E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP18;TN92　　**文献标识码**: A　　**文章编号**: 1006-8961(2025)12-3824-14

论文引用格式:Guo S X, Geng H Z, Su L, He S and Zhang X Y. 2025. Lightweight pyramid cross-attention network for orbital image defect detection. Journal of Image and Graphics, 30(12):3824-3837(郭斯栩, 耿慧拯, 粟栗, 何申, 张鑫月. 2025. 用于轨道图像缺陷检测的轻量级金字塔交叉注意力网络. 中国图象图形学报, 30(12):3824-3837)[DOI:10.11834/jig.240547]

用于轨道图像缺陷检测的轻量级金字塔交叉注意力网络

郭斯栩[1,2*],耿慧拯[1],粟栗[1],何申[1],张鑫月[1]

1. 中国移动通信有限公司研究院,北京　100032; 2. 大连交通大学中车学院,大连　116021

摘　要: **目的**　基于视觉的轨道缺陷检测方法大多存在高参数、计算复杂、检测速度慢以及精度低等缺陷,为了解决上述问题,提出一种使用RGB和深度图像进行轨道缺陷检测的轻量级金字塔交叉注意网络(lightweight pyramid cross-attention network for orbital image defect,LPCANet)。**方法**　LPCANet模型利用深度学习、RGB-D与显著性目标检测等技术,设计一种轻量级金字塔模块,能够从深度图像中提取多尺度特征图,而骨干模块从RGB图像中捕获金字塔特征细节;然后,将交叉注意力模块(cross-attention mechanism,CAM)应用于两种类型的特征映射;其次,利用空间特征提取子(spatial feature extractor,SFE)提高缺陷检测性能;最后,应用像素洗牌(pixel shuffle)操作恢复原始图像的大小。**结果**　在NEU-RSDDS-AUG、RSDD-TYPE1和RSDD-TYPE2 3种公开无服务RGB-D轨道数据集进行实验。结果表明,提出方法在NEU-RSDDS-AUG数据集的运行参数为9.90 M,计算量为2.50 G,模型大小为37.95 MB,运行速度为162.60帧/s,相比现有18种轨道缺陷检测方法,更为轻量化;与当前性能最优的CSEPNet相比,S-度量、交并比、最大F-度量、平均精度和平均绝对误差指标分别提高1.48%、0.86%、0.14%、0.03%和1.77%;在消融实验中,像素洗牌方法表现出明显优势,更适合LPCANet模型。深入分析各种骨干网络性能,实验表明,LPCANet模型不仅适用现有各种骨干网络,而且检测结果更加优秀。在非轨道数据集DAGM2007、MT和Kolektor-SDD2上进行实验,LPCANet模型在mAP、MAE与IOU指标均有提高,具备一定的泛用性。**结论**　提出的LPCANet模型综合了传统模型和深度学习模型的优点,在轨道缺陷图像检测领域具备良好的实际应用价值。

关键词: 轨道缺陷检测;深度学习;RGB-D;显著性目标检测;空间特征提取子(SFE)

Lightweight pyramid cross-attention network for orbital image defect detection

Guo Sixu[1,2*], Geng Huizheng[1], Su Li[1], He Shen[1], Zhang Xinyue[1]

1. China Mobile Research Institute, Beijing　100032, China;

2. Vehicle Engineering College, Dalian Jiaotong University, Dalian　116021, China

Abstract: **Objective** Most existing vision-based rail defect detection methods face challenges such as high parameter counts, computational complexity, slow detection speeds, and limited accuracy. Aiming to overcome these limitations, this paper introduces a lightweight pyramid cross-attention network (LPCANet) for orbital image defect detection using

收稿日期:2024-09-24;**修回日期**:2025-05-20;**预印本日期**:2025-05-27

*通信作者:郭斯栩　guosixu@chinamobile.com

基金项目:北京信息科技大学光电测量技术与仪器教育部重点实验室项目(OMTIKF2021001)

Supported by: Key Laboratory of the Ministry of Education for Optoelectronic Measurement Technology and Instrument, Beijing Information Science and Technology University (OMTIKF2021001)

RGB images and depth images. **Method** LPCANet adopts MobileNetv2 as its backbone network to extract multiscale feature maps from RGB images. Simultaneously, a lightweight pyramid module (LPM) is employed to extract similarly-sized feature maps from depth images. Each stage of the LPM comprises a sequence of operations including max pooling, a 3 × 3 convolutional layer, batch normalization, and ReLU activation, enabling efficient extraction of features from depth images. By leveraging deep learning, RGB-D technology, and salient object detection, LPCANet efficiently extracts multiscale feature representations from RGB and depth data. The LPM handles depth image features, while the backbone captures detailed pyramid features from RGB images. Subsequently, a cross-attention mechanism (CAM) is applied to integrate the feature maps from both modalities, enhancing the network's focus on relevant defect regions. Additionally, a spatial feature extractor (SFE) is introduced to further boost defect detection performance. Finally, a "pixel shuffle" operation is used to restore the output to the original image resolution. **Result** The proposed scheme was computationally evaluated using the PyTorch library in an environment equipped with an NVIDIA 3090 GPU, alongside several benchmark models for comparison. For the evaluation of LPCANet, three publicly available unsupervised RGB-D rail datasets were used: NEU-RSDDS-AUG, RSDD-TYPE1, and RSDD-TYPE2. Experimental results on the NEU-RSDDS-AUG dataset indicate that LPCANet achieves excellent efficiency, with 9.90 million parameters, a computational complexity of 2.50 G, a model size of 37.95 MB, and a running speed of 162.60 frames per second. Compared to 18 existing rail defect detection schemes, LPCANet exhibits superior lightness in performance. In particular, when compared against CSEPNet, the current best-performing model, LPCANet achieves improvements across several evaluation metrics: +1.48% in S_α, +0.86% in intersection over union (IOU), +0.14% in $F_\beta^{\max}$, +0.03% in mean average precision (mAP), and +1.77% in mean absolute error (MAE). An ablation study was conducted on four upsampling methods (interpolation, transposed convolution, patch merging, and "pixel shuffle") to evaluate their effectiveness within the LPCANet framework. Among these, the "pixel shuffle" method demonstrated clear advantages and was found to be the most suitable for the LPCANet model. Further ablation studies were conducted on four different components (backbone network, LPM, SFE, and CAM). The results indicate that CAM and SFE notably enhance the detection performance of LPCANet. An in-depth analysis of various backbone networks confirmed that LPCANet model is not only compatible with existing backbone networks but also consistently achieves superior detection results. Aiming to evaluate the model's generalization capability beyond rail datasets, experiments were also conducted on three non-rail defect datasets: DAGM2007, MT, and Kolektor-SDD2. The results show that LPCANet delivers improved performance across three key metrics: mAP, MAE, and IOU, demonstrating its potential for general-purpose defect detection tasks. **Conclusion** The LPCANet model proposed in this study effectively combines the advantages of traditional and deep learning approaches, demonstrating strong practical value in the field of rail defect image processing. In the future, this scheme will focus on further reducing the model size to achieve rapid detection speeds while ensuring further improvements in performance quality.

Key words: track defect detection; deep learning; RGB-D; significant object detection; spatial feature extractor(SFE)

0 引 言

轨道作为世界各地乘客和货物运输的重要系统发挥着至关重要的作用。然而，由于重载、天气侵蚀、材料缺陷和不当维护等多种因素，轨道容易受到损坏。这些潜在的损害行为会在轨道表面形成裂缝、磨损、腐蚀、疤痕和其他缺陷。因此，及时的轨道缺陷检测技术对于维持最佳轨道运输条件和降低安全风险至关重要。

传统的轨道表面缺陷检测主要依赖于人工肉眼检查，这种方法昂贵、劳动密集且容易出现主观变化（黎明 等，2004）。随着视觉技术的进步，缺陷检测方法可分为传统方法和深度学习方法两类。在传统方法中，利用图像中的关键显著特征，如边缘、颜色、纹理和形状，通过各种图像处理技术来区分异常物体（阎光伟 等，2023）。类似地，Manish 等人（2018）结合坎尼边缘检测和直方图分析来分析磨削表面完成度的像素强度分布和边缘条件。虽然传统检测技术可以识别出缺陷区域，但其效率和准确性可能无法满足工业要求。这是因为特征的手动设计和选择既耗时又费力，而且可能无法捕获精确缺陷检测所需的所有必要信息（Gao 等，2022）。相比之下，深度学习方法通过直接从数据集中学习相关特征表示来

消除对手动干预的需求。

卷积神经网络(convolutional neural network, CNN)目前在缺陷检测方面是一个很有前景的工具,因其擅长分析不规则和异常的图像特征,使得工业样本中非缺陷区域和缺陷区域之间的准确区分成为可能。Zhou等人(2021)探索了一种端到端的密集缺陷检测网络,编码器采用级联特征集成模块来融合多尺度特征图,而解码器则逐步将多尺度特征线索整合到最终的显著性图中。Yang等人(2022)结合差分盒计数和GrabCut算法来提高缺陷分割的速度和准确性。然而,纯卷积模型由于其有限的感受野大小而在全局能力上存在固有的限制。为了解决这个问题,Zhang等人(2022)引入一个全局上下文上采样模块,以从多尺度特征图中捕获长距离信息。这种方法可以有效地提取额外的边界细节和局部缺陷相似性。此外,Zhou等人(2024)创建了一个双注意力模块来全局捕获空间信息,并采用了一种相对位置初始化方法来增强局部信息的获取,以实现有效的钢表面缺陷检测。虽然上述方法在表面缺陷检测任务中表现出了令人印象深刻的性能,但由于单模型输入的限制,它们在同时处理RGB和深度图像方面存在局限。

由于深度图像能够详细提供几何描述性信息,几种深度视觉基础(RGB-D)模型(Li等,2021;Ding等,2022;Dong等,2022;Fan等,2020;Liu等,2020,2021;Wu等,2023;Cong等,2023;Jin等,2022)在显著性目标检测(significant object detection, SOD)中对RGB和深度图像的处理产生了兴趣。尽管现有的RGB-D模型(Huang和Gong,2024)已用于轨道缺陷检测任务,但仍面临参数、计算复杂性、推理速度、模型体积和检测性能方面的挑战(Fan等,2020;Liu等,2022a;Liu等,2022b)。为了解决这些问题,本文提出一种用于RGB-D轨道表面缺陷检测的轻量级金字塔交叉注意力网络(lightweight pyramid cross-attention network, LPCANet)。LPCANet选择经典的MobileNetv2(Dong等,2020)作为骨干网络,从RGB图像中提取多尺度特征图,同时采用轻量级金字塔模块(lightweight pyramid module, LPM)从深度图像中捕获相似大小的特征图。由于轨道深度图像的语义表示明显比传统公开的自然深度图像差,诸如通道压缩、池化增强和边缘监督等经典技巧无法提供足够的语义特征,如图1所示,第1阶段和第2阶段的轨道特征图与真实值(ground truth, GT)标签之间存在明显的语义差异,而第1阶段和第2阶段的常规特征图可以有效地传达物体的语义结构和边缘纹理。因此,使用交叉注意力机制融合两个相同分辨率的映射图,使颜色、纹理和深度信息之间产生交互。使用1×3和3×1卷积方法设计了空间特征提取子,以水平和垂直方式合并结构特征。

本文的主要贡献如下:1)提出一种新颖的轻量

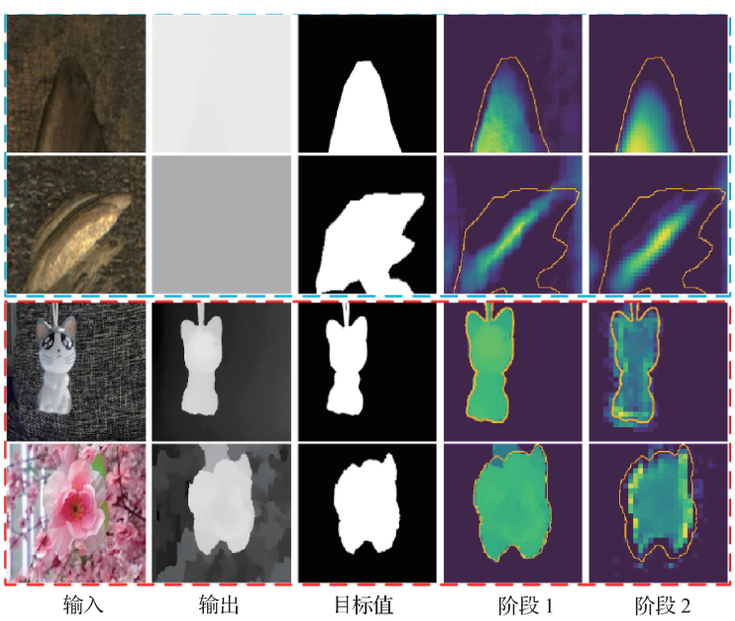


图1 轨道深度图像与普通深度图像的差异

Fig. 1 Differences between orbital depth images and normal depth images

级多尺度检测网络，用于轨道表面缺陷检测。该网络使用更少的参数和更低的计算复杂度，高效且准确地识别轨道图像缺陷。2）采用交叉注意力机制，使主干网络和 LPM 的特征图进行交互，从而实现 RGB 和深度图像中对应位置的显著轨道缺陷特征的融合；3）设计了一个空间特征提取子（spatial feature extractor，SFE），用于有效地提取轨道缺陷物体的几何结构。

1 相关知识

显著性目标检测（SOD）是计算机视觉中识别和突出图像中最显著的目标或区域的一项基本任务。通常，研究人员使用深度学习算法分析图像中的形状、纹理、轮廓或其他视觉属性。例如，Ke 和 Tsubono（2022）强调显著目标边缘质量的重要性，提出轮廓—显著性混合模块，旨在提高分段图像中边界预测的准确性。考虑到 CNN 方法有限的全局感知机制，Ma 等人（2023）使用动态互补注意模块增强感受场，以感知超大或小尺寸的对象。Yun 和 Li（2022）提出一种结合特征增强和多尺度感受野的低照度目标检测算法。其他 SOD 模型（江泽涛 等，2023）利用注意机制促进与特征图中全局上下文的更好交互，以提高模型性能，但受到更多参数数量和计算复杂性的影响。在工业环境中，由于结构简单、计算效率高和易于实现，具有较低内存且使用量更简单的 SOD 模型通常更受青睐。此外，仅依赖单一输入模式会限制可用信息的多样性，从而限制 SOD 模型。

RGB-D 模型同时使用 RGB 和深度图像执行 SOD 任务。与 RGB 显著性目标检测模型不同，深度图像可以提供不同的视觉线索，从复杂的背景中区分显著目标。汤勃等人（2017）采用双主干分别从 RGB 和深度图像中提取特征映射。这些特征映射在解码阶段使用 CNN 诱导的细化单元和点感知交互模块进行融合。Wu 等人（2023）采用标准主干提取 RGB 和深度特征，在多尺度特征融合后对粗粒度和细粒度信息进行解码。为了减轻模型体积，Jin 等人（2022）引入了一种非对称双流轻量级网络，该网络在编码阶段将 RGB 和深度图像的通道信息连接在一起进行特征融合。Zeng 等人（2024）提出一种轻量级网络，该网络采用并行注意力转移卷积、特征融合和特征重构模块来有效地提取显著特征。然而，这些轻量级模型损害了一些模型性能（Radosavovic 等，2020；Sun 等，2021；Zeng 等，2024；Zhou 等，2021）。针对轨道缺陷图像而言，RGB 图像在缺陷区域和背景之间表现出较低的对比度，而深度图像缺乏语义细节，导致这些轻量级模型的检测结果不足。因此，专门用于轨道表面缺陷检测的 RGB-D 模型是当今研究的热点。

2 LPCANet 模型

2.1 概述

LPCANet 模型的设计流程如图 2 所示。首先使用轻量型网络 MobileNetv2 作为主干网络从 RGB 图像中提取 4 个尺度特征图，同时利用提出的并行轻量型网络（LPM），每个阶段由最大池化、3 × 3 卷积层、批归一化（batch normalization，BP）和 ReLU（rectified linear unit）激活函数组成的网络结构对深度图像进行特征提取。然后，将两个分支的尺度特征图通过交叉注意力模块（cross attention module，CAM）进行交叉注意力计算，并用 SFE 模块对交叉计算后特征图进行特征融合，以强化显著性结构特征，详细的各个模块的结构和计算方法将在下文阐述。

$\boldsymbol{I}^r \in \mathbf{R}^{H\times W\times 3}$ 和 $\boldsymbol{I}^d \in \mathbf{R}^{H\times W\times 1}$ 分别代表初始输入 RGB 和深度图像，H 和 W 分别代表高度和宽度。首先，利用 MobileNetv2 主干网络从 RGB 图像中提取多级特征图，即 $\boldsymbol{F}_i^r \in \mathbf{R}^{\frac{H}{i+1}\times\frac{W}{2^{i+1}}\times C_i^r}$，其中，$i \in [1, 2, 3, 4]$。对于深度图像，则采用并行轻量级金字塔模块（LPM）以获取其多尺度特征，即 $\boldsymbol{F}_i^d \in \mathbf{R}^{\frac{H}{2^{i+1}}\times\frac{W}{2^{i+1}}\times C_i^d}$。以第 1 阶段的特征图 $\boldsymbol{F}_1^r$ 和 $\boldsymbol{F}_1^d$ 为例，提出方法执行交叉注意力计算，以处理 $\boldsymbol{F}_1^r$ 与 $\boldsymbol{F}_1^d$ 之间的关联。为了确保计算的输出维度与主干网络相匹配，对其进行相应调整。随后，利用 SFE 从之前的输出中提取结构内容。为了保持整个网络中的信息流畅，在下采样之前，应用残差连接将交叉注意力模块（CAM）与 SFE 的输出连接在一起。经过此过程的 3 次重复，可以获得第 4 阶段的特征输出，即 $\boldsymbol{F}_4 \in \mathbf{R}^{\frac{H}{32}\times\frac{W}{32}\times C_4^r}$。为了与 GT 标签 $\in \mathbf{R}^{H\times W\times 1}$ 的分辨率保持一致，LPCANet 采用批归一化和逐点卷积的组合来调整 $\boldsymbol{F}_4$ 的通道维度至 1 024。最后，LPCANet 利用像素洗牌方法对输出进行上采样处理，以匹配 GT 标签的尺寸。

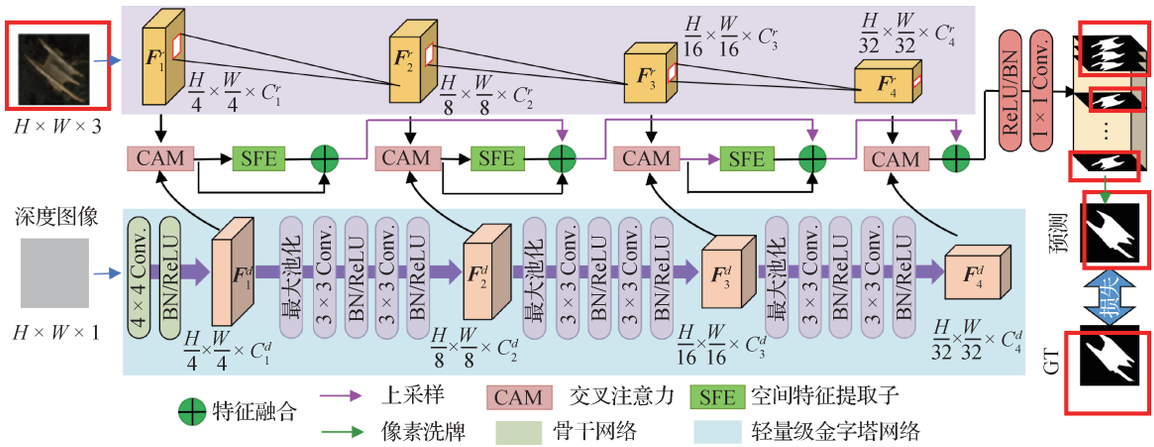


图2　LPCANet模型的流程

Fig. 2　Process of LPCANet model

2.2　轻量级金字塔模块

LPCANet选择一种轻量级金字塔网络提取深度特征，为与主干网络的特征图大小相匹配，采用一个尺寸为4×4的卷积核，步长为4，进而将图像尺寸从(H, W)减小到$\left(\frac{H}{4}, \frac{W}{4}\right)$。换言之，第1阶段的特征图分辨率标记为$\boldsymbol{F}_1^d \in \mathbf{R}^{\frac{H}{4}\times\frac{W}{4}\times C_1^d}$。接着，作为一个高效网络，其由两层组成，每层都包括有3×3卷积、批归一化和ReLU激活函数。随后，使用池化运算在相邻阶段之间进行降采样。为在保持模型性能的同时实现模型体积最小化，将第1阶段的初始投影维度设置为64维。在后续阶段中，投影维度成倍增加，为$C_{1,2,3,4}^d \in [64, 128, 256, 512]$。

2.3　交叉注意力模块

由于深度图像中的语义信息受限，当前常用的信息交互方法倾向于对信息进行压缩处理，但这种处理方式可能导致重要语义细节被过滤掉，造成语义失真。例如，诸如通道注意力之类的方法需要进行全局平均池化运算，以压缩来自宽度和高度维度的大量内容。使用交叉注意力机制对图像特征和深度特征进行交互，是一种在深度学习中提升模型性能的有效方法。LPCANet将平均池化方法应用于第1阶段和第2阶段的轨道和普通深度图像的特征图如图3所示。其中，普通深度图像的特征图像素点可以很好地与原始深度图像的特征分布对齐，而轨道深度图像的特征图像素点则未能有效地遵循原始深度图像的特征分布。

通过结合RGB图像和深度图像，LPCANet能够更准确地识别轨道上的缺陷。RGB图像提供了丰富的颜色和纹理信息，而深度图像则提供了关于物体形状和位置的三维信息。通过交叉注意力机制，模型能够融合这些信息，从而实现对轨道缺陷的准确检测。值得注意的是，在特征压缩过程中，轨道深度图像会丢失大量语义信息。为有效保留语义完整性，本文方法利用图2所示的交叉注意力机制在两个大小相同的特征图之间进行交互。每个第i阶段的特征交互形式为

$$Q_i^r = f_{\text{Linear}}(F_i^r) \tag{1}$$

$$K_i^d, V_i^d = f_{\text{Linear}}(F_i^d), f_{\text{Linear}}(F_i^d) \tag{2}$$

$$ca = f_{\text{Reshape}}\left(f_{\text{softmax}}\left(\frac{\hat{Q}_i^r \hat{K}_i^{d^{\mathrm{T}}}}{\sqrt{d_z}}\right)\hat{V}_i^d\right) \tag{3}$$

$$F_i^{ca} = f_{\text{Linear}}(ca) \in \mathbf{R}^{\left(\frac{H}{2^{i+1}}\times\frac{W}{2^{i+1}}\right)\times C_i^r} \tag{4}$$

式中，$\hat{Q}_i^r, \hat{K}_i^d, \hat{V}_i^d = f_{\text{Reshape}}((Q_i^r, K_i^d, V_i^d), N_h)$。运算符$f_{\text{Linear}}(\cdot)$表示线性投影，$f_{\text{Reshape}}(\cdot)$的目的是在多头注意力计算中修改输入或输出。$N_h$和$d_z$分别表示头的个数和每个头的维度。

2.4　空间特征提取子

为达到从深度图像中捕获结构细节的目的，在CAM的输出上运用结构SFE，如图4所示。高效结构提取涉及两个并行流，一个包含1×3卷积，后接批归一化；另一个包括3×1卷积，后接批归一化。这两个卷积核分别从水平和垂直方向收集轮廓信息。除此之外，提取子的输入端包括1×1卷积、批归一化和ReLU激活函数，而输出端通过对双流输出进行逐元素相加后，逆转输入过程。提取过程可描述为

$$f_i^{\text{in}} = ReLU(BN(Conv_{1\times1}(F_i^{\text{ca}}))) \tag{5}$$

$$f_{x\_\text{axis}} = Conv_{1\times3}(BN(f_i^{\text{in}})) \tag{6}$$

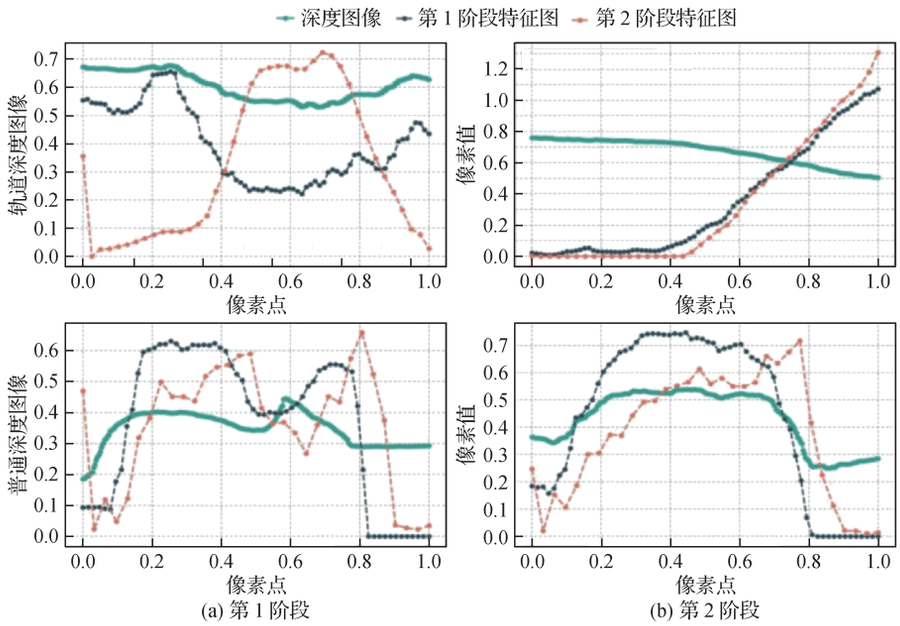


图 3　LPCANet 的特征图通道压缩后的像素分布

Fig. 3　Pixel distribution of LPCANet feature maps after channel compression ((a) stage 1; (b) stage 2)

$$f_{y\_axis} = Conv_{3\times1}(BN(f_i^{in})) \tag{7}$$

$$f_i^{out} = Conv_{1\times1}(BN(ReLU(f_{x\_axis} + f_{y\_axis}))) \tag{8}$$

式中，运算符 $Conv(\cdot)$、$BN(\cdot)$ 和 $ReLU(\cdot)$ 分别代表卷积运算、批归一化和 ReLU 激活函数。

为实现多个特征信息的高效传递，在下采样输入中采用残差连接，如图 2 所示。将 CAM、SFE 与每个下采样的输出逐元素相加，形成下一个下采样操作的输入。使用步长等于 2 的 4×4 卷积运算执行下采样过程。由于顶层的抽象语义信息可能无法直接与有缺陷对象的空间布局或几何形状完全对应。因此，在第 4 阶段，选择移除 SFE 以进一步对模型结构进行优化，进一步提高其轻量化。SFE 在各阶段的详细影响如表 3 所示。为了与 GT 标签的分辨率匹配，首先将第 4 阶段的输出维度线性投影到“d_model = 1 024”，然后采用像素洗牌（pixel shuffle）操作以恢复预测图像的尺寸。像素洗牌是一种基于深度学习的图像上采样技术，通过学习低高分辨率图像映射，将低分辨率图像转换为高分辨率图像，实现图像分辨率提升。整个过程的具体描述为

$$F_{down} = Conv_{4\times4}(f_4^{out}) \tag{9}$$

$$F_{mask} = PS(Conv_{1\times1}(ReLU(BN(F_{down}))) \tag{10}$$

式中，$F_{mask} \in \mathbf{R}^{H\times W\times1}$，运算符 $PS(\cdot)$ 表示像素洗牌运算。

2.5　损失函数

与以往采用多损失监督的研究不同，本文方法仅采用二元交叉熵损失函数对网络进行训练监督。需要指出的是，引入额外的辅助损失不仅会增加网络训练的复杂度，还需要进行额外的线性投影以对输出维度进行调整。损失函数的具体表述为

$$\mathcal{L}_{loss} = -\sum_{(a,b)}\left[p_{ab}\log(ab) + (1-p_{ab})\log(1-p_{ab})\right] \tag{11}$$

式中，p_{ab} 表示像素 (a, b) 的二元预测掩码，0 表示背

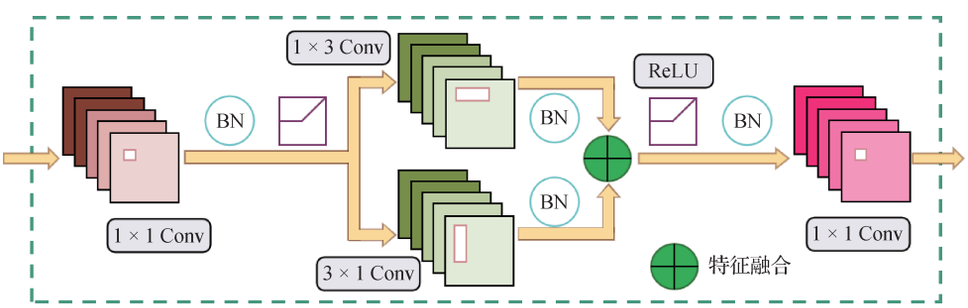


图 4　空间特征提取子

Fig. 4　SFE mode

景，1表示存在缺陷对象。类似地，q_{ab}表示相同位置的像素的GT标签。

3 实验及结果

3.1 实验环境

3.1.1 实现描述

在实验中，LPCANet模型在搭载NVIDIA 3090 GPU的环境下，利用PyTorch库对提出的模型以及其他对照模型进行计算。作为骨干网络，本文方法选用在IN-1k数据集上经过预训练的MobileNetv2模型，并将初始输入的RGB和深度图像调整大小至320×320像素。为了避免过拟合，在训练过程中采用多种数据增强技术，包括随机翻转、裁剪、旋转、高斯噪声和脉冲噪声。实验中选择AdamW优化器，初始学习率设定为0.000 1，设置动量为0.9、权重衰减为0.05。除此之外，采用余弦调度调整训练学习率。

3.1.2 数据集和评估指标

主体实验采用NEU-RSDDS-AUG公开无服务RGB-D轨道数据集，包括1个训练集和1个测试集，训练集由1 500幅图像组成，测试集包含362幅图像。数据集覆盖了轨道表面上多种缺陷类型，例如疤痕、裂缝、孔洞和焊接点。为了证明本文方法的有效性，增加RSDDs-TYPE1和RSDDs-TYPE2两个轨道数据集，并与现有方法进行比较。使用多项评价指标对分割结果的整体性能进行综合评估，包括S-度量（S_α，$\alpha = 0.5$）、交并比（intersection over union，IOU）、最大F-度量（$F_\beta^{\max}$）、最大E-度量（$E_\xi^{\max}$）、平均精度（mean average precision，mAP）和平均绝对误差（mean absolute error，MAE）。

1）S_α用于评估从预测掩码和GT标签中提取的区域和轮廓的结构相似性。

2）IOU负责衡量预测区域和GT区域之间的重叠面积。

3）$E_\xi^{\max}$反映两个分割图像之间的全局结构匹配程度。

4）$F_\beta^{\max}$通过对分割结果中的精度和召回率的评估来评估整体性能。

5）mAP通过综合权衡精确率和召回率，以评估模型的性能。

6）MAE计算了预测图像和GT图像之间的像素级误差。

3.2 实验结果

3.2.1 SOD方法对比

LPCANet与先进模型之间的计算负载比较如图5所示。可以看出，在轨道缺陷检测领域，

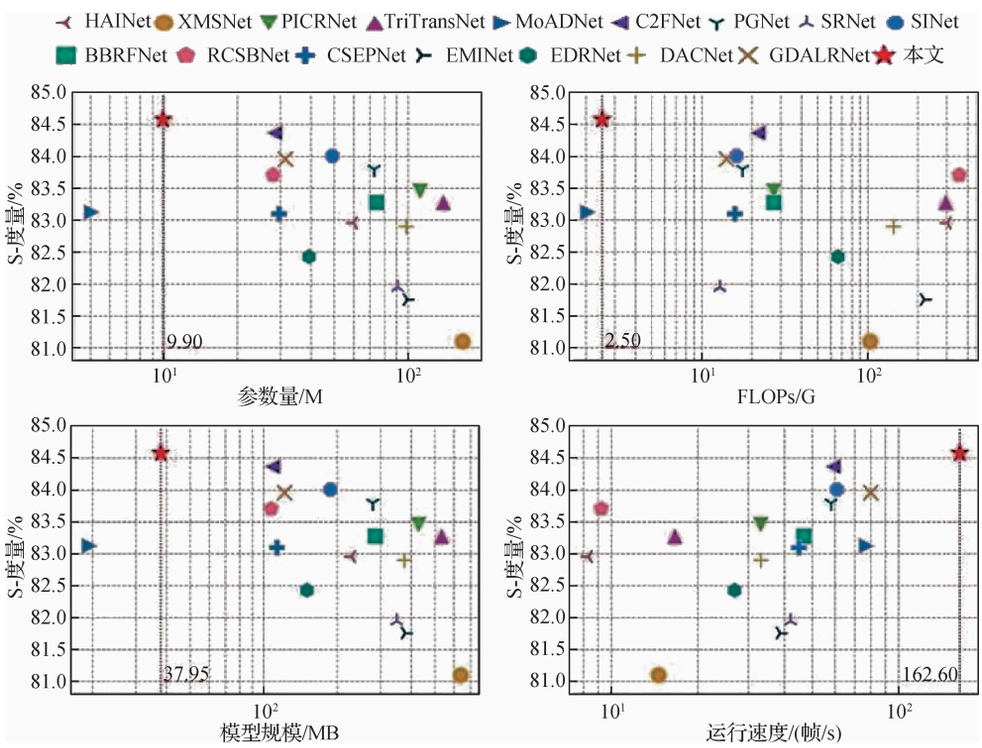


图5 LPCANet与先进模型之间的计算负载比较

Fig. 5 Comparison of computational load between LPCANet and state-of-the-art models

LPCANet的网络模型运行参数量为9.90 M，每秒浮点运算次数（floating point operations per second，FLOPs）为2.50 G，模型大小为37.95 MB，运行速度为162.60帧/s，相比现有先进的轨道缺陷检测模型，在性能上更为轻量化。

LPCANet与现行18种方法在NEU-RSDDS-AUG、RSDDs-TYPE1和RSDDs-TYPE2数据集的对比结果如表1—表3所示。

表1　提出的模型与其他18种方法在NEU-RSDDS-AUG数据集的定量比较

Table 1　Quantitative comparison between proposed model and 18 other methods on the NEU-RSDDS-AUG dataset

/%

方法	mAP	IOU	MAE	$F_\beta^{\max}$	$E_\xi^{\max}$	S_α
HAINet	92.45	81.50	7.51	87.70	91.89	82.96
CONet	86.48	71.91	13.11	79.21	84.32	75.19
XMSNet	91.31	80.11	9.06	84.80	89.47	81.10
PICRNet	92.25	81.75	7.46	87.70	91.41	83.46
TriTransNet	88.36	80.11	6.90	87.93	92.47	83.28
MoADNet	90.19	80.97	7.20	88.14	91.81	83.13
C2FNet	91.05	82.97	6.52	88.24	92.09	84.37
PGNet	93.18	81.94	7.82	87.05	92.21	83.80
SRNet	91.75	79.20	9.27	85.64	90.55	81.95
SINet	94.02	82.02	8.15	88.34	9.21	84.01
BBRFNet	93.73	81.71	7.79	88.31	91.83	83.28
RCSBNet	90.30	81.29	6.88	88.14	92.17	83.71
FDSNet	83.95	68.62	18.49	73.92	81.14	71.86
CSEPNet	94.40	82.22	8.88	88.43	92.37	83.10
EMINet	81.04	77.82	7.69	8.54	90.99	81.76
EDRNet	81.80	78.25	7.61	86.15	90.77	82.43
DACNet	83.03	79.40	6.95	88.68	91.76	82.90
GDALRNet	92.26	82.10	6.93	88.24	92.10	83.98
LPCANet	94.43	83.08	7.11	88.57	92.17	84.58

以NEU-RSDDS-AUG公开无服务RGB-D轨道数据集为例，LPCANet模型在mAP、IOU、$F_\beta^{\max}$和S_α指标上均表现优异，优于其他SOD模型。相较之下，LPCANet的mAP较SINet提高0.41%，S_α较SINet提高0.57%。LPCANet的IOU和$F_\beta^{\max}$分别比C2FNet提高0.11%和0.33%。尽管C2FNet模型在MAE方面略优于本文所提模型（6.52% vs 7.11%），但其需要更多的参数和更高的FLOPs（详见图5）。另外，PGNet模型在$E_\xi^{\max}$上略微提高0.04%，但需要超过7 000万个参数和约300 MB的模型大小。相较而言，LPCANet模型只需990万个参数和37.95 MB的模型大小。与之对比，LPCANet在所有指标上均显著优于SRNet模型，mAP、IOU、MAE、$F_\beta^{\max}$、$E_\xi^{\max}$和S_α分别提高2.68%、3.88%、2.16%、2.93%、1.62%和2.63%。此外，SRNet模型的参数数量和模型大小约9 000万和350 MB。BBRFNet和RCSBNet模型的性能评估不足且计算负载较高。综上所述，LPCANet模型在轨道缺陷检测中具有显著的性能优势。

表2　提出的模型与其他18种方法在RSDDs-TYPE1数据集的定量比较

Table 2　Quantitative comparison between proposed model and 18 other methods on the RSDDs-TYPE1 dataset

/%

方法	mAP	IOU	MAE	$F_\beta^{\max}$	$E_\xi^{\max}$	S_α
HAINet	83.77	76.45	12.34	78.45	87.65	82.34
CONet	81.53	72.89	16.78	82.34	84.32	74.56
XMSNet	84.56	81.23	9.56	75.67	89.10	81.23
PICRNet	87.12	74.56	14.23	86.78	82.78	83.78
TriTransNet	82.31	79.34	18.90	74.23	90.56	80.90
MoADNet	86.17	80.78	11.09	80.90	81.23	84.12
C2FNet	86.76	77.12	7.98	79.56	86.45	82.67
PGNet	87.44	71.90	15.45	84.12	88.98	81.45
SRNet	86.10	82.56	13.67	76.89	85.67	83.65
SINet	87.48	75.32	19.87	81.34	83.45	80.78
BBRFNet	86.29	78.67	8.45	77.65	91.78	84.98
RCSBNet	84.43	83.09	17.32	87.23	87.12	82.10
FDSNet	79.98	70.45	19.12	73.45	80.34	69.54
CSEPNet	87.50	73.87	12.78	85.67	89.56	83.21
EMINet	71.11	81.54	7.89	72.98	84.78	80.45
EDRNet	70.98	79.10	14.56	83.10	86.90	84.89
DACNet	72.68	72.34	16.34	88.09	82.12	82.76
GDALRNet	87.12	80.65	19.21	79.76	85.34	81.34
LPCANet	87.52	83.90	7.85	88.71	91.80	85.90

表3　提出的模型与其他18种方法在RSDDs-TYPE2数据集的定量比较

Table 3　Quantitative comparison between proposed model and 18 other methods on the RSDDs-TYPE2 dataset

/%

方法	mAP	IOU	MAE	$F_\beta^{\max}$	$E_\xi^{\max}$	S_α
HAINet	82.85	76.34	15.67	82.34	86.78	82.45
CONet	80.93	82.12	20.34	79.65	82.34	75.12
XMSNet	83.77	70.56	12.78	86.78	90.12	83.67
PICRNet	85.49	78.90	9.89	77.23	84.56	81.90
TriTransNet	79.91	81.23	18.45	84.56	89.78	84.23
MoADNet	84.98	74.87	14.23	90.12	81.90	80.78
C2FNet	95.30	79.45	11.56	88.78	91.23	85.65
PGNet	86.64	69.78	21.78	75.45	87.45	81.34
SRNet	84.67	83.65	8.67	81.90	85.67	83.78
SINet	87.10	73.21	16.90	85.34	88.32	82.10
BBRFNet	84.25	77.54	13.45	76.87	92.10	84.56
RCSBNet	82.78	80.12	19.12	89.23	83.45	80.98
FDSNet	77.69	75.98	23.46	83.45	80.76	70.12
CSEPNet	85.91	84.01	17.56	87.67	84.98	81.54
EMINet	70.08	71.34	10.34	78.98	91.54	83.21
EDRNet	70.43	72.67	21.23	80.10	82.12	85.90
DACNet	71.96	84.45	14.87	90.76	87.89	82.76
GDALRNet	86.65	78.23	16.54	84.54	90.34	84.02
LPCANet	86.63	84.47	8.44	91.20	92.11	86.61

3.2.2　SOD-D方法对比

本文对LPCANet模型与其他先进模型在RGB-D显著性目标检测方面的性能进行对比，如表1—表3所示。可以看出，LPCANet模型在各项指标上均超越HAINet、CONet、XMSNet、PICRNet和MoADNet模型。相较于CONet模型，所提模型在mAP、IOU、MAE、$F_\beta^{\max}$、$E_\xi^{\max}$和S_α指标上分别提升7.95%、11.17%、6.00%、9.36%、7.85%和9.39%。实验结果显示，CONet模型在结构完整性、定位准确性和分割精度等方面存在明显不足。相较于轻量级的MoADNet模型，虽然LPCANet模型具有更多的参数，但在mAP、IOU和S_α等方面仍然表现出明显优势，分别为94.43% vs. 90.19%、83.08% vs. 80.97%和84.58% vs. 83.13%。尽管MoADNet模型具有更少的参数、FLOPs和模型规模，但所提模型在推理速度上更为高效。虽然TriTransNet模型在$E_\xi^{\max}$得分上表现突出，且在MAE得分上位列第2，然而LPCANet模型在其他4项指标上表现优异。具体而言，相较于TriTransNet模型，LPCANet模型的mAP、IOU、$F_\beta^{\max}$和S_α分别提高6.07%、2.97%、0.64%和1.30%。TriTransNet模型的计算负荷较大，参数、FLOPs和模型大小分别达到约140 M、290 G和500 MB。与此类似，HAINet和XMSNet模型的计算成本也相当高。HAINet模型的FLOPs约为300 G，模型大小约为200 MB，而XMSNet模型的FLOPs约为160 G，模型大小约为600 M。综上所述，LPCANet模型在轨道缺陷检测方面表现卓越，与其他模型相比具有显著优势。

3.2.3　性能对比

针对轨道表面缺陷检测的特定模型进行性能对比，结果如表1所示。LPCANet性能超过列出的其他模型。与CSEPNet相比，LPCANet在mAP、IOU、MAE、$F_\beta^{\max}$和S_α指标上分别提高0.03%、0.86%、1.77%、0.14%和1.48%。此外，LPCANet模型在6个度量指标上均表现出明显的性能提升，平均提高10%。图5的FDSNet可视化结果揭示了其在完全分割缺陷区域方面的局限性。尽管EMINet、EDRNet、DACNet和GDALRNet模型可以成功检测到轨道表面的缺陷，但指标均低于本文提出的模型。CSEPNet、EMINet、EDRNet、DACNet和GDALRNet模型都不是轻量级模型，模型大小均超过100 MB。EMINet和DACNet模型的参数和FLOPs分别约为100 M和超过150 G，导致推理速度较慢。与之相反，LPCANet模型的计算负载相对较轻，计算运行速度为162.60帧/s。综上所述，LPCANet模型在检测轨道表面缺陷方面表现出色。

3.2.4　泛用性实验与比较

为了证明LPCANet模型的泛用性，本文增加了3个非轨道图像数据集DAGM2007（Yu等，2017）、MT（Aqeel等，2024）以及Kolektor-SDD2（Aqeel等，2025）进行实验测试，并与现存8种方法进行对比，结果如表4所示。对比的性能参数为mAP、MAE与IOU指标。通过实验结果可知，LPCANet模型对非轨道图像的缺陷检测依然有效，在mAP、MAE与IOU指标上均有提升，模型具有泛用性。

表4 非轨道数据集上提出的模型与其他8种方法之间的定量比较表

Table 4 Quantitative comparisons between the proposed model on non-orbit datasets and eight other methods

/%

方法	DAGM2007			MT			Kolektor-SDD2		
	mAP	IOU	MAE	mAP	IOU	MAE	mAP	IOU	MAE
XMSNet	89.34	88.90	8.54	92.76	91.78	10.23	86.54	87.65	9.78
PICRNet	91.23	81.23	11.65	87.45	79.56	8.90	90.12	82.78	10.45
SRNet	85.67	80.45	9.32	93.89	82.12	11.12	94.00	79.98	8.76
BBRFNet	88.78	82.34	10.89	90.34	80.67	8.15	85.98	81.01	9.54
RCSBNet	92.10	79.12	8.45	86.45	82.89	11.78	91.56	80.45	10.32
FDSNet	87.23	81.65	9.98	93.01	79.78	8.67	89.78	82.56	11.01
CSEPNet	90.56	80.98	11.45	85.12	81.34	9.23	92.34	82.78	8.29
GDALRNet	88.90	79.56	10.67	91.78	81.90	11.56	87.65	80.23	9.87
LPCANet	94.46	83.98	8.71	93.79	82.71	8.52	93.83	82.83	8.67

3.3 消融研究

3.3.1 SFE的影响

为探究所提出模型中SFE的影响，本文进行消融实验，即移除了不同阶段的SFE，结果如表5所示。SFE同时应用于第1、2、3阶段（即基线设置）时，模型各项指标除mAP外均表现最佳。需要指出的是，尽管仅在第1阶段使用SFE的mAP达到94.79%，但其他指标均为最低。此外，当SFE继续用于第4阶段时，模型的mAP（−0.22%）、IOU（−0.36%）、MAE（−0.11%）、$F_\beta^{\max}$（−0.36%）、$E_\xi^{\max}$（−0.27）和 S_α（−0.33%）与基线（表5第3行）相比均出现了下降。在第4阶段使用SFE会导致参数和FLOPs的增加，分别为+0.83 M和+0.09 G。因此，LPCANet选择基线设置作为最终模型配置。

表5 不同阶段空间特征提取子的影响

Table 5 Influence of SFE at different stages

第1阶段	第2阶段	第3阶段	第4阶段	参数量/M	FLOPs/G	运行速度/(帧/s)	mAP/%	IOU/%	MAE/%	$F_\beta^{\max}$/%	$E_\xi^{\max}$/%	S_α/%
√	–	–	–	9.69	2.40	92.38	94.79	82.16	8.41	86.62	91.50	83.54
√	√	–	–	9.70	2.41	83.88	94.15	82.55	7.48	87.61	91.63	83.73
√	√	√	–	9.90	2.50	162.60	94.43	83.08	7.11	88.57	92.17	84.58
√	√	√	√	10.73	2.58	84.81	94.21	82.72	7.22	88.21	91.90	84.25

注："√"表示使用SFE，"–"表示未使用。

3.3.2 上采样方法的影响

为了评估不同上采样方法的影响，LPCANet模型进行4种常见上采样方法（插值、转置卷积、补丁合并和像素洗牌）的消融实验，结果如图6所示，可以看出，在本文方法中，像素洗牌表现出明显优势。采用像素洗牌时，6项指标均达到最佳，相较排名第2的转置卷积，所提模型在mAP、IOU、MAE、$F_\beta^{\max}$、$E_\xi^{\max}$ 和 S_α 指标上分别提升0.34%、0.58%、0.48%、0.92%、0.79%和0.88%。因此，像素洗牌是更适合于LPCANet模型的上采样方法。

3.3.3 LPM通道维度分析

轨道深度图像的语义表征较差，可能会对模型性能产生负面影响，特别是在LPM中使用更高初始通道维度的情况。LPM通道维度的影响如图7所示。可以看出，将初始通道维度增加到96会导致模型性能下降。与通道维度64相比，这种增加导致mAP、IOU、MAE、$F_\beta^{\max}$、$E_\xi^{\max}$ 和 S_α 分别减少0.23%、0.24%、0.32%、0.73%、0.61%和0.58%。除此之

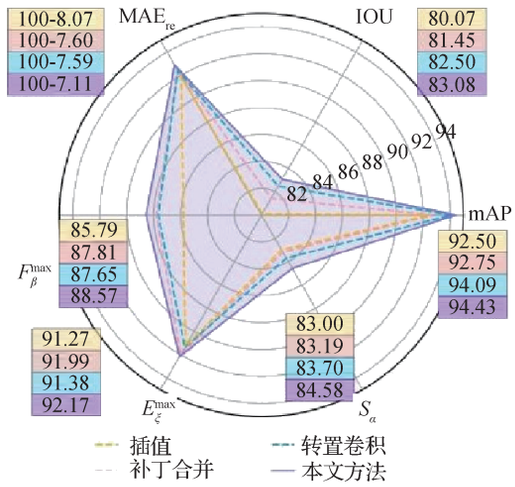


图6　不同采样方法的比较

Fig. 6　Comparison of different sampling methods

外，通道96会将模型参数从9.90 M增加到16.91 M，其不适合轻量级模型的要求。因此，本文方法在LPM中选择通道64。

3.3.4　不同组件的分析

本文提出的模型由骨干网络、LPM、SFE和CAM构成。为了评估它们的有效性，对不同组件进行消融实验，结果如图8所示。可以看出，单独测试骨干网络（Only Backbone）或单独LPM（Only LPM）执行检测任务时，模型的性能明显低于LPCANet。特别是，这两个组件的IOU分别降低了2.96%和6.55%，S_α分别降低了2.78%和5.61%。当移除CAM（No CAM）组件时，LPCANet模型的mAP、IOU、MAE、F_β^{max}、E_ξ^{max}和S_α明显下降。类似地，消除SFE组件（No SFE）导致IOU和F_β^{max}减少1.57%和7.06%。综上所述，CAM和SFE可以改善LPCANet的检测性能。

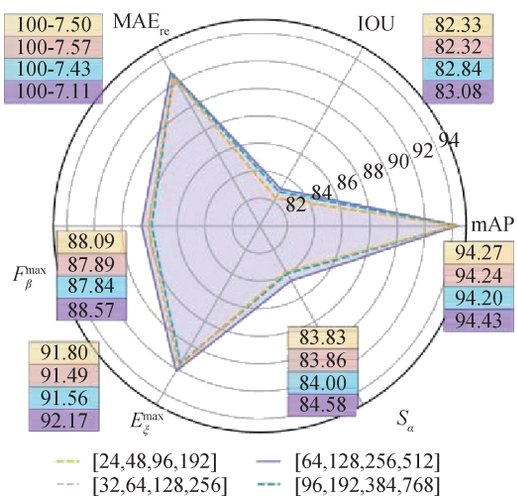


图7　LPM通道维度的影响

Fig. 7　Influence of LPM channel dimension

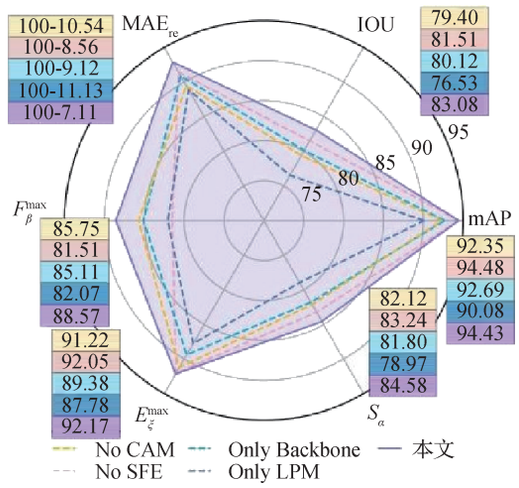


图8　不同组件对分割掩码的影响

Fig. 8　Influence of different components on segmentation mask

3.3.5　不同骨干网络的分析

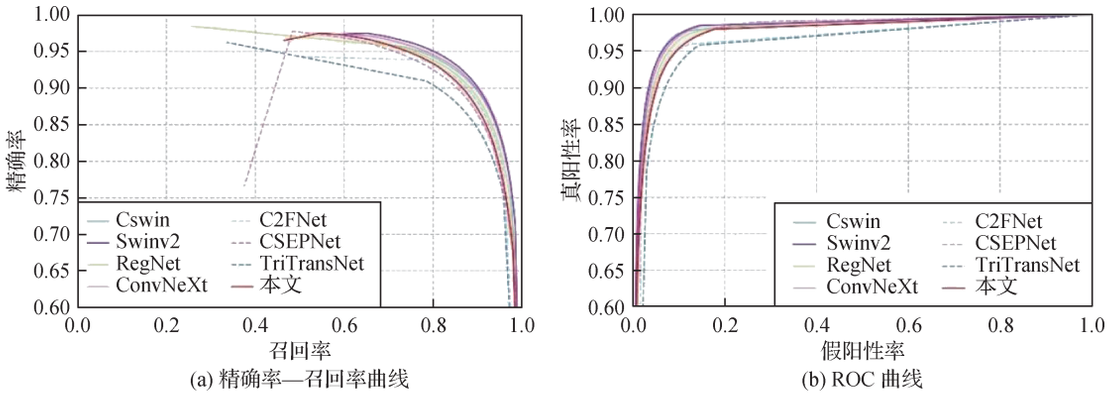
深入分析各种骨干网络性能时，LPCANet首先考虑其在模型学习图像显著特征方面的作用。不同骨干网络对分割结果的影响如表6所示。可以看出，Swinv2骨干网络最大化了所提模型的性能表现。相对于基线模型，Swinv2骨干网络将mAP、IOU、MAE、F_β^{max}、E_ξ^{max}和S_α分别提升0.76%、1.94%、

表6　不同骨干网络对分割结果的影响

Table 6　The impact of different backbone networks on segmentation results

模型	参数量/M	FLOPs/G	运行速度/(帧/s)	mAP/%	IOU/%	MAE/%	F_β^{max}/%	E_ξ^{max}/%	S_α/%
RegNet	53.32	9.79	96.48	93.56	83.43	6.41	89.74	93.34	84.89
ConvNeXt	66.63	14.38	101.40	94.72	84.35	5.80	90.23	93.99	86.31
Cswin	45.20	10.17	29.08	93.85	84.00	5.74	90.07	94.15	85.48
Swinv2	66.14	11.50	46.98	95.19	85.02	5.68	90.55	94.08	86.63
本文	9.90	2.50	162.60	94.43	83.08	7.11	88.57	92.17	84.58

1.43%、1.98%、1.91% 和 2.05%。同时，RegNet、ConvNeXt 和 Cswin 等其他骨干网络在 5 项指标上也都呈现显著改进。通过分析图 9 中的 PR（precision-recall）曲线和 ROC（receiver operating characteristic）曲线，发现 LPCANet 采用 Swinv2 骨干网络的模型展现出更为优异的性能。综上所述，LPCANet 模型不仅适用于现有的各种骨干网络，而且取得的检测结果更加优异。



(a) 精确率—召回率曲线　　(b) ROC 曲线

图 9　各种骨干网络、顶级 SOD、SOD-D、缺陷检测模型与本文所提 LPCANet 之间的精确率—召回率曲线和 ROC 曲线比较

Fig. 9　Comparison of PR and ROC curves between various backbone networks, top-level SOD, SOD-D, defect detection models, and the LPCANet proposed in this scheme ((a) PR curves; (b) ROC curves)

4　结　论

本文提出一种轻量级金字塔交叉注意力网络 LPCANet，用于处理 RGB 和深度图像，以实现针对轨道图像缺陷的检测。该网络由现有的骨干网络、提出的轻量级金字塔模块（LPM）、交叉注意力模块（CAM）和空间特征提取子（SFE）组成。利用交叉注意力计算方法将来自骨干网络和 LPM 的多尺度轨道特征图进行融合。随后，利用 SFE 从水平和垂直方向捕获轨道缺陷区域的结构纹理。最后采用像素洗牌方法对高级特征图进行上采样，以与标注的轨道缺陷标签的大小一致。实验结果表明，LPCANet 的所有组件均有效提升了模型性能。LPCANet 在轻量级体积（9.90 M 参数和 162.60 帧/s）下优于 18 种先进的模型。展望未来，计划进一步优化 LPCANet 的架构，旨在进一步减小模型体积，同时保持甚至提升当前的性能水平。这将涉及更精细的模型剪枝、量化以及更高效的网络结构设计，以期实现更快的轨道缺陷检测速度，确保在不影响检测精度的前提下，大幅提高轨道交通系统的安全监测效率与响应速度。此外，还将探索将 LPCANet 应用于更多类型的图像数据，拓宽其适用范围，为智能交通系统的发展贡献更多力量。

参考文献（References）

Aqeel M, Sharifi S, Cristani M and Setti F. 2024. Self-supervised learning for robust surface defect detection//Proceedings of the 5th International Conference on Deep Learning Theory and Applications. Dijon, France: Springer: 164-177 [DOI: 10.1007/978-3-031-66705-3_11]

Aqeel M, Sharifi S, Cristani M and Setti F. 2025. Self-supervised iterative refinement for anomaly detection in industrial quality control//Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. Porto, Portugal: [s.n.]: 173-183

Cong R M, Liu H Y, Zhang C, Zhang W, Zheng F, Song R and Kwong S. 2023. Point-aware interaction and CNN-induced refinement network for RGB-D salient object detection//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Canada: ACM: 406-416 [DOI: 10.1145/3581783.3611982]

Ding T, Li G Y, Liu Z and Wang Y K. 2022. Cross-scale edge purification network for salient object detection of steel defect images. Measurement, 199: #111429 [DOI: 10.1016/j. measurement. 2022. 111429]

Dong K, Zhou C J, Ruan Y H and Li Y Z. 2020. MobileNetV2 model for image classification//Proceedings of the 2nd International Conference on Information Technology and Computer Application. Guangzhou, China: IEEE: 476-480 [DOI: 10.1109/ITCA52113.2020.00106]

Dong X Y, Bao J M, Chen D D, Zhang W M, Yu N H, Yuan L, Chen D and Guo B N. 2022. CSWin transformer: a general vision trans-

former backbone with cross-shaped windows//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 12124-12134 [DOI: 10.1109/CVPR52688.2022.01181]

Fan D P, Ji G P, Sun G L, Cheng M M, Shen J B and Shao L. 2020. Camouflaged object detection//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 2777-2787 [DOI: 10.1109/CVPR42600.2020.00285]

Fan G R, Song K C and Yan Y H. 2020. EDRNet: encoder-decoder residual network for salient object detection of strip steel surface defects. IEEE Transactions on Instrumentation and Measurement, 69(12): 9709-9719 [DOI: 10.1109/TIM.2020.3002277]

Gao Y P, Li X Y, Wang X V, Wang L H and Gao L. 2022. A review on recent advances in vision-based defect recognition towards industrial intelligence. Journal of Manufacturing Systems, 62: 753-766 [DOI: 10.1016/j.jmsy.2021.05.008]

Huang L M and Gong A J. 2024. Surface defect detection for no-service rails with skeleton-aware accurate and fast network. IEEE Transactions on Industrial Informatics, 20(3): 4571-4581 [DOI: 10.1109/TII.2023.3327341]

Jiang Z T, Zhai F S, Qian Y, Xiao Y and Zhang S Q. 2023. Low illumination object detection combined with feature enhancement and multi-scale receptive field. Journal of Computer Research and Development, 60(4): 903-915 (江泽涛, 翟丰硕, 钱艺, 肖芸, 张少钦. 2023. 结合特征增强和多尺度感受野的低照度目标检测. 计算机研究与发展, 60(4): 903-915) [DOI: 10.7544/issn1000-1239.202111092]

Jin X, Yi K and Xu J. 2022. MoADNet: mobile asymmetric dual-stream networks for real-time and lightweight RGB-D salient object detection. IEEE Transactions on Circuits and Systems for Video Technology, 32(11): 7632-7645 [DOI: 10.1109/TCSVT.2022.3180274]

Ke Y Y and Tsubono T. 2022. Recursive contour-saliency blending network for accurate salient object detection//Proceedings of 2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 2940-2950 [DOI: 10.1109/WACV51458.2022.00143]

Li G Y, Liu Z, Chen M Y, Bai Z, Lin W S and Ling H B. 2021. Hierarchical alternate interaction network for RGB-D salient object detection. IEEE Transactions on Image Processing, 30: 3528-3542 [DOI: 10.1109/TIP.2021.3062689]

Li M, Ma C and Yang X Q. 2004. Detection of texture defects for machined surface. Journal of Image and Graphics, 9(3): 318-322 (黎明, 马聪, 杨小芹. 2004. 机械加工零件表面纹理缺陷检测. 中国图象图形学报, 9(3): 318-322) [DOI: 10.3969/j.issn.1006-8961.2004.03.011]

Liu N, Zhang N and Han J W. 2020. Learning selective self-mutual attention for RGB-D saliency detection//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 13756-13765 [DOI: 10.1109/CVPR42600.2020.01377]

Liu Z, Hu H, Lin Y T, Yao Z L, Xie Y D, Wei Y X, Ning J, Cao Y, Zhang Z, Dong L, Wei F R and Guo B N. 2022a. Swin transformer V2: scaling up capacity and resolution//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 12009-12019 [DOI: 10.1109/CVPR52688.2022.01170]

Liu Z, Mao H Z, Wu C Y, Feichtenhofer C, Darrell T and Xie S N. 2022b. A ConvNet for the 2020s//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 11976-11986 [DOI: 10.1109/CVPR52688.2022.01167]

Ma M C, Xia C Q, Xie C X, Chen X W and Li J. 2023. Boosting broader receptive fields for salient object detection. IEEE Transactions on Image Processing, 32: 1026-1038 [DOI: 10.1109/TIP.2022.3232209]

Manish R, Venkatesh A and Ashok S D. 2018. Machine vision based image processing techniques for surface finish and defect inspection in a grinding process. Materials Today: Proceedings, 5(5): 12792-12802 [DOI: 10.1016/j.matpr.2018.02.263]

Radosavovic I, Kosaraju R P, Girshick R, He K M and Dollár P. 2020. Designing network design spaces//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 10428-10436 [DOI: 10.1109/CVPR42600.2020.01044]

Sun Y J, Chen G, Zhou T, Zhang Y and Liu N. 2021. Context-aware cross-level fusion network for camouflaged object detection//Proceedings of the 30th International Joint Conference on Artificial Intelligence. Montreal, Canada: ijcai.org: 1025-1031 [DOI: 10.24963/ijcai.2021/142]

Tang B, Kong J Y and Wu S Q. 2017. Review of surface defect detection based on machine vision. Journal of Image and Graphics, 22(12): 1640-1663 (汤勃, 孔建益, 伍世虔. 2017. 机器视觉表面缺陷检测综述. 中国图象图形学报, 22(12): 1640-1663) [DOI: 10.11834/jig.160623]

Wu Z W, Wang J J, Zhou Z Y, An Z C, Jiang Q P, Demonceaux C, Sun G L and Timofte R. 2023. Object segmentation by mining cross-modal semantics//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Canada: ACM: 3455-3464 [DOI: 10.1145/3581783.3611970]

Yan G W, Zhou X J, Jiao R H and He H. 2023. Defect detection of tower bolts by fusion of priori information and feature constraints. Journal of Image and Graphics, 28(11): 3497-3508 (阎光伟, 周香君, 焦润海, 何慧. 2023. 融合先验信息和特征约束的杆塔螺栓缺陷检测. 中国图象图形学报, 28(11): 3497-3508) [DOI: 10.11834/jig.221077]

Yang H F, Wang Y Z, Hu J Y, He J T, Yao Z W and Bi Q S. 2022. Deep learning and machine vision-based inspection of rail surface

defects. IEEE Transactions on Instrumentation and Measurement, 71: #5005714 [DOI: 10.1109/TIM.2021.3138498]

Yu Z Y, Wu X J and Gu X D. 2017. Fully convolutional networks for surface defect inspection in industrial environment//Proceedings of the 11th International Conference on Computer Vision Systems. Shenzhen, China: Springer: 417-426 [DOI: 10.1007/978-3-319-68345-4_37]

Yun Y K and Lin W. 2022. SelfReformer: self-refined network with transformer for salient object detection [EB/OL]. [2024-09-24]. https://arxiv.org/pdf/2205.11283.pdf

Zeng Z H, Liu H J, Chen F L and Tan X H. 2024. AirSOD: a lightweight network for RGB-D salient object detection. IEEE Transactions on Circuits and Systems for Video Technology, 34(3): 1656-1669 [DOI: 10.1109/TCSVT.2023.3295588]

Zhang J, Ding R W, Ban M J and Guo T Y. 2022. FDSNeT: an accurate real-time surface defect segmentation network//Proceedings of 2022 IEEE International Conference on Acoustics, Speech and Signal Processing. Singapore, Singapore: IEEE [DOI: 10.1109/ICASSP43922.2022.9747311]

Zhou X, Zhou S H, Zhang Y C, Ren Z H, Jiang Z Y and Luo H F. 2024. GDALR: global dual attention and local representations in transformer for surface defect detection. Measurement, 229: #114398 [DOI: 10.1016/j.measurement.2024.114398]

Zhou X F, Fang H, Fei X B, Shi R and Zhang J Y. 2021. Edge-aware multi-level interactive network for salient object detection of strip steel surface defects. IEEE Access, 9: 149465-149476 [DOI: 10.1109/ACCESS.2021.3124814]

作者简介

郭斯栩，男，工程师，主要研究方向为图像处理、深度学习和密码学。E-mail: guosixu@chinamobile.com

耿慧拯，男，副高级工程师，主要研究方向为AI数据安全、图像处理。E-mail: genghuizheng@chinamobile.com

粟栗，男，教授级高级工程师，主要研究方向为AI数据安全。E-mail: suli@chinamobie

何申，男，教授级高级工程师，主要研究方向为AI数据安全。E-mail: heshen@chinamobile.com

张鑫月，女，工程师，主要研究方向为图像处理和AI数据安全。E-mail: zhangxinyue@chinamobile.com