

融合深度可分离卷积与高效注意力机制的轻量化 YOLOv11 车辆实时识别算法

宋梓棋¹ 杨丹^{1*} 王浩翔¹ 刘壹林¹

SONG Ziqi YANG Dan WANG Haoxiang LIU Yilin

摘要

针对智能交通系统中边缘计算设备对高实时性车辆检测算法的迫切需求,文章提出了一种通过对 YOLOv11 进行改进的轻量化车辆检测模型 YOLOv11-Light。首先,为降低计算复杂度,运用深度可分离卷积去重构主干网络,形成轻量化特征提取骨架;其次,为补偿轻量化导致的特征退化并提高模型在复杂场景中的鲁棒性,在特征融合颈部网络中嵌入一种轻量化坐标注意力模块,该模块可高效建模通道关系和长距离空间依赖;最后,构建了一个包含多种具有挑战性场景的综合车辆数据集 CityVehicles,并将其与公开数据集 UA-DETRAC 协同起来用于训练与评估。消融实验显示,所提出的两个模块协同工作,在精度与效率上实现了较好平衡。在对比实验当中, YOLOv11-Light 以 2.1×10^6 参数量和 4.3 MB 模型大小,在 UA-DETRAC 测试集上达到了 89.1% 的 mAP@0.5 和 35 帧/s (Jetson) 的检测速度,其综合性能明显优于 YOLOv5s、YOLOv8n 等主流轻量模型,为资源受限环境中的高精度实时车辆检测提供了有效的技术方案。

关键词

YOLOv11; 车辆检测; 轻量化网络

doi: 10.3969/j.issn.1672-9528.2026.03.013

0 引言

随着城市化进程不断加速以及机动车保有量在极大程度上的激增,智能交通系统(ITS)已经成为用来提升道路通行效率和保障交通安全的核心手段。车辆检测作为 ITS 视觉感知的基础,其性能会直接影响交通流量监控、违章行为识别、自动驾驶决策等高层次应用的效能^[1]。近年来,运用深度学习的目标检测算法,特别是以 YOLO 系列为代表的单阶段检测器,凭借其出色的实时性,在该领域得到了广泛应用。但是将高性能检测模型部署到实际的 ITS 环境,尤其是在算力、内存和功耗受限的边缘计算节点(如路口智能相机、车载计算单元),面临严峻挑战。一方面,为实现高精度,主流检测模型通常结构复杂、参数量极大(达数十兆甚至上百兆),难以满足边缘设备实时响应(> 30 帧/s)和低功耗需求;另一方面,真实交通场景复杂多变,车辆常呈现尺度差异大、相互遮挡严重、受光照和天气影响等特点,这对模型的鲁棒性提出了极高要求^[2]。所以,研究一种兼具高精度、高速度、强鲁棒性和小模型体积的车辆检测算法,具有重要

的理论价值和工程意义。

当前研究主要沿着两个方向开展:一是模型轻量化设计,二是特征表示增强。在轻量化方面, MobileNet 系列提出的深度可分离卷积操作,通过分解标准卷积操作,大幅降低了计算成本,成为轻量化 Backbone 设计的基石^[3]。 ShuffleNet 凭借通道混洗操作保证信息流通,常被用作 SSD 或 YOLO 主干的替代,以实现轻量化。在特征增强方面,注意力机制通过建模特征通道或空间位置的重要性,有效提高了模型对关键信息的捕捉能力。 SENet 的通道注意力以及 CBAM 的空间与通道混合注意力都证明了其有效性。然而,许多注意力模块会增加额外参数和计算量^[4],这与轻量化目标相悖。

YOLO 系列自身也在持续不断地进行进化。 YOLOv5 因其出色的工程友好性广受欢迎; YOLOv8 在精度和速度方面取得了更为良好的平衡;最新的 YOLOv11 (Ultralytics 发布)进一步优化了训练稳定性以及结构效率,其 nano 版本,即 YOLOv11n,是较为前沿的轻量化基准。然而,若直接将 YOLOv11n 应用于复杂车辆检测,其精度仍有提升空间,且计算量对部分超低功耗边缘设备来说仍显压力。

综上所述,现有工作尚未在最新的 YOLOv11 框架中,系统地将前沿的极致轻量化策略与高效注意力机制相结合,

1. 吉利学院智能科技学院 四川成都 641423

[基金项目] 吉利学院“三个校园”教学改革研究项目 (2025SGXYJG121)

专门针对复杂交通场景下的车辆检测任务进行优化^[5-6]。针对这一状况,本文提出一个全新的轻量化车辆检测模型,即YOLOv11-Light,其核心贡献如下:

(1) 结构的轻量化过程:设计了一种基于深度可分离卷积的改进型轻量化主干网络,即DS-Backbone,用作YOLOv11n原主干的替代,在确保特征提取能力的前提下,大幅压缩模型规模。

(2) 特征增强方面的轻量化:提出一种轻量化坐标注意力模块,即(lightweight coordinate attention, LCA),并且将它无缝地添加到颈部特征金字塔网络当中,这个模块凭借极低的开销增强了模型对车辆空间位置以及通道特征的判别能力。

(3) 数据以及验证环节:构建了一个覆盖多个不同场景、多个挑战的车辆数据集CityVehicles,并且鉴于此以及公开数据集开展了全面的消融实验与对比分析工作,来验证所提模型在精度、速度以及模型大小这些方面的综合优势,给边缘端的部署提供了可靠的方案。

1 改进的YOLOv11车辆识别算法

1.1 YOLOv11-nano 基线网络概述

YOLOv11-nano (YOLOv11n) 是YOLOv11系列中体积最小、速度最快的版本,整体架构由主干网络(Backbone)、颈部网络(Neck)和检测头(Head)三部分组成,如图1所示。

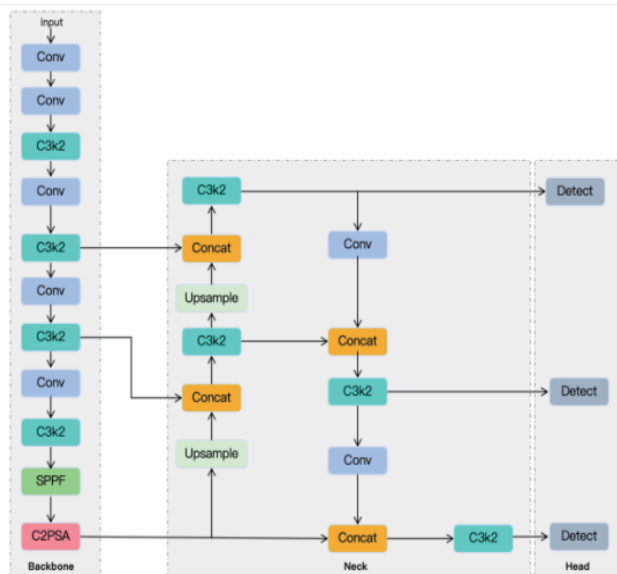


图1 YOLOv11n 模型架构图

从图1中可以看出,主干网络主要运用标准卷积以及跨阶段部分网络模块进行堆叠,从输入图像中提取多层次的特征图。颈部网络选用了路径聚合网络与特征金字塔网络相结合的结构,借助自上而下和自下而上的路径,融合来自主干

网络浅层的高分辨率且强位置信息特征以及深层的高语义信息特征,增强对不同尺度目标的检测能力。检测头使用解耦头结构,分别负责分类以及定位的工作,并且运用动态标签分配策略来进行训练。YOLOv11n作为基线,拥有良好的速度以及精度起点,然而它的主干网络存在大量 3×3 卷积,空间可进一步轻量化。

1.2 深度可分离卷积轻量化主干设计

由于标准卷积核同时对输入特征图的所有通道进行空间滤波并线性组合,其计算成本为 $C_{in}\times C_{out}\times K\times K$ (C 为通道数, K 为卷积核大小),对于特征提取主干,大量堆叠的卷积层是参数和计算量的主要来源。

深度可分离卷积(DSConv)由深度卷积(depthwise convolution, DWConv)和逐点卷积(pointwise convolution, PWConv)两步组成。DWConv对每个输入通道单独应用一个空间卷积核,PWConv(即 1×1 卷积)负责跨通道的信息组合。总计算量约为 $C_{in}\times K\times K+C_{in}\times C_{out}$ 。当 C_{out} 较大时,DSConv的计算量可降至标准卷积的 $1/C_{out}+1/K^2$ 左右,理论加速比显著。

本文构建了深度可分离卷积块(DSC Block)作为基本单元,如图2所示,该块由DWConv、PWConv、批量归一化(BatchNorm)和SiLU激活函数顺序构成。本研究用DSC Block替换YOLOv11n主干网络中所有非下采样的标准卷积块(即用于特征变换的卷积层),对于负责空间下采样的步长为2的卷积层,保留了标准卷积以保证下采样过程中的信息融合能力,改进后的主干网络称为DS-Backbone。

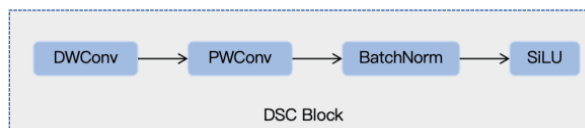


图2 深度可分离卷积块模型图

1.3 轻量化坐标注意力特征融合模块

轻量化主干网络虽然降低了计算负担,但也可能导致所提取特征的区分度下降,特别是对于有遮挡、小目标等困难样本。注意力机制是增强特征表达的有效工具,但SENet仅仅是对通道维度予以关注,CBAM等混合注意力引入了较多的参数。坐标注意力(coordinate attention, CA)通过将空间维度分解为两个一维的全局池化操作,一方面可以捕获跨通道的信息,另一方面也能够编码精确的位置信息,而且具有计算高效的特性。

本文设计了一个轻量化坐标注意力模块(LCA Module),其结构如图3所示。

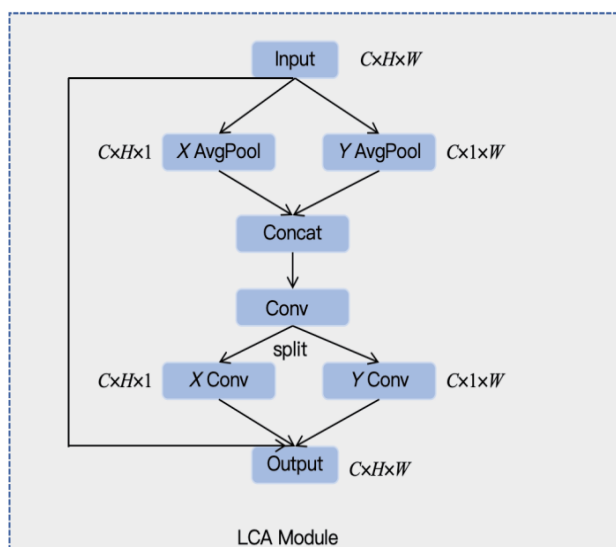


图3 轻量化坐标注意力模块图

给定输入特征图 $F \in \mathbb{R}^{C \times H \times W}$ ，模块首先沿水平 X 和垂直 Y 方向分别进行自适应平均池化，得到两个方向感知的特征图 $F_h \in \mathbb{R}^{C \times H \times 1}$ 和 $F_w \in \mathbb{R}^{C \times 1 \times W}$ 。接着，将二者拼接后通过一个共享的 1×1 卷积进行降维和特征交互，生成中间特征。随后，使用两个分离的 1×1 卷积分别生成水平方向和垂直方向的注意力权重图 G_h 和 G_w 。最后，将这两个权重图与原始输入特征图逐元素相乘，得到加权后的输出特征。过程可形式化表示为：

$$F_h = \text{AdaptiveAvgPool2D}_{(H,1)}(F) \quad (1)$$

$$F_w = \text{AdaptiveAvgPool2D}_{(1,W)}(F) \quad (2)$$

$$F_{\text{cat}} = \text{Concat}(F_h, F_w) \quad (3)$$

$$F_{\text{mid}} = \delta(\text{Conv1D}(F_{\text{cat}})) \quad (4)$$

$$G_h = \sigma(\text{Conv1D}_h(F_{\text{mid}})) \quad (5)$$

$$G_w = \sigma(\text{Conv1D}_w(F_{\text{mid}})) \quad (6)$$

$$F' = F \times G_h \times G_w \quad (7)$$

式中： δ 和 σ 分别为非线性激活函数和 Sigmoid 函数。

本文将 LCA 模块插入到颈部网络 PANet 结构的关键融合节点处，使网络在融合不同尺度特征时，能够自适应地强调与车辆目标相关的通道和空间位置。

1.4 YOLOv11-Light 整体网络结构

将上述改进进行整合，得到完整的 YOLOv11-Light 模型结构如图 4 所示。核心即 DS-Backbone 以及嵌入了 LCA 模块的增强型颈部网络，即 LCA-Neck。这种设计确保了模型获得轻量化优势，且借助注意力机制维持强大的特征判别能力，特别适合复杂交通场景中的车辆检测工作。

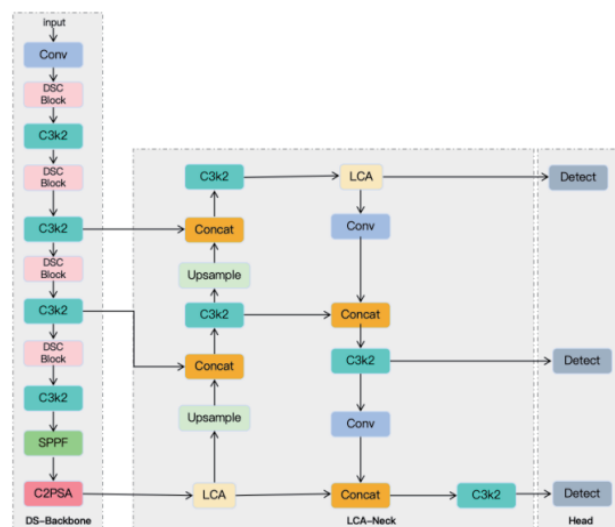


图4 YOLOv11-Light 整体网络结构图

2 实验设计

2.1 数据集

(1) UA-DETRAC 数据集：是一个大规模的公开车辆检测数据集，拥有 10 h 的视频，大概 14 万帧，对 121 万余辆车开展了标注工作。此数据集的场景囊括城市道路、高速公路等，含有不同的天气、光照条件以及交通密度。在本文中，按官方划分使用训练集（60%）、验证集（20%）以及测试集（20%）。

(2) CityVehicles 数据集：为了能够更为全面地对模型在多样化挑战下的性能进行评估，本文从公开的行车记录仪视频以及交通监控视频当中进行截取并且构建了这个数据集。拥有 5 200 张高分辨率图像（1 920 px×1 080 px），涵盖典型的挑战场景：

(1) 尺度变化：近处大车与远处小车；(2) 遮挡：车辆之间存在部分或严重遮挡；(3) 光照变化：白天强光、黄昏、夜间弱光及隧道入口；(4) 天气变化：晴天、雨天、雾天等天气情况；(5) 复杂背景：密集城市街区、立交桥、停车场等。

所有图像使用 LabelImg 工具进行精细标注，边界框仅包含车辆主体。按 7:1.5:1.5 的比例随机划分为训练集（3 640 张）、验证集（780 张）和测试集（780 张）。本数据集将与 UA-DETRAC 训练集混合使用，以提高模型泛化能力。

2.2 实验设置与评估指标

实验所采用的环境是 Ubuntu 20.04 以及 PyTorch 1.12.1 和 CUDA 11.6，训练以及测试都是在单张 NVIDIA Tesla V100 GPU 中开展，模拟边缘环境的速度测试则是在 NVIDIA Jetson Xavier NX 这一设备上。训练时输入图像统一调整为 640 px×640 px，运用了随机翻转、Mosaic、MixUp 等方面的

数据增强, 优化器为SGD, 动量为0.937, 权重衰减为0.000 5, 初始学习率为0.01, 可以采用余弦退火调度的方式来进行, Batch size 为32, 在混合数据集上进行300个epoch的训练, 所有的对比模型都运用相同设置来重新训练, 以确保公平。

本文采用以下评估指标:

(1) 精度指标: 平均精度均值 (mAP), 包括 IoU 阈值为0.5的 mAP@0.5 和 IoU 阈值从0.5到0.95间隔0.05的平均值 mAP@0.5:0.95。

(2) 效率指标: 模型参数量 (Params) 和在 Jetson Xavier NX 上的帧率 (frames per second, FPS)。

(3) 模型大小: 模型文件所占存储空间 (megabytes, MB)。

3 实验结果与分析

3.1 消融实验

为验证 DS-Backbone 和 LCA Module 各自的有效性 & 协同作用, 在 CityVehicles 测试集上进行了系统的消融实验, 结果如表1所示。

表1 在 CityVehicles 测试集上的消融实验结果

模型	DS- Backbone	LCA- Neck	mAP@0.5 /%	mAP@0.5:0.95 /%	Params /10 ⁶
YOLOv11n			87.9	65.2	2.6
模型1	√		86.2	63.5	2.0
模型2		√	88.7	66.1	2.7
Ours	√	√	87.5	64.9	2.1

表1中数据表明, 和基准模型 YOLOv11n 相较而言, 模型1仅用了 DS-Backbone 参数降低0.6, 这证实了深度可分离卷积在模型压缩以及加速方面所具备的极大优势, 然而 mAP 指标下降1.7%, 此为轻量化所带来的典型精度与速度之间的权衡; 模型2仅采用了 LCA-Neck, mAP@0.5 提升了0.8%, 显示出注意力机制有效地增强了特征, 提高了检测精度, 同时也带来参数以及计算量增加。Ours 为 YOLOv11-Light, 将两个模块进行结合之后, 和基线相比较, YOLOv11-Light 参数量减少了0.5, 精度方面相差不大。这充分证明了 DS-Backbone 与 LCA-Neck 的协同有效性: LCA 模块借助极小的计算开销, 有效补偿了轻量化主干所带来的特征信息损失, 最终实现了精度与效率近乎最优的平衡。

3.2 对比实验

把 YOLOv11-Light 和当下主流的轻量化目标检测模型, 在 UA-DETRAC 测试集上开展全面对比的工作, 其对比的相关结果如表2所示。并且所有的模型均在混合数据集之上进行训练, 在 UA-DETRAC 测试集之上进行评估工作, 保证对比的公平性。

表2 与主流轻量化模型的对比结果

模型	mAP@0.5 /%	mAP@0.5:0.95 /%	Params /10 ⁶	FPS /(帧·s ⁻¹)	模型大小 /MB
YOLOv5s	88.5	66.8	7.2	22	14.4
YOLOv8n	88.8	66.5	3.2	28	6.4
YOLOv10n	89.0	67.0	2.7	30	5.5
YOLOv11n	89.2	67.3	2.6	27	5.3
YOLOv11-Light (Ours)	89.1	67.1	2.1	35	4.3
MobileNetV3-SSD	84.7	60.1	5.4	38	21.2

从表2中可以看出, 精度方面 YOLOv11-Light 的 mAP@0.5 达到了89.1%, 和最强的基线模型 YOLOv11n (89.2%) 差不多是持平的状态, 并且显著高于其他经典轻量模型, 表明本文模型在把极致轻量化取得的同时, 成功实现并保持了 YOLOv11 系列的高精度水准。效率方面, YOLOv11-Light 在参数量 (2.1×10⁶) 以及模型大小 (4.3 MB) 上均为所有进行对比的 YOLO 系列模型当中最小的, 甚至比更早的 YOLOv8n 和 YOLOv10n 还要小, 其4.3 MB 的模型尺寸是极为适宜存储受限的边缘设备的。从速度方面来说, 在代表边缘计算平台的 Jetson Xavier NX 上, 帧率达到了35帧/s, 可以满足绝大多数实时视频流 (30帧/s) 的处理工作需求。综上所述, 相较于前沿轻量模型 YOLOv10n, 本文模型在精度相当的情况之下, 参数量得到减少且速度有所提升, 综合优势较为明显。即使与为移动端进行设计的 MobileNetV3-SSD 相比, YOLOv11-Light 在精度方面也有压倒性的优势, 同时还保持着极具竞争力的速度以及更小的模型体积。

4 结论

本文在边缘计算环境之下针对车辆实时检测这种应用需求, 提出轻量化改进模型 YOLOv11-Light。通过深度可分离卷积对主干网络加以重构, 极大程度上降低了模型这样那样的参数量以及计算复杂度; 把轻量化坐标注意力模块引入到特征融合颈部当中, 有效地增强了模型对于车辆目标关键特征的捕捉能力, 对轻量化所带来的精度损失进行了弥补。在自己所建立的 CityVehicles 数据集以及公开的 UA-DETRAC 数据集里进行的实验显示, 所提出的方法在模型体积、计算效率以及检测精度三个方面实现了较好平衡。YOLOv11-Light 凭借仅仅 2.1×10⁶ 的参数量以及 4.3 MB 的模型大小, 实现了同前沿轻量模型差不多甚至更优秀的检测精度, 并且在 GPU 以及边缘设备之上都收获了处于领先状态的速度, 拥有着良好的实际部署前景。

基于 YOLOv5s 的轻量级海洋塑料垃圾检测

刘炳如¹ 万丽莉^{1*}

LIU Bingru WAN Lili

摘要

针对海洋塑料垃圾检测模型复杂度高且实时性差的问题,文章基于 YOLOv5s 提出了一种改进的轻量化目标检测模型 YOLOv5s-MSL。首先,采用 MobileNetV4 替换原 CSPDarknet53 主干网络降低模型复杂度;然后,采用 SDI 多层次特征融合模块替换 Concat 模块,通过整合编码器生成的层级特征图来增强图像中的语义信息和细节信息,从而增强模型的精确度;最后,结合轻量化模块 CSPPC 替换颈部 C3 模块,实现模型在轻量化和检测准确性上的双重提升。仿真实验表明,在海洋塑料检测数据集上,YOLOv5s-MSL 模型的准确率达到 96.8%。模型大小仅为 5.0 MB,参数数量和运算浮点数相较于原 YOLOv5s 分别减少 56.55% 和 74.68%。

关键词

海洋塑料垃圾;轻量化;目标检测;CSPPC

doi: 10.3969/j.issn.1672-9528.2026.03.014

0 引言

海洋污染,尤其是海底垃圾问题,已成为全球性环境挑战。根据联合国环境规划署的数据,全球每年有大约 800 万吨塑料进入海洋^[1]。塑料垃圾在 5 年内就能到达海洋最深处。海洋环境复杂多变导致海洋塑料垃圾人工清理困难,有效监测和管理海洋塑料垃圾变得尤为重要^[2-3]。周冬梅等^[4]提出了 WG-YOLO 算法,其 mAP50 达到了 68.1%,mAP50:95 达到 48.7%。尽管该算法在水下目标检测领域展现出一定的有

效性,但仍存在检测精度较低且模型不够轻量化的情况。针对上述问题,本文提出了一种基于 YOLOv5s 的轻量级海洋垃圾检测模型 YOLOv5s-MSL,该模型以复杂海洋环境中的塑料垃圾数据集训练而成,在保持高精度的同时实现轻量化,为海洋垃圾实时监测提供了可行方案。

1 YOLOv5s-MSL 模型

1.1 YOLOv5s-MSL 网络结构

YOLOv5s-MSL 网络结构如图 1 所示,主干网络为 MobileNetV4,其采用多尺度特征金字塔结构,包含五个层次的特征图(P1~P5),分别对应不同下采样率的特征表示。其中,

1. 武汉轻工大学电气与电子工程学院 湖北武汉 430048

[基金项目] 武汉轻工大学科研项目资助(2025Y46)

参考文献:

- [1] 杜雨亭,何嘉.基于 YOLOv8 的雾天环境下车辆行人检测算法研究[J].软件导刊,2025,24(12):18-24.
- [2] 伍锡如,郝家琦,赵一波,等.复杂天气条件下基于 YOLO-CGT 的自动驾驶车辆检测[J].光学精密工程,2025,33(19):3135-3149.
- [3] 崔明义,冯治国,代建琴,等.基于深度可分离卷积和残差注意力模块的车道线检测方法[J].激光杂志,2024,45(4):81-87.
- [4] 马鸽,李洪伟,严梓维,等.基于多注意力的改进 YOLOv5s 小目标检测算法[J].工程科学学报,2024,46(9):1647-1658.
- [5] 牟毅,黄海松,李宜汀,等.改进 YOLOv11s 的无人机图

像小目标检测模型[J].电光与控制,2026,33(1):51-57.

- [6] 许牛琦,王金,付迪,等.基于 YOLOv11 改进的驾驶员视角下前方车头与车尾检测[J].信息技术与信息化,2025(11):106-109.

【作者简介】

宋梓棋(2006—),女,四川德阳人,本科在读,研究方向:深度学习、目标检测。

杨丹(1995—),通信作者(email:1253026013@qq.com),女,四川成都人,硕士研究生,教师、工程师,研究方向:深度学习、计算机视觉。

(收稿日期:2026-01-05 修回日期:2026-03-11)