张常斌, 孙连山, 唐景琰. 基于改进 YOLOv5 的轻量级手势识别算法[J]. 智能计算机与应用,2025,15(6):184-189. DOI: 10.20169/j.issn.2095-2163.250628

基于改进 YOLOv5 的轻量级手势识别算法

张常斌, 孙连山, 唐景琰
(陕西科技大学 电子信息与人工智能学院, 西安 710021)

摘 要: 针对嵌入式设备资源限制的问题,本文通过改进 YOLOv5 模型提出一种轻量级手势识别算法,旨在保证检测精度前提下,提高检测速度。本文模型 YOLOv5-GR 使用 GhostNet 网络作为特征提取网络,大幅减少了模型参数量;模型引入加权特征融合网络 Bi-FPN,增强特征融合能力,并使用 *SIoU* 损失函数替换 *CIoU* 损失函数,关注真实锚框与预测的角度信息,提升检测精度。此外,使用 K-Means++算法对 ASL 数据集聚类先验框,使其更加适合该数据集。实验结果表明,YOLOv5-GR 模型在保证检测精度与 YOLOv5 模型持平的前提下,模型参数量减少了 33.3%,实现了网络模型轻量化,满足在嵌入式资源受限设备上部署的需要。

关键词: 手势识别; YOLOv5; GhostNet; Bi-FPN; K-Means++
中图分类号: TP391.4 **文献标志码**: A **文章编号**: 2095-2163(2025)06-0184-06

Lightweight gesture recognition algorithm based on improved YOLOv5

ZHANG Changbin, SUN Lianshan, TANG Jingyan
(School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi′an 710021, China)

Abstract: In this paper, a lightweight gesture recognition algorithm is proposed to address the issue of resource constraints in embedded devices. The YOLOv5-GR model is developed by improving the YOLOv5 model, aiming to improve detection speed while ensuring detection accuracy. GhostNet is used as the feature extraction network, significantly reducing the model parameters. The model incorporates the Bi-FPN network for weighted feature fusion to enhance feature fusion capability. The *SIoU* loss function is introduced to focus on the angle information between the true anchor box and the prediction, thus improving detection accuracy. Additionally, the K-Means++ algorithm is used to cluster the prior boxes for the ASL dataset, making them more suitable for this dataset. Experimental results show that the YOLOv5-GR model achieves a 33.3% reduction in model parameters while maintaining detection accuracy comparable to the YOLOv5 model. This enables lightweight deployment of the network model on embedded devices with limited resources.

Key words: gesture recognition; YOLOv5; GhostNet; Bi-FPN; K-Means++

0 引 言

随着科学技术的发展,人机交互已经成为日常生活中必不可少的操作[1]。手势是人际交往的重要组成部分,具有直观形象和使用方便的特点,因此手势在人机交互中起到了重要的作用,可以应用在智慧医疗、智能家居、智慧城市等多个领域[2]。

目前,手势识别方法主要分为 2 种:基于穿戴设备和基于计算机视觉[3]。其中,基于穿戴设备的方法需要使用者利用数据手套等设备来采集手势数据,然后通过计算机进行手势识别,但这种方法存在设备昂贵且使用不方便的问题[4]。基于计算机视觉的方法使用摄像头来捕捉手势数据,不需要额外的设备[5]。

深度学习在目标检测、图像分类等领域取得了重大突破,其中如 YOLO 和 SSD 等算法在检测和分

基金项目: 陕西省自然科学基础研究计划(2023-JC-YB-581)。

作者简介: 张常斌(1998—),男,硕士研究生,主要研究方向:手势识别,智能灯光; 唐景琰(1999—),女,硕士研究生,主要研究方向:访问控制,智能家居。
通信作者: 孙连山(1977—),男,博士,教授,主要研究方向:软件工程,信息安全与隐私保护,数据溯源技术及应用,区块链技术。Email: sunlianshan@sust.edu.cn。

收稿日期: 2023-10-29

类问题上获得了较高的检测精确度。然而，随着网络层数的增加，对于嵌入式设备的资源要求也越来越高，因此，深度学习网络的轻量级优化是不可避免的[6]。

近年来，Hu 等学者[7]构建了手势数据集，并基于深度摄像头和轻量级卷积神经网络模型进行基于视频的手势识别。Ewe 等学者[8]提出一种基于轻量级 VGG16 和随机森林的混合网络架构的手势识别方法，主干网络采用卷积神经网络技术进行特征提取，并使用机器学习方法进行分类。Wang 等学者[9]提出了一种轻量级卷积神经网络 E-MobileNetv2 用于手势识别，将 ECA 模块添加到原始 MobileNetv2 网络模型中，并且添加了 R6-SELU 激活功能。Chen 等学者[10]通过将 YOLOv5 骨干网络中的 CSP1_x 模块替换为高效的层聚合网络，可以获得更丰富的梯度路径组合，以提高网络的学习表达能力和识别速度。

因此，本文提出了一种基于 YOLOv5s 网络模型的轻量级手势识别模型 YOLOv5-GR。该方法在保持了 YOLOv5s 模型在 ASL 手势数据集上高精度的同时，减少了模型参数，使得模型更适合在资源有限的嵌入式设备上进行部署[11]。该方法有效地解决模型大而复杂的问题，使得手势识别技术更加可靠和实用。

1　YOLOv5 网络模型

手势识别是人机交互的一种方式，会应用于在各种嵌入式设备上，对于模型大小具有较高的要求，所以选用 YOLOv5s 作为基本模型[12]。YOLOv5s 模型主要分为 3 部分。其中，输入端采用自适应锚计算、数据增强等方法，提高对小目标物体的检测精度[13]；主干网络使用 CSP 网络结构提取图形特征，头部网络主要解决主干提取特征映射的定位问题[14]。Neck 网络由 FPN 和 PAN 组成，是连接主干和头部的部分，对特征图进行细化和重构[15]。YOLOv5s 的网络结构如图 1 所示。

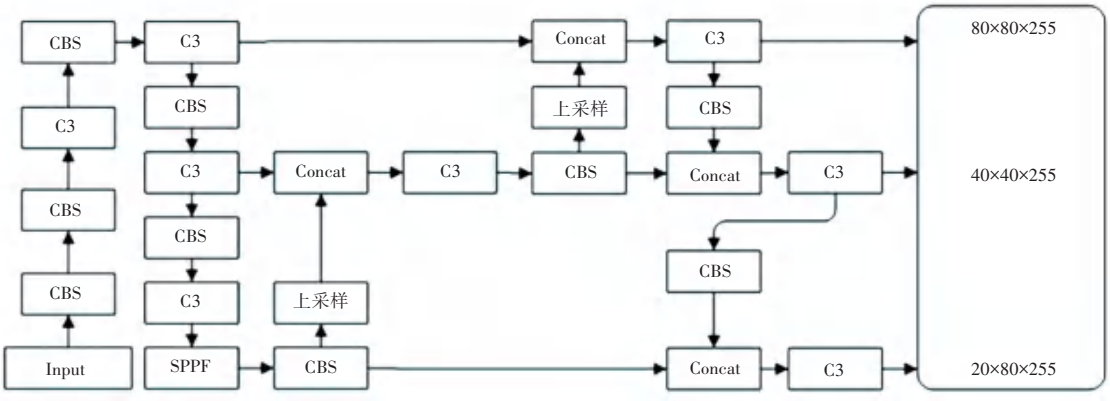

图 1　YOLOv5s 模型结构

Fig. 1　Structure of YOLOv5s model

2　模型改进

2.1　YOLOv5-GR 网络模型

为了实现模型的轻量化处理，本文提出的改进方案从主干网络、颈部、先验框和损失函数四个方面对 YOLOv5s 模型进行优化：主干网络中使用了 GhostNet 对主干特征提取网络进行重新设计；采用 Bi-FPN 网络来优化 Neck 网络，进行多尺度特征融合；为了使模型的先验框 Anchor 更适合 ASL 手势数据集，使用 K-means++算法对先验框尺寸进行重新聚类。此外，使用 *SIoU* 损失函数对 YOLOv5s 网络的损失函数 *CIoU* 进行替换。改进后的网络 YOLOv5-GR 整体如图 2 所示。

2.2　GhostNet 网络

在手势图片中存在非手势的干扰物品，YOLO 网络通过普通卷积操作会提取手势特征和非手势特征，在特征图中存在部分相似的特征，这些相似特征有助于提升模型的准确率；但这些相似特征在卷积过程中产生的冗余映射会消耗大量计算资源。

为了减少冗余映射带来的计算量，使用 GhostNet 网络进行特征提取。GhostNet 网络使用 Ghost Bottleneck 结构进行设计。Ghost Bottleneck 结构是以 Ghost 模块为基础，每个 Ghost Bottleneck 中都包含 2 个 Ghost 模块。第一个模块的作用是扩展特征通道数，而第二个模块则用于压缩特征通道数[16]。

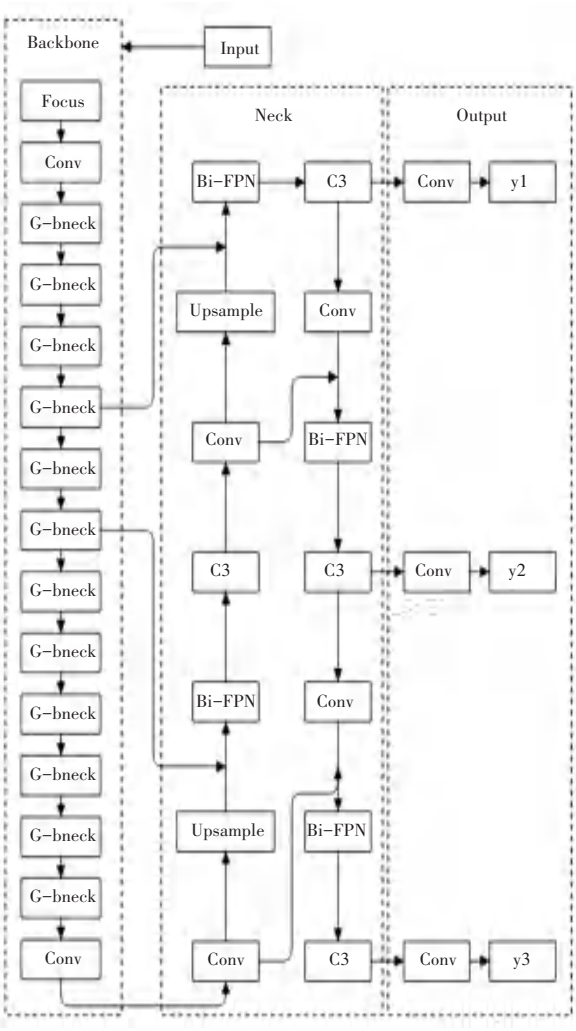

图 2 YOLOv5-GR 模型结构

Fig. 2 Structure of YOLOv5-GR model

Ghost 模块生成的特征图主要分为 2 个部分。第一部分使用部分普通卷积来生成部分特征图,第二部分对第一部分的特征图进行线性操作[17]。最后,将 2 组特征图在指定维度上进行拼接,以确保特征精度的同时减少整体计算量。Ghost 模块的结构如图 3 所示。

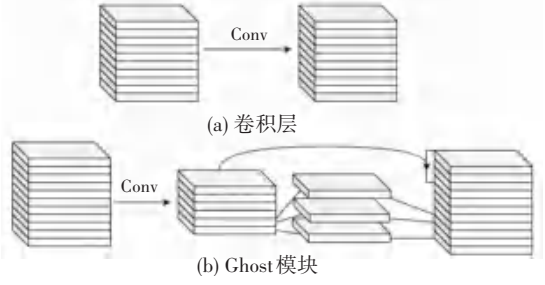

图 3 Ghost 模块结构

Fig. 3 Structure of Ghost module

2.3 Bi-PFN 网络模型

YOLOv5s 模型的特征融合网络是在之前 YOLO 网络的 FPN 基础之上增加路径聚合网络得到 PAN 进而得到的特征融合网络 PANet,虽然该网络能解决不同场景图像中对象尺度差异较大的问题,但并不是特征融合的最优解,此外,低级别的特征有助于大实例识别,但本文是对手势进行分类,并不需要进行实例识别,而且从底层结构到顶层特征还要经过一系列流程,这增加了获取准确定位信息的难度。因此使用 Bi-FPN 对 Neck 网络结构进行优化。

Bi-FPN 采用双向跨尺度连接,简化网络复杂度,在一条输入输出线上增加一条额外的连接,加强特征融合,并且在传统的 FPN 中加入了跳跃连接,使网络能够融合更多不同尺寸的特征,而且网络将减少只有一个输入的结点对特征网络的贡献。原理如图 4 所示[18]。

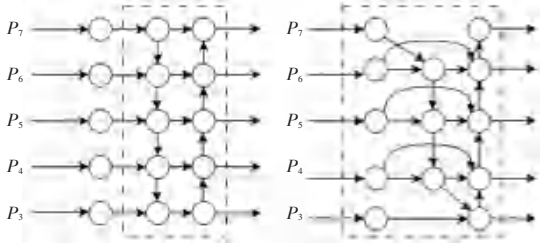

图 4 Bi-PFN 结构

Fig. 4 Structure of Bi-PFN

2.4 K-means++聚类先验框

YOLOv5s 网络中的 Anchor 尺寸是使用 K-means 算法在 COCO 数据集上生成的,但是 COCO 数据集有 80 种类别的物体,这些目标物体尺寸大小不一。本文的检测目标只有手势一种类别,而且手势的尺寸大小相近、宽高比相似,为了使优化算法更快速准确地预测出目标手势的位置,本文利用 K-means++聚类算法对 ASL 手势数据集的 Anchor 尺寸重新聚类。首先选取 k 个簇中心,然后计算各个目标与簇中心的距离,把目标分配给最近的簇中心,重复计算更新簇中心,直到簇中心不再改变,得到最终结果[19]。先验框对比结果见表 1。

表 1 先验框对比表

Table 1 Comparison table of prior box

类别	值
原始先验框	[10,13 16,30 33,23]
	[30,61 62,45 59,119]
	[116,90 156,198 373,326]
聚类的先验框	[85,142 142,266 169,191]
	[191,336 223,255 251,370]
	[296,188 347,286 368,426]

2.5　损失函数优化

YOLOv5s 采用 *CIoU* 作为定位损失函数。虽然该损失函数考虑了真实框和预测框的重叠面积以及中心点距离和长宽比，但 *SIoU* 损失函数在 *CIoU* 的基础上把预测框和真实框之间的向量角度作为惩罚项加入损失函数，防止预测框发生漂移，并使其更加贴合真实框，进一步提升网络的收敛速度和训练效果。因此本文对损失函数进行优化，使用 *SIoU* 替代 *CIoU*。

3　实验对比

3.1　实验环境

本文中实验的操作系统为 Linux 系统，训练框架为 PyTorch1.7.1，编译环境为 Python3.7，CPU 型号为 Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50 GHz，GPU 为 NVIDIA V100。

3.2　手势识别数据集

本文使用的数据集为 ASL(American Sign Language)手势数据集（https://universe.roboflow.com/meredith-lo-pmqx7/asl-project），该数据集的手势类别分别为字母 A~Z 共 26 种，共 11 355 张图片，分为训练集、测试集和验证集三部分，分别包括 9 246、789、1 320 张图片。具体数据集展示如图 5 所示。

图 5　数据集展示

Fig. 5　Presentation of the dataset

3.3　评价指标

实验使用召回率（*Recall*）、精准率（*Precision*）、平均精度 *AP*（Average Precision）、平均精度均值 *mAP*（mean Average Precision）来评价检测模型准确性。评价指标计算公式如下：

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$AP = \int_0^1 p(r)\,\mathrm{d}r \tag{3}$$

$$mAP = \frac{1}{m}\sum_{i=1}^{m} AP_i \tag{4}$$

其中，m 表示样本类别数；$p(r)$ 表示 *Precision* 以 *Recall* 为参数的一个函数；*TP* 表示被正确识别的正样本；*TN* 表示被正确识别的负样本；*FP* 表示负样本被错误识别为正样本；*FN* 表示正样本被错误识别为负样本。

3.4　实验结果与分析

针对原始模型 YOLOv5 和优化模型 YOLOv5-GR 在 ASL 手势数据集上进行对比。研究可知，原始模型在数据集上的 *mAP*@0.5 为 95.9%，模型参数大小为 13.86 M；优化模型 YOLOv5-GR 在数据集上的 *mAP*@0.5 值为 96.0%，模型参数为原始模型 YOLOv5s 的 2/3，大小仅为 9.25 M，实际检测结果如图 6 所示。

图 6　YOLOv5-GR 模型的检测结果

Fig. 6　Detection results of YOLOv5-GR model

为了验证本文模型 YOLOv5-GR 的各项改进模块对手势目标识别性能的影响，在同一实验环境下对 YOLOv5-GR 网络各个改进模块的有效性进行了实验，具体实验结果见表 2。

首先是 YOLOv5s 模型在 ASL 数据集进行实验，该模型在 ASL 数据集上的 *mAP*@0.5 值为 95.9%，虽然该网络模型已是 YOLOv5 的最简版本，但模型大小依旧可达到 13.86 M。

随即对改进的各版本网络模型进行实验。首先,使用 GhostNet 来降低 YOLOv5s 参数。改进后的网络模型大小降低了 1/3,仅为 9.24 M,但是 *mAP*@0.5 值降低了 2.5%。同时,该版本模型的召回率降低了 4.2%,精准率降低了 1.6%。

其次,使用 Bi-FPN 结构对网络进行优化。改进后的网络参数较版本 1 没有发生变化,仅仅增加了 0.01 M,但 *mAP*@0.5 值提升了 0.5%。

随后,使用 K-means++算法对 ASL 手势识别数据集的先验框进行了重新聚类,精准率提升了 1.9%,召回率提升了 3.8%,*mAP*@0.5 值提升了 1.8%。

最后,对网络损失函数进行优化,将 *CIoU* 替换为 *SIoU*,得到了最终网络版本 YOLOv5-GR。该网络的 *mAP*@0.5 值比原生 YOLOv5s 提高了 0.1%,并且模型参数下降了 1/3。

表 2 YOLOv5-GR 模型的消融实验

Table 2 Ablation experiment of YOLOv5-GR model

模型	GhostNet	Bi-FPN	K-means++	*SIoU*	*Precision*	*Recall*	*mAP*	参数/M
Version	×	×	×	×	0.936	0.931	0.959	13.86
	√	×	×	×	0.920	0.890	0.934	9.24
	√	√	×	×	0.920	0.882	0.939	9.25
	√	√	√	×	0.939	0.919	0.957	9.25
	√	√	√	√	0.939	0.920	0.960	9.25

至此,为了展示该模型的效果,在相同的运算环境下,将 YOLOv4、YOLOv3、SSD、YOLOv5s 算法以及本文算法 YOLOv5-GR 在 ASL 数据集上进行比较分析。不同算法的对比结果见表 3 所示。

表 3 不同算法的对比结果

Table 3 Comparison results of different algorithms

模型	*mAP*	参数/M
YOLOv5s	0.959	13.86
YOLOv4	0.915	244.44
YOLOv3	0.862	244.59
SSD	0.929	102.15
YOLOv5-GR	0.960	9.25
VGG-16	0.948	53.60

由表 3 可见,在检测精度上,YOLOv5-GR 模型和 YOLOv5s 模型的精度持平,在模型参数减少了 33.3%。除此之外,在检测精度和模型大小上,相比其他基准模型发现,YOLOv5-GR 模型均有一定的优势。因此 YOLOv5-GR 模型更适用于计算资源有限的嵌入式设备。

4 结束语

本文在 YOLOv5s 模型的基础上提出了一种轻量级手势识别算法 YOLOv5-GR。通过使用 GhostNet、Bi-FPN、*SIoU*、K-means++方法从主干网络、Neck 网络、先验框、损失函数等方面对 YOLOv5s 模型进行优化得到本文改进算法,YOLO-GR 模型的 *mAP* 值比原生的 YOLOv5s 提升了 0.1%,而模型的参数减少了 33.3%、现在仅为 9.25 M,可以提高模型的运算速度,使得该模型更加适用于资源和算力有限的嵌入式设备。

后续工作中,在保证准确率的条件下,提升召回率。并且在 ASL 手势数据集的基础上,还应扩充手势数据集,丰富手势目标的检测场景。除此之外,将继续从实际场景出发,不断研究和改进相关算法,设计出更轻量级的手势识别模型,使其在面对不同场景的情况下,更加适合部署在移动端和嵌入式设备。

参考文献

[1] GUO Lin, LU Zongxing, YAO Ligang. Human - machine interaction sensing technology based on hand gesture recognition: A review[J]. IEEE Transactions on Human-Machine Systems, 2021, 51(4): 300-309.

[2] OUDAH M, AL-NAJI A, CHAHL J. Hand gesture recognition based on computer vision: A review of techniques[J]. Journal of Imaging, 2020, 6(8): 73.

[3] ZHOU Weina, CHEN Kun. A lightweight hand gesture recognition in complex backgrounds[J]. Displays, 2022, 74: 102226.

[4] 王银,陈云龙,孙前来. 复杂背景下的手势识别[J]. 中国图象图形学报,2021,26(4):815-827.

[5] 辛文斌,郝惠敏,卜明龙,等. 基于 ShuffleNetv2-YOLOv3 模型的静态手势实时识别方法[J]. 浙江大学学报(工学版),2021,55(10):1815-1824.

[6] 卢迪,马文强. 基于改进 YOLOv4-tiny 算法的手势识别[J]. 电子与信息学报,2021,43(11):3257-3265.

[7] HU Pengli, TANG Chengpei, YIN Kang, et al. WiGR: A practical Wi - Fi - based gesture recognition system with a lightweight few-shot network[J]. Applied Sciences, 2021, 11(8): 3329.

[8] EWE E L R, LEE C P, KWEK L C, et al. Hand gesture recognition via lightweight VGG16 and ensemble classifier[J].

Applied Sciences，2022，12(15)：7643.
[9] WANG W，HE M，WANG X，et al. Medical gesture recognition method based on improved lightweight network [J]. Applied Sciences，2022，12(13)：6414.
[10] CHEN R，TIAN X. Gesture detection and recognition based on object detection in complex background [J]. Applied Sciences，2023，13(7)：4480.
[11] 黄凯雯，房宵杰，梅林，等. MPE-YOLOv5：面向边缘计算的轻量化 YOLOv5 手势识别算法[J]. 哈尔滨工业大学学报，2023，55(5)：1-13.
[12] HU Dunli，ZHU Jun，Liu Jiayu，et al. Gesture recognition based on modified Yolov5s[J]. IET Image Processing，2022，16(8)：2124-2132.
[13] HAO Bo，YIN Xingchao，YAN Junwei，et al. Gesture recognition in the complex environment based on Gan - St - YOLOv5 [J]. Journal of Northeastern University (Natural Science)，2023，44(7)：953-963.
[14] SAXENA S，PAYGUDE A，JAIN P，et al. Hand Gesture recognition using YOLO models for hearing and speech impaired people[C]//Proceedings of 2022 IEEE Students Conference on Engineering and Systems (SCES). Piscataway，NJ：IEEE，2022：1-6.
[15] YADAV Y G，KIRAN V S，THADIKAMALLA G A，et al. Real time sign language recognition using custom convolutional neural network and YOLOv5 [M]//DASSAN P，THIRUMAARAN S，SUBRAMANI N. Intelligent Computing，Smart Communication and Network Technologies，ICICSCNT 2023. Communications in Computer and Information Science. Cham：Springer，2024：157-171.
[16] WANG C，GUO J，WANG S，et al. Research on Gesture Recognition Algorithm Based on Lightweight YOLOv4 [C]// Proceedings of 2022 2nd International Conference on Computation，Communication and Engineering (ICCCE). Piscataway，NJ：IEEE，2022：74-78.
[17] 牛雅睿，武一，孙昆，等. 基于轻量级卷积神经网络的手势识别检测[J]. 电子测量技术，2022，45(4)：91-98.
[18] 李泰国，张英志，张天策，等. 基于改进 YOLOv5s 算法的列车驾驶员手势识别[J]. 铁道学报，2023，45(1)：75-83.
[19] 邢晋超，潘广贞. 改进 YOLOv5s 的手语识别算法研究[J]. 计算机工程与应用，2022，58(16)：194-203.