

doi:10.3969/j.issn.1003-3114.2025.05.015

引用格式:殷贤涛,胡波,李思照.基于YOLOv8n-Pose关键点特征增强估计算法[J].无线电通信技术,2025,51(5):1025-1035.[YIN Xiantao, HU Bo, LI Sizhao. Enhanced Estimation Algorithm Based on YOLOv8n-Pose Key Point Features[J]. Radio Communications Technology, 2025, 51(5): 1025-1035.]

基于 YOLOv8n-Pose 关键点特征增强估计算法

殷贤涛¹, 胡波^{1*}, 李思照²

(1. 贵州师范大学大数据与计算机科学学院, 贵州 贵阳 550025;

2. 哈尔滨工程大学计算机科学与技术学院, 黑龙江 哈尔滨 150001)

摘要: 现有关键点检测算法在光照变化、人员密集交叉遮挡等情况下, 易导致检测精度降低、漏检或关键点检测错位的问题。针对该问题, 提出基于 YOLOv8n-Pose 改进的 LBW-YOLOv8n-Pose 复杂环境多人人体姿态估计算法。通过在特征提取主干网络的快速空间金字塔池化 (Spatial Pyramid Pooling-Fast, SPPF) 层引入大核可分离注意力 (Large Separable Kernel Attention, LSKA), 增强图像特征表达能力和感知能力。在颈部网络引入加权双向特征金字塔网络 (Bidirectional Feature Pyramid Network, BiFPN) 进行重构, 提高多尺度特征融合效果, 并采用改进的 Wise-IoU 损失函数, 提升模型收敛速度与复杂场景下的鲁棒性。实验结果表明, 改进后模型在 MS-COCO2017 人体关键点数据集上精确率、召回率、平均精度值分别达到 85.7%、76.8%、81.7%, 相比原模型均有明显提升, 且能更精准、有效地检测复杂情况下多人人体关键点信息。

关键词: 姿态估计; 注意力机制; YOLOv8n-Pose

中图分类号: TP391.41; TP183

文献标志码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1003-3114(2025)05-1025-11



Enhanced Estimation Algorithm Based on YOLOv8n-Pose Key Point Features

YIN Xiantao¹, HU Bo^{1*}, LI Sizhao²

(1. School of Big Data and Computer Science, Guizhou Normal University, Guiyang 550025, China;

2. College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China)

Abstract: Existing key point detection algorithms tend to suffer from reduced detection precision, missed detections, or misaligned key points in scenarios with varying lighting conditions and dense crowds with overlapping figures. To address this issue, an improved LBW-YOLOv8n-Pose algorithm for multi-person pose estimation in complex environments is proposed based on YOLOv8n-Pose. By introducing the Large Separable Kernel Attention (LSKA) in the Spatial Pyramid Pooling-Fast (SPPF) layer of the feature extraction backbone network, the algorithm enhances the image feature representation and perception capabilities. A weighted Bidirectional Feature Pyramid Network (BiFPN) is incorporated in the neck network for reconstruction to improve the multi-scale feature fusion effect. Additionally, an improved Wise-IoU loss function is adopted to accelerate the model's convergence speed and enhance its robustness in complex scenarios. Experimental results show that the improved model achieves precision, recall, and average detection precision of 85.7%, 76.8%, and 81.7% respectively on the MS-COCO2017 human key point dataset, representing significant improvements over the original model. Moreover, it can more accurately and effectively detect key point information of multiple people in complex situations.

Keywords: pose estimation; attention mechanism; YOLOv8n-Pose

收稿日期: 2025-05-13

基金项目: 贵州省基础研究(自然科学)(黔科合基础-ZK[2023]一般 254)

Foundation Item: Basic Research (Natural Sciences) of Guizhou Province (Guizhou Science and Technology Cooperation-Basic-ZK[2023] General 254)

0 引言

近年来,得益于深度学习的快速发展,各类高效的特征提取与特征融合网络不断被提出与创新,使姿态估计算法在准确性和效率方面均取得了显著进展^[1]。二维姿态估计算法可以分为单人姿态估计算法与多人姿态估计算法。Toshev等^[2]在2014年首次提出了单人姿态估计DeepPose算法,该算法是首个将深度神经网络应用于人体关键点检测的算法框架,但其检测效率低,计算复杂度高。多人姿态估计算法主要分为自顶向下与自底向上2种方法。

自底向上的方法主要通过检测图像中所有的关键点,将关节点进行分组判断,再构建人体骨架,主要有OpenPose^[3]、HigherHRNet^[4]等算法。该方法的优点是计算量不会随着图像中人物的增加而增大,效率高,缺点是在关节点密集分布的场景中,由于相邻节点间几何特征相似度高且空间耦合性强,导致准确区分和关联各关节点具有较高难度,易引发局部节点空间位置误判或拓扑结构错位现象。

自顶向下的方法主要先检测出图像中的每个人物边界框,再对每个人物对象进行单独的姿态估计,主要有Alphapose^[5]算法,该算法可以灵活地使用各种目标检测算法检测人体边界框,常见的检测算法如Faster R-CNN^[6]、SSD^[7]、YOLO^[8]系列算法,其优点在于检测精度高,每个关节点在边框内进行检测,不会出现分组错误问题,缺点在于该算法属于两阶段算法,严重依赖第一阶段的目标检测算法,当出现复杂多人物场景时,计算量随着检测人物的增加而增加且训练较为繁琐,需要单独的训练目标检测模型与姿态估计模型,无法进行端到端的训练与优化^[9]。闻举等^[10]采用AlphaPose算法结合YOLOX目标检测模型进行打太极关键点检测研究,该方法属于两阶段算法,虽然实现了模型轻量化,但参数量(Params)与计算量仍存在较大优化空间。Maji等^[11]首次提出YOLO-Pose算法,该算法集成目标检测与姿态估计算法,可实现端到端的训练与优化,且结合了自顶向下与自底向上方法的优点,该方法不需要后续自底向上分组操作,但该方法的Params、时间复杂度仍存在较大优化空间,存在复杂环境下关键点空间位置易错位、边缘设备应用部署受限等缺点。Ultr-

alytics团队发布的YOLOv8n-Pose模型,虽然在关键点检测方面进行过优化,但在复杂的人物目标关键点检测场景中仍存在漏检或关键点错位问题。提出基于YOLOv8n-Pose单阶段多人姿态检测算法,通过高效的注意力机制、充分的特征信息融合及优化损失函数,解决在面对复杂多人姿态估计时出现的计算量大、精度低、漏检、关节点错位等问题。

1 模型改进

1.1 改进的YOLOv8n-Pose模型

YOLOv8n-Pose基于YOLOv8n模型,该模型适用于目标检测、图像分类、实例分割等任务,主要研究基于该模型的姿态检测模型YOLOv8n-Pose,具有体积小、速度快等优点。该模型结构主要包括输入层、主干网络、颈部网络以及检测头网络。输入部分对图像进行缩放至640 pixel×640 pixel大小,并对数据进行预处理与数据增强操作。数据增强主要通过对图像进行缩放、色调调整、马赛克增强、随机变换等,增强模型的鲁棒性。主干网络主要对图像进行特征提取;颈部网络采用特征金字塔网络(Feature Pyramid Network, FPN)^[12]与路径聚合网络(Path Aggregation Network, PAN)^[13]结合的双向特征金字塔网络,实现双向数据流,有效增强了不同特征层的数据融合;检测头部网络利用不同大小的特征图来获取不同大小对象的类别位置信息^[14]。

本文以YOLOv8n-Pose为基础网络模型,主要进行了如下改进:①首先在主干网络SPPF空间池化金字塔层创新性引入大核可分离注意力机制LS-KA,通过扩大感受野并增强长距离空间依赖关系捕捉能力,更加精准聚焦关键姿态区域以抑制背景干扰,增强模型特征提取能力;②在特征融合方面,颈部网络使用双向特征金字塔网络BiFPN构建高效跨尺度特征融合架构,通过加权特征融合方式增强高底层特征信息的交互,提升复杂场景下姿态特征的表达能力;③在损失函数方面优化为Wise-IoU,通过动态调整惩罚因子实现样本优化,结合边界框的尺度与位置信息,使得损失函数不仅关注目标重叠区域,还能对尺度不一、中心偏移问题进行优化,提升关键点定位精度,从而有效提高模型关键点检测的准确率。改进的YOLOv8n-Pose网络结构如图1所示。

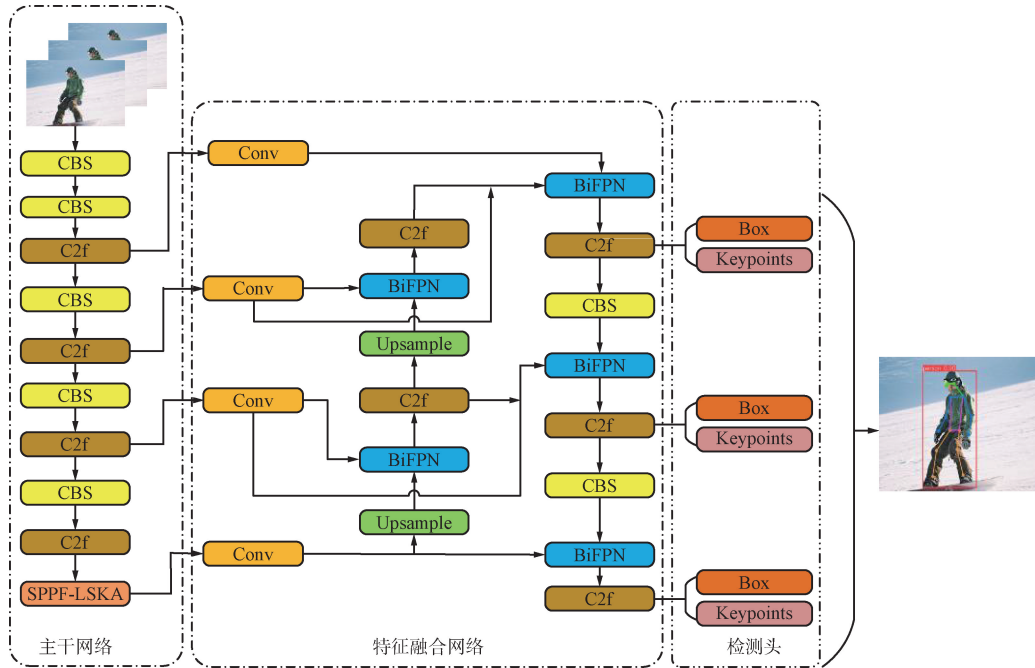


图1 YOLOv8n-Pose改进网络结构

Fig.1 Improved network structure of YOLOv8n-Pose

1.2 特征提取主干网络改进

YOLOv8n-Pose 的 SPPF 空间池化层主要作用减少特征图分辨率,对输入的特征图进行不同尺度的池化操作,并且进行多尺度特征图的融合,提升模型的感知能力与性能^[15]。但 YOLOv8n-Pose 在对多人密集环境时的关节点检测准确率降低,模型主要获取过多的局部信息,缺少全局信息。针对此问题,引入注意力机制方式,提高网络特征

提取能力,增强目标关键点检测能力,本文引入基于大核注意力(Large Kernel Attention, LKA)^[16],如图 2(a)所示;改进的 LSKA^[17],通过建立长距离依赖关系增强图像特征表达能力和感知能力,且在增强重要特征关注度的同时不会过度增加计算复杂度,有效提高多人密集遮挡场景的关键点检测准确率,如图 2(b)所示;改进的 SPPF-LSKA 结构如图 2(c)所示。

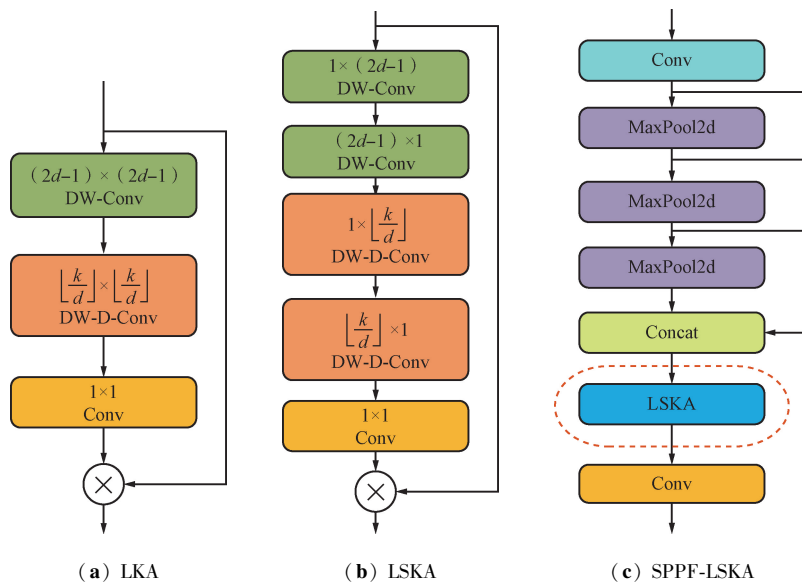


图2 改进的 SPPF 结构

Fig.2 Improved SPPF structure

与标准的 LKA 相比,LSKA 允许在注意力模块中直接使用具有大核的深度卷积层。LSKA 增强了输入图像的远程依赖性,且不会导致大量计算和内存占用。LKA 首先采用标准深度卷积来获取图像的局部上下文信息,专注于捕获图像的局部依赖性,更加精细地捕捉图像的细节信息,并缓解规则化网格采样导致的信息丢失问题。之后采用具有较大感受野的可扩展深度卷积来模拟远程依赖,有效捕获全局信息。然后将局部特征与全局特征融合后,输入到 1×1 的卷积中,以生成注意力图。最后将输出的注意力图与输入特征图作 Hadamard 乘积,使原始特征图中的每个元素均根据注意力图的值进行加权处理,从而突出重要的特征并抑制不重要的特征信息^[18]。

LKA 具体流程如图 2(a) 所示,LKA 将 $k \times k$ 大卷积核分解为 $(2d-1) \times (2d-1)$ 的标准深度卷积、 $(k/d) \times (k/d)$ 的深度可扩张卷积以及一个 1×1 卷积,分解原理^[17]如图 3 所示。

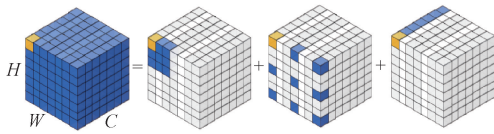


图 3 LKA 大核卷积分解原理

Fig. 3 Principle of LKA large kernel convolution decomposition

LSKA 结构如图 2(b) 所示,在 LKA 的基础上将深度卷积层的 2D 卷积核分解为级联的水平 1D 与垂直 1D 内核,可得到等效 LKA 结构,并且降低随 LKA 卷积核尺寸的增加而出现的参数与计算复杂度的二次增长,同时缓解了因大感受野结构带来的局部细节丢失问题。具体而言,LSKA 将 LKA 的 $(2d-1) \times (2d-1)$ 标准深度卷积分解为水平的 $1 \times (2d-1)$ 与垂直的 $(2d-1) \times 1$,再将 $(k/d) \times (k/d)$ 深度可扩张卷积分解为水平 $1 \times (k/d)$ 以及垂直的 $(k/d) \times 1$,最终将分解的卷积进行串联,分解结构如图 4 所示^[19-20]。

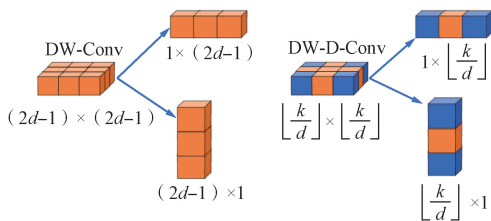


图 4 LSKA 卷积分离计算原理

Fig. 4 LSKA convolution separation calculation principle

LSKA 的计算如下:

$$\bar{Z}^C = \sum_{H,W} W_{(2d-1) \times 1}^C * \left(\sum_{H,W} W_{1 \times (2d-1)}^C * F^C \right), \quad (1)$$

$$Z^C = \sum_{H,W} W_{\lfloor \frac{k}{d} \rfloor \times 1}^C * \left(\sum_{H,W} W_{1 \times \lfloor \frac{k}{d} \rfloor}^C * \bar{Z}^C \right), \quad (2)$$

$$A^C = W_{1 \times 1} * Z^C, \quad (3)$$

$$\bar{F}^C = A^C \otimes F^C, \quad (4)$$

式中: d 为膨胀率, \bar{Z}^C 表示深度卷积核大小为 $(2d-1) \times (2d-1)$ 的输出, Z^C 表示深度可扩张卷积核大小为 $(k/d) \times (k/d)$ 的输出, A^C 表示使用 1×1 卷积得到的注意力特征图, \bar{F}^C 表示用得到的注意力特征图与原始输入的特征图进行 Hadamard 乘积。

在原始的 SPPF 层中,通过对输入特征执行不同程度的空间池化操作来捕获多尺度信息,但在高度复杂背景中捕获空间信息的能力仍显不足,限制了模型整体识别的准确性^[21]。LSKA 注意力机制使模型聚焦于输入特征的重要部分,本文在 SPPF 层的 Concat 之后引入 LSKA 得到 SPPF-LSKA,以便在 SPPF 完成池化与拼接操作后能够利用 LSKA 注意力^[22]。该融合既保留了 SPPF 模块对不同大小目标的检测优势,又通过 LSKA 模块的注意力加权,进一步增强了模型捕捉关键特征的能力。

1.3 BiFPN 特征融合网络改进

YOLOv8n-Pose 的颈部网络在基于 FPN 的基础上,引入了一条额外的自底向上的特征融合路径,创新组合形成 PANet 网络。FPN 引入了自顶向下的路径,从而将网络的深层特征信息融合至浅层特征层中,而自底向上的特征融合将浅层特征信息向上融合至更深层特征层中,二者巧妙结合,有效解决了特征信息丢失问题,增强了模型识别细微特征的能力^[23]。但对于图像中尺寸较小的人物,其关键点信息检测效果仍存在局限性。

针对该问题,通过引入 BiFPN^[24] 重构颈部层的特征融合网络,网络结构如图 5 所示。BiFPN 网络不仅包含了自顶向下与自底向上的 2 条特征信息流动路径,而且还进行双向跨尺度链接,增强各网络层的特征信息融合,使特征图拥有更多细节信息与位置信息。此外,BiFPN 为了应对不同分辨率下特征图对输出特征的贡献不平等,为每个输入特征增加了一个额外的可学习权重,使网络能够学习每个输入特征的重要性^[25],且采用快速归一化方法,

使权重值处于 0~1,提升模型在 GPU 上的训练速度,计算如下:

$$O = \sum_i \frac{W_i}{\varepsilon + \sum_j W_j} * I_i, \quad (5)$$

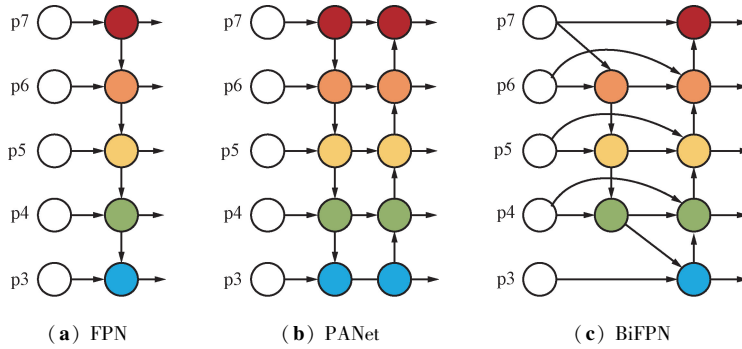


图 5 BiFPN 网络结构

Fig. 5 BiFPN network structure

1.4 优化损失函数 Wise-IoU

YOLOv8n-Pose 模型损失函数主要包括:① 边界框回归损失;② 姿态估计损失;③ 分类损失;④ 关键点损失。该模型的目标检测任务与关键点检测任务融合在一起,且关键点检测需在目标检测定位的区域内开展,所以关键点检测任务的准确性十分依赖目标检测任务的准确性。YOLOv8n-Pose 的边界框回归损失采用 CIoU^[26] 损失函数,CIoU 损失函数相较于传统的 IoU 损失函数,考虑了重叠面积、中心点距离及长宽比,并且该损失函数引入修正因子,增强损失函数对不同形状目标框的鲁棒性。CIoU 损失函数如下:

$$L_{CIoU} = 1 - IoU + \frac{p^2(b, b^{gt})}{c^2} + av, \quad (6)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2, \quad (7)$$

$$\alpha = \frac{v}{(1 - IoU) + v}, \quad (8)$$

式中: w 、 h 、 b 分别为预测边界框的宽度、高度、中心点位置, w^{gt} 、 h^{gt} 、 b^{gt} 分别为真实标注框的宽度、高度、中心点位置, c 为真实框与预测框最小外接矩形的对角线长度, v 用于衡量长宽比的相似性, α 为权重参数^[27]。虽然 CIoU 损失函数相较于 IoU 在边界框回归中有较好的效果,但参数 v 仅评价了长宽比的相似性,并不能够准确反映真实标注框与预测框宽度与高度的真实差异,从而导致高质量样本与低质量样本在边界框回归损失不平衡^[28]。同时,CIoU

式中: O 表示融合后的特征输出, W_i 与 W_j 均表示权重值, I_i 表示输入特征, $\varepsilon=0.0001$ 表示一个较小的常数,用于避免数值不稳定。BiFPN 最终融合方式为双向跨尺度连接与快速归一化融合,能够较好地识别小目标及关键点信息。

需计算预测框与真实框之间的中心距离、宽度、高度和面积的差异,计算复杂且不易优化^[29]。

为了应对上述 CIoU 产生的问题,本文使用 Wise-IoUv3^[30] 损失函数换 CIoU 损失函数,采用注意力函数结合动态非单调聚焦机制(Focusing Mechanism, FM)的策略,该机制使用离群值判断锚框的质量,并根据离群值分配梯度增益。Wise-IoUv3 为离群值较大的锚框分配较小的梯度增益,有效防止了低质量样本造成的较大梯度损失,平衡了高质量与低质量样本差异对模型的影响^[31],加快了模型收敛速度,提高了人体关键点提取性能。Wise-IoUv3 计算如下:

$$L_{IoU} = 1 - \frac{W_i H_i}{wh + w^{gt} h^{gt} - W_i H_i}, \quad (9)$$

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right), \quad (10)$$

$$L_{Wise-IoUv3} = r \cdot R_{WIoU} \cdot L_{IoU}, \quad r = \frac{\beta}{\delta \alpha^{\beta - \delta}}, \quad (11)$$

式中: r 表示非单调动态聚焦系数, $\beta = \frac{L_{IoU}^*}{L_{IoU}}$ 表示离群度, R_{WIoU} 表示注意力函数, W_g 与 H_g 分别表示真实框和预测框最小外接矩形的长与宽, W_i 与 H_i 分别表示真实框与预测框相交时的长与宽, L_{IoU}^* 表示 IoU 损失梯度增益, L_{IoU} 表示训练过程中运行的动态平均值。

2 实验及结果分析

2.1 实验数据集与环境

本文使用微软发布的 MS-COCO2017 公共数据

集,该数据集包含人体 17 个关键点的标注,其中用于人体姿态估计的训练图片 56 599 张,验证集图片 2 346 张。实验环境如表 1 所示。

表 1 实验环境
Tab. 1 Experimental environment

名称	配置
操作系统	Windows 10
内存/GB	767
CPU	Intel (R) Xeon(R) Gold 6342
GPU	NVIDIA RTX A6000 (47 GB)
开发环境	PyTorch 1. 10. 1, CUDA 11. 3

训练参数设置为:训练轮数(epoch)为 300,输入图像大小为 640 pixel×640 pixel,优化器为 SGD,初始学习率为 0. 01,周期学习率为 0. 01,线程数为 16,批处理大小为 64。

2. 2 模型评价指标

本文使用精确率(P)、召回率(R)、平均精度均值(mean Average Precision, mAP)作为模型评价标准。但在姿态估计的评价标准中,并不使用检测框的 IoU 判断真正例(TP)与伪正例(FP),而是通过计算预测关键点相对于真实关键点位置是否处于合理误差范围内。所使用对象关键点相似度 L_{oks} 计算为:

$$L_{oks} = \frac{\sum_i \exp\left(-\frac{d_i^2}{2s^2k_i^2}\right) \cdot \delta(v_i > 0)}{\sum_i \delta(v_i > 0)}, \quad (12)$$

式中: i 为关键点的序号, d_i^2 为实际与预测关键点位置的欧氏距离, v_i 为关键点是否可见, s^2 为检测体所占面积, k_i 为第 i 个关键点的归一化因子, δ 为脉冲函数,表示仅计算标注的可见关键点。精确率、召回率、平均精度均值计算如下:

$$P_{kpt} = \frac{TP_{kpt}}{TP_{kpt} + FP_{kpt}}, \quad (13)$$

$$R_{kpt} = \frac{TP_{kpt}}{TP_{kpt} + FN_{kpt}}, \quad (14)$$

$$AP = \int_0^1 P_{kpt} d(R_{kpt}), \quad (15)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N}, \quad (16)$$

式中: TP_{kpt} 表示正确识别,且关键点 L_{oks} 大于设置的阈值; FP_{kpt} 表示非关键点区域被错误识别为关键点,且 L_{oks} 大于设置的阈值; FN_{kpt} (伪反例)表示真实关键点区域被错误识别为非关键点,且 L_{oks} 小于设置的阈值。

2. 3 消融实验

为了验证改进的 YOLOv8n-Pose 网络中各改进部分的有效性,以 YOLOv8n-Pose 为基准模型,依次将各改进网络部分加入基准模型中,在同样的数据集下,保持各项参数一致,通过消融实验评估不同改进网络组件对最终模型的影响。消融实验结果如表 2 所示。

表 2 消融实验结果对比

Tab. 2 Comparison of ablation experiment results

实验	SPPF-LSKA	BiFPN	Wise-IoU	$P/\%$	$R/\%$	$mAP@0. 5/\%$	$mAP@0. 5; 0. 95/\%$	Params/M
1	×	×	×	83. 8	74. 1	79. 5	49. 2	3. 3
2	√	×	×	85. 2	74. 4	80. 1	49. 9	3. 5
3	√	√	×	84. 6	75. 3	81. 3	50. 9	4. 0
4	√	√	√	85. 7	76. 8	81. 7	51. 1	4. 0

由表 2 中实验 1 与实验 2 对比可知,加入改进的 SPPF-LSKA 网络后,平均精度均值提升 1. 4 个百分点,引入 LSKA 可有效提高精确率,且在实际效果中,对于紧密复杂的人体骨骼点检测效果优于原始模型。实验 2 与实验 3 结果表明,引入 BiFPN 特征融合网络后,模型召回率提升 0. 9 个百分点,较基准模型提升 1. 2 个百分点, $mAP@0. 5$ 与 $mAP@0. 5; 0. 95$ 较基准模型分别提升 1. 8 个百分

点、1. 7 个百分点。

实验 4 结果显示,引入 Wise-IoU 损失函数后,精确率、召回率及平均精度均值均有提升。实验数据如图 6 所示,最终改进模型较基准模型在精确率、召回率、 $mAP@0. 5$ 、 $mAP@0. 5; 0. 95$ 上分别提升 1. 9 个百分点、2. 7 个百分点、2. 2 个百分点、1. 9 个百分点,Params 仅增加 0. 7 M,最终模型的 Params 为 4 M。

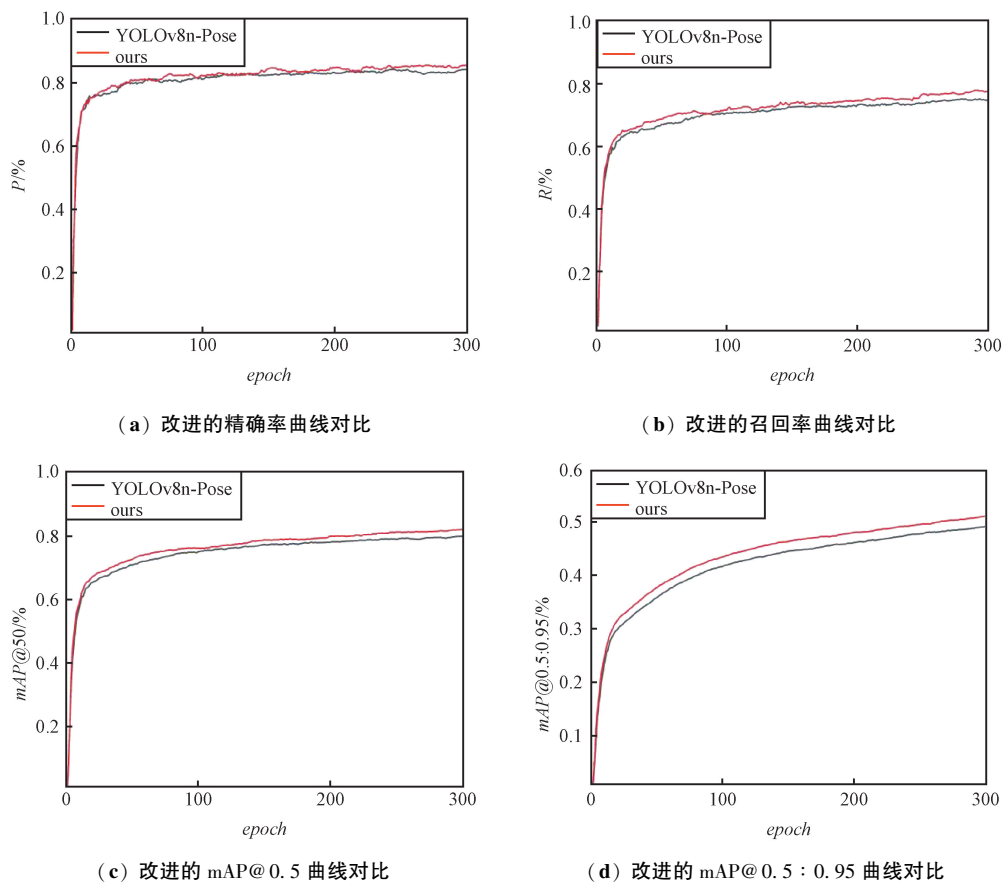


图6 模型对比实验数据

Fig. 6 Comparison of the experimental data of the model

2.4 主干网络对比实验

为了突显改进后的特征提取骨干网络的优势,在 MS-COCO2017 数据集上对常见的骨干网络进行性能对比实验。

以 YOLOv8n-Pose 作为基准模型,使用改进后的主干网络与 StarNet^[32]、FastNet^[33]、SPPF-FocalModulation^[34]、SPPF-AFI^[35] 主干网络进行对比实验,结果如表 3 所示。

表 3 特征提取骨干网络对比实验

Tab. 3 Comparative experiment of feature extraction backbone network

模型	Params/M	P/%	R/%	mAP@0.5/%	* mAP@0.5:0.95/%	GFLOPs
YOLOv8n-Pose	3.3	83.9	74.0	79.3	49.3	9.3
YOLOv8n-StarNet-Pose	2.5	82.9	72.9	78.1	47.0	7.6
YOLOv8n-FastNet-Pose	4.4	84.0	76.0	80.7	50.1	11.8
YOLOv8n-FocalModulation-Pose	3.4	84.2	74.5	79.9	50.0	9.4
YOLOv8n-AFI-Pose	3.2	84.8	74.2	80.2	49.6	9.2
YOLOv8n-Pose+SPPF_LSKA	3.5	85.2	74.4	80.4	49.9	9.5

由表 3 可以看出,改进后的特征提取骨干网络,在 Params 上仅比原始模型高 0.2 M,计算量高 0.2 GFLOPs。但在验证集各项指标均高于原始模型,其中精确率高 1.3 个百分点,召回率高 0.4 个百分点,mAP@0.5 高 1.1 个百分点。FastNet 骨干网

络的召回率高于其他骨干网络,但 Params 与计算量均高于其他模型。对比结果表明,改进后的骨干网络各项参数表现较优。

图 7 为 GradCAM 可视化热力图,直观展示了原始模型与融合 LSKA 注意力机制后的模型检测效果。



图 7 可视化热力图对比效果

Fig. 7 Comparison effect diagram of visualization heat map

由图 7(b)与图 7(c)的热力图对比结果可明显看出,无论是室内小目标、关键点遮挡区域,还是人物密集情况,融合 LSKA 注意力后,能更好地提取特征信息,特别是对小目标的感知能力,提高了模型检测精度。

2.5 特征融合网络对比试验

为了验证所改进的特征融合网络的先进性和有效性,将 BiFPN 特征融合网络与其他主流的特征融

合网络进行对比实验,实验结果如表 4 所示。可以看出,BiFPN 在召回率指标上明显高于其他 4 种结构,且在 $mAP@0.5$ 指标上也高于其他 4 种网络融合结构,相比 SlimNeck 特征融合网络提升 3.0 个百分点,表明 BiFPN 在特征提取与融合能力上更强。BiFPN 在 Params 与权重的对比中,虽高于其他对比模型,但 Params 与权重都在可接受范围内,综合对比,改进的特征融合网络综合效果较优。

表 4 特征融合网络对比实验

Tab. 4 Comparison experiment of feature fusion networks

模型	$P/\%$	$R/\%$	$mAP@0.5/\%$	$mAP@0.5:0.95/\%$	Params/M	权重/M
SlimNeck ^[36]	82.7	73.7	78.1	46.4	3.1	6.5
RepGFPN ^[37]	83.8	74.0	79.6	49.0	3.5	7.5
EfficientRepBiPAN ^[38]	82.1	73.6	78.2	47.2	3.0	6.4
ASF ^[39]	82.8	75.0	79.6	49.5	3.3	7.0
BiFPN	83.6	76.1	81.1	50.9	3.7	7.8

2.6 主流模型对比实验

将改进算法与主流算法在相同的验证集进行对比实验,实验数据如表 5 所示。平均精度均值最高的模型为 YOLOv7-W6-Pose,高达 94%,但 Params 远超其他模型,需要更高的算力与存储要求,不适合在资源有限的设备部署;改进模型相比于 YOLOv5s6-Pose 模型,其 Params 约为改进模型的 4 倍,但平均精度均值仅提高了 2.6 个百分点;改进模型相比于 YOLOx-Pose-tiny 模型,不仅 Params 低 2 M,且平均精度均值高约 2.4 个百分点。

表 5 主流模型对比实验

Tab. 5 Comparative experiments of mainstream models

模型	图像大小/ pixl	Params/M	$mAP@0.5/\%$
YOLOv5s6-Pose	640×640	15.0	84.3
YOLOx-Pose-tiny	416×416	6.0	79.3
YOLOv7-W6-Pose	960×960	80.0	94.0
YOLOv8n-Pose	640×640	3.2	79.5
本文模型	640×640	4.0	81.7

2.7 模型泛化验证

为了验证改进后模型的适用性与泛用性,在MPII公开数据集上进行额外实验,该数据集包含更复杂、多变、非常规的姿态动作,标注有16个关键点。训练集含10 306个样本,验证集含4 420个样本,实验训练环境保持一致,实验数据如表6所示。可以看出,改进后的模型在MPII数据集上的精确率、召回率和平均精度均值指标取得较好的结果。其中精确率与基准模型相差无几,但在 R 、 $mAP@0.5$ 和 $mAP@0.5:0.95$ 分别提高了1.6、0.6和0.9个百分点。通过一系列实验论证,改进后的模型在不同的数据集依然取得显著的实验效果,充分展现出较强的泛用性与适用性。

表6 模型泛化实验

Tab. 6 Model generalization experiment

单位:%

模型	P	R	$mAP@0.5$	$mAP@0.5:0.95$
YOLOv8n-Pose	85.3	80.0	83.8	46.4
YOLOv11n-Pose	85.0	80.0	83.8	45.9
YOLOv12n-Pose	84.5	79.0	82.3	43.9
本文模型	85.2	81.6	84.4	47.3

2.8 检测效果分析

在MS-COCO2017数据集上,通过对改进前后的模型进行不同环境的检测效果对比,主要包括人物关节节点不清晰、人物密集、人物遮挡等环境。对比检测效果如图8所示,左为原始模型,右为改进模型。



图8 模型对比效果

Fig. 8 Comparison effect of models

图8(a)中,在人物目标排列相对密集的关键点检测场景,改进模型相比原模型能减少目标漏检;图8(b)检测效果显示,改进模型在关节点被遮挡时也能较好检测出目标关键点;图8(c)中,在小目标及

关键点被遮挡的场景,改进模型相比原模型能检测出更多目标及被遮挡的关键点;图8(d)检测效果显示,针对极端姿态动作,原始模型检测出现关键点错位及漏检,改进模型检测效果更优;图8(e)中,针对光线

不充足的场景,原始模型出现严重的人物目标漏检,改进模型能检测更多目标,效果更优。图 8(f)中,以俯视角度检测目标关键点时,改进模型与原始模型均出现漏检、关键点错位甚至识别错误;图 8(g)中,针对人物密集无序且光照不足的情况,改进模型与原始模型均出现严重漏检;图 8(h)中,针对大密度人物目标场景,模型识别率均较低,出现严重漏检。

根据上述分析,本文模型在关键点遮挡、光线较暗且目标稀疏、极端姿态动作、人物密集有序的场景相比原模型有更好的检测效果,但在俯视角度检测、光线不足且人物密度大及更大密度人物目标关键点检测方面仍需改进。

3 结束语

针对现有关键点检测模型,在光照变化、人员密集交叉遮挡时存在检测精度降低、节点错位等问题,提出基于 YOLOv8n-Pose 模型改进的 LBW-YOLOv8n-Pose。在 SPPF 中引入 LSKA,能捕捉图像更广泛的上下文特征信息,增强对遮挡关键点的特征提取能力;采用 BiFPN 重构颈部网络,增强网络跨尺度特征融合能力,充分融合深浅层特征信息,增强模型对多尺度目标的检测与关键点提取能力;在此基础上,创新性应用 Wise-IoU 边界框损失函数,通过动态调整梯度权重,减少低质量样本产生的有害梯度,有效提升在复杂环境下的泛化能力与检测鲁棒性。实验结果表明,改进模型在 MS-COCO2017 关键点数据集上各项指标均有提升,且模型大小仅 4 M。尤其在关节点不清晰、人物密集、光线不足、遮挡环境下,改进模型检测效果较好,但在更大密度目标检测方面仍有较大改进空间。未来可结合时空图卷积开展人物动作识别研究,并应用于其他关键点检测领域。

参考文献

- [1] 王珂,陈启腾,陈伟,等. 基于深度学习的二维人体姿态估计综述[J]. 郑州大学学报(理学版), 2024, 56(4): 11-20.
- [2] TOSHEV A, SZEGEDY C. DeepPose: Human Pose Estimation via Deep Neural Networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 1653-1660.
- [3] CAO Z, SIMON T, WEI S E, et al. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 7291-7299.
- [4] CHENG B W, XIAO B D, WANG J, 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. Seattle: IEEE, 2020: 5386-5395.
- [5] FANG H S, LI J F, TANG H Y, 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.
- [6] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot Multibox Detector [C] // European Conference on Computer Vision. Amsterdam: Springer International Publishing, 2016: 21-37.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-time Object Detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779-788.
- [9] 傅裕,高树辉. 改进 YOLOv8s-Pose 多人姿态估计轻量化模型研究[J]. 计算机科学与探索, 2025, 19(3): 682-692.
- [10] 闻举,伊力哈木·亚尔买买提. 针对太极动作的轻量级人体姿态估计[J]. 计算机仿真, 2025, 42(3): 270-277.
- [11] MAJI D, NAGORI S, MATHEW M, et al. YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 2637-2646.
- [12] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2117-2125.
- [13] LIU S, QI L, QIN H F, et al. Path Aggregation Network for Instance Segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8759-8768.
- [14] YUAN Z J, SHAO P W, LI J R, et al. YOLOv8-ACU: Improved YOLOv8-Pose for Facial Acupoint Detection [J]. Frontiers in Neuroinformatics, 2024, 18: 1355857.
- [15] 高民,陈高华,古佳欣,等. FLM-YOLOv8: 一种轻量级的口罩佩戴检测算法[J]. 计算机工程与应用, 2024, 60(17): 203-215.
- [16] GUO M H, LU C Z, LIU Z N, et al. Visual Attention Network [J]. Computational Visual Media, 2023, 9(4): 733-752.

- [17] LAU K W, PO L M, REHMAN Y A U. Large Separable Kernel Attention; Rethinking the Large Kernel Attention Design in CNN [J]. Expert Systems with Applications, 2024, 236: 121352.
- [18] 王伊萌. 基于感受野的多层级特征融合交通标志检测算法研究[D]. 西安: 长安大学, 2024.
- [19] 史涛, 崔杰, 李松. 优化改进 YOLOv8 实现实时无人机车辆检测的算法[J]. 计算机工程与应用, 2024, 60(9): 79-89.
- [20] WANG Y, OUYANG C, PENG H, et al. YOLO-ALW: An Enhanced High-precision Model for Chili Maturity Detection[J]. Sensors, 2025, 25(5): 1405.
- [21] ZHOU Q, WANG Z, ZHONG Y W, et al. Efficient Optimized YOLOv8 Model with Extended Vision[J]. Sensors, 2024, 24(20): 6506.
- [22] WANG C J, WANG Y F. SLGA-YOLO: A Lightweight Castings Surface Defect Detection Method Based on Fusion-enhanced Attention Mechanism and Self-architecture [J]. Sensors, 2024, 24(13): 4088.
- [23] LI N, YE T R, ZHOU Z H, et al. Enhanced YOLOv8 with BiFPN-SimAM for Precise Defect Detection in Miniature Capacitors[J]. Applied Sciences, 2024, 14(1): 429.
- [24] TAN M, PANG R, LE Q V. Efficientdet: Scalable and Efficient Object Detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 10781-10790.
- [25] 王新蕾, 廖晨旭, 王硕, 等. 面向鱼眼摄像头目标检测的轻量级网络[J]. 激光与光电子学进展, 2025, 62(2): 441-455.
- [26] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression [C] // Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020, 34(7): 12993-13000.
- [27] 罗亮, 郎霄, 祖国庆, 等. 一种基于改进 YOLOv8n 的气缸套缺陷检测方法[J]. 中国机械工程, 2025, 36(5): 1054-1064.
- [28] YAO J S, SONG B, CHEN X Y, et al. Pine-YOLO: A Method for Detecting Pine Wilt Disease in Unmanned Aerial Vehicle Remote Sensing Images[J]. Forests, 2024, 15(5): 737.
- [29] 张朝阳, 张上, 胡益民, 等. 动态聚焦多维注意力遥感弱小目标检测[J]. 无线电通信技术, 2025, 51(1): 196-209.
- [30] TONG Z J, CHEN Y H, XU Z W, et al. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism [EB/OL]. (2023-01-24) [2025-05-12]. <https://arxiv.org/abs/2301.10051>.
- [31] RAN X S, HE S, LI R. Research on Fatigued-driving Detection Method by Integrating Lightweight YOLOv5s and Facial 3D Keypoints[J]. Sensors, 2023, 23(19): 8267.
- [32] MA X Y, DAI X L, BAI Y F, et al. Rewrite the Stars [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 5694-5703.
- [33] CHEN J R, KAO S H, HE H, et al. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 12021-12031.
- [34] YANG J W, LI C Y, DAI X Y, et al. Focal Modulation Networks [J]. Advances in Neural Information Processing Systems, 2022, 35: 4203-4217.
- [35] LI L H, ZHAO S W, SUN R Z, et al. AFI-Net: Attention-guided Feature Integration Network for RGBD Saliency Detection [J]. Computational Intelligence and Neuroscience, 2021(1): 8861446.
- [36] LI H L, LI J, WEI H B, et al. Slim-neck by GSConv: A Lightweight-design for Real-time Detector Architectures [J]. Journal of Real-time Image Processing, 2024, 21(3): 62.
- [37] XU X Z, JIANG Y Q, CHEN W H, et al. DAMO-YOLO: A Report on Real-time Object Detection Design [EB/OL]. (2022-11-23) [2025-05-12]. <https://arxiv.org/abs/2211.15444>.
- [38] LI C Y, LI L L, GENG Y F, et al. YOLOv6 v3.0: A Full-scale Reloading [EB/OL]. (2023-01-13) [2025-04-15]. <https://arxiv.org/abs/2301.05586>.
- [39] KANG M, TING C M, TING F F, et al. ASF-YOLO: A Novel YOLO Model with Attentional Scale Sequence Fusion for Cell Instance Segmentation [J]. Image and Vision Computing, 2024, 147: 105057.

作者简介:

殷贤涛 男, (1998—), 硕士研究生。主要研究方向: 嵌入式系统、AI。

(*通信作者) 胡波 男, (1980—), 硕士, 副教授。主要研究方向: 嵌入式系统、AI、模糊系统。

李思照 男, (1982—), 博士, 副教授。主要研究方向: AI 芯片、新型存储架构。