

基于 YOLOv5 架构改进的电磁信号目标识别方法

卢河旭¹, 陈丹伟¹, 罗圣美²

(1. 南京邮电大学 计算机学院, 南京 210023; 2. 中孚信息股份有限公司, 南京 211899)

摘要: 针对在时间及频段上重叠的多种电磁信号中识别出目标信号的问题, 提出了一种基于 YOLOv5 架构改进的电磁信号目标识别方法; 通过改进 YOLOv5 算法以提高对电磁信号时频图像上目标信号的检测精度; 采用 MobileNetV3 作为主干网络, 以实现骨干网络轻量化的同时平衡精度和速度; 通过在特征金字塔网络模块中引入自适应空间特征融合 (ASFF) 模块来将不同尺度的特征进行融合, 调整损失函数, 从而增强网络对不同尺度目标的感知能力; 将改进 YOLOv5 算法在自制电磁信号时频图像数据集上与原始 YOLOv5 算法和 SSD 算法进行对照试验; 实验表明, 改进后的方法对时频图像上目标信号的检测能力有明显增强, 具有一定的实用价值。

关键词: 深度学习; 目标检测; 电磁信号识别; 自适应空间特征融合; MobileNet

An Improved Recognition Method for Electromagnetic Signal Targets Based on YOLOv5 Architecture

LU Hexu¹, CHEN Danwei¹, LUO Shengmei²

(1. School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China;

2. Zhongfu Information Inc., Nanjing 211899, China)

Abstract: To address the problem of identifying target signals from multiple overlapping electromagnetic signals in both time and frequency domains, an electromagnetic signal target recognition method based on an improved YOLOv5 architecture is proposed. The improved YOLOv5 algorithm is used to increase the detection accuracy of target signals in time-frequency images of electromagnetic signals. The MobileNetV3 is used as the backbone network to achieve a balance between precision and speed with the backbone network lightweight. The adaptive spatial feature fusion (ASFF) module is introduced into the feature pyramid network module to merge the features of different scales and adjust the loss function, thereby enhancing the network's ability to perceive various scale targets. The improved YOLOv5 algorithm is compared with the original YOLOv5 algorithm and the SSD algorithm on a custom electromagnetic signal time-frequency image dataset. Experimental results show that the improved method significantly enhances the detection capability of target signals in time-frequency images, demonstrating a certain practical value.

Keywords: deep Learning; target detection; electromagnetic signal recognition; adaptive spatial feature fusion; MobileNet

0 引言

通信技术的进步不断推动无线通信系统向着更高的数据传输速度发展。然而, 频谱资源的固有限制与日益增加的需求之间的矛盾愈发显著。无线设备的广泛使用以及技术多样性的扩展进一步加剧了对频谱资源的需求^[1]。这要求通信技术不仅在提高数据传输效率方面取得进展, 同时也必须通过新的方式优化频谱资源的管理和利用。频谱资源作为一种有限和珍贵的资源, 目前一部分频谱过度使用, 而另一部分则未充分利用, 显示出资源配置的不均衡性^[2]。

在充斥着众多电磁信号的无线环境下, 通过识别并操作特定频段中的异常信号, 可以有效执行对无线频谱资源的管理。电磁信号的准确识别是无线频谱管理工作的基础和前提。然而, 与早期相比, 当前的信号环境已变得更加复杂。在过去, 电磁信号来源较少, 工作频率相对稳定, 信号类型简单, 使得信号的分析 and 识别相对容易。辐射源的用途多种多样, 例如雷达与通讯系统各分别在特定的频率上运行, 通常频率相差较大故而易于识别。然而, 随着科技的不断发展进步, 这些辐射源的总数以及所使用的信

收稿日期: 2024-05-17; 修回日期: 2024-06-18。

基金项目: 江苏省重点研发项目 (BE2022065-5)。

作者简介: 卢河旭 (1995-), 男, 硕士研究生。

陈丹伟 (1970-), 男, 博士, 教授。

罗圣美 (1970-), 男, 博士, 教授。

引用格式: 卢河旭, 陈丹伟, 罗圣美. 基于 YOLOv5 架构改进的电磁信号目标识别方法 [J]. 计算机测量与控制, 2024, 32(11): 243-250, 294.

号模式和频带日渐增加,致使原本分离的频率开始出现交叠。这导致在同一频带内更容易频繁遇到多重信号,进一步加剧了信道环境的复杂性。信号识别技术需要区分混杂环境下的各种信号,从而使接收设备能够更有效地对目标信号进行分析和处理^[3]。因此,开发精确且高效的信号识别技术对于应对当下及未来的频谱管理挑战显得至关重要。

通过短时傅里叶变换(STFT, short-time Fourier transform)可以将一维的时域电磁信号映射到时间和频率的二维平面^[4],从而更直观的反应电磁信号的时频联合特征^[5]。通过将电磁信号的时频特征表示在二维平面上,使得采用目标检测技术来进行电磁信号的检测识别成为可能。

作为深度学习领域中最具有代表性的模型,卷积神经网络(CNN, convolutional neural network)^[6]能够自动学习图像的高层特征,从而显著提高了检测的准确性和鲁棒性。AlexNet^[7]的成功应用证明了深度学习在视觉任务中的巨大潜力,随后更多的研究开始集中于如何优化和改进CNN模型,以适应不同的目标检测需求。2014年R-CNN^[8]算法将深度学习引入目标检测,通过先提取图像中的候选区域,再用CNN提取特征进行分类。此后, Fast R-CNN^[9]和 Faster R-CNN^[10]相继提出,提高了检测速度和精度。2015年YOLO^[11](YOLO, You Only Look Once)算法被提出,作为一种端到端的目标检测技术,YOLO将目标检测任务视为一个单一的回归问题,直接从图像像素映射到边界框坐标和类别概率。随后在YOLO算法基础上提出了YOLOv2^[12], YOLOv3^[13], YOLOv4^[14], YOLOv5等算法。其中YOLOv5算法吸收了前几代算法的优点取得了更好的检测速度和精度以及泛化能力。

近年来,随着深度学习技术在图像识别,目标检测和自然语言处理等领域的应用取得显著成果,许多研究人员选择使用深度学习技术来解决信号检测的问题。文献[15]提出了一种使用卷积递归神经网络(CRNN, Convolutional Recurrent Neural Network)来实现实际环境中的自动鸟类音频检测。该方法通过卷积层来提取高维的局部频移不变特征,递归层捕捉从短时间帧中提取特征之间的长期依赖性,取得了较好的鸟类音频检测效果。文献[16]提出了一种名为BN3的新型卷积神经网络,通过在输入和卷积层引入批量归一化以减轻过拟合,并在卷积层使用ReLU加速训练,从而提高P300信号的检测性能,并取得了优秀的检测性能。文献[17]使用YOLO算法在时频谱内检测和定位无线电信号,使用包含带有掩码和边界框注释的宽带频谱的标记训练数据来训练高效的无线电信号检测器,与传统的能量门限信号检测方案相比,该检测器在精度和灵活性方面都取得了更好的成绩。文献[18]使用Faster R-CNN的方法对在复杂背景噪声中精确检测微弱信号,在保证检测精度的同时,相比传统方案实现了更快的检测。

本文基于YOLOv5结构进行改进,引入了MobileNetV3^[19]作为主干网络提取特征以实现骨干网络轻量化的同时平衡精度和速度。为了提升网络对不同尺度目标

的感知能力,引入自适应空间特征融合(ASFF, Adaptively Spatial Feature Fusion)^[20]技术用于融合特征金字塔的不同尺度特征。同时采用了CIoU损失函数,该函数从重叠面积、中心点距离和预测边界框与真实边界框的长宽比3个方面,使损失函数更加关注边界框的形状,更加精确地描述了目标框的回归。

1 YOLOv5 算法结构概述

YOLOv5是一种单阶段的目标检测算法,主要由Backbone网络, Neck特征提取网络以及Head检测头3部分构成。如图1所示。

1.1 Backbone 主干网络

在YOLOv5中, Backbone的核心作用是从输入图像中提取特征,为后续的目标检测任务提供必要的视觉信息。YOLOv5的主干网络以CSPDarknet为基础,它是一种融合了多种技术的深度卷积网络。它起源于经典的Darknet架构,并融合了跨阶段局部网络(CSPNet, Cross Stage Partial Network)^[21]的设计,以实现结构上的优化。

CSPNet本身是为了增强CNN的性能而设计的一种结构。它的核心设计理念是在网络的特定阶段将特征图分割成两部分。一部分特征图直接传递到下一阶段,无需额外处理,形成一种“残差连接^[22]”,这有助于信息在网络中的高效流动。而另一部分特征图则会经过一系列卷积层进行深入的特征加工处理。在这些卷积层处理后,两部分特征图会在深度方向上重新合并,这样的结构不仅有助于降低计算复杂度,还能在减少参数数量的同时,保持或甚至提升网络的精度和处理效率。这种设计方式使得YOLOv5能够更有效地处理图像中的信息,尤其是在需要快速准确地进行目标检测的场景中。通过采用CSPNet的策略, YOLOv5的Backbone可以优化其性能,实现对复杂环境中多样目标的快速识别和准确检测。此外,通过这种分阶段处理和特征合并的方法, YOLOv5在处理不同尺寸和类型的目标时,显示出了更高的灵活性和适应性。

1.2 Neck 网络

YOLOv5的Neck网络位于Backbone和检测头之间,主要负责特征融合和增强,以提升检测性能。YOLOv5中采用了FPN^[23]和PAN^[24]的混合网络结构作为其颈部,旨在有效融合不同尺度的特征图,以改善对不同大小目标的检测能力。

FPN利用了深层网络中不同分辨率的特征图,通过自顶向下和横向连接的方式,构建了一个特征金字塔,将深层的语义特征传递到浅层,增强了语义信息,这样就能同时利用不同层次的特征进行目标检测。PAN是对FPN的补充,在FPN的后面添加了一个自底向上的金字塔结构,从而将浅层的强定位特征传递到深层来增强表达能力,从而提升对于小目标的检测性能。

1.3 Head 检测头

YOLOv5的检测头通过多个输出层来实现多尺度的预测,每个输出层对应不同大小的特征图用于检测不同尺寸

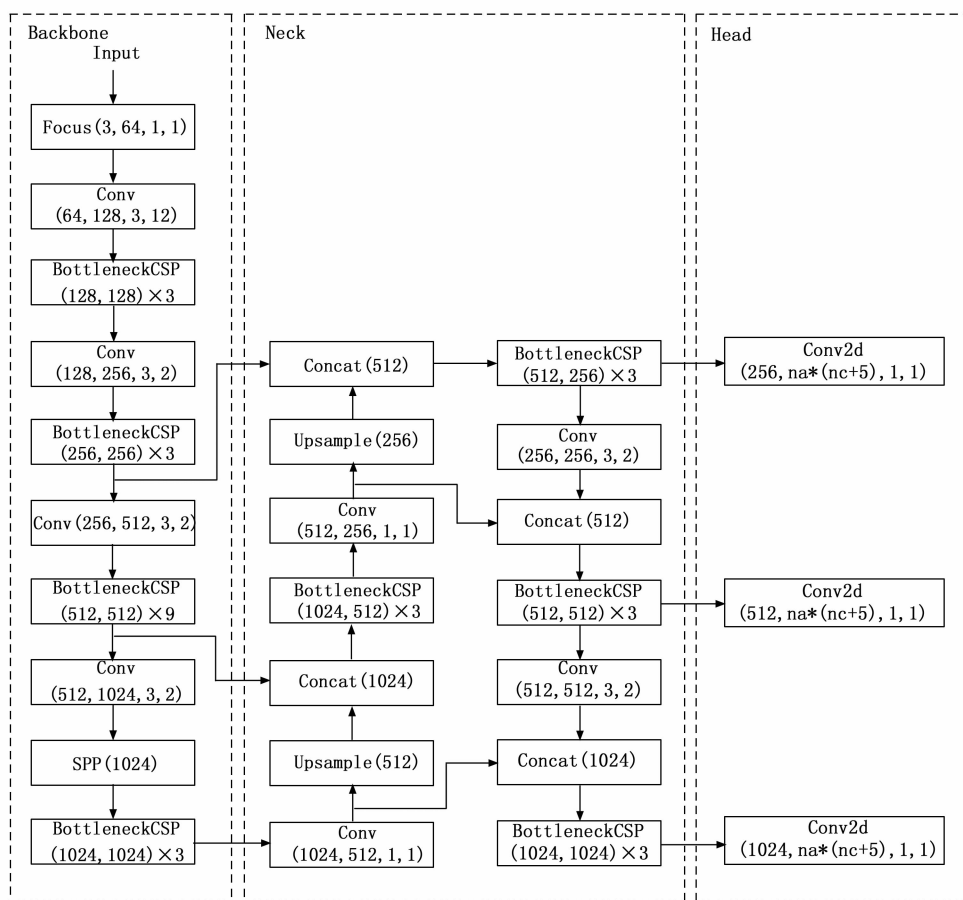


图 1 YOLOv5 结构图

的对象。输出层包含对于每个目标的边界框坐标, 区域所含有目标的置信度以及检测到目标的所属类别概率。同时 YOLOv5 使用结合了边界框回归损失、对象置信度损失和分类损失的复合损失函数, 从而实现对不同任务目标的优化, 使模型能够有效地学习在复杂场景中准确检测和分类不同的目标。

2 改进的 YOLOv5 算法

在使用宽带 IQ 数据进行信号检测时, 通常需要在频域内进行连续的带通滤波操作以切割频带, 从而获得各个分频段的 IQ 数据。随后研究者需要利用特定的检测算法来评估这些频段内的信号存在与否。整个流程包括频段的精细切割、自适应滤波组的设计以及带通滤波处理等步骤, 每一步都具有高度的技术复杂性并且对计算资源的需求极大, 这对实现实时和快速的信号检测带来了挑战。同时, 基于 IQ 数据的检测方法在确定样本内部信号的时域占用状况方面也存在一定困难, 因为信号可能在采集期间出现间断, 这使得使用单一样本很难准确判断信号的持续存在性。

作为具备历史数据追踪能力的时频谱图, 具有通过利用之前的输入来帮助分析当前信号状态的能力, 从而辅助进行信号的检测。此技术能够在时频谱图上框选标定出特

定的信号, 直观明了的显示所框选的信号在时间和频率上的占用情况, 从而提供时间和频率上双重维度的信息。此外, 时频分析技术同时还具有处理迅速且能即时反馈的特点。

针对以上挑战, 本文结合实际需求, 选用时频图像作为输入, 并结合目标检测技术, 提出了一种基于改进 YOLOv5 的电磁信号检测方法。实现了同时检测信号中所包含的多个目标信号, 并直观地展示这些信号的时域和频域信息, 这样的方法不仅提高了检测效率, 而且增强了信号分析的准确性和实用性。

本文在 YOLOv5 网络结构的基础上, 采用了更轻量化的 MobileNetV3 来替换原本的 CSPDarknet 主干网络, 在模型能够在保持较高准确度的同时减少了主干网络部分的参数量和计算复杂度。为了提升模型对于多尺度目标的检测能力, 在 Neck 部分引入了 ASFF 方法学习不同尺度特征图之间的联系来解决原本特征金字塔所输出的不同尺度特征图之间的不一致性的问题。采用 CIUO 作为损失函数, 该函数考虑了目标框之间的完整交叉, 并引入了修正因子, 以更精准地度量目标框之间的相似性, 使得模型在训练过程中能够更好地理解目标框的准确位置和形状。改进后的整体模型结构如图 2 所示。

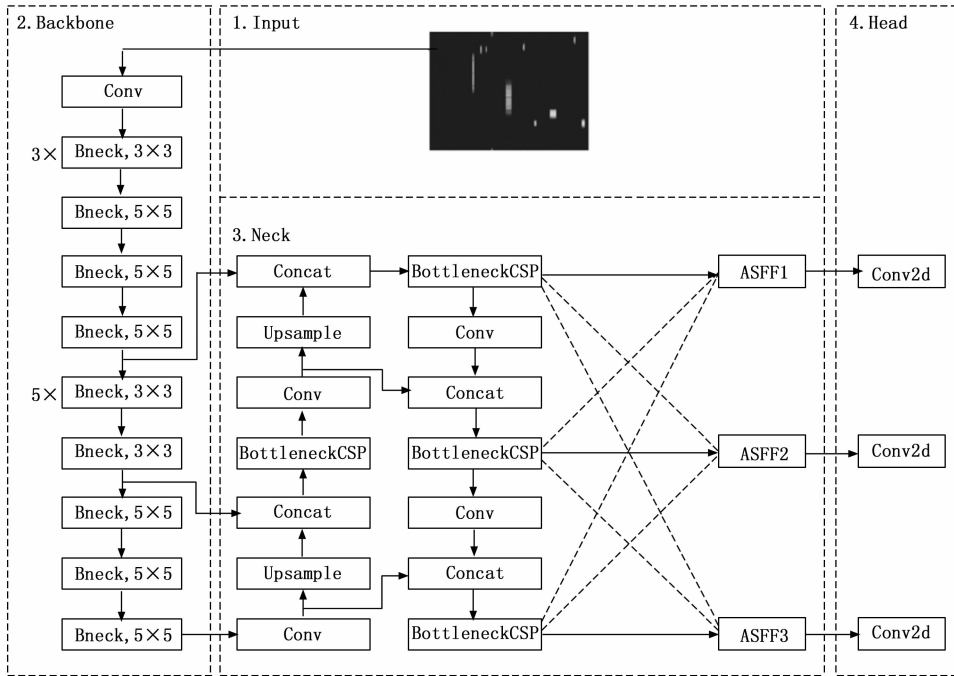


图 2 改进后的 YOLOv5 模型结构

下面分别对采用 MobileNetV3 作为主干网络、在 Neck 部分引入 ASFF 结构、采用 CIoU 作为损失函数这三项改进进行详细介绍。

2.1 MobileNetV3 网络结构

MobileNet 轻量化网络^[25]于 2017 年由谷歌提出，通过引入深度可分离卷积技术，将模型的参数数量及计算复杂度有效减少。MobileNetV2^[26]在继承了前代深度可分离卷积技术的基础上引入了通道扩充技术，倒残差结构和线性激活函数，在保持了模型轻量化的同时提高了对特征的提取能力。MobileNetV3 在 MobileNet 和 MobileNetV2 基础上进一步进行改进，在使用引入了轻量化注意力模块和新的 h-swish 激活函数，进一步提升了模型的性能。

2.1.1 Bneck 结构

Bneck 结构作为 MobileNetV3 的核心模块，主要实现了通道可分离卷积，SE 通道注意力模块以及残差结构的功能。结构如图 3 所示。

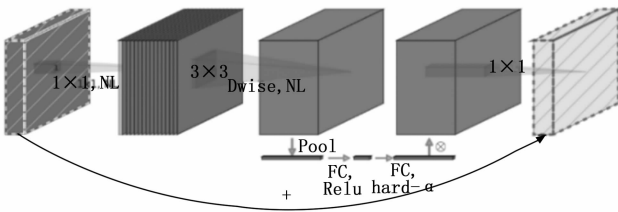


图 3 Bneck 结构图

输入特征图首先被分为两部分，一部分不经过任何额外处理作为残差边，另一部分首先经过 1×1 卷积扩展通道维数；然后扩展后的特征图经过深度可分离卷积处理；接

着对经过深度可分离卷积处理的特征图应用 SE 注意力机制，为特征图每个通道赋予权重，最后使用 1×1 卷积调整通道数并与残差边特征相加作为输出。

2.1.2 SE 注意力模块

SE 注意力机制^[27]是一种在深度学习领域中被广泛应用的技术，特别是在图像识别和分类任务中表现突出。该机制的核心思想是通过显著增强网络对特征通道重要性的感知能力，从而优化整个网络的性能。SE 注意力机制的工作流程可以分为两个主要阶段：压缩和激励。首先，在压缩阶段，网络会对每个通道的特征图进行全局平均池化，这一操作将每个二维特征图简化为一个单一的标量值，从而捕获通道级的全局信息。接着，在激励阶段，这些标量值被送入一个全连接层，该层通过学习不同通道间的复杂依赖关系，输出每个通道的重要性权重。通过这样的权重分配，SE 机制能够动态地调整各个通道的权重，使得模型在前向传播过程中更加关注于那些被认为是“更重要”的特征。这种选择性的关注显著提高了网络的表达能力，因为它允许模型侧重于对最终任务最为关键的信息，从而在各种视觉任务中实现更精准的识别和分类。结构如图 4 所示。

输入特征图首先进行全局池化变为一个 C 通道的 1×1 特征，该特征具有全局信息，然后对全局池化后的特征进行全连接变为 C/r 通道的 1×1 特征并使用 ReLU 激活，之后再进行一次全连接使特征变回 C 通道的 1×1 特征，最后使用 sigmoid 激活得到通道权重矩阵。

2.1.3 MobileNetV3 作为主干网络

本文通过采用经过修改的 MobileNetV3 网络来替换之

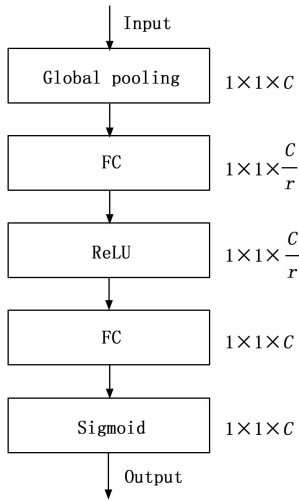


图 4 SE 注意力模块结构图

前的 CSPDarknet 主干网络, 从而使得主干网络模型在保持较好的特征提取能力的同时降低了网络的参数量以及计算的复杂度。经修改后的 MobileNetV3 结构参数如表 1 所示: 表中 Num 表示序号, Type 表示模块的类型, Exp size 表示倒残差结构中升维卷积核个数, out 表示倒残差结构中降维卷积核个数, Output 表示输出特征尺寸。

表 1 修改后的 MobileNetV3 结构参数表

Num	Type	Exp size	# out	Output
1	Conv2d	—	16	$512^2 \times 16$
2	Bneck, 3×3	16	16	$512^2 \times 16$
3	Bneck, 3×3	64	24	$256^2 \times 24$
4	Bneck, 3×3	72	24	$256^2 \times 24$
5	Bneck, 5×5	72	40	$128^2 \times 40$
6	Bneck, 5×5	120	40	$128^2 \times 40$
7	Bneck, 5×5	120	40	$128^2 \times 40$
8	Bneck, 3×3	240	80	$64^2 \times 80$
9	Bneck, 3×3	200	80	$64^2 \times 80$
10	Bneck, 3×3	184	80	$64^2 \times 80$
11	Bneck, 3×3	184	80	$64^2 \times 80$
12	Bneck, 3×3	480	112	$64^2 \times 112$
13	Bneck, 3×3	672	112	$64^2 \times 112$
14	Bneck, 5×5	672	160	$32^2 \times 16$
15	Bneck, 5×5	960	160	$32^2 \times 16$
16	Bneck, 5×5	960	160	$32^2 \times 16$

使用修改后的 MobileNetV3 主干网络, 输入大小为 $1\ 024 \times 1\ 024$ 像素的图像, 为了进行有效的特征提取和传递, 模型在主干网络的不同阶段选择了特定的 Bneck 模块的输出, 以供后续的 Neck 网络使用。分别取主干网络中第 6 个 Bneck 模块的输出: 此模块的输出特征图大小为 128×128 像素, 具有 40 个通道。这个尺寸和通道数的特征图能够提供较高分辨率的特征, 这对于检测较小的对象特别有用。第 12 个 Bneck 模块的输出: 输出特征图大小为 64×64 像

素, 带有 112 个通道。此级别的输出在分辨率和通道深度之间提供了一个平衡, 使得网络能够有效地处理中等大小的对象。第 15 个 Bneck 模块的输出: 最后, 这一模块的输出特征图尺寸为 32×32 像素, 拥有 160 个通道。较低的空间分辨率和较高的通道数输出适合捕捉更大尺寸的特征。

这 3 个输出点被选为进入 Neck 网络的输入, 主要是因为它们各自的特征图大小和通道数量能够代表不同层次的抽象, 适合多尺度的特征融合。在 Neck 部分, 这些多层次、多尺度的特征图将通过进一步的处理, 如特征融合技术, 来增强目标检测的精度和效率。这种设计确保了模型能够有效地检测不同大小的目标, 同时保持计算效率。

2.2 ASFF 结构

ASFF 是一种先进的特征融合方法, 主要应用于计算机视觉中的目标检测任务。其核心思想在于, 通过自适应地学习调整来自网络不同层次的特征融合权重, 使得模型能够根据不同的输入动态地进行权重分配和特征融合。这种自适应加权融合机制允许模型根据输入图像的具体内容和上下文, 优化特征使用, 从而提高信息利用效率和检测性能。具体来说, ASFF 策略通过引入可学习的融合参数来调整各层特征图的贡献比例。在实际操作中, 它考虑了来自网络底层的细粒度特征和高层的抽象特征, 并通过训练过程中不断优化这些特征的融合权重, 以适应不同尺度和复杂度的目标检测需求。这种方法尤其对于同时存在于图像中的小尺寸目标和大尺寸目标的情况表现出了极大的优势。

通过自适应调整特征融合的权重, ASFF 能够增强模型对小目标的检测能力, 这通常是传统目标检测模型中的一个挑战, 因为小目标可能在高层特征中丢失重要信息。同时, 对于大目标, ASFF 保持了足够的感受野和上下文信息, 确保检测的精度不会受损。因此, ASFF 提供的灵活性和适应性使其非常适合应用于各种尺寸和复杂度的目标检测任务中。ASFF 结构的引入提高模型的泛化能力和在复杂背景下的性能稳定性。在面对多变的环境和不同类型的干扰时, ASFF 能够通过优化特征融合策略, 减少误检和漏检的情况, 进一步提升整体的检测效率。

本文在 Neck 网络部分引入了 ASFF 模块用于提升模型检测能力。ASFF 模块的引入位于传统的 FPN 和 PAN 结构的输出之后, 主要目的是进一步优化特征融合的过程, 以增强模型对不同尺度目标的检测能力。具体地, FPN + PAN 结构已经通过多尺度的特征层来捕获从低到高不同级别的信息, 通过 ASFF 进一步增强这些特征层的融合是提高检测精度的关键。在 ASFF 模块中, 首先对 FPN + PAN 网络输出的 3 个不同层级的特征图进行尺寸和通道数的调整。这一步骤通过对较高分辨率的特征图进行下采样, 以及对较低分辨率的特征图进行上采样实现, 同时使用 1×1 的卷积核调整所有特征图的通道数, 确保它们在进行融合时维度一致。在调整后, 每个层级的特征图都会通过一个专门的网络层计算出一个权重向量。这些权重向量是基于网络学习到的对不同特征层的重要性评估, 使得 ASFF 模

块能够自适应地调整各层特征在融合过程中的贡献比重。随后，每个层级的特征图将与其相应的权重向量进行逐元素相乘，这一操作确保了重要的特征得到强调，而不那么重要的特征则被相应地抑制。最后，输出经自适应融合后含有所有层级信息的特征图用于后续的分类，改进后的结构如图 5 所示。

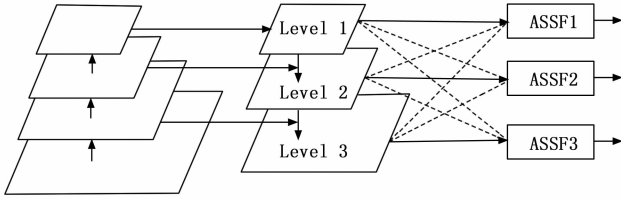


图 5 增加 ASFF 模块的 Neck 网络结构图

2.3 损失函数

YOLOv5 的损失函数是其训练过程的核心，负责衡量模型预测与实际标签之间的差异。为了训练模型以准确地预测目标的位置、大小和类别，YOLOv5 采用了一个复合损失函数，该损失函数由 3 个主要部分组成：边界框回归损失、对象置信度损失和分类损失。

对象置信度损失 L_{conf} 采用二元交叉熵损失来计算。这一部分的损失旨在优化模型对于目标存在与否的预测，从而确保模型能够准确识别出所包含目标的区域。计算公式如下：

$$Loss_{conf} = - \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{obj} [\hat{C}_i^j \log C_i^j + (1 - \hat{C}_i^j) \log (1 - C_i^j)] - \lambda_{noobj} \sum_{i=0}^{K \times K} \sum_{j=0}^M I_{ij}^{noobj} [\hat{C}_i^j \log C_i^j + (1 - \hat{C}_i^j) \log (1 - C_i^j)] \quad (1)$$

式中， K 表示网络最终输出的特征图被划分为 $K \times K$ 个单元格， M 表示每个单元格中对应的锚框数量， I_{ij}^{obj} 表示有目标的锚框， I_{ij}^{noobj} 表示无目标的锚框， λ_{noobj} 表示无目标锚框的置信度损失的权重系数。

分类损失 L_{class} 采用交叉熵损失来计算。这一部分的损失可以确保模型能准确分类所检测到的每个目标。计算公式如下：

$$Loss_{class} = - \sum_{i=0}^{K \times K} I_{ij}^{obj} \sum [\hat{P}_i^j \log p_i^j + (1 - \hat{P}_i^j) \log (1 - p_i^j)] \quad (2)$$

式中， p_i^j 为预测概率， \hat{P}_i^j 为网格中目标所属类别的实际概率。

原始 YOLOv5 使用的 GIoU 是一种改进的 IoU 计算方式，它不仅考虑了预测框与真实框的重叠区域，还包括了它们之间的非重叠区域，通过计算预测框和真实框的最小闭合区域来提高评估的全面性。尽管 GIoU 改进了 IoU 的一些限制，但它仍然主要关注真实框与预测框之间的重叠率，不足以很好全面描述目标框回归的问题。尤其是在预测框在真实框内部且预测框大小相同时，GIoU 将退化成 IoU，无法区分各预测框在空间上的相对位置。

本文使用了 CIoU 损失用于衡量预测的边界框（包括中

心点坐标、宽度和高度）与真实边界框之间的差异。CIoU 损失在 IoU 的基础上进行改进，考虑了边界框之间的重叠、中心点距离以及长宽比，从而提供了对边界框质量的更全面评估。边界框回归损失 L_{ciou} 计算公式如下：

$$L_{ciou} = 1 - (IoU - \frac{\rho^2 (b - b^{gt})}{c^2} - \alpha v) \quad (3)$$

其中： IoU 是预测框和真实框的交并比， b 是预测框的中心点， b^{gt} 是真实框的中心点， ρ 是预测框和真实框中心点之间的欧氏距离， c 是包含预测框和真实框的最小矩形区域的对角线距离， α 是权重系数， v 是预测框和真实框的长宽比之间的差异。

YOLOv5 的总损失函数是上述 3 个部分损失的加权和，即：

$$Loss = \lambda_1 L_{ciou} + \lambda_2 L_{conf} + \lambda_3 L_{class} \quad (4)$$

其中： λ_1 、 λ_2 、 λ_3 是用于平衡不同部分损失贡献的权重系数。这些权重系数可以根据具体任务和数据集进行调整，以优化模型的总体性能。

3 实验及结果分析

3.1 实验环境

本文实验在 Linux 操作系统下，使用 Intel Platinum P-8136 型 CPU、Nvidia RTX4090 显卡，Pytorch 深度学习框架。在同等配置下对 YOLOv5 及改进后的 YOLOv5 模型分别进行 400 轮次的训练。

3.2 实验数据集

本文实验数据是根据使用 USRP 设备所采集实际环境中的电磁信号，对原始电磁信号进行切分后使用 STFT 变换得到电磁信号的二维时频图像，得到的时频图像包含了大量的数据信息，其中也包括了噪声因素，通过对图像进行降噪减少图像中非目标信号的影，得到用于训练的原始时频图像。图 6 给出了降噪前包含有蓝牙信号和 WIFI 信号的时频图像示例，直观地显示了电磁信号在时间和频率上的特征，图 7 给出了经降噪处理后的时频图像示例，可以看到噪声被有效地抑制，信号的特征更为突出和清晰。

时频图像中包含有单个及多个混叠的蓝牙信号 (BT2.4) 和 WIFI 信号。使用 LabelImg 软件对其进行手工标注，标注完成后生成一个与标注图片名称相同的 xml 格式文件以保存标注信息，xml 文件中包含每一个被标注目标的标签类别以及该目标标记框坐标信息。数据集总数为 805 张时频图像，其中共含有蓝牙信号目标 2630 个，WIFI 信号目标 1146 个。选取其中 724 张用于训练集，81 张用于测试集。

3.3 评价指标

平均精度均值 (mAP, mean average precision) 是目标检测中常用的评价指标，它通过计算不同类别的平均精确度 (AP, average precision) 来衡量模型的整体性能。这一指标综合考虑了精确度和召回率，适用于多类别目标检测，尤其有效于类别繁多和数据不均衡的情况。

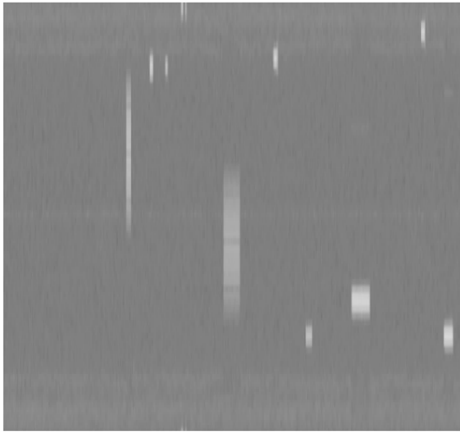


图 6 降噪前的时频图像

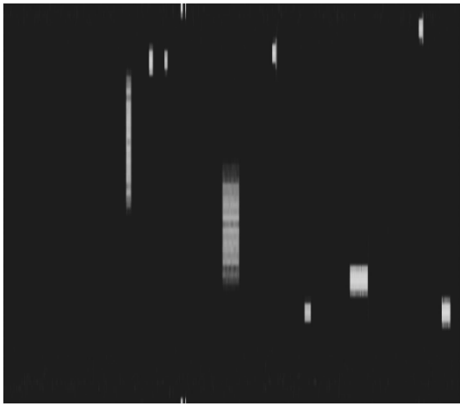


图 7 降噪处理后的时频图像

3.4 改进模型的消融实验

为了对不同的改进部分进行分析, 在保持参数不变的情况下进行消融实验, 消融实验结果如表 2 所示, 表中“√”表示在模型中使用了这种改进方式。由表 2 可知: 改进 1 通过使用 MoblieNetV3 替换原有的主干网络, 相较原始模型 mAP 仅下降 0.4 个百分点, 在实现了轻量化的同时兼顾了特征提取能力; 改进 2 在使用 MoblieNetV3 作为主干网络的同时引入了 ASFF 结构来融合特征金字塔的不同尺度特征, 有效提升了模型对于多尺度目标的检测能力, mAP 相较原始模型提升了 1.5 个百分点; 改进 3 在改进 2 的基础上使用 CIoU 作为损失函数解决了原有 GIoU 损失函数退化问题, 提供了对边界框质量的更全面评估, mAP 相较原始模型提升了 2 个百分点。消融实验结果表明了所提出改进方法的有效性。

表 2 不同改进方法消融实验对比表

算法	MobileNetV3 主干网络	ASFF 模块	损失函数	$mAP/\%$
YOLOv5				94.1
改进 1	√			93.7
改进 2	√	√		95.6
改进 3	√	√	√	96.1

3.5 实验结果对比分析

通过在自制电磁信号时频图像数据集对原始 YOLOv5 模型, SSD 模型和改进后的 YOLOv5 模型分别采用自适应矩估计 (Adam, adaptive moment estimation)^[28] 优化方法进行训练, 为了细致监控和评估模型性能, 在训练过程中每次迭代保存 $loss$ 值, 每 10 次迭代计算一次 mAP 值并保存模型参数权重文件。图 8 和图 9 分别展示了训练过程中 3 种不同算法的 mAP 和 $loss$ 。由两图可以看出, 随着训练轮数的增加, 3 个模型的 mAP 值逐渐提高, 在经过 400 轮的训练后, 3 个模型的表现趋于稳定, mAP 和 $loss$ 的波动减小, 这表明模型已经达到了收敛状态。

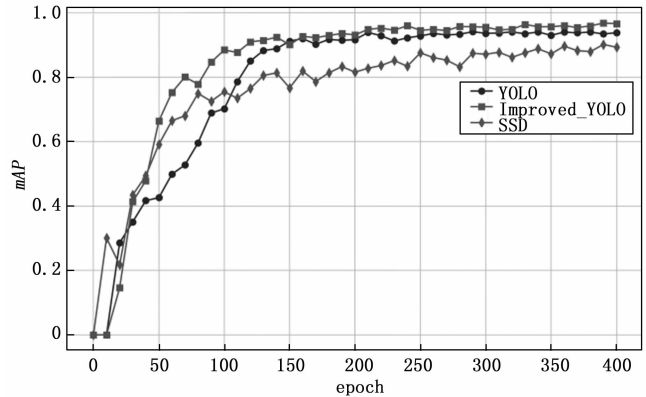


图 8 训练过程中 mAP 对比图

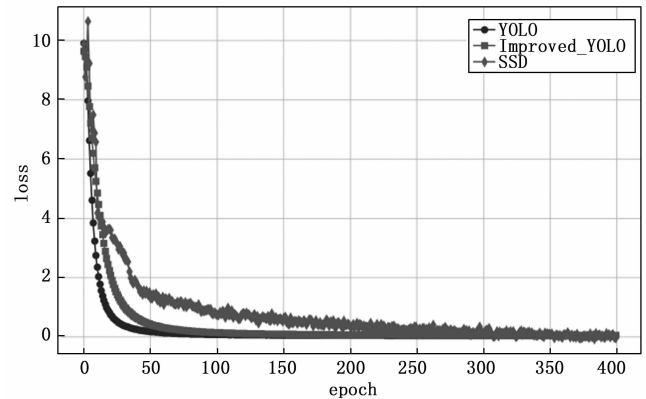


图 9 训练过程中损失对比图

为了更准确地评估和比较原始 YOLOv5 模型和 SSD 模型与改进 YOLOv5 模型在实际应用中的表现, 选取了第 200 轮, 250 轮, 300 轮, 350 轮, 400 轮时保存的权重文件。通过这些保存的权重, 在一个独立的测试集上进行了性能测试, 结果如表 3 所示, 从表中可以看出, 改进后的 YOLOv5 模型在所有选定的迭代点上均显示出比原始 YOLOv5 和 SSD 模型更高的检测精度。

3.6 检测结果可视化

为了综合评估和验证改进后的算法的性能, 分别取性能最佳的 YOLOv5 原算法, SSD 算法以及改进 YOLOv5 算法的模型权重进行对比测试, 同时取相同的图像进行识别

表 3 迭代次数与 mAP 对比表

算法	200	250	300	350	400
YOLOv5	91.4%	92.2%	94.1%	94.0%	93.7%
SSD	83.2%	83.3%	87.4%	88.8%	89.2%
改进后 YOLOv5	93.5%	95.9%	95.6%	95.7%	96.1%

并对比检测效果,来验证改进 YOLOv5 算法的检测性能。图 10 显示原始 YOLOv5 算法的识别结果,图 11 显示了 SSD 算法的识别结果,图 12 显示了改进 YOLOv5 算法的识别结果。从图 10 可以看出,尽管原始 YOLOv5 算法能够识别出多数目标,但它未能检测到具有细小特征的 BT2.4 目标。这可能是由于原始算法在处理小尺寸或细微特征方面的局限性。从图 11 可以看出 SSD 算法存在多个 BT2.4 目标和 WIFI2.4 目标的漏检,且选框较为粗糙,检测性能最差。相比之下,图 12 显示的改进后的 YOLOv5 算法不仅成功识别了 BT2.4 目标,还为其提供了更为精确的检测框。这表明改进的模型在特征提取和目标识别方面的能力得到了显著增强。结果表明,改进 YOLOv5 算法相比于原始 YOLOv5 算法和 SSD 算法在处理具有复杂特征的目标时具有更好的检测性能。

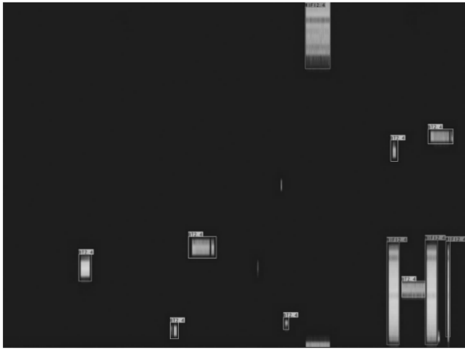


图 10 原始 YOLOv5 检测图

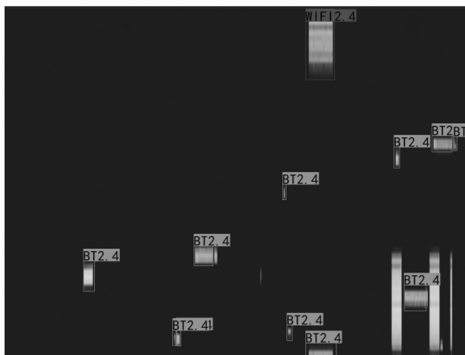


图 11 SSD 检测图

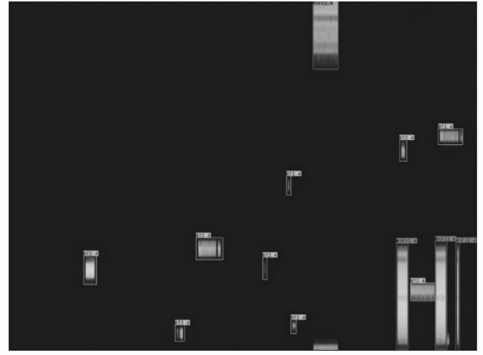


图 12 改进 YOLOv5 检测图

的同时保持了较高的精度。同时,引入 ASFF 模块使模型可以更加有效地利用不同层次的信息,增强了模型在各层次上的特征提取能力。引入 CIoU 损失函数使得目标框回归过程中更加稳定,收敛的精度更高。通过在自制电磁信号时频图像数据集上的对比实验,证明了改进 YOLOv5 算法相比原始 YOLOv5 算法以及 SSD 算法具有更好的识别效果。下一步的工作考虑如何使模型精确识别更多类型的目标信号,从而应用于更广泛的电磁环境中。

参考文献:

- [1] KULIN M, KAZAZ T, MOERMAN I, et al. End-to-end learning from spectrum data: A deep learning approach for wireless signal identification in spectrum monitoring applications [J]. IEEE Access, 2018, 6: 18484-18501.
- [2] 王雪松,李健兵,徐丰,等.电磁空间信息资源的认知与利用[J].中国科学基金,2021,35(5):682-687.
- [3] 孙文琦,吴建平,李贺武.宽带无线网络中的频谱资源管理综述[J].软件学报,2015,26(4):927-944.
- [4] THOMAZELLA, ROGÉRIO, et al. Digital signal processing for selfvibration monitoring in grinding: A new approach based on the time-frequency analysis of vibration signals [J]. Measurement, 2019 (145): 71-83.
- [5] 周鑫,何晓新,郑昌文.基于图像深度学习的无线信号识别[J].通信学报,2019,40(7):114-125.
- [6] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C] //Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2015: 1-9.
- [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [8] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [9] GIRSHICK R. Fast r-CNN [C] //Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.

(下转第 294 页)

4 结束语

本文提出了改进 YOLOv5 算法应用于在时间及频段上重叠的多种电磁信号中准确识别目标信号的问题,通过将 MobileNetV3 整合为主干网络,实现了对模型的轻量化