引用本文：王业恒，吴彰，赵永胜，等. 一种用于可见光通信信号调制格式识别的改进 YOLOv5s 算法[J]. 光通信技术，2024，48(3)：18-22.

一种用于可见光通信信号调制格式识别的改进 YOLOv5s 算法

王业恒，吴 彰，赵永胜，严志远，毛瑞霞，朱宏娜 *
（西南交通大学 物理科学与技术学院，成都 610031）

摘要：针对可见光通信信号在传输中易受信道环境和背景噪声干扰等因素影响调制格式识别精度的问题，提出一种用于可见光通信信号调制格式识别的改进 YOLOv5s（You Only Look Once）算法。首先，通过 YOLOv5s 算法网络输入端引入 Mixup 数据增强方式，将其与原网络中的 Mosaic 数据增强方式相结合，提升网络的鲁棒性，并增强算法在不同调制格式信号间的泛化能力；其次，将自适应空间特征融合（ASFF）引入到 Neck 网络中，充分提取不同层次的特征，提高检测精度。实验结果表明，在混合信噪比条件下，所提改进算法的平均精度均值（mAP）达到了 0.903，比原始 YOLOv5s 算法提升了 0.7%，且在信噪比为 20 dB 时 mAP 高达 0.993。

关键词：可见光通信；调制格式识别；YOLOv5s；Mixup 数据增强；自适应空间特征融合
中图分类号：TN911　**文献标志码**：A　**文章编号**：1002-5561(2024)03-0018-05
DOI：10.13921/j.cnki.issn1002-5561.2024.03.004　**开放科学（资源服务）标识码（OSID）**：

Improved YOLOv5s algorithm for modulation format recognition of visible light communication signal

WANG Yeheng, WU Zhang, ZHAO Yongsheng, YAN Zhiyuan, MAO Ruixia, ZHU Hongna*
（School of Physical Science and Technology, Southwest Jiaotong University, Chengdu 610031, China）

Abstract: Aiming at the problem that modulation format recognition accuracy is susceptible to factors such as channel environment and background noise interference in visible light communication signal transmission, this paper proposes an improved YOLOv5s(You Only Look Once) algorithm for modulation format recognition of visible light communication signals. Firstly, the Mixup data augmentation method is introduced at the input end of the YOLOv5s algorithm network, and it is combined with the Mosaic data augmentation method in the original network to enhance the robustness of the network and improve the generalization ability of the algorithm among different modulation format signals. Secondly, adaptively spatial feature fusion (ASFF) is introduced into the Neck network to fully extract features from different levels and improve detection accuracy. The experimental results indicate that under mixed signal-to-noise ratio conditions, the mean average precision(mAP) of the proposed improved algorithm reaches 0.903, representing a 0.7% improvement compared to the original YOLOv5s algorithm. Furthermore, the mAP reaches a high of 0.993 when the signal-to-noise ratio is 20 dB.

Key words: visible light communication, modulation format recognition, you only look once, Mixup data augmentation, adaptively spatial feature fusion

收稿日期：2024-02-29。
基金项目：中央高校基本科研业务费专项资金项目(202310613083)资助。
作者简介：王业恒(2003—)，男，河南洛阳人，本科生，现就读于西南交通大学物理科学与技术学院电子信息科学与技术专业，主要研究方向为机器学习、计算机视觉和光通信技术。
*** 通信作者**：朱宏娜(1979—)，女，博士，教授，主要研究方向为光纤通信、分布式光纤传感、机器学习和智能信息处理。

0 引言

近年来，人工智能(AI)技术迅速发展，深度学习作为一种强大的机器学习方法，已广泛应用于调制格式识别领域。深度神经网络能够自动提取并优化多维度特征，从而显著减少分类误差[1-3]。2019 年，XIANG Q 等人[4]提出了一种基于人工神经网络的联合精确信噪比(SNR)估计和调制格式识别方案。与卷积神经网络(CNN)和其它深度神经网络模型相比，该方案在识别性能上更胜一筹，同时所需计算资源更少。2020 年，

HE J 等人[5]设计了一种通过聚类和高斯模型对信号调制格式进行分类的方法，并在正交频分复用可见光通信系统中进行了实验验证。该方法在保持 100%识别精度的同时，所需的最小 SNR 比其它基于星座图的方法更低。2021 年，LV H J 等人[6]提出了一种基于信号幅值直方图的联合 SNR 监测和调制格式识别方法。该方法采用 CNN 模型，同时实现了 SNR 和调制格式的高精度识别。2022 年，GAO W 等人[7]设计了一种坐标合并算法，用于处理高维光脉冲传输数据。该算法在可见光通信系统中实现了更高的精度和更好的性能。2023 年，梁坤等人[8]针对光纤网络链路中的传输需求，提出了一种基于联合残差网络和 Bottleneck Transformer 的光纤调制格式识别方法。该方法在较宽的 SNR 范围内具有较高的识别准确率，能够有效应对传输损伤的影响。

但是，上述文献提到的识别方法存在调制信号转换流程繁琐、模型性能和训练复杂程度平衡不佳的问题。因此，本文提出一种用于可见光通信信号调制格式识别的改进 YOLOv5s(You Only Look Once)算法。

1 调制格式识别方法理论基础

可见光通信信号调制格式识别算法的总体流程图如图 1 所示。该识别过程主要包括以下 2 个步骤：

1)在图像生成模块中，首先生成各调制格式的时域 I、Q 信号，并通过 Matlab 中的 pspectrum 函数进行分段、加窗、傅里叶变换，从而计算每个时间段内的功率谱密度并合并，得到完整信号的短时功率谱密度估计。随后，该估计值被可视化并保存为时频图，进而生成 YOLOv5 算法所需的图像数据集。

2)在图像检测识别模块中，骨干网络负责从数据集图片中提取信号的特征，而 Neck 网络则进一步融合这些特征，使特征图的语义信息更加丰富。最终，经过这 2 个网络的协同工作，实现了不同调制信号的分类识别。

2 改进 YOLOv5s 算法

YOLO 算法是一种成熟的深度学习算法[9]，广泛应用于实时目标检测。它具备实时处理图像和视频的能力，并能同时检测多个不同类别的目标。其网络结构主要由四部分组成：输入端、骨干网络、改进后的 Neck 网络和 Head 输出端。与原始的 YOLOv5s 算法相比，本文主要在以下 2 个方面进行了改进：首先，在输入端引入了 Mixup 数据增强[10]方式，并将其与原有的 Mosaic 数据增强方式相结合，旨在提高网络的鲁棒性和增强模型在不同调制格式信号间的泛化能力。其次，将自适应空间特征融合(ASFF)[11]引入到了 Neck 网络中，通过自适应学习权重参数，实现对不同尺度特征的充分融合。改进后的网络整体结构如图 2 所示，其基本原理如下：首先，Mosaic-Mixup 模块对输入端的原始数据集进行数据增强；然后，骨干网络提取这些增强图像的特征；接下来，改进后的 Neck 网络将这些特征进行自适应融合；最后，Head 输出端对融合特征

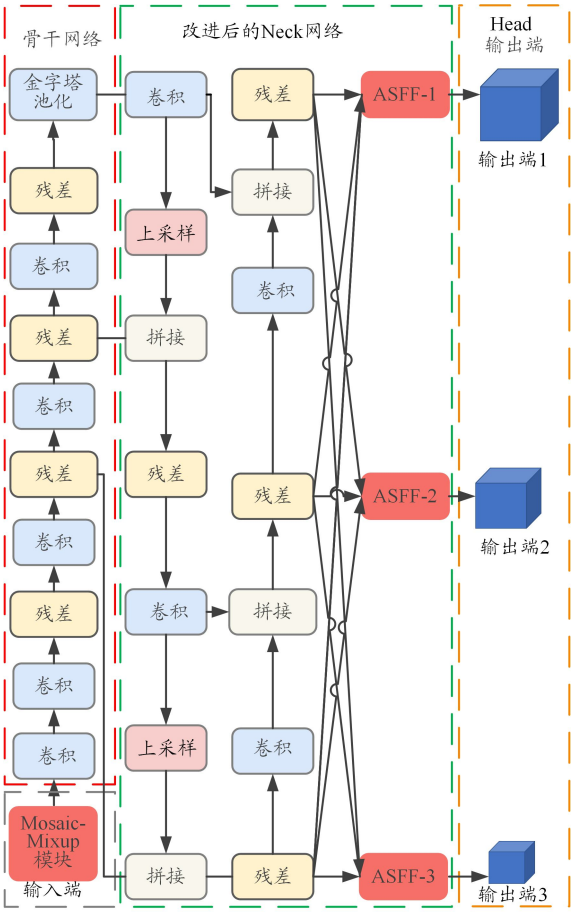

图 2 改进后的网络整体结构

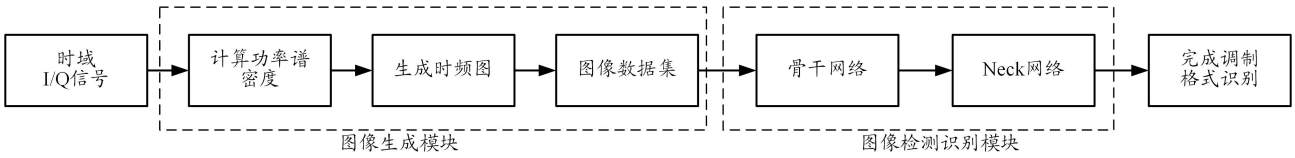

图 1 可见光通信信号调制格式识别方法流程

进行回归预测，输出最终结果。

2.1 数据增强的改进

YOLOv5s 算法在输入端默认采用 Mosaic 数据增强，通过对数据集中的 4 帧图像进行随机缩放、裁剪和拼接，有效丰富了检测数据集，从而缓解了数据集规模较小的问题。然而，在光通信信号时频图识别任务中，这种预处理操作可能产生无效背景，严重影响模型训练的准确度和效率。针对这一问题，本文所提改进算法在输入端结合了 Mosaic 和 Mixup 这 2 种数据增强方式。具体而言，数据集首先经过 Mosaic 数据增强处理，每 4 帧图像经过随机缩放和裁剪后拼接成 1 帧新的图像。随后，这些时频图再经过 Mixup 数据增强，按照邻域风险最小化原则，将不同类的图像以一定比例混合后作为训练数据送入深度学习网络进行训练。这种混合过程有效减少了随机裁剪拼接带来的无效背景，降低了模型在数据集中的泛化误差，使单帧图像的信息更加丰富，从而增强了模型在复杂环境下对目标的检测能力。这不仅提高了检测速度，还增强了改进 YOLOv5 算法识别模型的鲁棒性。

2.2 特征金字塔网络(FPN)的改进

YOLOv5s 算法的 Neck 网络采用了 FPN 和感知对抗网络(PAN)相结合的结构，如图 3 所示。FPN 通过自下而上的方式传递语义特征，而 PAN 则采用自上而下的方式传递定位特征。这 2 种机制的结合使得网络能够聚合来自不同层次的特征，从而显著增强了网络的特征融合能力。然而，当一帧图像中同时包含不同目标时，不同特征层次之间的特征冗余将降低特征金字塔的有效性，使得 FPN 与 PAN 相结合的融合方式无法充分利用不同尺度的特征。因此，本文在 Neck 网络层引入自适应空间特征融合(ASFF)，ASFF 结构如图 4 所示。

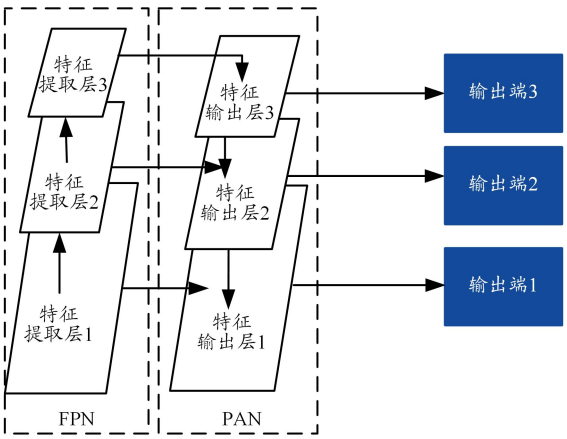

图 3 FPN 与 PAN 结合结构图

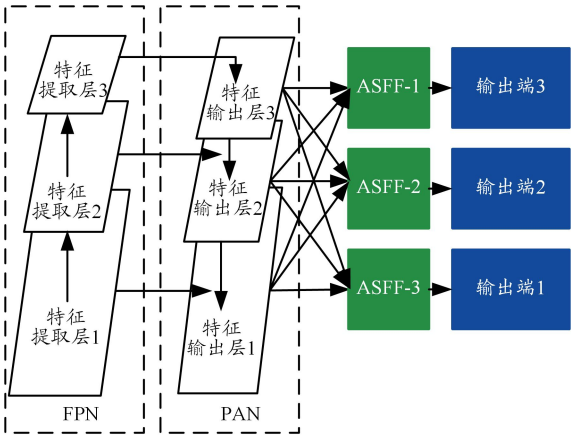

图 4 ASFF 结构图

ASFF 原理如下：首先，ASFF 结构输出 3 个不同尺度的特征层，这些特征层经过自适应融合后，形成了 3 个不同层级的 ASFF 特征层，分别命名为 ASFF-1、ASFF-2、ASFF-3。然后，ASFF 结构通过学习权重参数，能够有效地过滤掉其它层次的特征，仅保留对本层有用的信息进行组合，实现了各特征层的最优融合。这一特性使得后续的预测层能够充分利用不同层次的特征，从而提高了可见光通信调制格式识别的精度。

3 实验分析

3.1 数据集构建

本文主要针对可见光通信中的 4 种常见调制格式——二进制相移键控（BPSK）、四进制相移键控(QPSK)、八进制相移键控(8PSK)和 16 进制正交幅度调制(16QAM)进行研究。为了模拟真实的信道环境，本文在数据集生成过程中加入了多种信号干扰因素，如加性高期白噪声(AWGN)、多普勒频移、莱斯多径衰落和时钟偏移等。在 0~20 dB(步长为 2 dB)的 SNR 范围内，本文针对每种 SNR 条件生成 100 帧时频图，这样总共生成了 4 400 帧时频图。其中，3 520 帧用作训练集，880 帧用作测试集。实验采用 NVIDIA GeForce RTX 3090 GPU，在 Windows 10 操作系统下进行所有模型训练和测试，Pytorch 版本为 1.11.0，CUDA 版本为 11.3。此外，改进YOLOv5s 算法的参数设置如表 1 所示。

本文采用平均精度均值(mAP)来评估模型在调制格式识别方面的性能，并通过权重文件大小来体现模型在实际部署时的存储开销。本文所使用的 mAP 特指 mAP@0.5，即在交并比(IoU)为 0.5 时的 mAP。mAP 值越高，表明该目标检测模型在测试数据集上的检测效果越好。

表 1 改进 YOLOv5s 算法参数

参数	参数值
时频图分辨率	875×656
学习率	0.01
训练次数/次	300
批处理量/帧	64

3.2 实验结果及分析

3.2.1 消融实验

为了验证改进算法的有效性，本文以 YOLOv5s 算法为基础，在混合 SNR 条件下，对比分析了单独模块检测和组合模块检测下的不同效果，实验结果如表 2 所示。可以看出，在单独加入 Mixup 后，mAP 略有提升，这是因为单张图像的信息变得更加丰富，从而增强了网络在复杂环境下对目标的识别能力。当单独在 Neck 网络加入 ASFF 时，虽然 ASFF 能够充分提取不同层次的特征，提高网络对图像的表示能力，但由于数据集未经过 Mixup 数据增强处理，导致冗余的特征不利于模型识别，因此 mAP 略有下降。而当同时完成 Mixup+ASFF 的改进后，通过 Mixup 混合增强后的数据集再经过 ASFF 对不同层次特征的充分融合利用，mAP 显著提升至 0.903。这充分证明了本文所提改进算法的有效性。

表 2 算法消融实验结果

算法模块	mAP
YOLOv5s	0.896
YOLOv5s+Mixup	0.899
YOLOv5s+ASFF	0.892
YOLOv5s+Mixup+ASFF	0.903

3.2.2 不同算法性能对比实验

本文首先采用自制可见光通信调制格式数据集，训练了 SSD、Faster R-CNN、YOLOv3、YOLOv5s、YOLOv8s 这些主流的目标检测算法。接着，在混合 SNR 条件下，与本文提出的改进算法进行了对比，结果如表 3 所示。可以看出，改进 YOLOv5s 算法相较于 SSD 算法，mAP 提高了 52.3%，相较于 Faster R-CNN 算法提高了 43.1%，相较于 YOLOv3 算法提高了 1%。与原始的 YOLOv5s 和 YOLOv8s 相比，改进 YOLOv5s 算法在保持模型尺寸较小的同时，mAP 分别提高了 0.7%和 1%。

表 3 不同算法性能对比实验结果

算法	mAP	模型尺寸/MB
SSD	0.380	118
Faster R-CNN	0.472	314
YOLOv3	0.893	117
YOLOv5s	0.896	14.1
YOLOv8s	0.893	22.5
改进 YOLOv5s	0.903	24.17

3.3.3 不同 SNR 对比实验

SNR 是可见光调制信号传输和接收的关键因素。随着 SNR 的降低，时频图中的信号强度减弱，频谱展宽，甚至发生畸变，这使得信号与噪声的分离变得更加困难，进而影响了时频图的清晰度和可读性，最终对检测精度产生不利影响。因此，本文采用改进 YOLOv5s 算法，在多种 SNR 条件下对 4 种常见的调制信号进行了实验，实验结果如图 5 所示。可以看出，当 SNR 为 12 dB 时，mAP 为 0.903；当 SNR 为 20 dB 时，mAP 为 0.993。然而，当 SNR 降低至 8 dB 以下时，由于时频图中噪声的显著增加和信号的衰减，使得信号的识别细节变得模糊，可提取的识别特征减少，进而导致 mAP 明显降低。

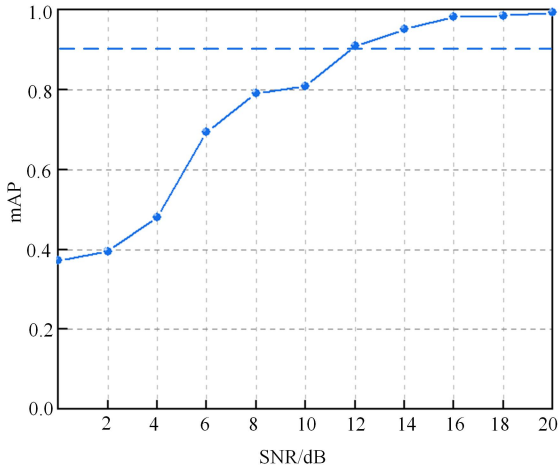

图 5 不同 SNR 实验结果图

4 结束语

本文提出了一种用于可见光通信信号调制格式识别的改进 YOLOv5s 算法，识别了可见光通信系统中常用的 4 种调制格式。实验结果表明：在混合 SNR 条件下，所提改进算法的 mAP 高达 0.903，比原始 YOLOv5s 算法提升了 0.7%；比 SSD 算法、Faster R-CNN 算法和 YOLOv3 算法分别提高了 52.3%、43.1%和 1%。本文方法为多类型的调制格式识别提供了可

行思路。在未来工作中，可以考虑对信号时频图进行降噪处理及进一步优化网络结构，以提升识别精度。

参考文献：

[1] TUNZE G B, HUYNH-THE T, LEE J M, et al. Sparsely connected CNN for efficient automatic modulation recognition [J]. IEEE Transactions on Vehicular Technology, 2020, 69(12): 15557–15568.

[2] ZENG Y, ZHANG M, HAN F, et al. Spectrum analysis and convolutional neural network for automatic modulation recognition [J]. IEEE Wireless Communications Letters: 2019, 8(3): 929–932.

[3] ANG R Q, PENG Y P, XIE W X, et al. Improved YOLOv4 small target detection algorithm with embedded SCSE module [J]. Journal of graphics, 2021, 42(4): 546–555.

[4] XIANG Q, YANG Y, ZHANG Q, et al. Joint and accurate SNR estimation and modulation format identification scheme using the Feature-Based ANN[J]. IEEE Photonics Journal, 2019, 11(4): 1–11.

[5] HE J, ZHOU Y, SHI J, et al. Modulation classification method based on clustering and gaussian model analysis for VLC system[J]. IEEE Photonics Technology Letters, 2020, 32(11): 651–654.

[6] LV H J, ZHOU X, HUO J H, et al. Joint SNR monitoring and modulation format identification on signal amplitude histograms using convolutional neural network[J]. Optical Fiber Technology: 2021, 61: 1–5.

[7] GAO W, XU C, XU Z, et al. Modulation format recognition based on coordinate transformation and combination in VLC system [C]//2022 Asia Communications and Photonics Conference (ACP), November 5–8, 2022, Shenzhen, China. Piscataway: IEEE, 2022: 1151–1155.

[8] 梁坤，刘战胜. 基于联合残差网络和 Bottleneck Transformer 的调制格式识别方法[J]. 光通信技术，2024，48(3)：13–17.

[9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27–30, 2016, Las Vegas, USA. New York: IEEE, 2016: 2–10.

[10] KAUR P, KHEHRA B S, MAVI E B S. Data augmentation for object detection: a review [C]//2021 IEEE International Midwest Symposium on Circuits and Systems(MWSCAS), August 9–11, 2021, Lansing, USA. New York: IEEE, 2021: 537–543.

[11] LIU S, HUANG D, WANG Y. Learning spatial fusion for single-shot object detection [EB/OL]. (2019–11–21) [2024–02–25]. https://arxiv.org/abs/1911.09516.