基于YOLOv5的静态手势识别检测模型

程亚龙,梁　军,邹雲宇

(华南师范大学 软件学院,广东 佛山 528225)

摘　要: 针对实时手势检测需求,提出一种基于YOLOv5的手势识别算法。通过采用轻量级主干网络MobileNetV3替代YOLOv5s中的CSPNet-53,优化后的主干网络整合了深度可分离卷积与SE注意力机制,形成模型M_YOLO_N(MobileNet_YOLOv5_NewIou)。与原始模型相比,M_YOLO_N的参数量减少了33%,计算复杂度(GFLOPs)降低了54%,在自制手势数据集上的mAP@0.5提升了2.4%。该模型不仅实现了轻量化,而且有效解决了实时检测问题。针对多尺度手势检测,保留SPPF模块,并引入归一化高斯瓦伦汀距离(NWD)技术,提出新的边界框损失函数NewIoU。在不增加参数的前提下,改进后的模型在多尺度手势检测中的置信度提升了20%。

关键词: YOLOv5;手势识别;深度可分离卷积;注意力机制

DOI: 10. 11907/rjdk. 232165　　开放科学(资源服务)标识码(OSID):

中图分类号: TP391　　**文献标识码:** A　　**文章编号:** 1672-7800(2024)011-0181-06

Detection Model for Static Gesture Recognition Based on YOLOv5

CHENG Yalong, LIANG Jun, ZOU Yunyu

(*School of Software*, *South China Normal University*, *Foshan* 528225, *China*)

Abstract: To meet the demand for real-time hand gesture detection, this paper presents a YOLOv5-based gesture recognition algorithm. By replacing CSPNet-53 in YOLOv5s with the lightweight MobileNetV3, the optimized backbone integrates depthwise separable convolutions and the SE attention mechanism, forming the M_YOLO_N model. Compared to the original, M_YOLO_N reduces parameters by 33% and decreases computational complexity by 54%. On a custom dataset, mAP@0.5 increased by 2.4%. This model achieves both lightweight design and real-time detection.For multi-scale detection, the SPPF module is retained, and the normalized Wasserstein distance (NWD) is introduced, proposes a new bounding box loss function NewIoU. Without increasing parameters, detection confidence improved by 20%.

Key Words: YOLOv5; gesture recognition; depth separable convolution; attention mechanism

0　引言

手势识别一直是国内外学者对于深度学习实际应用研究的热门方向之一,学者们通过多种方法提升手势识别的准确率。在静态手势识别方面,Jones等[1]通过肤色检测建立颜色分布模型,结合直方图和混合模型实现精准分割。胡宗承等[2]采用改进的轻量化MobileNetV2网络,使参数减少了27%,错误率降低了1.82%。凌滨等[3]融合深度信息与GrabCut算法构建新模型,解决了传统算法迭代时间长的问题。舒子超等[4]结合深度和肤色等特征信息,使模型的识别率达到了98.9%。刘淑萍等[5]从肤色检测出发,提出基于手指的分层识别策略,显著提高了识别率。在动态手势识别方面,Smedt等[6]使用Fisher核和多级时间金字塔从手势序列中提取有效的运动学描述符。Ohara等[7]通过Wi-Fi信号和隐马尔可夫模型提取手部运动分量和特征,并消除偏差。谷学静等[8]结合卷积神经网络和长短时记忆网络提出动态手势识别方法,平均识别率可达到92.5%。

然而,随着人们对识别准确率要求的不断提高,传统的深度神经网络模型由于具有复杂的参数和高计算需求,常常导致检测速度变慢,无法满足实时检测的需求。此

收稿日期: 2023-11-15

扫描二维码阅读全文:

基金项目: 广东省基础与应用基础研究基金项目(2022A1515140110,2021A1515110673, 2020B1515120089);佛山市高等教育高层次人才项目(2022)

作者简介: 程亚龙(2000-),男,CCF会员,华南师范大学软件学院硕士研究生,研究方向为图像识别、目标检测;梁军(1983-),男,博士,CCF高级会员,研究方向为机器学习、图论算法。本文通讯作者:梁军。

外，在实际应用中，多尺度、多目标的手势识别需求日益增长，而现有的静态和动态手势识别方法在应对这些挑战时仍存在不足，主要表现为实时性差和对多尺度目标检测的适应性不足。

因此，为了满足手势识别在实际应用中对于实时性、多尺度多目标同时检测的需求，本文对 YOLOv5 模型作出一些改进，提出一种基于 YOLOv5 目标检测的手势识别算法。采用轻量级主干网络 MobileNetV3[9]替换 YOLOv5s 中的 CSPNet-53[10]主干网，引入深度可分离卷积以及 SE 注意力机制，使模型参数量降低了 33%，计算复杂度降低了 54%，相比传统模型更加轻量化，且可以满足实时检测要求。本文提出的模型保留了原有的 Spatial Pyramid Pooling Fast（SPPF）模块，添加了 Normalized Wasserstein Distance（NWD）。SPPF 使模型对多尺度目标检测具有很好的普适性，NWD 则进一步提高了模型识别的准确率。该模型可以获得接近甚至高于部分传统模型的准确率。通过多次实验，在实验所用数据集上，修改后的算法在准确率和速度方面优于现有的 YOLOv5s 和部分相关领域模型，适用于实际应用的要求。

1 方法介绍

本文主要对 YOLOv5 模型进行了如下改进：

（1）使用 MobileNetv3 的主干网络代替原有的 CspNet53，引入深度可分离卷积（Depthwise Separable Convolution，DSConv）和 SE 注意力机制，同时保留了 SPPF 模块。新的主干网络具有更加轻量化的优点，模型参数量更少，有助于在终端上部署，也能够确保检测的实时性。

（2）使用 GIoU（Generalized Intersection over Union）替代原有的 CIoU（Complete Intersection over Union），并引入归一化高斯瓦伦汀距离（NWD）方法，两者构成新的边界框损失函数。GIoU 简单、高效，NWD 则能够提升检测结果的置信度，并增强多尺度检测效果。

1.1 数据集

相同的模型在不同数据集上的表现有很大差异，因此数据集对模型的识别效果有着很大影响。本文使用的静态手势识别数据集共分为 6 个类别，共 1 524 张图片。其中，训练集 995 张，验证集 529 张。手势类别及数量如表 1 所示。

Table 1 Gesture recognition data set

表 1 手势识别数据集

标签	类别	数量
0	4/D	320
1	OK	199
2	1/A	368
3	3/C	127
4	fine	200
5	2/B	310

1.2 特征提取网络

MobileNetv3_small 主干网络部分的相关参数如表 2 所示[9]。其中，exp size 表示第一个升维的卷积将维度升到多少维，out 表示输出维度，SE 表示是否使用注意力机制，NL 表示激活函数使用 Relu 或者 hard-swish，s 表示步距。

Table 2 Parameter of MobileNetv3_small backbone network part

表 2 MobileNetv3_small 主干网络部分参数

input	operator	exp size	out	SE	NL	s
224^2×3	Conv2d3x3	-	16	-	HS	2
112^2×16	bneck，3x3	16	16	C	RE	2
56^2×16	bneck，3x3	72	24	-	RE	2
28^2×24	bneck，3x3	88	24	-	RE	1
28^2×24	bneck，5x5	96	40	C	HS	2
14^2×40	bneck，5x5	240	40	C	HS	1
14^2×40	bneck，5x5	240	40	C	HS	1
14^2×40	bneck，5x5	120	48	C	HS	1
14^2×48	bneck，5x5	144	48	C	HS	1
14^2×48	bneck，5x5	288	96	C	HS	2
7^2×96	bneck，5x5	576	96	C	HS	1
7^2×96	bneck，5x5	576	96	C	HS	1
7^2×96	Conv2d1x1	-	576	C	HS	1

MobileNetv3 中的 Bneck 模块主要包含深度可分离卷积（DSConv），其过程如图 1 所示[11]。深度卷积对每个输入通道单独进行卷积操作，生成多个通道的特征图。逐点卷积（Pointwise Convolution）是在上述特征图的基础上应用 1×1 卷积，将输出通道数减少到所需的数量，可以显著减少参数量和计算成本。

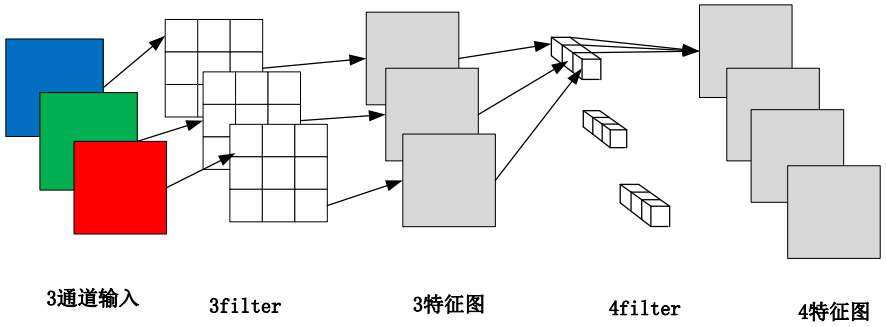

Fig. 1 Depthwise separable convolution

图 1 深度可分离卷积

SE注意力机制的主要步骤是将池化后每个通道的特征图转换为一个数字，以获取该通道的全局信息，如图2所示[12]。SE机制首先对输入通道进行全局平均池化，得到一个向量。其次，将池化后的向量通过全连接层，并应用ReLU激活函数。再次，经过一个全连接层并应用hard_sigmoid激活函数。最后，经过激活函数处理后的结果在0～1之间生成一个权重，以产生该通道的激励值。

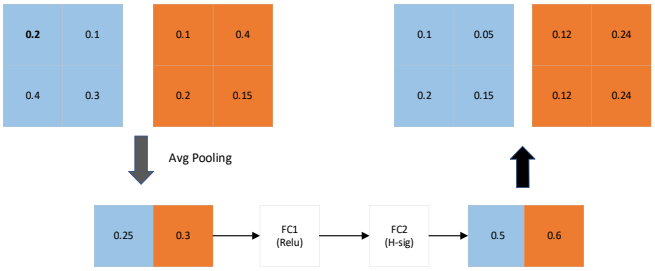

Fig. 2　SE attention mechanism

图2　SE注意力机制

MobileNetV3的倒残差结构通常使用倒残差块(Inverted Residual Block)进行构建，倒残差块内部包括深度可分离卷积和1×1卷积。这些组件共同构成MobileNetV3的架构，使其在保持轻量级的同时维持良好性能。MobileNetv3模块结构如图3所示[9]。

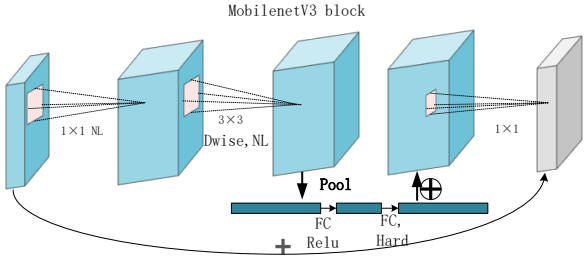

Fig. 3　MobileNetv3 module structure

图3　MobileNetv3模块结构

1.3　用于多尺度检测的SPPF模块

空间金字塔迅速池化模块(Spatial Pyramid Pooling Fast，SPPF)如图4所示[10]。从图中可以看出，输入图像大小为20×20×1 024，经过1×1卷积、归一化以及激活层后，输出图像的通道数降至512。经过不同次数的相同池化得到尺度不一的3张特征图，最后进行特征融合。融合后的图像再进行升维操作，将特征图还原为原有的20×20×1 024格式。该过程实现了在不改变特征图尺寸的情况下执行池化操作，从而捕捉不同尺度下的信息。

1.4　改进后的边界框损失函数

本文使用归一化高斯瓦伦汀距离方法计算分布距离[13]。$\mu_1 = N(m_1, \Sigma_1)$ 和 $\mu_2 = N(m_2, \Sigma_2)$，μ_1 和 μ_2 之间的Wasserstein距离可以表示为：

$$W_2^2(\mu_1,\mu_2) = \| \boldsymbol{m}_1 - \boldsymbol{m}_2 \|_2^2 + \| \boldsymbol{\Sigma}_1^{\frac{1}{2}} - \boldsymbol{\Sigma}_2^{\frac{1}{2}} \|_F^2 \tag{1}$$

式中，$\|.\|$是Frobenius范数。此外，对于从预测框 $A = (cx_a, cy_a, w_a, h_a)$ 与真实框 $B = (cx_a, cy_a, w_b, h_b)$ 建模的高斯分布 N_a 和 N_b，则可以表示为：

$$W_2^2(\mathcal{N}_a,\mathcal{N}_b) = \left\| \left[cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2} \right]^{\mathrm{T}}, \left[cx_b, cy_b, \frac{w_b}{2} \cdot \frac{h_b}{2} \right]^{\mathrm{T}} \right\|_2^2 \tag{2}$$

$W_2^2(\mathcal{N}_a,\mathcal{N}_b)$ 是一种距离之间的度量，所以不能直接用作相似度度量，即用0和1之间的某个值作为IoU。为了能够用一个小数表示IoU，需要对公式进行归一化操作，其中C是与数据集密切相关的常数。在下文实验中，根据经验设定产生了新的衡量标准：

$$NWD(\mathcal{N}_a,\mathcal{N}_b) = \exp\left(- \frac{\sqrt{W_2^2(\mathcal{N}_a,\mathcal{N}_b)}}{C} \right) \tag{3}$$

将GIoU[14]替换YOLOv5-6.0中原有的CIoU[10]，因为GIoU相比于CIoU更简单、稳定、通用，且能够更好地处理长宽比例问题。根据实验结果，得出GIoU的实验效果优于CIoU。

在计算边界框损失函数时，引入归一化高斯瓦伦汀距离以提升对小目标的检测效果。具体来说，将IoU损失与归一化高斯瓦伦汀距离损失进行权衡，得到最终的边界框分类损失函数NewIoU，其定义为0.5倍NWD加0.5倍GIoU。

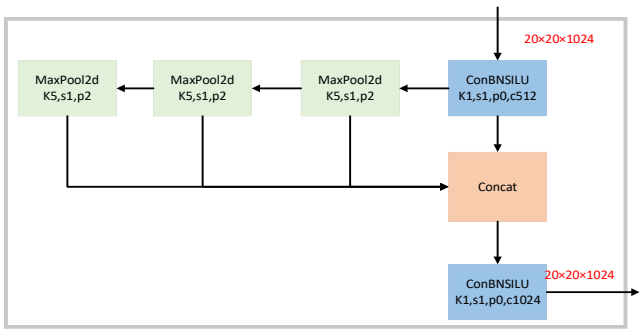

Fig. 4　SPPF module

图4　SPPF模块

1.5　M_YOLO_N整体框架

图5是经过上文改进后得到的M_YOLO_N(MobileNet_YOLOv5_NewIou)模型整体框架。从图5可以看出，与原有的YOLOv5s相比，其主要改变在于主干网进行了更换。

在更换了新的主干网后，引入深度可分离卷积与SE注意力机制。深度卷积对每个输入通道单独进行卷积操作，生成多个通道的特征图。逐点卷积是在上述特征图的基础上应用1×1卷积，将通道数减小到所需的输出通道数。该方式可以显著减少参数量和计算成本，达到轻量化模型的目的，从而确保手势识别的实时性。同时，保留SPPF模块以及改进损失函数，新的损失函数引入了新的真实边界框与预测边界框计算方法(归一化高斯瓦伦汀距离)，而原有的SPPF模块通过金字塔池化提取不同尺寸特征后再进行融合输出，以达到检测不同尺度目标的目的。

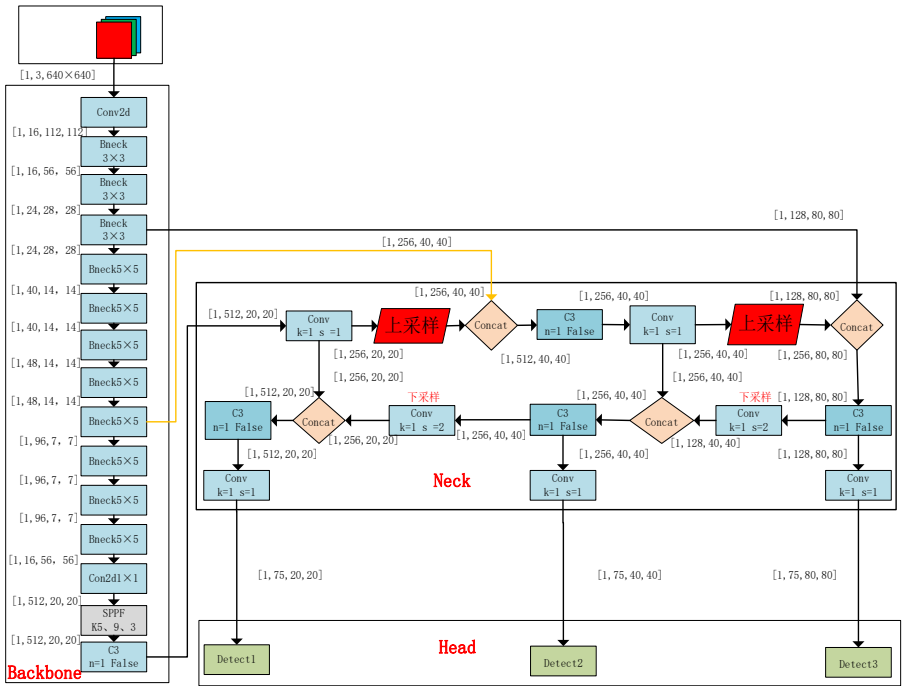

Fig. 5 M_YOLO_N model structure

图5 M_YOLO_N模型结构

Neck的改进不仅提高了模型检测精度、置信度，而且有效改善了模型的多尺度目标检测问题。

2 实验结果分析

2.1 实验配置

本文实验环境如下：NVIDIA Tesla T4，PyTorch1.7.0，CUDA 12.0，CPU为Intel（R）Xeon（R）CPU @2.20GHz。训练过程总共300轮，将学习率下降的方式设置为余弦退火，初始学习率设置为1E-2，且最小学习率为1E-5。训练模型时使用了MobileNetv3_small在ImagNet数据集上的权重作为预训练权重。网络训练参数如表3所示。改进后的模型采用YOLOv5中提供的数据预处理方式，包括随机尺度变换、随机移位、随机左右翻转、Mosaic等数据增强方式。

Table 3 Network training parameter settings

表3 网络训练参数设置

参数	设置值	说明
图像大小	640×640	输入图像的像素大小
Lr	0.01	学习率
Batch-size	4	一次训练的样本数量
Epoch	300	训练的总轮数
Cosine-lr	True	学习率余弦下降
Val-split	0.33	验证集所占比例

2.2 评价指数与评估标准

实验选取的评价指标有每个类别的平均精度（Average Precision，AP）、所有类别平均精度的平均值（Mean Average Precision，mAP）、模型参数量、推理计算复杂度。AP的计算公式中包含了精确率和召回率[15]。

精确率（precision）计算公式为：

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

召回率（Recall）计算公式为：

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

其中，将正样本预测为正的情况称为真正例（TP），将负样本预测为负的情况称为真负例（TN），将负样本预测为正的情况称为假正例（FP），将正样本预测为负的情况称为假负例（FN）。mAP的求解方法是先计算每个类别的平均精度（AP），再将所有类别的AP值进行平均计算，得到mAP值[15]。每个类别的AP值计算方法如下：

$$AP = \sum_{n=1}^{N} (R_n - R_{n-1}) P_n \tag{6}$$

式中，N表示正样本数量，R_n表示在前n个样本中的召回率，P_n表示在前n个样本中的精确率。

mAP的计算常常还会指定一个IoU的阈值，最常见的阈值是0.5。其含义是当IoU的值大于0.5，即会显示出预测框的置信度。因此，mAP@0.5表示IoU阈值为0.5时的mAP值，mAP@[0.5，0.95]表示IoU的阈值在0.5~0.95之间每隔0.05取一个值时的mAP值。

2.3 不同模型对比分析

为了说明改进后模型的优势，将本文模型M_YoLo_N与单节段目标检测模型SSD[16]、YOLOv6s[17]、YOLOv3-tiny[18]、YOLOv4-tiny[19]以及双阶段目标检测模型Faster R_CNN[20]进行比较，结果如表4所示。比较指标主要有模型的参数量、计算复杂度、帧率以及各类别的平均精度均值。

从表4可以看出，相比YOLOv3tiny、YOLOv4tiny，本文

Table 4　Comparative results of recognition and detection experiments of different algorithms on self-made data sets

表4　不同算法在自制数据集上的识别检测实验对比结果

模型	FPS	参数/10^6	GFLOPs	mAP@.5
YOLOv3[18]	63.36	56.3	156.3	0.931
YOLOv3t[18]	131.11	5.4	**3.4**	0.912
YOLOv4t[19]	138.27	10.7	6.3	0.926
SSD[16]	78.52	24.40	50.9	0.866
F_R_CNN[20]	55.31	45.43	114.7	0.944
YOLOv6s[17]	85.92	18.50	48.7	**0.983**
M_YOLO_N	**152.21**	**4.60**	7.2	0.978

模型虽然计算复杂度稍高，但是准确率有大幅提升，由91.2%提升到97.8%。同时，本文模型相比传统模型的参数量大大减小，本文模型的M_YOLO_N参数只有FasterR-CNN的1/10，但是准确率仍有所提升，达到了实验预期的轻量化目的。此外，本文模型在准确率略低的情况下，体积小，检测速度比YOLOv6s以及SSD更快，能够很好地实现对静态手势识别实时检测的目的。综合来说，改进后的模型M_YOLO_N的性能明显优于部分单阶段、双阶段以及轻量化目标检测模型。

将改进后的模型与YOLOv5的3个不同模型进行比较，结果如表5所示[21]。

Table 5　Experimental comparison results of YoLov5 models of different sizes on self-made datasets

表5　不同规模的YoLov5模型在自制数据集上的实验对比结果

模型	权重/10^6	参数/10^6	GFLOPs	mAP@.5
YOLOv5s	14.4	7.02	15.8	0.958
YOLOv5m	42.2	20.8	47.9	0.965
YOLOv5l	92.8	46.13	107.7	0.972
M_YOLO_N	**9.6**	**4.60**	**7.2**	**0.978**

从表5可以看出，原有的YOLO模型随着大小的增加，模型参数量、计算复杂度及权重大小都成倍增加，但是平均精度均值仅增加了1.4%。而M_YOLO_N的精确率要比YOLOv5l高0.06%，同时模型参数量由46.13 million降低至7.02 million，极大地降低了对终端资源的要求。为了达到手势识别实时检测这一预期目的，综合来看本文模型更符合实际需求，整体表现优势明显。由此可以得出，对于小型数据集，仅仅通过堆叠参数以提高精确度，不如针对性地优化模型效果更好。

2.4　消融实验

通过在原有模型下不断改进，以期达到更好的检测效果，因此进行了消融实验，结果如表6所示。其中，模型1是YOLOv5-6.0small的原有模型，模型2是将主干网更换为MobileNetv3-small后的模型（引入深度可分离卷积以及SE注意力机制），模型3是更换完主干网后添加了SPPF模块的模型，模型4是M_YOLO_N（MobileNet+SPPF+NewIoU）模型。

从表6可以看出，模型2相对于模型1引入了深度可分离卷积，导致模型参数量降低了近50%，计算复杂度也随之降低了57%。同时，模型2相比模型1的mAP提升了1.6%，在实现模型轻量化的同时，略提高了平均准确率。模型3在添加了用于多尺度特征提取的SPPF模块后，虽然模型体积有所增加，但模型对于手势识别的准确率也有所提高。模型4（M_YOLO _N）没有增加额外的参数量，但是由检测结果可知，检测效果仍有一定提高。M_YOLO_N实现了预期中的模型轻量化以及提升准确率的目的。

Table 6　Ablation experiment results

表6　消融实验结果

模型	参数量/10^6	mAP	GFLOPs
YOLOv5s	7.02	0.954	15.8
MobileNet_YOLO	3.73	0.970	6.5
Mobile_YOLO_SPPF	4.60	0.978	7.2
M_YOLO _N	4.60	0.978	7.2

改进前后的目标检测结果对比如图6所示。其中，左侧为原模型的检测结果，右侧为M_YOLO_N模型的检测结果。从图6以及消融实验结果可以看出，改进后模型的mAP提升了2.4%，表明手势识别的置信度显著提高。手势识别检测置信度的提升是一个积极的结果。保留SPPF模块并修改损失函数后，在不改变模型大小的情况下，多尺度目标检测结果也有明显改善。该方式不仅实现了实时检测的预期效果，而且改善了多尺度目标检测的效果。

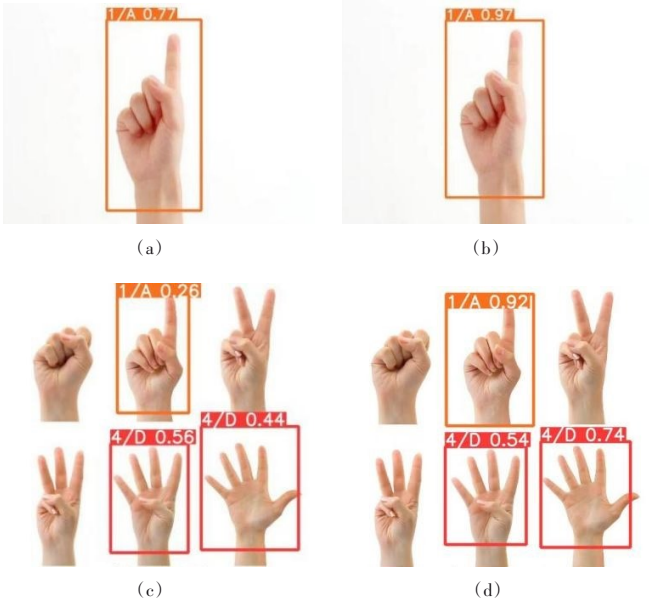

Fig. 6　Comparison of target detection results before and after improvement

图6　改进前后目标检测结果对比

3　结语

对于简单静态手势识别任务，YOLOv5网络模型存在参数冗余、模型结构复杂以及检测结果置信度不高等问题。本文进行了针对性优化，将更为轻量的MobileNetv3主干网替换原有主干网，从而实现模型的轻量化以及检测实时性的要求。本文模型M_YoLO_N保留了YOLOv5中的

SPPF模块,引入新的边界框损失函数计算方式,从而改善了手势识别的多尺度准确性以及检测效果。最终,M_YOLO_N模型相比原模型的参数量降低了33%,计算复杂度降低了54%,同时所有类别的平均精度提升了2.4%。

然而,本文的数据集尚存在不足,缺少多尺度手势识别图。因此,下一步的优化方向是完善数据集,并在完善后的数据集以及大型公开数据集上测试模型性能。此外,还可以尝试将模型部署到终端设备上,以满足实际应用中的实时检测要求。

参考文献:

[1] JONES M J, REHG J M. Statistical color models with application to skin detection [J]. International Journal of Computer Vision, 2002, 46: 81-96.

[2] HU Z C, ZHOU Y T, SHI B J, et al. A static gesture recognition algorithm combining attention mechanism and feature fusion [J]. Computer Engineering, 2022, 48(4): 240-246.
胡宗承,周亚同,史宝军,等. 结合注意力机制与特征融合的静态手势识别算法[J]. 计算机工程,2022,48(4):240-246.

[3] LING B, GUO Y, ZHAO Y H, et al. Fusion of color information and depth information for GrabCut image segmentation [J]. Computer Application and Software, 2020, 37(8): 188-193.
凌滨,郭也,赵永辉,等. 融合彩色信息和深度信息的GrabCut图像分割[J]. 计算机应用与软件,2020,37(8):188-193.

[4] SHU Z C, CAO S X, XIE D L, et al. Research on a novel method for semantic recognition of digital gestures based on three-dimensional visual features [J]. Journal of Electronic Measurement and Instrumentation, 2021,35(6): 124-130.
舒子超,曹松晓,谢代梁,等. 基于三维视觉特征的数字手势语义识别新方法研究[J]. 电子测量与仪器学报,2021,35(6):124-130.

[5] LIU S P, LIU Y, YU J, et al. Hierarchical static gesture recognition combining finger detection and HOG features [J]. Journal of Image and Graphics, China, 2015, 20(6): 781-788.
刘淑萍,刘羽,於俊,等. 结合手指检测和HOG特征的分层静态手势识别[J]. 中国图象图形学报,2015,20(6):781-788.

[6] SMEDT Q D , WANNOUS H , VANDEBORRE J P. Heterogeneous hand gesture recognition using 3D dynamic skeletal data[J]. Computer Vision and Image Understanding, 2019, 181: 60-72.

[7] OHARA K, MAEKAWA T, SIGG S, et al. Preliminary investigation of position independent gesture recognition using Wi-Fi CSI[C]//2018 IEEE International Conference on Pervasive Computing and Communications Workshops, 2018: 480-483.

[8] GU X J, ZHOU Z P, GUO Y C, et al. Dynamic gesture recognition method based on CNN-LSTM hybrid model[J]. Computer Applications and Software, 2021, 38(11): 205-209.
谷学静,周自朋,郭宇承,等. 基于CNN-LSTM混合模型的动态手势识别方法[J]. 计算机应用与软件,2021,38(11):205-209.

[9] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3 [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision,2019: 1314-1324.

[10] XIAO R Q. YOLOV5s-GTB: light-weighted and improved YOLOV5s for bridge crack detection[DB/OL]. https://arxiv. org/abs/2206. 01498.

[11] CHOLLET F. Xception: deep learning with depthwis separable convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017: 1251-1258.

[12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018: 7132-7141.

[13] WANG J, XU C, YANG W, et al. A normalized Gaussian Wasserstein distance for tiny object detection [DB/OL]. https://arxiv. org/abs/2110. 13389.

[14] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019: 658-666.

[15] BELLET A, HABRARD A, SEBBAN M. A survey on metric learning for feature vectors and structured data [DB/OL]. https://arxiv. org/abs/1306. 6709.

[16] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]// 14th European Conference on Computer Vision, 2016: 21-37.

[17] LI C, LI L, JIANG H, et al. YOLOv6: a single-stage object detection framework for industrial applications [DB/OL]. https://arxiv. org/abs/2209. 02976.

[18] REDMON J, FARHADI A. Yolov3: an incremental improvement [DB/OL]. https://arxiv. org/abs/1804. 02767.

[19] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection [DB/OL]. https://arxiv. org/abs/2004. 10934.

[20] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision,2015: 1440-1448.

[21] ZENG Y, GAO F Q. Defect detection algorithm for electronic component surface based on improved YOLOv5 [J]. Journal of Zhejiang University (Engineering Science), 2023, 57(3): 455-465.
曾耀,高法钦. 基于改进YOLOv5的电子元件表面缺陷检测算法[J]. 浙江大学学报(工学版),2023,57(3):455-465.

(责任编辑:黄 健)