文章编号：1001-4632（2025）06-0187-12
引用格式：秦勇，李威，曹志威，等. 无监督学习下悬挂式单轨列车前向异物检测方法[J]. 中国铁道科学，2025，46(6)：187-198.
Citation：QIN Yong，LI Wei，CAO Zhiwei，et al. An Unsupervised Learning Method for Forward Foreign Object Detection in Suspended Monorail Systems [J]. China Railway Science，2025，46 (6)：187-198.

无监督学习下悬挂式单轨列车前向异物检测方法*

秦 勇[1]，李 威[1,2]，曹志威[1]，连丽容[1,2]，高 阳[1,2]，徐 磊[3]，刘玉文[3]，赵长龙[3]

（1. 北京交通大学 先进轨道交通自主运行全国重点实验室，北京 100044；
2. 北京交通大学 交通运输学院，北京 100044；
3. 中车青岛四方机车车辆股份有限公司，山东 青岛 266111）

摘 要：针对悬挂式单轨列车前向未知异物检测精度低的问题，提出1种高效的端到端无监督学习检测方法DPMF-Flow。首先，通过跨尺度特征提取模块，并行捕获细粒度的局部细节与高层全局语义，以增强在复杂背景下对多尺度异物的表征能力；其次，通过双路径融合流模块分别提取局部和全局特征，并由互补的概率密度估计生成互补的隐变量，有效提升定位精度；最后，通过目标分割定位模块整合双路径隐变量信息，生成像素级目标定位热图。结果表明：在悬挂单轨异物入侵数据集中，所提方法的检测精度可达99.34%，并以66.04 FPS的帧率实时检测异物；与Padim和Fastflow等无监督方法相比，该方法具有显著的性能优势，在公开数据集MVTec AD上表现出良好的泛化性，可为轨道交通安全提供实用且先进的技术方案。

关键词：无监督目标检测；悬挂式单轨；列车安全；计算机视觉；深度学习
中图分类号：U216.3 **文献标识码**：A **doi**：10.3969/j.issn.1001-4632.2025.06.19

悬挂式单轨列车凭借其独特的结构特点，在中小城市、山地城市及旅游观光线路等特定场景中展现出显著优势，是改善城市交通拥堵的有效途径之一。然而，其运行环境对轨道异物侵入等异常事件的实时、高效检测提出了极高要求[1-2]。尽管计算机视觉通过车载相机图像为轨道异物检测提供智能化的检测方案[3]，但轨道异物在种类、形态与尺寸上极具多样性，加之未知类别异物的出现，对传统基于监督学习的目标检测方法构成严峻挑战。因此，无须海量人工标注的无监督目标检测，在解决此类问题上展现出重要的研究价值与应用潜力。

近年来，铁路异物检测的研究主要集中于有监督深度学习方法。YOLO系列变体[4-9]等主流模型已被广泛应用于障碍物检测、轨道组件识别和异物侵入检测等任务。为应对暴雨、大雪和雾霾等恶劣天气及夜间低照度、白天逆光等复杂光照条件导致的图像质量下降，研究者们提出改进YOLOv5以缓解雨雪遮挡[10]、结合Radon变换实现全天候检测[11]、改进边缘检测算法以适应低质量图像[12]、改进MaskR-CNN增强弱光环境下的定位能力[13]等诸多措施。

为克服纯视觉方法在复杂环境下的局限性，激光雷达（LiDAR）与视觉信息的多模态感知技术因其全天候特性和信息互补性而备受关注。LiDAR凭借高精度三维测距能力和三维点云信息，几乎不受光照和多数恶劣天气影响，能显著提升检测的鲁棒性。这类方法通过引入动态体素化[2]、多尺度体素特征聚合[14]等技术，提升了稀疏点云下小目标的检测性能。同时，通过提取列车轨行区域边界作为障碍物检测预警先验约束[15-16]，或设

收稿日期：2025-04-14；修订日期：2025-11-10 *本刊“中国科技期刊提能拓展计划项目”特约稿
基金项目：国家重点研发计划项目（2023YFB4302104-1）
第一作者：秦 勇（1971—），男，江苏徐州人，教授，博士研究生导师。E-mail：yqin@bjtu.edu.cn
通讯作者：曹志威（1993—），男，山东菏泽人，博士。E-mail：zhiwei@bjtu.edu.cn

计多模态融合网络[17−19]，显著提升了检测鲁棒性并减少误报、降低虚警率。

尽管如此，这些方法仍需依赖标注大规模高质量的数据，而轨道异物的多样性与未知性使其难以被穷举和充分标注，导致模型泛化能力受限，且多数悬挂式单轨列车仅配备摄像头，限制了多模态方案的广泛适用性。因此，探索仅依赖车载图像的无监督检测方法具有更强的现实意义。

在轨道交通领域，无监督检测技术主要应用于轨道裂纹、钢轨缺陷等结构性病害的自动化诊断，这些方法通常利用生成对抗网络[20]、特征嵌入[21]等技术建模正常图像的特征分布，计算偏差来检测异常。但该思路难以直接应用于异物检测，因为异物本身具有高度不可预测性，无法用固定的正常模式进行有效建模与识别。

在无监督目标检测通用领域中，基于图像重建的生成式方法虽取得了显著进展[20]，但这类方法依赖于对无目标图像的精确重建，在实际应用中，复杂场景下的图像重建往往难以达到理想效果，可能导致图像误判或产生虚假的检测结果。这种对重建质量的过度依赖限制了其在复杂场景下的鲁棒性。为了克服图像重建方法的局限，基于特征嵌入的方法[21−24]受到大量的关注，这类方法不再直接比较输入图像与重建图像的像素差异，而是将图像映射到高维特征空间，通过分析特征的分布判断图像中是否包含目标。这类方法将目标检测问题转化为特征空间的密度估计或距离度量问题，可在一定程度上避免对图像重建的依赖，提高了对复杂场景的适应能力。归一化流模型[25−28]作为1种重要的特征嵌入方法，通过一系列可逆变换将数据映射到高斯分布，从而实现概率密度的精确估计。Rudolph等[25]首次将归一化流引入目标检测领域，但其采用一维流结构会导致空间信息的损失。为了缓解这一问题，Gudovskiy等[26]和Rudolph等[27]尝试通过多尺度特征进行改进。Yu等[28]进一步探索了二维卷积归一化流，提出了Fastflow方法。然而，上述方法在处理复杂背景、小目标异物以及需要全局上下文信息的铁路场景时，仍然存在一定的局限性。

鉴于现有无监督目标检测方法在应对复杂侵限、小目标异物及全局上下文依赖等挑战时仍有局限，本研究提出1种面向悬挂式单轨系统无监督异物检测的双路径多尺度特征融合网络（Dual Path Multi-scale Fusion-Flow，DPMF-Flow）。该网络结合跨尺度特征提取模块、双路径融合流模块和目标分割定位模块，通过强化特征提取和融合、优化概率密度估计以及提升定位精度，实现高效、准确的异物检测。

1 无监督异物检测方法

1.1 整体结构

DPMF-Flow是1种基于双路径多尺度特征融合流的无监督异物检测方法，旨在解决悬挂式单轨复杂场景下未知类别异物的实时、准确检测难题，该方法的核心模型整体框架如图1所示。图中：y_i（$i=1,2,3$）为多尺度融合特征；α为自适应权重值；n为归一化流模块数量。该模型遵循端到端模式，主要由跨尺度特征提取模块、双路径融合流模块和目标分割定位模块3个核心部分级联而成。将工业相机采集的列车前向运行环境原始图像作为模型的输入。

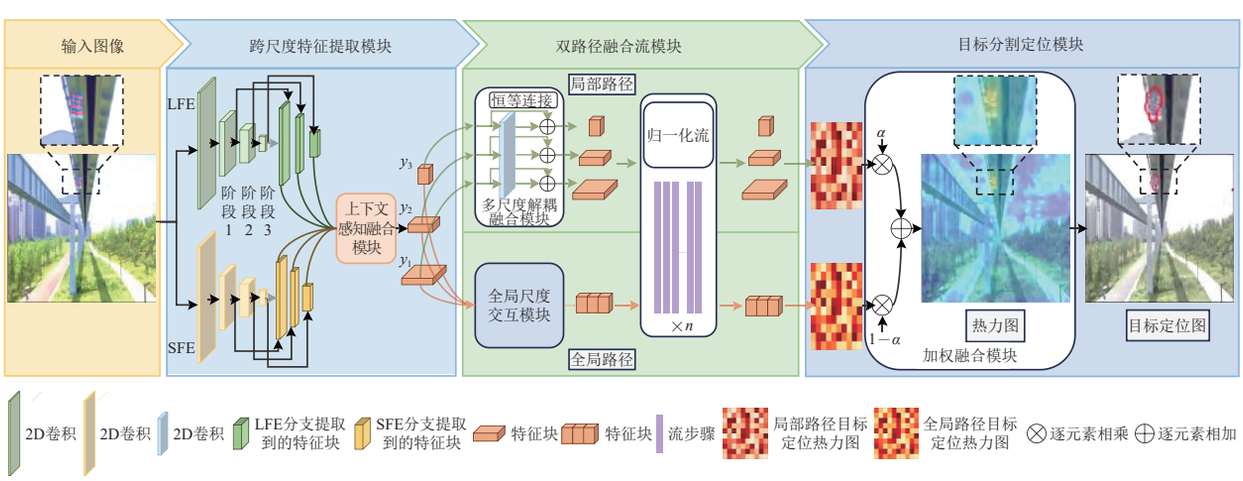


图1 DPMF-Flow模型整体框架

各模块具体功能为：①跨尺度特征提取模块负责提取输入图像的多尺度特征，采用局部特征提取（Local Feature Extraction，LFE）分支和语义特征提取（Semantic Feature Extraction，SFE）分支并行结构捕获异物边缘局部细节和轨道环境全局语义信息，2个分支的输出通过上下文感知融合模块进行自适应动态加权融合，生成一组信息丰富的多尺度特征，将其作为后续双路径融合流模块的输入；②双路径融合流模块负责对y_i进行概率密度估计，从而区分是否含有异物目标，该模块通过并行的局部和全局路径对特征进行处理，局部路径专注于保持各尺度特征的独立性，以精确定位小目标，全局路径则侧重于建立跨尺度的依赖关系，以理解全局上下文信息，通过一系列可逆的归一化流变换，最终将特征映射为2组分别对应局部和全局路径的互补隐变量并将其作为后续目标分割定位模块的输入；③目标分割定位模块负责将抽象的概率信息转化为直观的像素级定位结果，通过计算负对数似然、整合双路径信息并进行空间邻域最大池化等操作，最终生成高精度像素级目标定位热图，该热图清晰标示出异物的大小和位置，作为模型最终的输出结果传输给列车控制系统，实现对未知异物的有效预警和检测。

1.2　跨尺度特征提取模块

为应对悬挂式单轨交通场景的复杂性，捕捉图像中的局部细节信息和全局语义背景，并增强特征表示能力，引入跨尺度特征提取模块。该模块包括LFE分支、SFE分支和上下文感知融合模块。

LFE分支专注于捕获细粒度的局部图像细节，通过级联残差连接模块和下采样层，提取堆叠3个层的多尺度局部纹理特征y_i^{LFE}，其中$i=1,2,3$。SFE分支能够有效捕获图像的全局上下文信息和长距离依赖关系，通过堆叠Swin Transformer核心层在不同层间进行特征下采样，提取前3个层的多尺度特征y_i^{SFE}，其中$i=1,2,3$。SFE分支中Swin Transformer核心层通过窗口多头自注意力模块（W-MSA）和移位窗口多头自注意力模块（SW-MSA）进行计算[29]。W-MSA和MLP模块第l层的输出特征$\hat{s}_l$和s_l，以及SW-MSA和MLP模块第$l+1$层的输出特征$\dot{s}_{l+1}$和s_{l+1}为

$$\hat{s}_l=L_{\mathrm{N}}\left(A_{\mathrm{W-MSA}}(s_{l-1})\right)+s_{l-1} \tag{1}$$

$$s_l=L_{\mathrm{N}}\left(M_{\mathrm{LP}}(\hat{s}_l)\right)+\hat{s}_l \tag{2}$$

$$\dot{s}_{l+1}=L_{\mathrm{N}}\left(A_{\mathrm{SW-MSA}}(s_l)\right)+s_l \tag{3}$$

$$s_{l+1}=L_{\mathrm{N}}\left(M_{\mathrm{LP}}(\dot{s}_{l+1})\right)+\dot{s}_{l+1} \tag{4}$$

式中：$L_{\mathrm{N}}(\cdot)$为层归一化操作；$M_{\mathrm{LP}}(\cdot)$为多层感知机的前馈运算；$A_{\mathrm{W-MSA}}(\cdot)$为窗口多头自注意力操作；$A_{\mathrm{sW-MSA}}(\cdot)$为移位窗口多头自注意力操作；$s_{l-1}$为MLP模块第$l-1$层的输出特征。

为了自适应地调节和优化2个分支的双路径特征，并最大程度地保留和利用各自的优势信息，采用上下文感知融合模块充分融合双路径各自提取的互补性特征，其模块架构如图2所示。图中：y_i^{LFE}和y_i^{SFE}为输入特征；h'和w分别为特征的高度和宽度；Sigmoid为激活函数。上下文感知融合模块的核心机制为：首先，通过1×1大小的2D卷积模块对来自LFE和SFE分支的特征y_i^{LFE}和y_i^{SFE}进行通道对齐；然后，利用1个轻量级的门控卷积学习动态权重，该权重用于调节局部细节与全局语义特征在每个空间位置上的置信度平衡；最后，通过逐元素相乘和相加，将2个分支的互补信息进行有效融合，输出多尺度融合特征y_i作为后续模块的输入，其计算式为

$$y_i=g\otimes\left(C_{2d_{1\times1}}(y_i^{\mathrm{LFE}})\oplus y_i^{\mathrm{SFE}}\right) \tag{5}$$

式中：g为门控权重值；$C_{2d_{1\times1}}(\cdot)$为1×1大小的2D卷积操作；⊗表示逐元素相乘；⊕表示逐元素相加。

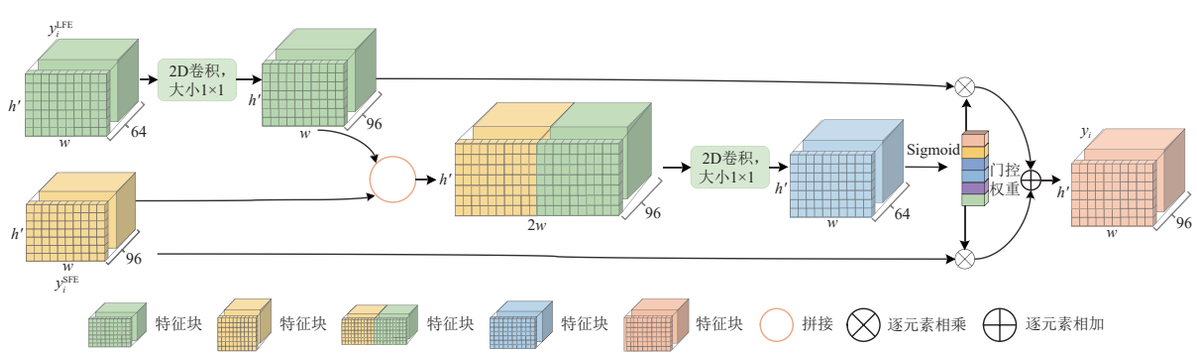


图2　上下文感知融合模块架构

1.3 双路径融合流模块

针对归一化流在处理跨尺度耦合时计算代价高昂且可能引入无关依赖关系以及将多尺度特征视为独立分量时导致尺度间信息利用不充分的情况，设计双路径融合流模块。该模块采用并行双路径策略处理融合后的多尺度特征：多尺度解耦融合模块聚焦局部关键信息，而全局尺度交互模块则捕获跨尺度全局上下文，最终经由归一化流进行处理，有效平衡模型性能与计算效率。

为了确保每个尺度的特征在进入归一化流之前互不干扰，从而尽可能多地保留局部目标信息，设计多尺度解耦融合模块，该模块对每个融合后的多尺度特征 y_i 分别应用 1 个独立的 1×1 大小的 2D 卷积模块进行通道变换，通过恒等映射连接，将变换后的特征与 y_i 相加，得到预处理后的特征 y_i' 并将其输入独立的归一化流模型中，得到相应的隐变量 h_i^{local} 和雅可比矩阵 $\boldsymbol{J}_i^{\text{local}}$。

为进一步降低计算复杂度，减少不必要的依赖关系，并增强特征表达能力，提出全局尺度交互模块，其架构如图 3 所示。图中：$y_i^{(0)}$ 为经过通道变换和下采样处理后的特征；$y_{\text{concat}}^{(i)}$ 为经过模块处理后的特征；y'_{concat} 为输出特征；$\boldsymbol{W}_i(i=1,\ 2,\ 3)$ 为特征权重；$s(y)$ 和 $b(y)$ 分别为缩放因子和平移因子。该模块首先通过 1×1 大小的 2D 卷积模块将

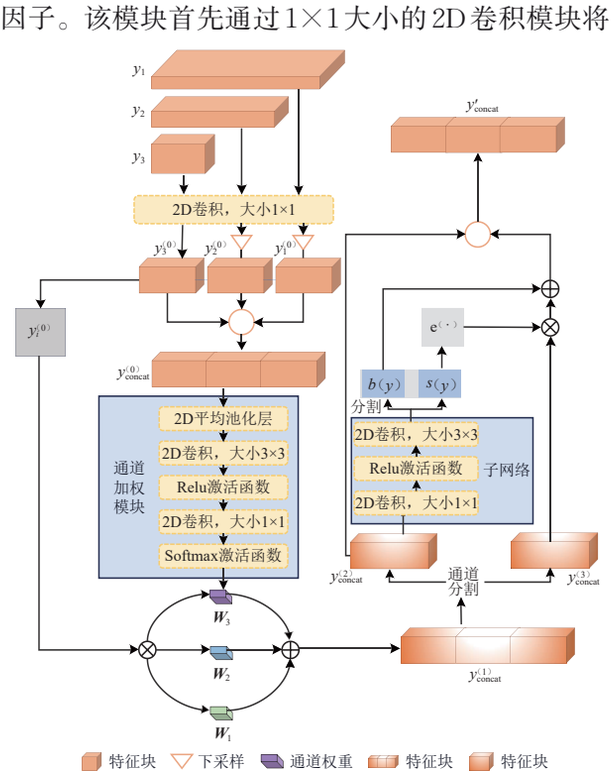


图 3 全局尺度交换模块架构

各尺度特征的通道数统一为 96，然后通过双线性插值将多尺度特征下采样至最小尺度，并将调整大小后的特征拼接成一个新特征 $y_{\text{concat}}^{(0)}$，接着通过 1 个由 2D 平均池化层、2D 卷积模块和激活函数组成的通道加权模块为每个尺度的特征图生成权重 $\boldsymbol{W}_i$，将这些权重分别应用于每个尺度的特征，并求和获得融合特征 $y_{\text{concat}}^{(1)}$ 为

$$y_{\text{concat}}^{(1)}=y_1^{(0)}\otimes\boldsymbol{W}_1\oplus y_2^{(0)}\otimes\boldsymbol{W}_2\oplus y_3^{(0)}\otimes\boldsymbol{W}_3 \tag{6}$$

特征 $y_{\text{concat}}^{(1)}$ 经过通道分割得到特征 $(y_{\text{concat}}^{(2)},\ y_{\text{concat}}^{(3)})$，通过由卷积模块构成的独立子网络预测缩放因子 $s(\cdot)$ 和平移因子 $b(\cdot)$，将特征 $y_{\text{concat}}^{(2)}$ 的一部分融入到特征 $y_{\text{concat}}^{(3)}$ 的另一部分进行仿射非线性变换，增强特征的表达能力，$s(\cdot)$ 和 $b(\cdot)$ 由子网络预测得到。经过非线性变换增强后的特征 y'_{concat} 为

$$y'_{\text{concat}}=C_{\text{concat}}\left(\mathrm{e}^{s\left(y_{\text{concat}}^{(2)}\right)}\otimes y_{\text{concat}}^{(3)}+b\left(y_{\text{concat}}^{(2)}\right),y_{\text{concat}}^{(2)}\right) \tag{7}$$

式中：C_{concat} 为拼接操作。

该网络采用 3×3 的 2D 卷积模块、ReLU 激活函数以及 3×3 的 2D 卷积模块的结构，最终将输出的特征 y'_{concat} 输入后续的归一化流层进行处理，以获得对应的全局隐变量 h^{global} 和雅可比矩阵 $\boldsymbol{J}^{\text{global}}$。

由于 2D 卷积神经网络在提取图像局部特征时保持平移不变性，双路径融合流模块内的完全双向可逆归一化流模块架构如图 4 所示，图中：y_0 为输入特征；y_a 和 y_b 为经过通道分割处理后的特征；$s_1(\cdot)$ 和 $b_1(\cdot)$ 为特征 y_a 经过处理后得到的缩放因子和平移因子；$s_2(\cdot)$ 和 $b_2(\cdot)$ 为特征 y_b 经过处理后得到的缩放因子和平移因子。y_0 经过通道随机重排和通道分割为特征 $(y_a,\ y_b)$，通过仿射耦合层使特征 y_a，y_b 的变换互相依赖，从而增强特征表达能力。其中，$s_1(\cdot)$，$s_2(\cdot)$，$b_1(\cdot)$ 和 $b_2(\cdot)$ 的子网络均由 LeakyRelu 激活函数和 2 个 3×3 大小的 2D 卷积模块构成。

1.4 目标分割定位模块

与传统生成像素级目标定位热图的方法不同，此处不直接使用归一化流输出的隐变量进行目标检测，而是设计 1 个目标分割定位模块，用于将全局和局部路径输出的隐变量转化为像素级目标定位热图。该模块首先计算隐变量 h_i^{local} 和 h^{global} 的负对数似然值（近似为沿通道维度的平方和），并通过指数函数将其转换为非归一化概率图的概率值 p_{rob}，其计算式为

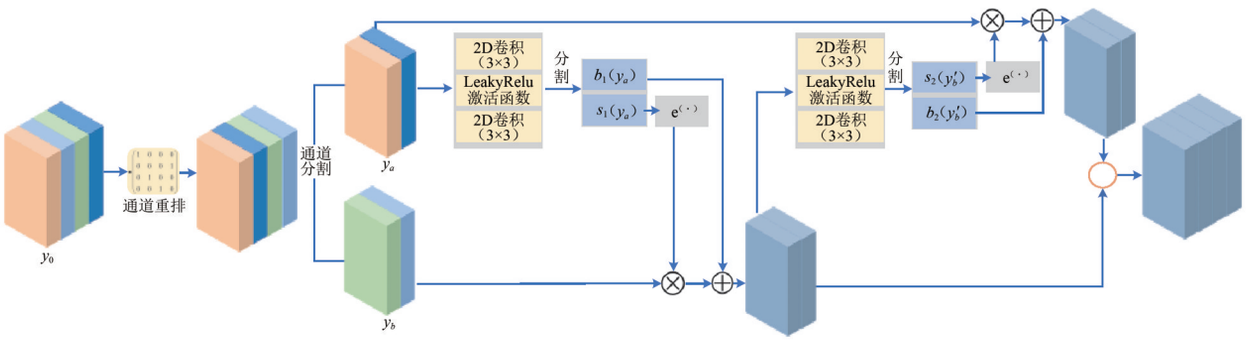


图 4　归一化流模块架构

$$p_{\text{rob}}=\begin{cases}\mathrm{e}^{\left(-\frac{1}{2}\sum_{k=1}^{C}\frac{(h_{i,k}^{\text{local}})^2}{C}\right)} & \text{若隐变量为局部隐变量}\\ \mathrm{e}^{\left(-\frac{1}{2}\sum_{k=1}^{C}\frac{(h_{k}^{\text{global}})^2}{C}\right)} & \text{若隐变量为全局隐变量}\end{cases}\tag{8}$$

式中：$h_{i,k}^{\text{local}}$ 为第 k 个通道的第 i 个局部隐变量；h_k^{global} 为第 k 个通道的全局隐变量；C 为通道总数。

该步骤的原理为：若样本不含异物目标，其隐变量 h_i 应接近高斯分布中心，因而具有较高的似然值，这导致其负对数似然值较低，经指数函数转换后，在非归一化概率图上呈现出较低的概率值；反之，含有目标的样本，其隐变量偏离高斯分布中心，将产生更高的概率值。因此，采用双线性插值将每个隐变量对应的概率图上采样至输入图像分辨率。为整合来自不同隐变量的目标信息并突出显示重要的目标区域，该模块将隐变量生成的概率图累加求平均，从而产生平均目标定位热图。此外，使用3×3大小的2D卷积模块进行空间邻域最大池化操作，以增强目标区域的显著性并抑制噪声干扰，分别得到全局路径目标定位热力图矩阵 $\boldsymbol{M}^{\text{global}}$ 和局部路径目标定位热力图矩阵 $\boldsymbol{M}_i^{\text{local}}$（$i=1,2,3$）。结合全局路径对大目标和上下文的感知能力与局部路径对小目标和细节的捕捉能力，通过自学习权重 α 来融合 $\boldsymbol{M}^{\text{global}}$ 和 $\boldsymbol{M}_i^{\text{local}}$，生成最终的高精度定位热图矩阵 $\boldsymbol{M}$ 为

$$\boldsymbol{M}=\alpha\otimes\boldsymbol{M}^{\text{global}}\oplus(1-\alpha)\otimes\boldsymbol{M}_i^{\text{local}}\tag{9}$$

式中：α 为自适应权重。

为实现对不含目标样本分布的精确建模，设计了联合局部与全局路径的损失函数 $\mathcal{L}(x)$。该损失函数通过最小化训练样本的负对数似然来优化模型。根据变量代换定理，输入数据 x 通过可逆函数 f 映射到隐变量，即 $h=f(x)$，隐变量 h 服从标准正态分布时，其对数似然 $\ln p(h)$ 与 h 的 L_2 范数平方成正比。因此，最小化负对数似然 $-\ln p(x)$ 等价于最小化 $-\ln p(h)-\ln|\boldsymbol{d}_{\text{et}}(\boldsymbol{J})|$。其中：$\boldsymbol{J}$ 为雅可比矩阵；$\boldsymbol{d}_{\text{et}}(\cdot)$ 为计算行列式。该损失函数整合了全局路径和局部路径的似然结果，强制模型同时关注局部细节和全局上下文信息来学习正常样本的分布。这种互补的约束使得模型对不含目标的建模更加全面，进而增强了鲁棒性，使其能更准确地识别出真正偏离正常模式的异物。该损失函数 $\mathcal{L}(x)$ 的表达式为

$$\begin{aligned}\mathcal{L}(x)=&-(\ln p_X(x)^{\text{local}}+\ln p_X(x)^{\text{global}})=\\&\frac{\left\|h_i^{\text{local}}+h^{\text{global}}\right\|_2^2}{2}-\ln|\boldsymbol{d}_{\text{et}}(\boldsymbol{J}_i^{\text{local}})|\times\\&\ln|\boldsymbol{d}_{\text{et}}(\boldsymbol{J}^{\text{global}})|\end{aligned}\tag{10}$$

式中：$p_X(\cdot)$ 为似然值，$\ln p_X(x)^{\text{global}}$ 和 $\ln p_X(x)^{\text{local}}$ 分别为全局和局部路径的隐空间特征的对数似然；h^{global} 和 h_i^{local} 为全局路径和局部路径获得的隐变量；$\boldsymbol{J}^{\text{global}}$ 和 $\boldsymbol{J}_i^{\text{local}}$ 为全局和局部路径获得的雅可比矩阵。

2　数据处理

2.1　数据采集

利用在湖北省武汉市悬挂式单轨上部署的单目工业相机设备，实地采集列车前向运行环境数据。通过对所采集数据进行标注，构建用于异物入侵检测的数据集。为了保证数据采集的稳定性，相机安装于固定设备上，以减少来自轨道和列车运行振动造成的影响，设备外观如图 5 所示。该数据集由 3 600 张图像数据组成，所有图像的原始分辨率大小为 1 920 像素×1 080 像素。其中标注的侵入物体为尺寸、材质、形状、纹理、运动特性等方面有显著差异并且是悬挂式单轨场景中的代表性样本，包括无人机、气球、塑料布等。为提升模型在复杂环境下的鲁棒性，对原始数据进行了模拟恶劣天气、不同光照条件以及轻微图像扰动等多种数据增强操



（a）固定设备



（b）工业相机

图 5 采集设备外观

作，通过上述数据增强方法，数据集扩充为约 6 000 张不含异物的正常样本图像和约 2 400 张包含各类侵入异物的异物样本图像。

MVTec AD [30] 数据集是 1 个广泛应用于无监督目标检测与定位研究的工业级基准数据集，旨在模拟真实世界的产品质量检测场景。该数据集包含 15 个不同的类别，涵盖 10 种物体和 5 种纹理，共计 5 354 张高分辨率图像（分辨率为 700 像素×700 像素至 1 024 像素×1 024 像素），其中训练集包含 3 629 张无缺陷图像，测试集包含 1 725 张混合了正常与多种缺陷类型（如裂纹、划痕等）的 70 种图像，并提供像素级标注。

2.2 试验环境

试验采用 Ubuntu 20.04 操作系统进行测试，利用 Anomalib 深度学习库进行算法开发，CPU 型号为 i9-10980XE，GPU 型号为 NVIDIA RTX 3090Ti，RAM 为 24 GB。输入图像分辨率均为 256 像素×256 像素，训练集和测试集的划分比例为 1∶4，统一设置归一化流为 $n=2$，学习率为 0.001。在数据集上进行试验时，上述参数均保持不变。

2.3 评估指标

无监督异物检测评估包含 2 个层面的任务：图像级分类和像素级定位。前者判断图像是否包含异物，并采用 AUROC（ROC 曲线下面积）进行评估；后者识别异物的具体像素大小和位置，通过检测异物样本的图像级和像素级指标 AUROC、检测异物样本的像素级定位指标 AUPRO（PRO 曲线下面积）和 F_1 分数指标进行评估。

AUROC 用于评估算法区分“含异物”与“不含异物”样本的性能。其计算式为

$$A_{\mathrm{AUROC}}=\int_0^1 (T_{PR})\mathrm{d}(F_{PR}) \tag{11}$$

其中，

$$F_{PR}=\frac{F_{\mathrm{P}}}{F_{\mathrm{P}}+T_{\mathrm{N}}}$$

$$T_{PR}=R=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{N}}}$$

式中：F_{PR} 为误检的异物样本比率；T_{PR} 为正确检测的异物样本比率；T_{P} 为正确检测的异物样本数；F_{P} 为漏检的异物样本数；T_{N} 为正确检测的不含异物样本数；R 为检测异物样本的召回率。

异物样本的检测精度 P 的计算式为

$$P=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{P}}} \tag{12}$$

F_1 为精度 P 和召回率 R 的调和平均数，用于综合评估模型在检测到异物样本时的准确性和完整性。其计算式为

$$F_1=2\frac{PR}{P+R} \tag{13}$$

AUPRO 用于评估图像中每个区域的预测结果与实际结果的重叠程度，通过计算 Precision-Recall 曲线下的面积来反映模型在不同召回率下的平均精确率，其计算式为

$$A_{\mathrm{AUPRO}}=\int_0^1 P(R)\mathrm{d}R \tag{14}$$

式中：$P(R)$ 为以召回率 R 为自变量的精度函数。

3 试验结果与分析

3.1 悬挂式单轨异物入侵数据集上的试验结果

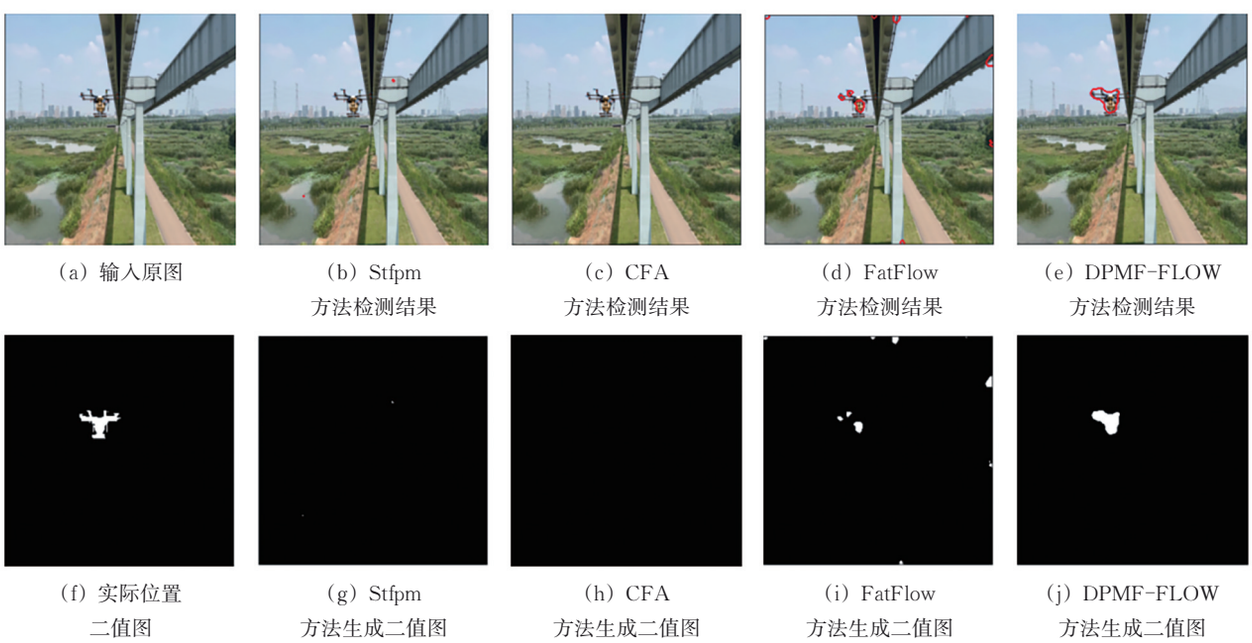
为验证 DPMF-Flow 在悬挂式单轨异物检测任务的有效性，选取近几年的先进无监督目标检测算法与 DPMF-Flow 进行对比，结果见表 1。

表 1 悬挂式单轨异物入侵数据集对比结果

方法名称	图像级指标/%	像素级指标/%			FPS
	AUROC	AUPRO	AUROC	F_1	
Padim[24]	71.51	67.67	96.09	23.45	46.88
Stfpm[22]	75.37	77.89	97.77	26.12	56.03
CFA[31]	69.92	59.59	94.95	26.64	61.74
DSR[32]	46.48	18.98	51.02	0.28	32.73
Csflow[27]	74.75	41.17	85.18	2.27	55.63
Cflow[26]	62.73	61.83	94.02	1.60	54.40
Fastflow[28]	68.03	67.27	94.63	22.09	61.75
DPMF-Flow	82.64	91.65	99.34	41.44	66.04

由表 1 可知：DPMF-Flow 网络在悬挂式单轨异物入侵数据集上的有监督目标检测任务中表现出优越的性能，与其他所有方法相比，在图像级别，DPMF-Flow 的 AUROC 达到 82.64%，是所比较方法中最高的；在像素级别，DPMF-Flow 的 AUROC 达到 99.34%，AUPRO 达到 91.65%，F1Score 达到 41.44% 也明显优于其他方法。证明 DPMF-Flow 在具有挑战性的悬挂式单轨环境中检测和定位目标侵入的有效性。同时 DPMF-Flow 的 FPS 达到了 66.04，满足许多实际应用中的实时处理要求。

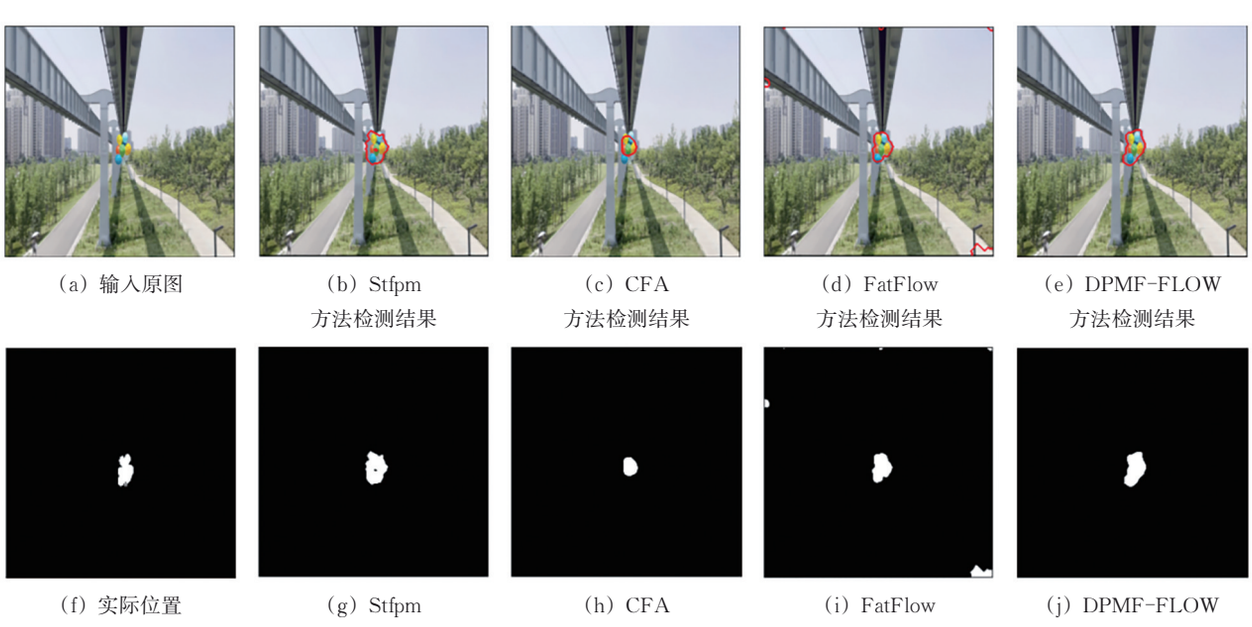
为了更直观地展示 DPMF-Flow 的效果，将上述方法和 DPMF-Flow 方法在悬挂式单轨异物入侵数据集的检测结果进行可视化对比，可视化结果如图 6—图 8 所示，将判别到的异物用红色圈出。由图 6—图 8 可知：DPMF-Flow 方法能够实现对轨道异物的准确定位，且误报率较低，相比之下，其余方法或存在误报率偏高的问题，或在目标边界轮廓的提取上不够精确，甚至出现部分目标漏检的情况。



（a）输入原图　（b）Stfpm 方法检测结果　（c）CFA 方法检测结果　（d）FatFlow 方法检测结果　（e）DPMF-FLOW 方法检测结果

（f）实际位置二值图　（g）Stfpm 方法生成二值图　（h）CFA 方法生成二值图　（i）FatFlow 方法生成二值图　（j）DPMF-FLOW 方法生成二值图

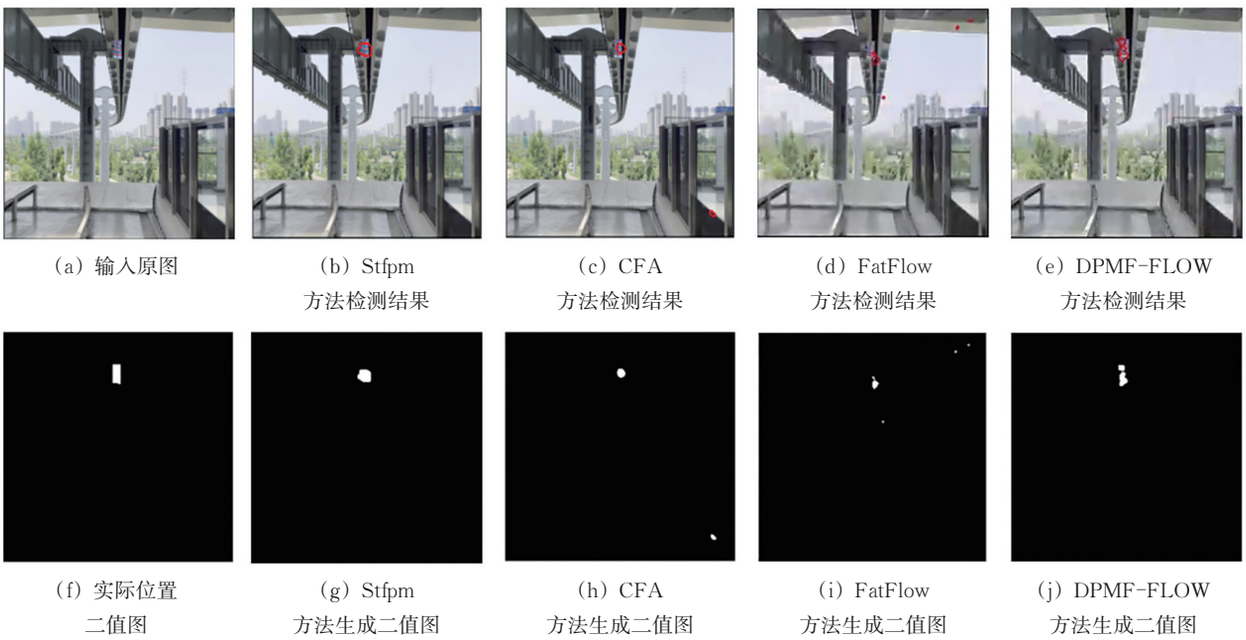
图 6　悬挂式单轨异物入侵数据集中无人机检测结果可视化对比



（a）输入原图　（b）Stfpm 方法检测结果　（c）CFA 方法检测结果　（d）FatFlow 方法检测结果　（e）DPMF-FLOW 方法检测结果

（f）实际位置二值图　（g）Stfpm 方法生成二值图　（h）CFA 方法生成二值图　（i）FatFlow 方法生成二值图　（j）DPMF-FLOW 方法生成二值图

图 7　悬挂式单轨异物入侵数据集中气球检测结果可视化对比



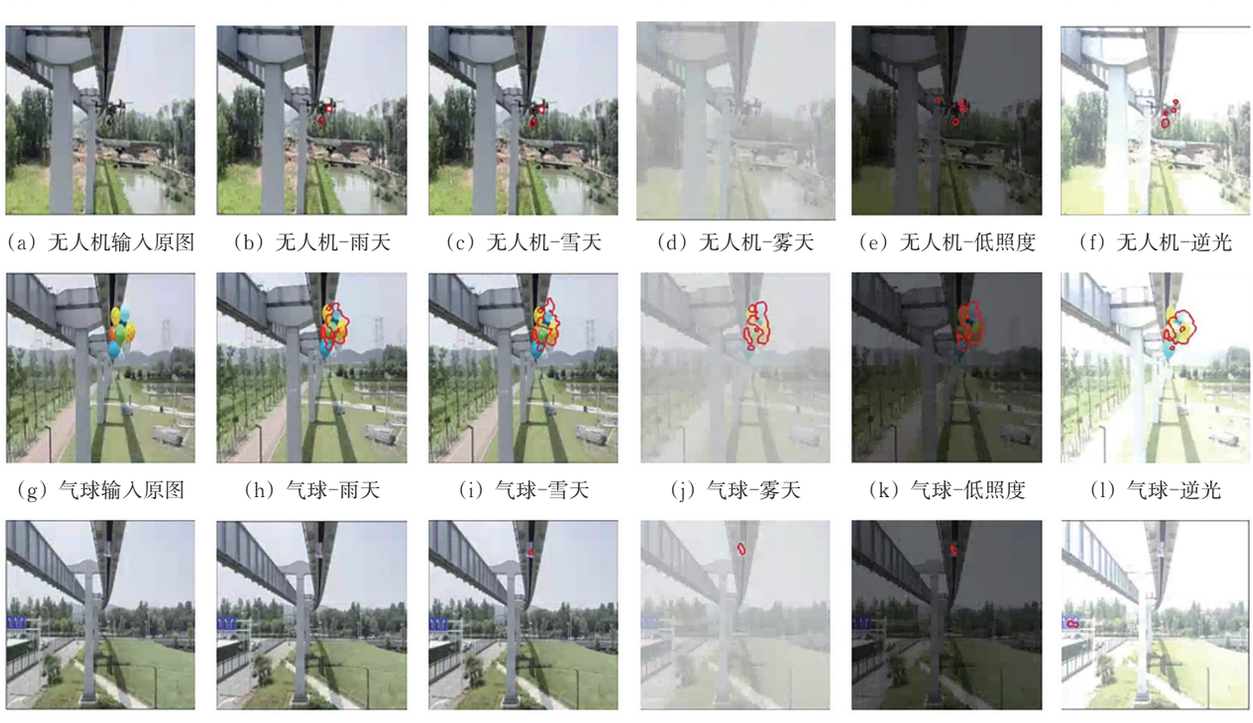
（a）输入原图 （b）Stfpm 方法检测结果 （c）CFA 方法检测结果 （d）FatFlow 方法检测结果 （e）DPMF-FLOW 方法检测结果

（f）实际位置二值图 （g）Stfpm 方法生成二值图 （h）CFA 方法生成二值图 （i）FatFlow 方法生成二值图 （j）DPMF-FLOW 方法生成二值图

图8 悬挂式单轨异物入侵数据集中塑料布检测结果可视化对比

为了进一步探究算法局限性，分析了典型的失败案例，DPMF-Flow算法在恶劣天气和复杂光照下的局限性可视化对比如图9所示。由图9可知：在恶劣天气或者复杂光照条件下，由于异物尺寸小且轨道与异物区的颜色接近、分度低，可能漏检部分异物像素或异物区域，甚至会产生误报。这些案例表明，该算法虽然在多数情况下有效，但在处理边界模糊和低对比度目标时仍存在不足。未来的改进方向包括引入轨道区域分割作为空间先验约束，以更精确地聚焦于轨行区内的异常，从而进一步降低虚警率和漏检率。

为了更深入地评估DPMF-Flow方法在处理不同尺寸异物时的性能，对检测结果进行细化分析。在实际运营中，对不同大小和距离异物的检测能力



（a）无人机输入原图 （b）无人机-雨天 （c）无人机-雪天 （d）无人机-雾天 （e）无人机-低照度 （f）无人机-逆光

（g）气球输入原图 （h）气球-雨天 （i）气球-雪天 （j）气球-雾天 （k）气球-低照度 （l）气球-逆光

（m）塑料布输入原图 （n）塑料布-雨天 （o）塑料布-雪天 （p）塑料布-雾天 （q）塑料布-低照度 （r）塑料布-逆光

图9 DPMF-Flow方法在恶劣天气和复杂光照下的局限性可视化对比

是衡量方法实用性的关键指标。虽然采用单目视觉方案难以直接估算物理距离，但异物在图像中所占的像素尺寸仍可间接反映其相对大小或距离。因此，参照业界通用的COCO数据集和相关团体标准对目标尺寸进行划分，将自建数据集中的异物样本按照其像素面积分为大、中、小3种尺寸，本试验所构建的数据集仅涵盖中、小2种尺寸的异物。

分别统计DPMF-Flow方法对中、小尺寸目标的检测精度和召回率，并与其他代表性方法进行对比，结果见表2。

表2　不同方法在不同尺寸异物检测性能分析

方法名称	小尺寸目标(＜32像素×32像素)		中尺寸目标(≥32像素×32像素,≤96像素×96像素)	
	检测精度	召回率	检测精度	召回率
Padim[24]	0.754	0.310	0.921	0.897
Stfpm[22]	0.727	0.427	0.800	0.941
CFA[31]	0.750	0.126	0.846	0.579
Fastflow[28]	0.500	0.178	0.730	0.844
DPMF-Flow	0.879	0.473	1.000	1.000

从表2可以看出，所有方法对中尺寸目标的检测性能均较好，DPMF-Flow方法在处理中、小尺寸目标时表现出最优的性能，尤其是处理最具挑战性的小目标检测任务中的优势更为突出，这充分证明该方法在捕捉细微异物方面的优势。

3.2　MVTec AD数据集上的试验结果

为验证DPMF-Flow的泛化性，将该方法和其他方法在MVTec AD数据集上进行对比，试验结果见表3。由表3可知：在MVTec AD数据集上，DPMF-Flow的性能显著领先于对比方法，其平均AUROC和平均AUPRO分别达到98.6%和96.7%，与表现次优的FastFlow方法[28]相比，这2项指标分别提升了0.2%和2.3%；该优势在网格、瓷砖及木材等类别上尤为突出，以瓷砖类别为例，DPMF-Flow的平均AUROC和平均AUPRO比FastFlow提升了1.2%和5.8%。这充分证明DPMF-Flow方法在无监督目标检测任务中的优越性，以及其在不同类型数据上的良好泛化能力。

表3　MVTec AD数据集上的像素级指标对比结果（AUROC; AUPRO）　%

方法名称	地毯	网格	皮革	瓷砖	木材	胶囊	榛子	平均值
Padim[24]	99.1;96.2	97.3;94.6	99.2;97.8	94.1;86.0	94.9;91.1	98.5;93.5	98.2;92.6	97.3;93.1
US[33]	93.5;87.9	89.9;95.2	97.8;94.5	92.5;94.6	92.1;91.1	96.8;96.8	98.2;96.5	94.4;93.8
SPADE[23]	97.5;94.7	93.7;86.7	97.6;97.2	87.4;75.9	88.5;87.4	99.0;93.7	99.1;95.4	94.7;90.1
Csflow[27]	95.2;83.6	94.5;78.7	89.3;81.1	90.7;73.0	86.3;60.5	98.1;89.3	96.1;82.8	92.9;78.4
FastFlow[28]	99.4;97.3	98.3;94.8	99.5;99.2	96.3;86.9	97.0;94.1	99.1;91.2	99.1;91.2	99.1;91.2
DPMF-Flow	99.3;97.7	98.9;96.5	99.7;99.5	97.5;92.7	96.7;96.4	99.1;96.5	98.7;97.5	98.6;96.7

3.3　消融试验

为验证DPMF-Flow中提出的每个模块的有效性，进行消融试验，结果见表4。表中："√"表示在基准模型的基础上采用该模块进行改进。由表4可知：3个模块均未使用的情况下，模型的异物检测能力较差；引入跨尺度特征提取模块显著提升了模型性能，图像级指标AUROC提高了12.41%，像素级指标AUPRO大幅提高了22.26%，证明了双路径特征提取器和多尺度特征融合策略对于增强特征表示能力的至关重要性；加入双路径融合流模块后产生了像素级检测性能的提升，像素级指标AUPRO和F_1分别提高了6.21%和8.64%，验证了双路径流在精细化目标度量方面的有效性；引入目标分割定位模块后像素级指标AUROC和AUPRO分别提高了2.39%和0.71%，表明该模块在生成目标定位热图和提高定位精度方面发挥了积极作用；当3个模块同时被使用时，DPMF-Flow方法取得了最佳性能，像素级指标AUPRO达到了最高值91.65%，这表明模块之间存在协同效应，共同促成了DPMF-Flow方法最终的优越性能。

表4 不同模块对评价指标的消融试验 %

基础模型名称	跨尺度特征提取模块	双路径融合流模块	目标分割定位模块	图像级指标	像素级指标		
				AUROC	AUPRO	AUROC	F_1
FastFlow				68.03	67.27	94.63	22.09
	√			80.44	89.53	99.22	39.52
		√		71.64	73.48	95.32	30.73
			√	69.03	67.98	97.02	23.60
	√	√		82.52	91.25	99.33	40.50
	√		√	81.71	90.81	99.32	41.08
		√	√	79.26	82.34	97.60	31.27
	√	√	√	82.64	91.65	99.34	41.44

4 结 论

（1）本研究提出了1种高效的无监督前向异物检测方法（DPMF-Flow），设计了3个模块以应对轨道交通中的异物检测面临的挑战：双分支跨尺度特征提取模块通过并行处理，同时兼顾局部纹理细节与全局场景上下文，解决了复杂背景下小目标异物难以识别的问题；双路径融合流模块对2条特征路径进行独立的概率密度建模，生成互补的隐变量，克服了单一特征尺度信息不足的局限，显著提升了异物定位的准确性；目标分割定位模块则高效融合了多路径信息，生成了高质量的像素级定位结果。

（2）通过试验验证了DPMF-Flow方法的卓越性能，其检测精度可达99.34%，并以66.04 FPS的帧率进行实时检测。在自建的悬挂式单轨异物入侵数据集中，该方法在各项评价指标上均显著优于其他主流无监督方法。同时，在公开基准MVTec AD上的领先表现，也彰显了其泛化性和先进性。

参 考 文 献

[1] QIN Y, CAO Z W, SUN Y F, et al. Research on Active Safety Methodologies for Intelligent Railway Systems [J]. Engineering, 2023, 27: 266-279.

[2] 曹志威,戈轩宇,秦勇,等.基于激光雷达的列车前向障碍物检测方法[J].中国铁道科学,2024,45(6):183-193.
(CAO Zhiwei, GE Xuanyu, QIN Yong, et al. Forward Obstacle Detection Algorithm for Train Based on LiDAR [J]. China Railway Science, 2024, 45 (6): 183-193. in Chinese)

[3] CAO Z W, QIN Y, JIA L M, et al. Railway Intrusion Detection Based on Machine Vision: a Survey, Challenges, and Perspectives [J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25 (7): 6427-6448.

[4] 王楠,侯涛,牛宏侠.多尺度特征融合的铁轨异物入侵检测研究[J].西安交通大学学报,2024,58(9):139-153.
(WANG Nan, HOU Tao, NIU Hongxia. Research on Railway Track Foreign Object Intrusion Detection Based on Multi-Scale Feature Fusion [J]. Journal of Xi'an Jiaotong University, 2024, 58 (9): 139-153. in Chinese)

[5] 付浩辰,白堂博,许贵阳,等.基于改进YOLOv8的轨道板裂缝检测算法[J].北京交通大学学报,2024,48(6):133-143.
(FU Haochen, BAI Tangbo, XU Guiyang, et al. Crack Detection in Track Slabs Based on an Improved YOLOv8 Algorithm [J]. Journal of Beijing Jiaotong University, 2024, 48 (6): 133-143. in Chinese)

[6] 姜香菊,冯海照,李涛.基于FasterNet和YOLOv8s改进的铁路异物入侵快速检测方法[J].北京交通大学学报,2024,48(5):39-48.
(JIANG Xiangju, FENG Haizhao, LI Tao. Improved Rapid Detection Method for Foreign Object Intrusion on Railroads Based on FasterNet and YOLOv8s [J]. Journal of Beijing Jiaotong University, 2024, 48 (5): 39-48. in Chinese)

[7] 孟彩霞,王兆楠,石磊,等.改进YOLOv5s的铁路异物入侵检测算法[J].小型微型计算机系统,2024,45(4):879-886.
(MENG Caixia, WANG Zhaonan, SHI Lei, et al. Improved Railroad Foreign Object Intrusion Detection Algorithm for YOLOv5s [J]. Journal of Chinese Computer Systems, 2024, 45 (4): 879-886. in Chinese)

[8] 刘力,苟军年.基于YOLO v4的铁道侵限障碍物检测方法研究[J].铁道科学与工程学报,2022,19(2):528-536.
(LIU Li, GOU Junnian. Research on Detection Method of Railway Intrusion Obstacles Based on YOLO v4 [J]. Journal of Railway Science and Engineering, 2022, 19 (2): 528-536. in Chinese)

[9] 白堂博,段嘉明,杨建伟,等.基于边缘AI计算的轨道扣件状态定量化检测方法[J].中国铁道科学,2025,46(1):60-69.
(BAI Tangbo, DUAN Jiaming, YANG Jianwei, et al. Quantitative Detection of Rail Fasteners Based on Edge AI

Computing [J]. China Railway Science, 2025, 46 (1): 60-69. in Chinese)

[10] 马晓君,王栋,刘德胜,等.基于改进YOLOv5的降雪天气高速列车障碍物检测[J].计算机仿真,2025,42(1):155-161,451.
(MA Xiaojun, WANG Dong, LIU Desheng, et al. Obstacle Detection of High-Speed Train in Snowfall Weather Based on Improved YOLOv5 [J]. Computer Simulation, 2025, 42 (1): 155-161, 451. in Chinese)

[11] 王前选,梁习锋,刘应龙,等.缓变异物入侵铁路线路视觉检测方法[J].中国铁道科学,2014,35(3):137-143.
(WANG Qianxuan, LIANG Xifeng, LIU Yinglong, et al. Visual Detection Method for the Invasion of Slowly Changing Foreign Matters to Railway Lines [J]. China Railway Science, 2014, 35 (3): 137-143. in Chinese)

[12] PERIĆ S, MILOJKOVIĆ M, STAN S D, et al. Dealing with Low Quality Images in Railway Obstacle Detection System [J]. Applied Sciences, 2022, 12 (6): 3041.

[13] 陈永,卢晨涛,王镇.红外弱光环境下多尺度密集注意力铁路异物检测[J].铁道学报,2022,44(7):63-71.
(CHEN Yong, LU Chentao, WANG Zhen. Detection of Railway Foreign Object Intrusion under Infrared Low Light Environment Based on Multi-Scale Dense Attention Network [J]. Journal of the China Railway Society, 2022, 44 (7): 63-71. in Chinese)

[14] LIAN L R, CAO Z W, QIN Y, et al. RAE3D: Multiscale Aggregation-Enhanced 3D Object Detection for Rail Transit Obstacle Perception [J]. IEEE Transactions on Industrial Informatics, 2025, 21 (5): 4221-4232.

[15] SHEN T, ZHOU J H, YUAN T F, et al. LiDAR-Based Urban Three-Dimensional Rail Area Extraction for Improved Train Collision Warnings [J]. Sensors, 2024, 24 (15): 4963.

[16] 潘文波,李程,袁希文,等.面向地铁场景的激光雷达目标检测研究[J].机车电传动,2022(2):89-96.
(PAN Wenbo, LI Cheng, YUAN Xiwen, et al. Research on Lidar Target Detection for Metro Scene [J]. Electric Drive for Locomotives, 2022 (2): 89-96. in Chinese)

[17] LIU W T, WANG Y F, YU G Z, et al. RailFusion: a Lidar-Camera Data Interaction Network for 3-D Railway Object Detection [J]. IEEE Transactions on Intelligent Transportation Systems, 2025, 26 (8): 12761-12773.

[18] FAN Z W, CHENG G Z, SUN T, et al. A Method for Railway Intrusion Objects Detection with LiDAR and Image Fusion [C]// Proceedings of the 2024 8th International Conference on Computer Science and Artificial Intelligence. Beijing: ACM, 2024: 39-44.

[19] SHEN T, XIE Y X, YUAN T F, et al. A Detection Methods with Image Recognition for Specific Obstacles in the Urban Rail Area [J]. IEEE Access, 2024, 12: 142772-142783.

[20] 何庆,刘震,王启航,等.基于改进Skip-GAN的高速铁路巡检图像异常检测[J].铁道学报,2024,46(9):121-128.
(HE Qing, LIU Zhen, WANG Qihang, et al. Anomaly Detection of High-Speed Railway Inspection Images Based on Improved Skip-GAN [J]. Journal of the China Railway Society, 2024, 46 (9): 121-128. in Chinese)

[21] GAO Y, CAO Z W, QIN Y, et al. Segmented-Aggregated Framework with Internal-External Constraint Mechanism for Unsupervised Track Anomaly Detection [J]. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 2523612.

[22] WANG G, HAN S, DING E, et al. Student-Teacher Feature Pyramid Matching for Anomaly Detection [C]// Proceedings of the British Machine Vision Conference. UK: BMVA, 2021.

[23] COHEN N, HOSHEN Y. Sub-Image Anomaly Detection with Deep Pyramid Correspondences [J]. arXiv: 2005.02357 [cs.CV], 2020.

[24] DEFARD T, SETKOV A, LOESCH A, et al. PaDiM: a Patch Distribution Modeling Framework for Anomaly Detection and Localization [C]// Pattern Recognition. ICPR International Workshops and Challenges. Cham: Springer International Publishing, 2021: 475-489.

[25] RUDOLPH M, WANDT B, ROSENHAHN B. Same Same but DifferNet: Semi-Supervised Defect Detection with Normalizing Flows [C]// 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). New York: IEEE Press, 2021: 1907-1916.

[26] GUDOVSKIY D, ISHIZAKA S, KOZUKA K. CFLOW-AD: Real-Time Unsupervised Anomaly Detection with Localization via Conditional Normalizing Flows [C]// 2022 IEEE/CVF, CVF Winter Conference on Applications of Computer Vision (WACV). New York: IEEE Press, 2022: 1819-1828.

[27] RUDOLPH M, WEHRBEIN T, ROSENHAHN B, et al. Fully Convolutional Cross-Scale-Flows for Image-Based Defect Detection [C]// 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). New York:

IEEE Press, 2022: 1829-1838.
[28] YU J, ZHENG Y, WANG X, et al. FastFlow: Unsupervised Anomaly Detection and Localization via 2D Normalizing Flows [J]. arXiv:2111.07677 [cs.CV], 2021.
[29] LIU Z, HU H, LIN Y, et al. Swin Transformer V2: Scaling up Capacity and Resolution [C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2022: 11999-12009.
[30] BERGMANN P, FAUSER M, SATTLEGGER D, et al. MVTec AD-a Comprehensive Real-World Dataset for Unsupervised Anomaly Detection [C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2019: 9584-9592.
[31] LEE S, LEE S, SONG B C. CFA: Coupled-Hypersphere-Based Feature Adaptation for Target-Oriented Anomaly Localization [J]. IEEE Access, 2022, 10: 78446-78454.
[32] ZAVRTANIK V, KRISTAN M, SKOČAJ D. DSR-a Dual Subspace Re-Projection Network for Surface Anomaly Detection [C]// Computer Vision-ECCV 2022. Lecture Notes in Computer Science. Cham: Springer, 2022: 539-554.
[33] BERGMANN P, FAUSER M, SATTLEGGER D, et al. Uninformed Students: Student-Teacher Anomaly Detection with Discriminative Latent Embeddings [C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2020: 4182-4191.

An Unsupervised Learning Method for Forward Foreign Object Detection in Suspended Monorail Systems

QIN Yong[1], LI Wei[1, 2], CAO Zhiwei[1], LIAN Lirong[1, 2], GAO Yang[1, 2], XU Lei[3], LIU Yuwen[3], ZHAO Changlong[3]

(1. State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing 100044, China;
2. School of Transportation, Beijing Jiaotong University, Beijing 100044, China;
3. CRRC Qingdao Sifang Co., Ltd., Qingdao Shandong 266111, China)

Abstract: To address the challenge of low accuracy of detecting unknown forward foreign objects in suspended monorail systems, an efficient, end-to-end unsupervised learning detection method, termed DPMF-Flow is proposed. Initially, a Cross-Scale Feature Extraction module is employed to concurrently capture fine-grained local details and high-level global semantics, thereby enhancing the feature representation of multi-scale foreign objects against complex backgrounds. Subsequently, a Dual-Path Fusion Flow module is used to extract local and global features respectively, and the generative complementary latent variables are estimated through complementary probability density, effectively improving localization precision. Finally, an Object Segmentation and Localization module integrates the latent variables information from dual-path to generate pixel-level object localization heatmaps. The results show that on a Suspended Monorail Foreign Object Intrusion dataset, the proposed method achieves a detection accuracy of 99.34%, and detects foreign objects in real-time at a frame rate of 66.04 FPS. Compared to other unsupervised methods such as Padim and Fastflow, the approach demonstrates a significant performance advantage. Furthermore, it exhibits strong generalization on the public MVTec AD dataset, offering a practical and advanced technical solution for rail transit safety.

Key words: Unsupervised object detection; Suspended monorail; Train safety; Computer vision; Deep learning

（责任编辑 杨婧婕）