光电技术及应用

DOI：10.16818/j.issn1001－5868.2024073103

层内调节特征金字塔的密集人群姿态估计算法

谷学静[1,2]，郭志斌[1,2*]

（1. 华北理工大学 电气工程学院，河北 唐山 063210；2. 唐山市数字媒体工程技术研究中心，河北 唐山 063210）

摘　要： 针对现有姿态估计算法对密集人群检测存在漏检和误检等问题，提出一种改进 YOLOv8sPose 的密集人群姿态估计算法 YOLOv8Pose-DC。首先，设计一种集中式层内调节特征金字塔网络，采用并联方式把可变形注意力机制和 CASPPF 结合起来，通过自上而下的方式对金字塔网络进行全局集中调节，增加网络中全局表示的空间权重，使得改进算法能够获得全面且具有区分性的特征表示；其次，提出多尺度双检测头结构，减少计算量的同时提高模型检测效率；然后，使用 DySample 模块，提高模型上采样效率；最后，加入上下文感知模块，提高模型全局信息关联能力，并抑制无用背景突出人物特征。实验结果表明，相较于基准模型，YOLOv8Pose-DC 的 $mAP@0.5$ 提升 3.1%，召回率提升 4.2%。所设计的算法性能有较大提升，满足生产需要。

关键词： 姿态估计；YOLOv8sPose；密集人群；集中式层内调节特征金字塔；DySample

中图分类号： TP391.4　**文章编号：** 1001－5868(2024)06－0931－08

Dense Crowd Pose Estimation Algorithm for In-layer Adjustment Feature Pyramid

GU Xuejing[1,2], GUO Zhibin[1,2]

(1. College of Electrical Engineering, NorthChina University of Science and Technology, Tangshan 063210, CHN;
2. Tangshan Digital Media Engineering Technology Research Center, Tangshan 063000, CHN)

Abstract: To address the issues of missed and false detections in existing dense crowd pose estimation algorithms, an improved YOLOv8sPose algorithm for dense crowd pose estimation, namely, YOLOv8Pose-Dense Crowd (YOLOv8Pose-DC), is proposed. First, a centralized intrinsic adjustment feature pyramid network is designed, which combines deformable attention mechanisms and coordinate attention-based spatial pyramid pooling fast (CASPPF) in a parallel manner, globally focusing and adjusting the pyramid network from top to bottom, thereby increasing the spatial weight of global representation within the network. This enables the improved algorithm to obtain comprehensive and distinctive feature representations. Second, a multi-scale dual detection head structure is proposed, reducing computational complexity while enhancing model detection efficiency. Furthermore, the DySample module is utilized to improve the upsampling efficiency of the model. Finally, a spatial context aware module (SCAM) is added to enhance the model's ability in associating global information and suppressing irrelevant background features, to highlight human characteristics. Compared to the baseline model, YOLOv8Pose-DC increases mAP@0.5 by 3.1% and recall rate by 4.2%. The designed algorithm significantly improves performance and fully meets production requirements.

Key words: pose estimation; YOLOv8sPose; dense crowds; centralized intra-layer

收稿日期：2024－07－31.

基金项目：河北省自然科学基金高端钢铁冶金联合研究基金专项项目(F2017209120)；唐山市沉浸式虚拟环境基础创新团队项目(18130221A).

* **通信作者：**郭志斌

adjustment feature pyramid; DySample

0 引言

近年来,随着机器视觉、深度学习以及人工智能算法的不断提出和完善,特别是深度学习技术的兴起与进步,使得姿态估计技术得到了显著突破。目前常见的姿态估计技术有二维姿态估计技术[1-2]和三维姿态估计技术[3-4],其中三维姿态估计技术难度大、对设备的要求高且计算复杂度高实时性差,而二维姿态估计技术成本、计算复杂度低、实时性好,且近几年得到了快速的发展。在 YOLOPose 之前,基于深度学习的二维姿态估计主要有自上而下[5-6]和自下而上[7-8]这两种方法。自上而下的方法是先检测到人体,再进行人体关键点标注,这种方法具有较高的精度,但计算复杂度高、对模型依赖强,且当场景发生变化或目标不符合预设模型时,性能可能降低。自下而上的方法是先检测到所有人的关键点,再把关键点聚类到每个人,这种方法具有适应性强和通用性强等特点,但其精度低、计算开销大且缺乏可解释性。YOLOPose[9]是 Maji 等提出的一种基于 YOLO[10]框架的无热图联合检测方法,这种方法不需要自下而上的后处理来将检测到的关键点分组到骨架中,也与自上而下的方法不同,减去了多次前向传递。此模型在一次前向传递中会把检测多人的边界框和与其相对应的二维姿势联合起来,从而带来自上而下和自下而上方法结合的最佳效果。YOLOv8Pose 是继 YOLOPose 之后,采用无锚框检测、新的卷积核操作、新的 C2f 模块设计等方式优化了原有模型结构。此模型精度领先同类网络,在保持实时性的基础上能实现更高的识别精度。

Lin 等[11]提出了特征金字塔网络(PFN)。Quan 等[12]对原有特征金字塔进行了优化,提出集中式特征金字塔(CFP)。Tan 等[13]提出了一种加权双向特征金字塔(BiFPN)。Jing 等[14]对 BiFPN 网络进行了改进,提出双向自适应特征金字塔网络。Toshev 等[15]提出了 DeepPose 算法,这是首次利用深度神经网络(DNN)进行特征提取和关键点坐标回归的人体姿态估计方法。Ma 等[16]对 FPN 网络进行了改进,提出了跨层级联方法。Wang 等[17]对 YOLOv8 的 Backbone 进行改进,提出了安全装备检测算法 Wear-YOLO。Zhao 等[18]通过改进 Backbone 和颈部网络中的 C2f 模块,提出用于航拍小目标检测方法 CRP-YOLO。Li 等[19]结合 CNN 和 Transformer 的优势提出 BOT 模块,并针对检测道路损伤提出轻量型道路损伤检测算法 YOLOv8-RD。但在面对检测密集人群时,现有模型会出现误检漏检等问题。

为解决上述问题,以 YOLOv8sPose 为基准,设计出针对检测密集人群的算法 YOLOv8Pose-DC。主要的工作如下:

(1)引入可变形注意力 DA(Deformable Attention);

(2)设计了坐标注意力机制 CA(Coordinate Attention)与空间特征金字塔 SPPF(Spatial Pyramid Pooling Fast)相结合的新模块 CASPPF;

(3)设计了集中式层内调节特征金字塔,先把 DA 和 CASPPF 模块并联起来,再通过自上而下的方式和浅层信息进行特征融合,从而增加网络全局信息权重;

(4)提出多尺度双检测头结构;

(5)DySample 代替上采样;

(6)引入上下文感知模块 SCAM(Spatial Context Aware Module)。

1 基于 YOLOv8Pose 姿态估计算法

YOLOv8Pose 的结构图如图 1 所示。YOLO 是一种流行的目标检测算法,由 Redmon 等于 2015 年首次提出[10]。YOLOv8 是 YOLO 系列的一种算法,适用于目标检测、图像分类、实力分割、关键点检测等领域,其中本文专注于关键点检测。YOLOv8Pose 的网络结构分为四个部分,分别是输入(Input)、骨干网络(Backbone)、颈部(Neck)、检测头(Head)。

YOLOv8Pose 的 Input 部分是指模型接收的原始图像数据。该部分主要负责将输入图像放缩到训练所需要的尺寸,并进行数据预处理和数据增强操作。预处理包括图像归一化和缩放,确保输入尺寸和像素范围的一致性。数据增强采用缩放、色调调整、马赛克增强等技术。Input 部分还采用了无锚点机制来直接预测对象中心点,降低了模型复杂性和对预定义锚点大小和形状的依赖。Backbone 部分是 YOLOv8Pose 中的特征提取器,负责从输入图像中提取特征,其中包含 Conv,C2f 和 SPPF 模块。

新的 C2f 结构应用残差特征进行学习，丰富梯度流信息；SPPF 可将任意大小的特征图转换为固定大小的特征向量。Neck 是连接 Backbone 和 Head 的部分，主要作用是对骨干网络提取的特征进行进一步的处理和融合。其结构遵循特征金字塔网络(FPN)和路径聚合网络(PAN)，有效整合了自上而下和自下而上的信息流并提高检测性能。Head 是 YOLOv8Pose 中负责生成最终目标检测结果的部分。该模块利用不同大小的特征图来获取不同大小对象的类别和位置信息，它应用了分布式焦点损失(DFL)的概念，减少了参数大小和计算复杂度。YOLOv8Pose 系列在精度和速度方面表现良好，所以本实验采用 YOLOv8sPose 为基准，并在此基础上进行改进，进一步提升检测性能。

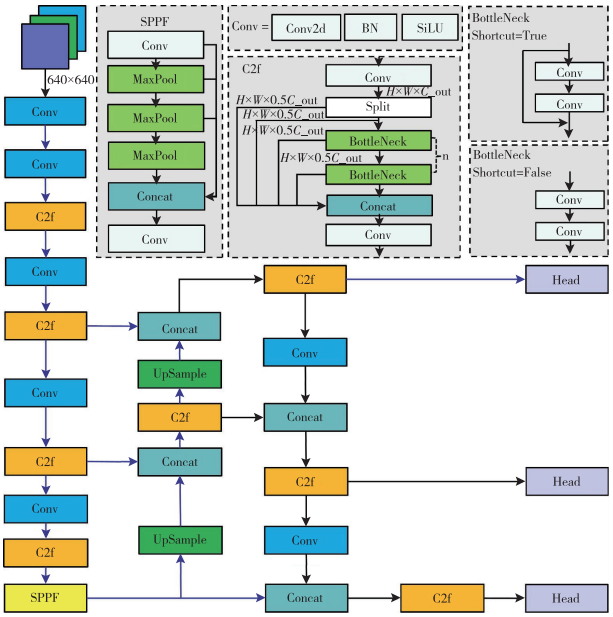


图 1　YOLOv8Pose 结构图

2　YOLOv8Pose-DC 的设计

2.1　网络整体结构

改进后的 YOLOv8Pose-DC 由主干网络、集中式层内调节特征金字塔和多尺度双检测头等三部分组成，其网络总体结构如图 2 所示。图中虚线框内、红色箭头和红色模块表示改动部分。左侧虚线框是把 CASPPF 和 DA 并联，目的是提高特征提取能力的同时使模块拥有捕捉长程依赖关系(全局信息)的能力。中间红线部分是把在深层中获得的特征表示通过上采样到与低级特征相同的空间尺度，然后通过 Concat 特征融合方式，就能在自上而下的路径中增加特征金字塔每一层全局表示的权重，从而获得全面且具有区分性的特征表示。中间红色模块是 DySample 上采样，能够提高上采样效率。右侧红色方块是加入的 SCAM 模块，用来提高全局信息关联能力。右侧虚线框内，采用 40×40 和 20×20 网络的检测头，因为原有 80×80 网络的检测头虽然能检查更小的目标，但是其计算量大，而且对小目标的敏感度不如 40×40 网络的检查头，因此，综合选择 40×40 和 20×20 网络的双检测头。

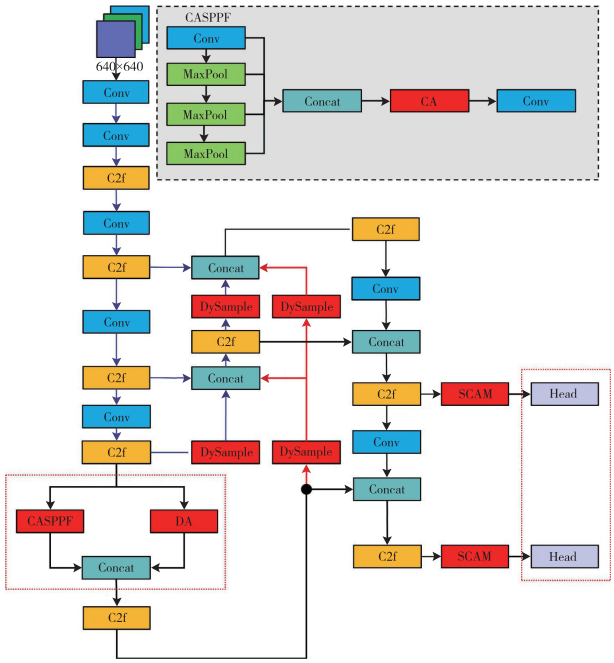
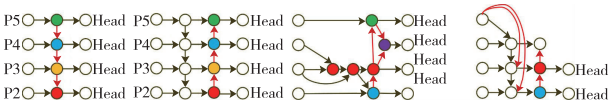


图 2　YOLOv8Pose-DC 结构图

2.2　集中式层内调节特征金字塔

所设计的集中式层内调节特征金字塔，主要包含层内调节层、DA 模块和 CASPPF 模块。FPN 提出了一种自上而下的路径来组合多尺度特征，如图 3(a)所示。PANet[20]在 FPN 之上添加一个额外自上而下的路径聚集网络，如图 3(b)所示。NAS-FPN[21]利用神经网络架构搜索自动设计特征网络拓扑，如图 3(c)所示。以上网络都过分关注层间特征交互，而忽略了层内特征调节，但经验证明，对金字塔进行层内调节是有益处的。于是在 CFP 网络的启发下设计了集中式层内调节特征金字塔。考虑到最深层的特征通常包含浅层特征中稀缺的最为抽象的特征表示，所以在实现中，把深层中获得的特征上采样到与浅层特征相同的空间尺度，然后沿着空间维度与相关的浅层特征相连接。这样使得金字塔能够在自上而下的路径中增加特征金字塔每一层全局表示的空间权重，金字塔能有效提取全面且有区分性的特征表示。所设计的集中式层内调节金字塔

如图 3(d)所示。



(a) FPN (b) PAFPN (c) NAS-FPN (d) 本文

图 3 特征金字塔网络

可变形注意力变换器(Deformable Attention Transformer,DAT)是 Xia 等[22]在 2022 年提出的,本文将其引入所设计的集中式层内调节特征金字塔中,旨在捕捉全局长程依赖性。DAT 结构如图 4 所示。实现原理为:给定输入特征图 $x \in R^{H\times W\times C}$,把生成的 $p \in R^{H_G\times W_G\times 2}$ 均匀网络作为参考。网格输入大小是输入特征图按因子 r 下采样得到的,其公式为 $H_G=H/r$,$W_G=W/r$。参考点的值是线性间隔的 2D 坐标$\{(0,0),\cdots\cdots,(H_G-1,W_G-1)\}$,然后根据网格形状 $W_G\times H_G$ 将它们标准化到范围$[-1,+1]$,其中$(-1,-1)$表示左下角,$(+1,+1)$表示右上角。为了获得每个参考点的偏移量,对特征图进行线性投影以获得查询标记,公式为 $q=xW_q$,然后将其输入到轻量级子网络 $\theta_{offset}(\cdot)$中以生成偏移量 $\Delta p=\theta_{offset}(q)$。为了稳定训练过程,通过一些预定义的因子 s 来缩放 Δp 的幅度,以防止偏移变得太大,即 $\Delta p \leftarrow s\tanh\Delta p$。然后在变形点的位置对特征进行采样并作为键和值。投影矩阵为:

$$q=xW_q,\tilde{k}=\tilde{x}W_k,\tilde{v}=\tilde{x}W_v \tag{1}$$

$$\Delta p=\theta_{offset}(q),\tilde{x}=\varphi(x;p+\Delta p) \tag{2}$$

式中,$\tilde{k}$ 和 $\tilde{v}$ 分别表示变形的键和值嵌入。采用 $\varphi(.;.)$设置为双线性插值,使其可微:

$$\varphi(z;(p_x,p_y))=\sum_{(r_x,r_y)} g(p_x,r_x)g(p_y,r_y)\cdot z[r_y,r_x,:] \tag{3}$$

式中,$g(a,b)=\max(0,1-|a-b|)$并与$[r_x,r_y]$索引 $z\in R^{H\times W\times C}$ 上的所有位置。由于 g 仅在最接近(p_x,p_y)的 4 个积分点上不为零,可将公式(3)简化为 4 个位置上的加权平均值。与现有方法类似,对 q,k,v 进行多头注意力并采用相对位置偏移 R。注意力头的输出表示为:

$$z^{(m)}=\sigma(q^{(m)}\tilde{\kappa}^{(m)T}/\sqrt{d}+\varphi(\hat{B};R))\tilde{v}^{(m)} \tag{4}$$

式中,$\varphi(\hat{B};R)\in R^{HW\times H_GW_G}$。每个头部特征连接在一起并通过 W_o 投影。

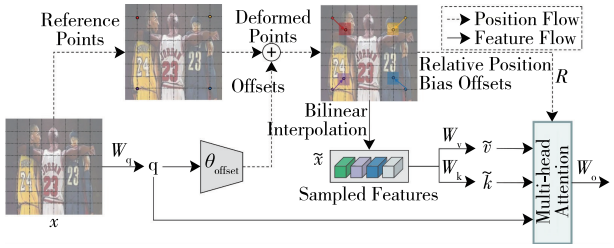


图 4 可变形注意力模块

CASPPF 的结构如图 5 所示。SPPF 模块是在 YOLOv5 的 SPP 结构上进行改进得到的。通过减少需要进行池化的特征图的数量,解决了 SPP 在计算过程中一直存在的冗余问题。此外,与 SPP 相比,SPPF 允许每个池化层具有不同的池化因子,这使得模型可以更加灵活地处理不同尺寸的目标。CASPPF 是在原有的 SPPF 的残差网络后加入 CA 组成。CA 不仅可以使得 SPPF 在进行特征提取时能够更加聚焦在重要信息上,从而提升 SPPF 的特征提取能力,而且可以增强模型对于噪声和干扰的抵抗能力,使得模型具有更强的适应能力,即提高模型的鲁棒性。

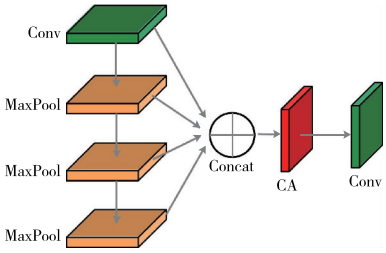


图 5 CASPPF 模块结构图

2.3 双检测头结构

双检查头结构如图 6 所示。为了降低模型计算量,提高模型对密集人群的检查效率,设计了双检查头结构。原 YOLOv8sPose 中有三个检测头,分别对应分辨率为 80×80,40×40 和 20×20。其中分辨率为 80×80 的检测头的计算量大,从而降低了模型的检测速度和检测效率。实验采用 40×40 和 20×20 的双检测头结构,这种结构计算量较小,大大提高了模型的检测速度和检测效率,并通过消融实验证明了双检查头设计方案的可行性。

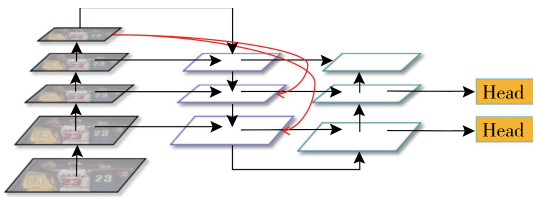


图 6 双检测头结构图

2.4 DySample 模块

DySample[23]模块是一款超轻量且高效的动态上采样器。DySample 采样器是从点采样的角度设计的，不需要高分辨率引导作为输入，也不需要除 PyTorch 之外的任何额外 CUDA 包、内存占用、FLOP 和参数量，在保证轻量化的同时提高了采样效率。DySample 模块的原理图如图 7 所示。

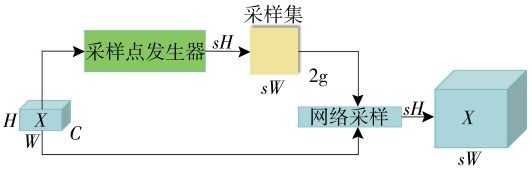


图 7　DySample 原理图

采样网络：给定大小为 $C\times H_1\times W_1$ 的特征图 X 和大小为 $2\times H_2\times W_2$ 的采样集 S，其中第一维的 2 表示 x 和 y 坐标，网络采样函数使用 S 中的位置来重新将假设的双曲性插值 X 采样大小为 $C\times H_2\times W_2$ 的 X'。定义式为：

$$X'=\text{grid_sample}(X,S) \tag{5}$$

给定上采样比例因子 s 和大小为 $C\times H\times W$ 的特征图 X，通过输入和输出通道数分别为 C 和 $2s^2$ 的线性层来生成大小为 $2s^2\times H\times W$ 的偏移量 O，然后使用像素洗牌将其重塑为 $2\times sH\times sW$。即采样集合 S 是偏移量 O 与原始采样网络 G 之和。公式表达为：

$$O=\text{linear}(X) \tag{6}$$

$$S=G+O \tag{7}$$

2.5 空间上下文感知模块

SCAM[24]是一种用于计算视觉任务的特征增强模块，旨在提高跨渠道、跨空间的全局关联能力，其主要思想是利用局部和全局的空间上下文信息来调整特征图中的响应，以更好地捕获图像中的语义信息和结构信息。全局上下文信息可以用来表示像素跨空间关系，从而抑制无用的背景并增强物体和背景之间的区分度，进而提高检测的准确性。SCAM 主要由三个分支组成，如图 8 所示。第一个分支是使用 GAP 和 GMP 来集成全局信息。第二个分支使用 1×1 卷积生成特征图的线性变换结果。第三个分支使用 1×1 卷积来简化查询和键的倍数，此卷积为了和第二个分支中的卷积区分开，把它命名为 QK。随后第一和第三分支分别与第二分支进行矩阵乘法。获得的两个分支分别代表跨通道和跨空间的上下文信息。最后，在这两个分支上使用 Hadamard 积得到 SCAM 的输出。在每一层中，像素级空间上下文可以表示为：

$$Q_i^j=P_i^j+a_i^j\sum_{j=1}^{N_i}\left[\frac{\exp(w_{qk}P_i^j)}{\sum_{n=1}^{N_i}\exp(w_{qk}P_i^n)}\cdot w_vP_i^j\right] \tag{8}$$

$$a_i^j=\frac{\exp([\text{avg}(P_i);\max(P_i)P_i^j])}{\sum_{n=1}^{N_i}\exp([\text{avg}(P_i);\max(P_i)P_i^n])}\cdot w_v \tag{9}$$

式中，P_i^j 和 Q_i^j 分别表示 i 层特征图中第 j 个像素输入和输出。N_i 表示像素总数。w_{qk} 和 w_v 是用于投影特征图的线性变换矩阵，通过 1×1 卷积进行简化。avg(·)和 max(·)分别执行 GAM 和 GMP。GAM 和 GMP 可以指导特征图选择具有重要信息的通道，这使得 SCAM 能够了解有关通道维数的上下文信息。

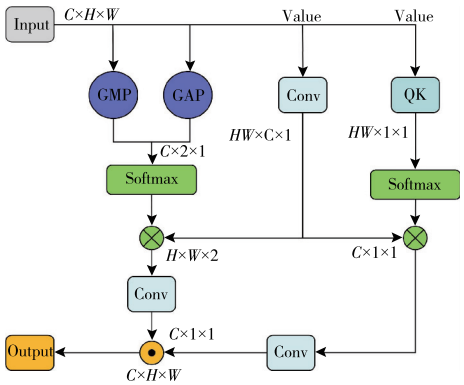


图 8　SCAM 原理图

3　实验

3.1 实验数据集与实验环境

本次实验的数据集是在 MS COCO(Common Objects in Context)2017 数据集中随机选择 20 000 张图片及其对应的标签，其中包括 18 000 张训练集、2 000 张验证集，分别涵盖单人、多人、小目标、大目标等多类型的图片。

本实验是在 AutoDL 服务器上进行，CPU 的配置为 12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU，GPU 的配置为 RTX 3080，CUDA 版本为 11.3，PyTorch 版本为 1.11.0，Python 版本为 3.8。设置 YOLOv8pose 训练参数，其中，初始学习率为 0.01，迭代次数为 200 轮，批处理大小为 16，输入图像分辨率为 640×640。

3.2 评价指标

为了准确评估深度学习模型在姿态检测任务上

的表现，使用目标检测任务领域内公认的评估标准—mAP，即数据集中所有类别精度平均概念，作为评价检测准确性的关键指标。

精准度 P(Precision)、召回率 R(Recall)、平均精度均值(mAP)和所有类别在特定 IoU 阈值下的平均精度(AP)。公式如下：

$$P=\frac{TP}{TP+FP} \tag{10}$$

$$R=\frac{TP}{TP+FN} \tag{11}$$

$$AP=\int_0^1 P(R)\mathrm{d}R \tag{12}$$

$$mAP=\frac{\sum_{i=1}^{n}AP_i}{n} \tag{13}$$

式中，TP 表示被正确分类的正例个数，FP 表示被错误分类的负例个数，FN 表示被错误分类的正例个数，TN 表示被正确分类的负例个数。$mAP@0.5$ 是指 IoU 设置为 0.5 时所有类别的平均 AP，$mAP@0.5:0.95$ 是指不同 IoU 阈值下的平均 mAP。IoU 范围为 0.5 至 0.95，步长为 0.05。FPS 表示每秒检测到的图片，数值越高，检测速度越快。计算量 $GFLOPs$ 表示计算复杂程度，数值越大，计算复杂度越高。

3.3 消融实验

为了研究检测头设计对检测性能的影响，在 YOLOv8Pose-DC 的基础上设计了四组不同检测头组合进行对比试验，结果如表 1 所示。通过对四组不同检测头组合进行对比，可以直观地看出，40×40 和 20×20 双检测头组合在 COCO-2017 节选数据集上平均精度等各项精度指标最高且计算量也最小，该组合能够以更小的计算量，获得更高的精度。由此证明，选择 40×40 和 20×20 双检测头组合是最合适的。

表 1 不同检测头组合对实验效果的影响

Grid scale	$GFLOPs$	P/%	R/%	$mAP@0.5$/%	$mAP@0.5:0.95$/%
80×80,40×40,20×20	35.5	84.6	77.4	80.8	49.3
80×80,40×40	33.6	83.8	75.7	80.1	47.7
80×80,20×20	32.4	84.6	76.4	81.3	50.2
40×40,20×20	32.0	85.5	77.4	82.1	50.9

消融实验如表 2 所示，实验表明每个改进点对模型性能都有所提升。表 2 中一共展示了 8 组实验，其中模型 a 是 YOLOv8s-Pose 的基准模型。模型 b 是对基准模型结构改进，包含集中式层内调节特征金字塔和双检测头，此模型 $mAP@0.5$ 和 $mAP@0.5:0.95$ 分别提高了 2.3 个百分点和 1 个百分点，证明了其改进的有效性。模型 c 是把基准模型的 UpSample 替换成 DySample，$mAP@0.5$ 提高了 0.9 个百分点，证明 DySample 上采样对模型精度提升是有帮助的，并且计算量也有所下降，提高了模型的上采样效率。模型 d 是在基准模型的 Neck 加入 SCAM 模块，此模型 $mAP@0.5$ 提高了 0.9 个百分点，证明 SCAM 模型提高模块全局关联能力有助于提高模型的检测精度。模型 e,f,g 分别是在模型结构改进的基础上加入 DySample 模块、模型结构改进的基础上加入 SCAM 模块和 DySample 模块与 SCAM 模块同时加入到基准模型的情况，以上三个模块对精度提升都有贡献，证明三个模块两两组合对模块精度提升没有负面影响。模型 h 是所改进的 YOLOv8Pose-DC 模型，通过表格可以看出，改进后的模型在精度方面都有所提升，其中模型的 P，R，$mAP@0.5$ 和 $mAP@0.5:0.95$ 分别提升了 0.3 个百分点、4.2 个百分点、3.1 个百分点和 1.8 个百分点。综上所述，本消融实验证明了改进模型的可行性，实现了模型各项精度指标均提高的效果。

表 2 消融实验

模型	$GFLOPs$	P/%	R/%	$mAP@0.5$/%	$mAP@0.5:0.95$/%	FPS
a	30.2	85.2	73.2	79.0	49.1	30
b	31.6	85.1	76.6	81.3	50.1	27
c	30.2	83.2	75.8	79.9	48.3	29
d	31.1	84.9	75.4	81.2	49.7	26
e	31.3	84.4	76.9	81.3	50.0	27
f	32.3	85.2	77.4	81.6	50.7	25
g	31.1	85.4	75.9	80.8	49.5	28
h	32.0	85.5	77.4	82.1	50.9	25

3.4 对比实验

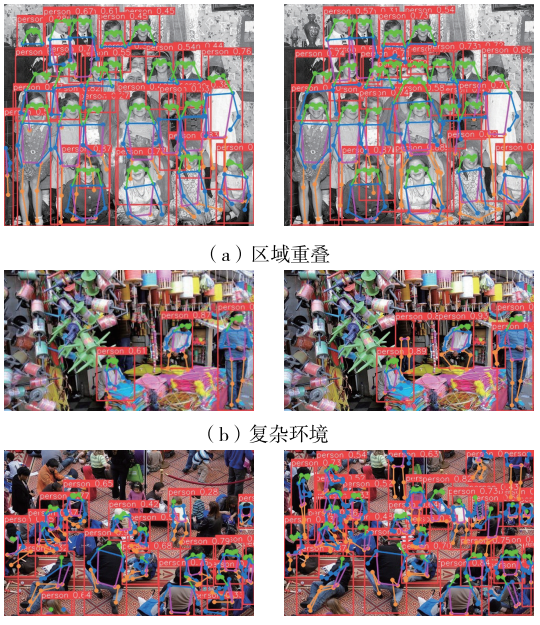
为更加客观展示改进模型在密集人群姿态识别中的效果，对当前主流算法进行对比实验，其中包括 Openpose，Lightweight OpenPose，HRNet，HigherHRNet-w32，YOLOv8nPose 和 YOLOv8sPose，同一选择节选 COCO2017 数据集。评价指标选择 P，R，$mAP@0.5$，$mAP@0.5:0.95$ 和 FPS。结果如表 3 所示。

从表中可以看出，与各个模型相比，YOLOv8Pose-DC 的 P，R，$mAP@0.5$ 和 $mAP@0.5:0.95$ 最优。其中，在 $mAP@0.5$ 指标下 YOLOv8Pose-DC 比 Openpose 高 12.2%，比 Lightweight OpenPose 高 14.9%，比 HRNet 高

14.9%，比 HigherHRNet-w32 高 9.2%，比 YOLOv8nPose 高 8.4%；在 $mAP@0.5:0.95$/%指标下 YOLOv8Pose-DC 比 Openpose 高 2.3%，比 Lightweight OpenPose 高 2.7%，比 HRNet 高 2.7%，比 HigherHRNet-w32 高 3%，比 YOLOv8nPose 高 10.9%。从而可以证明改进模型的可行性。改进前后模型对比图如图 9 所示。

表 3　对比实验

模型	P/%	s	$mAP@0.5$/%	$mAP@0.5:0.95$/%	FPS
Openpose[25]	72.1	67.2	78.9	50.5	10
Lightweight OpenPose[26]	59.1	—	67.2	48.2	23
HRNet[27]	82.9	71.9	79.4	47.7	9
HigherHRNet-w32[28]	63.6	—	72.9	47.9	13
YOLOv8nPose	82.7	69.7	73.7	40.0	39
YOLOv8sPose	85.2	73.2	79.0	49.1	30
YOLOv8Pose-DC	85.5	77.4	82.1	50.9	25



（a）区域重叠

（b）复杂环境

（c）杂乱人群

图 9　改进前后模型效果对比

如图 9 为三组对比图，在每组图中，左侧为原算法检测的效果图，右侧为本文算法检测的效果图。如图 9(a)，在重叠场景下，原算法把右上角的红色瓶子和左下角的毛巾误检成了人，并且没有检测出中间的圣诞老人，而本文算法能在完美解决误检漏检情况的同时，检测精度都有所上升，比如右上对角小男孩的检测精度从 45%上升到了 54%，体现了模型较好的检测能力。如图 9(b)，在复杂环境场景下，原算法没有检测到中间灰色衣服的人，而本文算法能很好地区分人体和环境且精度也有所提升，比如对左下角穿白衣服的老人的检测精度从 61%上升到了 89%，由此可见改进后的模型能很好地把任务和背景区分出来。如图 9(c)，在杂乱人群情境中，原算法没有检测出右上角和左上角的人物，而本文算法能够大大改善原算法所出现的问题，检测到了大部分人物，表明了改进算法对杂乱人群有较好的检测能力。

4　结束语

针对现有算法对密集人群检测存在误检漏检等问题，本文对 YOLOv8sPose 模型进行改进，并将改进后的模型命名为 YOLOv8Pose-DC。设计集中式层内调节特征金字塔，其中包括 DA 模块、CASPPF 模块和层内调节层，使得模型不仅有捕获全局信息的能力，而且能够得到全面且具有区分性的特征表示，解决误检问题。设计多尺度双检测头，提高模型检测效率且相对减少了模型的计算量。DySample 模块替代上采样，提高上采样效率。在颈部加入 SCAM 模块，提高模型的全局信息关联能力。

在节选 COCO2017 数据集上的实验结果表明，与原算法相比，YOLOv8Pose-DC 的 $mAP@0.5$ 提升了 3.1%，R 提升 4.2%，P 提升了 0.3%，满足准确性的要求。通过消融实验、对比试验和对比实验图可知，改进模型大大改善了对密集人群检测所出现的漏检和误检等问题，满足检测要求。在接下来工作中，将聚焦于在提高检测速度的基础上，继续提高检测精度，构建快速高精度密集人群检测算法。

参考文献：

[1] Güler R A, Neverova N, Kokkinos I. Densepose: Dense human pose estimation in the wild[C]// Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7297-7306.

[2] Papaioannidis C, Mademlis I, Pitas I. Fast CNN-based single-person 2D human pose estimation for autonomous systems[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 33(3): 1262-1275.

[3] Tong M L, Han H, Zhu W. Shared latent dynamical structure for three-dimensional human pose estimation[J]. Science China Information Sciences, 2011, 54: 1375-1382.

[4] Wang H, Shi Q, Shan B. Three-dimensional human pose estimation with spatial-temporal interaction enhancement transformer[J]. Applied Sciences, 2023, 13(8): 5093.

[5] Nguyen T D, Kresovic M. A survey of top-down approaches for human pose estimation[J]. arXiv preprint arXiv: 2202.02656, 2022.

[6] Chan Y, Wang Z, Peng Y, et al. Cascaded pyramid network

for multi-person pose estimation[C]// Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7103-7112.

[7] Carreira J, Zakharov D. PoseNet: A convolutional network for real-time 6-DOF camera relocalization [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(9): 1881-1893.

[8] Yin R, Yin J. Global relation modeling and refinement for bottom-up human pose estimation[J]. arXiv preprint arXiv: 2303.14888, 2023.

[9] Maji D, Nagori S, Mathew M, et al. Yolo-pose: Enhancing Yolo for multi person pose estimation using object keypoint similarity loss[C]// Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 2637-2646.

[10] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]// Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[11] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C]// Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.

[12] Quan Y, Zhang D, Zhang L, et al. Centralized feature pyramid for object detection [J]. IEEE Trans. on Image Processing, 2023, 32: 4341-4354.

[13] Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C]// Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.

[14] 井方科, 任红格, 李 松. 基于多尺度特征融合的小目标交通标志检测[J/OL]. 激光与光电子学进展, 1-13[2024-05-26].
Jing Fangke, Ren Hongge, Li Song. Small target traffic signdetection based on multi-scale feature fusion [J/OL]. Advances in Laser and Optoelectronics, 1-13[2024-05-26].

[15] Toshev A, Szegedy C. Deeppose: Human pose estimation via deep neural networks[C]// Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2014: 1653-1660.

[16] 马明旭, 马 宏, 宋华伟. 基于YOLO-Pose的城市街景小目标行人姿态估计算法[J]. 计算机工程, 2024, 50(4): 177-186.
Ma Mingxu, Ma Hong, Song Huawei. Pedestrian pose estimation algorithm based on YOLO-Pose in urban street view [J]. Computer Engineering, 2024, 50(4): 177-186.

[17] 王 茹, 刘大明, 张 健. Wear-YOLO:变电站电力人员安全装备检测方法研究[J]. 计算机工程与应用, 2024, 60(9): 111-121.
Wang Ru, Liu Daming, Zhang Jian. Wear-YOLO: Research on detection method of safety equipment for power personnel in substation[J]. Computer Engineering and Applications, 2024, 60(9): 111-121.

[18] 赵志宏, 郝子晔. 改进YOLOv8的航拍小目标检测方法: CRP-YOLO[J/OL]. 计算机工程与应用: 1-15[2024-06-09].
Zhao Zhihong, Hao Ziye. Improved YOLOv8 aerial small target detection method: CRP-YOLO [J/OL]. Computer Engineering and Applications: 1-15[2024-06-09].

[19] 李 松, 史 涛, 井方科. 改进YOLOv8的道路损伤检测算法[J]. 计算机工程与应用, 2023, 59(23): 165-174.
Li Song, Shi Tao, Jing Fangke. Improved road damage detection algorithm of YOLOv8[J]. Computer Engineering and Applications, 2023, 59(23): 165-174.

[20] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]// Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, 2018: 8759-8768.

[21] Ghiasi G, Lin T Y, Le Q V. Nas-fpn: Learning scalable feature pyramid architecture for object detection[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2019: 7036-7045.

[22] Xia Z, Pan X, Song S, et al. Vision transformer with deformable attention[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2022: 4794-4803.

[23] Liu W, Lu H, Fu H, et al. Learning to upsample by learning to sample[C]// Proc. of the IEEE/CVF Inter. Conf. on Computer Vision, 2023: 6027-6037.

[24] Zhang Y, Ye M, Zhu G, et al. FFCA-YOLO for small object detection in remote sensing images[J]. IEEE Trans. on Geoscience and Remote Sensing, 2024, 62: 1-15.

[25] Gao Z, Hidalgo G, Simon T, et al. OpenPose: realtime multi-person 2D pose estimation using part affinity fields[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(1): 172-186.

[26] Osokin D. Real-time 2D multi-person pose estimation on CPU: Lightweight openpose[C]// International Conference on Pattern Recognition Applications and Meth, 2019: 53996245.

[27] Sun K, Xiao B, Liu D, et al. Deep high-resolution representation learning for human pose estimation [C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2019: 5686-5696.

[28] Cheng B W, Xiao B, Wang J D, et al. HigherHRNet: Scale-aware representation learning for bottom-up human pose estimation[C]// Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2020: 5386-5395.

作者简介:

谷学静(1972—),女,河北省唐山市人,博士,教授,主要研究方向为数字媒体技术、人机交互、虚拟人及智能Agent、人工心理和人工情感;

郭志斌(1972—),男,山东省泰安市人,硕士研究生,主要研究方向为姿态识别。