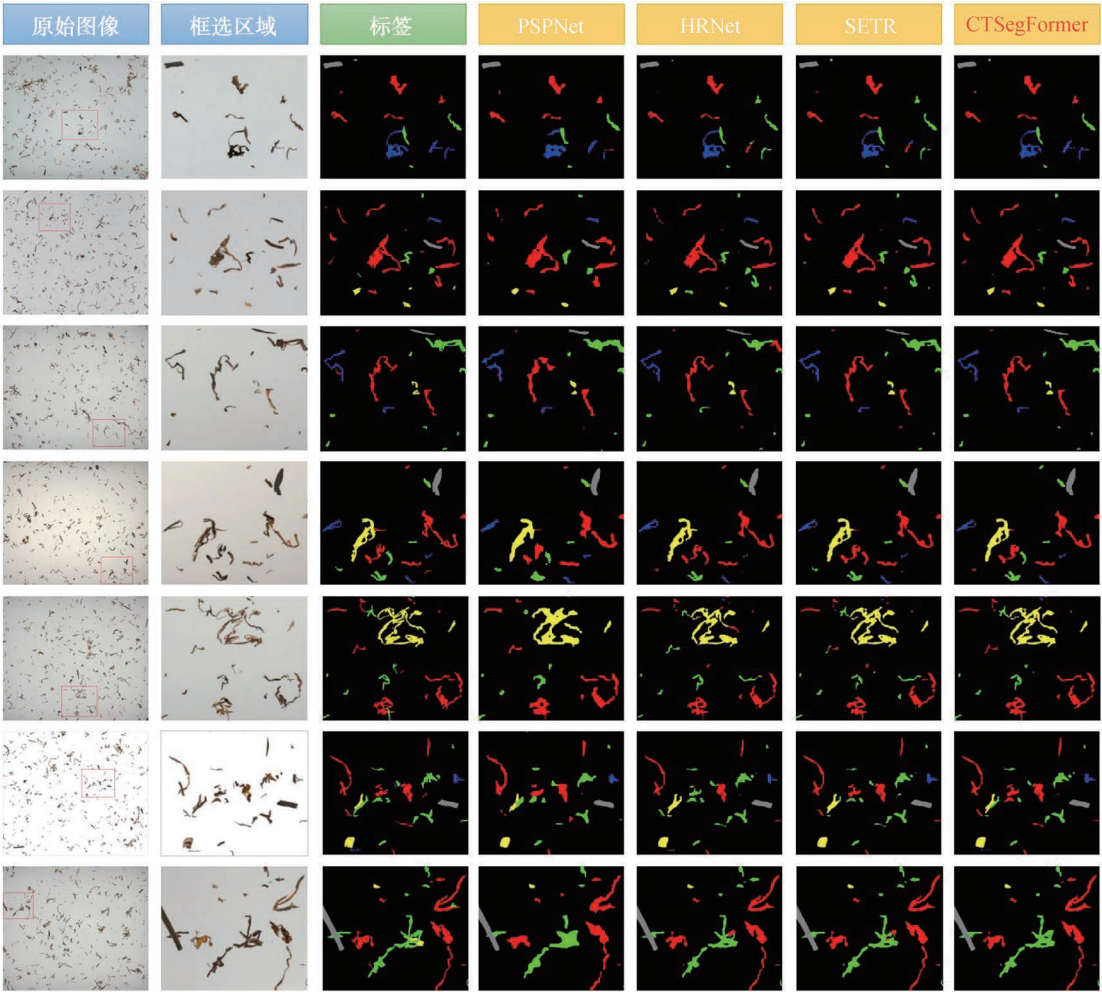


图 6 不同模型分割烟丝的可视化结果

Fig. 6 Visualization results of tobacco shred segmentation by different models

表 3 MFPP、PAC 和 PDS 模块的消融实验结果对比
Table 3 Comparative ablation study results of MFPP, PAC and PDS modules

模型	<i>mIoU</i> /%	<i>mPA</i> /%	<i>PA</i> ₁ /%	<i>PA</i> ₂ /%	<i>PA</i> ₃ /%	<i>PA</i> ₄ /%	<i>PA</i> ₅ /%	参数量/M
SegFormer(ViT-B2)	70.4	81.9	83.3	82.6	79.2	84.2	80.2	27.4
SegFormer+MFPP	<i>DC</i> _{<i>i</i>} = 2 : 1 : 1	73.8	86.3	86.8	84.3	82.6	86.1	36.2
	<i>DC</i> _{<i>i</i>} = 1 : 2 : 1	73.8	86.3	86.6	84.2	82.7	86.3	36.2
	<i>DC</i> _{<i>i</i>} = 1 : 1 : 2	73.7	86.3	86.8	84.2	82.6	86.3	36.2
	<i>DC</i> _{<i>i</i>} = 1 : 1 : 1	73.8	86.4	86.8	84.3	82.6	86.3	36.2
SegFormer+PAC	71.8	85.1	85.9	84.6	82.1	86.3	86.5	28.3
SegFormer+MFPP+PAC	75.0	87.3	86.9	85.7	84.7	87.4	91.8	37.1
SegFormer+MFPP+PAC+PDS (CTSegFormer-Base)	76.1	87.6	87.9	84.8	83.9	88.0	93.5	37.1

小于 1.2,说明本文模型的性能提升比较稳定。

其次,对 EEAA 中的 MLP 项和卷积项(公式①)的相关参数设置做了对比实验,有关 EEAA 模

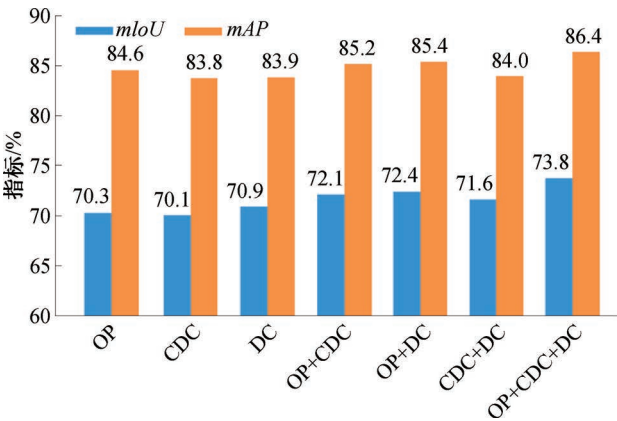


图 7 3 种卷积单独和组合的性能对比
Fig. 7 Performance comparison of three convolutions used individually and in combination

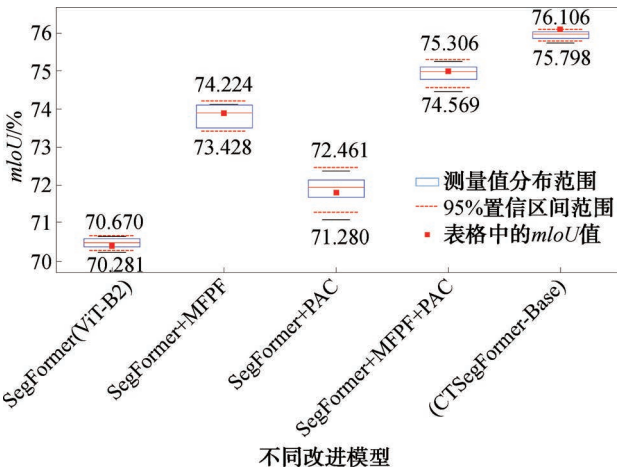


图 8 不同改进模型的置信度区间
Fig. 8 Confidence intervals of different improved models

块的消融实验结果见表 4。由表 4 可知,MLP 的隐含层神经元个数多一些性能会更好,而卷积项的卷积核尺寸设置得小一些更有利于性能提升,故采用 1024 个神经元,3×1 的小卷积核以保证模型的整体性能。

PDS 模块性能分析:表 3 也呈现了 CTSegFormer 模型在解码器部分使用和未使用 PDS 模块的实验结果。在未使用 PDS 模块的情况下,解码器是 SegFormer 的轻量级解码器,只包含一些 MLP 层和双线性插值。使用 PDS 模块后,*mPA* 提升了 0.3%,*mIoU* 提升了 1.1%,表明采用渐进式的双线性插值上采样可以保留更多的细节特征,提升 *mIoU*,同时深度监督的训练策略能够确保不同尺寸的分支都可以学习到烟丝的结构和细节局部特征,进而提升了模型的综合性能。

损失函数性能分析:有关损失函数的消融实验结果见表 5。由表 5 可知,使用 Loss 替换 Focal Loss 时(表 5 最后 1 行),每个指标都取得了小幅度提

表 4 有关 EEAA 模块的消融实验结果
Table 4 Ablation study results on the EEAA module

EEAA 的 MLP 隐含层 神经元个数	EEAA 卷积项的 卷积核大小	<i>mIoU</i> /%
256	3×1	75.22
	5×1	75.37
	7×1	75.03
512	3×1	75.81
	5×1	75.69
	7×1	75.43
1024	3×1	76.12
	5×1	75.98
	7×1	75.83

升,其中 $mIoU$ 和 mPA 分别小幅提升了 0.5% 和 0.6%, $PA5$ 提升幅度较大,提升了 1.2%。这说明 GHMLoss 的引入能够小幅度改善烟丝样本不平衡的情况,为模型带来更高的分割准确度。由公式⑤的损失函数中 ω_i 取不同加权值时的结果可知,固定取 $\omega_1 = 1, \omega_2 = 1, \omega_3 = 1, \omega_4 = 2$ 综合效果最好。

Transformer 模块尺寸性能分析: 有关 Transformer 骨架大小的消融实验结果见表 6,遵循 Segformer 原文设置,表中显示 3 种 Transformer 模块尺寸 Vit-B0、Vit-B1 和 Vit-B2。使用 Vit-B0 的 CTSegFormer 命名为 CTSegFormer-Tiny,使用 Vit-B1 的 CTSegFormer 命名为 CTSegFormer-Small,由表 6 可知,相比与之对应的基准模型 SegFormer (Vit-B0) 和 SegFormer (Vit-B1), $mIoU$ 分别提升了 3.8% 和 4.1%,这说明本文提出的改进模块在轻量级 SegFormer 模型上同样能够提升分割精度。但是两者提高幅度均低于 SegFormer (Vit-B2),说明 Transformer 模块的全局信息提取能力同样不可忽视。

输入图像的分辨率: 除实验中所使用的图像分辨率外,还进一步测试了 1200×1024 、 1024×768 和 768×576 3 种更低分辨率的样本图像,验证 CTSegFormer-Base 对输入图像尺寸变化的鲁棒性。实验发现,随着分辨率的降低,纹理丢失对类别的辨识影响很大,导致 CTSegFormer-Base 的 $mIoU$ 分别下降了 1.13%、4.12% 和 9.20%。

2.5 烟丝占比测定结果分析

在完成语义分割后,统计各类像素的个数作为各类烟丝在摄像机下的投影面积,再分别乘以事先测定的各类烟丝面密度比,即得到烟丝质量比,归一化后可得到组分占比 r 。使用相对误差来衡量占比预测的性能,第 i 类烟丝的相对误差定义如下:

$$err = \frac{|r_i - \bar{r}_i|}{\bar{r}_i}$$

其中, $\bar{r}_i$ 为第 i 类真实占比。为验证组分占比的预测精度,选取了 4 组不同占比的混合烟丝,由各类纯丝和梗签混合得到。配料时,先称取各类别烟丝的质量,计算占比真实值。混合时先用镊子随机夹取放入器皿,而后再轻轻摇晃,以保证烟丝不破损。摇匀后轻轻倒在柔性振动台上,振动至烟丝能均匀摊开,再用手机拍摄烟丝图像。

CTSegFormer 模型对真实烟丝样品的组分占比测定结果见表 7。由表 7 可知,4 组样品所有类别烟丝的相对误差最大为 5.96%。而第 6、7 列中用格式 (误差源 1 误差,误差源 2 误差) 给出的数据进一步列出了 2 个主要误差源的绝对误差和相对误差。这 2 个主要误差是由烟丝识别与分割 2 部分产生。误差源 1 来源于测试图像欠清晰、小碎片烟丝与其他烟丝粘连,以及部分稀有形态烟丝样本在训练集中覆盖不足等;误差源 2 来源于烟丝的自缠绕和遮挡,即使压平且可见部分分割正确,烟丝面积统计也会存在一定的误差,影响烟丝组分占比的预测结果。总体来看,4 组样品所有类别烟丝的相对误差最大为第 2 组样品中的膨胀丝,为 5.96%,4 组样品中叶丝的平均相对误差最小,其他 3 类都在某一组样品中出现了较大的相对误差。

2.6 占比测定性能和模型鲁棒性分析

近年来,行业内一般期望自动化烟丝组分占比测定能够达到与人工测量接近或更低的相对误差,即不超过 5%。因此,本文选取了基于高光谱和太赫兹的烟丝组分占比测定方法作对比。经测定,基于高光谱的测定方法^[6] 相对误差范围为 0.71% ~ 8.36%,基于太赫兹的测定方法^[30] 的相对误差范围为 1.70% ~ 10.00%,本文测定方法利用可见光谱段,相对误差范围为 0.42% ~ 5.96%。

由于不同品牌手机的摄像头在靶面类型、尺寸等工艺参数上存在差异,其成像特性会对烟丝样品的拍摄效果产生直接影响。为验证本文所建模型的鲁棒性,用 CTSegFormer 模型对国内外 5 种主流品牌手机 (iPhone、vivo、小米、华为、OPPO) 所拍摄的同一组烟丝的语义分割结果如图 9 所示,CTSegFormer 对不同手机拍摄同一组烟丝样品组分占比测定结果见表 8。由图 9 和表 8 可知,拍摄图像质量主要影响模型对烟丝类型的识别精度和语义分割的准确性。从相对误差数据可以看出,各品牌手机对应的测定误差大小与人眼主观评判的图像质量基本排序一致。因使用不同品牌手机拍摄而引起的每一类烟丝占比预测的最大和最小相对误差的差异分别为:叶丝 4.90%,膨胀丝 2.17%,薄片丝 1.37%,梗丝 2.88%,在本文测定方法的相对误差范围内,且远小于高光谱和太赫兹测定方法的误差上限,说明本文所建模型 CTSegFormer 的鲁棒

表 5 有关损失函数的消融实验结果
Table 5 Ablation study results on loss functions

模型	$mIoU$	mPA	PA_1	PA_2	PA_3	PA_4	PA_5	%
CTSegFormer-Base(Focal)	75. 6	87. 0	87. 6	83. 9	83. 5	86. 8	93. 0	
CTSegFormer-Base(GHM)所有 ω_i 在 1~2 范围自适应学习	75. 8	87. 4	87. 5	83. 2	82. 2	86. 2	93. 1	
CTSegFormer-Base(GHM) $\omega_1=2,\omega_2=1,\omega_3=1,\omega_4=1$	75. 7	87. 5	87. 1	84. 5	83. 3	85. 8	92. 7	
CTSegFormer-Base(GHM) $\omega_1=1,\omega_2=2,\omega_3=1,\omega_4=1$	75. 3	87. 1	87. 3	84. 2	83. 5	86. 4	92. 4	
CTSegFormer-Base(GHM) $\omega_1=1,\omega_2=1,\omega_3=2,\omega_4=1$	75. 8	87. 0	87. 6	84. 0	83. 2	86. 3	92. 1	
CTSegFormer-Base(GHM) $\omega_1=1,\omega_2=1,\omega_3=1,\omega_4=2$	76. 1	87. 6	87. 9	84. 8	83. 9	88. 0	93. 5	

表 6 有关 Transformer 骨架大小的消融实验结果
Table 6 Ablation study results on Transformer backbone sizes

模型	$mIoU$ /%	mPA /%	PA_1 /%	PA_2 /%	PA_3 /%	PA_4 /%	PA_5 /%	参数量/M
SegFormer(Vit-B0)	62. 4	76. 4	79. 8	76. 2	72. 2	75. 9	77. 9	3. 7
SegFormer(Vit-B1)	68. 5	81. 1	82. 9	81. 9	77. 9	83. 9	79. 2	13. 7
SegFormer(Vit-B2)	70. 4	83. 9	83. 3	82. 6	79. 2	84. 2	80. 2	27. 4
CTSegFormer-Tiny	66. 2	80. 0	81. 7	81. 4	81. 2	76. 0	79. 7	6. 2
CTSegFormer-Small	72. 6	84. 9	85. 1	83. 4	83. 1	85. 9	87. 0	23. 4
CTSegFormer-Base	76. 1	87. 6	87. 9	84. 8	83. 9	88. 0	93. 5	37. 1

表 7 CTSegFormer 对真实烟丝样品的组分占比测定结果
Table 7 Results of component proportion analysis in real tobacco samples by CTSegFormer

编号	烟丝种类	真实质量/g	实际组分占比/%	预测组分占比/%	绝对误差/% (误差源 1,误差源 2)	相对误差/% (误差源 1,误差源 2)
真实样品 1	叶丝	0. 30	42. 86	42. 16	0. 70 (0. 21,0. 49)	1. 63 (0. 49,1. 14)
	膨胀丝	0. 10	14. 29	14. 45	0. 16 (0. 09,0. 07)	1. 12 (0. 63,0. 49)
	薄片丝	0. 20	28. 57	29. 09	0. 51 (0. 13,0. 38)	1. 79 (0. 46,1. 33)
	梗丝	0. 10	14. 29	14. 31	0. 02 (0. 008,0. 012)	0. 14 (0. 06,0. 08)
真实样品 2	叶丝	0. 40	40. 00	40. 17	0. 17 (0. 04,0. 13)	0. 42 (0. 10,0. 32)
	膨胀丝	0. 20	20. 00	21. 19	1. 19 (0. 73,0. 46)	5. 96 (3. 65,2. 31)
	薄片丝	0. 20	20. 00	19. 88	0. 12 (0. 05,0. 07)	0. 60 0. 265,0. 335)
	梗丝	0. 20	20. 00	18. 86	1. 14 (0. 55,0. 59)	5. 70 (2. 75,2. 95)
真实样品 3	叶丝	0. 40	57. 14	56. 33	0. 81 (0. 44,0. 37)	1. 42 (0. 77,0. 65)
	膨胀丝	0. 1	14. 29	14. 21	0. 07 (0. 03,0. 04)	0. 49 (0. 23,0. 26)
	薄片丝	0. 1	14. 29	15. 03	0. 74 (0. 22,0. 52)	5. 18 (1. 54,3. 64)
	梗丝	0. 1	14. 29	13. 82	0. 46 (0. 31,0. 15)	3. 22 (2. 17,1. 05)
真实样品 4	叶丝	0. 4	44. 44	44. 78	0. 34 (0. 11,0. 23)	0. 77 (0. 25,0. 52)
	膨胀丝	0. 1	11. 11	11. 59	0. 48 (0. 26,0. 22)	4. 32 (2. 34,1. 98)
	薄片丝	0. 2	22. 22	21. 77	0. 45 (0. 009,0. 441)	2. 03 (0. 04,1. 98)
	梗丝	0. 2	22. 22	21. 86	0. 36 (0. 17,0. 19)	1. 62 (0. 77,0. 86)

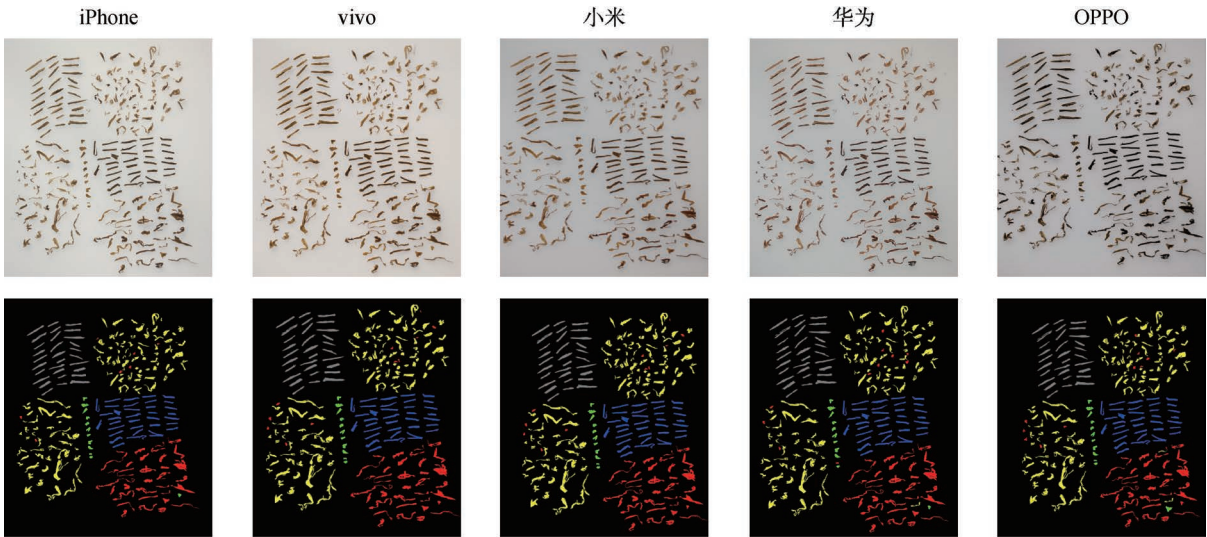


图9 5种主流品牌手机拍摄同一组烟丝样品的原图像及语义分割结果

Fig. 9 Original images and semantic segmentation results of the same cut tobacco sample photographed by five mainstream-brand mobile phones

表8 CTSegFormer 对不同手机拍摄同一组烟丝样品的组分占比测定结果

Table 8 CTSegFormer’s measurement results of component proportions for the same set of tobacco shred samples captured by different mobile phones

样品	真实占比	iPhone		vivo		小米		华为		OPPO	
		预测占比	相对误差	预测占比	相对误差	预测占比	相对误差	预测占比	相对误差	预测占比	相对误差
叶丝	31.84	32.64	2.51	32.54	2.20	33.49	5.18	34.10	7.10	33.98	6.72
膨胀丝	8.32	8.41	1.08	8.49	2.04	8.19	1.26	8.12	2.40	8.59	3.25
薄片丝	22.64	22.71	0.31	22.81	0.75	22.43	0.93	22.34	1.33	22.26	1.68
梗丝	37.20	36.24	2.58	36.16	2.80	35.89	3.52	35.44	4.73	35.17	5.46

性较好,具有可行性。

3 结论

本文提出了基于CTSegFormer模型的烟丝组分占比测定方法,以SegFormer为基础进行改进,并设计了MFPF、PAC和PDS模块,EEAA模块在引入少量线性层和一维卷积层的前提下,增强了SwiftFormer^[19]中加性注意力机制的特征提取能力,在保持计算复杂度O(N)的同时弥补了Efficient Self-attention因特征维度放缩带来的性能下降。PDS模块则是引入了渐进式的双线性插值进行上采样,恢复特征分辨率的过程中,能够更好地保留烟丝的细节和边缘信息,同时多层监督信号能帮助模型在不同特征层次上学到更丰富的特征表示,提高模型的泛化能力并加速收敛过程。同时,构建了数据集CTobaccoDataset-v1用于评估模型性能,在该

数据集上的大量对比和消融实验结果表明,CTSegFormer在烟丝语义分割性能上优于主流模型,所设计的模块能够有效提升模型性能,在此基础上,测得烟丝组分占比测定误差小于6%,验证了该方法的有效性与可行性。后续研究将进一步考虑更多种类的烟丝样本,以提升模型在不同烟丝条件下的泛化能力;同时,开展模型轻量化研究,以满足实际检测场景中的部署需求。此外,还将探索模型在不同品牌烟丝之间的迁移应用,进一步提高方法的适用范围和稳定性。

参考文献:

[1] Gao Z Y, Wang A, Ding H, et al. Tobacco material composition recognition method based on convolutional neural network[J]. Tobacco Science & Technology, 2017, 50(9):68-75.
[2] 王小明,魏甲欣,马飞,等. 基于AdaBoost集成学习的烟

- 丝组分识别[J]. 食品与机械, 2022, 38(3): 205-211.
- Wang X M, Wei J X, Ma F, et al. Identification of cut tobacco components based on AdaBoost ensemble learning [J]. Food & Machinery, 2022, 38(3): 205-211.
- [3] 刘栋, 陈越立, 李华杰, 等. 取样量、取样次数对烟丝混合均匀性检测的影响[J]. 郑州轻工业学院学报(自然科学版), 2013(1): 45-49.
- Liu D, Chen Y L, Li H J, et al. Influence of sampling volume and frequency on the mixing uniformity test of cut filler [J]. Journal of Zhengzhou University of Light Industry (Natural Science), 2013(1): 45-49.
- [4] 李斌, 蔡佳校, 何邦华, 等. 一种基于热分析技术测定烟丝中梗丝含量的方法: CN201710577469. 5[P]. 2017-10-20.
- Li B, Cai J X, Hua B H, et al. A method for determining stem content in cut tobacco based on thermal analysis technology: CN201710577469. 5[P]. 2017-10-20.
- [5] 寇霄腾, 张勇, 张卉, 等. 基于 RGB 图像特征的卷烟梗丝掺配比例检测[J]. 食品与机械, 2021, 37(9): 78-82.
- Kou X T, Zhang Y, Zhang H, et al. Detection of blending proportion of cut tobacco stem based on RGB image features [J]. Food & Machinery, 2021, 37(9): 78-82.
- [6] 杨天卓, 何晋, 吴恋恋, 等. 基于高光谱成像的烟丝掺配比例检测研究[J]. 轻工学报, 2025, 40(3): 115-126.
- Yang T Z, He J, Wu L L, et al. Detection of tobacco blend ratio based on hyperspectral imaging [J]. Journal of Light Industry, 2025, 40(3): 115-126.
- [7] 刘松岳, 俞世康, 赵宇, 等. 基于深度语义分割的重叠烟叶分级方法[J]. 计算机与数字工程, 2023, 51(7): 1645-1650.
- Liu S Y, Yu S K, Zhao Y, et al. Overlapping tobacco ranking method based on deep semantic segmentation [J]. Computer and Digital Engineering, 2023, 51(7): 1645-1650.
- [8] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale [EB/OL]. V2. arXiv (2021-06-03) [2025-06-30]. <https://doi.org/10.48550/arxiv.2010.11929>.
- [9] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C] // Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.
- [10] Xie E Z, Wang W H, Yu Z D, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers [EB/OL]. V3. arXiv (2021-10-28) [2025-06-30]. <https://doi.org/10.48550/arxiv.2105.15203>.
- [11] 高延海. 基于改进 SegFormer 的语义分割方法的研究 [D]. 青岛: 青岛理工大学, 2024.
- Gao Y H. Research on semantic segmentation method based on improved SegFormer [D]. Qingdao: Qingdao University of Technology, 2024.
- [12] 王晓杰, 陈少康, 闫皓炜, 等. 融合注意力机制的 SegFormer 遥感影像道路识别[J]. 计算机系统应用, 2024, 33(11): 186-193.
- Wang X J, Chen S K, Yan H W, et al. Road recognition in remote sensing images using SegFormer fused with attention mechanism [J]. Computer Systems and Applications, 2024, 33(11): 186-193.
- [13] 胡涛涛, 李屹旭, 张俊. 基于注意力机制改进的 SegFormer 遥感图像语义分割方法[J]. 激光杂志, 2024, 45(7): 130-136.
- Hu T T, Li Y X, Zhang J. Improved SegFormer network semantic segmentation method for remote sensing images based on attention mechanism [J]. Laser Journal, 2024, 45(7): 130-136.
- [14] 田雪伟, 汪佳丽, 陈明, 等. 改进 SegFormer 网络的遥感图像语义分割方法[J]. 计算机工程与应用, 2023, 59(8): 217-226.
- Tian X W, Wang J L, Chen M, et al. Improved SegFormer network based method for semantic segmentation of remote sensing images [J]. Computer Engineering and Applications, 2023, 59(8): 217-226.
- [15] 蒋仕新, 唐椿程, 杨建喜, 等. 基于改进 SegFormer 的混凝土桥梁表观病害轻量级识别方法[J]. 中国公路学报, 2024, 37(2): 77-87.
- Jiang S X, Tang C C, Yang J X, et al. Lightweight detection method of surface damage on concrete bridges based on improved SegFormer [J]. China Journal of Highway and Transport, 2024, 37(2): 77-87.
- [16] 张铭泉, 张泽恩, 曹锦纲, 等. 结合 SegFormer 与增强特征金字塔的文本检测方法[J]. 智能系统学报, 2024, 19(5): 1111-1125.
- Zhang M Q, Zhang Z E, Cao J G, et al. Text detection method combining SegFormer with an enhanced feature pyramid [J]. CAAI Transactions on Intelligent Systems, 2024, 19(5): 1111-1125.
- [17] Yu Z T, Zhao C X, Wang Z Z, et al. Searching central difference convolutional networks for face anti-spoofing [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 5294-5304.
- [18] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C] // Computer Vision-ECCV 2018. Cham: Springer, 2018: 833-851.
- [19] Shaker A, Maaz M, Rasheed H, et al. SwiftFormer: efficient additive attention for transformer-based real-time mobile vision applications [C] // 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France. IEEE, 2023: 17379-17390.
- [20] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [C] // Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham: Springer, 2015: 234-241.

[21] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481–2495.

[22] Li B Y, Liu Y, Wang X G, et al. Gradient harmonized single-stage detector[C]. // Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 8577–8584.

[23] Wang W H, Xie E Z, Li X, et al. PVT v2: improved baselines with pyramid vision transformer[J]. Computational Visual Media, 2022, 8(3): 415–424.

[24] Mehta S, Rastegari M. Separable self-attention for mobile vision transformers[PP/OL]. arXiv(2022-06-06)[2025-06-30]. <https://doi.org/10.48550/arxiv.2206.02680>.

[25] Zheng S X, Lu J C, Zhao H S, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers[C]. // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA. IEEE, 2021: 6877–6886.

[26] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA. IEEE, 2015: 3431–3440.

[27] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network[C]. // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017: 6230–6239.

[28] Sun K, Xiao B, Liu D, et al. Deep high-resolution representation learning for human pose estimation[C]. // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE, 2019: 5686–5696.

[29] Cheng M M, Guo M H, Hou Q B, et al. SegNeXt: rethinking convolutional attention design for semantic segmentation[C]. // Advances in Neural Information Processing Systems 35. New Orleans, Louisiana, USA. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2022: 1140–1156.

[30] 周博, 朱文魁, 王赵改, 等. 基于太赫兹时域光谱技术的烟草组分识别[J]. 农业工程学报, 2022, 38(10): 310–316.

Zhou B, Zhu W K, Wang Z G, et al. Identification of tobacco materials based on terahertz time-domain spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(10): 310–316.

CTSegFormer: a semantic segmentation network for cut tobacco component proportion measurement

Ding Duo¹, Li Chunxiu², Jin Yong¹, Xin Yiding³, Tan Ge⁴, Wang Yibin⁵, Xu Wenwu⁵

1. Technology Center, Chinese Tobacco Industry Hunan Co., Ltd., Changsha 410000, China;
2. Changde Tobacco Factory, Chinese Tobacco Industry Hunan Co., Ltd., Changde 415000, China;
3. Information Center, Chinese Tobacco Industry Hunan Co., Ltd., Changsha 410000, China;
4. Raw Material Procurement Center, Chinese Tobacco Industry Hunan Co., Ltd., Changsha 410000, China;
5. Nanjing Joint Science & Technology Co., Ltd., Nanjing 210000, China

Abstract:【Objective】To improve the efficiency and accuracy of component proportion measurement for tobacco shreds.【Methods】A semantic segmentation model named CTSegFormer is proposed. Taking SegFormer as the backbone, the model designs parallel fusion of multiple types of convolutions, an enhanced efficient additive attention mechanism, and a progressive deep supervision module. It achieves effective recognition and segmentation of four categories of tobacco shreds including lamina shreds, reconstituted tobacco shreds, stem shreds, expanded shreds, as well as stem sticks. On this basis, accurate measurement of tobacco shred component proportions is realized, and a tobacco shred component measurement dataset is constructed for testing and validation.【Results】Compared with existing semantic segmentation models, CTSegFormer achieves the highest mIoU in tobacco shred segmentation and recognition. Meanwhile, the relative error of proportion prediction is no more than six percent when testing multiple samples with different component ratios.【Conclusion】The proposed CTSegFormer model can be well applied to the task of tobacco shred component proportion measurement, and it is of great significance for monitoring the quality of finished cigarettes.

Key words: tobacco shred; component proportion; semantic segmentation; SegFormer

[责任编辑: 刘丹丹 刘春奎]