智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

基于序列分析的多模态石化VOCs烟羽语义分割

王子豪, 夏秀山, 曹洋, 张锟宇

引用本文:
王子豪, 夏秀山, 曹洋, 等. 基于序列分析的多模态石化VOCs烟羽语义分割[J]. 智能系统学报, 2025, 20(6): 1420-1431.
WANG Zihao, XIA Xiushan, CAO Yang, et al. Multimodal sequence-based petrochemical VOCs plume semantic segmentation[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(6): 1420-1431.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202501034

DOI: 10.11992/tis.202501034
网络出版地址: https://link.cnki.net/urlid/23.1538.tp.20250908.1744.002

基于序列分析的多模态石化 VOCs 烟羽语义分割

王子豪[1], 夏秀山[1], 曹洋[2], 张锟宇[3]
(1. 中国科学技术大学 先进技术研究院, 安徽 合肥 230031; 2. 中国科学技术大学 自动化系, 安徽 合肥 230027; 3. 合肥综合性国家科学中心 人工智能研究院, 安徽 合肥 230088)

摘 要: 石化挥发性有机化合物 (volatile organic compounds, VOCs) 烟羽在红外成像下表现出形态扭曲多变、边缘模糊和半透明的特性, 直接使用现有的图像语义分割方法难以提取气体特征, 分割效果不佳。为此本文提出一种结合上下文序列图像的多模态石化 VOCs 烟羽分割方法, 利用烟羽边缘的扩散特性提取目标帧的前后帧运动扩散矢量, 通过叠加运动信息增强 VOCs 烟羽边缘特征。利用 VOCs 在可见光下不成像的特点, 设计自适应权重模块融合可见光和红外光图像特征, 进一步增强烟羽特征, 过滤背景干扰。引入一种基于区域代理的烟羽分割解码器, 加强烟羽边缘和中心特征的关联性, 同时降低烟羽分割计算量。此外, 本文构建了石化 VOCs 可见光与红外视频数据集, 在数据集上的实验结果表明, 与基线网络相比, 本文方法计算效率提高了 1.81 帧/s, 同时分割精度提高了 3.53%。

关键词: VOCs 烟羽; 气体检测; 语义分割; 运动信息; 扩散; 多模态特征融合; 红外图像; 边缘模糊
中图分类号: TP391 **文献标志码**: A **文章编号**: 1673−4785(2025)06−1420−12

中文引用格式: 王子豪, 夏秀山, 曹洋, 等. 基于序列分析的多模态石化 VOCs 烟羽语义分割 [J]. 智能系统学报, 2025, 20(6): 1420–1431.
英文引用格式: WANG Zihao, XIA Xiushan, CAO Yang, et al. Multimodal sequence-based petrochemical VOCs plume semantic segmentation[J]. CAAI transactions on intelligent systems, 2025, 20(6): 1420–1431.

Multimodal sequence-based petrochemical VOCs plume semantic segmentation

WANG Zihao[1], XIA Xiushan[1], CAO Yang[2], ZHANG Kunyu[3]
(1. Institute of Advanced Technology, University of Science & Technology of China, Hefei 230031, China; 2. Department of Automation, University of Science & Technology of China, Hefei 230027, China; 3. Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, China)

Abstract: Petrochemical volatile organic compounds (VOCs) plumes manifest distorted and changeable shapes, blurred edges, and translucency under infrared imaging. The implementation of existing image semantic segmentation methods in the direct application context presents significant challenges in the extraction of gas features, resulting in suboptimal outcomes. To address this, this paper proposes a multimodal petrochemical VOCs plume segmentation method (MPPS) that incorporates contextual sequences. Initially, the diffusion characteristics of the plume edge are utilized to extract the motion diffusion vectors of the previous and subsequent frames of the target frame. Subsequently, the edge features of the VOC plume are enhanced by superimposing motion information. Second, an adaptive weight module is designed to leverage the non-imaging characteristics of VOCs in visible light. This module fuses visible and infrared image features, further enhancing plume features and filtering background interference. Finally, a region-based proxy plume segmentation decoder is introduced to enhance the correlation between edge and center features of the plume while reducing the computational load of plume segmentation. Furthermore, this paper constructs a visible and infrared petrochemical VOCs video dataset. Experimental results on this dataset demonstrate that MPPS improves computational efficiency by 1.81 frames per second and segmentation accuracy by 3.53% compared to baseline networks.

Keywords: VOCs plume; gas detectors; semantic segmentation; motion information; diffusion; multimodal feature fusion; infrared imaging; blurred edge

收稿日期: 2025−01−27. **网络出版日期**: 2025−09−09.
基金项目: 安徽省重点研究与开发计划项目 (2022107020030).
通信作者: 夏秀山. E-mail: xiaxiushan@iat.ustc.edu.cn.

石化气体包括烷烃、烯烃等种类繁多的挥发性有机化合物 (volatile organic compounds, VOCs),

是石油化工、医药化工等行业的基础原材料。近年来, 因石化气体泄漏导致的生产安全和环境污染事故时有发生, 给人们的生命、财产安全带来了巨大风险。因此, 及时有效地检测石化气体泄漏具有巨大的社会和经济价值。借助石化 VOCs 烟羽对特定红外波段的“指纹”吸收特性[1-2], 通过红外相机可以捕捉裸眼不可见的石化 VOCs 烟羽泄漏。现阶段石油化工企业多采用人工携带红外设备巡检可疑泄漏源的方式, 效率低下尚未实现快速高效的自动化检测[3]。基于视觉算法自动分割石化 VOCs 泄漏烟羽对于定位泄漏源和估计泄漏量至关重要, 但当前研究仍然十分有限。

近年来, 通用图像分割领域持续取得显著进展, Shojaief 等[4] 通过在轻量级 U-Net 架构中利用空洞空间金字塔池化融合可见光特征, 提高夜间交通场景红外成像下对行人、道路、汽车等目标的分割精度; Huo 等[5] 利用玻璃对可见光透明但对热能不透明的特性, 构建基于注意力多模态融合的玻璃分割算法; Li 等[6] 设计 K-Net 网络, 利用一组可学习核关联视频序列上下同一对象特征, 实现对人物、房屋、树木等目标的追踪; Yang 等[7] 基于 Mamba 将长程时空表征压缩至多尺度序列, 结合边界感知仿射约束, 增强对长序列医学图像如甲状腺、乳腺、结肠息肉等目标的分割能力。但是, 上述通用图像分割方法针对固体目标, 并不完全适用于石化 VOCs 烟羽。这是由于石化 VOCs 烟羽运动是一种湍流运动, 表现为形态扭曲多变、边缘透明渐变模糊, 与上述方法的分割目标有较大差异。

目前针对气体的烟羽检测方法可以分为传统图像分割方法和基于深度学习的方法。传统图像分割通过综合烟羽的颜色、纹理和动态特征, 利用先验规则和数学模型分割烟羽, 如基于阈值[8]、基于聚类[9] 和基于小波分析[10] 等, 不需要大规模训练就可以在简单的场景下取得良好的效果, 但其依赖浅层特征和专家知识, 难以适应复杂的场景。基于深度学习的方法能够通过神经网络自动提取烟羽的深层特征表征, 例如: Wang 等[11] 构建包含多个成像光谱的森林火灾烟羽数据集, 采用结合注意力机制的 Smoke-Unet 网络学习深层次烟羽特征; Yuan 等[12] 设计包含颜色通道的三维注意力模块以增强对烟羽空间特征表达能力, 提高对烟羽边界的定位精度。尽管石化 VOCs 烟羽红外成像下的运动特征与可见光烟羽存在相似性[13], 但红外成像缺乏颜色空间信息, 其烟羽纹理特征仅表现为辐射强度分布。这种物理成像特性的本质差异使得上述方法不能有效区分石化 VOCs 烟羽和背景。

当前针对红外烟羽的分割方法研究仍不充分。何自芬等[14] 在 YOLACT(you only look at coefficients)[15] 基础上引入注意力机制来增强特征提取能力, 但该方法基于单幅图像, 未考虑气体运动信息, 提取的特征鲁棒性较差。江逸远等[16] 将气体溯源定位视为关键点检测任务, 使用三维高斯分布对同一烟羽对象的像素时空分布进行建模, 提取气体扩散特征, 实现溯源和分割, 但该方法假设视频中存在可见的稳定泄漏点, 所适用的烟羽场景特征单一。综上, 为了提取鲁棒的 VOCs 烟羽特征以克服现有方法的局限性, 本文提出一种基于序列分析的多模态石化 VOCs 烟羽分割方法 (multimodal petrochemical VOCs plume segmentation method, MPPS)。该方法能够融合 VOCs 泄漏烟羽特有的红外运动扩散矢量信息和可见光模态的互补信息, 增强烟羽特征并抑制背景干扰。本文的主要贡献如下:

1) 提出一种运动感知边缘增强 (motion-aware edge enhancement, MAEE) 模块。与侧重于目标跟踪或简单帧间差分/叠加的时序分割方法[17-18] 相比, MAEE 利用 VOCs 烟羽在时间上从中心到边缘扩散的物理特性, 以目标帧为中心, 计算其前后帧相对于目标帧的烟羽粒子扩散矢量[19], 进而将参考帧烟羽特征根据扩散轨迹传播至目标帧, 实现对目标帧烟羽区域的运动感知增强。

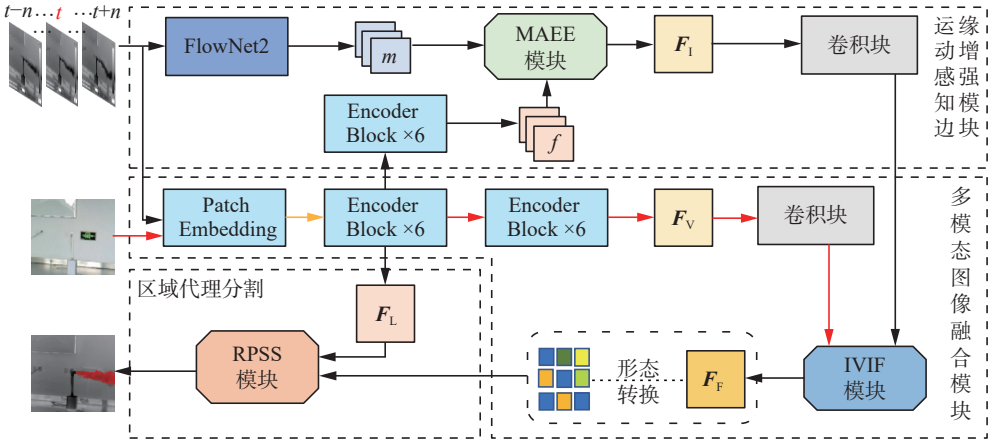
2) 提出一种相似度引导的红外光与可见光图像融合 (infrared and visible image fusion, IVIF) 模块。不同于主流多模态融合策略平等对待所有模态信息或仅关注共存的显著特征[20], IVIF 利用 VOCs 气体仅在红外成像, 而背景干扰在双模态下均成像的特点, 将可见光信息作为“背景鉴别器”直接、高效地抑制大部分背景干扰。

3) 此外, 引入一种基于区域代理的语义分割 (region-proxy-based semantic segmentation, RPSS) 模块。将逐像素分割问题转化为基于区域代理的分类问题[21], 与传统的超像素代理策略相比, 这些代理区域具有灵活的几何形状, 是可学习的语义单元, 能自适应地贴合 VOCs 烟羽无规则的形态, 实现在不牺牲分割精度的前提下提升烟羽分割的解码效率。

1　多模态石化 VOCs 分割网络

本文提出的基于序列分析的多模态石化 VOCs 分割网络, 如图 1 所示。特征提取网络的核心结构采用 Vision Transformer(ViT)的 Patch Embed-

ding 和 Encoder Block 部分[22]。其中，红外运动信息提取器由光流提取网络 FlowNet2[23] 和 MAEE 模块组成，可见光特征融合器包含特征提取网络和 IVIF 模块，解码器主要由 RPSS 模块组成。首先，将待检测的红外图像$\boldsymbol{I}_N, N=t-n, t-(n-1), \cdots, t+n$，同时输入到 FlowNet2 和特征提取网络中，获得目标帧的运动扩散矢量序列$\boldsymbol{M}_N$和特征图序列$\boldsymbol{F}_N$，MAEE 模块参考运动扩散矢量$\boldsymbol{M}_N$将参考帧$\boldsymbol{F}_N$中烟羽中心区的特征传播到目标帧的烟羽边缘区，得到红外特征图$\boldsymbol{F}_{\mathrm{I}}$。可见光模态下的目标帧图像通过特征提取网络得到特征图$\boldsymbol{F}_{\mathrm{V}}$。随后，IVIF 模块通过卷积将$\boldsymbol{F}_{\mathrm{I}}$和$\boldsymbol{F}_{\mathrm{V}}$映射到同一特征空间，之后加权融合可见光特征以得到最终的特征图$\boldsymbol{F}_{\mathrm{F}}$。在解码部分，RPSS 模块利用目标帧的低层次语义图$\boldsymbol{F}_{\mathrm{L}}$，通过多尺度计算像素点与特征点之间的亲和度，获取特征点所代理的区域范围，然后对$\boldsymbol{F}_{\mathrm{F}}$的特征点进行线性分类，得到区域类别，最终生成完整的烟羽分割结果。



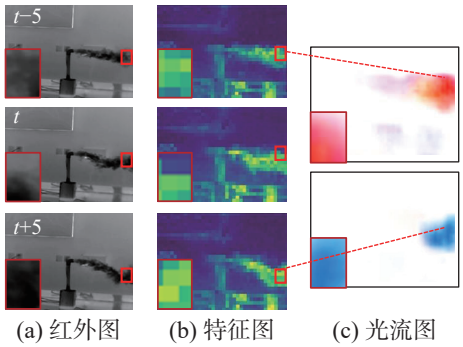
红色线条代表可见光图像处理路径，黑色线条代表红外图像处理路径，黄色线条代表两者的共同路径

图 1　多模态烟羽分割网络结构示意

Fig. 1　Schematic of multimodal plume segmentation network structure

1.1　运动感知边缘增强

红外视频中 VOCs 烟羽根据浓度差异可分为浓度较高、特征明显的烟羽中心区和浓度较低、特征不明显的烟羽边缘区。烟羽运动的本质是中心区域的高浓度烟羽通过湍流扩散作用向边缘区域迁移的过程，该过程中浓度梯度逐渐减小，红外烟羽特征也随之减弱，如图 2 所示。



(a) 红外图　(b) 特征图　(c) 光流图

图 2　烟羽扩散示意

Fig. 2　Schematic of smoke diffusion

烟羽的运动信息表征了其从中心区向边缘区扩散的矢量信息。当受风速、风向等外界因素干扰时，扩散矢量 (如光流幅值) 显著增强[24]。如图 2(b) 所示，中心区域的特征权重明显高于边缘区域，且对于目标帧t的烟羽边缘区灰度值和烟羽特征相比于第$t-5$帧和第$t+5$帧的烟羽中心区都不明显，而对应的边缘区的光流特征明显，体现了烟羽的扩散过程。基于运动扩散矢量，将前后帧$(t-5\sim t+5)$中心区域特征传播至边缘区域，可有效增强边缘特征。

本文提出的 MAEE 模块，如图 3 所示，分为参考帧浓烟特征传播模块和目标帧边界特征增强模块。首先利用参考帧光流图提取前后K个参考帧中烟羽中心区的特征，传播到参考帧特征图的烟羽边缘区。然后根据特征图不同位置的相似度将参考帧特征加权到目标帧特征上，从而增强目标帧边缘区烟羽特征。

参考帧浓烟特征传播模块　光流是空间运动物体在观察成像平面上的像素运动的瞬时速度，在时间间隔很小的视频连续帧之间可以看作目标点的位移。通过光流计算网络 (例如 FlowNet2) 可以计算出参考帧$\boldsymbol{I}_j$相对于目标帧$\boldsymbol{I}_i$的矢量光流图$\boldsymbol{M}_{j\to i}$，其中任意元素$u=(x,y)$代表了当前像素点相对目标帧的位移。根据光流图$\boldsymbol{M}_{j\to i}$计算参考帧$\boldsymbol{I}_j$的坐标位移图$\boldsymbol{C}_{j\to i}$的过程为

$$\boldsymbol{C}_{j\to i}=\boldsymbol{C}_{\mathrm{blank}}+\boldsymbol{M}_{j\to i}=\begin{bmatrix}\boldsymbol{w}_{\mathrm{coor}}\\ \boldsymbol{h}_{\mathrm{coor}}\end{bmatrix}+\boldsymbol{M}_{j\to i} \tag{1}$$

式中：$\boldsymbol{C}_{\mathrm{blank}}$是维度为$w\times h$的原始坐标图，每个元

素代表当前横坐标和纵坐标(w_{coor}, h_{coor}), 将其和光流图$\boldsymbol{M}_{j\to i}$相加可以得到位移后的坐标图$\boldsymbol{C}_{j\to i}$。利用$\boldsymbol{C}_{j\to i}$对参考帧$\boldsymbol{I}_j$的特征图$\boldsymbol{F}_j$在相应坐标进行特征采样, 可以将烟羽中心区特征传播到烟羽边缘区, 得到图$\boldsymbol{D}_{j\to i}$; 进一步, 为了减少背景干扰和冗余信息对特征融合的影响, 将$\boldsymbol{D}_{j\to i}$输入到全卷积子网络中, 并使用注意力机制增强烟羽部分的特征表达, 得到的投影特征$\boldsymbol{E}_{j\to i}$确保了在后续相似度计算中所关注的是烟羽的真实特征而非背景干扰。

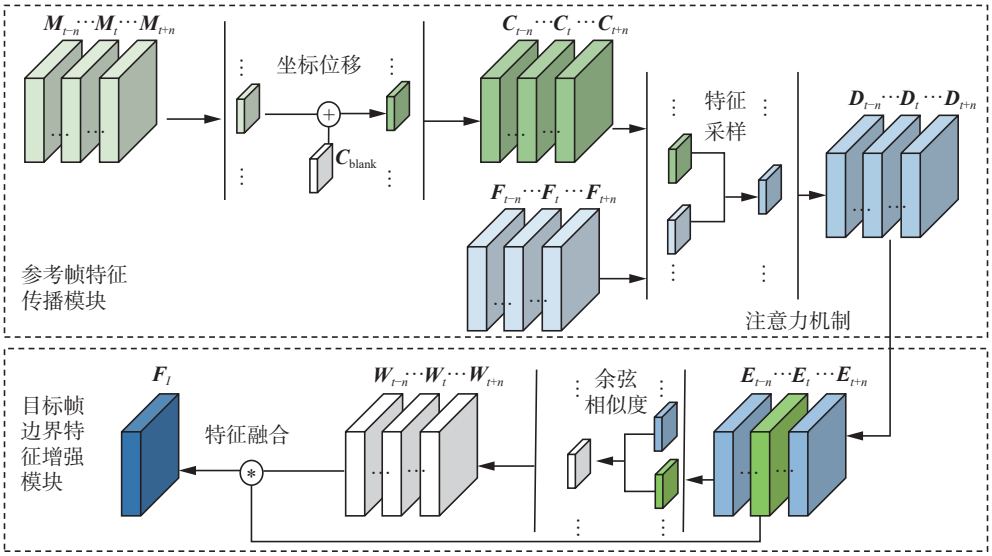


图 3　MAEE 模块结构

Fig. 3　Structure of MAEE module

注意力机制公式为

$$\boldsymbol{E}_{j\to i}=\varepsilon\left(\boldsymbol{D}_{j\to i}\right)=M_{\mathrm{S}}\left(M_{\mathrm{C}}\left(f^{3\times3}\left(\boldsymbol{D}_{j\to i}\right)\right)\right) \tag{2}$$

式中: $\boldsymbol{E}_{j\to i}$代表$\boldsymbol{D}_{j\to i}$的映射特征, 嵌入的子网络$\varepsilon(\cdot)$包括一个3×3的$f^{3\times3}(\cdot)$全卷积网络和后面相邻卷积注意力模块 (convolutional block attention module, CBAM)[25], CBAM 由通道注意力模块$M_{\mathrm{C}}(\cdot)$和空间注意力模块$M_{\mathrm{S}}(\cdot)$顺序连接组成。

目标帧边界特征增强模块　在经过特征传播之后, 参考帧的烟羽边缘区已经包含了烟羽中心区的特征信息, 接下来根据该特征信息增强目标帧的烟羽边缘区特征。特征增强过程为

$$\boldsymbol{F}_{\mathrm{I}}=\sum_{j=i-k}^{i+k}\boldsymbol{W}_{j\to i}\boldsymbol{E}_{j\to i} \tag{3}$$

式中: $[i-k,i+k]$为参考帧范围, 聚合过程中$\boldsymbol{W}_{j\to i}$会对参考帧$\boldsymbol{I}_j$赋予权重, 权重大小与目标帧特征相似度成正相关, 目的是对烟羽边缘区的特征增强。对于不同的参考帧$\boldsymbol{E}_{j\to i}$计算相对目标帧$\boldsymbol{E}_i$的不同位置的权重$\boldsymbol{W}_{j\to i}$, 然后聚合得到最终的红外特征图$\boldsymbol{F}_{\mathrm{I}}$。本文使用余弦相似度[26]计算$\boldsymbol{W}_{j\to i}$, 计算公式分别为

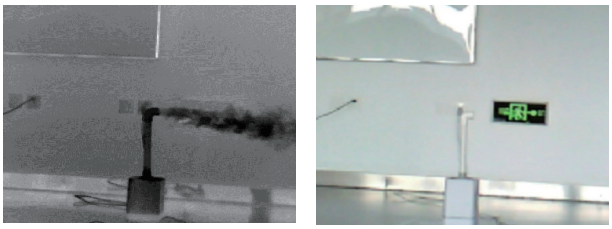
$$\boldsymbol{W}_{j\to i}=\sigma\left(\boldsymbol{S}_{j\to i}\right)=\frac{\mathrm{e}^{\boldsymbol{S}_{j\to i}}}{\sum_{x=i-k}^{i+k}\mathrm{e}^{x\to i}} \tag{4}$$

$$\boldsymbol{W}_{j\to i}=\cos(\theta)=\frac{\boldsymbol{E}_{j\to i}\boldsymbol{E}_i}{\|\boldsymbol{E}_{j\to i}\|\|\boldsymbol{E}_i\|} \tag{5}$$

式中: $\boldsymbol{E}_{j\to i}\boldsymbol{E}_i$是参考帧特征图和目标帧特征图的点积, $\|\boldsymbol{E}_{j\to i}\|\|\boldsymbol{E}_i\|$是各自的范数。使用余弦相似度算法$\cos(\cdot)$计算得到$\boldsymbol{S}_{j\to i}$, 使用 Softmax 函数$\sigma(\cdot)$归一化处理得到$\boldsymbol{W}_{j\to i}$, 使得所有权重和为 1, 有助于强调相似度较高的特征。

1.2　相似度引导的红外与可见光图像融合

红外图像捕捉的是物体的红外波段辐射, 而可见光图像捕捉的是反射的可见光光谱, 这种物理特性的差异导致成像效果存在显著不同。石化 VOCs 烟羽对特定红外波段具有“指纹”吸收特性, 而在可见光下透明不可见, 如图 4 所示。对比图 4(a) 和图 4(b) 可见, VOCs 烟羽泄漏区域的红外与可见光图像结构相似度较低, 而其他区域相似度较高。因此, 可以通过图像结构相似度引导多模态间的特征融合。



(a) 红外成像　　(b) 可见光成像

图 4　红外及可见光 VOCs 成像

Fig. 4　Infrared and visible VOCs imaging

为此，本文提出基于相似度引导的多模态 IV-IF 模块，优化背景噪声的过滤和增强烟羽区域的特征。首先，如图 1 所示，采用双分支特征提取网络来提取多模态的特征图。该网络的嵌入层 (Patch Embedding) 和前 6 个编码层 (Encoder Block) 共享参数，从而学习到两种模态 (红外与可见光) 中相似的低级别特征，如边缘信息和纹理特征。这种共享设计能够有效捕捉两种模态中的共性特征，同时减少模型参数量[27]。接下来，网络的后 6 个编码器块是独立的，分别用于提取红外图像中的 VOCs 红外辐射特征和可见光图像中的视觉反射特征。如图 5 所示，前两行分别为红外和可见光模态下不同层次的特征图，第 3 行为多模态不同层次特征的相似度图，可以看出共享层提取到的背景信息会表现为相似度较高的区域，而独立层则专注于提取烟羽区域的深层特征，能够有效区分不同模态下的烟羽特征。这种设计使得网络能够在保留两种模态特有信息的同时，增强对烟羽区域的敏感度。

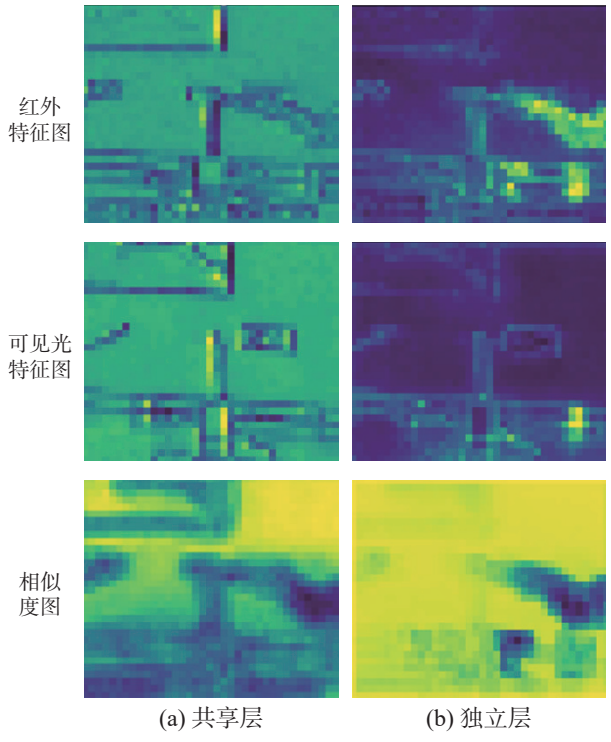


图 5　红外及可见光图像特征相似度对比

Fig. 5　Infrared and visible image feature similarity comparison

在此基础上融合模块，如图 6 所示，输入双分支特征提取网络得到的可见光特征图 $\boldsymbol{F}_{\mathrm{V}}$ 和经过 MAEE 模块增强后的红外特征图 $\boldsymbol{F}_{\mathrm{I}}$。由于两种模态在特征空间上存在差异，因此使用余弦相似度算法计算相似度矩阵 $\boldsymbol{S}$，以量化它们在结构上的相似程度，进一步将 $\boldsymbol{S}$ 归一化，将 $\boldsymbol{1}-\boldsymbol{S}$ 作为权重矩阵 $\boldsymbol{W}$，利用权重矩阵对红外模态的特征图进行逐元素相乘加权，得到加权后的特征图 $\boldsymbol{M}_{\mathrm{I}}$，从而使得特征图中相似度较低的区域 (即 VOCs 烟羽区域) 得到强化，同时减少背景干扰。加权后的特征图和原始红外特征图 $\boldsymbol{F}_{\mathrm{I}}$ 在通道维度上进行拼接，以确保性能的提高。最后，拼接后的特征图经过卷积网络和激活函数的处理，进一步降维和激活特征图中 VOCs 烟羽区域。特征融合过程为

$$\boldsymbol{F}_{\mathrm{M}}=\mathrm{ReLU}\left(f^{3\times3}\left(\mathrm{concat}\left(\boldsymbol{W}\boldsymbol{F}_{\mathrm{I}},\boldsymbol{F}_{\mathrm{I}}\right)\right)\right) \tag{6}$$

$$\boldsymbol{W}=\boldsymbol{1}-\mathrm{standard}\left(\cos\left(\boldsymbol{F}_{\mathrm{I}},\boldsymbol{F}_{\mathrm{V}}\right)\right) \tag{7}$$

式中：$\boldsymbol{F}_{\mathrm{I}}$ 代表融合了运动扩散矢量的红外特征图，$\boldsymbol{F}_{\mathrm{V}}$ 代表可见光特征图，$\boldsymbol{F}_{\mathrm{M}}$ 是融合得到的特征图，ReLU(·) 代表 ReLU（Rectified linear unit）激活函数，$f^{3\times3}(\cdot)$ 是 3×3 的卷积网络，concat(·) 是通道拼接操作，standard(·) 是归一化操作。

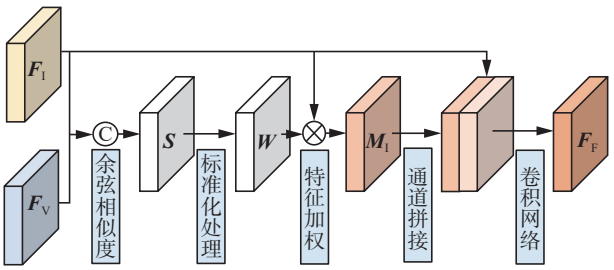


图 6　IVIF 模块结构

Fig. 6　Structure of IVIF module

1.3　基于区域代理的语义分割

Zhang 等[21] 用区域代理特征分类解码的方式代替传统的逐像素分类。区域代理特征将图像划分为可学习的、形状灵活的区域代理块，具体来说，该方法利用超像素分割原理，将特征点作为区域代理，通过多尺度卷积操作计算每个像素与周围特征点的亲和度生成亲和度图。亲和度图能够自适应地将像素聚合到局部细节和全局上下文都得到体现的区域代理中。这种区域代理块可以自适应地贴合石化 VOCs 烟羽的无规则形态，为此本文引入一种基于区域代理的语义分割模块即 RPSS，用区域代理特征分类解码的方式代替逐像素分类，同时通过减少逐像素计算提升计算效率。

RPSS 将特征图的每一个特征点 f 看作原图中一部分区域 R 的代理特征，对特征点进行分类，可以得到 R 的类别 C。通过计算原图中每一个像素点对区域 R 的亲和度，可以得到其属于类别 C 的概率。即 RPSS 分为区域亲和度计算和区域特征分类两个模块。

区域亲和度计算模块　其网络结构如图 7 虚线框内部分所示。输入为包含较多局部纹理、边缘和形状等细节信息的底层特征图，其维度为

$N\times D$, 其中D表示特征图通道数, $N=H\times W$为特征图的空间尺寸。为了显式地捕捉不同尺度上的局部和全局特征, 将特征图输入到空洞空间金字塔池化 (atrous spatial pyramid pooling, ASPP)[27] 模块进行多尺度特征提取。随后通过卷积模块将特征维度调整为$H\times W\times khw$, 并进一步通过维度变换得到$Hh\times Ww\times k$的亲和度图, 其中Hh和Ww分别为原图的高度和宽度, k表示像素p周围相邻的区域个数。此过程中, 原图被划分为H行W列的网格, 每个网格单元尺寸大小为$h\times w$, 特征图的特征点f对应一个网格作为代理的区域R, 像素p属于周围k个区域的概率由输出的亲和度图直接表示。像素p的亲和度$q_a(p)$, 满足归一化条件:

$$\sum_{a=1}^{k} q_a(p)=1 \tag{8}$$

式中: $a\in N_p$为像素p周围的区域, 如图 7 所示。

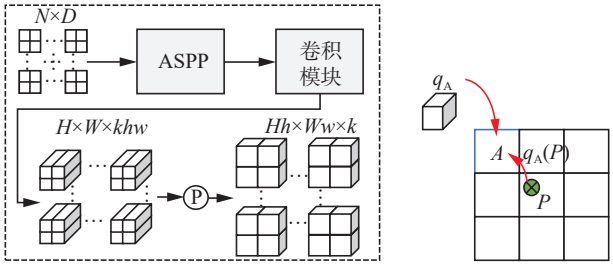


图 7　区域亲和度计算模块示意

Fig. 7　Schematic of regional affinity calculation module

区域特征分类模块　因为将特征点f看作区域R的代理特征, 且获得了待测图像素点p对周围特征点f的亲和度图$\boldsymbol{Q}$, 所以将每个特征点f进行线性分类即可得到 VOCs 的语义分割结果。维度为$N\times D$的特征图经过线性分类器之后变为$N\times 2$ (VOCs 烟羽前景和背景两类) 的类别图$\boldsymbol{Y}$, 其中$N=H\times W$。结合亲和度图$\boldsymbol{Q}$可以得到每个像素点p的类别计算公式为

$$C(p)=\sum_{a=1}^{k} y(a)q_a(p) \tag{9}$$

式中: $y(a)\in\boldsymbol{Y}$是领域a的类别, $q_a(p)\in\boldsymbol{Q}$是像素点p对于领域a的亲和度。这样就得到了最终的分割概率图, 并且由于只使用了一个线性分类器, 计算量相对于逐像素分割大大降低。可以简单证明如下: 假设使用L个卷积解码器进行上采样分类, 第i层的卷积核为$K_i\times K_i$, 包含M_i个卷积核, 输入特征图尺寸$H_i\times W_i\times D_i$的计算量为

$$\partial_{\text{pixel}}=\sum_{i=1}^{L} H_i\times W_i\times D_i\times K_i^2\times M_i \tag{10}$$

使用区域特征分类的计算量为

$$\partial_{\text{region}}=H\times W\times D \tag{11}$$

式中: $H\times W\times D$为特征图的尺寸, 可以看出$\partial_{\text{region}}\ll\partial_{\text{pixel}}$, 区域代理分割计算复杂度远低于逐像素分割。

2　模型分割结果及分析

2.1　损失函数设计

总的损失函数L_{total}由两部分组成, L_{IVIF}用于监督 IVIF 模块, L_{S}用于监督最后的分割结果:

$$L_{\text{total}}=\alpha L_{\text{IVIF}}(\boldsymbol{M}_{\text{S}},\boldsymbol{S})+\beta L_{\text{S}}(\boldsymbol{F}_{\text{F}},\boldsymbol{G}) \tag{12}$$

式中: α、β为权重参数; $\boldsymbol{F}_{\text{F}}$是最后的分割预测图; $\boldsymbol{G}$为标注图; $\boldsymbol{M}_{\text{S}}$为多模态融合模块生成的相似度权重矩阵, 范围在 [0,1]; $\boldsymbol{S}$为标注的目标相似度矩阵, 利用标注图$\boldsymbol{G}$生成, 气体区域相似度为 0, 背景区域相似度为 1。

损失函数L_{IVIF}是对比学习损失用于衡量红外和可见光特征之间的差异, 使模型输出的相似度接近目标差异矩阵的值, 保证在不同区域的特征差异能被有效学习到, 公式为

$$L_{\text{IVIF}}(\boldsymbol{M}_{\text{S}},\boldsymbol{S})=\frac{1}{N}\sum_{i=1}^{N}\left(s_i\cdot\max(0,m-m_i)^2+(1-s_i)\cdot m_i^2\right) \tag{13}$$

式中: N为图像中像素总数, m_i代表$\boldsymbol{M}_{\text{S}}$中位置$i$处的相似度, s_i代表标注的相似度, 超参数m为烟羽区域不相似特征的最大余弦相似度。在第 1 部分$s_i\cdot\max(0,m-m_i)^2$中, 当s_i为 1 时表明该位置是背景区域, 希望$\boldsymbol{M}_{\text{S}}$和$\boldsymbol{S}$的特征相似度$m_i$尽可能大, 如果$m_i$的距离小于$m$则推动模型将相似特征对的余弦相似度增大。第 2 部分$(1-s_i)\cdot m_i^2$, 当s_i为 0 时表示该位置为烟羽区域, 希望该位置特征相似度m_i尽可能小, 促使模型将不同模态下烟羽区域特征映射到不同空间。

L_{S}是二分类交叉熵损失函数, 用于监督最后的烟羽分类, 公式为

$$L_{\text{S}}(\boldsymbol{F}_{\text{F}},\boldsymbol{G})=-\frac{1}{N}\cdot\sum_{i=1}^{N}\left(f_i\log g_i+(1-f_i)\log(1-g_i)\right) \tag{14}$$

式中: f_i为预测图$\boldsymbol{F}_{\text{F}}$第 i 个像素点, g_i为标注图$\boldsymbol{G}$第i个像素点。

网络的最终目的是为了最小化解码层的损失, 所以将α和β权重根据损失值动态调整, 公式为

$$\alpha=\frac{L_{\text{IVIF}}}{(L_{\text{IVIF}}+L_{\text{S}})},\quad \beta=\frac{L_{\text{S}}}{(L_{\text{IVIF}}+L_{\text{S}})} \tag{15}$$

如果发现L_{IVIF}损失较大, 则说明模型在学习相似特征时存在困难, 此时增大α的值; 反之L_{IVIF}

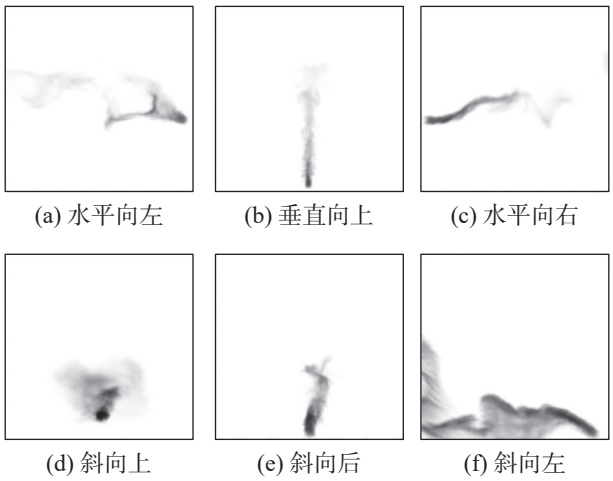
损失较小，则说明在根据特征图进行解码时存在困难，则增大β的值。动态调整这两个超参数可以让模型根据不同的训练阶段或样本情况自适应地优化，使得不同类型的损失对模型的训练发挥最大作用。

2.2　实验环境设置

所有实验均在 Ubuntu 22.04.3 LTS 版本操作系统上进行，CPU 为 Intel(R) Core(TM) i7-10700K，GPU 为单张 NVIDIA GeForce RTX 3090 24 GB，训练及测试代码均基于 MMCV 深度学习框架实现，实验训练批次设置为 16 000，动量为 1.0，学习率采用“Poly”策略，初始学习率设置为 0.000 06，训练过程中随着迭代次数增加而逐渐衰减学习率，采用权重衰减的 AdamW (Adam + Weight decay)[28] 算法训练模型。

2.3　数据集

目前针对 VOCs 烟羽分割尚无公开数据集，本文按照 PASCAL VOC 数据集的格式建立了场景丰富的烟羽分割数据集，整体数据集由合成数据和实地拍摄数据组成。合成数据使用 Blender 软件生成，在水平面中前、后、左、右等 17 个方向的气体飘散模型模拟 VOCs 烟羽，如图 8 所示，共截取 4 000 帧连续图像。实验室在多个背景和多个烟羽扩散条件下拍摄 32 段红外 VOCs 扩散视频，截取 1 000 帧连续图像进行手工标注。训练集和测试集的比例为 4∶1。每个场景下均拍摄了同一视角的可见光视频，并进行了图像对齐处理。



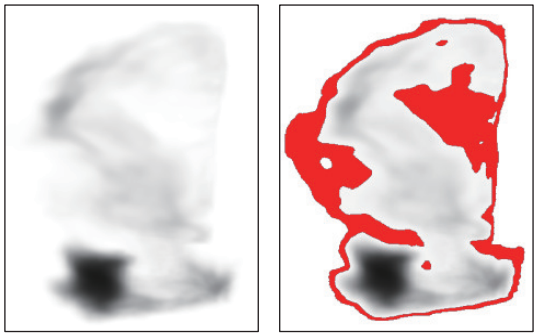
(a) 水平向左　(b) 垂直向上　(c) 水平向右

(d) 斜向上　(e) 斜向后　(f) 斜向左

图 8　VOCs 烟羽合成示意

Fig. 8　Schematic of VOCs gas synthesis

另外，对合成烟羽图像进行进一步处理，合成烟羽包含浓度极低和背景无法区分的区域，将该区域去除，避免生成标注图时引入噪声，如图 9 所示，红色掩码区域为去除的难以分辨的模糊烟羽边缘区域。



(a) 合成气体　(b) 合成气体后处理

图 9　合成数据后处理

Fig. 9　Synthetic data post-processing

2.4　参数分析

在 MAEE 模块中参考帧的数量K的取值会影响烟羽边缘特征的增强效果。为探讨合适的取值，在训练和测试环节设置不同的K值组合来进行实验，实验结果如表 1 所示。组别 1 的分割结果最差，这是由于 MAEE 模块能够提取到的烟羽边缘特征与输入的参考帧数量相关，参考帧数量过少时能够获取的烟羽特征不明显，进而影响最后的分割效果。固定训练参考帧K_{train}，逐渐增加测试参考帧K_{test}的大小如组别 (1~6) 和组别 (7、9~12) 可以发现 IoU 逐渐提高，但是当参考帧增大到一定程度后 IoU（intersection over union）提高微小如组别 (3~6) 和组别 (9~12) 说明融合的特征达到一定程度之后，继续增加参考帧无法提供更多的有效特征。从表中分割速度列可以看到提高参考帧 K 的数量会增加光流计算和特征提取的任务量，降低模型实时性，并且从表中可以看到K_{test}相同时K_{train} 增加对分割效果提升较少，但是实际训练时间会大大增加，所以综合分割效果和计算效率，本文选择K_{train}为 2，K_{test}为 9。

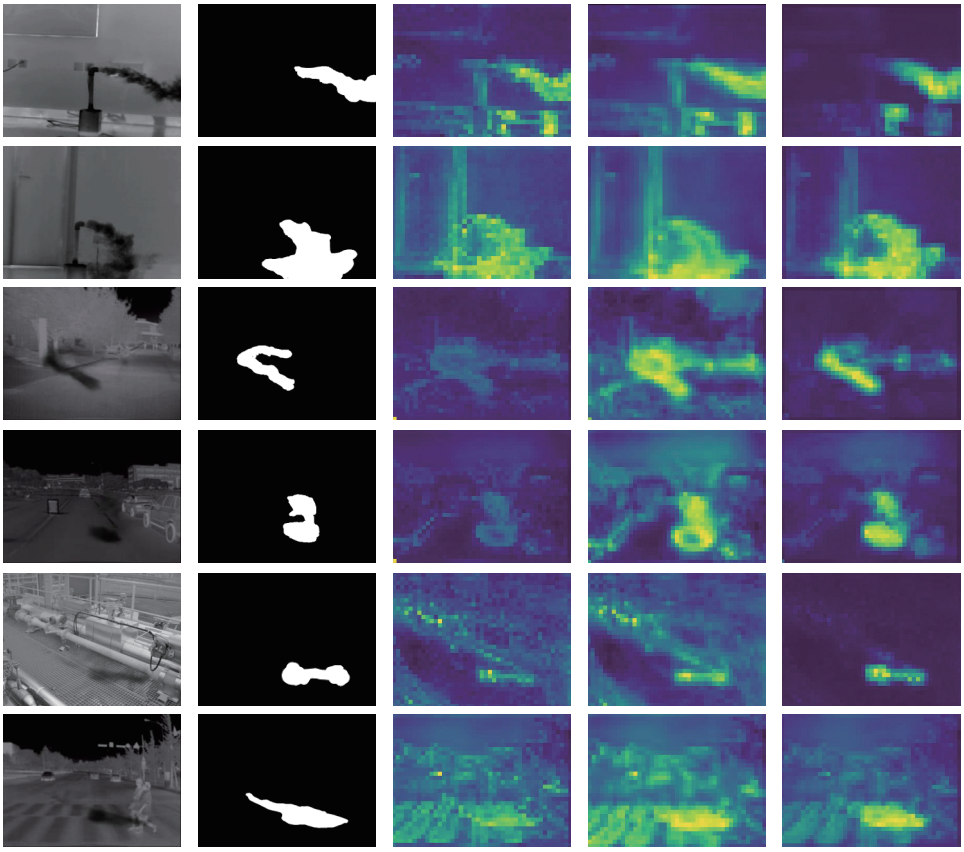
表 1　不同参考帧数目组合的实验结果

Table 1　Experimental results of different combinations of reference frame numbers

组别	K_{train}	K_{test}	IoU/%	分割速度/(帧/s)
1	2	5	85.82	33.67
2	2	7	86.45	31.75
3	2	9	87.04	29.94
4	2	11	87.04	28.66
5	2	13	87.03	25.31
6	2	15	87.05	22.96
7	5	5	85.91	33.67
8	5	7	86.43	31.75
9	5	9	87.04	29.94
10	5	11	87.05	28.66
11	5	13	87.04	25.31
12	5	15	87.06	22.96

2.5　训练可视化

为验证提出的模型是否提取到了前后参考帧的运动扩散矢量, 以及融合可见光特征信息是否有效激活烟羽区域特征, 实验中对烟羽图像、标签和相关特征图进行了可视化, 如图 10 所示: (a) 为烟羽图像; (b) 为烟羽图像的标注图$\boldsymbol{G}$; (c) 为特征提取网络获取的特征图$\boldsymbol{F}$; (d) 为融合了前后帧运动扩散矢量的特征图$\boldsymbol{F}_{\mathrm{I}}$; (e) 为融合可见光特征的特征图$\boldsymbol{F}_{\mathrm{F}}$, 即最后分割所使用到的特征图。可以看出, 融合了前后帧运动扩散矢量的特征图$\boldsymbol{F}_{\mathrm{I}}$相比于$\boldsymbol{F}$能够有效地增强烟羽边缘特征, 烟羽特征边界更平滑。且从第 4~6 行可以明显看出由于前后帧特征的累加, $\boldsymbol{F}_{\mathrm{I}}$烟羽主体部分的特征激活也变得更加强烈。进一步观察$\boldsymbol{F}_{\mathrm{F}}$会发现, 由于参考了可见光特征区域, 特征图中背景的特征强度被大大削弱, 烟羽特征和背景的对比更加明显, 与标注图$\boldsymbol{G}$的差异最小, 这说明模型可以结合前后参考帧的运动扩散矢量完善边缘位置的特征信息, 而且输入的可见光模态能够有效地激活烟羽位置的特征, 从而提升烟羽分割效果。



(a) 烟羽图　(b) 标注图 $\boldsymbol{G}$　(c) 特征图 $\boldsymbol{F}$　(d) 运动特征图 $\boldsymbol{F}_{\mathrm{I}}$　(e) 融合特征图 $\boldsymbol{F}_{\mathrm{F}}$

图 10　可视化分析

Fig. 10　Visualization analysis

2.6　定量分析

本文用交并比 (IoU)[29]、参数量作为标准来评价烟羽分割算法的性能, IoU 是评估语义分割模型性能的核心指标, 其通过计算预测结果与真实标注的重叠程度来衡量分割精度; 模型参数量是指网络中所有可学习参数的总和, 是衡量模型复杂度的关键指标之一。为了对比测试, 本文搭建经典的分割网络 DeepLab 系列[30]、RGB-T(RGB-thermal)多模态网络 GMNet(graded-feature multilabel-learning network)[31] 等, 并针对本文模型使用的 ViT-T/16[22] 骨架网络搭建 Segmenter[32] 网络作为基线网络, 计算出各自的烟羽交并比和参数量, 其结果如表 2 所示。简单的 U-Net[33] 网络和 SegNet 网络[34] 对于红外成像下的 VOCs 分割表现较差, DeepLab+[30] 使用空洞卷积和 ASPP 之后, 特征提取能力得到提高, 分割效果有所改善。GMNet 使用了多层次的融合策略能够利用可见光模态的背景上下文信息, 且使用边界损失, 分割精度在非 Transformer 架构中是最高的; Segmenter 使用了 Transformer 的骨架网络, 能够提取长距离的烟羽特征, 有助于更好地提取烟羽边缘特征, 提高分割精准度。并且从表中数据可以看出, 本文方法相比使用了相同 ViT-T/16 骨架网络的基线网络 Segmenter, 参数量减少了 1.30×10^{6}, 分

割速度提升了 1.81 帧/s, 同时 IoU 提高了 3.53%。

表 2　分割结果对比

Table 2　Comparison of segmentation results

方法	骨架网络	参数量/10^6	分割速度/(帧/s)	IoU/%
U-Net	FCN	29.06	35.88	62.37
SegNet	Vgg-16	3.72	56.64	64.27
D.LabV3+	MobileNetV2	16.38	44.72	80.25
GMNet	ResNet50	28.63	36.97	82.93
Segmenter	ViT-T/16	6.71	28.13	83.51
本文方法	ViT-T/16	5.41	29.94	87.04

2.7　消融分析

为验证本文提出的各个模块的有效性, 对模型中提出的 MAEE、IVIF、RPSS 模块进行了消融实验, 以烟羽的 IoU 和分割速度作为评价指标。实验结果如表 3 所示, 消融实验的基准为同样使用 ViT-T/16 作为特征提取网络的 Segmenter, 对应实验组别 1。从表中可以看出, 所有模块都对分割效果 IoU 起到提升的作用, 下面对消融实验进行详细分析。

表 3　消融实验结果

Table 3　Results of ablation experiments

组别	MAEE	IVIF	RPSS	IoU/%	分割速度/(帧/s)
1	—	—	—	83.51	28.13
2	√	—	—	85.33	20.59
3	—	√	—	84.60	22.28
4	—	—	√	83.64	56.36
5	—	√	√	85.70	44.86
6	√	—	√	86.14	32.46
7	√	√	—	86.90	17.58
8	√	√	√	87.04	29.94

1) 单模块效果分析: 为验证单模块对模型的提升作用, 设计了组别 2、3、4, 分别加入 MAEE、IVIF、RPSS 模块观察分割结果和分割速率。组别 2 相比组别 1, IoU 提升了 1.82%, 说明加入运动信息融合模块之后能够利用到前后帧的运动扩散矢量, 然后加强到当前帧的烟羽特征上; 组别 3 相比于组别 1, IoU 提升了 1.09%, 说明根据多模态之间的相似度计算调整烟羽特征的强度可以明显地提高分割的精度, 正如前面分析所述, 可以激活烟羽区域特征削弱背景特征; 组别 4 相比于组别 1 在保持分割效果提升的同时, 显著地提高了单帧的分割效率, 表明区域分割解码器中直接对特征区域线性分类能够大幅减少计算量。

2) 多模块效果分析: 为了验证多个模块之间组合起来的效果, 设计了多组组合实验。首先对比组别 1、2、3、7 发现, 同时加入 MAEE 和 IVIF 模块的效果提升比单独使用组别 2、3 的提升效果加起来都多, 说明两个模块的提升性作用是正向叠加的。其次结合组别 2、3、5、6 发现, MAEE 模块引入的前后参考帧和 IVIF 引入的可见光参考帧增加的计算负担可以被 RPSS 模块有效缓解, 同时还能够和其他模块共同提升效果。最后对比组别 1 和组别 8 可以发现, 本文算法对比基线网络 Segmenter 方法, 在引入了参考帧之后分割速率仍有小幅度提升, 同时分割精度提高了 3.53%。

综合分析可知, MAEE 模块和 IVIF 模块可以很好地提升烟羽分割性能, RPSS 模块可以作为优化模块附加到网络中降低模型复杂度。

2.8　定性分析

为更直观地观察模型的分割性能, 选择经典的语义分割网络 U-Net、SegNet、GMNet、Segmenter 和本文提出的网络模型在标注的 VOCs 烟羽测试集上进行更为直观的定性分析, 分割结果如图 11 所示。可以看出, 在针对红外 VOCs 的场景下, 其他网络难以过滤背景干扰, 可能产生较大的分割误差或者不能很好地拟合烟羽的边界, 在图中圈出了其对比真实值分割的不足。首先, 除本文方法外, 其余测试的分割网络都存在错误分割的现象: 单模态网络几乎都存在会把背景中的干扰信息误当作烟羽特征, 说明在红外场景下无法有效地过滤特征和烟羽相似的干扰信息; 多模态网络 GMNet 由于 VOCs 只在红外成像的特殊性, 在和可见光模态融合时同样会引入来自可见光背景的干扰, 导致在部分场景下同样存在背景误分割问题, 如第 5 列场景发生了较大的误识别, 无法排除全部的背景干扰。此外, U-Net 和 SegNet 在每一幅图中都有烟羽主体部分错误分割, 而且边缘分割比较粗糙, 部分图像存在较大的分割错误。GMNet 网络得益于其多策略的监督函数, 烟羽主体避免了大面积的分割错误, 并且在小烟羽场景下更好, 如第 3、4 列场景。Segmenter 的分割效果优于 GMNet, 例如在 1~4 列场景中的边缘细节更加清楚, 能够在不同尺度上捕捉特征, 对烟羽细小部分也能较好捕捉, 仅在第 5、6 列场景中存在较严重的过分割现象。本文提出的网络虽然也存在一定的分割误差, 但相比于其他的网络, 其 IVIF 模块通过融合可见光信息有效过滤了背景干扰, 激活 VOCs 烟羽主体特征;

MAEE 模块将前后帧特征根据运动扩散矢量传播到烟羽边界, 更好地保留了烟羽的边界等细节信息。分割的结果在保证烟羽主体的完整的情况下, 对烟羽边界的拟合更加精细、完整和平滑, 从而体现了提出的运动信息融合模块、多模态权重自适应模块和区域解码模块的有效性。

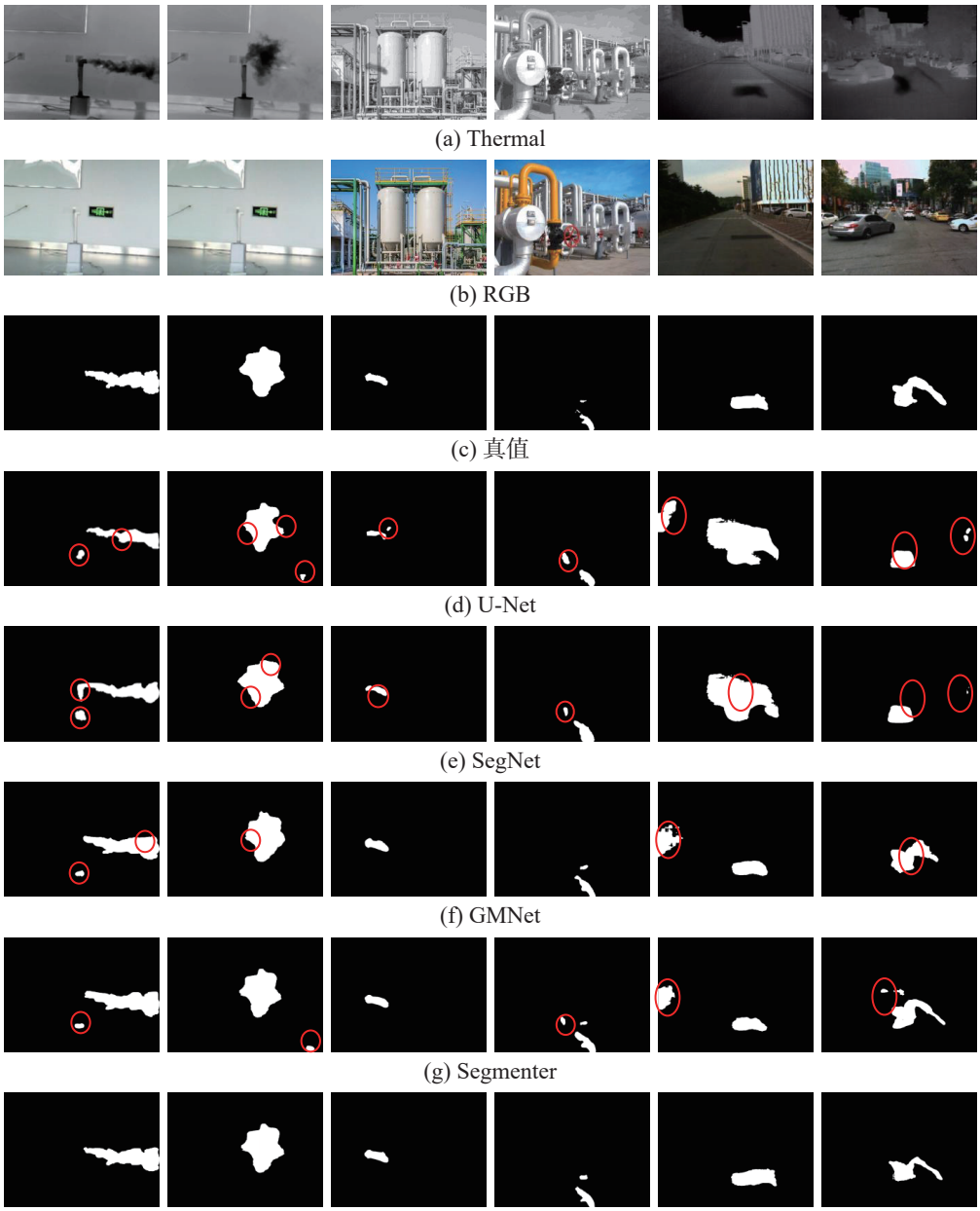


图 11　不同网络分割结果对比

Fig. 11　Comparison of different network segmentation results

3　结束语

针对石化 VOCs 烟羽在红外成像条件下存在因弱分辨力低和低对比度导致的现有烟羽检测算法分割效果不佳和边界拟合粗糙问题, 本文提出了基于序列分析的多模态石化 VOCs 烟羽分割网络。首先, 运动信息融合模块 MAEE 通过计算参考帧的运动扩散矢量信息, 将烟羽中心区特征扭曲到烟羽边缘区, 增强了网络对于模糊边界特征的识别能力。其次, 可见光融合模块 IVIF, 引入 VOCs 的可见光模态特征可以更好地激活烟羽区域的特征, 过滤背景区域的干扰。最后, 引入基于区域代理的分割模块 RPSS, 将逐像素的解码改为区域解码。通过在标注的 VOCs 数据集上开展网络实验, 本文提出的网络在多模态损失函数的监督下对比基线网络能够在保持分割速率的情况下显著提高烟羽的分割精度, 并且能拟合出更好的烟羽边界, 为进一步的研究提供了思路, 具有良好的理论和应用价值。未来通过进一步深入研

究气体扩散动态特性，以提升模型对于复杂气体行为的建模能力。

参考文献：

[1] 张振杰, 李志平, 张苗苗. 红外成像技术在石化装置易挥发性气体泄漏检测中的应用[J]. 山东化工, 2015, 44(12): 159−162.
ZHANG Zhenjie, LI Zhiping, ZHANG Miaomiao. Infrared thermal imaging technology in petrochemical device application of volatile gas leak detection[J]. Shandong chemical industry, 2015, 44(12): 159−162.

[2] 迟晓铭. 石化企业气体泄漏红外成像检测技术实验研究与分析[J]. 红外技术, 2024, 46(8): 947−956.
CHI Xiaoming. Experimental research and analysis of infrared imaging detection technology for gas leakage in petrochemical enterprises[J]. Infrared technology, 2024, 46(8): 947−956.

[3] WANG Jingfan, TCHAPMI L P, RAVIKUMAR A P, et al. Machine vision for natural gas methane emissions detection using an infrared camera[J]. Applied energy, 2020, 257: 113998.

[4] SHOJAIEE F, BALEGHI Y. EFASPP U-Net for semantic segmentation of night traffic scenes using fusion of visible and thermal images[J]. Engineering applications of artificial intelligence, 2023, 117: 105627.

[5] HUO Dong, WANG Jian, QIAN Yiming, et al. Glass segmentation with RGB-thermal image pairs[J]. IEEE transactions on image processing, 2023, 32: 1911−1926.

[6] LI Xiangtai, ZHANG Wenwei, PANG Jiangmiao, et al. Video K-Net: a simple, strong, and unified baseline for video segmentation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 18825−18835.

[7] YANG Yijun, XING Zhaohu, YU Lequan, et al. Vivim: a video vision mamba for medical video segmentation [EB/OL]. (2024−01−25)[2025−08−29]. https://arxiv.org/abs/2401.14168.

[8] DENG Xing, YU Zhongming, WANG Lin, et al. Smoke image segmentation based on color model[J]. Journal on innovation and sustainability, 2015, 6(2): 130.

[9] MA Zongfang, CAO Yonggen, SONG Lin, et al. A new smoke segmentation method based on improved adaptive density peak clustering[J]. Applied sciences, 2023, 13(3): 1281.

[10] YE Shiping, BAI Zhican, CHEN Huafeng, et al. An effective algorithm to detect both smoke and flame using color and wavelet analysis[J]. Pattern recognition and image analysis, 2017, 27(1): 131−138.

[11] WANG Zewei, YANG Pengfei, LIANG Haotian, et al. Semantic segmentation and analysis on sensitive parameters of forest fire smoke using smoke-unet and landsat-8 imagery[J]. Remote sensing, 2022, 14(1): 45.

[12] YUAN Feiniu, SHI Yu, ZHANG Lin, et al. A cross-scale mixed attention network for smoke segmentation[J]. Digital signal processing, 2023, 134: 103924.

[13] 洪少壮, 胡英, 于宏伟. 基于多特征的红外成像VOCs气体检测[J]. 计算机仿真, 2021, 38(3): 374−379.
HONG Shaozhuang, HU Ying, YU Hongwei. Infrared imaging VOCs gas detection based on multi-feature[J]. Computer simulation, 2021, 38(3): 374−379.

[14] 何自芬, 曹辉柱, 张印辉, 等. 融合注意力分支特征的甲烷泄漏红外图像分割[J]. 红外技术, 2023, 45(4): 417−426.
HE Zifen, CAO Huizhu, ZHANG Yinhui, et al. Infrared image segmentation of methane leaks incorporating attentional branching features[J]. Infrared technology, 2023, 45(4): 417−426.

[15] BOLYA D, ZHOU Chong, XIAO Fanyi, et al. YOLACT: real-time instance segmentation[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 9157−9166.

[16] 江逸远, 谷小婧, 顾幸生. 基于红外视频的VOCs泄漏源定位与气羽实例分割[J]. 华东理工大学学报 (自然科学版), 2024, 50(5): 695−707.
JIANG Yiyuan, GU Xiaojing, GU Xingsheng. VOCs leakage source location and gas plume instance segmentation based on infrared video[J]. Journal of East China University of Science and Technology, 2024, 50(5): 695−707.

[17] ZHOU Tianfei, LI Jianwu, LI Xueyi, et al. Target-aware object discovery and association for unsupervised video multi-object segmentation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 6981−6990.

[18] GARG S, GOEL V, KUMAR S. Unsupervised video object segmentation using online mask selection and space-time memory networks[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020.

[19] WU Yuanlu, CHEN Minghao, WO Yan, et al. Video smoke detection base on dense optical flow and convolutional neural network[J]. Multimedia tools and applications, 2021, 80(28): 35887−35901.

[20] ZHANG Yifei, SIDIBÉ D, MOREL O, et al. Deep multimodal fusion for semantic image segmentation: a survey[J]. Image and vision computing, 2021, 105: 104042.

[21] ZHANG Yifan, PANG Bo, LU Cewu. Semantic segmentation by early region proxy[C]//2022 IEEE/CVF Confer-

ence on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 1248−1258.

[22] DOSOVITSKIY A, BEYER L, KOLESNIKOVA, et al. An image is worth 16×16 words: transformers for image recognition at scale[EB/OL]. (2020−10−22)[2024−01−01]. https://arxiv.org/abs/2010.11929.

[23] ILG E, MAYER N, SAIKIA T, et al. FlowNet 2.0: evolution of optical flow estimation with deep networks[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 1647−1655.

[24] ZHU Liying, WANG Ang, JIN Fang. Using image processing technology and general fluid mechanics principles to model smoke diffusion in forest fires[J]. Fluid dynamics & materials processing, 2021, 17(5): 1213−1222.

[25] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 3−19.

[26] LUO Chunjie, ZHAN Jianfeng, XUE Xiaohe, et al. Cosine normalization: using cosine similarity instead of dot product in neural networks[C]//Artificial Neural Networks and Machine Learning–ICANN 2018. Cham: Springer International Publishing, 2018: 382−391.

[27] 周晓君, 高媛, 李超杰, 等. 基于多目标优化多任务学习的端到端车牌识别方法[J]. 控制理论与应用, 2021, 38(5): 676−688.
ZHOU Xiaojun, GAO Yuan, LI Chaojie, et al. Multi-objective optimization based multi-task learning for end-to-end license plates recognition[J]. Control theory & applications, 2021, 38(5): 676−688.

[28] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization[C]//International Conference on Learning Representations, [S.l.]: OpenReview.net, 2020.

[29] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge[J]. International journal of computer vision, 2010, 88(2): 303−338.

[30] CHEN L C, ZHU Yukun, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 833−851.

[31] ZHOU Wujie, LIU Jinfu, LEI Jingsheng, et al. GMNet: graded-feature multilabel-learning network for RGB-thermal urban scene semantic segmentation[J]. IEEE transactions on image processing, 2021, 30: 7790−7802.

[32] STRUDEL R, GARCIA R, LAPTEV I, et al. Segmenter: transformer for semantic segmentation[C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 7242−7252.

[33] 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 International Publishing, 2015: 234−241.

[34] 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.

作者简介:

王子豪, 硕士研究生, 主要研究方向为计算机视觉、图像分割。E-mail: wzh8096@mail.ustc.edu.cn。

夏秀山, 副研究员, 主要研究方向为计算机视觉、多模态信息处理。主持、参与国家和省部级科研项目 10 余项, 发表学术论文 10 余篇。E-mail: xiaxiushan@iat.ustc.edu.cn。

曹洋, 副教授, 博士生导师, 主要研究方向计算机视觉、智能机器人。主持国家重点研发计划项目子课题、国家自然科学基金项目等, 获中国自动化学会科技奖一等奖 1 项, 发表学术论文 50 余篇。E-mail: forrest@ustc.edu.cn。