

基于改进的YOLOv8毫米波图像目标分割方法

高晨凯, 钱子凡, 李文杰, 叶学义

杭州电子科技大学通信工程学院, 浙江 杭州

收稿日期: 2026年5月18日; 录用日期: 2026年6月22日; 发布日期: 2026年6月30日

摘要

现阶段的毫米波图像检测多依赖于检测框, 未充分利用目标轮廓信息, 限制了检测系统的性能。针对这一问题, 本文提出一种基于改进的YOLOv8分割模型, 通过轮廓信息提升检测精度, 是对现有先进技术的一次成功的、针对特定应用场景的优化集成。首先, 引入了感受野坐标注意力卷积模块强化目标轮廓特征提取。其次, 通过分析毫米波图像中小目标为主的分布特性, 新增一层专门用于增强小目标检测能力的检测层。最后, 采用自适应阈值的标签分配解决小目标IoU敏感的问题, 进一步确保模型对小目标轮廓的精准识别。实验结果表明, 该方法在box mAP50上提升了2.6%, 在mask mAP50上提升了10.4%, 在毫米波图像检测中精度优势显著。

关键词

毫米波图像检测, 感受野坐标注意力卷积, 自适应阈值

A New Method for Millimeter-Wave Image Target Segmentation Using the Tuned YOLOv8

Chenkai Gao, Zifan Qian, Wenjie Li, Xueyi Ye

School of Communication Engineering, Hangzhou Dianzi University, Hangzhou Zhejiang

Received: May 18, 2026; accepted: June 22, 2026; published: June 30, 2026

Abstract

At present, millimeter-wave image detection mostly relies on detection frames and fails to make full use of the target contour information, which limits the performance of the detection system. To address this issue, this paper proposes an improved YOLOv8 segmentation model to enhance detection accuracy through contour information. Firstly, the Receptive Field Coordinate Attention

Convolution module was introduced to enhance the extraction of target contour features. Secondly, by analyzing the distribution characteristics dominated by small targets in millimeter-wave images, a new detection layer specifically designed to enhance the detection capability of small targets is added. Finally, adaptive threshold label allocation is adopted to address the issue of IoU sensitivity of small targets, further ensuring the model's accurate recognition of the contours of small targets. The experimental results show that this method improves by 2.6% on box mAP50 and by 10.4% on mask mAP50, and has a significant advantage in accuracy in millimeter-wave image detection.

Keywords

Millimeter-Wave Image Detection, Receptive Field Coordinate Attention Convolution, Adaptive Threshold

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着近年公共交通枢纽如机场、火车站和地铁站安全需求提升，传统安检技术的缺陷愈发凸显。例如，X 射线对人体有辐射危害；金属探测器对于非金属和隐藏在衣物下的物品不够敏感等。而毫米波成像作为一种新兴的安检技术因其独特的优势而备受关注。

毫米波是指频段在 30 GHz 至 300 GHz 之间的电磁波，其波长较短(1 mm~10 mm)，能有效地穿透衣物和其他表面物体[1]，从而提供隐藏在物体表面下的信息。且对人体无害，并能够在不损害隐私的情况下检测到潜在的危险物品。这使得毫米波成像成为一种非接触式、高效且隐私友好的安检技术。

过去数年，已有多项研究针对毫米波图像隐藏目标检测展开。早期方法多基于机器学习或数学模型方法来检测物体[2]。但这些方法仅在一些具有特定约束的小数据集上进行评估，难以应对如噪声背景与物体相似度高、小目标占多数，成像不稳定等实际场景。而深度学习在毫米波物体检测这方面取得了重大突破。如 Liu 等人在 Faster RCNN[3]的基础上提出了一种上下文嵌入目标检测网络[4]，并在特征图进行了具有不同扩张率的扩张卷积操作以提升在小目标上的检测性能；Pang 等人采用了 Yolo 系列算法中的 Yolov3 让毫米波图像从检测精度、检测速度和计算资源的角度上使得毫米波图像的实时检测更为有效和可行[5]；Su 等人通过引入新颖的卷积结构 Wavelet-Conv 和注意力机制 Wavelet-Attention [6]，该结构在保持信息完整性的同时，将高频和低频带的特征分离，通过有效地利用频带特征提高了毫米波图像中小目标的召回率和平均精度，同时保持了具有竞争力的推理速度。

现有研究大多聚焦于目标检测，较少利用图像分割获取更精细的轮廓信息。轮廓信息对刀具、枪支等形状敏感类违禁品识别尤为关键，然而在当前毫米波图像分割领域中，目前仍缺少成熟的深度学习方案。因此，本文以 YOLOv8 [7]网络为基础框架，进行场景化优化集成，主要聚焦于分割隐藏物体方面，分割所带有的轮廓信息更有利于识别像枪、刀这类违禁品，提高检测系统的效果和可靠性。首先，引入了基于感受野注意力卷积操作(Receptive Field Coordinate Attention Convolution, RFCACConv) [8]，以取代标准卷积操作。RFCACConv 利用空间特征为卷积核提供有效的注意力权重，能够捕捉不同位置带来的信息差异，从而提升卷积神经网络的性能，而且几乎不增加额外的计算成本和参数。其次，针对毫米波图像中小目标分布的特性，对原始的 Yolov8 检测层进行了调整。随着网络深度增加，特征分辨率降低，语义信息增加，但位置信息变得更加粗糙，感受野也变得更大，这往往会导致忽略小目标的特征。为了提升

小目标检测性能，增加了更有利于检测小目标的高分辨率浅层特征图。最后，采用自适应阈值的方法 (Adaptive Training Sample Selection, ATSS) [9] 来进行正负样本的标签分配，这种自适应的策略可以根据具体数据集和目标特点进行调整，从而更好地平衡正负样本的比例，避免过度拟合或欠拟合的问题，同时不会引入任何额外的计算开销。

本文属于现有先进技术的场景化优化集成，而非底层网络结构创新，旨在为毫米波安检提供高精度、抗干扰能力强的分割方案。

2. 相关工作

2.1. 毫米波图像检测

毫米波图像检测可分为传统方法与深度学习方法两类。传统方法依赖人工设计的特征，如边缘、纹理、形状等，抗干扰能力较弱，难以适应毫米波图像噪声高、目标尺寸小、轮廓模糊的特点。深度学习方法则可分为两类：一类是先提取候选区域再进行分类与回归的两阶段检测方法；另一类是直接特征图上完成预测的一阶段检测方法；此外还有能够同时完成目标检测与轮廓分割的实例分割方法。现有毫米波图像检测研究多采用一阶段检测框架，能够兼顾检测精度与运行速度，但普遍缺少针对小目标和轮廓特征的系统性优化设计。

2.2. 小目标检测技术

多尺度特征融合、高分辨率浅层设计、自适应样本分配与注意力增强，是小目标检测的主流优化思路。多尺度融合通过高低层特征互补，强化小目标细节表达；高分辨率浅层特征可保留更多空间信息，避免细小特征丢失；自适应样本分配能根据目标特性动态调整正负样本阈值，缓解小目标对交并比敏感的问题；注意力增强则聚焦关键区域、抑制背景干扰，提升弱特征辨识度。

在单阶段检测框架中，已有多种面向小目标的优化方案，包括增设高分辨率特征层、改进特征融合结构、优化样本分配策略以及引入轻量化注意力模块。本文综合毫米波图像噪声高、小目标占比大、轮廓模糊的特点，选用高分辨率特征层、感受野坐标注意力模块、自适应样本分配策略的组合方案，兼顾轻量高效、低参数量、实时性与针对性，有效解决毫米波图像中小目标检测与轮廓提取的核心难点。

2.3. 注意力机制对比

常见的注意力机制对比如 CBAM、CA、ECA、RFCACConv 等。CBAM 是通道 + 空间注意力，它的参数量较大，计算负担较大；CA 是坐标注意力，虽然轻量但轮廓捕捉能力弱；ECA 是高效通道注意力，它仅仅关注通道的维度，但缺乏空间位置感知；而 RFCACConv 感受野坐标注意力卷积融合空间位置与感受野信息，几乎不增参数量与计算量，轮廓特征提取能力最优。因此，本文选择 RFCACConv 替代标准卷积，新增 P2 高分辨率层，采用 ATSS 自适应标签分配，形成适配毫米波安检场景的优化集成方案。

3. 改进的 YOLOv8 模型

基于 YOLOv5 [10]，YOLOv8 做出了全面升级：第一，主干网络采用了 CSPDarknet [11] 的增强版本，同时引入了 C2f 模块，增强了特征提取和融合。该模块将特征图分为两部分，分别经过不同的卷积层处理，然后再合并，将高级特征与上下文信息结合。这种设计在保持准确性的同时降低了计算复杂度。此外，其采用先进的激活函数如 SiLU (Swish) [12]，替代了 Leaky ReLU，提升了梯度流和特征表达能力。第二，在 Neck 部分沿用路径聚合网络(PANet) [13]，以改善网络不同层之间的信息流和特征融合。PANet 基于特征金字塔网络(FPN) [14] 设计，增加了一个额外的自底向上路径，与传统的自顶向下路径相结合。

这种设计增强了不同特征层级之间的信息流动,改善了网络在多尺度上检测对象的能力,并提升了对小物体检测的性能。第三,在 Head 部分采用无锚点(anchor free)架构[15],使用解耦的头部来分别独立处理目标性、分类和回归任务,使每个分支能够专注于自己的任务,从而提高了模型的整体精度。第四,在损失函数部分使用 CIoU [16]和 DFL [17]损失函数来处理边界框损失,以及二元交叉熵损失[18]来处理分类损失。兼顾检测精度和小目标性能。此外, YOLOv8 还引入了混合精度训练技术[19]和超参数优化。混合精度结合了 16 位和 32 位浮点运算,以加速训练过程、减少内存使用,同时保持模型的准确性。混合精度训练特别适用于计算能力和内存有限的边缘设备。

本文基于原始 YOLOv8 框架进行了三项关键改进(图 1)。

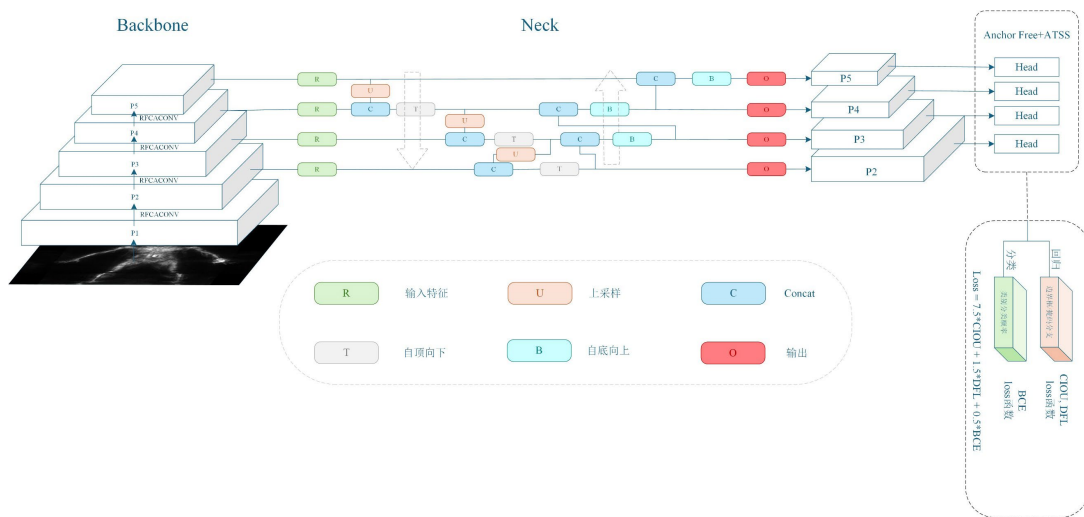


Figure 1. Construction of the proposed network
图 1. 网络整体结构

3.1. 感受野注意力卷积操作(Receptive Field Coordinate Attention Convolution, RFCACONV)

标准的卷积操作是通过滑动窗口和共享参数来提取特征信息,解决使用全连接层构建的神经网络所面临的问题,如参数过多和计算负担高的困扰。但无法感知不同空间位置的信息差异,最终在一定程度上限制了卷积神经网络的性能。空间注意力机制被广泛用于提升卷积神经网络的性能,但它通常会带来更大的参数量以及复杂度。

为了解决以上问题,本文引入感受野注意力卷积操作(RFCACONV)来替代标准的卷积操作,将空间注意力与卷积核权重自适应结合,以空间特征为卷积核提供注意力权重,几乎不增加任何计算成本和参数。为了直观地展示这个过程,在 1×1 卷积操作前插入了空间注意机制。通过注意力映射对输入特征图进行加权操作,最终通过 1×1 卷积操作提取出感受野中的滑动特征信息。整个过程可以简单表示如下:

$$\begin{aligned} F_1 &= X_1 \times A_1 \times K \\ F_2 &= X_2 \times A_2 \times K \\ &\vdots \\ F_N &= X_N \times A_N \times K \end{aligned}$$

我们将 $A_i \times K$ 的值作为新的卷积核参数, 1×1 卷积操作在特征提取中的参数共享问题得到了解决。如图 2 所示,它展示了 1×1 卷积核和空间注意力机制的组合。这种组合本质上解决了参数共享的问题,具体来说,输出特征是通过乘以感受野滑块、注意力权重和 1×1 卷积核参数得到的。

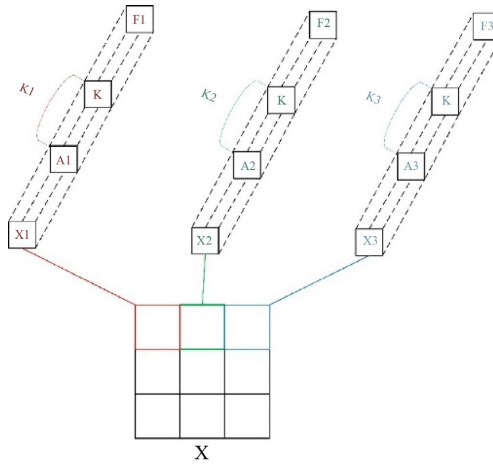


Figure 2. The convolution kernel parameters, and it varies for each sliding block of the receptive field
图 2. 卷积核参数 $K_i = A_i \times K$, 并且每个感受野滑动块中是不同的, 即: $K_1 \neq K_2 \neq K_3 \neq \dots \neq K_N$

它的具体流程如图 3 所示该方法首先通过组卷积和批归一化生成特征图, 将其重排为适用于注意力机制的格式。接着, 沿高度和宽度方向自适应平均池化, 生成特征图并拼接, 经过点卷积和激活函数处理后切分。最后, 通过卷积计算高度和宽度的注意力权重, 归一化后与特征图相乘, 再通过 $k \times k$ 卷积输出结果。

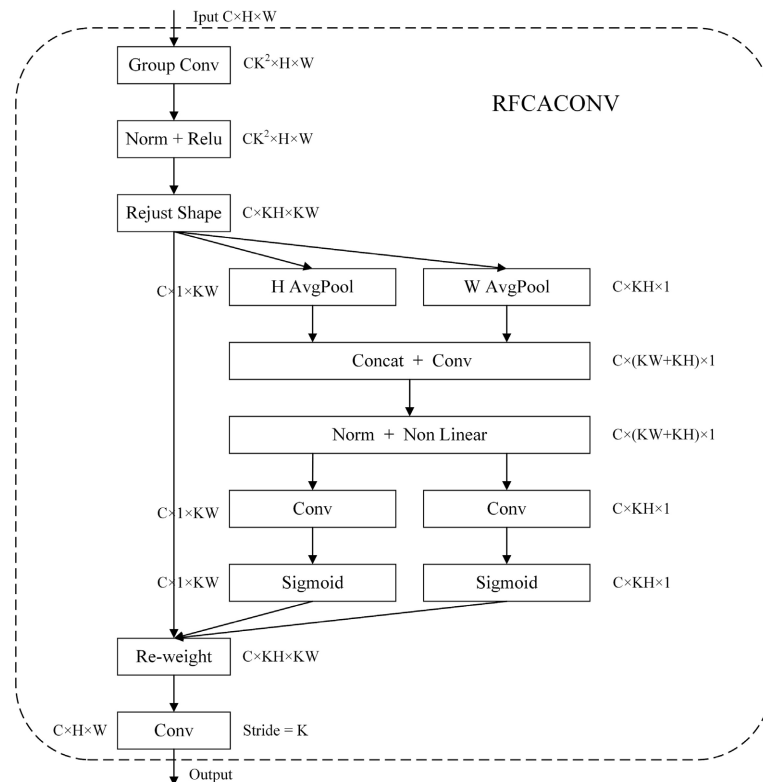


Figure 3. Receptive Field Coordinate Attention Convolution (RFCACONV)
图 3. RFCACONV

3.2. 目标检测层调整

为了了解毫米波图像目标大小的分布特性，本文在自标注的毫米波分割数据集上，对总计 5500 个目标的大小进行了统计。统计结果如图 4 所示，目标的大小分布被划分为四个范围：面积小于 16^2 个像素、面积位于 16^2 至 32^2 个像素、面积位于 32^2 至 48^2 个像素，以及面积大于 48^2 个像素。

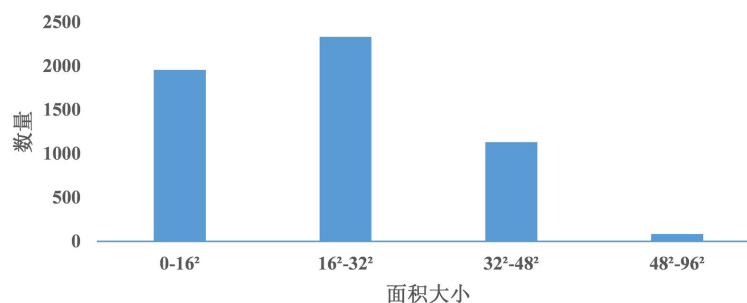


Figure 4. Target distribution characteristic statistics
图 4. 目标分布特性统计

根据统计结果显示，有超过 75% 的目标小于 32^2 个像素，小目标占比极高，这也是毫米波图像最显著的特点。这种目标分布特性表明，在开发毫米波图像的目标检测算法时，需要特别关注对小目标的检测和分割精度。

原始 YOLOv8 采用 P3、P4、P5 三层特征金字塔，分别对应不同大小的目标：P3 层负责检测较小的目标，P4 层处理中等大小的目标，而 P5 层则主要用于检测较大的目标。然而，由于毫米波图像中的小目标占比较高，许多目标的面积都小于 32^2 个像素，甚至 16^2 个像素，仅依赖 P3 层可能不足以捕捉到这些细小的特征信息。因此，为了提升毫米波图像中小目标的检测性能，本文增加高分辨 P2 浅层特征层，可以获得更高分辨率的特征，捕捉到更细微的目标细节。如图 5 所示，这种结构的调整将提高 YOLOv8 模型在小目标检测任务中的适应性，特别是在毫米波图像中小目标占比较高的情况下，这种改进将显著增强模型的性能，满足实际应用的需求。

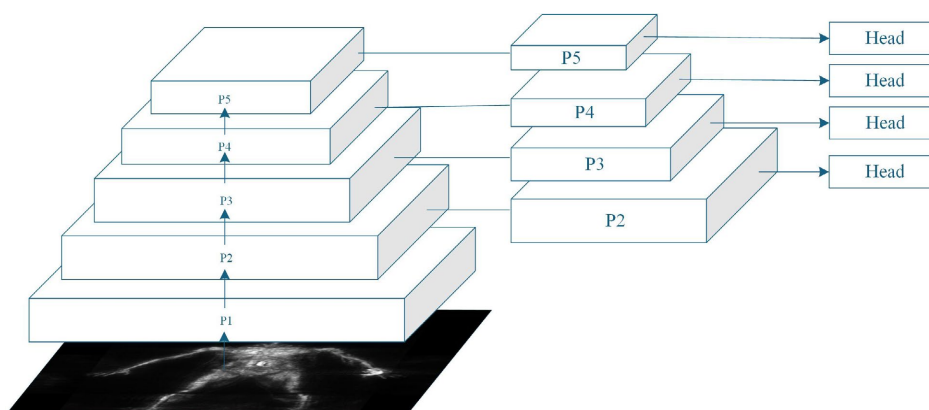


Figure 5. Multi-scale feature fusion module
图 5. 多尺度特征融合模块

3.3. 自适应阈值的标签分配方法

通常样本选择策略存在一些敏感的超参数，例如基于锚框的检测器中的 IoU 阈值和基于无锚点的检

测器中的尺度范围。在设置完这些超参数后,所有的真实值框都必须按照固定的规则选择它们的正样本,这些规则适用于大多数物体,但是一些较为特殊的物体可能会被忽略。尤其是毫米波图像大多数的目标都是小目标,而小目标对于 IoU 的变化极其敏感,预测中的微小差异往往会导致 IoU 损失的巨大变化。如图 6 所示,两个 10×10 的框,仅仅只有 1 行 1 列不重叠,它的 IoU 却只有 0.68。如果是两个 100×100 的框只有 1 行 1 列不重叠,它的 IoU 高达 0.976。

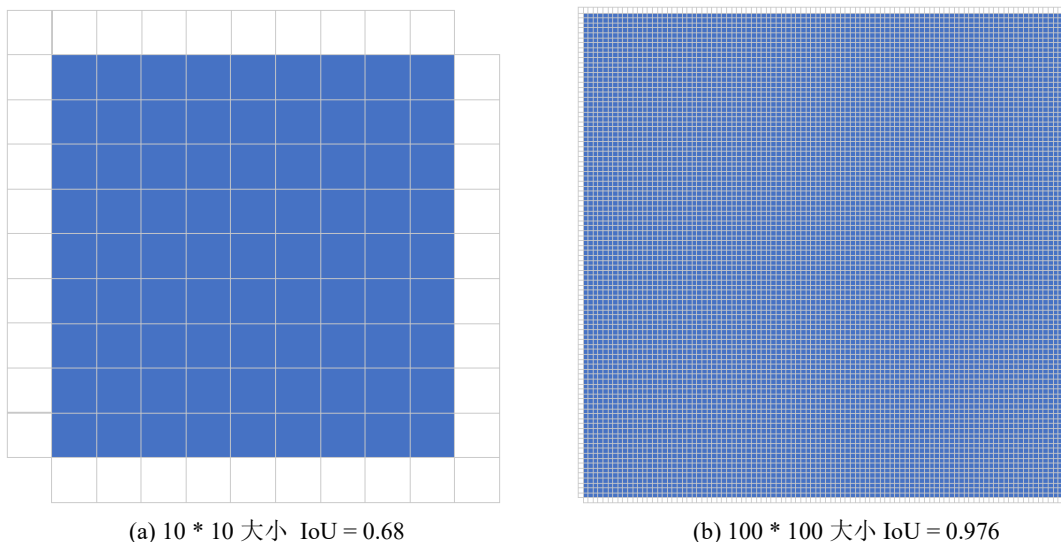


Figure 6. IoU threshold for differently sized objects with single row/column gap
图 6. 不同大小目标在仅有一行一列不相交情况下的 IoU 阈值

因此,这些超参数的不同设置将会产生非常不同的结果。本文采用自适应阈值标签分配,自动计算动态阈值,实现正负样本的精准划分,不引入额外计算开销,可以显著提升小目标的训练稳定性与分割精度。算法 1 描述了该方法如何对输入图像进行工作。对于图像上的每个真实框 g ,我们首先找出其候选正样本。如第 3 至 6 行所述,在每个金字塔层上,我们根据 L2 距离选择 k 个中心最接近 g 中心的锚框。假设有 L 个特征金字塔层,则真实框 g 将有 $k \times L$ 个候选正样本。之后,我们计算这些候选框与真实框 g 之间的 IoU,即第 7 行中的 D_g ,以及均值和标准差,分别为第 8 行和第 9 行中的 m_g 和 v_g 。利用这些统计数据,可以得到该真实框 g 的 IoU 阈值,即第 10 行中的 $t_g = m_g + v_g$ 。最后,我们选择 IoU 大于或等于阈值 t_g 的候选框作为最终的正样本。值得注意的是,正样本的中心限制在真实框内,如第 12 行所示。此外,如果一个锚框被分配给多个真实框,则将选择 IoU 最高的那个,其余的为负样本。

具体算法的伪代码如下:

算法 1: 自适应阈值的标签分配算法:

输入:

图像上的真实边界框集合 G

特征金字塔的层数 L

是第 i 层金字塔中的锚框集合 A_i

所有锚框的集合 A

具有鲁棒性的超参数 k

输出:

一组正样本 P

一组负样本 N

续表

```

1:  for 每一个真实框  $g \in G$  do
2:      构建一个空集合用于存储与真实边界框  $g$  相关的候选正样本:  $C_g \leftarrow \emptyset$ ;
3:      for 对于每个层级  $i \in [1, L]$  do
4:          用基于 L2 距离的规则从集合  $A_i$  中选择  $k$  个锚框得到  $S_i$ , 这些锚框的中心最接近真实边界框  $g$ 
          的中心。
5:          将选出的锚框集合添加到  $C_g$  中:  $C_g = C_g \cup S_i$ ;
6:      end for
7:      计算  $C_g$  与  $g$  的 IoU:  $D_g = \text{IoU}(C_g, g)$ ;
8:      计算  $D_g$  的平均值:  $m_g = \text{Mean}(D_g)$ ;
9:      计算  $D_g$  的标准差:  $v_g = \text{Std}(D_g)$ ;
10:     计算真实边界框  $g$  的 IoU 阈值:  $t_g = m_g + v_g$ ;
11:     for 每个候选样本  $c \in C_g$  do
12:         if  $\text{IoU}(c, g) \geq t_g$  and  $c$  的中心在  $g$  内 then
13:             添加  $c$  到正样本集合  $P$  中:  $P = P \cup c$ ;
14:         end if
15:     end for
16: end for
17: 将未被选中的锚框集合设为负样本集合:  $N = A - P$ ;
18: return  $P, N$ ;

```

4. 实验结果及分析

4.1. 数据集

本文的研究主要基于自标注的毫米波分割数据集和公开的毫米波检测数据集。由于目前几乎没有公开可用的毫米波图像分割数据集, 为了支持本研究的需求, 同时确保数据集的准确性, 本文采用了预训练好的 SAM (Segment Anything Model) [20]模型与人工标注相结合的方式对数据集进行分割。通过先使用 SAM 模型进行初步自动分割, 再由人工进行修正和校对, 保证了分割数据的精度, 从而为研究提供了可靠的数据基础。

自标注的数据集共计包含 4000 张图像, 其中 324 张为背景样本。数据集中总共标注了 5500 个目标, 共分为 6 个类别, 分别是刀类、爆炸物、液体、枪、打火机以及其它。各类别的具体数量如表 1 所示。同时为了确保数据集能够真实反映实际应用场景, 目标在图像中被随机且均匀地分布在不同的位置, 并且模特的性别比例均衡。在实验中, 数据集按照 8:2 随机的划分为训练集以及验证集。

Table 1. Number of samples for each category in the dataset

表 1. 数据集中不同类别的具体数量

类别	刀类	爆炸物	液体	枪	打火机	其它	总计
数量	1422	1008	934	858	500	778	5500

4.2. 结果及分析

4.2.1. 实验细节

算法的实验环境是基于 Python 3.8 下实现, 使用了 PyTorch 1.13.1 版本的深度学习框架。训练、验证

和测试过程在一对 NVIDIA GeForce RTX 3090 GPU 上进行, CPU 型号为 AMD R9-3900X, 操作系统为 Ubuntu 22.04。在训练参数方面, 训练的轮次设置为 100 轮, 初始学习率为 0.01, 使用 SGD 作为优化器和自动混合精度(AMP)加速训练, 前 3 轮用于预热, 学习率在最后一轮衰减为 $1e-4$ 。数据增强方面, 除了传统的色调、饱和度和亮度的颜色增强和图像的平移和缩放, 还使用了马赛克增强来增加数据集的多样性, 并于最后 10 轮关闭马赛克增强。考虑到主干结构的修改, 网络没有使用 YOLOv8 的预训练模型。同时为了性能和推理速度之间的平衡, 我们将深度系数和宽度系数设置为 0.33 和 0.5, 使用 YOLOv8 中的 small 模型进行实验。

在这项研究中, 本文使用精确度(Precision)、召回率(Recall)以及 mAP (Mean Average Precision)来评估模型的准确性。精确度衡量正确预测的比例, 即模型在所有预测中有多少是准确的; 召回率表示正确识别的类别占实际类别总数的比例, 反映模型的识别能力。mAP@0.5 则通过计算交并比(IoU)阈值为 0.5 时, 各类别的平均精度(AP), 用于评估模型在特定阈值下的整体性能。

4.2.2. 实验结果分析

为了验证本文方法的有效性, 本文通过大量实验对比了多种主流的图像分割方法在毫米波分割数据集上的表现。这些方法包括图像分割领域的经典算法, 如 Mask RCNN [21]、Cascade RCNN [22]、QueryInst [23], 以及单阶段算法 Yolov5、原始的 Yolov8 和 Yolov9 [24]。为了保证公平性, 所有的模型均采用预训练模型, 同时在 box AP、mask AP, 以及模型的参数量多个方面进行了全面对比分析。结果如表 2 所示:

Table 2. Comparison of different methods on mm wave segmentation datasets

表 2. 在毫米波分割数据集上不同方法之间的比较

模型	骨干网络	参数量	Box mAP@0.5	Mask mAP@0.5
Mask RCNN	R-50-FPN	44M	60.4	60.9
Cascade RCNN	R-50-FPN	77.04M	64.3	63.2
QueryInst	R-50-FPN	176.12M	74.4	72.6
Yolov5	CSPDarknet53 PAN	9.32M	88.4	78.1
Yolov8	CSPNet + PAN	11.24M	89.8	79.0
Yolov9	CSPELAN PAN	8.24M	87.3	75.4
本文	-	10.78M	92.4	89.5

实验结果表明, 本文方法在毫米波分割数据集上的表现有了显著提升, 尤其是在 mAP@0.5 性能指标上取得了显著进步。与原始 YOLOv8 相比, 本文方法的 mAP@0.5 提高了 2.6%, mAP@0.5 更是大幅提升了 10.4%。在参数量可控的情况下实现最优性能。

在图 7 中, 本文展示了基于不同方法的多种可视化测试的结果。可视化结果表明: 传统模型对“相对较大、边缘清晰”的目标表现良好, 但对“体积较小、低频出现”的物体, 存在较大随机性, 对于模糊和小型物体, 易出现漏检、误分类与轮廓偏差; 本文方法在小目标与遮挡目标上分割更完整、边缘更精准。

4.2.3. 消融实验

为了评估模型改进部分对性能的具体影响, 本文通过三个不同的实验设置, 对模型的改进策略进行了逐步验证。所有实验均在相同的训练参数下进行, 并使用 FP16 精度和 batch=1 进行推理测试。结果如表 3 所示:

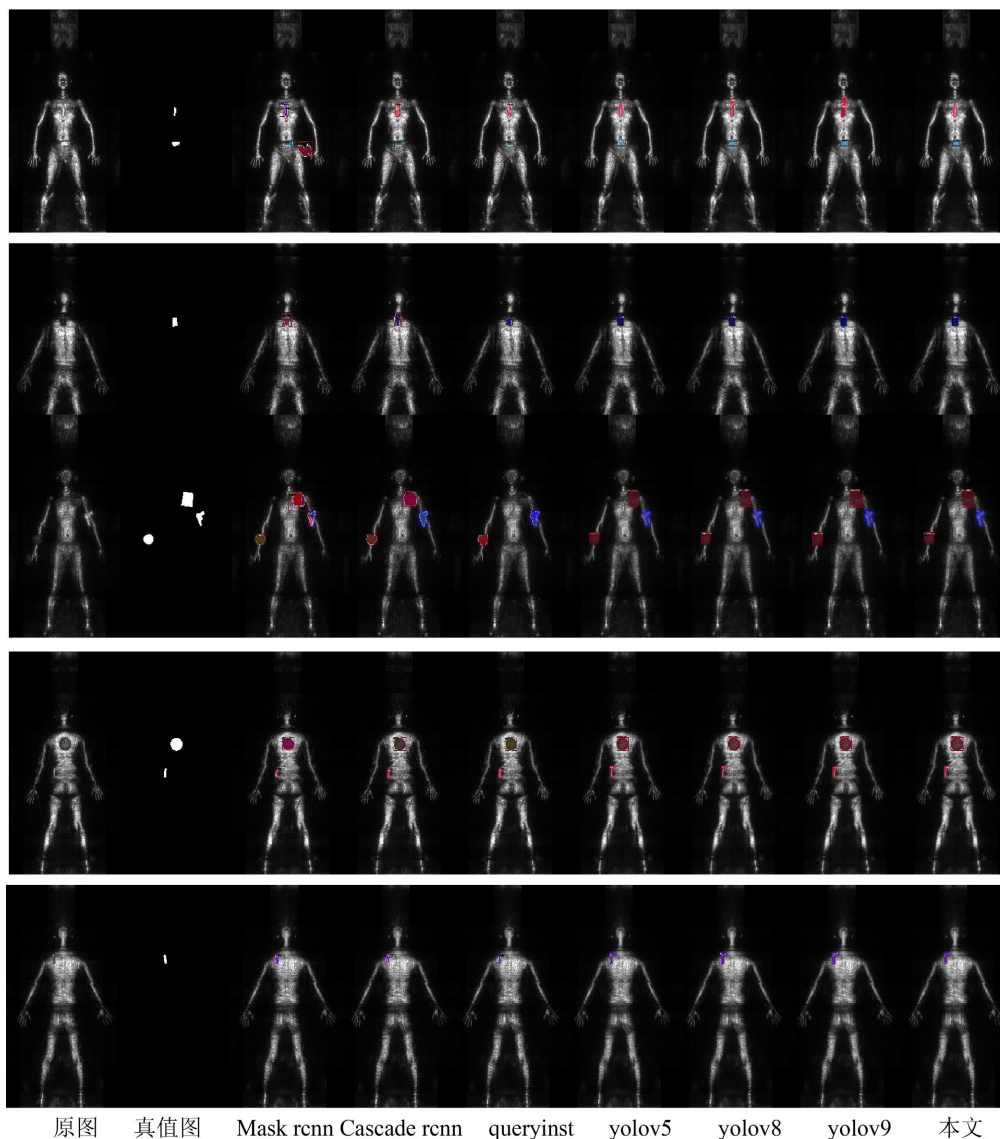


Figure 7. Visualization results for multiple instances using different methods
图 7. 不同方法对多个实例可视化结果

Table 3. Ablation results on mm wave segmentation datasets
表 3. 毫米波分割数据集消融实验结果

模型	参数量/M	复杂度/GFLOPs	Box mAP@0.5	Mask mAP@0.5	延迟/ms	FPS
Yolov8	11.24	42.4	89.8	79.0	7.89	126.6
+P2	10.71	87.4	90.6	86.0	10.89	91.8
+RFCACnv	10.77	87.8	91.8	88.1	12.69	78.8
+ATSS	10.77	87.8	92.4	89.5	12.69	78.8

首先，测试了增加 P2 层对模型的影响。结果显示，尽管计算复杂度和推理时间有所提升，但带来的是 mAP 显著上升，尤其是在小目标检测方面表现突出。这表明引入 P2 层显著提升小目标检测与分割精

度。其次，我们引入了 RFCACnv 模块来增强特征提取能力。消融实验结果表明，该模块在引入后，仅对计算成本产生轻微影响，却带来了 2.1% 的 Mask mAP 提升。表明 RFCACnv 轻量高效，大幅增强轮廓特征。最后，实验评估了 ATSS 策略的影响。引入 ATSS 后，对模型的推理速度和复杂度没有任何影响，Mask mAP 却提升了 1.4%。这一结果证明，ATSS 的标签分配有助于模型更准确地选择训练样本，进而提高特征的判别能力，增强了对目标的检测性能。

总的来说，这些改进共同提升了模型的整体性能，尽管模型复杂度有所增加，但检测速率依然达标。经过这些改进后，模型在复杂场景中的精度显著提高，尤其是在需要精确分割和检测小目标的应用中表现优异。此外，召回率和精确率被视为毫米波检测的重要评估指标，它们分别反映了模型在检测中的灵敏度和准确性。为了深入理解模型改进的影响，本文以 IOU = 0.5 为基准，对分割的精确率以及召回率进行了详细分析。结果如表 4 所示：

Table 4. Detected results for different targets of mm wave segmentation datasets

表 4. 毫米波分割数据集不同目标检测结果

类别	YOLOv8			本文			Δ
	P	R	Mask AP@0.5	P	R	Mask mAP@0.5	Δ Mask AP@0.5
刀类	60.5	58.7	55.9	77.1	70.1	72.8	+16.9
爆炸物	94.4	88.9	95.3	96.1	86.9	95.3	-
液体	76.9	62.9	66.9	91.5	76.6	88.5	+21.6
枪	94.5	90.4	95.3	92.7	91.5	95.5	+0.2
打火机	73.6	63.8	67.6	90.8	77.7	87.1	+19.5
其它	95.1	92.6	93.2	98.4	95.3	97.5	+4.3

经过改进后的模型在总体召回率和精确率上均有所提升。特别是在处理难以检测的小目标时，召回率的显著提升显示了改进策略在提高检测范围上的优势。具体到各个类别，改进后的模型对不同目标的表现有所不同。对于大多数类别，如爆炸物、枪和其它(包含手机等金属物品)，平均精度值均有小幅提升，这表明模型改进策略在这些类别上的表现更加稳健和一致。然而，对于像打火机和藏在衣物下的液体和刀这些小像素目标类别，召回率和精确率的提升尤为显著。打火机的精确率从 63.8% 跃升至 90.8%，刀类的精确率从 60.5% 升至 77.1%，液体的精确率则从 76.9% 提升至 91.5%。这些结果表明，模型改进策略对毫米波图像核心难点针对性极强。

4.3. 在公开数据集实验

为了进一步验证我们方法的有效性，我们在毫米波公开数据集上进行了额外的实验。在原文献[25]中，作者探索了多种先进的方法，并采用了基于 RetinaNet + P2 的方法以获得最高性能。在文献[6]中，Su 等人提出的 YOLOv8-wavelet 方法也在该数据集上取得了显著的成绩，达到了 66.53 的精度。为了确保与这些方法的公平比较，本文遵循相同的数据集划分，并直接引用原文献中提供的结果，仅针对边界框结果进行分析，并使用原文献中的 mAP@0.5 作为评价指标，取得结果如表 5 所示，本文的方法较当前最先进的的方法在 mAP 上提高了约 3.35%，这一显著的提升验证了本文提出的方法在不同场景中的稳定性和泛化能力。

Table 5. Results on public millimeter-wave datasets
表 5. 毫米波公开数据集结果

模型	骨干网络	Box mAP@0.5
YOLOv3	Darknet-53	39.29
YOLOv4	CSPDarknet53	41.39
FRCN-OHEM	VGG16	42.32
RetinaNet	ResNet-50	54.58
RetinaNet	ResNet-101	47.99
RetinaNet+P2	ResNet-50	60.32
YOLOv8	CSPNet+	58.96
YOLOv8-Wavelet	-	66.53
本文	-	69.88

5. 结论

本文提出了一个基于 YOLOv8 的毫米波分割优化集成模型,通过感受野注意力卷积操作、增加 P2 小目标特征层、自适应阈值的标签分配三项改进,有效解决毫米波图像噪声高、小目标多、轮廓难分割等问题。其中, RFCACnv 模块在几乎不影响计算成本的情况下,通过捕捉目标的空间位置信息以提升特征能力的表达;引入 P2 层,进一步细化特征金字塔以增强对小目标的检测能力,确保在不同尺度上捕捉到更为细腻的目标特征,从而提升检测精度;自适应阈值的标签分配策略,针对小目标在 IoU 敏感性上的特性,以动态调整阈值,优化正负样本的匹配过程。经过改进的模型在性能评估中表现出显著提升,其中 box mAP50 提升了 2.6%,而 mask mAP50 则提高了 10.4%,在实际应用场景中具备高精度、高实用性的应用价值。

参考文献

- [1] Du, K., Wang, W., Nian, F., *et al.* (2016) Concealed Objects Detection in Active Millimeter-Wave Images. *Systems Engineering and Electronics*, **38**, 1462-1469. (In Chinese).
- [2] Haworth, C.D., De Saint-Pern, Y., Clark, D., Trucco, E. and Petillot, Y.R. (2006) Detection and Tracking of Multiple Metallic Objects in Millimetre-Wave Images. *International Journal of Computer Vision*, **71**, 183-196. <https://doi.org/10.1007/s11263-006-6275-8>
- [3] Ren, S.Q., He, K.M., Girshick, R.B., *et al.* (2015) Faster R-CNN: Towards Realtime Object Detection with Region Proposal Networks. *Proceedings of the International Conference on Neural Information Processing System*, Istanbul, Turkey, 9-12 November 2015, 91-99.
- [4] Liu, T., Zhao, Y., Wei, Y., Zhao, Y. and Wei, S. (2019) Concealed Object Detection for Activate Millimeter Wave Image. *IEEE Transactions on Industrial Electronics*, **66**, 9909-9917. <https://doi.org/10.1109/tie.2019.2893843>
- [5] Pang, L., Liu, H., Chen, Y. and Miao, J. (2020) Real-Time Concealed Object Detection from Passive Millimeter Wave Images Based on the Yolov3 Algorithm. *Sensors*, **20**, Article 1678. <https://doi.org/10.3390/s20061678>
- [6] Su, Y., Tan, W., Dong, Y., Xu, W., Huang, P., Zhang, J., *et al.* (2024) Enhancing Concealed Object Detection in Active Millimeter Wave Images Using Wavelet Transform. *Signal Processing*, **216**, Article ID: 109303. <https://doi.org/10.1016/j.sigpro.2023.109303>
- [7] Jocher, G., Chaurasia, A. and Qiu, J. (2023) YOLO by Ultralytics. <https://github.com/ultralytics/ultralytics>
- [8] Zhang, X., Liu, C., Yang, D., Song, T., Ye, Y., Li, K. and Song, Y. (2023) RFACnv: Innovating Spatial Attention and Standard Convolutional Operation. arxiv: 2304.03198.
- [9] Zhang, S., Chi, C., Yao, Y., Lei, Z. and Li, S.Z. (2020) Bridging the Gap between Anchor-Based and Anchor-Free Detection via Adaptive Training Sample Selection. 2020 *IEEE/CVF Conference on Computer Vision and Pattern*

- Recognition (CVPR)*, Seattle, 13-19 June 2020, 9756-9765. <https://doi.org/10.1109/cvpr42600.2020.00978>
- [10] Jocher, G., *et al.* (2020) Yolov5. <https://github.com/ultralytics/yolov5>
- [11] Wang, C., Mark Liao, H., Wu, Y., Chen, P., Hsieh, J. and Yeh, I. (2020) CSPNet: A New Backbone That Can Enhance Learning Capability of CNN. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, 14-19 June 2020, 1571-1580. <https://doi.org/10.1109/cvprw50498.2020.00203>
- [12] Elfving, S., Uchibe, E. and Doya, K. (2018) Sigmoid-weighted Linear Units for Neural Network Function Approximation in Reinforcement Learning. *Neural Networks*, **107**, 3-11. <https://doi.org/10.1016/j.neunet.2017.12.012>
- [13] Liu, S., Qi, L., Qin, H., Shi, J. and Jia, J. (2018) Path Aggregation Network for Instance Segmentation. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 8759-8768. <https://doi.org/10.1109/cvpr.2018.00913>
- [14] Lin, T., Dollar, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 936-944. <https://doi.org/10.1109/cvpr.2017.106>
- [15] Cheng, G., Wang, J., Li, K., Xie, X., Lang, C., Yao, Y., *et al.* (2022) Anchor-Free Oriented Proposal Generator for Object Detection. *IEEE Transactions on Geoscience and Remote Sensing*, **60**, 1-11. <https://doi.org/10.1109/tgrs.2022.3183022>
- [16] Zheng, Z., Wang, P., Ren, D., Liu, W., Ye, R., Hu, Q., *et al.* (2022) Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation. *IEEE Transactions on Cybernetics*, **52**, 8574-8586. <https://doi.org/10.1109/tcyb.2021.3095305>
- [17] Li, X., Wang, W., Wu, L., *et al.* (2020) Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection. *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Vancouver, 6-12 December 2020, 21002-21012.
- [18] Mao, A., Mohri, M. and Zhong, Y. (2023) Cross-Entropy Loss Functions: Theoretical Analysis and Applications. *Proceedings of the 40th International Conference on Machine Learning*, Honolulu, 23-29 July 2023, 23803-23828.
- [19] Micikevicius, P., Narang, S., Alben, J., *et al.* (2017) Mixed Precision Training. arXiv: 1710.03740.
- [20] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., *et al.* (2023) Segment Anything. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 3992-4003. <https://doi.org/10.1109/iccv51070.2023.00371>
- [21] He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017) Mask R-CNN. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 2980-2988. <https://doi.org/10.1109/iccv.2017.322>
- [22] Cai, Z. and Vasconcelos, N. (2018) Cascade R-CNN: Delving into High Quality Object Detection. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 6154-6162. <https://doi.org/10.1109/cvpr.2018.00644>
- [23] Fang, Y., Yang, S., Wang, X., Li, Y., Fang, C., Shan, Y., *et al.* (2021) Instances as Queries. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 6890-6899. <https://doi.org/10.1109/iccv48922.2021.00683>
- [24] Wang, C., Yeh, I. and Mark Liao, H. (2024) YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. arXiv: 2402.13616.
- [25] Liang, D., Xue, F. and Li, L. (2021) Active Terahertz Imaging Dataset for Concealed Object Detection. arXiv: 2105.03677.