小 型 微 型 计 算 机 系 统
Journal of Chinese Computer Systems
DOI:10.20009/j.cnki.21-1106/TP.2023-0391
2025 年 1 月 第 1 期
Vol.46 No.1 2025

改进 YOLOpose 的轻量化多人姿态检测模型

张欣毅[1],张运楚[1,2],王 菲[1],刘一铭[1]

[1](山东建筑大学 信息与电气工程学院,济南 250101)
[2](山东省智能建筑技术重点实验室,济南 250101)
E-mail:zzzxxxyyy777@163.com

摘 要: 二维人体姿态估计对安全生产、智能交互等研究都有重要的意义.针对目前的人体姿态估计模型计算量大、检测速度慢等问题,本文提出一种基于 YOLOpose 模型的轻量化改进算法.首先引入运算更精巧的 GSConv 卷积模块替换普通 Conv 卷积,大大降低模型计算量和复杂度;然后用 CARAFE 模块替换 UPSample 模块,完成上采样工作,同时引入 CBAM 注意力机制模块以避免模型轻量化带来的精度降低的问题.实验结果表明,YOLOpose 模型经过上述轻量化改进后,模型体量降低为 135.6MB,降低了约 15.8%,GFLOPS 降为了 86.9,降低了约 15.0%,模型计算量显著降低,再加入 CBAM 注意力机制对模型精度影响较小,改进后模型既可以保证识别的准确度,又可以实现检测算法的轻量化.

关 键 词: 姿态估计;YOLOpose;轻量化;GSConv 卷积;CARAFE 模块
中图分类号: TP391 **文献标识码:** A **文 章 编 号:** 1000-1220(2025)01-0167-06

Lightweight Multiplayer Pose Detection Model with Improved YOLOpose

ZHANG Xinyi[1],ZHANG Yunchu[1,2],WANG Fei[1],LIU Yiming[1]

[1](School of Information and Electrical Engineering,Shandong Jianzhu University,Jinan 250101,China)
[2](Shandong Key Laboratory of Intelligent Buildings Technology,Jinan 250101,China)

Abstract: 2D human posture estimation is of great significance for safety production,intelligent interaction and other research. Aiming at the current human posture estimation model with large computation and slow detection speed,this paper proposes a lightweight improvement algorithm based on the YOLOpose model. Firstly,the GSConv convolution module,which is more delicate in operation,is introduced to replace the ordinary Conv convolution,which greatly reduces the model computation and complexity; then the UPSample module is replaced by the CARAFE module to complete the up-sampling work,and at the same time,the CBAM attention mechanism module is introduced to avoid the problem of reduced accuracy brought by the model lightweighting. The experimental results show that after the YOLOpose model is improved by the above lightweighting,the model volume is reduced to 135.6MB,which is reduced by about 15.8%,and the GFLOPS is reduced to 86.9,which is reduced by about 15.0%,with a significant reduction in the model computation volume,and then the addition of the CBAM attention mechanism has a small effect on the model accuracy,and the improved model can ensure the accuracy of recognition and also realize the lightweight of the detection algorithm.

Keywords: attitude estimation;YOLOpose;lightweighting;GSConv convolution;CARAFE module

0 引 言

姿态估计是计算机视觉研究热点之一,广泛应用于无人驾驶、智能交互、安全检测等领域.人体关键点检测是行人姿态估计和状态检测的方法之一,通过人体关键点检测可以实现对行人所处状态的精确判断,利用关键点坐标可以计算人体各部位位置及判断人体目前所处的状态为站立、蹲下、躺平等.目前常见的姿态估计方法主要有 2D 姿态估计和 3D 姿态估计技术,其中 3D 姿态估计依托的设备成本高,需要捕捉更多的关键点形成空间三维坐标模型.而 2D 姿态估计需要的成本更低,实现更简单,且近几年发展速度较快,应用场景更广泛,有更高的研究价值.基于传统方法的人体姿态估计主要是基于图结构和形变部件模型,建立图模型中各部件连通性,同时利用人体运动学的约束条件不断优化模型进而实现人体姿态估计.但该方法模型结构单一,可识别姿态具有局限性,适用范围有较大限制.

基于深度学习的 2D 姿态估计主要有自上而下和自下而上两种方法,其检测机制、优缺点对比如表 1 所示.

Alexander Toshev 等人[1]提出了 DeepPose 人体姿态估计算法,首次采用深度神经网络(Deep Neural Networks,DNNs)进行特征提取和关键点坐标回归,但该算法精确度不稳定.Papandreou 等人[2]提出的 G-RMI 采用了一种基于关键点的独特置信估计形式代替基于目标框的评分,同时用基于关键点的新非最大抑制(NMS)代替基于人体的较粗糙的 NMS.Kumar[3]等人提出了一种基于 SSD(Single Shot MultiBox Detector)的多人姿态估计模型,该方法使用多任务权重共享架

收稿日期:2023-11-11 收修改稿日期:2023-12-10 基金项目:国家自然科学基金项目(62003191)资助. 作者简介:张欣毅,女,1999 年生,硕士研究生,研究方向为模式识别、计算机视觉;张运楚,男,1968 年生,博士,教授,研究方向为模式识别、智能信息处理;王 菲,女,2001 年生,硕士研究生,研究方向为图像处理、机器视觉;刘一铭,女,1999 年生,硕士研究生,研究方向为目标检测、深度学习.

构来联合训练检测和姿势估计,这种模块化架构可以适应不同的任务. Fang 等人[4]设计了一种区域多人姿态估计(RMPE)框架,使用对称空间变压器网络(SSTN)和参数位姿非最大抑制(PNMS)以减少网络冗余,以便在存在不准确的人体边界框的情况下进行姿态估计;Chen 等人[5]提出了一种名为层叠金字塔网络(Cascaded Pyramid Network,CPN)的新型网络结构算法,首先通过 GlobalNet 特征金字塔网络定位容易识别的关键点,再利用 RefineNet 网络整合 GlobalNet 的各级特征来确定不易识别的关键点;Sun 等人[6]提出了一个可以持续保持高分辨率表征的网络,从高分辨率子网络开始逐渐添加高分辨率到低分辨率的子网络,形成更多阶段,并并行连接多分辨率子网络,反复进行多尺度融合,使每个从高分辨率到低分辨率的表征反复接收来自其他并行表征的信息,从而形成丰富的高分辨率表征. 但自上而下的方法检测网络复杂,不支持实时工程的实现.

表 1 2D 姿态估计方法对比

Table 1 Comparison of 2D attitude estimation methods

策略	机制	优点	缺点
自上而下	先检测人体,再进行单人姿态估计	精度较高,对低分辨率人体检测效果较好	时间复杂度与图片中检测到的人数成正比
自下而上	先检测所有人的关键点,再将关键点聚类到每个人	一次性检测所有关键点,效率高,模型小	由于关键点相似度高,目前准确度较低

MoveNet 是 Beletti F 等人[7]设计的集成在 MediaPipe 上的人体姿态估计模型,是一种自下而上的单人姿态方法,速度较快,但无法实现多人检测. 李佳[8]提出了一种基于高斯响应热图编码图像的自底向上的多人姿态估计方法,设计了推理高斯热图的多阶段卷积神经网络实现骨骼关键点和其位置信息的匹配;Luo[9]提出了一种尺度自适应热图回归方法,可以实现自适应调整每个关键点的标准偏差,可以解决尺度和标签模糊的问题;Cheng[10]等人提出了 HigherHRNet,这是一种利用高分辨率特征金字塔学习尺度感知表征的自下而上人体姿态估计方法,利用多分辨率监督进行训练,并利用多分辨率聚合进行推理,能够解决自下而上的多人姿势估计中的尺度变化难题,并能更精确地定位关键点;上述提出的都是基于热图的姿态估计方法,但依赖于热图和复杂后处理分组的算法具有不可避免的量化误差. 还有 Kreiss[11]等人利用部件强度场定位身体部位,利用部件关联场将身体部位彼此关联得以完整人体,也是由于这种定位方法限制了其识别精度.

自下而上的方法对于多人检测更具优势,且自下而上方法的研究近几年也发展较快,出现了性能较好的模型,如 OpenPose[12]、YOLOpose(You Only Look Once Pose)[13]等. OpenPose 是自下而上的多人姿态估计方法的典型模型,但是 OpenPose 运行速度极慢,难以满足工业现场实时检测的要求,因此 Openpose 并不是行人姿态估计的上上之选. 而 Debapriya Maji 提出的 YOLOpose 模型很好的解决了这一问题,推理所需消耗的计算成本相对较小,且精度较高,但对于实时要求较高的应用场景,如安全实时监测、无人机实时监测等情况下难以满足实际工业应用要求,所以需要对该模型在不影响精度的情况下,降低模型参数,提高检测效率.

本文提出了一种基于改进 YOLOpose 模型的多人姿态估计模型,包括用运算更精巧的 GSConv 卷积模块替换普通 Conv 卷积,用 CARAFE 模块替换 UPSample 模块,完成上采样工作,同时,避免模型轻量化带来的精度降低的问题引入 CBAM 注意力机制模块. 由此,改进 YOLOpose 模型在计算参数量、模型精度上都有较好的保证,满足工业现场部署的要求.

1 YOLOpose 姿态检测模型

YOLOpose 模型是在 YOLOv7[14]框架基础上同时实现目标框检测和关键点检测,主要由 4 部分组成,分别为 Input 输入端、Backbone 网络、Neck 网络和 Head 网络. 其网络结构如图 1 所示. 首先图片进入输入端,图片大小为 640×640,输入

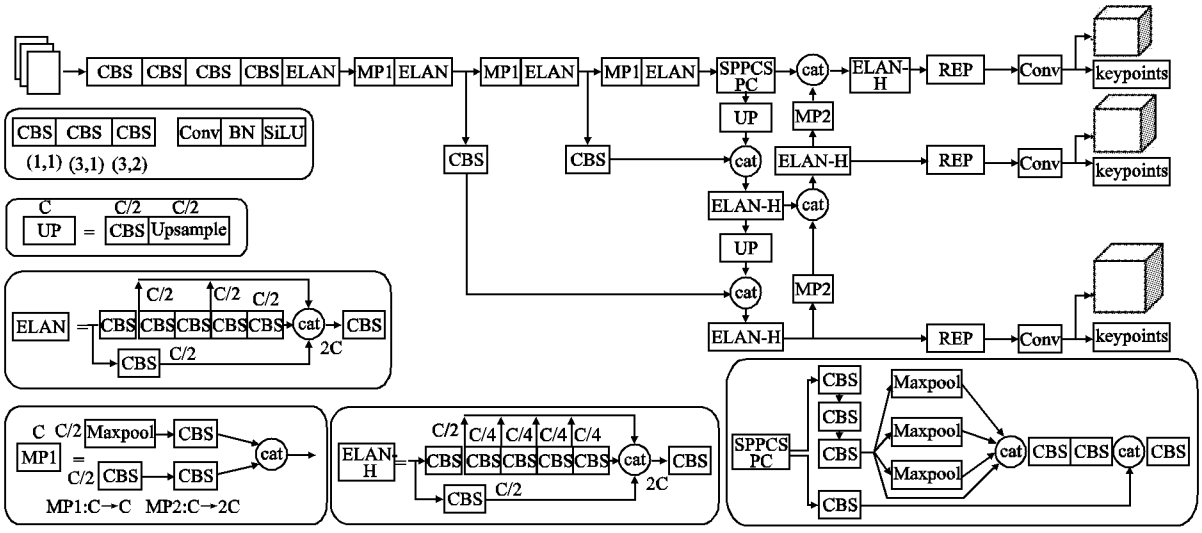


图 1 YOLOpos 网络结构图

Fig. 1 YOLOpose network structure diagram

到 Backbone 网络中进行预训练,完成特征提取,然后经 Neck 层网络对图像进行分类,输出 3 层不同大小的特征图,最后调融合各特征层特征,融合其位置信息和语义信息,最后 Head 整通道数输出人体框和关键点预测结果.

训练前需要根据网络结构设计数据集,预测输出为 17 个关键点,则数据集中图片标记 17 个关键点,对应标签号如图 2 所示,其中,0 代表鼻子、1 代表左眼、2 代表右眼、3 代表左耳、4 代表右耳、5 代表左肩、6 代表右肩、7 代表左肘、8 代表右肘、9 代表左腕、10 代表右腕、11 代表左胯、12 代表右胯、13

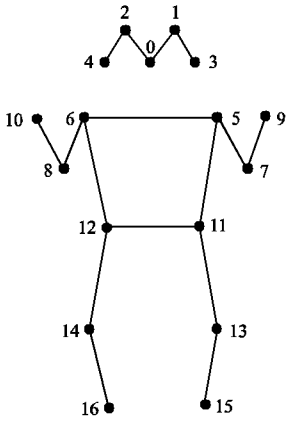


图 2　17 个人体关键点图

Fig. 2　17 key points of the human body diagram

代表左膝、14 代表右膝、15 代表左踝、16 代表右踝. 将标记全部人体框和各人体框内 17 个关键点的图像输入 YOLOpose 网络中,经过各部分运算输出,输出也为标记 17 个关键点和人体框的图像.

2　轻量化 YOLOpose 姿态检测模型

2.1　GSConv 模块

标准卷积(Standard Convolution,SConv)是对 3 个通道同时操作,卷积核的个数等于输出通道数,卷积核的通道数等于输入通道数,所以使用过多的标准卷积对图像进行特征提取时,会造成参数量的累积、特征的冗余,层数越深,影响越大. Ghost Conv 模块是 Han K 等人[15]提出的卷积模块,在提取到有效特征的同时降低了参数和计算消耗. 其操作分为两步:分别是少量卷积和线性变换操作,最后将两个操作得到的特征图拼接在一起输出. Ghost Conv 因其出色的性能,多被用于计算机视觉模型轻量化的研究中,但 Ghost Conv 模块在其第 2 步操作中却会丢失大量的通道信息,为克服此缺点,Li H 等人[16]提出了 GSConv 模块,GSConv 结构如图 3 所示.

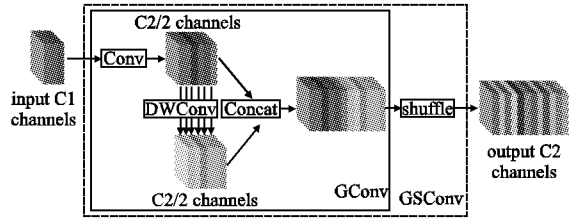


图 3　GSConv 模块结构图

Fig. 3　GSConv module structure diagram

设输入通道数为 C1,输出通道数为 C2. 首先经过一个标准卷积,通道数变为 C2/2,再经过一个深度可分离卷积,通道数不变,最后将两次卷积的结果进行拼接和混洗. 最后的混洗操作,能够将通道信息进行均匀打乱,增强提取到的语义信息,加强特征信息的融合,提高图像特征的表达能力. 网络在 Neck 层进行特征融合时,语义信息也会不断地向下传输,当特征图的高宽和通道数被不断压缩和扩展时,会导致部分语义信息的丢失,影响最后的预测,具有一定的局限性. 本文在网络的 Head 层引入 GSConv 模块,使用 GSConv 模块代替标准卷积来进行上采样和下采样,降低模型的参数量和计算量,并最大程度保证采样效果,保证训练精度.

2.2　CARAFE 模块

YOLOpose 原模型中采用的上采样模块为 UPSample 模块,该模块主要是采用最近邻插值的方法进行上采样,即通过将距离目标像素点最近的像素值赋给该目标像素点放大图像,但是这种方法不能增加像素信息,也损失了原图像像素点间的渐变关系. 而 CARAFE 模块[17]通过上采样核预测和特征重组可以完成上采样,同时更具优势,如 CARAFE 可以对目标像素点集合更大范围的信息,支持具体内容具体处理的操作,具有自适应内核,且计算量极小,具有明显的优越性.

CARAFE 上采样结构如图 4 所示,该过程主要分为两个模块,核预测模块和特征重组模块.

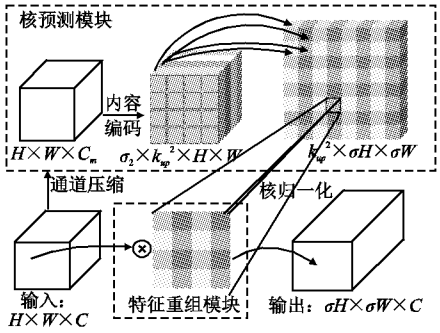


图 4　CARAFE 模块结构图

Fig. 4　CARAFE module structure diagram

CARAFE 模块的算法步骤如下:

1)对输入特征图进行通道压缩,输入尺寸为 $H\times W\times C$,压缩处理后尺寸为 $H\times W\times C_m$,大大减小了计算量.

2)将通道压缩后的特征图进行上采样核预测,利用大小为 $k_{en}\times k_{en}$ 的卷积核对上述压缩特征图进行内容编码得到尺寸为 $\sigma_2\times k_{up}^2\times H\times W$ 的特征图,然后在通道维上展开,此时尺寸变为 $k_{up}^2\times\sigma H\times\sigma W$.

3)利用 Softmax 函数进行归一化处理,使得上采样核的权重之和为 1.

4)将输入特征图与预测的上采样核进行卷积运算得到最终的上采样结果.

CARAFE 上采样过程的参数量如式(1)所示:

$$P=C\times C_m+\sigma^2k_{up}^2(\sigma^2k_{en}^2C_m+1)+\sigma^2k_{up}^2C \quad (1)$$

其中,$k_{en}\times k_{en}$ 为内容编码的卷积核尺寸,$k_{up}\times k_{up}$ 为预测的上采样核尺寸. CARAFE 模块的自适应性大大减轻了冗余计算量,且与 UPSample 等其他上采样方法相比,可以融合更多信息,采样效果更优越.

2.3　CBAM 注意力机制模块

YOLOpose 模型经过上述轻量化改进后,训练好的模型体量有明显降低,但模型精度也会有相应损失,为了减少这种

损失、提高模型精度,本文设计在SPPCSPC模块增加CBAM[18]注意力机制.

CBAM(Convolutional Block Attention Module)是一个相对轻量级模块,在模型中,输入特征图有两个相互独立的维度,分别为通道和空间,将注意力映射图与输入特征图进行乘积,可实现对图像的自适应特征注意集中. CBAM 模块的结构如图 5 所示.

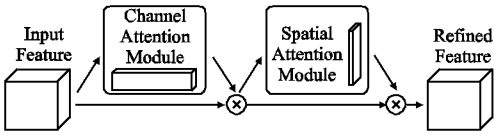


图 5 CBAM 模块结构图

Fig. 5 CBAM module structure diagram

通道注意力机制主要关注目标的特征信息,对于一个特征图 I,其通道注意力计算公式如下:

$$Fc(I) = \sigma Conv2[Conv1(I^c_{avg})] + \sigma Conv2[Conv1(I^c_{max})] \tag{2}$$

其中,I^c_{avg}为全局平均池化,I^c_{max}为全局最大池化,σ 为加权值.

空间注意力机制主要关注目标的位置信息,其特征图 Fs(I)计算公式如下:

$$Fs(I) = \sigma[Conv^{7\times7}(I^s_{avg} + I^s_{max})] \tag{3}$$

而 CBAM 为通道注意力机制和空间注意力机制结合组成,其计算公式如下:

$$I' = Fs[Fc(I) \times I] \times [Fc(I) \times I] \tag{4}$$

CBAM 注意力机制在深度学习架构中可以实现即插即用,同时可以以端到端方式进行训练,因此可以加入到 YOLOpose 模型中.

3 实验与结果分析

3.1 数据集与实验环境

本文实验使用微软构建的 MS COCO(Microsoft Common Objects in Context)2017 数据集,主要用于目标检测、分割、关键点检测等,包括训练集 56599 张、验证集 2346 张,其中包括单人和多人、大目标和小目标等多类型图片.

实验环境使用 Windows10 操作系统,CPU 型号为 Intel Core i9-10900X,使用 NVIDIA GeForce GTX 3090 显卡进行运算,CUDA 版本为 11.6,Pytorch 版本为 1.13.0,Python 语言环境为 3.8. 设置 YOLOpose 训练的参数,其中,初始学习率(learning rate)为 0.01,批处理大小(batch size)为 16,迭代轮数(epochs)为 200,输入图像分辨率为 640×640.

3.2 评价指标

为了准确评估深度模型在姿态检测图像上的检测性能,本文采用检测算法评估公认度最高的 mAP(mean Average Precision,平均精度均值),即数据集中各类精度的平均值,在目标检测任务中作为衡量检测精度的重要指标.

P(precision,准确率)、R(recall,召回率)、AP、mAP 定义如下:

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

$$AP = \int_0^1 P(r)\,dr \tag{7}$$

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{k} \tag{8}$$

其中 TP 为真正例,FP 为假正例,TN 为真反例,FN 为假反例.

mAP@0.5 表示 IOU 阈值为 0.5 时的平均精度,主要用于评估模型的识别能力. FPS 表示每秒检测到的图像数,值越高,检测速度越快. 计算量 FLOPs 表示模型复杂程度,体现模型体量大小. 综上,本文主要用 mAP、FPS 和 FLOPs 来作为模型的评价指标.

3.3 消融实验

为验证本文改进算法的有效性,在相同数据集上进行了消融实验. 在 YOLOpose 原模型的基础上,依次进行修改:引入 CARAFE 上采样模块、引入 GSConv 卷积模块、引入 CBAM 注意力机制. 该实验主要采用 mAP@0.5、mAP@0.5:0.95、GFLOPs 和模型大小作为评价指标,消融实验结果如表 2 所示.

表 2 消融实验

Table 2 Ablation experiments

Model	mAP@0.5	mAP@0.5:0.95	GFLOPs	模型大小/MB
YOLOpose	0.937	0.724	102.2	161.1
YOLOpose + CARAFE	0.94	0.728	102.3	161.3
YOLOpose + CARAFE + GSConv	0.935	0.72	86.9	135.6
YOLOpose + CARAFE + GSConv + CBAM	0.936	0.723	87.1	135.7

由表 2 可以看出,加入 CARAFE 上采样模块后模型精度有所提高,模型体量也有所上升;再加入 GSConv 卷积模块后,YOLOpose 模型经过上述轻量化改进后,训练好的模型体量从 161.1MB 降低为 135.6MB,降低了约 15.8%,GFLOPS 由 102.2 降为了 86.9,降低了约 15.0%,模型计算量显著降低,检测速度提高了 29%,模型精度由 93.7% 降低为 93.6%,影响较小;再加入 CBAM 注意力机制,模型精度有所提高. 综上所述,改进后模型体量显著降低,实现轻量化目标,满足计算需求,同时模型精度影响不大.

3.4 主流算法对比实验

为了更加客观展现本文选择的 YOLOpose 模型在多人姿态估计方面的效果,首先进行主流算法对比实验,将本文选择的 YOLOpose 模型与其他主流模型的算法对比,进行对比实验的模型包括 Deeppose、Openpose、Alphapose、HRNet、YOLOpose 和本文改进的 YOLOpose 算法,统一选择 COCO2017 数据集进行实验. 评价指标选择 mAP@0.5、FPS,这两个参数即可实现对模型精度和速度的性能对比,mAP@0.5 越大,模型准确度越高,FPS 越大,模型检测速度越快,实验结果如表 3 所示.

从表 3 可以看到,本文选择的 YOLOpose 模型的 FPS 指标最高,大大超过了其他模型的指标,改进后检测速度明显超过改进模型前,同时可实现 Deeppose 实现不了的多人检测,

且该模型并没有因为检测速度提高而明显降低模型精度,可

表 3　主流算法性能对比

Table 3　Performance comparison of mainstream algorithms

模型名称	mAP@0.5(%)	FPS
HRNet[6]	86.3	9
Openpose	84.9	10
Alphapose[19]	89.6	12
Deeppose	61.8	15
YOLOpose	93.7	31
改进 YOLOpose	93.6	39

以达到实时检测的效果,因此实验构建的二维人体姿态估计模型在检测精度 mAP 和速度综合指标上都有较好的性能.

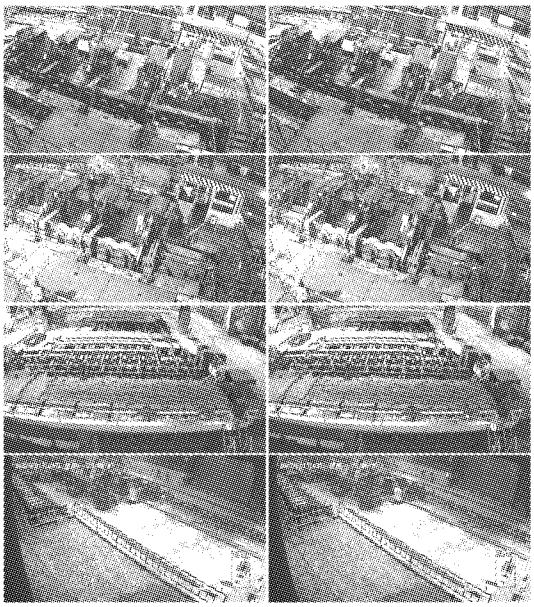
3.5　改进算法对比实验

为了验证本文采取的改进算法的有效性,进行改进算法对比实验,实验模型包括将改进的 YOLOpose 模型与 COCO 2017 数据集上其他改进模型的平均精度和检测速度进行比较,实验结果如表 4 所示.

表 4　改进算法性能对比

Table 4　Performance comparison of improved algorithms

模　型	mAP@0.5(%)	FPS
HigherHRNet[10]	88.2	7
EfficientHRNet-H[20]	82.6	23
LightweightOpenPose[21]	62.8	26
LCSA-YOLOpose[22]	89.0	33
YOLOpose	93.7	31
改进 YOLOpose	93.6	39



(a)改进前　　(b)改进后

图 6　改进前后模型检测对比

Fig. 6　Comparison of model detection before and after improvement

从表 4 可以看出,本文改进的 YOLOpose 模型 mAP 值大大超过其他改进模型,且检测速度明显较快,与同样针对 YOLOpose 的改进模型 LCSA-YOLOpose 相比模型精度和速度都有明显优势,证明本文采用的引入 GSConv 卷积、CARAFE 上采样模块对轻量化模型有较好的效果.

将改进前后模型对图片进行检测对比,如图 6 所示.

图 6 中左侧为 YOLOpose 原模型训练后检测效果图,右侧为改进模型训练后检测效果图,对比实验选择 4 组复杂环境、复杂姿态对照组. 第 1 组为多人姿态估计,且一人为蹲下姿势,改进前后检测较准确,改进后蹲姿工人膝盖位置检测更准确;第 2 组环境光照不均且为多人检测,改进前存在漏检,改进后模型增加了注意力机制,对小目标检测更有优势,所以将全部人准确检出;第 3 组采集图片为广角图,存在微变形,且目标较小,模型人体框及关键点可准确检出;第 4 组图片分辨率较低,人体框及关键点均准确检出,但改进后模型由于轻量化带来的精度损失有所体现,置信度有微弱降低. 虽然模型改进后带来一定的精度降低,但注意力机制的引入使小目标检测更加准确,同时模型检测速度大幅下降,满足实时检测需求,更适合在工业现场部署.

4　结束语

由于现有的姿态估计模型计算量和参数量都较大,不利于设备在现场实时检测,针对这个问题,本文对 YOLOpose 模型进行轻量化改进,将 GSConv 卷积模块、CARAFE 上采样模块引入 YOLOpose 模型中,大大提高了检测速度,同时将 CBAM 注意力机制引入模型来提升检测精度. 消融实验展现了 YOLOpose 引入各模块的优势,与其他主流算法对比展现了 YOLOpose 模型的优越性,与其他改进模型对比表现了本文改进模型的效果,最后通过模型改进前后训练后检测效果表现了实际检测精度良好,综上本文改进的 YOLOpose 模型在精度和速度上均表现优异,符合轻量化要求,更适宜部署在要求实时检测是设备中.

References:

[1] Toshev Alexander, Szegedy Christian. DeepPose: human pose estimation via deep neural networks [C]//Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014: 1653-1660.

[2] Papandreou George, Zhu Tyler, Kanazawa Nori, et al. Towards accurate multi-person pose estimation in the wild [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 4903-4911.

[3] Kumar Chandan, Ramesh Jayanth, Chakraborty Bodhisattwa, et al. VRU pose-SSD: multiperson pose estimation for automated driving [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(17): 15331-15338.

[4] Fang Haoshu, Xie Shuqin, Tai Yuwing, et al. RMPE: regional multi-person pose estimation [C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017: 2334-2343.

[5] Chen Yilun, Wang Zhicheng, Peng Yuxiang, et al. Cascaded pyramid network for multi-person pose estimation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 7103-7112.

[6] Sun Ke, Xiao Bin, Liu Dong, et al. Deep high-resolution representation learning for human pose estimation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recogni-

tion(CVPR),2019:5693-5703.
[7] Beletti F, Chen Y H, Oerlemans A, et al. MoveNet single pose: model card[EB/OL]. Google Ronny Votel (Google Research), Next-Generation Pose Detection with MoveNet and TensorFlow. js, https://storage. googleapis. com/movenet/MoveNet. Singlepose\%20Model\%20card. pdf,2021.
[8] LI J. A study of bottom-up multi-person posture estimation methods [D]. Hefei: University of Science and Technology of China,2021.
[9] Luo Zhengxiong, Wang Zhicheng, Huang Yan, et al. Rethinking the heatmap regression for bottom-up human pose estimation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR),2021:13264-13273.
[10] Cheng Bowen, Xiao Bin, Wang Jingdong, et al. HigherHRNet: scale-aware representation learning for bottom-up human pose estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR),2020:5386-5395.
[11] Kreiss Sven, Bertoni Lorenzo, Alahi Alexandre. PifPaf: composite fields for human pose estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),2019:11969-11978.
[12] Cao Zhe, Simon Tomas, Wei ShihEn, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2017:7291-7299.
[13] Maji Debapriya, Nagori Soyeb, Mathew Manu, et al. YOLO-pose: enhancing YOLO for multi person pose estimation using object keypoint similarity loss[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022:2637-2646.
[14] Wang ChienYao, Bochkovskiy Alexey, Mark Liao HongYuan. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023:7464-7475.
[15] Han Kai, Wang Yunhe, Tian Qi, et al. GhostNet: more features from cheap operations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Seattle,2020:1577-1586.
[16] Li Hulin, Li Jun, Wei Hanbing, et al. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles[J]. 2022, doi:10.48550/arXiv:2206.02424.
[17] Wang Jiaqi, Chen Kai, Xu Rui, et al. CARAFE: content-aware ReAssembly of FEatures[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision,2019:3007-3016.
[18] Woo Sanghyun, Park Jongchan, Lee JoonYoung, et al. CBAM: convolutional block attention module[C]//European Conference on Computer Vision,2018:3-19.
[19] Fang H S, Li J, Tang H, et al. AlphaPose: whole-body regional multi-person pose estimation and tracking in real-time[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(6):7157-7173.
[20] Neff Christopher, Sheth Aneri, Furgurson Steven, et al. EfficientHRNet efficient scaling for lightweight high-resolution multi-person pose estimation[J]. 2020, doi:10.48550/arXiv.2007.08090.
[21] Osokin Daniil. Real-time 2D multi-person pose estimation on CPU: lightweight OpenPose[J]. 2018, doi:10.48550/ arXiv:1811.12004.
[22] WANG M H, XU W M, JIANG H K. An improved lightweight human pose estimation algorithm[J]. Liquid Crystal and Display, 2023,38(7):955-963.

附中文参考文献:

[8] 李 佳. 自底向上的多人姿态估计方法研究[D]. 合肥: 中国科学技术大学,2021.
[22] 王名赫,徐望明,蒋昊坤. 一种改进的轻量级人体姿态估计算法[J]. 液晶与显示,2023,38(7):955-963.