

基于视觉/激光雷达轻量化目标三维检测与定位方法

蒋承伶¹, 黄郑¹, 沈超², 马洲俊¹, 张明晖²

(1.国网江苏省电力有限公司 南京 210024; 2.国网江苏省电力有限公司无锡供电分公司 无锡 214000)

摘要: 目标三维检测与定位是无人机精确巡检与避障的重要保障。传统方法通常将目标检测、定位通过同一网络实现, 存在网络结构复杂、计算量大、实时部署困难等问题。针对于此, 提出了一种基于网络解耦的轻量化目标三维检测与定位方法: 首先, 提出了一种基于高效特征提取与增强注意力机制的轻量化视觉二维目标检测网络, 降低网络整体参数数量的同时提高对多种目标的泛化检测能力; 其次, 提出了一种具有跨层连接与辅助损失的视觉/激光雷达融合深度补全网络, 实现高精度稠密深度图估计; 最后, 设计了检测目标像素/深度对齐方案, 通过坐标变换实现目标三维空间位置的精确计算。试验结果表明, 相对于 YOLOv9 目标检测算法, 本文算法目标检测精度提高了 5.1%; 相对于 AVOD 目标定位算法, 本文方法目标三维定位精度提高了 45%。同时, 本文算法在无人机端侧运行频率达到 36 帧/秒, 相对于 AVOD 提升了 76%, 在无人机目标检测领域具有较好应用参考价值。

关键词: 目标检测; 激光雷达; 机器视觉; 深度学习; 目标定位

中图分类号: TP273 **文献标志码:** A

A Lightweight 3D Object Detection and Localization Method Based on Visual/LiDAR Fusion

JIANG Chengling¹, HUANG Zheng¹, SHEN Chao², MA Zhoujun¹, ZHANG Minghui²

(1.State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210024, China; 2.Wuxi Power Supply Company of State Grid Jiangsu Electric Power Co., Ltd., Wuxi 214000, China)

Abstract: Accurate 3D object detection and localization are critical for UAV-based inspection and obstacle avoidance. Traditional methods often integrate detection and localization within a unified network, resulting in complex architectures, high computational costs, and challenges in real-time deployment. To address these issues, we propose a lightweight 3D object detection and localization framework based on network decoupling. First, a lightweight 2D object detection network is designed, incorporating efficient feature extraction and an enhanced attention mechanism, which significantly reduces the number of parameters while improving generalization across diverse target types. Second, we introduce a visual/LiDAR fusion-based depth completion network with cross-layer connections and auxiliary loss functions to achieve high-precision dense depth map estimation. Finally, a pixel/depth alignment scheme is developed to accurately compute 3D spatial positions of detected objects via coordinate transformation. Experimental results demonstrate that, compared to the YOLOv9 detection algorithm, the proposed method improves object detection accuracy by 14%, and enhances 3D localization accuracy by 45% over the AVOD framework. Moreover, the proposed approach achieves a processing rate of 36 frames per second on UAV edge devices, representing a 90% increase over AVOD, highlighting its practical value for real-time UAV-based object detection applications.

Keywords: object detection; LiDAR; machine vision; deep learning; object positioning

基金项目: 国网江苏省电力有限公司科技项目 (J2023014)。

收稿日期: 2025-xx-xx; 修订日期: 2025-xx-xx

引言

近年来,无人机(Unmanned aerial vehicle, UAV)巡检在社会民生中得到了广泛应用,相比传统人工巡检,其在提升安全性、降低成本、快速性方面具有显著优势^{[1][2]}。无人机执行巡检任务时,高效可靠避障是保障其安全飞行的重要基础^{[3][3]}。障碍物的识别与定位是无人机避障的关键技术,需要对障碍物目标进行准确、快速定位。

视觉与激光雷达是无人机常用的感知传感器,在视觉目标识别定位方面,Yolov7^[4]、CFN^[5]、Yolov9^[7]等基于深度学习的算法通过卷积神经网络提取图像特征,结合核相关滤波^[8]等跟踪算法实现动态障碍物的实时检测与二维定位。然而,视觉系统受限于单目仅感知二维信息,难以获取精确的深度信息,导致对障碍物尺寸、距离的估计存在误差,难以满足精细化巡检任务需求。

在激光雷达目标识别定位方面,主流方法常将原始点云通过体素化或投影转换为结构化数据输入,结合三维边界框回归算法完成障碍物的分类与定位^[9]。Yao 提出了一种几何分布与三维形状特征匹配及并行扫描检测跟踪方法,实现了复杂干扰场景下多目标跟踪^[10]。Li 提出了通过分析并改进点基方法中提案尺寸特征忽略问题,提升了现有三维点云检测器的性能^[11]。这类方法能够输出障碍物的三维几何参数和空间坐标。但受限于激光雷达的物理特性,点云数据在远距离区域呈现显著稀疏化特征,面对小体积、低反射率材质障碍物时,有效点云数量可能不足。此外,激光雷达无法感知物体的颜色、纹理等丰富的外观信息,限制了障碍物目标检测可靠性。

相对于采用单一视觉或激光雷达传感器,视觉提供高分辨率纹理信息,激光雷达补充深度数据,通过传感器互补可实现稠密三维检测。Liu 等人提出了 BEVFusion 框架将多模态特征投影至鸟瞰视图统一融合,结合目标运动模型优化定位精度^[12]。Lin 等人提出了包含双流雷达特征提取网络和跨注意力多层融合模块的 RCBEVDet 识别网络,提高了目标检测效率^[13]。Tao 等人提出了一种 Swin Transformer 与 UNet 的跳跃连接架构及 LI/IL 多模态融合模块增强的 AVOD 检测方法,提高了低质量图像条件下的三维目标检测精度^[14]。然而,激光雷达/视觉融合方法面临以下难点:一方面,激光雷达/视觉信息处理计算复杂度高。同时处理图像和点云数据,并进行跨模态特征对齐与融合,常需复杂的网络结构和大量计算资源,机载平台受限于功耗体积限制,难以满足实时性需求。另一方面,激光雷达/视觉神经网络训练成本高。多模态数据对齐困难,构建精确标定数据集需复杂硬件同步与大量人工标注,显著增加成本。

针对上述问题,部分研究者通过结合二维目标检测技术与激光雷达的深度信息实现障碍物的三维定位,该方法计算简单能够降低运算开销。为解决点云的稀疏性问题,该类方法通过融合图像数据和稀疏的点云数据获取稠密深度图。Wang 等人提出了一种基于潜在空间学习的深度补全网络 Sparse2Dense,提升了在稀疏特征下的深度估计性能^[17]。Wouter 等人提出了基于 RGB 引导与不确定性融合的稀疏 LiDAR 补全方法,实现了稀疏噪声条件下深度预测的鲁棒性提升^[18]。但上述方法计算效率较低,且易在小尺寸障碍物或细边缘产生漏检,无法满足无人机巡检需求。

针对于此,本文提出了一种基于视觉/激光雷达的轻量化检测与定位方法。将目标检测与位置估计解耦,通过深度可分离卷积与混合降采样模块,在降低计算负载的同时保留多模态特征细节信息互补。主要创新点如下:

1) 设计了一种适用于无人机端侧的障碍物目标检测与三维坐标解耦解算方案,通过改进 Yolov9 二维检测和高精度深度补全网络信息,基于传感器参数精确获取检测目标的三维空间位置。

2) 提出了一种基于强化注意力机制的轻量化 Yolov9 目标检测网络,设计了深度可分离卷积高效聚合网络 (Depthwise separable convolution-efficient layer aggregation network, DSC-ELAN) 与增强注意力模块 (Enhanced convolutional block attention module, E-CBAM)。在 ELAN 模块的基础上引入深度可分离卷积,设计了强化注意力机制模块与混合下采样 (Mixed-downsampling, MDS),提高了网络整体的特征提取能力与目标检测精度下,减少网络主干特征提取模块的参数数量,有效降低了无人机平台的计算负担。

3) 构建了具有高效特征提取模块与跨层连接的多尺度深度补全网络,通过融合图像与点云数据输出稠密深度图,同时引入辅助深度损失,提高深度补全精度。

1 视觉/激光雷达轻量化融合的目标定位算法

本文轻量化目标定位算法整体工作流程分为四个部分:多模态融合模块、目标检测模块、深度补全模块以及坐标解算模块。

- 在多模态融合模块中,摄像头与激光雷达之间的外参以及摄像头的内参通过联合标定套件进行标定获取。基于此将雷达点云投影到图像平面并对对应区域进行裁剪对齐,获取融合特征图。随后,将融合特征图分别送入目标检测模块与深度补全模块。
- 在目标检测模块中,通过 YOLO 架构的目标检测网络对输入数据进行特征提取融合与解码,获取障碍物目标的二维检测信息。
- 在深度补全模块中,通过编码器-解码器网络对稀疏点云数据进行补全获取稠密深度信息。
- 在坐标解算模块中,二维目标检测结果与稠密深度图对齐,获取障碍物目标的尺度数据,基于传感器参数完成反向坐标解算,得到目标的三维空间位置。

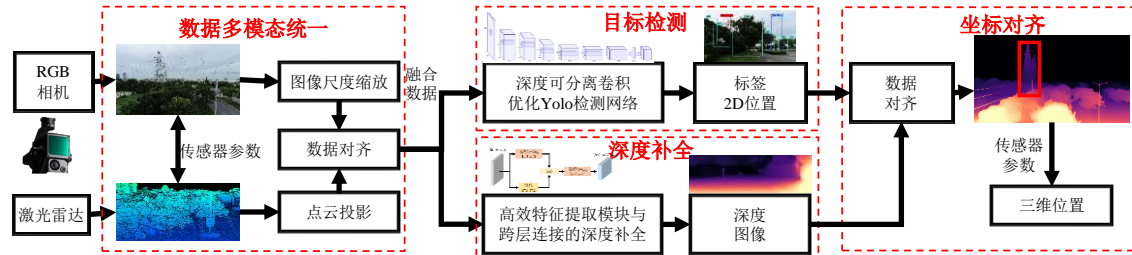


图 1 视觉/激光雷达轻量化融合的目标定位算法示意图

Fig.1 Schematic diagram of the lightweight vision/LiDAR fusion-based target localization algorithm

1.1 深度可分离卷积目标检测网络优化设计

对于摄像头和激光雷达,其输出数据的形状分别为 $(H,W,3)$ 和 $(N,3)$,其中 H 和 W 表示图像的高度和宽度, N 表示激光雷达获取点的数量。基于相机内参 K 和外参 M 。对点云进行投影变换和坐标计算,得到形状为 $(H,W,1)$ 的稀疏深度图,与图像堆叠得到形状为 $(H,W,4)$ 的融合特征图并作为目标检测网络和深度补全网络的输入。

在实际无人机巡检中,会存在多种不同尺寸、形状的障碍物,对目标检测网络的性能提出了更高的要求。考虑到无人机平台有限的计算资源,以及需要运行深度补全网络和三维坐

标解算算法，本文在保证目标检测网络的特征提取能力的同时，实现了轻量化设计，并将输入图像大小设置为 416×416。目标检测网络的整体框架如图 2 所示。

模型以 YOLO 为基础，该系列算法广泛应用于嵌入式平台，在检测速度和精度之间有良好的平衡，具有较强的工程实用性，尤其适用于无人机端的目标检测任务。整体网络分为三部分：特征提取网络、特征融合网络、预测网络。在特征提取阶段，采用了 ELAN 模块，通过 4 条并行路径从输入特征图中提取特征并进行整合，提供丰富的特征信息。但其中两条路径产生的特征较为相似，并且整个模块包含大量 3×3 卷积，导致网络参数数量和计算负担显著增加。因此，基于深度可分离卷积设计了更轻量的 DSC-ELAN 模块代替原有的高效层聚合网络（Efficient layer aggregation networks, ELAN），并对特征提取路径进行优化。在特征融合阶段，遵循自底向上和自顶向下的双向融合原则，基于注意机制优化并嵌入了 E-CBAM 模块，该模块在融合更多梯度信息的同时，提供了空间和通道注意力信息。最后，头部网络在三个不同的尺度上输出预测的边界框坐标、类别和置信度。

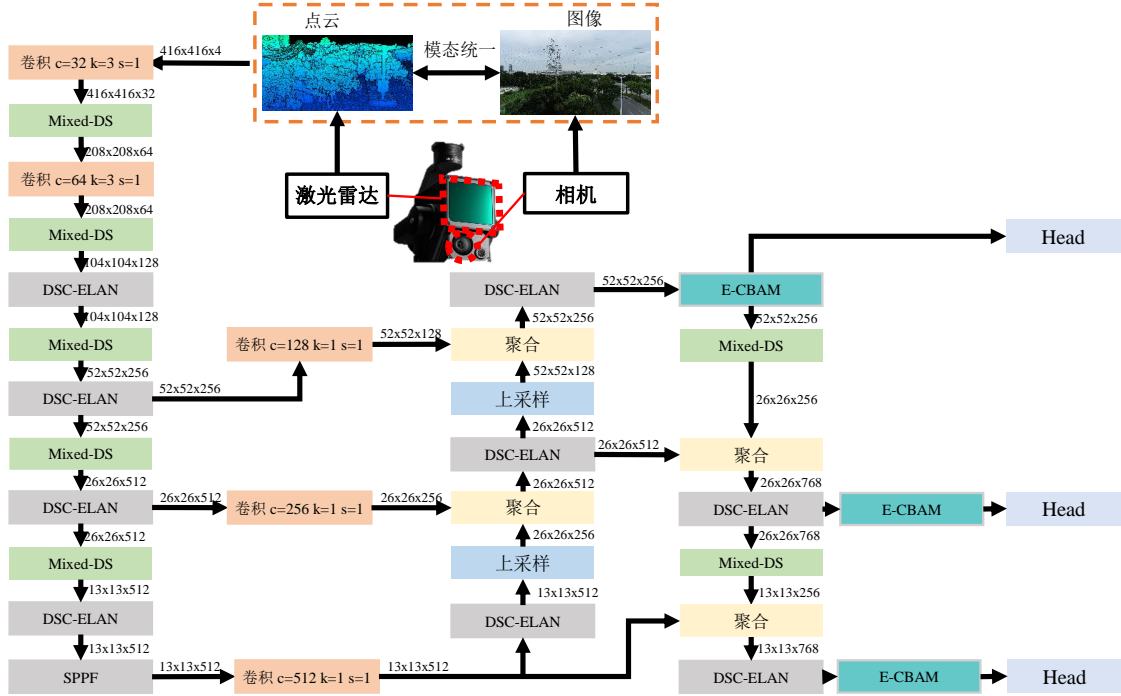


图 2 深度可分离卷积目标检测网络架构

Fig.2 Architecture of the depthwise separable convolution-based object detection network

在 DSC-ELAN 模块中，深度可分离卷积是一种高效的卷积计算方法，其将标准卷积分解为深度卷积和点卷积。对于 3×3 卷积核，深度卷积在每个通道上独立执行 3×3 卷积操作，而不会在通道之间混合信息，而点卷积使用 1×1 卷积操作来混合通道信息并实现特征融合。根据参数计算公式，若网络输入通道数为 C_{in} ，输出通道数为 C_{out} ，输入特征图宽度和高度为 W 和 H ，卷积核大小为 K ，则该卷积操作的计算量为：

$$O_{conv} = H \times W \times C_{in} \times C_{out} \times K \times K \quad (1)$$

对于深度可分离卷积，深度卷积部分的计算量为 $H \times W \times C_{in} \times K \times K$ ，点卷积部分的计算量为 $H \times W \times C_{in} \times C_{out}$ 。因此，深度可分离卷积的总计算量为：

$$O_{dsconv} = H \times W \times C_{in} \times K \times K + H \times W \times C_{in} \times C_{out} \quad (2)$$

若 C_{in} 和 C_{out} 相同，用深度可分离卷积模块替换标准卷积模块可得到以下计算比较：

$$\begin{aligned}\frac{O_{dsconv}}{O_{conv}} &= \frac{H \times W \times C_{in} \times K \times K + H \times W \times C_{in} \times C_{in}}{H \times W \times C_{in} \times C_{in} \times K \times K} \\ &= \frac{C_{in} \times K^2 + C_{in}^2}{C_{in}^2 \times K^2} = \frac{1}{C_{in}} + \frac{1}{K^2}\end{aligned}\quad (3)$$

当通道数较大时，可以省略公式中的第一项，将参数量对比简化为：

$$\frac{O_{dsconv}}{O_{conv}} \approx \frac{1}{K^2} \quad (4)$$

与标准卷积相比，深度可分离卷积的计算量可降至 1/9，使得深度可分离卷积适合部署在计算资源有限的无人机嵌入式平台上。

ELAN 模块和 DSC-ELAN 模块的设计对比如图 3 所示。ELAN 模块基于跨层连接的概念，首先将输入特征图分成两条通路。一条通路通过 1×1 卷积进行降维，而另一条通路首先执行 1×1 卷积进行降维，然后使用四个标准 3×3 卷积进行特征提取。在 DSC-ELAN 模块中，ELAN 模块中的标准卷积操作被深度可分卷积替代。此外，其中一条路径包括一个额外的 3×3 深度可分离卷积，以增强网络提取局部特征的能力。最后，将所有四条路径的特征图聚合起来，为了保持相同数量的输入和输出通道，并避免由于通道扩展而产生的计算冗余，使用 1×1 卷积进行通道压缩。

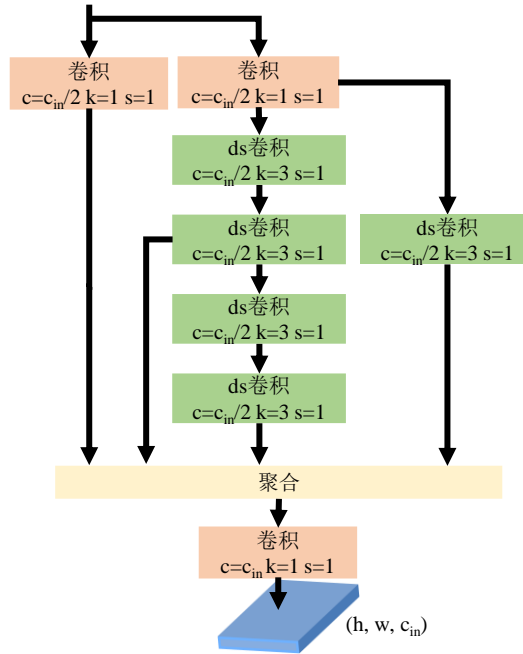


图 3 DSC-ELAN 网络结构图

Fig.3 Architecture of the DSC-ELAN module

对于原始 ELAN 模块，如果输入输出通道数均为 C_{in} ，则计算量可根据参数计算公式得出：

$$O_{ELAN} = 12 \times H \times W \times C_{in} \times C_{in} \quad (5)$$

经计算，DSC-ELAN 模块的计算成本为：

$$\begin{aligned}O_{DSC-ELAN} &= \frac{17}{4} \times H \times W \times C_{in} \times C_{in} \\ &+ \frac{45 \times H \times W \times C_{in}}{2}\end{aligned}\quad (6)$$

在实际的网络架构中，通道数较大，公式中的后一项可以忽略。在该情况下，与 ELAN 模块相比，计算成本可减少 65%：

$$\begin{aligned} \frac{O_{DSC-ELAN}}{O_{ELAN}} &= \frac{\frac{17}{4} \times H \times W \times C_{in} \times C_{in} + \frac{45 \times H \times W \times C_{in}}{2}}{12 \times H \times W \times C_{in} \times C_{in}} \\ &\approx \frac{\frac{17}{4} \times H \times W \times C_{in} \times C_{in}}{12 \times H \times W \times C_{in} \times C_{in}} = \frac{17}{48} \end{aligned} \quad (7)$$

下采样常通过步幅为 2 的最大池化来实现。然而会导致信息丢失。为解决此问题，本文设计了混合下采样（Mixed-downsampling, MDS）模块，其网络结构如图 4 所示。MDS 模块采用双路下采样方式。一条路径利用 3×3 最大池化操作，步幅为 2，以保留局部极大值，增强边缘和形状特征的提取。另一条路径采用 3×3 卷积操作，步幅为 2，执行下采样，提供更平滑的特征映射，并补充池化操作造成的可能信息损失。随后，对两路特征进行聚合，并通过 1×1 卷积来调整通道的数量。这种方法在减少池化操作带来信息损失的同时增强了特征表示能力和多尺度特征适应性。

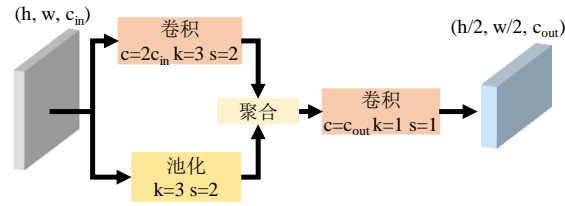


图 4 混合下采样网络结构

Fig.4 Architecture of the hybrid downsampling module

通过将 MDS 模块和 DSC-ELAN 模块与基础网络架构相结合，在特征提取阶段对相应模块进行替换和优化，逐步实现高维特征提取，输出三种不同尺度的特征图。利用空间金字塔池化融合模块进一步增大感受野范围，通过 1×1 卷积来压缩通道数量，避免后续特征融合阶段过大的计算量。在特征融合阶段，同时使用自顶向下和自底向上的过程。DSC-ELAN 模块取代原有 3×3 卷积层，进一步降低计算成本，使得该网络更适合计算资源有限的无人机障碍物检测任务。

在完成主要的特征提取和特征融合过程后，为了进一步增强无人机在复杂环境中检测不同尺度和形状障碍物的能力。本文引入了 E-CBAM，其在通道注意力（Channel attention module, CAM）和空间注意力机制（Spatial attention module, SAM）的基础上增加了快速连接路径，快速连接路径为特征传递提供捷径，有效避免中间层特征信息在堆叠计算中被稀释或丢失，增强梯度流动性，减少了信息丢失，避免了 sigmoid 激活函数引起的梯度消失问题，提供了更稳定的训练效果。网络结构如图 5 所示。

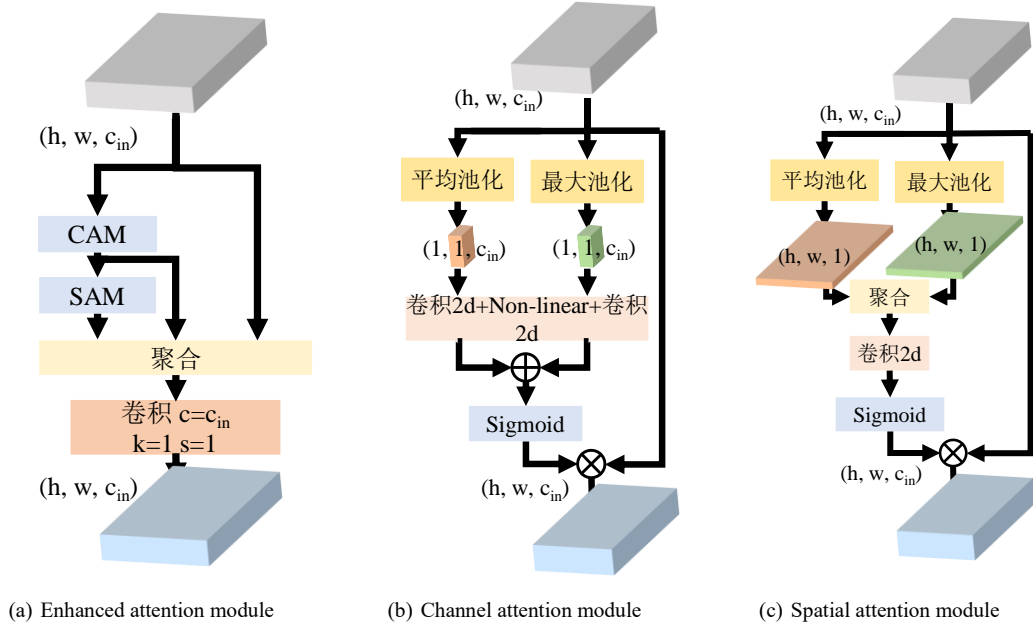


图 5 E-CBAM 网络结构

Fig.5 Network architecture of the E-CBAM

E-CBAM 包含原始通道注意模块，该模块使用全局信息来学习不同通道的重要性。其通过加权运算增强关键特征通道的响应，抑制不相关或冗余的通道，以提高模型的表达能力。对于输入特征图 $F \in R^{C \times H \times W}$ ，使用全局平均池化和全局最大池化提取全局特征，使模型能够专注于整个特征图的通道信息，具体如下：

$$F_{avg}(c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W F_c(i, j)$$

$$F_{max}(c) = \max_{i=1}^H \max_{j=1}^W F_c(i, j)$$
(8)

随后，通过具有共享权重的 1×1 卷积层，激活函数层和归一化函数，生成通道权重。将这些权重与输入特征图相乘，得到输出特征图 F_{CAM} ：

$$F_{CAM} = F \times \sigma(W_2(\delta(W_1[F_{avg}(c) + F_{max}(c)])))$$
(9)

式中 σ 表示 Sigmoid 函数， W_1 和 W_2 表示两个卷积层的权重， $[F_{avg}(c) + F_{max}(c)]$ 表示两个特征映射的共享权重， δ 表示 ReLU 激活函数。

在检测过程中，障碍物目标的形状、大小、位置变化较大，背景复杂。因此，仅依靠通道注意力模块可能不足以实现精确的目标定位。空间注意力模块用于增强特征图中关键空间位置的响应，使模型在检测过程中更多地关注障碍物目标所在的区域，减少不相关背景的影响。对于输入特征图 $F \in R^{C \times H \times W}$ ，在通道维度上进行池化操作获得空间注意特征图：

$$F_{avg}(i, j) = \frac{1}{C} \sum_{c=1}^C F_c(i, j)$$

$$F_{max}(i, j) = \max_{c=1}^C F_c(i, j)$$
(10)

然后，将两个特征图连接起来，并通过带有权重 W_3 的卷积层进行传递。应用 Sigmoid 归一化函数生成空间注意权重，与输入特征图相乘得到输出特征图 F_{SAM} ：

$$F_{SAM} = F \times \sigma(W_3[F_{avg}(i, j); F_{max}(i, j)])$$
(11)

$[F_{avg}(i, j); F_{max}(i, j)]$ 表示两个特征图沿通道维度的拼接。

此外，为了提高网络在更大范围内捕获信息的能力，E-CBAM 在这些模块之间增加了两

条额外的快速连接路径。对于输入特征图 $F \in R^{C \times H \times W}$ ，有三条信息路径输出到模块的末端。沿着通道维度进行堆叠后，通道通过 1×1 卷积再次被压缩。若通道注意模块的输出特征图为 $M \in R^{C \times H \times W}$ ，则 E-CBAM 的输出可表示为：

$$F_{E-CBAM} = W[M; M_{SAM}; F_{skip}] \quad (12)$$

式中， $M = F_{CAM}$ 和 F_{skip} 表示来自快速连接路径的特征图， W 表示 1×1 卷积层的权重。

通过添加这些快速连接路径，加速了信息传播，减少了 CAM 和 SAM 中的信息损失。同时也缓解了梯度消失问题，从而获得了更稳定的训练效果。E-CBAM 模块通过通道注意与空间注意机制增强特征表示，提升模型对关键通道的响应能力和关键区域的识别能力。通过引入快速连接路径，增强信息流动性，缓解特征损失与梯度消失问题。整体结构轻量，适合部署在资源受限的无人机平台中。

1.2 多尺度深度补全网络优化设计

三维激光雷达点云投影到图像平面后，有效像素区域占比较低，大部分区域缺少深度信息。因此，本文采用多尺度神经网络对稀疏深度图像进行补全，获得密集深度图像。整体网络架构如图 6 所示。该网络采用编码器-解码器结构，对于统一模态的图像和点云数据，将生成的四通道特征图输入编码器-解码器网络，输出表示稠密深度图像的单通道特征图，其中每个像素的值对应于该位置的深度信息。为了提高算法的精度，在编码器网络和解码器网络相应特征提取阶段之间构建跨层连接，直接将特征图从编码器阶段输出到解码器阶段。

通常编码器和解码器都使用标准卷积模块进行特征提取，对于在目标检测部分已经设计的高效特征提取模块 DSC-ELAN 和下采样模块 MDS，可直接替换应用到深度补全网络中。对于编码器部分，首先使用 3×3 卷积层将输入特征图维度增加到 32，随后依次应用 DSC-ELAN 和 MDS 模块，实现四个不同尺度下的高维特征提取。对于解码器部分，使用上采样和 DSC-ELAN 模块逐步恢复图像尺度和细节，最终使用具有单个输出通道的 1×1 卷积层生成稠密深度图。

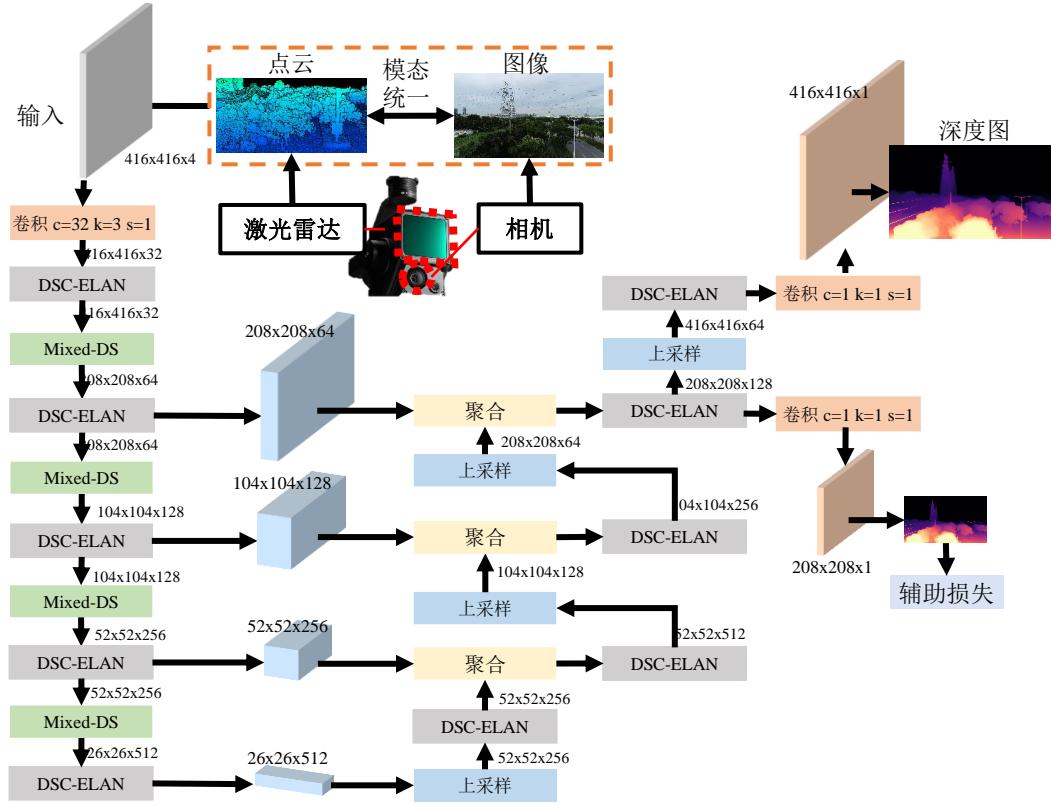


图6 多尺度深度补全网络架构

Fig.6 Architecture of the multi-scale depth completion network

为了增强深度图的边缘细节并使物体边界更清晰，选择最近邻插值进行上采样。该方法直接复制最近的像素值来填充高分辨率网格，而不涉及复杂的插值计算，相比双线性插值的平滑处理和反卷积可能引入伪影，最近邻插值在边缘保留方面具有更好的表现，尤其适合边界清晰的障碍物补全任务。

对于网络的单通道输出特征图 D ，选择 MSE 作为损失函数，其公式为：

$$Loss_{mse} = \frac{\sum_{d \in D} I(d)(d_{pred} - d_{gt})^2}{\sum_{d \in D} I(d)} \quad (13)$$

其中 $I(d)$ 为指示函数，当真实标签像素没有有效深度值时，指示函数为 0，否则为 1。 d_{pred} 表示网络输出， d_{gt} 表示真实标签值。

在实际训练过程中，为加速训练和提供额外的梯度信息，本文设计了辅助损失函数，以增强模型在不同尺度上的深度补全能力。在 208×208 尺度上添加 1×1 卷积层输出该尺度上的深度图预测，然后使用对应深度标签的 MSE 损失对其进行监督。这个深度标签通过对原始比例尺深度图进行降采样获得，以确保尺度一致性。因此，总损失函数为：

$$Loss = \lambda_{main} L_{main} + \lambda_{aux} L_{aux} \quad (14)$$

其中， λ_{main} 和 λ_{aux} 分别为主损失和辅助损失的权重参数。在试验中，通过调整 λ_{aux} 来平衡辅助监督对深度补全的影响，确保在加速训练收敛的同时，保持全尺度深度补全性能。

1.3 三维坐标解算方案设计

通过物体检测算法和深度补全算法获得障碍物的二维检测结果和深度图后，结合两类信息计算障碍物的三维空间位置。假设障碍物在图像坐标系中的位置为 (u, v) ，相机的内参记

为 K ，障碍物在相机坐标系中的三维位置为 (X_c, Y_c, Z_c) ，则可以使用下式计算出障碍物的三维位置：

$$d \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = K \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} \quad (15)$$

其中 f_x 和 f_y 分别为 x 和 y 轴方向上的焦距， c_x 和 c_y 分别为原点在 x 和 y 轴方向上的偏移量。

2 试验结果与分析

2.1 试验设置

本研究通过实际电力巡检场景下的无人机应用验证所提出算法的有效性。试验平台的硬件配置如下所述：算法训练环节采用 AMD Ryzen 9 7950X@4.50 GHz CPU 与 NVIDIA GeForce RTX 4090 GPU 构成的运算单元，系统环境为 Ubuntu 18.04，用于算法的训练与优化。巡检无人机搭载多模态传感器阵列，包含摄像头及激光雷达，其机载计算平台采用 NVIDIA Orin NX 嵌入式计算单元，具备 200 TOPS 的峰值计算性能。现场试验为典型城市复杂电磁环境下的低空电力巡检场景，无人机按预设航线在距地面 50 米高度执行自主巡检任务。试验过程中机载系统以 20Hz 采样频率同步获取前方 150 米探测范围内障碍物的 1280×720 像素 RGB 图像数据及 16 线三维点云信息。图 7 为系统硬件架构及试验场景示意图。



图 7 试验平台与实验环境示意图

Fig.7 Schematic diagram of the experimental platform and environment

2.2 障碍物检测试验

在本文算法框架中，若无法检测到障碍物的二维位置，则无法进行后续的坐标求解，因此二维检测算法的性能对障碍物最终的三维定位效果起着决定性的作用。首先对障碍物检测算法的性能进行测试，障碍类型包括路灯、电线杆、电塔和树木。对于路灯和树木，训练图像从网络数据集中获取，并辅以真实数据采集。对于电塔和电线杆，则通过视觉摄像头和激光雷达采集数据。试验共采集 4314 对图像和点云数据，将点云数据投影到图像平面上，利用传感器之间的外参获得深度图数据。对于训练参数，图像大小设置为 416x416，Batch Size 为 32，优化器为 Adam，初始学习率为 0.001，训练 200 个周期。

障碍物目标检测试验中使用 AP 和 mAP 作为算法的评价指标，不同算法在电力场景下的检测效果会由于泛化能力不同而导致性能产生差异，为全面对比 YOLO 系列不同算法在轻量化低算力场景下的表现，验证本文算法在不同模型架构上的鲁棒性和兼容性，选择 YOLOv6^[15]、YOLOv7^[4]和 YOLOv9^[7]为主要对比算法。其分别代表了 Anchor-free 设计、动态标签分配策略、可逆梯度传播等不同阶段的技术创新，通过横向对比，可验证本文算法在

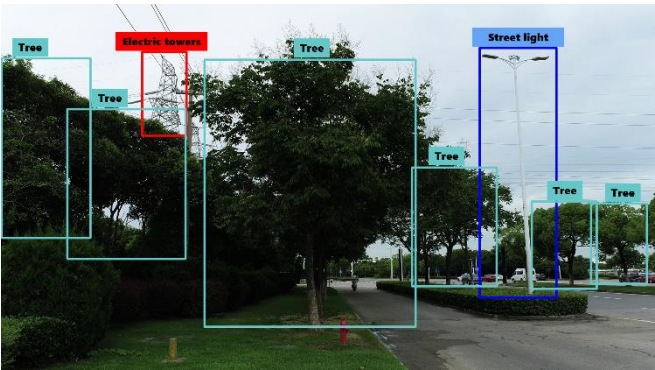
不同模型架构上的鲁棒性与兼容性。四种算法使用相同的超参数，试验结果如表 1 所示。算法的检测结果如图 8 所示。

表 1 目标检测算法精度对比

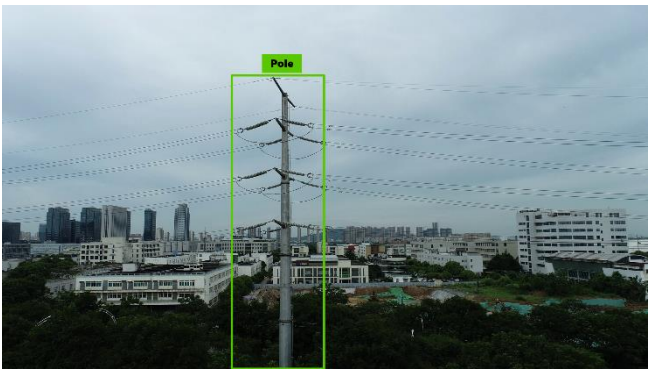
Table 1 Comparison of depth completion accuracy				
类别	YOLOv6	YOLOv7	YOLOv9	本文算法
路灯	0.72	0.74	0.76	0.79
电线杆	0.68	0.72	0.75	0.79
电塔	0.71	0.76	0.77	0.82
树木	0.80	0.83	0.85	0.87
mAP	0.73	0.76	0.78	0.82

表 2 算法消融试验精度比较

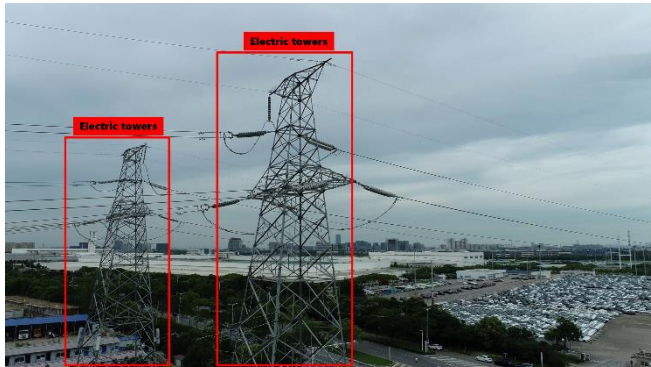
Table 2 Accuracy comparison of algorithmic ablation experiments	
算法	mAP
YOLOv9	0.78
YOLOv9+DSC-ELAN+MDS	0.81
YOLOv9+DSC-ELAN+MDS+E-CBAM (本文算法)	0.82



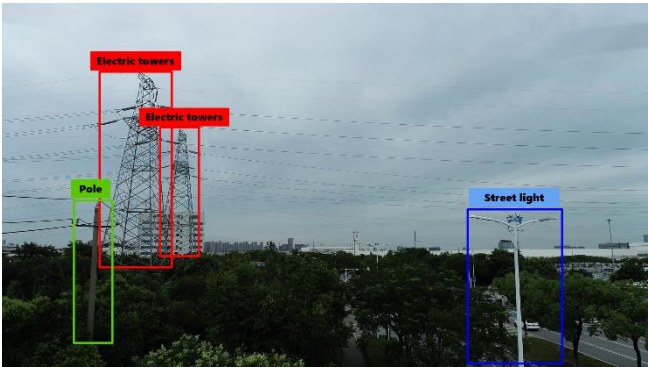
(a) Detection of streetlights, trees and transmission towers



(b) Detection of poles



(c) Detection of transmission towers



(d) Detection of transmission towers, poles and streetlights

图 8 实际电力巡检中障碍物目标检测结果图

Fig.8 Obstacle detection results in practical power inspection scenarios

综合以上图表可以得出以下结论：

1) 本文算法较对比算法中具有较高的 AP 值。从表 1 可以看出，对于路灯、电线杆、电塔、树木这四种障碍物类别，本文算法 AP 绝对值比 YOLOv6 算法分别提高了 7%、11%、

11%、7%，mAP 值提高了 9%。比 YOLOv7 分别提高了 5%、7%、6%、4%，mAP 值提高了 6%，比 YOLOv9 分别提高了 3%、4%、5%、2%，mAP 值提高了 4%，相对提高了 5.1%。

2) 本文所提特征提取增强模块与融合网络模块可有效改进目标识别性能。从表 2 可以看出，加入 DSC-ELAN 与 MDS 模块后，mAP 提升 3%，进一步加入 E-CBAM 模块后，mAP 进一步提升 1%，证明了改进特征提取和特征融合网络模块的有效性。

2.3 深度补全试验

由于像素级深度标定难度较大，因此，基于公共数据集验证深度补全性能。选取 KITTI 深度补全数据集作为训练数据^[16]，其包括原始雷达点云、视觉图像和传感器参数，共提供 85898 个训练样本、1000 个验证样本和 1000 个测试样本。训练的 Batch Size 大小为 8，初始学习率为 0.001，其他超参数设置为默认值，共训练 40 个 Epoch。使用均方根误差（Root mean square error，RMSE）和平均绝对误差（Mean absolute error，MAE）作为评估指标。KITTI 验证数据集上的误差结果和实际处理时间如表 3 所示。

表 3 深度补全精度速度对比

Table 3 Comparison of depth completion accuracy and speed			
算法	RMSE (mm)	MAE (mm)	时间(ms)
Sparse2Dense ^[17]	815	250	87
FusionNet ^[19]	773	215	23
Improving Single-bra ^[20]	830	233	27
多尺度估计算法	761	209	11
本文算法	755	203	11

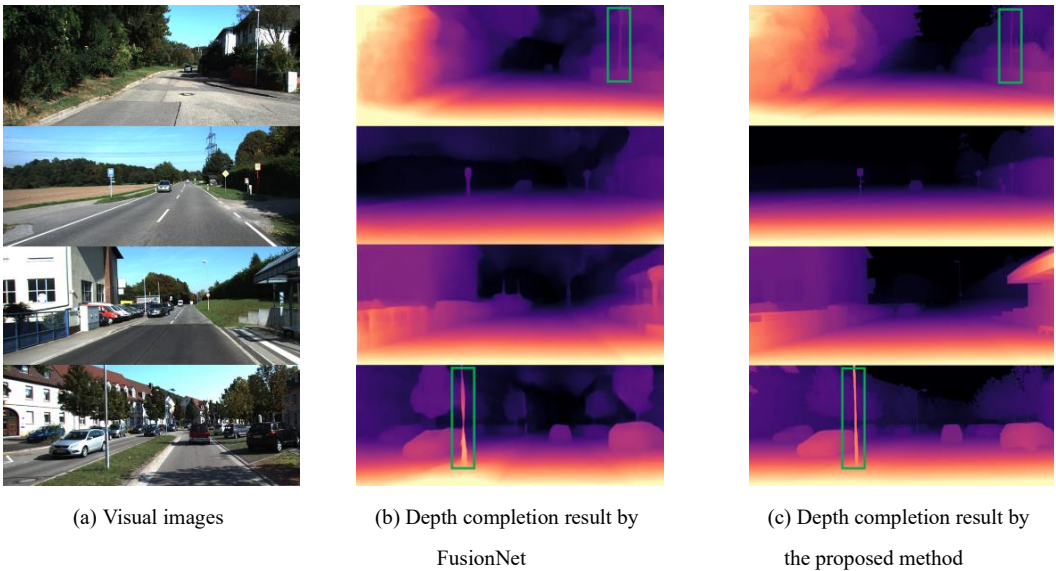


图 9 KITTI 数据集深度补全图像

Fig.9 Depth completion images from the KITTI dataset

综合以上图表可以得出以下结论：

1) 本文算法较对比算法实现了更高精度的深度估计。如表 3 所示，相较于 Sparse2Dense、FusionNet 和 Improving Single-bra 三类算法，所提方法在 RMSE 指标分别降低 6.63%、2.33%

和 8.31%，MAE 指标分别降低 16.40%、2.79%和 10.30%；引入 Aux Loss 后，综合精度达到最佳值 755 mm，较基础版本再提升 0.79%。设计的多层次特征融合结构和置信度引导机制，能有效提升稀疏深度点的传播精度。

2) 本文算法较对比算法具有更高的实时性。在输入分辨率 416×416 条件下，所提方法推理耗时仅为 FusionNet 算法的 48%、Improving Single-bra 算法的 40.7%。该效率提升主要得益于网络主干中部署的 DSC-ELAN 轻量化模块，其深度可分离卷积结构可使计算量减少约 83%，同时通过动态通道修剪机制自适应优化计算路径。

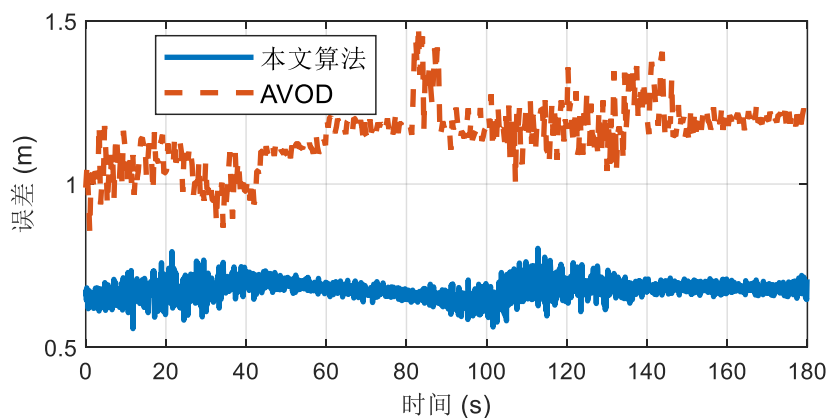
2.4 障碍物定位试验

在实际场景中开展障碍物目标识别与三维定位性能验证，试验中采集并标注了 4314 对点云和视觉图像，使用相同的参数进行训练，训练后将算法部署在无人机平台上。本文选择 AVOD^[14]作为比较算法，该算法通过融合图像与点云数据进行三维检测直接获得障碍物的三维空间位置。利用欧几里得距离公式计算真实坐标位置 (x, y, z) 与算法得到的坐标位置 (x', y', z') 之间的最终定位误差。在城市和道路环境中检测并定位多个障碍物后的距离误差如表 4 所示。对两种算法的运行时间进行测试，结果如表 5 所示。

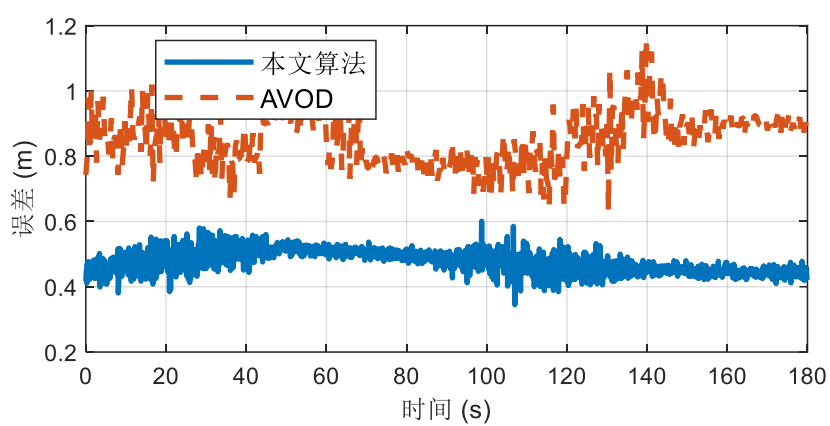


图 10 电力环境深度补全图像

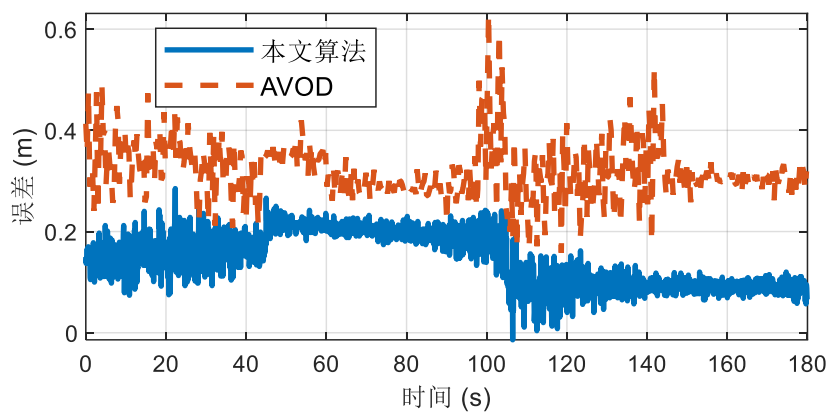
Fig.10 Depth completion images in the power environment



(a) Localization error along the X-axis



(b) Localization error along the Y-axis



(c) Localization error along the Z-axis

图 11 三维目标估计误差曲线图

Fig.11 3D object estimation error curves

表 4 障碍物定位算法精度比较

Table 4 Comparison of obstacle localization algorithm accuracy

类别	AVOD/m	算法/m
路灯	1.23	0.68
电线杆	0.89	0.53
电塔	1.61	0.86
树	1.34	0.71
平均误差	1.27	0.70

表 5 算法处理时间比较

Table 5 Comparison of algorithm processing time	
算法	处理时间/ms
AVOD	117
本文检测算法	17
本文定位算法	11

综合上述图表，可以得出以下结论：

1) 本文算法可有效保留障碍物目标细节边界。由图 10 可视化对比可知，绿框标定区域的深度跳变边界清晰度高，对电线杆等狭窄物体的深度连续性保持良好。

2) 本文算法相较 AVOD 具有更高的障碍物定位精度。从图 11 可以看出，本文方法对比 AVOD 方法具有更高的定位精度。如表 4 所示，在电力巡检典型障碍物定位任务中，本文方法较 AVOD 算法在路灯、电线杆、电塔、树木四类目标的定位误差分别降低 44.72%、40.45%、46.58%、47.01%，平均定位精度提升 45%。改进的深度补全网络通过融合多尺度几何特征与纹理特征，有效增强了复杂电力巡检场景下的空间感知能力，实现了亚米级精度的障碍物定位。

3) 本文算法相较 AVOD 具有更优的实时性能。如表 5 所示，AVOD 算法单帧处理耗时为 117ms，而本文通过解耦式架构设计，将目标检测与定位阶段耗时分别控制在 17ms 和 11ms，总处理时间仅为 28ms，处理效率提升 76.07%。结合轻量化网络设计与硬件加速策略，最终实现 36fps 的实时处理帧率，满足电力巡检系统对实时响应的要求。

3 结束语

本文通过提出一种用于三维障碍物识别定位的视觉/激光雷达解耦融合方法，实现了在有限算力的无人机端侧实时三维目标检测。在检测阶段，考虑到无人机计算资源有限，引入 DSC-ELAN 模块、MDS 模块和 E-CBAM 模块，在减少网络参数数量的同时增强网络的特征提取能力。此外，在设计的高效特征提取模块基础上，构建了具有跨层连接和辅助损失函数的编码器-解码器网络，以提高深度补全算法的精度。获取到障碍物检测信息与深度信息后，在障碍物定位阶段设计了高效精简的坐标计算方法，确定障碍物的三维空间位置。试验结果表明，本文算法在电力巡检环境中对不同类型的障碍物具有更高的定位精度和更快的处理速度，具有较好的工程应用价值和推广潜力。

参考文献：

- [1] Wang Z, Gao Q, Xu J, et al. A review of UAV power line inspection[C]//Advances in Guidance, Navigation and Control: Proceedings of 2020 International Conference on Guidance, Navigation and Control, ICGNC 2020, Tianjin, China, October 23–25, 2020. Springer Singapore, 2022: 3147-3159.
- [2] 李亚飞,赵瑞.城市复杂环境下多目标无人机路径规划研究[J].南京航空航天大学学报,2024,56(06):1002-1012.
Li Y, Zhao R. Research on Multi-target UAV Path Planning in Urban Complex Environments. [J]. Journal of Nanjing University of Aeronautics & Astronautics/Nanjing Hangkong Hangtian Daxue Xuebao, 2024, 56(6):1002-1012.
- [3] 张辉,杜瑞,钟杭,等.电力设施多模态精细化机器人巡检关键技术及应用[J].自动化学报,2025,51(01):20-42.

- Zhang H, Du R, Zhong H, et al. The Key Technology and Application of Multi-modal Fine Robot Inspection for Power Facilities[J]. *Zidonghua Xuebao/Acta Automatica Sinica*, 2025,51(01):20-42.
- [4] Bashir N, Boudjit S, Dauphin G, et al. An obstacle avoidance approach for UAV path planning[J]. *Simulation modelling practice and theory*, 2023, 129: 102815.
- [5] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2023: 7464-7475.
- [6] 张玉,尚志华,郭晓楠,等.小样本图像分类中的类别信息融合网络[J].*南京航空航天大学学报*,2022,54(04):715-722.
- [7] Zhang Y, Shang Z, Guo X, et al. Category Information Fusion Network in Few-Shot Image Classification. [J]. *Journal of Nanjing University of Aeronautics & Astronautics/Nanjing Hangkong Hangtian Daxue Xuebao*, 2022, 54(4):715-722.
- [8] Wang C Y, Yeh I H, Mark Liao H Y. Yolov9: Learning what you want to learn using programmable gradient information[C]//*European conference on computer vision*. Cham: Springer Nature Switzerland, 2024: 1-21.
- [9] Zheng L, Tang M, Chen Y, et al. Improving multiple object tracking with single object tracking[C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021: 2453-2462.
- [10] Zamanakos G, Tsochatzidis L, Amanatiadis A, et al. A comprehensive survey of LIDAR-based 3D object detection methods with deep learning for autonomous driving[J]. *Computers & Graphics*, 2021, 99: 153-181.
- [11] Yao Z, Chen X, Xu N, et al. LiDAR-based simultaneous multi-object tracking and static mapping in nearshore scenario[J]. *Ocean Engineering*, 2023, 272: 113939.
- [12] Li Z, Wang F, Wang N. Lidar r-cnn: An efficient and universal 3d object detector[C]//*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021: 7546-7555.
- [13] Liu Z, Tang H, Amini A, et al. Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation[C]//*2023 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2023: 2774-2781.
- [14] Lin Z, Liu Z, Xia Z, et al. RCBEVDet: radar-camera fusion in bird's eye view for 3D object detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024: 14928-14937.
- [15] Tao Z, Wan Y, Hou J, et al. Improved AVOD Model Based on a Bidirectional Fusion Mechanism for 3D Object Detection[C]//*Proceedings of the 2024 3rd International Symposium on Intelligent Unmanned Systems and Artificial Intelligence*. 2024: 309-312.
- [16] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. *arXiv preprint arXiv:2209.02976*, 2022.
- [17] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? the kitti vision benchmark suite[C]//*2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012: 3354-3361.
- [18] Wang T, Hu X, Liu Z, et al. Sparse2Dense: Learning to densify 3d features for 3d object detection[J]. *Advances in Neural Information Processing Systems*, 2022, 35: 38533-38545.
- [19] Van Gansbeke W, Neven D, De Brabandere B, et al. Sparse and noisy lidar completion with rgb guidance and uncertainty[C]//*2019 16th international conference on machine vision applications (MVA)*. IEEE, 2019: 1-6.
- [20] Eldesokey A, Felsberg M, Khan F S. Confidence propagation through cnns for guided sparse depth regression[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2019, 42(10): 2423-2436.

[21] Wang Y, Zhang G, Wang S, et al. Improving depth completion via depth feature upsampling[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 21104-21113.

作者简介:

	蒋承伶(1986-), 男, 高级工程师, 研究方向: 电力人工智能及无人机智能运检技术研究, E-mail: jcl@js.sgcc.com.cn。		黄 郑(1990-), 通信作者, 男, 高级工程师, 研究方向: 电力无人机智能运检技术研究, E-mail: huangzhg10@mail.163.com。		沈 超(1987-), 男, 高级工程师, 研究方向: 电力无人机智能运检技术研究, E-mail: 527200783@qq.com。
	马洲俊(1986-), 男, 高级工程师, 研究方向: 电力人工智能技术研究, E-mail: mazjl@js.sgcc.com.cn。		张明晖(1991-), 男, 工程师, 研究方向: 输电线路无人机智能运检技术, E-mail: zhangmh5@js.sgcc.com.cn。		