

文章编号 1004-924X(2025)14-2291-12

多分辨率特征融合的人体下肢关键点检测

夏晓华*, 向浩鸣, 陈 坚, 冯鑫淼, 邱法博, 王耀耀

(长安大学 陕西省高速公路施工机械重点实验室, 陕西 西安 710064)

摘要: 现有的人体关键点检测模型缺乏对高分辨率特征的关注, 训练中使用的数据集分辨率较低且标注误差较大, 导致其在步态分析等任务中检测结果不稳定, 定位精度较差。针对上述问题, 提出一种基于多分辨率特征融合的人体下肢关键点检测模型, 以高分辨率图像作为网络输入, 利用微调的 MobiliNet v1 网络结合注意力机制提取全局低分辨率特征, 初步预测关键点位置, 通过与之并行的浅层网络提取局部高分辨率特征, 然后采用连续残差结构与注意力机制将不同分辨率特征融合, 提升预测关键点的准确性, 并有效缓解高分辨率图像带来的高计算量问题。通过预标记的方式制作高分辨率、高精度的人体下肢关键点数据集以确保模型训练的准确性。以模型复杂度、检测速度、检测精确率以及检测误差为评价指标, 与其他经典和先进的方法进行实验对比。结果表明, 所提模型的测试检测率达到 95.2%, 均优于 Lightweight-OpenPose, HRNet-W32, HRNet-W48, YOLO-Pose, RTMPose 和 SimCC 模型, 检测精确率提升了 4.1%~83.6%, FPS 提升了 7.6~13.9。证明了提出的模型在高精度人体下肢关键点检测中的有效性。

关键词: 人体下肢关键点检测; 多分辨率特征融合; 注意力机制; 预标记

中图分类号: TP394.1 **文献标识码:** A

doi: 10.37188/OPE.20253314.2291

CSTR: 32169.14.OPE.20253314.2291

Human lower limb keypoint detection based on multi-resolution feature fusion

XIA Xiaohua*, XIANG Haoming, CHEN Jian, FENG Xinmiao, QIU Fabo, WANG Yaoyao

(Key Laboratory of Expressway Construction Machinery of Shaanxi Province,
Chang'an University, Xi'an 710064, China)

* Corresponding author, E-mail: xhxia@chd.edu.cn

Abstract: The existing human keypoint detection models lack attention to high-resolution features, and the dataset used in training has low resolution and large annotation errors. This causes an unstable detection result and poor positioning accuracy in tasks such as gait analysis. A human lower-limb keypoint-detection model based on multi-resolution feature fusion was proposed to address the above issues. High-resolution images were adopted as network inputs, and a fine-tuned MobileNet v1 network combined with an attention mechanism was employed to extract global low-resolution features. Preliminary keypoint positions were predicted, after which local high-resolution features were extracted through a parallel shallow network. Subsequently, features at different resolutions were fused via a continuous residual structure and an attention mechanism, thereby improving keypoint-prediction accuracy and effectively alleviating the

收稿日期: 2025-03-12; 修订日期: 2025-06-18.

基金项目: 国家自然科学基金 (No. 61901056); 陕西省重点研发计划项目 (No. 2025CY-YBXM-545)

computational burden imposed by high-resolution images. In this work, a high-resolution, high-precision dataset of human lower-limb keypoints was created through pre-labeling to ensure reliable model training. Model complexity, detection speed, detection accuracy, and detection error were evaluated and compared with other classic and state-of-the-art methods through experiments. The results show that the test detection rate of the proposed model reaches 95.2%, which is better than the comparison methods of Lightweight-OpenPose, HRNet-W32, HRNet-W48, YOLO-Pose, RTMPose and SimCC, and the detection accuracy is increased by 4.1%-83.6%, and the FPS is increased by 7.6-13.9. The effectiveness of the proposed model in high-precision human lower limb key point detection is demonstrated.

Key words: human lower limb keypoint detection; multi-resolution feature fusion; attention mechanism; pre labeling

1 引言

人体关键点检测是计算机视觉领域中一项基础且具有挑战性的任务,旨在识别并定位图像或视频中人体的关键点位置,比如髋关节和膝关节等身体部位^[1],在视频监控、智能交互等领域有着广泛的应用^[2-4],也是人体步态分析的主要方法之一^[5]。

利用人体关键点检测模型提取关键点坐标,并根据坐标计算相应的步态参数,从而实现人体步态特征的量化评估,可以为全膝关节置换、中风后偏瘫等患者的诊断和治疗提供客观有效的理论支持^[6-7],让医生快速、准确地掌握患者的术后康复情况,完善治疗方案^[8]。Ota 等人^[9]和 Diaz-Arancibia 等人^[10]使用由 Cao^[11]提出的人体关键点检测模型 OpenPose,实现了在人体行走过程中计算膝关节角度和踝关节角度。Stenum 等人^[12]采用基于 OpenPose 的方法测量人体的时空和运动学步态参数,并通过实验验证了人体关键点检测算法在测量人类步态数据方面的可行性与发展潜力。Zhang 等人^[13]使用由 Sun^[14]提出的 HRNet 网络构建关键点自动检测模型,实现了人体下肢的关键角度的自动测量,其结果与医生的测量结果比较吻合。

将人体关键点检测模型应用在人体步态分析领域时,尽管检测过程具有非侵入性、智能化程度高等优点,但是其检测结果存在不稳定、有效性较差的问题^[15-16]。利用该类模型检测到的人体关键点不够准确,导致计算得到的数据不平滑,存在不少的跳变数据点,而这种跳变数据点在滤波、插值等后处理之后依然与原始数据存在

一定偏差^[17-18]。究其原因主要为两方面,一方面,现有模型训练时所用到的公开数据集采用的图像规格不统一,其中部分图像人体占比较小,且采用人工标注,关键点选取具有较强的主观性,导致标注不够准确;另一方面,现有模型通常使用分辨率较低的图像作为网络输入,虽然保证了网络运行速度,但是丢失了高分辨率的特征,致使精度下降。

为解决人体关键点检测模型在实际应用中存在的上述问题,本文提出一种以检测高精度的人体下肢关键点为目标,适用于高分辨率图像的人体下肢关键点检测模型。本文的主要贡献为:

(1) 将高分辨率图像作为网络输入,提取局部高分辨率特征,为关键点预测提供准确的特征信息。

(2) 采用融合不同分辨率特征的方式,使模型兼顾精度与速度,缓解了高分辨率图像为了获得全局感受野需要进行深层卷积与池化带来的高计算量问题。

(3) 利用注意力机制增强不同通道特征之间的信息交互,提高模型的准确性。

(4) 通过预标记的方式,制作高精度、高分辨率的人体下肢关键点数据集,为网络训练提供可靠的数据来源。

2 模型设计

在神经网络中,高分辨率的图像中蕴含着更加丰富的特征信息,对位置的敏感性更强,有利于检测任务;而低分辨率的图像可以使模型拥有较小的计算量,更加轻量化,且相较高分辨率图

像,利用低分辨率图像得到的特征拥有更大的感受野^[19]。为了使模型不仅能够有效利用图像中的高分辨率信息,还具有好的检测速度,以较小的计算资源得到高精度的检测结果,本文提出的人体下肢关键点检测模型采用两个分支分别提取图像的高分辨率、低分辨率两种特征。本文提出了一种具有双分支结构的下肢关键点检测模型,其网络结构如图1所示。

模型的第一个分支以低分辨率图像作为输入,提取具有全局感受野的低分辨率特征,通过浅层网络结构在较低计算资源下预测目标点位置(创新性地缓解了传统高分辨率图像为获取全局感受野所需的深层卷积与池化带来的高计算量问题)。其中全局感受野使模型理解关键点的具体结构和空间关系,有利于区分人体左、右两侧相似的关键点;而第二个分支采用更轻量的浅层网络,从高分辨率图像中提取局部细节特征,在保持实时性的同时获取亚像素级精度的位置信息;最终通过双分辨率特征融合机制,将低分辨率分支的空间关系预测结果与高分辨率特征

的位置细节相结合,利用后者高精度信息对前者进行动态优化调整,实现兼顾速度与精度的关键点定位。

在低分辨率特征提取分支,首先将输入的高分辨率图像采用平均池化的方式降采样至原来的1/4大小,得到输入图像所对应的低分辨率图像。进而选用轻量化网络 MobileNet v1^[20]的前23层作为特征提取器,并对其结构进行轻微调整,如表1所示,用于提取低分辨率特征。然后将提取到的低分辨率特征上采样4倍恢复分辨率,并在特征降维后接入高效通道注意力机制模块(Efficient Channel Attention, ECA)^[21],最终生成低分辨率图像对应的关键点热图。

在高分辨率特征提取分支,采用低参数数量和计算量的深度可分离卷积层提取局部范围内的高分辨率特征,并将其与低分辨率热图对齐。在分支的起始与终止位置,利用残差网络模块增强网络的非线性表达能力和鲁棒性,残差网络模块由 $1\times 1, 3\times 3, 3\times 3$ 三层卷积构成,第一层的输出与第三层的输出跳变连接,其中第三层采用膨胀

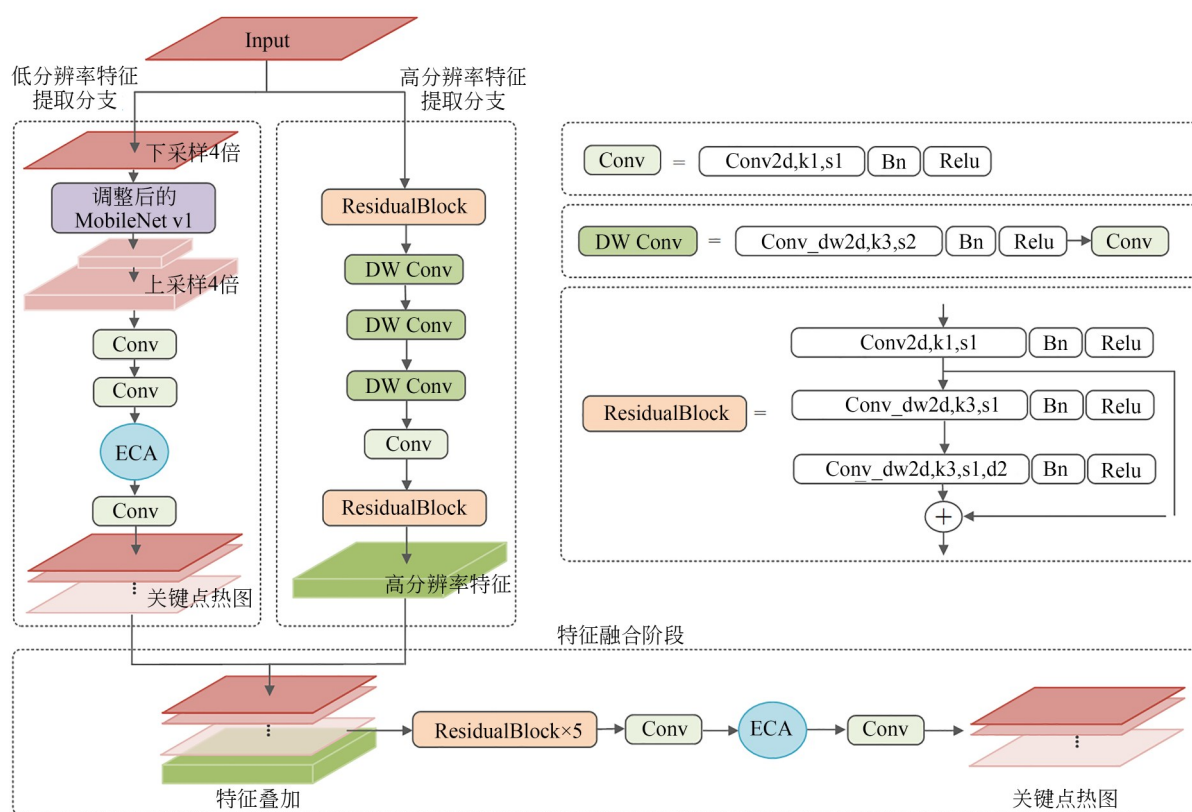


图1 网络结构图

Fig. 1 Network structure diagram

表 1 调整后的 MobileNet v1 网络结构

Tab. 1 Adjusted MobileNet v1 network structure

Layers	输入	Operator	输出通道数	步长
1	$224^2 \times 3$	Conv, 3×3	32	2
2	$112^2 \times 32$	Conv_dw, 3×3	32	1
3	$112^2 \times 32$	Conv, 1×1	64	1
4	$112^2 \times 64$	Conv_dw, 3×3	64	2
5	$56^2 \times 64$	Conv, 1×1	128	1
6	$56^2 \times 128$	Conv_dw, 3×3	128	1
7	$56^2 \times 128$	Conv, 1×1	128	1
8	$56^2 \times 128$	Conv_dw, 3×3	128	2
9	$28^2 \times 128$	Conv, 1×1	256	1
10	$28^2 \times 256$	Conv_dw, 3×3	256	1
11	$28^2 \times 256$	Conv, 1×1	256	1
12	$28^2 \times 256$	Conv_dw, 3×3	256	1
13	$28^2 \times 256$	Conv, 1×1	512	1
14~23	$28^2 \times 512$	Conv_dw, 3×3	512	1
	$28^2 \times 512$	Conv, 1×1	512	1

值为 2 的膨胀卷积,用于增大特征的感受野。

在特征融合阶段,将低分辨率图像的热图与高分辨率特征通过五个残差网络模块进行融合,在特征降维后同样接入 ECA 注意力机制模块,并生成最终的关键点热图。

本文网络中使用的 ECA 模块是一种用以捕捉局部跨通道交互信息的轻量注意力机制模块,通过自适应权重分配增强关键通道的特征响应,从而提升模型对下肢关键点(如髋关节、膝关节)细微纹理差异的敏感性,创新性地解决了传统方法在复杂人体遮挡场景下的特征混淆问题。如图 2^[21]所示,在低分辨率特征提取阶段与特征融合阶段中,将特征通道数降为 128 通道后接入 ECA 模块,利用其跨通道交互机制强化与解剖结构强相关的通道特征,例如在骨盆区域检测时同步增强骨骼边缘与软组织对比度的特征表达,最终通过通道级特征校准提升热图预测的准确性。

由于标注的数据为坐标信息,而网络输出结果为关键点热图,所以需要将标注数据转化为 q 张 $m \times n$ 的目标热图, q 是关键点个数, m 和 n 为热

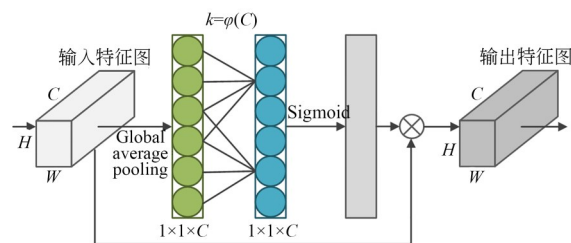


图 2 ECA 模块

Fig. 2 ECA module

图大小,与网络输出热图保持一致。在构建目标热图的过程中,标注点附近的像素点与标注点相似度较高,如果直接将其标为负样本,会给网络训练带来干扰,不利于网络收敛。通常以标注点为中心构建高斯分布,距离目标点越近,激活值越大,可以为网络的训练增加一个方向性的引导。对于某一个标注点 (x_i, y_i) ,其对应的目标热图上一 (x, y) 上的值为:

$$h_i(x, y) = e^{-\frac{(x-x_i)^2 + (y-y_i)^2}{2\sigma^2}}, \quad (1)$$

其中: σ 是高斯分布的标准差,取值与热图大小成正比。

本文采用均方误差 (Mean Square Error, MSE)^[22]作为损失函数, MSE 函数是凸函数, 具有全局最小值, 且将误差平方, 放大了较大误差的影响, 能够加快收敛速度。其计算公式为:

$$L_{\text{MSE}} = \frac{\sum_{i=1}^q \sum_{y=1}^m \sum_{x=1}^n (h_i(x, y) - \hat{h}_i(x, y))^2}{q \cdot m \cdot n}. \quad (2)$$

3 数据集与数据增强

为了尽可能地保证数据集的丰富性, 降低背景等因素的影响, 以提高关键点检测的准确性, 本文在数十种不同背景下拍摄人体下肢图像构建数据集, 人体与相机之间呈多种角度, 以及被拍摄者处于站立或行走两种状态。数据集的图像为 800×1280 分辨率的 RGB 图像。

由于公开数据集在标注过程中, 采用多人标注, 标注关键点的位置取决于标注者的主观

选择, 对于关键点位置的判断具有很强的随机性, 不同人标注的关键点位置存在差异, 且同一个人标注不同图像的同一个关键点, 其标注结果也存在位置偏差^[23]。本文采用预先做标记的方式, 提高数据集标注的客观性。在拍摄图像之前, 在被拍摄者的关键点处粘贴直径为 3 mm 的圆形标记物, 然后使用 Labelme 软件对图像中标记物的圆心进行标注。在标注完数据集后, 利用标注的坐标信息对原图像进行滤波处理以消除图像中的标记物信息, 以滤波后无标记物的图像和先前标注的坐标信息作为数据集训练模型, 防止预先标记的标记物对训练结果产生影响。

本文以人体左右的髋关节、膝关节、踝关节以及脚的中趾一共 8 个关键点为例, 构建了包含 1 665 张拍摄图像的人体下肢关键点数据集。数据集的部分图像及其标注信息如图 3 所示。

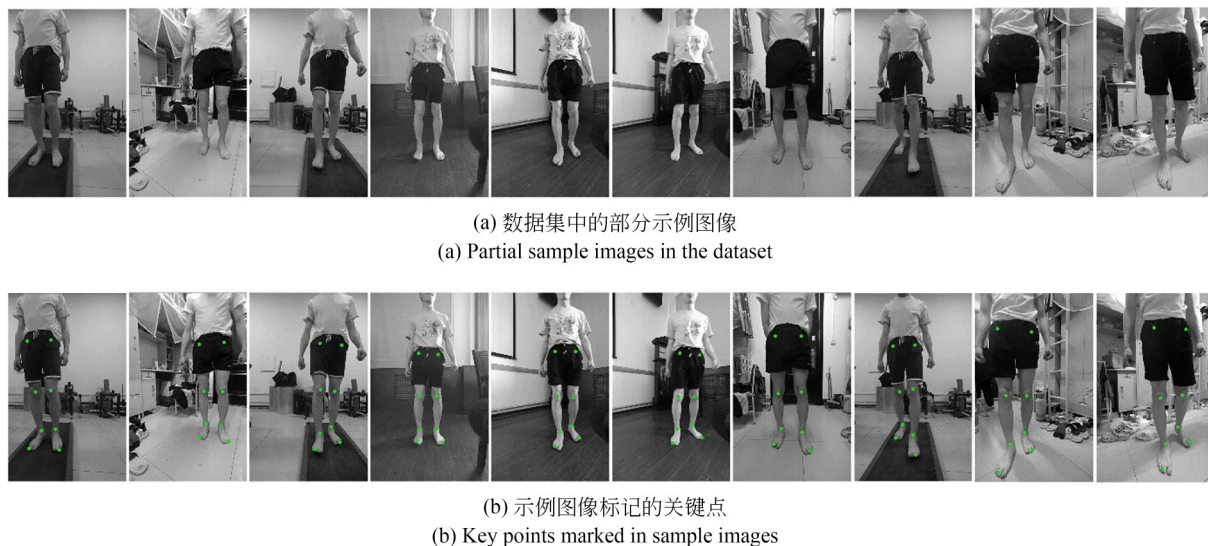


图 3 数据集的示例图像及其标注信息

Fig. 3 Sample images in the dataset and their annotation information

为了提高训练数据的多样性以改善模型的鲁棒性和泛化能力, 本文在训练深度学习模型时引入数据增强技术。采用改变图像色彩属性的数据增强方法, 在 $\pm 10\%$ 的范围内随机调整数据集中图像的亮度、对比度和饱和度, 生成具有不同属性的图像。通过这种增强方式模拟真

实世界中不同环境和条件下的图像变化, 使得模型能够更好地适应因光线等因素引起的图像变化。此外, 采用添加高斯随机噪声和椒盐随机噪声的方式扩充测试集, 其中高斯随机噪声的方差为 0.000 25, 随机生成 3 次; 椒盐随机噪声的密度为 0.3%, 随机生成 2 次, 最终将原始的

240 张测试集扩充至 1 440 张。经数据增强处理后,本文数据集共有 4 530 张图像,其中包含 2 730 张训练集,360 张验证集以及 1 440 张测试集。

4 实验结果与分析

4.1 评价指标与实验环境

目前最常用的人体关键点检测性能度量指标是 OKS(Object Keypoints Similarity)。公开数据集中,由于不同关键点的标注难度不同,标注误差也较大,OKS 对不同关键点的评价要求存在一定差异且相对较低。而本文制作的数据集中,不同关键点的标注难度相同,且标注精度高,为了准确评价模型在该数据集上的检测结果,本文选用以下评价指标:

(1) 检测率(Detection Rate, DR),即检测正确的关键点数量与关键点总数之间的比值,检测到的关键点和真值之间的欧氏距离小于设定阈值则认为检测正确。该指标用于描述模型检测正确的关键点的比例,其计算公式为:

$$DR = \frac{t}{N}, \quad (3)$$

其中: t 是检测到正确的关键点数, N 为关键点总数。

(2) 平均欧氏距离(Mean Euclidean Distance, MED)是检测到的关键点和真值之间欧氏距离的平均值,可以有效地反映预测值与真实值之间的误差,其计算公式为:

$$MED = \frac{1}{N} \sum_{i=1}^N d_i, \quad (4)$$

其中: d_i 是检测到的第 i 个关键点与真实值之间的欧氏距离。

(3) 均方根误差(Root Mean Square Error, RMSE),是检测到的关键点和真值之间的均方差距离,可以判断误差的分布情况,反映了模型的稳定性,其计算公式为:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (d_i - \bar{d}_i)^2}. \quad (5)$$

此外,还通过对比网络的参数量(Params)、浮点运算数(Floating Point Operations, FLOPs)

和每秒检测帧数(Frames Per Second, FPS)来衡量模型的复杂度与检测速度。

本文以 Windows 11 系统作为实验的基础平台,CPU 型号为 AMD Ryzen 7 4800H with Radeon Graphics,软件环境为 Python 3.6,Conda 11.7 和 Pytorch 1.10.0,在 NVIDIA GeForce GTX 1660 Ti 显卡上运行。

4.2 客观数据分析

将本文提出的模型和公开数据集中表现优异的 Lightweight-OpenPose 模型^[24]、HRNet-W32 模型^[14]、HRNet-W48 模型^[14]、YOLO-Pose 模型^[25]、RTMPose 模型^[26]和 SimCC 模型^[27]进行对比分析,由于公开数据集中标注的关键点不统一且误差较大,所以各个模型均采用本文制作的数据集进行训练并对比。考虑到各个模型原本的输入图像分辨率不同,为了更加有效地评价本文提出的模型的检测结果,设计两组对比试验,第一组将数据集中的图像调整至各个模型原论文中所采用的分辨率进行训练,第二组将所有模型都采用高分辨率图像进行训练,测试集上的实验结果如表 2 所示,其中计算检测率时的距离阈值设置为 5 像素。

4.2.1 直线图元的测量算法

由表 2 可知,现有的人体关键点检测模型采用原论文中的分辨率图像训练后,SimCC 模型在检测率 DR 指标下的效果最好,达到了 46.8%,HRNet-W48 模型则在平均欧氏距离 MED 与均方根误差 RMSE 指标下表现最佳,分别达到了 6.79 pixels 和 3.61 pixels,而本文模型的 DR 达到 95.2%,MED 与 RMSE 指标分别为 2.29 pixels 和 1.38 pixels,相较其他模型检测结果更加准确且稳定。在模型的复杂度与检测速度方面,本文模型的 Params 为所有模型中最小,仅 3.8 M,模型复杂度最低,FLOPs 与 FPS 分别为 45.1 G 和 17.2,模型的运算量较大,处理速度较慢,这与输入图像的分辨率最高有关,而采用与本文模型输入分辨率相近的图像进行训练的 YOLO-Pose 模型的 FLOPs 与 FPS 分别为 117.2 G 和 3.3,与之相比,本文模型的复杂度与检测速度具有明显优势。

表 2 各模型的测试结果对比
Tab. 2 Comparison of test results of each model

训练方式	Model	Input size	DR/%	MED/pixel	RMSE/pixel	Params/M	FLOPs/G	FPS
采用原论文 中的分辨率 图像训练	Lightweight-OpenPose	456×256	26.9	8.77	5.54	4.0	8.7	48.6
	HRNet-W32	256×192	33.3	6.89	3.79	28.5	7.7	27.7
	HRNet-W48	256×192	34.5	6.79	3.61	63.6	15.8	25.2
	YOLO-Pose	640×640	12.5	13.29	7.84	15.0	20.3	16.9
	RTMPose	256×192	28.2	13.89	17.91	8.9	6.7	41.2
	SimCC	256×192	46.8	7.13	8.32	64.1	5.5	21.5
采用高分辨 率图像训练	Lightweight-OpenPose	1 280×800	91.1	2.87	4.29	4.0	65.9	9.6
	HRNet-W32	1 280×800	78.8	3.64	5.45	28.5	160.5	8.5
	HRNet-W48	1 280×800	78.7	4.26	4.45	63.6	328.5	5.2
	YOLO-Pose	960×960	15.3	10.84	8.37	76.8	117.2	3.3
	RTMPose	1 280×800	34.5	12.52	18.63	8.9	56.7	9.4
	SimCC	1 280×800	86.3	2.93	3.69	64.1	115.5	4.7
	Ours	1 280×800	95.2	2.29	1.38	3.8	45.1	17.2

4.2.2 采用高分辨率图像训练

将所有模型都采用高分辨率图像进行训练,基于热图检测的 Lightweight-OpenPose 模型、HRNet-W32 模型和 HRNet-W48 模型的检测率 DR 分别为 91.1%,78.8% 和 78.7%,相较采用原论文中低分辨率图像训练得到的模型分别提高了 64.2%,44.3% 和 44.2%,MED 指标分别减小了 5.9 pixels,3.25 pixels,2.53 pixels,说明采用高分辨率图像训练模型能够有效提高模型检测结果的准确性。与上述模型不同,RTM-Pose 模型和 SimCC 模型将关键点定位视为分类任务,无需基于热图的模型所依赖的高分辨率特征图和复杂上采样层,避免了大量的计算损耗,模型参数量和计算量较低,但由于 RTMPose 模型聚焦于多人姿态的实时估计,相较于 SimCC 模型注重单人姿态的精准检测,其检测速度虽快,但精度有所下降。在所有对比模型中,YOLO-Pose 模型在训练原分辨率图像和高分辨率图像得到的模型效果在 DR 指标下都为最低值,这与其依赖描点匹配的方式有关,部分小目标会因锚点覆盖不足导致关键点定位出现偏差。与上述模型中检测效果最好的 Lightweight-OpenPose 模型相比,本文模型的检测率 DR 高出 4.1%,MED

和 RMSE 指标分别减少了 20.2% 和 67.8%,检测结果更加准确且稳定。在模型的复杂度与检测速度方面,本文模型以 3.8 M 参数量和 45.1 G FLOPs 实现 17.2 FPS,其计算复杂度为所有模型中最低,而其他模型最高仅 9.6 FPS,证明了在输入图像分辨率相同的条件下本文模型的检测速度明显优于其他模型。

4.3 自动聚焦高度测量消融实验分析

本文模型在结构上采用了融合不同分辨率特征的方法来提高模型的检测性能,同时使用了高效通道注意力机制模块 ECA 增加了不同通道之间的特征交互。通过消融实验研究这两种方法在模型训练中的作用。分别设置了四种方案对模型进行训练,方案一为去除网络结构中的高分辨率特征提取分支,仅利用低分辨率图像训练模型;方案二为去除网络结构中的高分辨率特征提取分支,并去除低分辨率分支中的上、下采样操作,仅利用高分辨率图像训练模型;方案三为去除 ECA 模块训练模型;方案四采用本文提出的完整的网络结构训练模型。各个方案下模型在测试集上的检测结果如表 3 所示。

对比表 3 中方案一、方案二和方案四这三种

表 3 消融实验结果

Tab. 3 Ablation test results

方案	DR/%	MED/pixel	RMSE/pixel	Params/M	FLOPs/G	FPS
方案一	42.9	6.75	4.99	3.3	28.2	36.8
方案二	88.5	2.54	4.83	3.3	54.5	14.8
方案三	93.3	2.48	1.59	3.8	45.1	17.1
方案四	95.4	2.28	1.37	3.8	45.1	17.1

方案下的模型检测结果可知,由于方案一仅利用低分辨率图像进行训练,缺乏高分辨率特征,检测率DR仅42.9%,MED和RMSE分别为6.75 pixels和4.99 pixels;方案二仅利用高分辨率图像进行训练,DR为88.5%,MED和RMSE分别为2.54 pixels和4.83 pixels,由于输入图像分辨率的提高,DR比方案一提高了45.6%;而使用完整模型训练的方案四的检测结果相较方案一和方案二检测率DR分别提高了52.5%和6.9%,MED分别降低了66.2%和10.2%,RMSE分别降低了72.5%和71.5%,而且采用的融合高、低

分辨率特征的方式,相较仅采用高分辨率图像训练的方案二,模型运算量FLOPs降低了17.2%,FPS由14.8提高至17.1。对比方案三和方案四的检测结果,ECA模块提高了2.1%的模型检测率,降低了检测误差,MED和RMSE分别降低了8.1%,13.8%,使模型的检测结果更加准确且稳定。

4.4 可视化分析

为了对不同模型的检测结果进行更加直观的分析,图4和图5展示了本文模型与其他几种模型针对测试集中经典图像的检测结果,第一

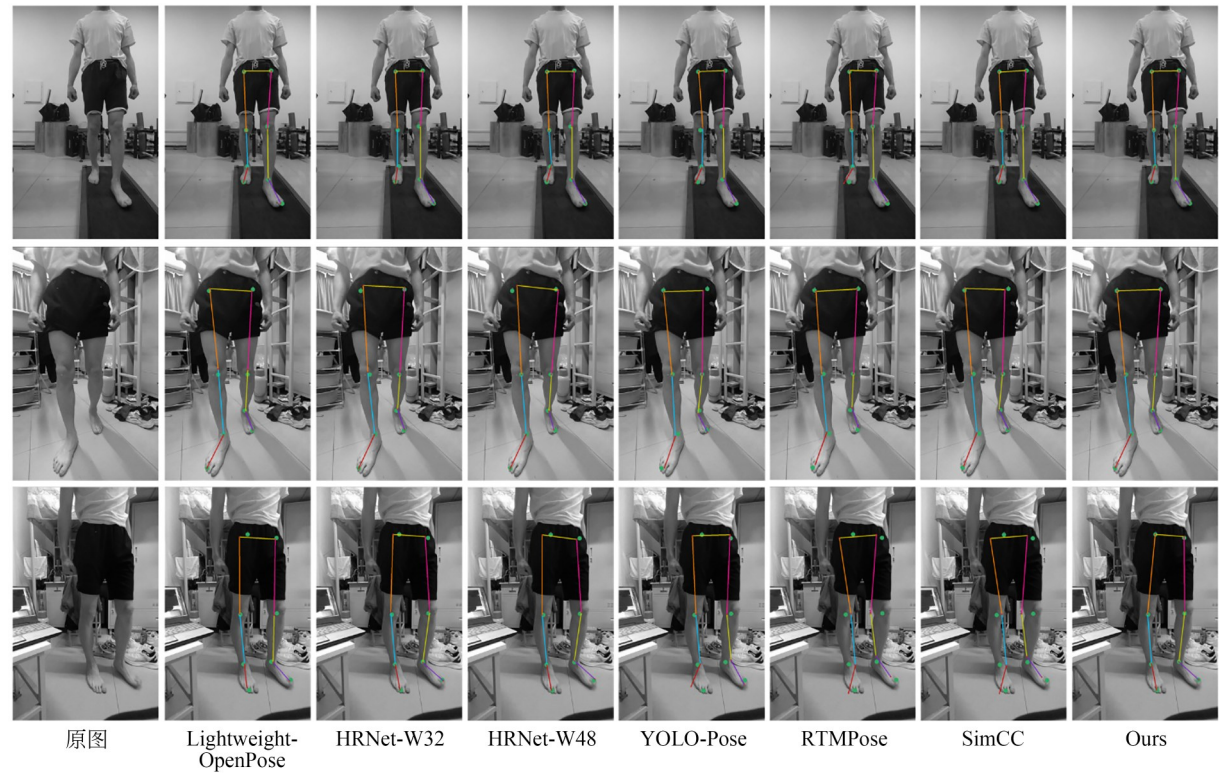


图 4 采用原论文中的分辨率图像训练的检测结果对比

Fig. 4 Comparison of the detection results of the resolution image training in the original paper

列为原图,其他各列为每种模型的检测结果。其中,图 4 为各个模型采用原论文中的分辨率图像训练之后的检测结果,图 5 为各个模型采用高分辨率图像训练之后的检测结果。为了直观展现检测结果,在可视化过程中将图像灰度

化,而标注信息及检测结果用彩色的点和线段表示。图中绿色的点为标注的关键点位置,红色的点以及连接红色点的彩色线段表示检测到的关键点位置及其连接方式(彩图见期刊电子版)。

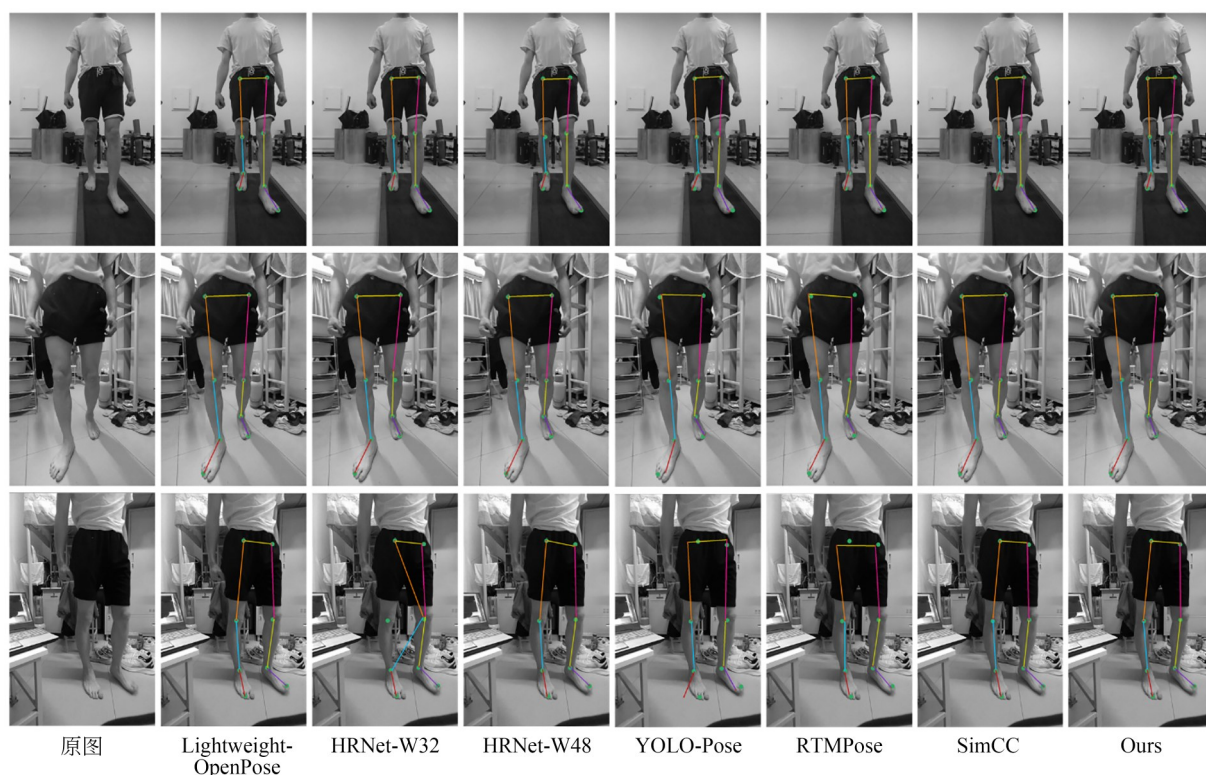


图 5 采用高分辨率图像训练的检测结果对比

Fig. 5 Comparison of detection results using high-resolution image training

由图 4 可知,采用低分辨率图像训练的 Lightweight-OpenPose, HRNet-W32 以及 HRNet-W48 模型可以检测出关键点的大致位置,但检测结果与标注的位置差异较大,即使对于特征明显的关键点,比如脚趾,依然无法准确检测。而 YOLO-Pose, RTMPose 以及 SimCC 模型不仅存在检测位置不准确的问题,个别关键点还存在检测错误的情况。本文模型的检测结果与标注信息重合度较高,检测位置准确,可见其用于高精度的关键点检测是非常有效的。

由图 5 可以看出,相较于低分辨率图像,采用高分辨率图像训练得到的 Lightweight-OpenPose, HRNet-W48, RTMPose 以及 SimCC 模型可以提升关键点的位置准确性。但由于参与训练的图像分辨率较高,与提出者建议的分辨率不

一致,引发两类典型失效模式:基于热图检测的 HRNet-W32 因局部特征歧义性增强,出现解剖学结构误判,如左右腕关节位置互换;基于描点匹配的 YOLO-Pose 模型受限于小目标上锚点覆盖不足,导致细粒度关键点误检。而本文模型的检测结果相较其他模型依然更准确,且全局特征可以很好表征关键点在图像中的全局位置信息,其预测结果不仅完整覆盖解剖学关键点,且严格遵循人体左右对称性约束,不会存在无法区分左右关键点的情况。

如图 6 为本文模型的检测结果,展示了模型针对具有不同背景、人体姿态图像的检测效果。由图 6 可知,本文模型检测到的各个关键点的位置准确,检测结果稳定,不受背景、人体姿态等因素影响。

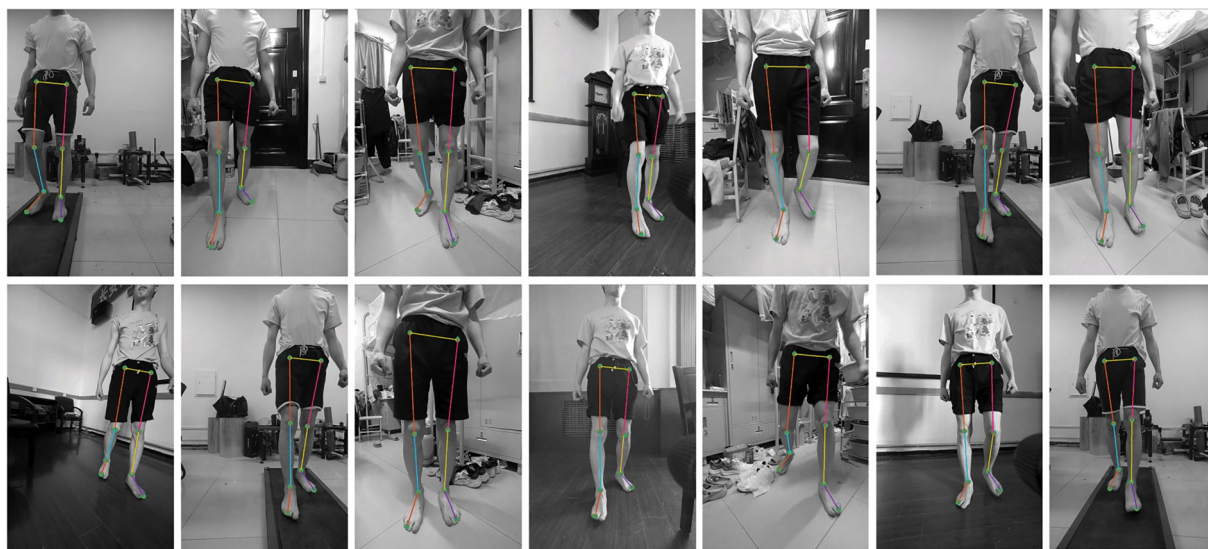


图6 本文模型检测结果

Fig. 6 Detection results of the model in this document

5 结 论

本文提出了一种人体下肢关键点检测模型, 现有的人体关键点检测模型通常采用低分辨率的输入图像进行训练, 检测结果不稳定, 定位精度较低, 在对精度要求较高的应用场景中检测效果较差。本文使用融合高、低分辨率特征的方式构建了轻量级人体下肢关键点检测模型, 并与先进模型进行对比实验。结果表明, 提出的模型在数据集上的检测率达到 95.2%, 与 Lightweight-OpenPose, HRNet-W32, HRNet-W48, YOLO-Pose, RTMPose 和 SimCC 模型相比, 检测精确率

提升了 4.1%~83.6%, FPS 提升了 7.6~13.9。证明了提出的模型在高精度人体下肢关键点检测中的有效性。

作者贡献声明:

夏晓华:方法的提出, 论文审核与编辑;
向浩鸣:方法的完善, 数据采集与实验设计, 论文整体修改;
陈坚:论文初稿撰写, 实验设计;
冯鑫森:方法的提出, 论文初稿撰写;
邱法博:实验的开展, 数据整理;
王耀耀:实验数据分析。

参考文献:

- [1] 张宇, 温光照, 米思娅, 等. 基于深度学习的二维人体姿态估计综述[J]. 软件学报, 2022, 33(11): 4173-4191.
ZHANG Y, WEN G Z, MI S Y, *et al.* Overview on 2D human pose estimation based on deep learning [J]. *Journal of Software*, 2022, 33 (11): 4173-4191. (in Chinese)
- [2] 李建更, 谢海征. 基于姿态估计的人体异常行为识别算法[J]. 北京工业大学学报, 2022, 48(7): 710-720.
LI J G, XIE H Z. Human abnormal behavior recognition algorithm based on pose estimation[J]. *Journal of Beijing University of Technology*, 2022, 48 (7): 710-720. (in Chinese)
- [3] 王璇, 吴佳奇, 阳康, 等. 煤矿井下人体姿态检测方法[J]. 工矿自动化, 2022, 48(5): 79-84.
WANG X, WU J Q, YANG K, *et al.* Human posture detection method in coal mine [J]. *Journal of Mine Automation*, 2022, 48 (5): 79-84. (in Chinese)
- [4] 郑奇, 郭立泉, 陈静, 等. 基于计算机视觉的居家康复训练评估算法[J]. 小型微型计算机系统, 2022, 43(11): 2336-2341.
ZHENG Q, GUO L Q, CHEN J, *et al.* Home rehabilitation training evaluation algorithm based on computer vision [J]. *Journal of Chinese Computer Systems*, 2022, 43 (11): 2336-2341. (in Chinese)

- [5] KOUR N, Sunanda, ARORA S. Computer-vision based diagnosis of Parkinson's disease *via* gait: a survey[J]. *IEEE Access*, 2019, 7: 156620-156645.
- [6] WANG Z H, WANG J C, ZHENG S, *et al.* Three-dimensional gait characteristics of patients after unilateral total knee arthroplasty [J]. *Medicine*, 2021, 100(34): e26968.
- [7] CIMOLIN V, VISMARA L, FERRARIS C, *et al.* Computation of gait parameters in post stroke and Parkinson's disease: a comparative study using RGB-D sensors and optoelectronic systems[J]. *Sensors*, 2022, 22(3): 824.
- [8] MUNTEANU M V, STANCIU M D, VLASE S. Image correlation to predict the gait recovery of patients in the post-surgery period [J]. *Procedia Manufacturing*, 2020, 46: 499-508.
- [9] OTA M, TATEUCHI H, HASHIGUCHI T, *et al.* Verification of validity of gait analysis systems during treadmill walking and running using human pose tracking algorithm[J]. *Gait & Posture*, 2021, 85: 290-297.
- [10] DÍAZ-ARANCIBIA J, CÓRDOVA M, ARANGO-LÓPEZ J, *et al.* Automatic gait analysis through computer vision: a pilot study[J]. *Neural Computing and Applications*, 2025, 37 (11) : 7619-7639.
- [11] CAO Z, SIMON T, WEI S H, *et al.* Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017, *Honolulu, HI, USA*. IEEE, 2017: 1302-1310.
- [12] STENUM J, ROSSI C, ROEMMICH R T. Two-dimensional video-based analysis of human gait using pose estimation[J]. *PLoS Computational Biology*, 2021, 17(4): e1008935.
- [13] 张子健, 马剑雄, 柏豪豪, 等. 基于深度学习构建模型识别下肢全长正位X线片中的下肢力线关键点及自动测量关键角度[J]. *中国医学影像技术*, 2022, 38(6): 901-906.
ZHANG Z J, MA J X, BAI H H, *et al.* Automatic measurement model for recognizing key points and automatic measurement of key angles of lower limb alignment on anteroposterior full-length lower limb radiograms based on deep learning[J]. *Chinese Journal of Medical Imaging Technology*, 2022, 38(6): 901-906. (in Chinese)
- [14] SUN K, XIAO B, LIU D, *et al.* Deep high-resolution representation learning for human pose estimation[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 15-20, 2019, *Long Beach, CA, USA*. IEEE, 2019: 5686-5696.
- [15] TAY C Z, LIM K H, PHANG J T S. Markerless gait estimation and tracking for postural assessment [J]. *Multimedia Tools and Applications*, 2022, 81(9): 12777-12794.
- [16] SETHI D, BHARTI S, PRAKASH C. A comprehensive survey on gait analysis: History, parameters, approaches, pose estimation, and future work [J]. *Artificial Intelligence in Medicine*, 2022, 129: 102314.
- [17] WADE L, NEEDHAM L, MCGUIGAN P, *et al.* Applications and limitations of current markerless motion capture methods for clinical gait biomechanics[J]. *PeerJ*, 2022, 10: e12995.
- [18] 刘驰, 周晋, 陈雨. Openpose算法在步态分析中的可靠性研究[J]. *皮革科学与工程*, 2021, 31(1): 65-68.
LIU C, ZHOU J, CHEN Y. Reliability of openpose algorithm in gait analysis[J]. *Leather Science and Engineering*, 2021, 31(1): 65-68. (in Chinese)
- [19] 何志超, 赵龙章, 陈闯. 用于人脸表情识别的多分辨率特征融合卷积神经网络[J]. *激光与光电子学进展*, 2018, 55(7): 370-375.
HE Z C, ZHAO L Z, CHEN C. Convolution neural network with multi-resolution feature fusion for facial expression recognition[J]. *Laser & Optoelectronics Progress*, 2018, 55(7): 370-375. (in Chinese)
- [20] DEBNATH B, O'BRIEN M, YAMAGUCHI M, *et al.* Adapting Mobilenets for Mobile Based Upper Body Pose Estimation[C]. 2018 *15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. November 27-30, 2018, *Auckland, New Zealand*. IEEE, 2018: 1-6.
- [21] WANG Q L, WU B G, ZHU P F, *et al.* ECA-net: efficient channel attention for deep convolutional neural networks[C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 13-19, 2020, *Seattle, WA, USA*. IEEE, 2020: 11531-11539.

- [22] TIAN Y J, SU D, LAURIA S, *et al.* Recent advances on loss functions in deep learning for computer vision [J]. *Neurocomputing*, 2022, 497: 129-158.
- [23] LI Y, WANG C X, CAO Y, *et al.* Human pose estimation based in-home lower body rehabilitation system [C]. 2020 *International Joint Conference on Neural Networks (IJCNN)*. July 19-24, 2020, Glasgow, UK. IEEE, 2020: 1-8.
- [24] OSOKIN D. Real-time 2D multi-person pose estimation on CPU: lightweight openpose [C]. *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods*. February 19-21, 2019. Prague, Czech Republic. SCITEPRESS-Science and Technology Publications, 2019: 744-748.
- [25] MAJI D, NAGORI S, MATHEW M, *et al.* YO-LO-pose: enhancing YOLO for multi person pose estimation using object keypoint similarity loss [C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 19-20, 2022, New Orleans, LA, USA. IEEE, 2022: 2636-2645.
- [26] 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.
- [27] LI Y J, YANG S, LIU P D, *et al.* SimCC: a simple coordinate classification perspective for human pose estimation [C]. *Computer Vision-ECCV 2022*. Cham: Springer Nature Switzerland, 2022: 89-106.

作者简介:



夏晓华(1987—),男,山东临沂人,博士,副教授,2009年于中国石油大学(华东)获得学士学位,2015年于西安交通大学获得博士学位,主要从事机器视觉与光机电一体化方面的研究。
E-mail:xhxia@chd.edu.cn

作者简介:



向浩鸣(1998—),男,陕西汉中,人,博士研究生,分别在2020年和2024年于长安大学获得学士、硕士学位,主要从事机器视觉方面的研究。E-mail: 2021125044@chd.edu.cn