DOI: 10.13382/j.jemi.B2508731

基于改进 YOLOv13n-Pose 的心肺复苏识别算法*

谢天宇[1] 何赟泽[2] 邓海平[1] 邓堡元[1] 宋殿义[3] 方学林[4]

(1. 湖南大学人工智能与机器人学院 长沙 410082;2. 湖南大学电气与信息工程学院 长沙 410082;
3. 国防科技大学军政基础教育学院 长沙 410072;4. 湖南辰超科技有限公司 长沙 410007)

摘 要:心肺复苏(cardiopulmonary resuscitation, CPR)是提升心脏骤停患者存活率的关键。传统 CPR 培训与考核主要依赖人工观察,存在主观性强、量化数据缺失等问题,而现有的方法在 CPR 特有的肢体严重遮挡及精细动作的捕捉上存在精度不足。针对以上问题,提出了一种基于改进 YOLOv13n-Pose 的心肺复苏姿态估计算法。该算法在构建的多人多场景的 CPR 姿态估计数据集上进行验证。首先,针对操作者手臂与躯干形成的严重自遮挡问题,引入双层路由注意力机制(BiLevel routing attention, BRA),通过动态稀疏感知增强网络在复杂场景下的全局特征提取能力。其次,设计混合增强上采样模块(hybrid enhanced Upsample, HEUpsample)替换颈部上采样结构,利用空间偏移与通道重排策略减少特征图放大过程中的细节丢失,提升关键点定位精度。实验结果表明,改进模型在 CPR 数据集上的行为识别与姿态估计 mAP 分别达到 95.8%和 88.5%,较原始模型分别提升 2.3%和 3.6%。相较于主流算法,该模型能保持较高的推理速度,同时降低了计算量,并在 OCHuman 公开数据集上验证了良好的泛化性。

关键词: YOLOv13n-Pose;心肺复苏;姿态估计;双层路由注意力机制;混合增强上采样

中图分类号: TP391.4;TN911 **文献标识码**: A **国家标准学科分类代码**: 520.6040

Cardiopulmonary resuscitation recognition algorithm based on improved YOLOv13n-Pose

Xie Tianyu[1] He Yunze[2] Deng Haiping[1] Deng Baoyuan[1] Song Dianyi[3] Fang Xuelin[4]

(1. School of Artificial Intelligence and Robotics, Hunan University, Changsha 410082, China; 2. School of Electrical and Information Engineering, Hunan University, Changsha 410082, China; 3. Undergraduate School, National University of Defense Technology, Changsha 410072, China; 4. Hunan Chenchao Technology Co., Ltd., Changsha 410007, China)

Abstract: Cardiopulmonary resuscitation (CPR) is critical for improving the survival rate of patients with cardiac arrest. Traditional CPR training and assessment mainly rely on manual observation, which suffers from strong subjectivity and a lack of quantitative metrics. Moreover, existing methods show insufficient accuracy in handling the severe limb occlusions and fine-grained motion capture that are characteristic of CPR. To address these challenges, this paper proposes a CPR pose estimation algorithm based on an improved YOLOv13n-Pose model. The proposed method is validated on a self-constructed CPR pose estimation dataset covering multiple subjects and scenarios. First, to tackle the severe self-occlusion formed by the rescuer's arms and torso, a Bi-Level routing attention (BRA) mechanism is introduced, which enhances the network's global feature extraction capability in complex scenes through dynamic sparse perception. Second, a hybrid enhanced upsample (HEUpsample) module is designed to replace the neck upsampling structure. By incorporating spatial offset and channel shuffle strategies, the module reduces detail loss during feature map upsampling and improves keypoint localization accuracy. Experimental results demonstrate that the improved model achieves mAP scores of 95.8% and 88.5% for CPR action recognition and pose estimation on the CPR dataset, representing improvements of 2.3% and 3.6% over the baseline model, respectively. Compared with mainstream methods, the proposed approach maintains a high inference speed while reducing computational cost, and it also demonstrates strong generalization ability on the public OCHuman dataset.

Keywords: YOLOv13n-Pose; CPR; pose estimation; BRA; HEUpsample

收稿日期: 2025-09-08 Received Date: 2025-09-08

*基金项目: 湖南省科技创新项目(2023RC1039)资助

0 引 言

心肺复苏(cardiopulmonary resuscitation, CPR)是抢救心脏骤停患者、提升生存率的关键技术,其施救质量直接决定了抢救的成功率。目前 CPR 培训与考核主要依赖专业导师的现场观察与主观评分。然而,即便是专业人员也难以仅凭肉眼精确捕捉例如按压频率或按压深度等指标的微小偏差,导致考核结果存在主观性强、量化数据缺失等问题[1-2]。虽然基于压力传感器的模拟人模型可以获取部分数据,但其高昂的成本限制了大规模普及。因此,利用计算机视觉技术实现低成本、非接触式且高精度的 CPR 识别,成为急救医学与人工智能交叉领域的迫切需求[3]。

人体姿态估计技术旨在通过图像数据识别人体的关键关节点与姿态结构,为运动行为的精确分析提供支撑[4]。将姿态估计技术引入 CPR 考核系统,可以实现对操作者动作的非接触式识别与定量分析。人体姿态估计主要分为两类,以 OpenPose[4]、HigherHRNet[5]、DEKR[6]等为代表的自底向上算法和以 Mask R-CNN[7]、RMPE[8]、YOLO-Pose[9]等为代表的自顶向下算法。自底向上的算法直接在图像中检测出多个人体关键点,利用肢体走向辅助关键点的聚类等局部关键点组合形成完整的人体姿势,这种方法虽然运行效率较高,但在处理 CPR 这种肢体密集交叉的场景时,极易因局部关键点聚类错误导致姿态畸变。自顶向下的算法则是通过先在图像中进行人体的边界框检测,再对每个人进行单独的关键点检测,但对算力要求严苛且计算效率较低。随着 YOLO 系列算法的不断创新优化[10-16],其目标检测与姿态估计的性能得到了显著提升。尤其是 Lei 等[17]提出的 YOLOv13,融合了超图增强机制和全流程特征分发策略,在实时目标检测中展现了优异性能,为解决上述矛盾提供了新的思路。

尽管现有通用模型表现优异,但将其直接应用于 CPR 智能考核仍面临着高精度要求与复杂遮挡环境之间的矛盾。CPR 操作具有极强的特殊性,操作者需双手紧扣并垂直按压患者胸部,这种动作导致手腕、手肘与躯干之间产生严重的肢体遮挡与自遮挡,极易造成关键点丢失。此外,在长时间的按压过程中,施救者因疲劳产生的姿态变形往往非常隐蔽,普通网络难以在全局感受野下捕捉这种细微的非刚性形变。目前使用姿态估计算法实现对心肺复苏的质量评估仍处于起步阶段,心肺复苏对于操作者动作的标准与流程的正确有着严格的要求,加大了算法识别的精度要求。范莹莹等[18]提出一种改进的多阶段三维人体姿态估计算法和基于 LSTM-FCN 的动作评价模型,该方法将姿态估计和行为识别分为两个网络模型对心肺复苏进行识别与评分,对计算资源要求高且识别精度不高。靳杰[19]提出一种基于 OpenPose 的心肺复苏识别算法,但仅能识别关节点而无法进行行为识别。陈姝琪等[20]提出基于骨骼点数据的行为识别算法,但没有涉及如何从 RGB 图像生成骨骼点数据。张敏[21]利用人体三维重建技术生成三维人体骨骼数据并进行 CPR 动作评估,但需要较大的计算开销。张斯琪[22]基于 OpenPose 实现对 CPR 中人工呼吸动作的判别,但缺乏 CPR 全流程的识别。行为识别方面,何赟泽等[23]提出了一种改进 DETR 的行为识别算法,任丹彤等[24]提出了一种基于双光融合的行为识别算法,但都无法获取骨骼点数据,无法提供针对肢体角度的精细化指导。

综上,为实现对心肺复苏操作者与假人进行精确地姿态估计,并在检测人体框的同时进行动作识别,为心肺复苏提供动作指导,本文提出了一种基于改进 YOLOv13n-Pose 的姿态估计算法。改进原有网络的注意力机制,引入了双层路由注意力机制进行更好的特征提取,并提出了一种混合增强上采样模块来替换原有的上采样模块,提升原有网络上采样后的特征表达质量。实验证明,改进方法可以实现对心肺复苏姿态估计的高精度,自动化检测,对于姿态估计的研究具有积极意义。

1 原始模型

YOLOv13 的网络结构包括了主干网络、颈部网络和检测头。其中,YOLOv13 采用深度可分离卷积(depthwise separable convolution)与改进的轻量模块(如 DS-Bottleneck、DS-C3k2 等)构建高效的主干网络,以低计算代价提取多尺度特征图;引入基于超图的自适应相关增强机制(hypergraph-enhanced adaptive correlation enhancement, HyperACE)模块,通过将特征图中的像素点建模为超图节点,利用可学习的超边进行高阶全局关系建模与跨尺度特征增强;进一步地,为实现跨阶段的特征协同与信息流动,YOLOv13 采用了全流程聚合与分发范式(full pipeline aggregation-and-distribution, FullPAD)机制,将增强后的特征分别分发至主干连接、颈部结构及检测头之间的多个关键位置,有效地实现了整个网络中的细粒度信息流和表示协同。在输出阶段,YOLOv13 结合多尺度检测头对融合特征进行解码,实现对不同尺寸目标的精确定位与分类。得益于上述设计,YOLOv13 在保持较小模型体积和高推理速度的同时,具备良好的泛化能力,为人体关键点检测及结构改进提供了坚实的基础。

2 改进算法

针对心肺复苏考核场景中动作规范性要求高、肢体

自遮挡严重以及应用设备计算资源受限等特点，本文选用参数量较小的 YOLOv13n 模型作为基础架构。为了克服原始网络在处理复杂非刚性形变时的特征提取局限，并提升对关键点的定位精度，本文重点对主干网络与特征融合网络进行了针对性优化。YOLOv13n 中的主干网络所采用的 A2C2f 模块虽然具备一定的局部建模能力，但在面对大尺度目标或复杂上下文关系时，其感受野与信息交互能力存在一定局限，难以有效捕捉远距离依赖信息。因此本文采用了一种双层路由注意力机制（BiLevel routing attention，BRA）[25] 来替换原有的 A2C2f 模块，用于在引导特征交互过程中能够自适应地挖掘关键信息区域，从而提升主干网络特征提取能力与鲁棒性。同时，为了提升 YOLOv13n 的颈部网络的特征融合性能，本文构建了一种混合增强上采样模块（hybrid enhanced upsample，HEUpsample）用于替换原有的 Upsample 上采样模块。图 1 为本文的网络模型结构。

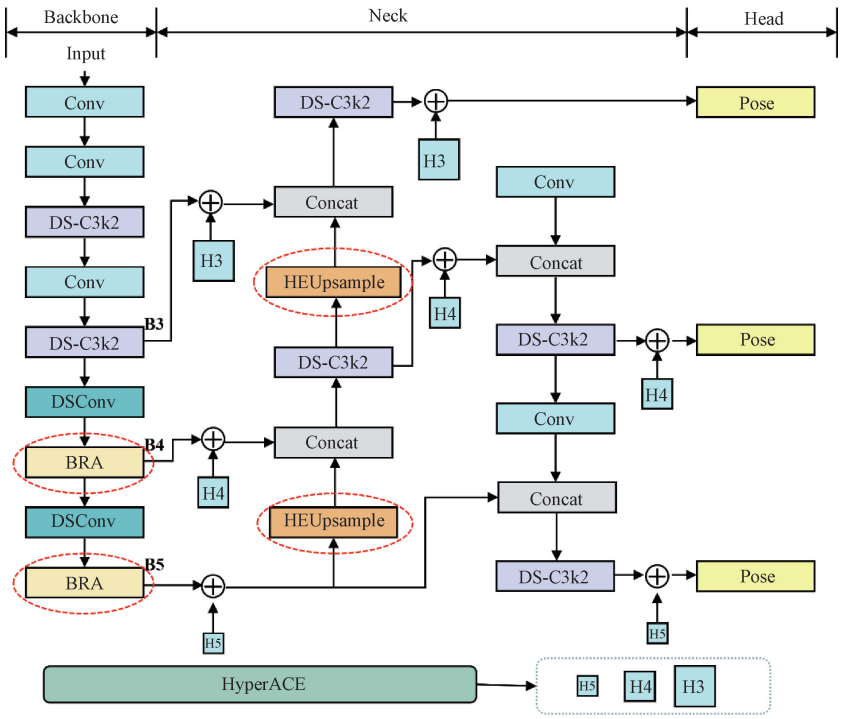


图 1　本文网络结构

Fig. 1　Framework of ours

2.1　双层路由注意力机制

在 CPR 操作过程中，操作者需双手紧扣并垂直按压患者胸部，这种特定的姿态导致手部、腕部与胸部之间形成高频的重叠区域，极易造成特征混淆。此外，原始 YOLOv13n 主干网络中的 A2C2f 模块虽然具备局部特征融合能力，但在面对这种严重的肢体自遮挡时，难以有效捕捉肩部、肘部与手腕之间的长距离关系，容易导致关键点识别丢失。

为此，本文在主干网络的深层特征提取阶段，引入双层路由注意力机制（BRA）替换原有的 A2C2f 模块。BRA 的核心优势在于其动态稀疏感知能力，能够跨越局部遮挡，通过全局关联来增强关键部位的特征表达。

BRA 是一种动态的、查询感知的稀疏注意力机制，核心思想是在粗粒度区域级别中过滤大部分不相关的键值对，仅保留少量的被路由区域，然后在这些路由区域中集中的进行细粒度的 token-to-token 的注意力计算。结构如图 2 所示。与 CBAM、SE 等广泛使用的卷积注意力模块相比，BRA 不仅局限于对局部空间或通道维度的加权，而是继承了 Vision Transformer 捕捉长距离依赖的能力，能够有效跨越遮挡区域，建立起肩、肘、腕等关键点之间的全局几何约束。

首先，BRA 将输入特征图 $X \in \boldsymbol{R}^{H\times W\times C}$ 划分为 $S\times S$ 个不重叠区域，则每个区域包含 $\frac{HW}{S^2}$ 个特征向量，该步骤将输入特征重塑为 $X^{\gamma} \in \boldsymbol{R}^{S^2\times\frac{HW}{S^2}\times C}$，再通过线性映射获得查询矩阵、键矩阵和值矩阵 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V} \in \boldsymbol{R}^{S^2\times\frac{HW}{S^2}\times C}$，计算公式如式（1）所示。

$$\boldsymbol{Q} = X^{\gamma}W^{q},\boldsymbol{K} = X^{\gamma}W^{\kappa},\boldsymbol{V} = X^{\gamma}W^{v} \tag{1}$$

式中：W^{q}、W^{κ}、$W^{v} \in \boldsymbol{R}^{C\times C}$ 分别是查询矩阵、键矩阵和值矩阵对应的映射权重。

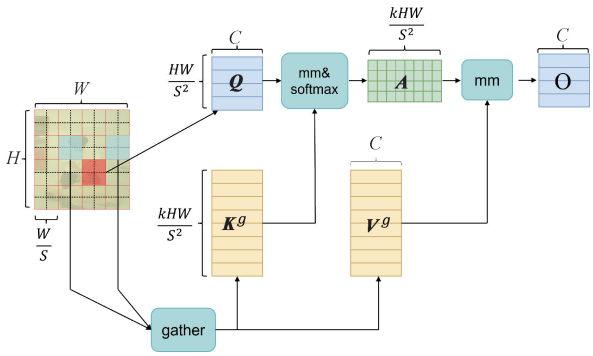


图 2 BRA 模块结构

Fig. 2 BRA module structure

接着,通过构建有向图推导区域间的注意力关系,分别对查询矩阵 $\boldsymbol{Q}$ 和键矩阵 $\boldsymbol{K}$ 进行逐区域的平均计算得到区域级别的查询矩阵和键矩阵 $\boldsymbol{Q}^{\gamma},\boldsymbol{K}^{\gamma}\in\boldsymbol{R}^{S^2\times C}$,将 $\boldsymbol{Q}^{\gamma}$ 和转置 $\boldsymbol{K}^{\gamma}$ 两个矩阵进行矩阵乘法计算得到区域间关联图的邻接矩阵 $\boldsymbol{A}^{\gamma}$。邻接矩阵 $\boldsymbol{A}^{\gamma}$ 中的元素表示区域间的语义关联强度,根据邻接矩阵 $\boldsymbol{A}^{\gamma}$ 保留每个区域的最关联的前 k 个连接得到路由索引矩阵 $\boldsymbol{I}^{\gamma}$,如式(2)所示。

$$\boldsymbol{A}^{\gamma}=\boldsymbol{Q}^{\gamma}(\boldsymbol{K}^{\gamma})^{\mathrm{T}},\boldsymbol{I}^{\gamma}=topIndex(\boldsymbol{A}^{\gamma}) \tag{2}$$

最后,利用路由索引矩阵 $\boldsymbol{I}^{\gamma}$ 聚合键矩阵和值矩阵便于加速计算,将聚合的键值矩阵进行最终细粒度的 token-to-token 的注意力计算,如式(3)所示。

$$\begin{aligned}&\boldsymbol{K}^{g}=gather(\boldsymbol{K},\boldsymbol{I}^{\gamma}),\\&\boldsymbol{V}^{g}=gather(\boldsymbol{V},\boldsymbol{I}^{\gamma}),\\&O=Attention(\boldsymbol{Q},\boldsymbol{K}^{g},\boldsymbol{V}^{g})+LCE(\boldsymbol{V})\end{aligned} \tag{3}$$

式中:$\boldsymbol{K}^{g}$ 为聚合键矩阵;$\boldsymbol{V}^{g}$ 为聚合值矩阵;O 为输出;$LCE(\cdot)$ 为局部上下文增强模块[26]。

2.2 混合增强上采样模块

在姿态估计任务中,颈部网络负责将深层语义信息与浅层细节信息进行融合,其上采样操作直接影响关键点的像素级定位精度。CPR 考核对动作标准有严格要求,这意味着模型必须具备极高的边缘还原能力。然而,YOLOv13n 原有的上采样模块采用最近邻插值,这种粗糙的插值方式容易在特征图放大过程中丢失细节,导致肢体边缘模糊,进而影响关节点坐标的回归精度。

为了解决上述问题,本文构建了一种混合增强上采样模块(HEUpsample)替换 Neck 结构中的标准上采样模块。HEUpsample 是一种集成空间与通道信息融合的轻量化模块,核心思想是通过结合深度可分离卷积、通道重排机制以及空间偏移混合策略,实现结构感知与信息融合的统一。与 CARAFE 等上采样操作相比,HEUpsample 通过引入通道重排与空间偏移这种无参数的操作,以极低的计算代价实现了跨通道与跨空间的特征交互。这种设计既避免了普通插值的语义缺失,又规避了复杂算子的高延迟,在精度与速度之间取得了最佳平衡,更契合 CPR 实时纠错的应用场景。

首先,HEUpsample 将输入特征图 $X\in\boldsymbol{R}^{H\times W\times C}$ 经过上采样操作 $Up(\cdot)$,将特征图的空间尺寸提升至两倍 $X^{\rho}=\boldsymbol{R}^{2H\times 2W\times C}$,并使用深度可分离卷积 $DWC(\cdot)$ 来增强上采样后的特征图,如式(4)所示。

$$X^{\alpha}=DWC(Up(X)) \tag{4}$$

接着使用通道重排机制(channel shuffle,CS)来提升跨通道的信息交互能力。该机制将输入特征 $X^{\alpha}=\boldsymbol{R}^{2H\times 2W\times C}$ 划分为 G 个通道组得到 $X^{\beta}=\boldsymbol{R}^{2H\times 2W\times G\times\frac{C}{G}}$,并在 $\left(G,\frac{C}{G}\right)$ 维度重排其通道顺序得到 $X^{\beta}=\boldsymbol{R}^{2H\times 2W\times\frac{C}{G}\times G}$,接着恢复得到 $X^{\delta}=\boldsymbol{R}^{2H\times 2W\times C}$。将经过通道重排机制的特征图 $X^{\delta}=\boldsymbol{R}^{2H\times 2W\times C}$ 进行空间偏移(spatial shift,SS)融合,把输入通道分为 4 组 $x_1,x_2,x_3,x_4\in\boldsymbol{R}^{2H\times 2W\times\left(\frac{C}{4}\right)}$,分别沿着高度维度和宽度维度进行循环平移,以捕获不同方位的空间特征,如图 3 所示。

3 实验配置与数据集

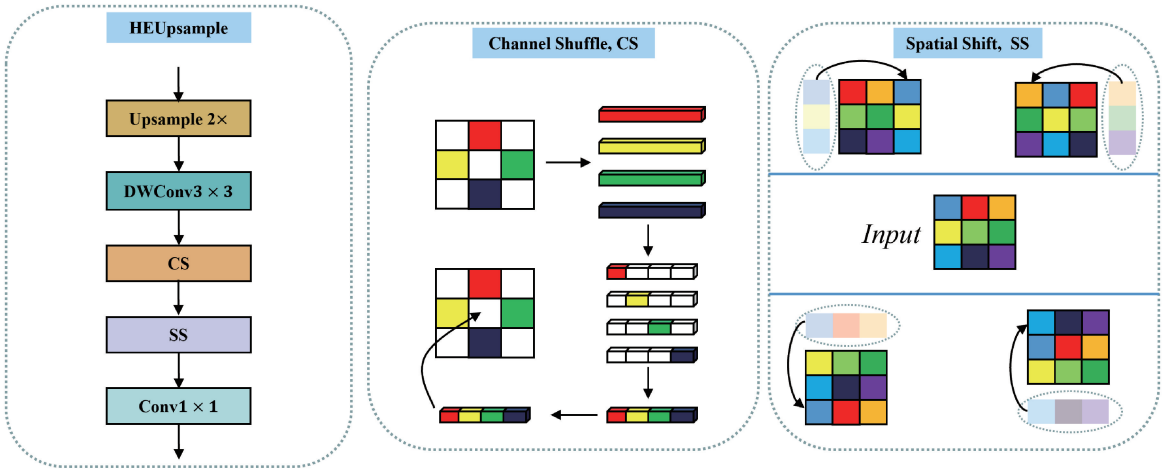
3.1 实验配置

本实验所采用的 CPU 为 INTEL(R) XEON(R) PLATINUM 8551C,显卡为 NVIDIA GeForce RTX4090。软件环境为 CUDA12.2,操作系统为 Linux。网络模型基于 Pytorch2.0.1 框架,Python 版本为 3.8。所有的网络模型训练 100 轮,批次大小为 32,初始学习率为 0.01。

3.2 CPR 姿态估计数据集

为了构建具有代表性的 CPR 姿态估计数据集,本文模拟真实急救考核环境采集了分辨率为 1 920×1 080、帧率 30 fps 的视频数据。CPR 姿态估计数据集组成如表 1 所示。采集场景包括模拟急救室、半户外自然光及复杂背景环境,以确保数据的多样性。经抽帧与去重处理后,共筛选出 4 539 张原始图像,并在训练阶段引入 Mosaic 与随机旋转增强以提升样本鲁棒性。数据标注采用 LabelMe 工具,遵循 MS COCO 规范。标注内容包括操作者 17 个关键点的坐标定位,以及根据心肺复苏标准流程划分评估环境、判断意识、判断呼吸、呼救、摆体位、开放气道、去除分泌物、胸外按压、人工呼吸、判断结果和其他共 11 个操作者的动作识别,对假人划分为假人衣物未打开,假人衣物打开 2 个状态。最终将数据集按视频来源随机划分为训练集 3 756 张与验证集 783 张,用于模型的训练与评估。

图 4 所示为操作者对应的 10 个主要动作类别,不包含其他动作类别。其中,前 5 个动作类别中假人模型为衣物未打开,后 5 个动作类别中假人模型为衣物已打开。



(a) HEUpsample模块
(a) HEUpsample module

(b) 通道重排模块
(b) Channel shuffle module

(c) 空间偏移模块
(c) Spatial shift module

图 3　HEUpsample 模块结构

Fig. 3　HEUpsample module structure

(a) 评估环境
(a) Evaluate_environment

(b) 判断意识
(b) Judge_consciousness

(c) 判断呼吸
(c) Judge_breathing

(d) 呼救
(d) Call_help

(e) 摆体位
(e) Put_posture

(f) 开放气道
(f) Open_airway

(g) 去除分泌物
(g) Remove_obstruction

(h) 胸外按压
(h) Press_chest

(i) 人工呼吸
(i) Artificial_respiration

(j) 判断结果
(j) Evaluate_result

图 4　CPR 姿态估计数据示例

Fig. 4　Sample data of CPR pose estimation dataset

构建好的数据集囊括常见的使用场景,包含多个角度,考虑了现实考核中会出现的所有情况,并且严格按照心肺复苏标准规定的动作进行拍摄采集,能够满足本文实验的验证需要。

3.3　OCHuman 数据集

OCHuman 数据集是一个专注于极端遮挡场景下的

人体姿态估计数据集。该数据集共有 5 081 张图像，13 360 个人体标注实例，每个实例均提供边界框与 17 个关键点的标注。与常见的人体姿态估计数据集相比，OCHuman 数据集的显著特点在于其极高的遮挡比例，大多数实例的重叠程度超过 0.5，平均遮挡强度接近 0.57。这种复杂场景有效反映了多目标密集分布及强遮挡情况下的识别难点，因此 OCHuman 数据集被广泛用作遮挡鲁棒性评估的标准数据集，对检测与姿态估计算法的泛化能力提出了更高要求。本文采用按 8∶2 的比例对其随机划分训练集和验证集。

表 1 CPR 姿态估计数据集组成

Table 1 Composition of the CPR pose estimation dataset

类名	标签名	标签数量
评估环境	evaluate_environment	184
判断意识	judge_consciousness	109
判断呼吸	judge_breathing	253
呼救	call_help	97
摆体位	put_posture	86
开放气道	open_airway	303
去除分泌物	remove_obstruction	246
胸外按压	press_chest	1 208
人工呼吸	artificial_respiration	375
判断结果	evaluate_result	271
其他	others	1 407
假人衣物未打开	dummy	2 366
假人衣物打开	dummy_open	2 173

4 实验结果

4.1 评价指标

为了全面评估模型的性能，本文采用了以下几个评价指标。分别是召回率(recall)、阈值为 50 时的平均精度均值(mAP)、模型参数量(Params)、浮点运算次数(GFLOPs)和每秒传输帧数(FPS)。其中，平均精度均值包括了基于边界框(box)的行为识别精度(mAP(B))和基于关键点(pose)姿态估计精度(mAP(P))。

召回率为真实值为正样本中，预测为正样本的概率，公式如式(5)所示。

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

式中：TP 为真阳性；FN 为真阴性。

平均精度均值由召回率与准确度(precision)决定，准确度为真实值为正样本占所有预测值为正样本的比例，准确度和平均精度均值的公式如式(6)所示。

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

$$mAP = \frac{1}{C}\sum_{k=0}^{C}\int_{0}^{1} P(r)\,\mathrm{d}r \tag{6}$$

式中：FP 假阳性；$P(r)$ 为召回率为横坐标，准确度为纵坐标的曲线；C 为类别数量。

浮点运算次数为每秒 10 亿次的浮点运算数，用于衡量算法模型的计算量。模型参数量表示型中所有可学习参数的总和，决定了模型的复杂度和容量。FPS 表示模型每秒能够完成推理处理的图像帧数，是评估算法在实际设备上运行速度与实时性能的核心指标。

4.2 CPR 姿态估计实验结果

为验证本文算法在心肺复苏特定场景下的有效性，本文使用 YOLOv13n、YOLOv12n、YOLOv11n、YOLOv10n、HighterHRNet、DEKR 和 RTMPose[27] 这些目前主流的自顶向下和自底向上的姿态估计算法对本文构建的姿态估计数据集进行对比实验，实验结果如表 2 所示。

表 2 CPR 姿态估计数据集对比实验结果

Table 2 Comparative experimental results on CPR pose estimation dataset

模型	Recall/%	mAP(P)/%	Params/($\times 10^6$)	GFLOPs	帧率/fps
本文	84.8	88.5	2.5	7.3	180.1
YOLOv13n	81.5	84.9	2.7	7.3	177.0
YOLOv12n	83.8	87.0	2.8	6.9	169.4
YOLOv11n	86.9	87.6	2.9	7.4	187.6
YOLOv10n	83.5	87.6	2.9	8.9	193.4
HighterHRNet	49.7	32.6	28.6	74.8	58.2
DEKR	78.1	47.3	29.6	45.4	40.4
RTMPose	88.7	86.0	3.3	0.4	130.0

从表 2 可以看出，本文提出的改进模型在综合性能上展现出显著优势。在检测精度方面，本文算法的姿态估计 mAP(P)达到了最高的 88.5%。在模型复杂度与实时性方面，得益于对主干网络的轻量化重构，本文模型的参数量仅为 2.5×10^6，是所有对比算法中最低的。同时，模型的推理速度高达 180.1 fps，虽然略低于单纯追求速度的 YOLOv10n，但远超 HigherHRNet 与 DEKR，且完全满足急救考核场景下对 30 fps 实时视频流的处理需求。虽然 RTMPose 计算量极低，但其在 CPR 特定数据集上的 mAP(P)和 FPS 均低于本文算法。相比于计算量极大的 HigherHRNet，本文算法以仅 7.3 GFLOPs 的计算成本实现了更高的精度，验证了其在计算资源受限的嵌入式设备上部署的可行性与高效性。

为了验证本文提出算法的检测效果，特别选取了相同图像在对比算法和改进算法进行推理检测，结果如图 5 所示。可以看到在行为识别方面原始 YOLOv13n 算法和本文提出的算法在行为识别方面的检测效果一致，在姿态估计方面，原始 YOLOv13n 对于假人的关节点检测出现了严重的误差，而本文提出的改进模型的检测效果良好。

图 5　CPR 姿态估计可视化对比

Fig. 5　Visualization comparison of CPR pose estimation

4.3　OCHuman 实验结果

本文在公开数据集 OCHuman[28] 上与其他姿态估计算法进行对比实验。OCHuman 数据集是专门为复杂场景下的人体姿态估计任务设计的标注数据集,共有 5 081 张图像。本文同样按 8∶2 的比例随机划分数据集,实验结果如表 3 所示。实验结果表明本文提出的算法在 OCHuman 数据集的结果优于其他算法,具有较好的泛化能力。

表 3　OCHuman 数据集对比实验结果

Table 3　Comparative experimental results on OCHuman dataset

模型	Recall/%	mAP(P)/%	Params/($\times10^6$)	GFLOPs	帧率/fps
本文	61.4	56.4	2.5	7.3	180.1
YOLOv13n	59.8	55.0	2.7	7.3	177.0
YOLOv12n	61.4	55.9	2.8	6.9	169.4
YOLOv11n	60.4	56.0	2.9	7.4	187.6
YOLOv10n	62.2	55.7	2.9	8.9	193.4
HighterHRNet	30.7	54.1	28.6	74.8	58.2
DEKR	62.9	38.4	29.6	45.4	40.4
RTMPose	64.0	53.6	3.3	0.4	130.0

特别选取相同图像在对比算法和改进算法进行推理检测,结果如图 6 所示。可以看原始 YOLOv13n 和其他算法对于被遮挡情况下的人体关键点相较于本文提出的改进模型存在着识别不到,精度不高的情况,可以看出本文提出的改进模型具有较好的泛用性。

4.4　消融实验

为验证本文设计的双层路由注意力机制(BRA)与混合增强上采样模块(HEUpsample)对算法性能的提升效果及其实时性表现,在相同的软硬件环境下进行了消融对比实验,结果如表 4 所示。首先,实验 A 在 YOLOv13n 基准模型的基础上引入 BRA 模块替换主干网络深层的 A2C2f 结构。实验结果表明,该改进使得姿态估计的平均准确率 mAP(P)从 84.9%提升至 87.2%,同时行为识别准确率也有所增长。值得注意的是,得益于 BRA 机制的动态稀疏路由策略,网络能够自适应过滤 CPR 场景中的冗余背景特征,这使得模型参数量从 2.7×10^6 降低至 2.4×10^6,计算量从 7.3 GFLOPs 降至 6.6 GFLOPs,推理速度更是从 177.0 fps 显著提升至

图 6 OCHuman 可视化对比

Fig. 6 Visualization comparison of OCHuman

190.2 fps,验证了 BRA 在解决肢体遮挡问题的同时有效减轻了计算负担。

其次,实验 B 在基准模型上单独引入 HEUpsample 替换颈部上采样模块。结果表明,该模块对关键点定位精度的提升尤为明显,mAP(P)达到 87.8%,mAP(B)提升至 95.2%。虽然由于 HEUpsample 引入了通道重排与空间偏移计算,导致模型参数量微增至 2.8×10^6,帧率轻微下降至 164.6 fps,但该速度仍远超实时检测需求,且精度的显著提升证明了其在恢复手臂、手腕等细微关节点边缘细节方面的必要性。

最后,本文提出的改进模型同时融合了 BRA 与 HEUpsample 模块。实验结果显示,该模型实现了最佳的综合性能,行为识别 mAP(B)达到 95.8%,姿态估计 mAP(P)高达 88.5%,较基准模型分别提升了 2.3% 和 3.6%。更为关键的是,BRA 模块带来的轻量化优势有效抵消了 HEUpsample 增加的计算开销,使得最终模型的参数量低于原始 YOLOv13n,且推理速度达到 180.1 fps,反而优于原始模型的 177.0 fps。这表明本文提出的算法并非单纯的模块堆砌,而是通过架构的互补优化,在大幅提升 CPR 检测精度的同时实现了更高效的推理,完美满足了急救考核系统对高精度与低延迟的双重严苛要求。

表 4 消融实验结果

Table 4 Results of ablation test

	BRA	HEUpsample	Recall(B)/%	mAP(B)/%	Recall(P)/%	mAP(P)/%	Params/($\times10^6$)	GFLOPs	帧率/fps
Baseline			87.9	93.5	81.5	84.9	2.7	7.3	177.0
A	√		93.0	95.0	86.4	87.2	2.4	6.6	190.2
B		√	89.2	95.2	82.5	87.8	2.8	8.0	164.6
本文	√	√	90.5	95.8	84.8	88.5	2.5	7.3	180.1

为了进一步验证本文选用的双层路由注意力机制(BRA)与混合增强上采样模块(HEUpsample)在心肺复苏特定场景下的性能优势,本文设计了横向对比实验。在保持 YOLOv13n 基础架构不变的前提下,分别将本文提出的模块与当前主流的同类改进算子进行替换对比,重点考察模型在精度、参数量与实时性之间的平衡能力,实验结果如表 5 所示。

在注意力机制方面,针对主干网络的特征提取优化,实验将 BRA 模块与主流的 CBAM(convolutional block attention module)[29] 进行了对比。实验数据表明,YOLOv13n 引入 CBAM 后,虽然帧率提升至 197.1 fps,且姿态估计 mAP(P)微增至 85.2%,但其参数量有所增加。

相比之下,本文使用的 BRA 机制优势更为显著,由于动态稀疏路由策略,BRA 能够更精准地捕捉因肢体遮挡产生的长距离依赖,使得 mAP(P)大幅提升至 87.2%。此外,BRA 模块有效实现了网络的轻量化重构,模型参数量降至 2.4×10^6,计算量降至 6.6 GFLOPs。尽管 BRA 的 FPS 略低于 CBAM,但仍远超实时检测需求,且在精度与模型轻量化方面表现出明显的压倒性优势。

在上采样模块方面,针对颈部网络的特征融合优化,实验对比了本文提出的 HEUpsample 与常用的 CARAFE(content-aware ReAssembly of FEatures)[30]模块。CARAFE 算子凭借较大的感受野,将 mAP(P)提升至 86.9%,但其代价是模型参数量增加至 2.9×10^6。而本文设计的 HEUpsample 模块通过高效的通道重排与空间偏移策略,在参数量仅为 2.8×10^6 的情况下,实现了更高的检测精度。同时,在实时性方面,两者差距微小。这表明 HEUpsample 在保证推理速度几乎不受影响的前提下,通过更优的特征重建能力,实现了比 CARAFE 更高的姿态估计精度与更低的空间复杂度。

表 5　不同模块性能评估对比

Table 5　Comparative evaluation of different modules

	模型	mAP(B)/%	mAP(P)/%	Params/($\times10^6$)	GFLOPs	帧率/fps
Baseline	YOLOv13n	93.5	84.9	2.7	7.3	177.0
注意力机制	YOLOv13n + CBAM	94.8	85.2	2.8	7.3	197.1
	YOLOv13n + BRA	95.0	87.2	2.4	6.6	190.2
上采样	YOLOv13n + CARAFE	95.1	86.9	2.9	7.5	166.8
	YOLOv13n + HEUpsample	95.2	87.8	2.8	8.0	164.6

4.5　工程部署可行性分析

基于本文模型 2.5×10^6 的超低参数量和 7.3 GFLOPs 的计算量,该算法完全满足在嵌入式设备上的实时运行要求。在实际 CPR 考核场景中,可以将该模型部署于移动端 App 或考场监控终端,配合简单的单目摄像头即可实时捕捉学员动作。系统依据识别到的关键点坐标计算大臂与地面的角度,一旦检测到角度小于 90°或频率异常,即可通过语音实时纠正,从而以低成本替代昂贵的传感器模拟人设备。

5　结　论

针对心肺复苏考核中对动作识别高精度与设备低算力的双重挑战,本文提出了一种基于改进 YOLOv13n-Pose 的姿态估计算法。通过集成双层路由注意力机制(BRA),模型有效克服了 CPR 操作中频繁出现的肢体互遮挡难题;配合混合增强上采样模块(HEUpsample),显著改善了关节点边缘特征的恢复质量。实验表明,改进算法在自建 CPR 数据集及 OCHuman 数据集上均取得了优于 YOLOv13n、HigherHRNet 等主流模型的综合性能,在实现 88.5%姿态估计精度的同时,仅占用 2.5×10^6 参数量,推理速度高达 180.1 fps。本研究不仅验证了算法的有效性,更证明了其在移动终端及边缘计算设备上部署的可行性。该算法可作为核心视觉引擎嵌入急救培训模拟人或智能考核 APP 中,实现对按压频率、深度及姿势规范性的实时反馈。未来的工作将致力于结合时序动作分析网络,进一步实现对 CPR 全流程操作质量的连续性动态评估。

参考文献

[1] NIELSEN A M, ISBYE D L, LIPPERT F K, et al. Quality of cardiopulmonary resuscitation performed by trained bystanders: A systematic review [J]. Resuscitation, 2013, 84(12): 1591-1598.

[2] CHENG A, BROWN L L, DUFF J P, et al. Improving cardiopulmonary resuscitation with a CPR feedback device and refresher simulations (CPR CARES study): A randomized clinical trial[J]. JAMA Pediatrics, 2015, 169(2): 137-144.

[3] YANG J. A review of AI-based technologies for CPR training and evaluation [J]. Artificial Intelligence in Medicine, 2024, 150: 102592.

[4] 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, 2017: 7291-7299.

[5] CHENG B, XIAO B, 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, 2020: 5386-5395.

[6] GENG Z, SUN K, XIAO B, et al. Bottom-up human pose estimation via disentangled keypoint regression[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 14676-14686.

[7] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.

[8] FANG H S, XIE S, TAI Y W, et al. RMPE: Regional multi-person pose estimation[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2334-2343.

[9] 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, 2022: 2637-2646.

[10] JOCHER G, CHAURASIA A, STOKEN A, et al. Ultralytics/YOLOv5: v6.2-YOLOv5 classification models, apple m1, reproducibility, clearml and deci.ai integrations[J]. Zenodo, 2022, DOI: 10.5281/zenodo.7002879.

[11] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. ArXiv preprint arXiv:2209.02976, 2022.

[12] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. 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, 2023: 7464-7475.

[13] WANG C Y, YEH I H, LIAO H Y M. YOLOv9: Learning what you want to learn using programmable gradient information[C]. European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 1-21.

[14] WANG A, CHEN H, LIU L, et al. YOLOv10: Real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.

[15] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements[J]. ArXiv preprint arXiv:2410.17725, 2024.

[16] TIAN Y J, YE Q X, DOERMANN D S. YOLOv12: Attention-centric real-time object detectors[J]. ArXiv preprint arXiv:2502.12524, 2025.

[17] LEI M Q, LI S Q, WU Y H, et al. YOLOv13: Real-time object detection with hypergraph-enhanced adaptive visual perception[J]. ArXiv preprint arXiv: 2506.17733, 2025.

[18] 范莹莹. 基于多目摄像机的心肺复苏人体姿态估计和评价方法的分析与研究[D]. 南昌:南昌大学, 2022.
FAN Y Y. Analysis and research on CPR posture estimation and evaluation method based on multi-camera system[D]. Nanchang: Nanchang University, 2022.

[19] 靳杰. 基于 OpenPose 的智能胸外按压培训系统的设计与实现[D]. 天津:天津大学, 2022.
JIN J. Design and implementation of an intelligent chest compression training system based on OpenPose[D]. Tianjin: Tianjin University, 2022.

[20] 陈姝琪, 曹江涛, 赵挺, 等. 基于关节点数据的双人交互行为识别[J]. 电子测量与仪器学报, 2020, 34(6):124-130.
CHEN SH Q, CAO J T, ZHAO T, et al. Two-person interaction recognition based on joint point data[J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(6): 124-130.

[21] 张敏. 基于轻量化特征保持骨架生成的心肺复苏动作评估应用研究[D]. 南昌:南昌大学, 2024.
ZHANG M. Research on CPR action assessment application based on lightweight feature-preserving skeleton generation[D]. Nanchang: Nanchang University, 2024.

[22] 张斯琪. 基于 OpenPose 的人工呼吸动作判别及其培训系统的研究[D]. 天津:天津大学, 2023.
ZHANG S Q. Research on artificial respiration action recognition and its training system based on OpenPose[D]. Tianjin: Tianjin University, 2023.

[23] 何赟泽, 谯灵俊, 王洪金, 等. 基于改进 DETR 的智慧车间人员典型行为识别算法[J]. 电子测量与仪器学报, 2024, 38(9): 76-84.
HE Y Z, QIAO L J, WANG H J, et al. A typical behavior recognition algorithm for smart workshop personnel based on an improved DETR[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(9): 76-84.

[24] 任丹彤, 何赟泽, 刘贤金, 等. 面向智慧工厂的双光融合车间人员行为识别方法[J]. 测控技术, 2022, 41(8): 9-15.
REN D T, HE Y Z, LIU X J, et al. Dual-light fusion based workshop personnel behavior recognition method for smart factories[J]. Measurement & Control Technology, 2022, 41(8): 9-15.

[25] ZHU L, WANG X, KE Z, et al. BiFormer: Vision transformer with bi-level routing attention[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 10323-10333.

[26] REN S, ZHOU D, HE S, et al. Shunted self-attention via multi-scale token aggregation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 10853-10862.

[27] JIANG T, LU P, ZHANG L, et al. RTMPose: Real-time multi-person pose estimation based on MMPose[J]. ArXiv preprint arXiv:2303.07399, 2023.

[28] ZHANG S H, LI R, DONG X, et al. Pose2Seg: Detection free human instance segmentation [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 889-898.

[29] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision, 2018: 3-19.

[30] WANG J, CHEN K, XU R, et al. CARAFE: Content-aware reassembly of features [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3007-3016.

作者简介

谢天宇,2024 年于河海大学获得学士学位,现为湖南大学硕士研究生,主要研究方向为图像处理、机器视觉。

E-mail: s240900697@ hnu. edu. cn

Xie Tianyu received his B. Sc. degree from Hohai University in 2024. Now he is a M. Sc. candidate of Hunan University. His main research interests include image processing and machine vision.

何赟泽(通信作者),2012 年于国防科学技术大学获得博士学位,现为湖南大学教授,主要研究方向为嵌入式人工智能与边缘计算、红外热成像与机器视觉。

E-mail: hejicker@ 163. com

He Yunze (Corresponding author) received his Ph. D. degree from the University of Defense Science and Technology in 2012. Now he is a professor at Hunan University. His main research interests include embedded artificial intelligence and edge computing, infrared thermal imaging and machine vision.