

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2025)08-2775-15

论文引用格式: Sun Z B, Hu S, Zhang F and Zhou Y. 2025. Improved YOLOv7-tiny with long short range dependency feature pyramid network. Journal of Image and Graphics, 30(8):2775-2789(孙中彬, 胡帅, 张帆, 周勇. 2025. 长短程依赖特征金字塔的YOLOv7-tiny改进. 中国图象图形学报, 30(8):2775-2789)[DOI:10.11834/jig.240490]

# 长短程依赖特征金字塔的YOLOv7-tiny改进

孙中彬<sup>1,2\*</sup>, 胡帅<sup>1,2</sup>, 张帆<sup>3</sup>, 周勇<sup>1,2</sup>

1. 矿山数字化教育部工程研究中心, 徐州 221116; 2. 中国矿业大学计算机科学与技术学院, 徐州 221116;

3. 浪潮卓数大数据产业发展有限公司, 济南 250011

**摘要:** 目的 YOLOv7-tiny (you only look once version 7-tiny) 成为实时目标检测领域的常用方法, 由于其轻量化网络架构设计和较少的参数量, 整个训练过程在单个网络中进行, 检测速度快且不需要使用滑动窗口或候选区域, 在资源受限、实时性要求高的任务中表现优异。然而, YOLOv7-tiny 在特征融合阶段存在相邻层特征融合时信息丢失和非相邻层特征信息差异两个问题。为了解决上述问题, 提出一种长短程依赖特征金字塔网络 LSRD-FPN (long short range dependency feature pyramid network), 并基于该网络对 YOLOv7-tiny 方法进行改进。方法 LSRD-FPN 包括两个关键组成部分: 局部短程依赖机制 SRD (short range dependency) 和全局长程依赖机制 LRD (long range dependency)。局部短程依赖机制通过改进上采样方式和引入注意力机制, 有效缓解了特征融合过程中信息丢失的问题; 全局长程依赖机制通过引入跨层连接模块, 将主干网络的多尺度特征缩放、融合并分配到检测阶段的不同层级特征。LSRD-FPN 不仅增强了模型的特征表达能力, 而且提升了其在多尺度目标检测任务的性能表现。结果 选用两个不同场景和规模的数据集进行实验。实验结果表明, 相较于 YOLOv7-tiny, 本文方法的 mAP 分别取得 1.3% 和 0.5% 的性能提升。与参数量相当的 YOLOv5-s 和 YOLOv8-n 相比, mAP 指标在 TDD (traffic detection dataset) 数据集上分别提升 2.6% 和 0.2%, 在 Cmudsodd (coal mine underground drilling site object detection dataset) 数据集上分别提升 2.1% 和 4.4%。结论 本文提出的长短程依赖特征金字塔网络解决了 YOLOv7-tiny 在特征融合阶段存在的相邻层特征融合时信息丢失问题和非相邻层特征信息差异问题, 提升了 YOLOv7-tiny 方法的检测性能, 并优于两种参数量相当的方法 YOLOv5-s 和 YOLOv8-n。

**关键词:** 目标检测; 特征融合; 特征金字塔; YOLOv7-tiny; 多尺度特征

## Improved YOLOv7-tiny with long short range dependency feature pyramid network

Sun Zhongbin<sup>1,2\*</sup>, Hu Shuai<sup>1,2</sup>, Zhang Fan<sup>3</sup>, Zhou Yong<sup>1,2</sup>

1. Mine Digitization Engineering Research Center of the Ministry of Education, Xuzhou 221116, China;

2. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China;

3. Inspur Zhuoshu Big Data Industry Development Co., Ltd., Jinan 250011, China

**Abstract: Objective** In recent years, YOLOv7-tiny has become a commonly used method in real-time object detection. The entire training process of this method is conducted in a single network due to its lightweight network architecture design

收稿日期: 2024-09-02; 修回日期: 2024-12-23; 预印本日期: 2024-12-30

\* 通信作者: 孙中彬 zhongbin@cumt.edu.cn

基金项目: 中央高校基本科研业务费专项资金资助(2021QN1075)

Supported by: Fundamental Research Funds for the Central Universities(2021QN1075)

and fewer parameters. The method offers fast detection speed without relying on sliding windows or region proposals, which makes it suitable for tasks with limited resources and high real-time requirements. However, YOLOv7-tiny has two problems in the feature fusion stage: one is information loss in adjacent layer feature fusion, and the other involves differences in non-adjacent layer feature information. Specifically, YOLOv7-tiny uses the traditional nearest neighbor upsampling method in adjacent layer feature fusion, which may lead to jagged edges in the generated feature map. This condition reduces the quality and expression ability of the feature map. The problem of non-adjacent layer feature differences occurs during the bidirectional fusion process of YOLOv7-tiny using feature pyramids. The unique information of upper and lower layers is gradually "diluted". As a result, feature maps contain different scale information in the feature extraction and detection stages, which may seriously affect the ability of the model to detect large- or small-scale objects. **Method** This study proposes a long short range dependency feature pyramid network (LSRD-FPN) to solve the two problems. The network will be employed to improve the YOLOv7-tiny method. LSRD-FPN consists of two key components: the local short range dependency (SRD) mechanism and the global long range dependency (LRD) mechanism. SRD improves the upsampling method and introduces an attention mechanism. It uses the lightweight feature upsampling method CARAFE instead of the traditional nearest neighbor upsampling method, with an increase of only approximately 20 000 parameters. In addition, adding a non-parametric attention mechanism SimAM after local feature fusion aims to enhance feature representation and enhance perceptual range, which effectively reduces the problem of information loss during the feature fusion process. LRD is inspired by the ResNet and Libra R-CNN models by introducing cross layer connection modules. In this study, multi-scale feature maps of different resolutions in the backbone network are scaled and adjusted to the same scale. Then, these maps are fused and assigned to different levels in the detection stage. The extreme scale object feature information of the backbone network is directly input into the detection stage. This improvement enhances not only the feature expression ability of the model but also its performance in multi-scale object detection tasks. **Result** The training process of this study is conducted under the Ubuntu 20.04.4 LTS operating system, with a GPU configured as an NVIDIA RTX 3090 and a graphics memory size of 24 GB. The input image is fixed to  $640 \times 640$  pixels, the batch size is set to 16, and 100 epochs are trained. Other parameter settings are set using the default YOLOv7-tiny settings. The method proposed in this study is compared on two datasets with different scenarios and quantities, namely, the Traffic Detection Dataset TDD and the Coal mine underground drilling site object detection dataset Cmudsodd. This experiment uses YOLOv7-tiny as the benchmark and embeds LSRD-FPN into the YOLOv7-tiny. After 100 epochs of training, the experimental results show that the method achieves performance improvements of 1.3% mAP and 0.5% mAP compared with the benchmark model YOLOv7-tiny on the TDD and Cmudsodd datasets, respectively. Despite significant performance improvements, the number of parameters remains at a relatively low level. This study conducts ablation experiments on two sub models of LSRD-FPN, namely, LRD and SRD. The local SRD mechanism achieves improvements of 0.6% mAP and 0.2% mAP on the TDD and Cmudsodd datasets, respectively. The global LRD mechanism achieves improvements of 0.7% mAP and 0.3% mAP on the TDD and Cmudsodd datasets, respectively. Compared with other real-time object detection algorithms with the same number of parameters, the algorithm proposed in this study improves the TDD dataset by 2.6% mAP compared with YOLOv5-s and by 0.2% mAP compared with YOLOv8-n. In contrast to the two algorithms, the Cmudsodd dataset shows improvements of 2.1% mAP and 4.4% mAP. In addition, the frame per second (FPS) of the model proposed in this study is higher than 160, which meets the requirements of real-time detection tasks. Therefore, the proposed method not only improves performance but also exhibits rapid deployment, which can be more quickly applied to practical scenarios. **Conclusion** The proposed LSRD-FPN method can effectively improve the detection performance of the object detection model while involving fewer parameters and floating-point operations to ensure that the model meets the requirements of real-time detection speed. In addition, LSRD-FPN can be applied to not only the YOLOv7-tiny model but also other object detection models. The plug and play nature of LSRD-FPN eases its deployment to other object detection models and results in performance improvements.

**Key words:** object detection; feature fusion; feature pyramid; YOLOv7-tiny; multiscale feature

## 0 引言

目标检测是计算机视觉领域一项至关重要的任务,旨在自动识别图像或视频中的物体,并准确预测它们的位置和类别。相较于图像分类任务,目标检测要求不仅仅是对图像中的物体进行识别,还需要进行精确的位置标记,在多个实际应用场景中得到了广泛应用,包括视频监控(叶晨等,2022)、异常检测(李晨玉等,2024)、红外目标检测(Wu等,2023)、自动驾驶(李熙莹等,2023)和医学图像分析(史彩娟等,2024)等领域。

随着深度学习技术的崛起,基于深度学习的目标检测算法成为主要的研究方向,根据其检测方式和思想的不同,通常可以细分为两类:一阶段算法(Sirisha和Sudha,2023)和两阶段算法(Zou等,2023)。Redmon等人(2016)提出一阶段实时目标检测算法YOLO(you only look once),推动了实时目标检测领域的快速发展。YOLO方法的改进版本很多(Terven等,2023),比较有代表性的是YOLOv7(Wang等,2023),其在各领域中有广泛的应用,例如Zhou等人(2023)和Chen等人(2023)将YOLOv7应用于农业检测获得了显著的效果。YOLOv7-tiny作为YOLOv7轻量级版本,模型参数量仅6M左右,更有利于部署到嵌入式设备、移动设备或边缘设备,并且广泛应用于资源受限、实时性要求高的任务。Zhang等人(2023)以及齐向明和董旭(2023)分别将YOLOv7-tiny应用于无人机检测和钢材表面缺陷检测领域,均实现了边缘设备的高效部署。

YOLOv7-tiny方法的网络结构可分为3个部分:特征提取阶段、特征融合阶段和检测阶段。特征融合通过使用特征金字塔(Lin等,2017)或路径聚合网络(Liu等,2018)进行,在主干网络特征提取阶段后引入自上而下融合通路,利用语义信息丰富的高层特征逐层增强缺少语义信息的低层特征,但这会引发相邻层特征融合时特征信息丢失问题。图1给出了YOLOv7-tiny采用的特征金字塔融合方法。具体而言,特征 $B_3$ 经过金字塔池化得到输出特征 $N_3$ ,然后通过最近邻上采样操作将其调整为与特征 $B_2$ 相同的尺度以便后续融合。然而,最近邻上采样可能会导致图像或特征图失真,因为它简单地复制最近的像素值或特征值,可能导致生成的图像或特征图

出现锯齿状边缘,从而降低图像质量和特征表达能力。另外,经过特征金字塔的深度特征融合,来自特征提取阶段和检测阶段的相同尺度特征会出现非相邻层特征信息差异问题,鉴于特征提取阶段与检测阶段并非直接相邻,伴随特征融合的推进,高层和低层特征所具有的独特信息会逐渐“稀释”。这一现象极有可能对检测图像中的大尺度或小尺度物体造成严重影响。

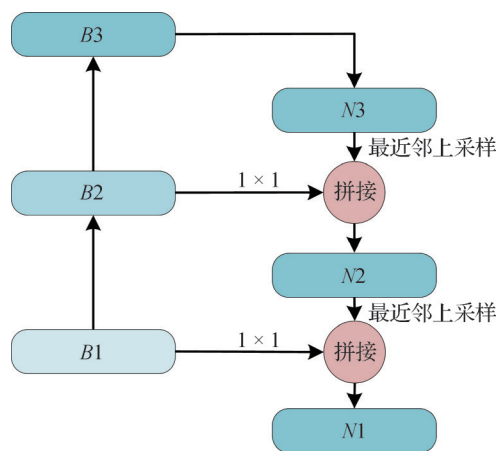


图1 YOLOv7-tiny特征金字塔融合方法

Fig. 1 Feature pyramid fusion method in YOLOv7-tiny

针对上述问题,国内外研究学者提出一些方案。阳丽莎等人(2025)提出改进的ELAN(efficient layer aggregation network)结构替代YOLOv7-tiny特征融合阶段的卷积结构,提升了模型的多尺度检测能力。邱云飞和辛浩(2024)在处理YOLOv7-tiny特征信息丢失问题时,提出一种并行双路径主干下全局特征融合方法,保证了特征信息之间的相互补充。Yang等人(2023)将Biformer注意力机制引入YOLOv7-tiny的特征融合阶段,通过在局部融合之前对特征实施注意力机制,减少特征信息丢失情况。Wang等人(2024)为了解决特征丢失问题并且充分利用多尺度信息,提出RepGFPN(reparameterized generalized feature pyramid network)代替YOLOv7-tiny的特征融合阶段,实现了检测性能的提升。Ma等人(2023)在YOLOv7-tiny方法的基础上利用ULSAM(ultra-lightweight subspace attention mechanism)注意力机制减少目标特征的丢失,并且使用P2BiFPN(P2 Bi-directional feature pyramid network)提升多尺度特征融合能力。She等人(2023)为了解决特征信息丢失问题,在YOLOv7-tiny方法中引入通道注意力机制,

并且建立从主干网络到特征金字塔网络的跨层连接通路,增强特征融合阶段的小尺度物体的位置信息,提升小尺度物体的定位精度。然而,特征信息丢失和特征信息差异问题是同时存在于实时目标检测算法特征融合阶段的,现有研究并未同时关注这两个问题,这在一定程度上影响检测阶段特征的优劣,最终影响检测性能。

为了解决上述两个问题,本研究提出长短程依赖特征金字塔网络(long short range dependency feature pyramid network, LSRD-FPN),并对 YOLOv7-tiny 方法进行了改进。LSRD-FPN 通过建立局部短程依赖机制(short range dependency, SRD)解决相邻层特征融合时特征信息丢失问题,并同时采用全局长程依赖机制(long range dependency, LRD)解决非相邻层特征信息的差异问题。在实验中,本文方法在 TDD 和 Cmudsodd 数据集上的平均精度均值(mean average precision, mAP)分别提升了 1.3% 和 0.5%,并优于两种参数量相当的方法 YOLOv5-s 和 YOLOv8-n。此外通过消融实验分析发现,LRD 和 SRD 都有效地提升了整体模型的检测性能。LRD 和 SRD 的设计考虑了参数数量的控制,适合于实时目标检测的应用场景。

本文的主要贡献如下:1)提出长短程依赖特征金字塔网络 LSRD-FPN,并对 YOLOv7-tiny 方法进行了改进,有效提升了 YOLOv7-tiny 方法的检测性能,优于两种参数量相当的方法 YOLOv5-s 和 YOLOv8-n; 2)提出 SRD 局部短程依赖机制,采用特征上采样和无参注意力机制,解决了相邻层特征融合时的信息丢失问题,增强了 YOLOv7-tiny 的特征表示能力; 3)提出 LRD 全局长程依赖机制,增强检测阶段特征图的语义信息和细节信息,解决了非相邻层特征信息差异问题,提升了 YOLOv7-tiny 的多尺度检测能力。

## 1 方法

### 1.1 网络整体架构

YOLOv7-tiny 存在两个问题:相邻层特征融合时特征信息丢失和非相邻层特征信息差异。本研究以这两个问题为切入点,设计了基于长短程依赖机制的特征金字塔网络 LSRD-FPN,并将其替代 YOLOv7-tiny 的特征金字塔网络,旨在提升 YOLOv7-

tiny 方法的特征信息表达能力,从而提高模型的多尺度检测性能。基于 LSRD-FPN 的 YOLOv7-tiny 详细网络架构如图 2 所示。

在图 2 中, YOLOv7-tiny 接收  $640 \times 640$  像素尺寸的 RGB 三通道图像作为输入。为了减少参数量,输入图像首先经过两次  $3 \times 3$  步长为 2 的卷积核进行下采样,将特征图尺寸减小为输入图像的  $1/4$ 。然后,经过 ELAN-Tiny 和 Maxpooling 结构,进一步对图像进行下采样,并增加多尺度表达能力,得到不同尺度的 3 幅特征图,将这 3 幅特征图作为特征融合阶段的输入。在特征融合阶段,不同层级特征图特征融合采用局部短程依赖机制 SRD,旨在解决相邻层特征融合时特征信息丢失问题。此外,提出自特征融合阶段到检测阶段的全局长程依赖机制 LRD,利用主干网络特征增强检测阶段特征,提升模型多尺度检测能力。

### 1.2 局部短程依赖机制

局部短程依赖机制旨在解决相邻层特征融合时出现的特征信息丢失问题,其原因在于两幅输入特征图的分辨率存在差异。为了能够让两幅特征图更好地进行特征融合,需要使用上采样将低分辨率特征图调整到与高分辨率特征图相对应的尺度。

YOLOv7-tiny 在进行特征融合时,采用最近邻上采样调整不同特征图的尺寸,如图 3 所示。假设原始输入特征图的尺寸为宽  $W$  和高  $H$ ,若进行两倍图像上采样操作,将生成宽  $W'$  ( $W' = 2W$ ) 和高  $H'$  ( $H' = 2H$ ) 的特征图,生成特征图的  $2 \times 2$  区域将由原始特征图的单个元素填充。

具体而言,最近邻上采样通过选择最接近目标像素位置的原始像素进行插值。这种方法计算简单、不会引入额外参数,但会导致图像出现锯齿状边缘。此外,这些特征融合通常位于网络的特征融合阶段,不可避免地对检测阶段的特征以及检测结果造成一定影响。本研究提出局部短程依赖机制 SRD,SRD 使得模型更专注于局部特征的学习和提取,通过引入可学习特征上采样算子与注意力机制,从而更好地捕获物体的特征,提高输出特征的表达能力。

图 4 比较了 YOLOv7-tiny 特征融合模块和基于 SRD 的特征融合模块。具体而言,本研究引入特征上采样算子 CARAFE (content-aware reassembly of features)(Wang 等,2019),通过内容感知的特征重组

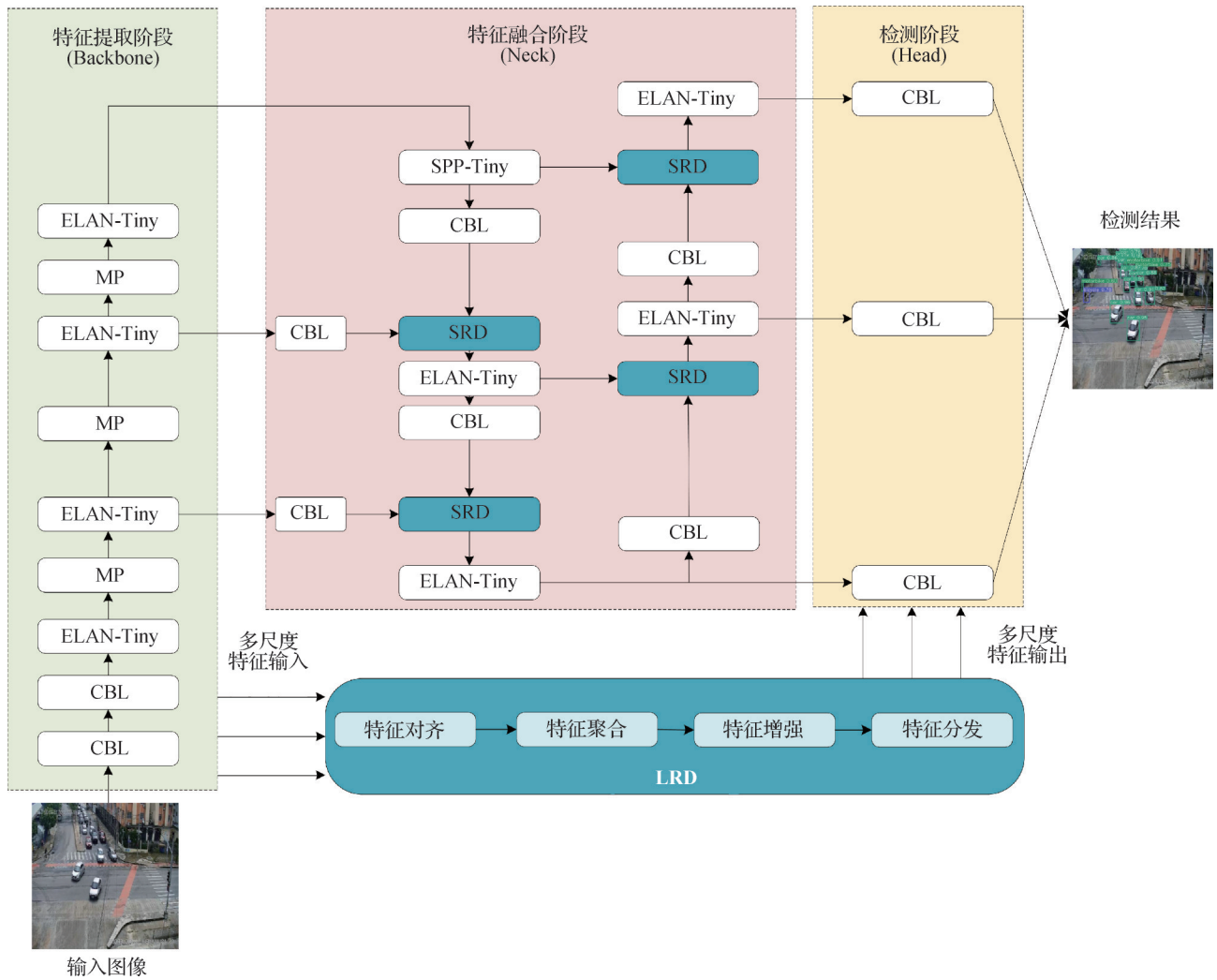


图2 改进模型的网络架构图

Fig. 2 Network architecture diagram of improved model

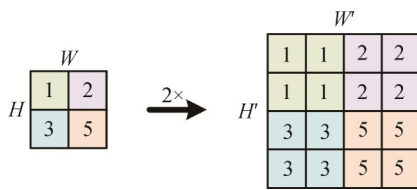
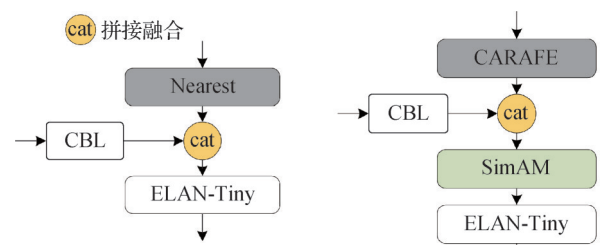


图3 最近邻上采样示意图

Fig. 3 Schematic diagram of nearest neighbor upsampling

方式对特征图进行特征上采样,这样低分辨率特征图经过上采样后将会产生更少的信息丢失。在融合操作之后,使用SimAM(simple, parameter-free attention module)(Yang等,2021)为特征图分配三维的注意力权重,以建立输入数据不同位置之间的上下文信息关联,从而使模型能够更好地理解图像的整体语义。

CARAFE是一种轻量级的上采样方式,相较于传统的最近邻插值、双线性插值和转置卷积等,能更



(a) YOLOv7-tiny特征融合模块 (b) 基于SRD的特征融合模块

图4 局部特征融合方法比较

Fig. 4 Comparison of local feature fusion methods

((a) YOLOv7-tiny feature fusion module;  
(b) SRD based feature fusion module)

好地保留图像细节的语义信息。尽管CARAFE是可学习的上采样方式,但是嵌入网络之后引入的参数量相对较少。CARAFE包括核预测模块和内容感知重组模块,如图5所示。核预测模块首先对输入特

征进行通道压缩,显著减少了后续卷积操作的参数量。然后,将输出特征传入内容编码器进行编码,并且生成重组核。核归一化对每个重组核应用 softmax 函数,以确保核内数值总和为1。内容感知重组模块将输出特征图的每个位置映射到输入特征图,并通过与预测出的上采样核进行点积操作,生成输出特征图。这一过程使得模型能够更关注局部的相关特征信息,从而增强了特征图的整体语义表达能力。

SimAM是一种简单而高效的注意力机制,旨在便捷地将注意力机制整合到深度学习模型中。与经典注意力机制使用基于一维的通道注意力机制和二

维的空间注意力机制不同,SimAM使用三维权重,为每个像素分配唯一的权重。在不引入额外参数的前提下,SimAM可以有效提升网络的性能。

SRD由CARAFE上采样和SimAM注意力机制组合而成,在引入少量参数的前提下,显著提高目标检测模型的检测性能。在特征金字塔自上而下通路中,低分辨率特征图通过上采样生成高分辨率特征图,SRD使用CARAFE上采样代替最近邻上采样,并对融合后的特征使用注意力机制以增强特征表示。在特征金字塔自下而上通路中,并未使用上采样方式,而是利用步长为2的 $3 \times 3$ 卷积核进行下采样,SRD仅对特征融合结果使用注意力机制。

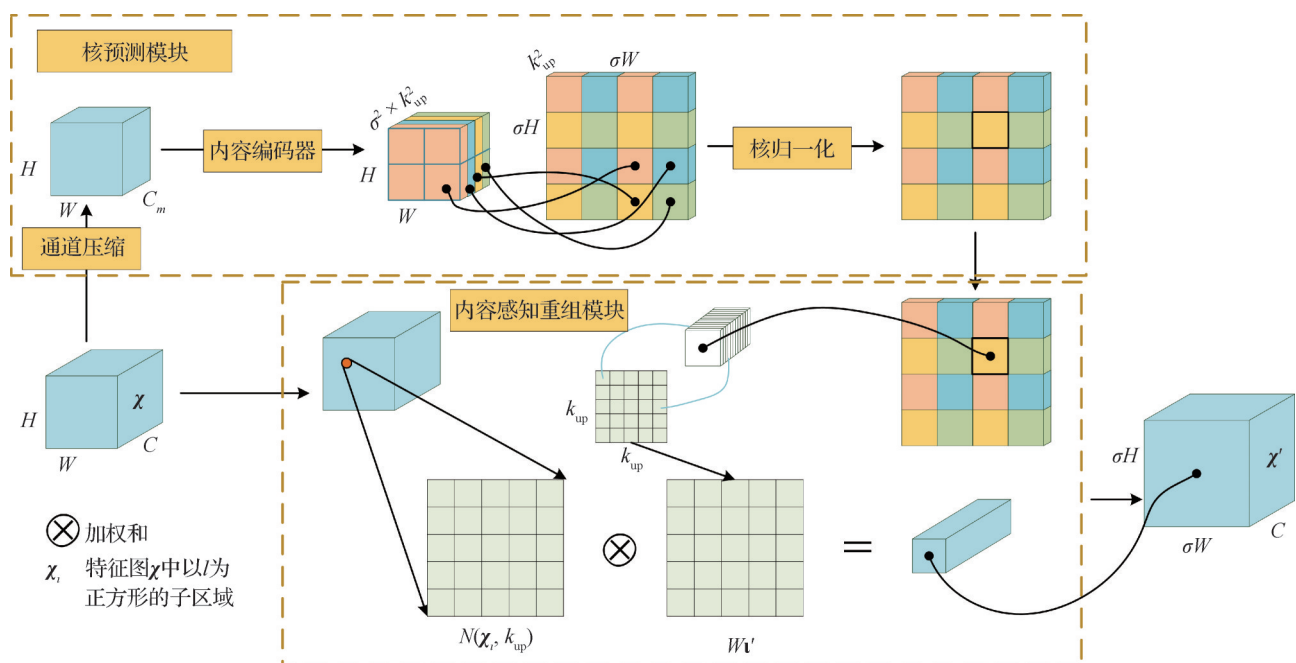


图5 CARAFE上采样结构示意图

Fig. 5 Schematic diagram of CARAFE upsampling structure

### 1.3 全局长程依赖机制

随着特征金字塔网络自上而下和自下而上双向融合的过程,与主干网络相对应尺寸的特征图在检测阶段通常会出现语义信息和细节信息的丢失,同时也会与主干网络生成的特征存在一定的差异。这一现象主要由于特征融合过程中不同尺度特征之间的相互影响所导致。为解决主干网络特征提取阶段和检测阶段特征出现差异的问题,基于 ResNet (residual neural network) (He 等, 2016) 和 Libra R-CNN (Libra region-based convolutional neural network) (Pang 等, 2019) 等方法思路,本研究提出一种基于

跨层连接的全局长程依赖机制LRD。

LRD采用了ResNet的残差思想,建立了从特征提取阶段到检测阶段的全局长程依赖机制。同时,受到Libra R-CNN的启发,LRD将主干网络的多尺度特征进行融合,并将其分配到检测阶段,然后与特征融合阶段的输出特征进一步融合。但与Libra R-CNN架构不同的是,LRD是以跨层连接的方式融合到特征金字塔网络,而不是完全替代特征金字塔网络,因此检测阶段特征具有更强的多尺度表达能力,并且提升对极端尺度物体检测效果。

全局长程依赖模块LRD的示意图如图6所示,

其中,  $B$ 、 $N$ 、 $H$  分别表示特征提取阶段(backbone)、特征融合阶段(neck)和检测阶段(head)各层级特征。为了确保模型具有较低的数量参数并提升检测的实时性, 在设计LRD时以轻量化为主要思想。首先对齐多尺度特征, 利用  $2 \times 2$  池化窗口的最大池化操作将  $B1$  特征调整至  $B2$  尺度, 并且利用 CARAFE 上采样调整  $B3$  特征至  $B2$  尺度, 以获取同一尺度但具有不同语义信息和细节信息的3个层级特征。然后将  $B2$  特征与缩放后的  $B1$  和  $B3$  特征拼接, 得到聚合后的特征  $B'$ 。对  $B'$  特征使用  $1 \times 1$  卷积调整通道长度, 以实现特征跨通道的交互并减少后续计算参数量, 得到富语义、富细节的输出特征  $H'$ 。对  $H'$  特征采用与聚合方式相反的操作, 生成不同尺度的3个层级特征, 并将特征分发到检测阶段的输出特征  $H1$ 、 $H2$  和  $H3$ , 增强检测阶段特征的多尺度检测能力。

## 2 实验与分析

### 2.1 数据集

为了验证改进方法的有效性, 实验选用 Kaggle 平台提供的开源数据集 TDD (traffic detection dataset) (FSMVU, 2023) 和科学数据银行平台提供的开源数据集 Cmudsodd (coal mine underground drilling site object detection dataset) (Zhou 等, 2024)。如图7所示, TDD 数据集涵盖了来自不同国家交通摄像头

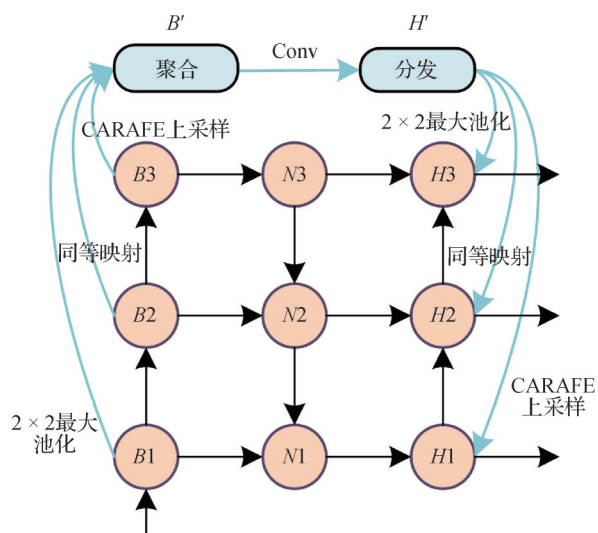


图6 LRD机制示意图  
Fig. 6 LRD mechanism schematic diagram

捕捉到的图像, 每幅图像都拥有精心标注的物体边界框标签, 用以识别各种对象, 包括行人和多种交通工具。该数据集包括在不同天气条件下拍摄的交通场景图像, 适用于真实世界的应用场景。Cmudsodd 数据集则以煤矿井下钻场打钻现场为场景。该数据集包括了夹持器、钻机卡盘和煤矿工人等多类检测目标。煤矿井下光照不足等环境因素的影响对目标检测任务带来了极大的挑战。利用 TDD 数据集和 Cmudsodd 数据集可以有助于评估本文方法在不同应用场景、不同环境因素下的鲁棒性。

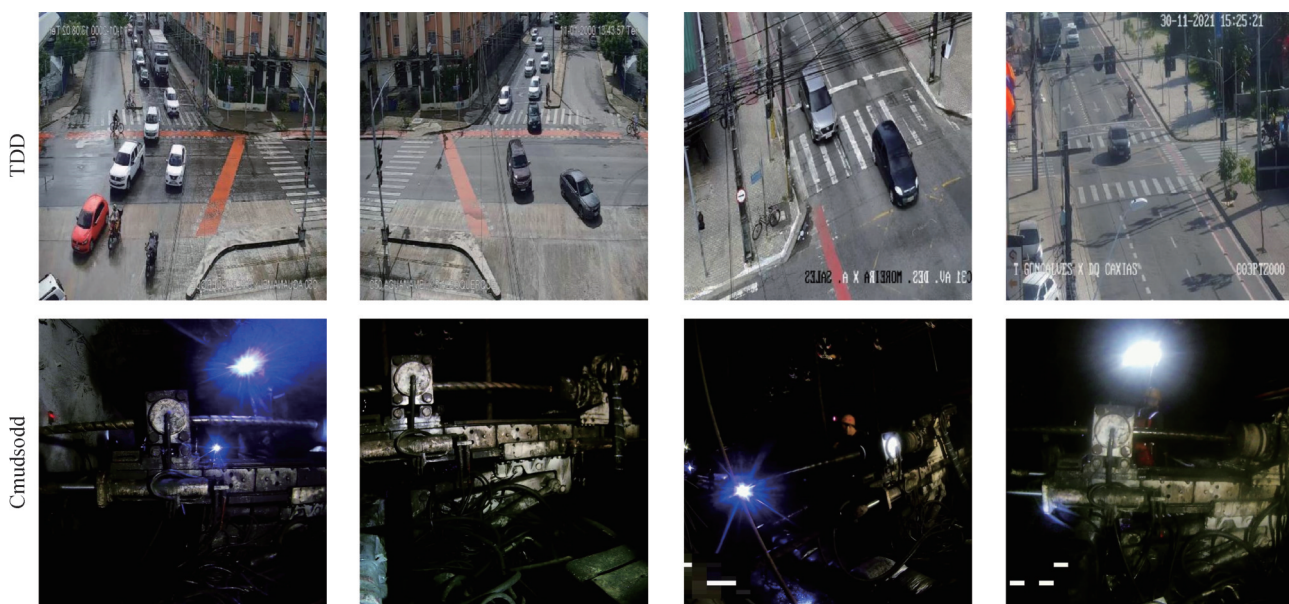


图7 TDD数据集和Cmudsodd数据集图像  
Fig. 7 TDD dataset and Cmudsodd dataset images

TDD数据集和Cmudsodd数据集的数据统计信息如表1所示。通过表中数据可以发现,两个数据集中物体尺度分布存在差异,并且平均每

幅图像标签数量也不同,从而有效验证模型在不同场景下、不同任务难度下的多尺度检测能力。

表1 TDD和Cmudsodd数据集比较  
Table 1 Comparison of TDD and Cmudsodd datasets

数据集	图像总数/幅	训练集/幅	验证集/幅	测试集/幅	目标类别	标签总数	平均每幅图像标签数	物体尺度占比(s:m:l)
TDD	6 633	5 805	549	279	5	79 628	12	1:0.25:0.39
Cmudsodd	70 683	49 477	7 068	14 138	5	262 345	3.7	1:0.47:3.23

注:s为小尺度,m为中尺度,l为大尺度。

## 2.2 评价指标

实验采用了目标检测领域通用的几种评价指标,分别分析了不同阈值和尺度情况下的检测精度和模型浮点运算数。平均精度(average precision, AP)表示单个类别precision-recall曲线下的面积,其数值越高,表明模型性能越好。mAP表示对所有类别的AP值取平均,从而综合反映了模型在整个数据集上的性能表现。AP0.5和AP0.75分别表示将交并比(intersection over union, IoU)值大于0.5和0.75的预测边界框划分为正样本,并且根据这些样本计算AP值。在存在多尺度物体的数据集中,通常使用APs(小尺度AP)、APm(中尺度AP)和APl(大尺度AP)描述不同尺度物体的检测效果。GFLOPs(Giga floating-point operations per second)主要用于表征模型整体的计算规模,是衡量模型计算复杂程度的关键指标。

## 2.3 实验设置

实验的训练过程在Ubuntu 20.04.4 LTS操作系统下进行,GPU配置为单张NVIDIA RTX 3090,显存大小为24 GB。CPU配置为Intel Xeon Platinum 8358P,PyTorch版本为2.0.0+cu118。具体的训练参数如表2所示。

其他参数设置与YOLOv7-tiny方法默认设置相同。值得注意的是,本研究使用K-means为TDD数据集和Cmudsodd数据集分别聚类产生9种不同尺度的先验框(anchors),其中为检测大、中、小物体各分配3个anchors。由于TDD数据集和Cmudsodd数据集物体尺度和形状分布不同,因此anchors设置也存在一定的差异。两个数据集的具体anchors设置如表3所示。

## 2.4 实验结果

### 2.4.1 本文方法与基础方法比较

本文提出长短程依赖机制LSRD-FPN取代

表2 模型训练参数设置

Table 2 Model training parameter settings

参数	参数值
输入图像尺寸/像素	640 × 640
批大小	16
训练总轮数	100
数据加载器数量	8
优化器	SGD
预训练权重	None
线性学习率	0.01
余弦退火学习率	0.1
mosaic	1.0
mixup	0.15
paste_in	0.15

表3 模型训练anchors设置

Table 3 Model training anchors settings

尺度	TDD	Cmudsodd
小	10 × 13	25 × 21
小	16 × 30	28 × 35
小	33 × 23	51 × 48
中	30 × 61	158 × 22
中	62 × 45	57 × 70
中	59 × 119	285 × 26
大	116 × 90	67 × 124
大	156 × 198	301 × 48
大	373 × 326	122 × 211

YOLOv7-tiny中的特征金字塔网络(feature pyramid network)FPN,并且在两个数据集上对所提出的方法进行了实验比较,实验结果如表4所示。表4结果表

明,基于本文提出的LSRD-FPN可以实现在不同场景数据集的mAP提升。在TDD交通检测数据集上,基于长程依赖特征金字塔的方法可以实现1.3% mAP的提升,并且在0.5和0.75阈值下本文方法分别提升0.9% mAP和2.7% mAP。尽管小尺度物体的检测性能略有下降,但是在中、大尺度物体检测精度均有提升。在Cmudsodd数据集上呈现同样的提升效果,本文方法与基础方法相比提升0.5% mAP,虽然小尺度物体的性能存在小幅下降,其他指标的检测精度均有不同程度的提升。因此,基于LSRD-

FPN的YOLOv7-tiny方法,与基准YOLOv7-tiny方法相比,在牺牲部分小尺度物体的特征表示能力时增强了中、大尺度物体的特征表示,实现中、大尺度物体检测性能以及整体检测性能的提升。

尽管本文方法在各类评价指标下取得了不错的结果,证明LSRD-FPN增强了YOLOv7-tiny的特征表示能力,然而为了更加直观展示本文方法在特征层面的优势,本研究使用Grad-CAM (gradient-weighted class activation mapping)(Selvaraju等,2017)对TDD数据集的训练过程进行可视化分析,结果如图8所示。

表4 本文方法与基础方法对比结果

Table 4 Comparison results between our method and the basic method

数据集	模型	参数量	mAP/%	AP0.5/%	AP0.75/%	APs/%	APm/%	API/%	GFLOPs
TDD	YOLOv7-tiny(FPN)	6 025 812	61.1	89.5	66.8	<b>35.3</b>	68.5	74.2	<b>13.2</b>
	YOLOv7-tiny(LSRD-FPN)	6 681 716	<b>62.4</b>	<b>90.4</b>	<b>69.5</b>	34.8	<b>69.6</b>	<b>79.5</b>	15.1
Cmudsodd	YOLOv7-tiny(FPN)	6 025 812	60.4	94.3	64.5	<b>15.7</b>	52.8	64.3	<b>13.2</b>
	YOLOv7-tiny(LSRD-FPN)	6 681 716	<b>60.9</b>	<b>94.4</b>	<b>65.6</b>	15.6	<b>53.3</b>	<b>65.0</b>	15.1

注:加粗字体表示不同数据集各列最优结果。

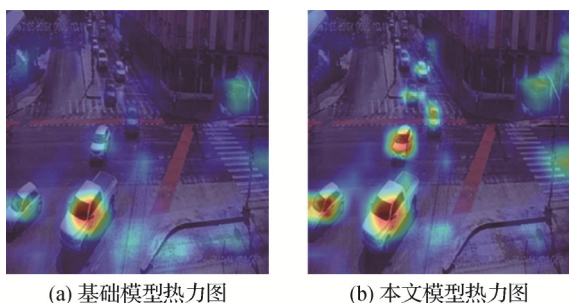


图8 基础模型与本文模型特征图可视化

Fig. 8 Visualization of feature maps for basic model and our model ((a) basic model heatmap; (b) our model heatmap)

图8(a)为YOLOv7-tiny的热力图,而图8(b)则展示了应用LSRD-FPN的YOLOv7-tiny的热力图。其中,红色区域表示该区域对最终预测结果贡献大,具有最高的关注度;黄色区域表示模型对该区域关注度较高;蓝色区域表示该区域的影响较小。两幅热力图均显示出对大尺度物体的较高关注度,这与表4中的结果相符,其体现在API指标明显高于APs和APm,这是因为大尺度物体通常具有丰富的语义信息和细节特征,目标检测算法更容易检测出这些大尺度物体。然而,无论是使用FPN还是LSRD-FPN,小尺度物体的检测效果依然较差。小尺度物

体检测往往过度依赖上下文信息,并且小尺度物体在图像中所占的像素较少,因此其特征不够显著,这就导致小尺度物体检测效果不佳。然而,值得注意的是,在应用LSRD-FPN的模型中,热力图显示出对中尺度物体具有较高的置信度。这表明本研究提出的改进方法对中尺度物体关注度更高,因此更有可能检测到中尺度物体,这在评价指标中体现为APm指标的提升。

图9和图10分别给出了在TDD和Cmudsodd数据集上的示例检测效果,分别展示了输入图像、利用基础方法YOLOv7-tiny的检测效果和基于LSRD-FPN的YOLOv7-tiny检测效果。

通过对上述两种检测结果进行比较,可观测到提出的方法在提升物体置信度、降低漏检率方面优于基础方法。图9第1行中基础方法car的置信度为0.66,本文方法car的置信度为0.82;图10第1行中基础方法helmet的置信度为0.70,本文方法helmet的置信度为0.72。此外,对比图9第2行中的两个检测结果,发现基础方法漏检了一个行人,而本文方法将该行人检测了出来(图9(c)中第2行的红色标记区域);对比图10第2行中的两个检测结果,发现基础方法漏检了一个钻井管,而本文方法检测出了



(a) 原始图像 (b) 基础方法 (c) 本文

图9 本文方法和基础方法在TDD数据集检测效果对比  
Fig. 9 Comparison of detection effects between our method and the basic method on TDD dataset((a) original images; (b) basic method; (c) ours)

该钻井管(图10(c)第2行的drill-pipe物体)。这表明与基础方法相比,本文方法具有更高的准确性、更强的鲁棒性以及更优的适应性。

#### 2.4.2 本文方法与参数量同等方法比较

本研究旨在分析基于LSRD-FPN的YOLOv7-tiny方法的有效性。为了确保对比实验的准确性和科学性,将基于LSRD-FPN的YOLOv7-tiny方法与其他相似参数量或处于同一数量级的方法进行对比,并保持训练参数一致。实验中选取了YOLOv5-s和YOLOv8-n两种对比方法,表5列出不同方法在TDD和Cmudsodd数据集上的性能结果,并对比分析了mAP、AP0.5、AP0.75、APs、APm、API和计算复杂度GFLOPs等多个性能指标。

对比结果显示,本文方法在TDD和Cmudsodd数据集分别取得了62.4% mAP和60.9% mAP。相对于YOLOv5-s方法,本文方法在所有指标提升显著,



(a) 原始图像 (b) 基础方法 (c) 本文

图10 本文方法和基础方法在Cmudsodd数据集检测效果对比  
Fig. 10 Comparison of detection effects between our method and the basic method on Cmudsodd dataset((a) original images; (b) basic method; (c) ours)

并且模型浮点运算数更少。尽管YOLOv8-n在TDD数据集上取得了与本文方法相近的检测精度,但其检测性能不稳定,尤其在Cmudsodd数据集上表现较差,这可能是由于YOLOv8-n方法的鲁棒性较低,难以应对Cmudsodd数据集中的极端场景。此外,本文方法在AP0.5和AP0.75指标上也展现了优异的表现。与其他先进的目标检测方法相比,本文方法在处理极端场景的目标以及多尺度目标方面具有明显优势。这得益于我们在模型设计中引入的创新模块和优化策略,使得模型能够更有效地提取目标特征,提高检测精度。

#### 2.4.3 消融实验

本研究对LSRD-FPN的组成部分进行了消融实验,分别对SRD的注意力机制和上采样方法以及LRD进行了详细的分析与比较。同时,为了验证本

表5 本文方法与同等方法比较结果

Table 5 Comparison results between our method and equivalent methods

数据集	模型	参数量	mAP/%	AP0.5/%	AP0.75/%	APs/%	APm/%	API/%	GFLOPs
TDD	YOLOv5-s	7 033 114	59.8	90.1	65.4	34.3	67.0	71.8	16.0
	YOLOv8-n	3 011 823	62.2	88.2	<b>70.0</b>	<b>35.3</b>	69.0	73.8	<b>8.2</b>
	本文	6 681 716	<b>62.4</b>	<b>90.4</b>	69.5	34.8	<b>69.6</b>	<b>79.5</b>	15.1
Cmudsodd	YOLOv5-s	7 033 114	58.8	93.3	62.0	12.0	51.0	62.5	16.0
	YOLOv8-n	3 011 823	56.5	92.3	58.3	<b>16.6</b>	52.2	59.0	<b>8.2</b>
	本文	6 681 716	<b>60.9</b>	<b>94.4</b>	<b>65.6</b>	15.6	<b>53.3</b>	<b>65.0</b>	15.1

注:加粗字体表示不同数据集各列最优结果。

研究提出的LSRD-FPN方法在不同场景中的有效性,本研究在TDD和Cmudsodd数据集上进行消融

实验。表6展现了在TDD和Cmudsodd数据集上不同模块的消融实验结果。

表6 本文方法消融实验结果

Table 6 The results of the ablation experiment using the proposed method

数据集	CARAFE	SimAM	LRD	mAP/%	AP0.5/%	AP0.75/%	APs/%	APm/%	API/%	GFLOPs
TDD	×	×	×	61.1	89.5	66.8	35.3	68.5	74.2	13.2
	√	×	×	61.4	89.5	67.7	33.1	68.7	75.5	13.2
	√	√	×	61.7	89.7	67.6	34.1	68.8	73.7	13.2
	√	√	√	<b>62.4</b>	<b>90.4</b>	<b>69.5</b>	34.8	<b>69.6</b>	<b>79.5</b>	15.1
Cmudsodd	×	×	×	60.4	94.3	64.5	15.7	52.8	64.3	13.2
	√	×	×	60.6	<b>94.4</b>	65.3	12.4	52.9	64.7	13.2
	√	√	×	60.6	94.3	65.0	15.1	53.0	64.5	13.2
	√	√	√	<b>60.9</b>	<b>94.4</b>	<b>65.6</b>	15.6	<b>53.3</b>	<b>65.0</b>	15.1

注:加粗字体表示不同数据集各列最优结果,“√”表示采用,“×”表示未采用。

如表6实验结果所示,每个数据集的第1行表示基准方法YOLOv7-tiny,第2、3行表示对SRD子模块的消融实验,第4行表示应用所有模块,即使用LSRD-FPN的检测结果。在TDD数据集,LSRD-FPN带来1.3% mAP的提升,并显著提高了AP0.5、AP0.75以及中大尺度物体的精度。从小尺度物体角度看,仅使用CARAFE上采样虽然会带来mAP提高,但是小尺度物体检测指标APs明显下降,分析发现这可能是CARAFE的内核权重无法准确捕捉到小尺度目标细微的特征变化。但是随着注意力机制和全局长程依赖机制LRD的引入,一方面,引入注意力机制聚焦图像中与目标相关的关键区域和特征;另一方面,来自特征提取阶段的特征进一步增强多尺度检测能力,包括APs在内的所有评价指标都在稳定提升。检测速度仍保持在160幅/s以上,这完全满足实时检测场景的任务需求。这一结果不仅保持了模型的复杂度和实时性,而且在此基础上提升了YOLOv7-tiny方法的检测精度。在TDD数据集上,组合使用SRD(CARAFE、SimAM)和LRD能明显提升mAP,说明这两个模块的结合对整体检测性能提升起到了关键作用。本文方法在Cmudsodd数据集上的性能增长趋势与TDD数据集上相似,但是Cmudsodd数据集性能增长幅度较小。这主要是因为Cmudsodd数据集的检测难度较大,该数据集的图像以深色和黑色为主,而矿灯的照射角度不同,导致物体出现不同程度的阴暗和光影,这对目标检测模

型带来巨大的挑战,即使Cmudsodd数据集检测难度大,但基于LSRD-FPN的方法依然在多个指标取得了提升。但是分析发现,在Cmudsodd数据集上小尺度物体检测性能偏低,且APs仅有15%左右,远小于其他尺度物体。通过对Cmudsodd数据集取样分析和观察得出结论,由于小尺度物体在图像中占的像素较少,并且数据集场景复杂,小尺度物体特征信息在深度卷积操作中更易丢失。在Cmudsodd数据集上,尽管不同组合之间的mAP差异不大,但使用本文方法仍然表现最佳,进一步证明了该组合的有效性。

#### 2.4.4 SRD局部短程依赖机制分析

SRD的核心在于引入注意力机制和改进上采样方法,旨在提升融合后特征图的特征表示能力,并解决相邻层特征融合时特征信息丢失问题。该部分的研究重点放在如何在保证SRD参数量的前提下,提升模型整体的检测精度。下面分别从上采样方法和注意力机制分析SRD的参数量和精度。

上采样方法是多尺度特征融合时的关键步骤,同时也是导致特征信息丢失的关键所在。首先,本研究对不同的上采样方式进行横向比较,并在TDD和Cmudsodd数据集上进行了实验验证,实验设置与表2保持一致。本研究分别对最近邻上采样、双线性插值和CARAFE上采样方法进行实验比较。其中,最近邻上采样是YOLO系列使用的上采样方法,双线性插值则在两个方向上进行线性插值,能够在

不引入额外参数的前提下生成更具表示能力的特征图。而CARAFE上采样方式是一种轻量化的可学习上采样方式,其具有两个参数,即 $k_{\text{encoder}}$ 表示内容编码器的卷积核大小。 $k_{\text{up}}$ 表示核预测模块内重组内核大小。在保证模型轻量化的前提下,本研究设置了3个不同的CARAFE参数组合,例如CARAFE13表示 $k_{\text{encoder}} = 1, k_{\text{up}} = 3$ ,并对这些实验结果进行比较,实验结果如表7所示。

实验结果显示,在TDD和Cmudsodd数据集上,基于可学习上采样方法CARAFE的模型相较于最近

邻上采样和双线性插值,在mAP评价指标上表现更加出色,并且在大尺度和中尺度物体上呈现更优性能。实验结果进一步表明,CARAFE参数 $k_{\text{encoder}}$ 和 $k_{\text{up}}$ 的增大会提高mAP值,但是由于卷积核尺寸增加,模型参数数量也会显著增加,这样带来的mAP收益却微乎其微,为了方便模块部署到不同方法中,这里的设计将以尽可能轻量化为主。因此,本研究选择了CARAFE参数 $k_{\text{encoder}} = 1, k_{\text{up}} = 3$ 作为SRD局部短程依赖机制的上采样改进部分,并用其代替YOLOv7-tiny中的最近邻上采样方法。

表7 不同上采样方法在两个数据集的性能

Table 7 Performance of different upsampling methods on two datasets

数据集	上采样方法	参数量	mAP/%	AP0.5/%	AP0.75/%	APs/%	APm/%	APl/%	GFLOPs
TDD	最近邻上采样	6 025 812	61.1	89.5	66.8	35.3	68.5	74.2	13.2
	双线性插值	6 025 812	61.2	89.5	67.1	34.1	68.5	74.9	13.2
	CARAFE13	6 043 108	61.4	89.5	67.7	33.1	68.7	75.5	13.2
	CARAFE33	6 079 972	61.6	89.5	67.7	34.2	68.7	73.3	13.3
	CARAFE35	6 153 956	61.5	89.5	67.1	34.5	68.7	74.7	13.5
Cmudsodd	最近邻上采样	6 025 812	60.4	94.3	64.5	15.7	52.8	64.3	13.2
	双线性插值	6 025 812	60.6	94.3	64.9	12.7	52.9	64.4	13.2
	CARAFE13	6 043 108	60.6	94.4	65.3	12.4	52.9	64.7	13.2
	CARAFE33	6 079 972	60.8	94.4	65.5	13.0	53.3	64.6	13.3
	CARAFE35	6 153 956	60.7	94.3	65.4	12.3	52.9	64.3	13.5

SRD局部短程依赖机制相对于基准方法YOLOv7-tiny,不仅修改了上采样方法,还创新性地引入了注意力机制。本研究对使用各种注意力机制的检测效果进行了详细的分析。在此过程中,本研究选取了6种重要的注意力机制方法,并对这些方法进行了全面的测试,具体结果见表8。

根据表8的结果显示,诸如MLCA、SimAM和EMA等注意力机制的应用,能够有效提升模型的准确性。然而,采用具有超参数的注意力机制会增加模型的参数量,并且对边缘设备造成一定的负担。SimAM注意力机制无需引入额外参数即可为特征图分配3D注意力权值,这对保持模型轻量化至关重要,并且在两个数据集上都获得了较高的mAP。将SRD应用于其他目标检测算法时,不引入额外参数的特性,有助于保持算法的计算复杂度。一般而言,按照经验,引入注意力机制通常会提高模型的性能。然而,实验结果并非总是如此。这是因为注意力机

制引入了额外的计算和参数,可能增加了模型的复杂性。如果模型超参数设置不当,模型的性能可能会下降。因此,在使用注意力机制时,需要谨慎选择超参数,并仔细评估其对模型性能的影响。

#### 2.4.5 LRD全局长程依赖机制分析

本研究对全局长程依赖LRD模块的有效性进行了实验,一种是建立从特征提取阶段到检测阶段的跨层连接,该设置来源于BiFPN(Tan等,2020)的连接方式,各层级特征相对独立并且没有特征信息交互,这种方式类似于ResNet残差方式,本节将其命名为Identity。另一种是基于多尺度特征融合方法的LRD机制,同样是建立从特征提取阶段到检测阶段的多尺度跨层连接,但是各层级特征通过融合、增强和分配等操作,语义信息和细节信息更加丰富,更有利于YOLOv7-tiny进行多尺度检测。本研究对这两种长距离依赖机制的性能进行了分析,实验结果如表9所示。

表8 不同注意力机制在两个数据集的模型性能

Table 8 Model performance of different attention mechanisms on two datasets

数据集	注意力机制	参数量	mAP/%	AP0.5/%	AP0.75/%	APs/%	APm/%	API/%	GFLOPs
TDD	无	6 043 108	61.4	89.5	67.7	33.1	68.7	75.5	13.2
	+Biformer(Zhu等,2023)	7 697 636	61.2	89.1	66.6	34.1	68.7	73.8	77.9
	+CBAM(Woo等,2018)	6 094 704	61.3	89.5	66.3	34.7	68.4	74.7	13.3
	+LSK(Li等,2023)	6 748 156	61.4	88.9	67.3	32.8	68.6	74.0	14.8
	+MLCA(Wan等,2023)	6 043 184	61.6	89.8	68.0	34.0	68.7	74.8	13.4
	+EMA(Ouyang等,2023)	6 107 684	61.8	89.5	67.6	33.8	69.0	72.2	14.3
	+SimAM(Yang等,2021)	6 043 108	61.7	89.7	67.6	34.1	68.8	73.7	13.2
Cmudsodd	无	6 043 108	60.6	94.4	65.3	12.4	52.9	64.7	13.2
	+Biformer(Zhu等,2023)	7 697 636	60.3	94.4	64.4	14.4	52.5	64.4	77.9
	+CBAM(Woo等,2018)	6 094 704	60.6	94.3	65.0	12.8	53.1	64.2	13.3
	+LSK(Li等,2023)	6 748 156	61.0	94.5	65.7	13.3	53.7	64.3	14.8
	+MLCA(Wan等,2023)	6 043 184	60.5	94.5	64.8	15.5	52.9	64.4	13.4
	+EMA(Ouyang等,2023)	6 107 684	60.5	94.3	65.0	13.3	53.0	64.4	14.3
	+SimAM(Yang等,2021)	6 043 108	60.6	94.3	65.0	15.1	53.0	64.5	13.2

表9 长程依赖机制在两个数据集的性能

Table 9 The performance of LRD on two datasets

数据集	长程依赖方式	参数量	mAP/%	AP0.5/%	AP0.75/%	APs/%	APm/%	API/%	GFLOPs
TDD	无	6 043 108	61.7	89.7	67.6	34.1	68.8	73.7	<b>13.2</b>
	+Identity	6 388 964	61.1	89.1	66.5	32.5	68.6	74.5	13.9
	+LRD	6 681 716	<b>62.4</b>	<b>90.4</b>	<b>69.5</b>	<b>34.8</b>	<b>69.6</b>	<b>79.5</b>	15.1
Cmudsodd	无	6 043 108	60.6	94.3	65.0	15.1	53.0	64.5	<b>13.2</b>
	+Identity	6 388 964	60.7	<b>94.5</b>	64.9	12.6	<b>53.3</b>	64.6	13.9
	+LRD	6 681 716	<b>60.9</b>	94.4	<b>65.6</b>	<b>15.6</b>	<b>53.3</b>	<b>65.0</b>	15.1

注:加粗字体表示不同数据集各列最优结果。

根据表9的结果显示,仅将主干网络对应尺度的特征附加到检测阶段并未带来精度升高,反而出现了大多数评价指标的下降。相反,将基于多尺度融合、富语义信息和富细节信息的特征添加到检测阶段能够取得更好的效果。不同尺度物体的检测精度同时得到了提升,这是因为多尺度特征被分配到了不同的检测头。

### 3 结论

本文从解决YOLOv7-tiny特征融合阶段存在的

两个问题出发,提出一种基于长程依赖的特征金字塔方法LSRD-FPN。应用LSRD-FPN的YOLOv7-tiny方法在交通检测和煤矿井下检测场景数据集上的实验表明,该方法展现了出色的性能,在两个数据集上分别提升1.3% mAP和0.5% mAP,并且LSRD-FPN的两个子模块在YOLOv7-tiny方法上也有不同程度的提升效果。与此同时,与近年来流行的实时目标检测算法进行比较,提出的方法也表现出了良好的性能,这在一定程度上解决了YOLOv7-tiny特征融合阶段出现的信息丢失和信息差异问题。

后续的研究方向将以实现检测速度与检测精度

的平衡为着手点。首先,探索和设计新的网络架构: CNN 和 Transformer 作为目标检测领域常用的基础架构,分别在速度和精度方面具有一定的优势,如何将 CNN 和 Transformer 架构的优点融合,最终实现速度与精度的同步提升,是未来重要的研究方向之一。其次,可以通过模型剪枝、蒸馏等技术,减小模型的计算复杂度和参数规模,从而在保持较高检测精度的同时提高检测速度。

## 参考文献 (References)

- Chen J C, Ma B X, Ji C, Zhang J, Feng Q C, Liu X and Li Y J. 2023. Apple inflorescence recognition of phenology stage in complex background based on improved YOLOv7. *Computers and Electronics in Agriculture*, 211: #108048 [DOI: 10.1016/j.compag.2023.108048]
- FSMVU. 2023. Traffic detection dataset [DB/OL]. Roboflow Universe [2024-09-02].  
<https://universe.roboflow.com/fsmvu/traffic-detection-zthfz>
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Li C Y, Hong D F and Zhang B. 2024. Deep unfolding network for hyperspectral anomaly detection. *National Remote Sensing Bulletin*, 28(1): 69-77 (李晨玉, 洪丹枫, 张兵. 2024. 深度展开网络的高光谱异常探测. *遥感学报*, 28(1): 69-77) [DOI: 10.11834/jrs.20233075]
- Li Y X, Hou Q B, Zheng Z H, Cheng M M, Yang J and Li X. 2023. Large selective kernel network for remote sensing object detection//*Proceedings of 2023 IEEE/CVF International Conference on Computer Vision*. Paris, France: IEEE: 16748-16759 [DOI: 10.1109/ICCV51070.2023.01540]
- Li X Y, Ye Z H, Wei S K, Chen Z, Chen X T, Tian Y H, Dang J W, Fu S J and Zhao Y. 2023. 3D object detection for autonomous driving from image: a survey——benchmarks, constraints and error analysis. *Journal of Image and Graphics*, 28(6): 1709-1740 (李熙莹, 叶芝桢, 韦世奎, 陈泽, 陈小彤, 田永鸿, 党建武, 付树军, 赵耀. 2023. 基于图像的自动驾驶3D目标检测综述——基准制约因素和误差分析. *中国图象图形学报*, 28(6): 1709-1740) [DOI: 10.11834/jig.230036]
- Lin T Y, Dollár P, Girshick R, He K M, Hariharan B and Belongie S. 2017. Feature pyramid networks for object detection//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA: IEEE: 936-944 [DOI: 10.1109/CVPR.2017.106]
- Liu S, Qi L, Qin H F, Shi J P and Jia J Y. 2018. Path aggregation network for instance segmentation//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 8759-8768 [DOI: 10.1109/CVPR.2018.00913]
- Ma L, Zhao L Y, Wang Z X, Zhang J and Chen G F. 2023. Detection and counting of small target apples under complicated environments by using improved YOLOv7-tiny. *Agronomy*, 13(5): #1419 [DOI: 10.3390/agronomy13051419]
- Ouyang D L, He S, Zhang G Z, Luo M Z, Guo H Y, Zhan J and Huang Z J. 2023. Efficient multi-scale attention module with cross-spatial learning//*Proceedings of 2023 IEEE International Conference on Acoustics, Speech and Signal Processing*. Rhodes Island, Greece: IEEE: 1-5 [DOI: 10.1109/ICASSP49357.2023.10096516]
- Pang J M, Chen K, Shi J P, Feng H J, Ouyang W L and Lin D H. 2019. Libra R-CNN: towards balanced learning for object detection//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE: 821-830 [DOI: 10.1109/CVPR.2019.00091]
- Qi X M and Dong X. 2023. Improved YOLOv7-tiny algorithm for steel surface defect detection. *Computer Engineering and Applications*, 59(12): 176-183 (齐向明, 董旭. 2023. 改进YOLOv7-tiny的钢材表面缺陷检测算法. *计算机工程与应用*, 59(12): 176-183) [DOI: 10.3778/j.issn.1002-8331.2302-0191]
- Qiu Y F and Xin H. 2024. Target detection algorithm based on global feature fusion in parallel dual path backbone. *Journal of Frontiers of Computer Science and Technology*, 18(12): 3247-3259 (邱云飞, 辛浩. 2024. 并行双路径主干下全局特征融合的目标检测算法. *计算机科学与探索*, 18(12): 3247-3259) [DOI: 10.3778/j.issn.1673-9418.2312050]
- Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 779-788 [DOI: 10.1109/CVPR.2016.91]
- Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D and Batra D. 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 618-626 [DOI: 10.1109/ICCV.2017.74]
- She F F, Hong Z Y, Zeng Z Q and Yu W H. 2023. Improved traffic sign detection model based on YOLOv7-tiny. *IEEE Access*, 11: 126555-126567 [DOI: 10.1109/ACCESS.2023.3331426]
- Shi C J, Zheng Y F, Ren B J, Kong F Y and Duan C Y. 2024. Single-domain generalized breast tumor detection in X-ray images. *Journal of Image and Graphics*, 29(3): 725-740 (史彩娟, 郑远帆, 任弼娟, 孔凡跃, 段昌钰. 2024. 单域泛化X-ray乳腺肿瘤检测. *中国图象图形学报*, 29(3): 725-740) [DOI: 10.11834/jig.230279]
- Sirisha M and Sudha S V. 2023. A review of deep learning-based object detection current and future perspectives//*Proceedings of the 3rd International Conference on Sustainable Expert Systems*. Singapore, Singapore: Springer: 937-951 [DOI: 10.1007/978-981-19-

- 7874-6\_69]
- Tan M X, Pang R M and Le Q V. 2020. EfficientDet: scalable and efficient object detection//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 10778-10787 [DOI: 10.1109/CVPR42600.2020.01079]
- Terven J, Córdova-Esparza D M and Romero-González J A. 2023. A comprehensive review of yolo architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, 5(4): 1680-1716 [DOI: 10.3390/make5040083]
- Wan D H, Lu R S, Shen S Y, Xu T, Lang X L and Ren Z J. 2023. Mixed local channel attention for object detection. *Engineering Applications of Artificial Intelligence*, 123: #106442 [DOI: 10.1016/j.engappai.2023.106442]
- Wang C Y, Bochkovskiy A and Liao H Y M. 2023. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 7464-7475 [DOI: 10.1109/CVPR52729.2023.00721]
- Wang J Q, Chen K, Xu R, Liu Z W, Loy C C and Lin D H. 2019. CARAFE: content-aware reassembly of features//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 3007-3016 [DOI: 10.1109/ICCV.2019.00310]
- Wang L, Bai J X, Wang P and Bai Y N. 2024. Research on pedestrian detection algorithm in industrial scene based on improved YOLOv7-tiny. *IEEJ Transactions on Electrical and Electronic Engineering*, 19(7): 1203-1215 [DOI: 10.1002/tee.24075]
- Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2\_1]
- Wu X, Hong D F and Chanussot J. 2023. UIU-Net: U-Net in U-Net for infrared small object detection. *IEEE Transactions on Image Processing*, 32: 364-376 [DOI: 10.1109/TIP.2022.3228497]
- Yang L S, Li M J, Hu J W and Wang D X. 2025. Strip steel surface defect detection algorithm based on improved YOLOv7-tiny. *Computer Engineering*, 51(1): 208-215 (阳丽莎, 李茂军, 胡建文, 王鼎湘. 2025. 基于改进YOLOv7-tiny的带钢表面缺陷检测算法. *计算机工程*, 51(1): 208-215) [DOI: 10.19678/j.issn.1000-3428.0068397]
- Yang L X, Zhang R Y, Li L D and Xie X H. 2021. SimAM: a simple, parameter-free attention module for convolutional neural networks//Proceedings of the 38th International Conference on Machine Learning. Virtual Event: ACM: 11863-11874
- Yang Z J, Feng H L, Ruan Y P and Weng X. 2023. Tea tree pest detection algorithm based on improved YOLOv7-tiny. *Agriculture*, 13(5): #1031 [DOI: 10.3390/agriculture13051031]
- Ye C, Lu T Y, Xiao Y H, Lu H and Yang Q H. 2022. Maritime surveillance videos based ships detection algorithms: a survey. *Journal of Image and Graphics*, 27(7): 2078-2093 (叶晨, 逯天洋, 肖漓灏, 陆海, 杨群慧. 2022. 海事监控视频舰船目标检测研究现状与展望. *中国图象图形学报*, 27(7): 2078-2093) [DOI: 10.11834/jig.200674]
- Zhang L H, Xiong N, Pan X H, Yue X D, Wu P and Guo C P. 2023. Improved object detection method utilizing YOLOv7-tiny for unmanned aerial vehicle photographic imagery. *Algorithms*, 16(11): #520 [DOI: 10.3390/a16110520]
- Zhou J L, Zhang Y Y and Wang J P. 2023. RDE-YOLOv7: an improved model based on YOLOv7 for better performance in detecting dragon fruits. *Agronomy*, 13(4): #1042 [DOI: 10.3390/agronomy13041042]
- Zhou W, Dong L H, Ye O, She X Y, Duan X Y, Peng Z K, Wang S Q, Zhao N N and Guo X P. 2024. Coal mine underground drilling site object detection dataset [DB/OL]. [2024-09-02]. <https://cstr.cn/31253.11.sciencedb.j00001.01020>
- Zhu L, Wang X J, Ke Z H, Zhang W and Lau R. 2023. BiFormer: vision transformer with bi-level routing attention//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 10323-10333 [DOI: 10.1109/CVPR52729.2023.00995]
- Zou Z X, Chen K Y, Shi Z W, Guo Y H and Ye J P. 2023. Object detection in 20 years: a survey. *Proceedings of the IEEE*, 111(3): 257-276 [DOI: 10.1109/JPROC.2023.3238524]

## 作者简介

孙中彬,男,预聘副教授,主要研究方向为不平衡数据挖掘、目标检测和工业视觉异常检测。

E-mail: zhongbin@cumt.edu.cn

胡帅,男,硕士研究生,主要研究方向为目标检测。

E-mail: hushuai0801@cumt.edu.cn

张帆,男,高级工程师,主要研究方向为云计算、大数据技术和应用软件工程。E-mail: tangliwen@inspur.com

周勇,男,教授,主要研究方向为机器学习、人工智能、数据科学与工程。E-mail: yzhou@cumt.edu.cn