



(21) 申请号 202410444233.4

(22) 申请日 2023.07.21

(65) 同一申请的已公布的文献号

申请公布号 CN 118506397 A

(43) 申请公布日 2024.08.16

(62) 分案原申请数据

202310899391.4 2023.07.21

(73) 专利权人 北京优创新港科技股份有限公司

地址 100000 北京市海淀区上地东路35号1

号楼2层1-202-238

(72) 发明人 李继凯

(74) 专利代理机构 成都余行专利代理事务所

(普通合伙) 51283

专利代理师 邢智勇

(51) Int.Cl.

G06V 40/10 (2022.01)

G06V 10/25 (2022.01)

G06V 10/52 (2022.01)

G06V 10/80 (2022.01)

G06V 10/82 (2022.01)

G06N 3/0464 (2023.01)

G06N 3/08 (2023.01)

(56) 对比文件

CN 109977912 A, 2019.07.05

CN 113673354 A, 2021.11.19

审查员 安静

权利要求书3页 说明书10页 附图5页

(54) 发明名称

基于单个模型高效率识别人体位置和关键点检测方法

(57) 摘要

本发明公开了基于单个模型高效率识别人体位置和关键点检测方法,该方法包括,步骤1、构建人体关键点估计网络,所述人体关键点估计网络包括多尺度特征提取模块、检测头模块、人体位置框检测模块、感兴趣区域对齐模块和人体关键点检测模块;步骤2、利用训练数据集训练所述人体关键点估计网络,得到人体关键点估计模型;步骤3、将待检测图像输入所述人体关键点估计模型,输出不同尺度的人体关键点预测坐标;步骤4、对不同尺度的所述人体关键点预测坐标进行融合,得到人体关键点坐标。本发明是在自顶而下的基础上次加以改进,采用一次前向推理的结构,即只需要单个模型就可实现人体位置的检测和人体关键点的检测,确保了检测精度的同时,又提高了运行效率。

步骤1, 构建人体关键点估计网络, 人体关键点估计网络包括多尺度特征提取模块、检测头模块、人体位置框检测模块、感兴趣区域对齐模块和人体关键点检测模块

步骤2, 利用训练数据集训练人体关键点估计网络, 得到人体关键点估计模型

步骤3, 将待检测图像输入人体关键点估计模型, 输出不同尺度的人体关键点预测坐标

步骤4, 对不同尺度的人体关键点预测坐标进行融合, 得到人体关键点坐标

1. 基于单个模型高效率识别人体位置和关键点检测方法,其特征在于,包括:

步骤1、构建人体关键点估计网络,所述人体关键点估计网络包括多尺度特征提取模块、检测头模块、人体位置框检测模块、感兴趣区域对齐模块和人体关键点检测模块;

所述多尺度特征提取模块用于接受输入图像并输出多个尺度的特征图;

所述检测头模块用于接受特征图并输出多个尺度的检测头特征图;

所述人体位置框检测模块用于接受检测头特征图并输出多个尺度的预测人体位置框;

所述感兴趣区域对齐模块用于接受检测头特征图、人体位置框并输出多个尺度的对齐特征图;

所述人体关键点检测模块用于接受对齐特征图并输出不同尺度的人体关键点预测坐标;

步骤2、利用训练数据集训练所述人体关键点估计网络,得到人体关键点估计模型;

所述步骤2包括:

步骤2.1,选择所述训练数据集,所述训练数据集为coco数据集;

步骤2.2,所述人体位置框为真实人体位置框或所述预测人体位置框,所述真实人体位置框为coco数据集中标注的人体位置框,训练过程中实时获取所述真实人体位置框以及所述预测人体位置框;

当所述真实人体位置框与所述预测人体位置框的 $IOU \leq 80\%$ 时,所述人体关键点估计网络的感兴趣区域对齐模块输入的人体位置框为真实人体位置框;

当所述真实人体位置框与所述预测人体位置框的 $IOU > 80\%$ 时,所述人体关键点估计网络的感兴趣区域对齐模块输入的人体位置框为真实人体位置框和预测人体位置框;

步骤2.3,采用梯度下降法训练所述人体关键点估计网络,得到人体关键点估计中间模型;

步骤2.4,对所述人体关键点估计中间模型的感兴趣区域对齐模块进行调整,使其输入的所述人体位置框为所述预测人体位置框,得到人体关键点估计模型;

步骤3、将待检测图像输入所述人体关键点估计模型,输出不同尺度的人体关键点预测坐标;

所述步骤3具体步骤为:

步骤31,将待检测图像输入所述多尺度特征提取模,输出多个尺度的特征图;

步骤32,将所述特征图输入所述检测头模,输出多个尺度的检测头特征图;

步骤33,将所述检测头特征图输入所述人体位置框检测模块,输出多个尺度的预测人体位置框;

步骤34,将所述检测头特征图以及所述预测人体位置框输入所述感兴趣区域对齐模块,输出多个尺度的对齐特征图;

步骤35,将所述对齐特征图输入人体关键点检测模块,输出多个尺度的人体关键点预测特征图;

步骤36,在每个尺度的所述人体关键点预测特征图中,取每个通道中最大值的坐标作为人体关键点预测坐标,得到不同尺度的人体关键点预测坐标;

步骤4、对不同尺度的所述人体关键点预测坐标进行融合,得到人体关键点坐标;

所述步骤4包括:

步骤41,将不同尺度中的人体关键点预测坐标 $\{X_{D,i}, Y_{D,i}\}$ 映射回待检测图像中,得到不同尺度的人体关键点映射坐标 $\{X_{D,i}^{org}, Y_{D,i}^{org}\}$;

步骤42,将不同尺度的人体关键点映射坐标 $\{X_{D,i}^{org}, Y_{D,i}^{org}\}$ 求均值,如下:

$$X_i^{org} = \frac{1}{D} \sum_D X_{D,i}^{org} \quad Y_i^{org} = \frac{1}{D} \sum_D Y_{D,i}^{org}$$
 得到最终的人体关键点坐标, $\{X_i^{org}, Y_i^{org}\}$, 其中, $i = 0, \dots, 16$; D 为不同尺度;

所述多尺度特征图提取模块输出4个尺度的特征图,所述检测头模块输出4个尺度的检测头特征图,所述人体位置框检测模块输出4个尺度的预测人体位置框;所述感兴趣区域对齐模块输出4个尺度的对齐特征图,所述人体关键点检测模块输出4个不同尺度的17个人体关键点预测坐标;

所述多尺度特征图提取模块包括Backbone特征提取模块、自上而下多尺度特征融合模块以及自下而上多尺度特征融合模块;

所述Backbone特征提取模块采用多个残差网络块,得到至少4个尺度的特征图,输出分辨率从小到大的最后4个尺度的特征图,分别为: $BONE_{end}$, $BONE_{end-1}$, $BONE_{end-2}$, $BONE_{end-3}$;

所述自上而下多尺度特征融合模块,输出4个尺度的特征图,分辨率从小到大分别为 UP_{end} , UP_{end-1} , UP_{end-2} , UP_{end-3} , 其中, UP_{end} 通过对 $BONE_{end}$ 进行卷积操作得到; UP_{end-1} 通过对 UP_{end} 进行转置卷积并加上 $BONE_{end-1}$ 的卷积操作后得到; UP_{end-2} 通过对 UP_{end-1} 进行转置卷积并加上 $BONE_{end-2}$ 的卷积操作后得到; UP_{end-3} 通过对 UP_{end-2} 进行转置卷积并加上 $BONE_{end-3}$ 的卷积操作后得到;

所述自下而上多尺度特征融合模块,输出4个尺度的特征图,分辨率从小到大分别为 $DOWN_{end}$, $DOWN_{end-1}$, $DOWN_{end-2}$, $DOWN_{end-3}$, 其中, $DOWN_{end-3}$ 通过对 UP_{end-3} 卷积操作得到; $DOWN_{end-2}$ 通过对 $DOWN_{end-3}$ 卷积操作并加上 UP_{end-2} 的卷积操作后得到; $DOWN_{end-1}$ 通过对 $DOWN_{end-2}$ 卷积操作并加上 UP_{end-1} 的卷积操作后得到; $DOWN_{end}$ 通过对 $DOWN_{end-1}$ 卷积操作并加上 UP_{end} 的卷积操作后得到;

所述检测头模块通过对4个尺度的特征图分别做卷积操作,输出4个尺度的检测头特征图,分别为: $Head_{end}$, $Head_{end-1}$, $Head_{end-2}$, $Head_{end-3}$;

所述人体位置框检测模块分别对4个尺度的检测头特征图做卷积操作,得到4个尺度的预测人体位置框,分别为 BOX_{end} , BOX_{end-1} , BOX_{end-2} , BOX_{end-3} , 为每个尺度中的每个像素预测至少1个anchor,每个anchor预测6个值,分别为人体位置框的中心位置的横坐标、人体位置框的中心位置的纵坐标、人体位置框相对中心位置宽的偏移量、人体位置框相对中心位置高的偏移量、目标存在的置信度 box_{conf} 以及人体类别置信度 $human_{conf}$;

所述人体关键点检测模块对4个尺度的对齐特征图采用卷积操作,得到4个尺度的人体关键点预测特征图,分别为 $POINT_{end}$, $POINT_{end-1}$, $POINT_{end-2}$, $POINT_{end-3}$, 所述人体关键点预测特征图的通道数为17;

所述步骤1-4自顶而下的基础上次加以改进,采用一次前向推理的结构,包括了人体位置框检测和人体关键点检测,即只需要单个模型就可实现人体位置的检测和人体关键点的

检测,确保了自顶而下技术方案的准确性的同时,又提高了运行效率。

2.根据权利要求1所述的基于单个模型高效率识别人体位置和关键点检测方法,其特征在于,所述预测人体位置框的特征图的通道数为18,为每个像素预测3个anchor。

3.根据权利要求1所述的基于单个模型高效率识别人体位置和关键点检测方法,其特征在于,所述步骤2中,总误差公式为: $Loss_{total} = \alpha Loss_{box} + (1 - \alpha) Loss_{point}$ 其中, $Loss_{total}$ 为总误差, $Loss_{box}$ 为人体位置框总误差, $Loss_{point}$ 为人体关键点总误差, $\alpha \in (0, 1)$ 为用以权衡人体位置框总误差和人体关键点总误差的权重;

所述人体位置框总误差公式为: $Loss_{box} = \sum_{D,i,j,k} (1 - SIoU(BoX_{gt}^{D,i,j,k}, BoX_{pred}^{D,i,j,k}))$ 采用SIoU来计算所述真实人体位置框和所述预测人体位置框之间的交集;

其中, D 表示不同的尺度, i 和 j 分表示人体位置框的中心位置的横、纵坐标, k 表示anchor的序号, $BoX_{gt}^{D,i,j,k}$ 表示真实人体位置框, $BoX_{pred}^{D,i,j,k}$ 表示预测人体位置框;

所述人体关键点总误差公式为: $Loss_{point} = \sum_{D,m,n,h} (x_{m,n,D,h} - \hat{x}_{m,n,D,h})^2$ 其中, h 为人体关键点的序号, m 和 n 分别为人体关键点的横、纵坐标, $x_{m,n,D,h}$ 为人体关键点真实的高斯值, $\hat{x}_{m,n,D,h}$ 为人体关键点预测的高斯值。

基于单个模型高效率识别人体位置和关键点检测方法

技术领域

[0001] 本发明申请为申请日2023年7月21日,申请号为:202310899391.4,名称为“一种人体关键点检测方法及系统”的发明专利申请的分案申请。

[0002] 本发明涉及计算机视觉技术领域,尤其涉及一种人体关键点检测方法及系统。

背景技术

[0003] 人体关键点检测通过对人体主要骨骼点(关键点)的位置预测,实现人体姿态估计。人体姿态估计是计算机视觉领域的基础任务之一,可广泛应用于动作识别、行人跟踪、自动驾驶、人机交互等领域。行业内的技术采用分为自底而上(bottom-up)和自顶而下(top-down)两种思路。

[0004] 自底而上的方法直接从原图检测人体关键点,再根据不同关键点间的关系进行分组。自顶而下的方法先检测行人区域,再对每个行人进行关键点检测,可视为行人检测和单人姿态估计的组合。通常自顶而下的技术方案的准确性要高于自底而上,但是效率却要低于自底而下。

[0005] 为了实现高精度的人体关键点检测,发明专利一种多人人体姿态估计方法(CN111339903B),公开了采用自底向上与自顶向下两种方法结合的两阶段训练方法,训练深度特征人体关键点检测模型,首先进行自底向上的方法训练该深度特征人体关键点检测模型,然后进行自顶向下的方法训练该深度特征人体关键点检测模型,最终利用两阶段训练好深度特征人体关键点检测模型检测人体关键点,采用自底向上训练深度特征人体关键点检测模型,输出人体关键点检测置信度热图;采用自顶向下训练深度特征人体关键点检测模型,识别裁剪后的图像中的所有人体关键点。

[0006] 上述方法,可实现图像中多人体关键点的识别,但模型训练过程复杂,特提出本发明,本发明是在自顶而下的基础上次加以改进,在确保自顶而下技术方案的准确性的同时,又提高了运行效率。

发明内容

[0007] 本发明的发明目的是针对现有技术的不足,提供了一种人体关键点检测方法及系统,可精确实现人体关键点的估计,适用范围广。

[0008] 第一方面,本申请提供一种人体关键点检测方法,包括:

[0009] 步骤1,构建人体关键点估计网络,所述人体关键点估计网络包括多尺度特征提取模块、检测头模块、人体位置框检测模块、感兴趣区域对齐模块和人体关键点检测模块;

[0010] 所述多尺度特征提取模块用于接受输入图像并输出多个尺度的特征图;

[0011] 所述检测头模块用于接受特征图并输出多个尺度的检测头特征图;

[0012] 所述人体位置框检测模块用于接受检测头特征图并输出多个尺度的预测人体位置框;

[0013] 所述感兴趣区域对齐模块用于接受检测头特征图、人体位置框并输出多个尺度的

对齐特征图；

[0014] 所述人体关键点检测模块用于接受对齐特征图并输出不同尺度的人体关键点预测坐标；

[0015] 步骤2,利用训练数据集训练所述人体关键点估计网络,得到人体关键点估计模型；

[0016] 步骤3,将待检测图像输入所述人体关键点估计模型,输出不同尺度的人体关键点预测坐标；

[0017] 步骤4,对不同尺度的所述人体关键点预测坐标进行融合,得到人体关键点坐标。

[0018] 进一步地,所述多尺度特征图提取模块输出4个尺度的特征图,所述检测头模块输出4个尺度的检测头特征图,所述人体位置框检测模块输出4个尺度的预测人体位置框；所述感兴趣区域对齐模块输出4个尺度的对齐特征图,所述人体关键点检测模块输出4个不同尺度的17个人体关键点预测坐标。

[0019] 进一步地,所述多尺度特征图提取模块包括Backbone特征提取模块、自上而下多尺度特征融合模块以及自下而上多尺度特征融合模块。

[0020] 进一步地,所述Backbone特征提取模块采用多个残差网络块,得到至少4个尺度的特征图,输出分辨率从小到大的最后4个尺度的特征图,分别为: $BONE_{end}$, $BONE_{end-1}$, $BONE_{end-2}$, $BONE_{end-3}$ ；

[0021] 所述自上而下多尺度特征融合模块,输出4个尺度的特征图,分辨率从小到大分别为 UP_{end} , UP_{end-1} , UP_{end-2} , UP_{end-3} ,其中, UP_{end} 通过对 $BONE_{end}$ 进行卷积操作得到； UP_{end-1} 通过对 UP_{end} 进行转置卷积并加上 $BONE_{end-1}$ 的卷积操作后得到； UP_{end-2} 通过对 UP_{end-1} 进行转置卷积并加上 $BONE_{end-2}$ 的卷积操作后得到； UP_{end-3} 通过对 UP_{end-2} 进行转置卷积并加上 $BONE_{end-3}$ 的卷积操作后得到；

[0022] 所述自下而上多尺度特征融合模块,输出4个尺度的特征图,分辨率从小到大分别为 $DOWN_{end}$, $DOWN_{end-1}$, $DOWN_{end-2}$, $DOWN_{end-3}$,其中, $DOWN_{end-3}$ 通过对 UP_{end-3} 卷积操作得到； $DOWN_{end-2}$ 通过对 $DOWN_{end-3}$ 卷积操作并加上 UP_{end-2} 的卷积操作后得到； $DOWN_{end-1}$ 通过对 $DOWN_{end-2}$ 卷积操作并加上 UP_{end-1} 的卷积操作后得到； $DOWN_{end}$ 通过对 $DOWN_{end-1}$ 卷积操作并加上 UP_{end} 的卷积操作后得到。

[0023] 进一步地,所述检测头模块通过对4个尺度的特征图分别做卷积操作,输出4个尺度的检测头特征图,分别为: $Head_{end}$, $Head_{end-1}$, $Head_{end-2}$, $Head_{end-3}$ 。

[0024] 进一步地,所述人体位置框检测模块分别对4个尺度的检测头特征图做卷积操作,得到4个尺度的预测人体位置框,分别为 BOX_{end} , BOX_{end-1} , BOX_{end-2} , BOX_{end-3} ,为每个尺度中的每个像素预测至少1个anchor,每个anchor预测6个值,分别为人体位置框的中心位置的横坐标、人体位置框的中心位置的纵坐标、人体位置框相对中心位置宽的偏移量、人体位置框相对中心位置高的偏移量、目标存在的置信度 box_{conf} 以及人体类别置信度 $human_{conf}$ 。

[0025] 进一步地,所述预测人体位置框的特征图的通道数为18,为每个像素预测3个anchor。

[0026] 进一步地,所述感兴趣区域对齐模块采用感兴趣区域对齐技术,输出4个尺度的对齐特征图,分别为 $ALIGN_{end}$, $ALIGN_{end-1}$, $ALIGN_{end-2}$, $ALIGN_{end-3}$ 。

[0027] 进一步地,所述人体关键点检测模块对4个尺度的对齐特征图采用卷积操作,得到

4个尺度的人体关键点预测特征图,分别为 $POINT_{end}$, $POINT_{end-1}$, $POINT_{end-2}$, $POINT_{end-3}$,所述人体关键点预测特征图的通道数为17。

[0028] 进一步地,所述步骤2包括:

[0029] 步骤21,选择所述训练数据集,所述训练数据集为coco数据集;

[0030] 步骤22,所述人体位置框为真实人体位置框或所述预测人体位置框,所述真实人体位置框为coco数据集中标注的人体位置框,训练过程中实时获取所述真实人体位置框以及所述预测人体位置框;

[0031] 当所述真实人体位置框与所述预测人体位置框的 $IOU \leq 80\%$ 时,所述人体关键点估计网络的感兴趣区域对齐模块输入的人体位置框为真实人体位置框;

[0032] 当所述真实人体位置框与所述预测人体位置框的 $IOU > 80\%$ 时,所述人体关键点估计网络的感兴趣区域对齐模块输入的人体位置框为真实人体位置框和预测人体位置框;

[0033] 步骤23,采用梯度下降法训练所述人体关键点估计网络,得到人体关键点估计中间模型;

[0034] 步骤24,对所述人体关键点估计中间模型的感兴趣区域对齐模块进行调整,使其输入的所述人体位置框为所述预测人体位置框,得到人体关键点估计模型。

[0035] 进一步地,所述步骤2中,总误差公式为:

$$[0036] \quad Loss_{total} = \alpha Loss_{box} + (1 - \alpha) Loss_{point}$$

[0037] 其中, $Loss_{total}$ 为总误差, $Loss_{box}$ 为人体位置框总误差, $Loss_{point}$ 为人体关键点总误差, $\alpha \in (0, 1)$ 用以权衡人体位置框总误差和人体关键点总误差的权重。

[0038] 进一步地,所述人体位置框总误差公式为:

$$[0039] \quad Loss_{box} = \sum_{D,i,j,k} (1 - SIoU(Box_{gt}^{D,i,j,k}, Box_{pred}^{D,i,j,k}))$$

[0040] 采用SIoU来计算所述真实人体位置框和所述预测人体位置框之间的交集;

[0041] 其中,D表示不同的尺度,i和j分表示人体位置预测特征图的横纵坐标,k表示anchor的序号, Box_{gt} 表示真实人体位置框, Box_{pred} 表示预测人体位置框。

[0042] 进一步地,所述人体关键点总误差公式为:

$$[0043] \quad Loss_{point} = \sum_{D,m,n,h} (x_{m,n,D,h} - \hat{x}_{m,n,D,h})^2$$

[0044] 其中,D为不同的尺度,h为人体关键点的序号,m和n分别为人体关键点预测特征图的横纵坐标, $x_{m,n,s,h}$ 为人体关键点真实高斯值, $\hat{x}_{m,n,s,h}$ 为人体关键点预测坐标。

[0045] 进一步地,所述人体关键点真实高斯值根据coco数据集中标定的人体关键点坐标,采用高斯分布对所述人体关键点坐标进行变换获得。

[0046] 进一步地,人体关键点的采用高斯分布来表示,将人体的关键点真实位置坐标的映射到不同尺度的人体关键点预测特征图,并对坐标并取整,设为;设置高斯圆的半径,设为R;在人体关键点预测特征图上,以为圆心,半径为R填充高斯函数计算值。

[0047] 进一步地,所述步骤3具体步骤为:

[0048] 步骤31,将待检测图像输入所述多尺度特征提取模,输出多个尺度的特征图;

[0049] 步骤32,将所述特征图输入所述检测头模,输出多个尺度的检测头特征图;

[0050] 步骤33,将所述检测头特征图输入所述人体位置框检测模块,输出多个尺度的预测人体位置框;

[0051] 步骤34,将所述检测头特征图以及所述预测人体位置框输入所述感兴趣区域对齐模块,输出多个尺度的对齐特征图;

[0052] 步骤35,将所述对齐特征图输入人体关键点检测模块,输出多个尺度的人体关键点预测特征图;

[0053] 步骤36,在每个尺度的所述人体关键点预测特征图中,取每个通道中最大值的坐标作为人体关键点预测坐标,得到不同尺度的人体关键点预测坐标。

[0054] 进一步地,所述步骤4包括:

[0055] 步骤41,将不同尺度的所述人体关键点预测坐标映射回所述待检测图像中,得到不同尺度的人体关键点映射坐标;

[0056] 步骤42,将不同尺度的所述人体关键点映射坐标求均值,得到所述人体关键点坐标。

[0057] 第二方面,本申请还提供一种人体关键点检测系统,包括人体关键点估计网络构建模块、网络训练模块、人体关键点坐标预测模块、人体关键点坐标融合模块;

[0058] 所述人体关键点估计网络构建模块用于构建人体关键点估计网络,所述人体关键点估计网络包括多尺度特征提取模块、检测头模块、人体位置框检测模块、感兴趣区域对齐模块和人体关键点检测模块;

[0059] 所述多尺度特征提取模块用于接受输入图像并输出多个尺度的特征图;

[0060] 所述检测头模块用于接受特征图并输出多个尺度的检测头特征图;

[0061] 所述人体位置框检测模块用于接受检测头特征图并输出多个尺度的预测人体位置框;

[0062] 所述感兴趣区域对齐模块用于接受检测头特征图、人体位置框并输出多个尺度的对齐特征图;

[0063] 所述人体关键点检测模块用于接受对齐特征图并输出不同尺度的人体关键点预测坐标;

[0064] 所述网络训练模块利用训练数据集训练所述人体关键点估计网络,得到人体关键点估计模型;

[0065] 人体关键点坐标预测模块,通过将待检测图像输入所述人体关键点估计模型,输出不同尺度的人体关键点预测坐标;

[0066] 人体关键点坐标融合模块,通过将不同尺度的所述人体关键点预测坐标进行融合,得到人体关键点坐标。

[0067] 本发明的有益效果在于:

[0068] (1) 本发明是在自顶而下的基础上次加以改进,采用一次前向推理的结构,包括了人体位置框检测和人体关键点检测,即只需要单个模型就可实现人体位置的检测和人体关键点的检测,确保了自顶而下技术方案的准确性的同时,又提高了运行效率。

[0069] (2) 设计了集成多尺度特征提取模块、检测头模块、人体位置框检测模块、感兴趣区域对齐模块和人体关键点检测模块的人体关键点检测网络,使用多通道、多阶段的模式对多尺度特征图进行分析,得到不同尺度的人体关键点预测坐标,最后以多尺度融合的方法

式将多尺度的人体关键点预测坐标进行融合,得到更加全面且有效的人体关键点坐标,实现人体关键点的高精度识别。

[0070] (3) 集合在密集人群的应用场景中,本发明可实现多人的人体框和关键点同时识别,能够准确且快速地完成人体关键点的估计,无需额外的模型预先提取人体框,算法流程简便,运算量低,满足各种低算力设备的轻量化部署要求。

附图说明

[0071] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据提供的附图获得其他的附图。

[0072] 图1为一种人体关键点检测方法的流程示意图。

[0073] 图2为人体关键点估计网络结构。

[0074] 图3为一种人体关键点检测方法中步骤S2的流程示意图。

[0075] 图4为步骤S3的流程示意图。

[0076] 图5为一种人体关键点检测系统的组成框图。

具体实施方式

[0077] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述;显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例,基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0078] 如图1所示,本发明实施例提供了一种人体关键点检测方法,包括以下步骤:

[0079] 步骤1,构建人体关键点估计网络,人体关键点估计网络包括多尺度特征提取模块、检测头模块、人体位置框检测模块、感兴趣区域对齐模块和人体关键点检测模块;

[0080] 多尺度特征提取模块用于接受输入图像并输出多个尺度的特征图;

[0081] 检测头模块用于接受特征图并输出多个尺度的检测头特征图;

[0082] 人体位置框检测模块用于接受检测头特征图并输出多个尺度的预测人体位置框;

[0083] 感兴趣区域对齐模块用于接受检测头特征图、人体位置框并输出多个尺度的对齐特征图;

[0084] 人体关键点检测模块用于接受对齐特征图并输出不同尺度的人体关键点预测坐标。

[0085] 步骤2,利用训练数据集训练人体关键点估计网络,得到人体关键点估计模型。

[0086] 步骤3,将待检测图像输入人体关键点估计模型,输出不同尺度的人体关键点预测坐标。

[0087] 步骤4,对不同尺度的人体关键点预测坐标进行融合,得到人体关键点坐标。

[0088] 本实施例,在自顶而下的基础上加以改进,采用一次前向推理的结构,包括了人体位置框检测和人体关键点检测,即只需要单个模型就可实现人体位置的检测和人体关键点的检测,确保了自顶而下技术方案的准确性的同时,又提高了运行效率。

[0089] 不同尺度的人体关键的覆盖的相同区域的语义信息不同,为了实现多尺度的人体关键点特征分析,设计了多尺度特征图提取模块,多尺度特征提取模块用于输出至少两个尺度的特征图。

[0090] 本实施例中,多尺度特征图提取模块输出4个尺度的特征图,检测头模块输出4个尺度的检测头特征图,人体位置框检测模块输出4个尺度的预测人体位置框;感兴趣区域对齐模块输出4个尺度的对齐特征图,人体关键点检测模块输出4个不同尺度的17个人体关键点预测坐标。

[0091] 设计了集成多尺度特征提取模块、检测头模块、人体位置框检测模块、感兴趣区域对齐模块和人体关键点检测模块的人体关键点检测网,使用多通道、多阶段的模式对多尺度特征图进行分析,得到不同尺度的人体关键点预测坐标,最后以多尺度融合的方式将多尺度的人体关键点预测坐标进行融合,得到更加全面且有效的人体关键点坐标,实现人体关键点的高精度识别。

[0092] 如图2所示,多尺度特征图提取模块包括Backbone特征提取模块、自上而下多尺度特征融合模块以及自下而上多尺度特征融合模块。

[0093] Backbone特征提取模块采用多个残差网络块,得到至少4个尺度的特征图,输出分辨率从小到大的最后4个尺度的特征图,分别为: $BONE_{end}$, $BONE_{end-1}$, $BONE_{end-2}$, $BONE_{end-3}$;

[0094] 自上而下多尺度特征融合模块,输出4个尺度的特征图,分辨率从小到大分别为 UP_{end} , UP_{end-1} , P_{end-2} , UP_{end-3} ,其中, UP_{end} 通过对 $BONE_{end}$ 进行卷积操作得到; UP_{end-1} 通过对 UP_{end} 进行转置卷积并加上 $BONE_{end-1}$ 的卷积操作后得到; UP_{end-2} 通过对 UP_{end-1} 进行转置卷积并加上 $BONE_{end-2}$ 的卷积操作后得到; UP_{end-3} 通过对 UP_{end-2} 进行转置卷积并加上 $BONE_{end-3}$ 的卷积操作后得到;

[0095] 自下而上多尺度特征融合模块,输出4个尺度的特征图,分辨率从小到大分别为 $DOWN_{end}$, $DOWN_{end-1}$, $DOWN_{end-2}$, $DOWN_{end-3}$,其中, $DOWN_{end-3}$ 通过对 UP_{end-3} 卷积操作得到; $DOWN_{end-2}$ 通过对 $DOWN_{end-3}$ 卷积操作并加上 UP_{end-2} 的卷积操作后得到; $DOWN_{end-1}$ 通过对 $DOWN_{end-2}$ 卷积操作并加上 UP_{end-1} 的卷积操作后得到; $DOWN_{end}$ 通过对 $DOWN_{end-1}$ 卷积操作并加上 UP_{end} 的卷积操作后得到。

[0096] 将图像输入多尺度特征图提取模块输出依次经过Backbone特征提取模块、自上而下多尺度特征融合模块以及自下而上多尺度特征融合模块,输出4个尺度的特征图,分辨率从小到大分别为 $DOWN_{end}$, $DOWN_{end-1}$, $DOWN_{end-2}$, $DOWN_{end-3}$ 。

[0097] 将 $DOWN_{end}$, $DOWN_{end-1}$, $DOWN_{end-2}$, $DOWN_{end-3}$ 输入检测头模块,检测头模块通过对4个尺度的特征图分别做卷积操作,输出4个尺度的检测头特征图,分别为: $Head_{end}$, $Head_{end-1}$, $Head_{end-2}$, $Head_{end-3}$ 。

[0098] 将 $Head_{end}$, $Head_{end-1}$, $Head_{end-2}$, $Head_{end-3}$ 输入人体位置框检测模块,人体位置框检测模块分别对4个尺度的检测头特征图做卷积操作,得到4个尺度的预测人体位置框,分别为 BOX_{end} , BOX_{end-1} , BOX_{end-2} , BOX_{end-3} ,为每个尺度中的每个像素预测至少1个anchor,每个anchor预测6个值,分别为人体位置框的中心位置的横坐标、人体位置框的中心位置的纵坐标、人体位置框相对中心位置宽的偏移量、人体位置框相对中心位置高的偏移量、目标存在的置信度 box_{conf} 以及人体类别置信度 $human_{conf}$ 。

[0099] 本实施例中,预测人体位置框的特征图的通道数为18,为每个像素预测3个

anchor。

[0100] 每个预测人体位置框的特征图的通道数为18,18表示为每个人体位置框预测3个anchor,每个anchor预测6个值。

[0101] 将 $Head_{end}$, $Head_{end-1}$, $Head_{end-2}$, $Head_{end-3}$ 、人体位置框输入感兴趣区域对齐模块,感兴趣区域对齐模块采用感兴趣区域对齐技术,输出4个尺度的对齐特征图,分别为 $ALIGN_{end}$, $ALIGN_{end-1}$, $ALIGN_{end-2}$, $ALIGN_{end-3}$ 。

[0102] 具体地,在预测阶段,将不同尺度的检测头特征图和预测人体位置框分别采用感兴趣区域对齐技术,得到不同尺度的对齐特征图,通过 $Head_{end}$ 和 BOX_{end} 得到 $ALIGN_{end}$,通过 $Head_{end-1}$ 和 BOX_{end-1} 得到 $ALIGN_{end-1}$,通过 $Head_{end-2}$ 和 BOX_{end-2} 得到 $ALIGN_{end-2}$,通过 $Head_{end-3}$ 和 BOX_{end-3} 得到 $ALIGN_{end-3}$ 。

[0103] 人体关键点检测模块对4个尺度的对齐特征图 $ALIGN_{end}$, $ALIGN_{end-1}$, $ALIGN_{end-2}$, $ALIGN_{end-3}$ 采用卷积操作,得到4个尺度的人体关键点预测特征图,分别为 $POINT_{end}$, $POINT_{end-1}$, $POINT_{end-2}$, $POINT_{end-3}$,人体关键点预测特征图的通道数为17,17表示人体的17个关键点,即每个通道表示一个人体关键点。

[0104] 为了建立高精度的人体关键点估计模型,通过步骤2,采用coco数据集对人体关键点估计网络进行训练,如图3所示,步骤2包括:

[0105] 步骤21,选择训练数据集,训练数据集为coco数据集。

[0106] 可选的训练数据集有像COCO、MPII和CrowdPose这样的公共数据集,人体关键点检测属于比较复杂一类的问题。为神经网络模型建立一个合适的数据集是很困难的,图像中每个人的每个关节都必须定位和标记,这是一项琐碎而费时的任务。本实施例采用COCO数据集,它有大约80类图像和大约250000个人物实例。

[0107] 步骤22,人体位置框为真实人体位置框或预测人体位置框,真实人体位置框为coco数据集中标注的人体位置框,训练过程中实时获取真实人体位置框以及预测人体位置框;

[0108] 当真实人体位置框与预测人体位置框的 $IOU \leq 80\%$ 时,人体关键点估计网络的感兴趣区域对齐模块输入的人体位置框为真实人体位置框;

[0109] 当真实人体位置框与预测人体位置框的 $IOU > 80\%$ 时,人体关键点估计网络的感兴趣区域对齐模块输入的人体位置框为真实人体位置框和预测人体位置框。

[0110] 在人体关键点估计网络训练的初始阶段,由于模型预测的人体位置框不准确,此时,感兴趣区域对齐模块的只接收真实人体位置框,如图2所示,输入包括真实人体位置框,真实人体位置框输入感兴趣区域对齐模块。

[0111] 随着模型的不断训练,当真实人体位置框与通过模型获得的预测人体位置框的IoU值超过80%时,感兴趣区域对齐模块的将同时接收真实人体位置框和预测人体位置框。增加了样本的多样性和准确性。

[0112] 步骤23,采用梯度下降法训练人体关键点估计网络,得到人体关键点估计中间模型。

[0113] 步骤24,对人体关键点估计中间模型的感兴趣区域对齐模块进行调整,使其输入的人体位置框为预测人体位置框,得到人体关键点估计模型。

[0114] 通过步骤2.3获得了训练好的人体关键点估计中间模型,该模型感兴趣区域对齐

模块输入的人体位置框为真实人体位置框和预测人体位置框,但在利用训练好的模型进行人体关键点检测的时候,不存在人工标定的真实人体位置框,只有模型获取的预测人体位置框,基于此,通过步骤2.4对人体关键点估计中间模型的感兴趣区域对齐模块的输入进行调整,将真实人体位置框在感兴趣区域对齐模块中去掉,获得最终的人体关键点估计模型。

[0115] 步骤2中,训练过程中的总误差公式为:

$$[0116] \quad Loss_{total} = \alpha Loss_{box} + (1 - \alpha) Loss_{point}$$

[0117] 其中, $Loss_{total}$ 为总误差, $Loss_{box}$ 为人体位置框总误差, $Loss_{point}$ 为人体关键点总误差, $\alpha \in (0, 1)$ 用以权衡人体位置框总误差和人体关键点总误差的权重。

[0118] 本实施例中,同时对模型预测的人体的位置框和人体关键点信息进行约束,这样保证了最终的人体位置关键点估计模型更加准确。

[0119] 人体位置框总误差公式为:

$$[0120] \quad Loss_{box} = \sum_{D,i,j,k} (1 - SIoU(Box_{gt}^{D,i,j,k}, Box_{pred}^{D,i,j,k}))$$

[0121] 采用SIoU来计算真实人体位置框和所述预测人体位置框之间的交集;

[0122] 其中, D 表示不同的尺度, i 和 j 分表示人体位置预测特征图的横纵坐标, k 表示 anchor 的序号, Box_{gt} 表示真实人体位置框, Box_{pred} 表示预测人体位置框。

[0123] 人体关键点总误差公式为:

$$[0124] \quad Loss_{point} = \sum_{D,m,n,h} (x_{m,n,D,h} - \hat{x}_{m,n,D,h})^2$$

[0125] 其中, D 为不同的尺度, h 为人体关键点的序号, m 和 n 分别为人体关键点预测特征图的横纵坐标, $x_{m,n,D,h}$ 为人体关键点真实高斯值, $\hat{x}_{m,n,D,h}$ 为人体关键点预测坐标。

[0126] 本实施例中,采用均方差来计算真实人体关键点和预测人体关键点之间的误差。

[0127] 人体关键点真实高斯值根据coco数据集中标定的人体关键点坐标,采用高斯分布对人体关键点坐标进行变换获得。

[0128] 具体地,人体关键点的采用高斯分布来表示,将人体的关键点真实位置坐标的映射到不同尺度的人体关键点预测特征图,并对坐标并取整,设为;设置高斯圆的半径,设为 R ;在人体关键点预测特征图上,以为圆心,半径为 R 填充高斯函数计算值。

[0129] 通过步骤2获得训练好的人体位置关键点估计模型,通过步骤3获得待检测图像中人体的不同尺度的人体关键点预测坐标,如图4所示,步骤3具体步骤为:

[0130] 步骤31,将待检测图像输入多尺度特征提取模,输出多个尺度的特征图;

[0131] 步骤32,将特征图输入检测头模,输出多个尺度的检测头特征图;

[0132] 步骤33,将检测头特征图输入人体位置框检测模块,输出多个尺度的预测人体位置框;

[0133] 步骤34,将检测头特征图以及预测人体位置框输入感兴趣区域对齐模块,输出多个尺度的对齐特征图;

[0134] 步骤35,将对齐特征图输入人体关键点检测模块,输出多个尺度的人体关键点预测特征图;

[0135] 步骤36,在每个尺度的人体关键点预测特征图中,取每个通道中最大值的坐标作

为人体关键点预测坐标,得到不同尺度的人体关键点预测坐标。

[0136] 步骤4用于实现不同尺度的人体关键点预测坐标的融合,步骤4包括:

[0137] 步骤41,将不同尺度的人体关键点预测坐标映射回待检测图像中,得到不同尺度的人体关键点映射坐标;

[0138] 步骤42,将不同尺度的人体关键点映射坐标求均值,得到人体关键点坐标。

[0139] 具体地,步骤41,将不同尺度中的人体关键点预测坐标 $\{X_{D,i}, Y_{D,i}\}$ 映射回待检测图像中,得到不同尺度的人体关键点映射坐标 $\{X_{D,i}^{org}, Y_{D,i}^{org}\}$;

[0140] 步骤42,将不同尺度的人体关键点映射坐标 $\{X_{D,i}^{org}, Y_{D,i}^{org}\}$ 求均值,如下:

$$[0141] \quad X_i^{org} = \frac{1}{D} \sum_D X_{D,i}^{org}$$

$$[0142] \quad Y_i^{org} = \frac{1}{D} \sum_D Y_{D,i}^{org}$$

[0143] 得到最终的人体关键点坐标, $\{X_i^{org}, Y_i^{org}\}$, 其中, $i=0, \dots, 16$ 。

[0144] 17个关键点所对应的人体部位包括:0:鼻子、1:左眼、2:右眼、3:左耳、4:右耳、5:左肩、6:右肩、7:左肘、8:右肘、9:左腕、10:右腕、11:左胯、12:右胯、13:左膝、14:右膝、15:左踝、16:右踝。

[0145] 集合在密集人群的应用场景中,本实施例可实现多人的人体框和关键点同时识别,能够准确且快速地完成人体关键点的估计,无需额外的模型预先提取人体框,算法流程简便,运算量低,满足各种低算力设备的轻量化部署要求。

[0146] 第二方面,本实施例还提供一种人体关键点检测系统,包括人体关键点估计网络构建模块、网络训练模块、人体关键点坐标预测模块、人体关键点坐标融合模块;

[0147] 人体关键点估计网络构建模块用于构建人体关键点估计网络,人体关键点估计网络包括多尺度特征提取模块、检测头模块、人体位置框检测模块、感兴趣区域对齐模块和人体关键点检测模块;

[0148] 多尺度特征提取模块用于接受输入图像并输出多个尺度的特征图;

[0149] 检测头模块用于接受特征图并输出多个尺度的检测头特征图;

[0150] 人体位置框检测模块用于接受检测头特征图并输出多个尺度的预测人体位置框;

[0151] 感兴趣区域对齐模块用于接受检测头特征图、人体位置框并输出多个尺度的对齐特征图;

[0152] 人体关键点检测模块用于接受对齐特征图并输出不同尺度的人体关键点预测坐标;

[0153] 网络训练模块利用训练数据集训练人体关键点估计网络,得到人体关键点估计模型;

[0154] 人体关键点坐标预测模块,通过将待检测图像输入人体关键点估计模型,输出不同尺度的人体关键点预测坐标;

[0155] 人体关键点坐标融合模块,通过将不同尺度的人体关键点预测坐标进行融合,得到人体关键点坐标。

[0156] 需要说明的是,本实施例中的各模块(或单元)是逻辑意义上的,具体实现时,多个模块(或单元)可以合并成一个模块(或单元),一个模块(或单元)也可以拆分成多个模块(或单元)。

[0157] 本实施中的一种人体关键点检测系统实现了自动化识别且成本低廉、简单便捷、效果明显,其应用实现了在密集人群的应用场景中,能够准确且快速地实现人体关键点的估计,具有降低工作人员工作量,且提高工作效率的优势。

[0158] 本领域普通技术人员可以理解,实现上述实施例方法中的全部或部分流程是可以通程序来指令相关的硬件来完成,所述的程序可以存储于计算机可读取存储介质中,该程序在执行时,可包括如上各方法的实施例的流程。其中,所述的存储介质可为磁碟、光盘、只读存储记忆体(Read-OnlyMemory,ROM)或随机存储记忆体(Random Access Memory, RAM)等。

[0159] 以上所述仅为发明的优选实例而已,并不用于限制发明,尽管参照前述实例对发明进行了详细的说明,对于本领域的技术人员来说,其依然可以对前述各实例记载的技术方案进行修改,或者对其中部分技术特征进行等同替换。凡在发明的精神和原则之内,所做的修改、等同替换等均应包含在发明的保护范围之内。

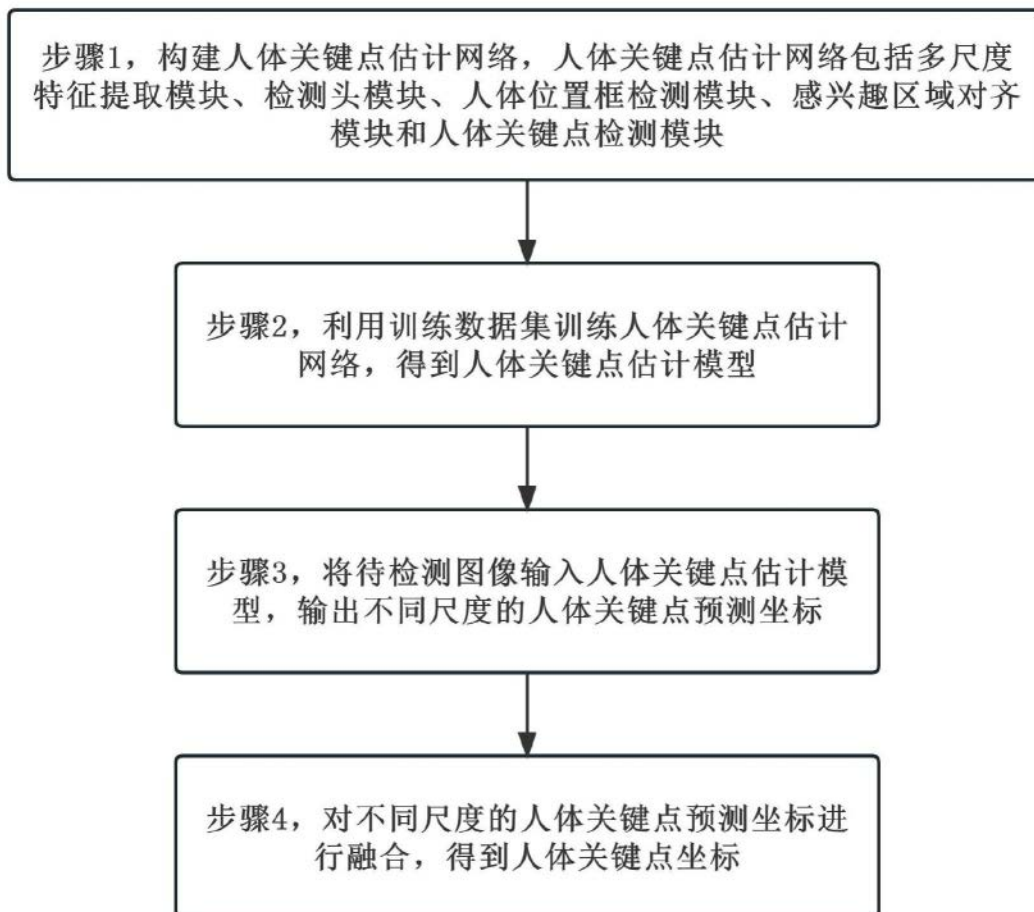


图1

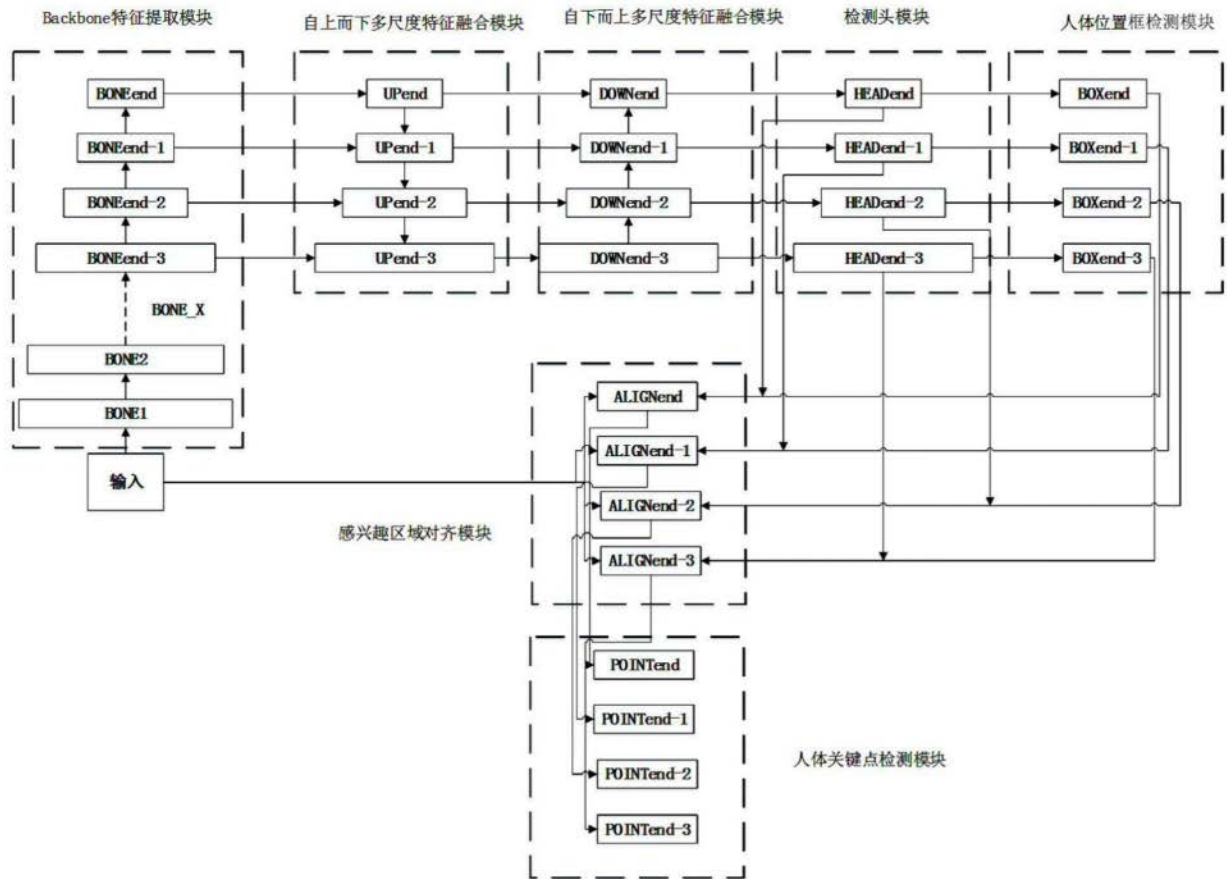


图2

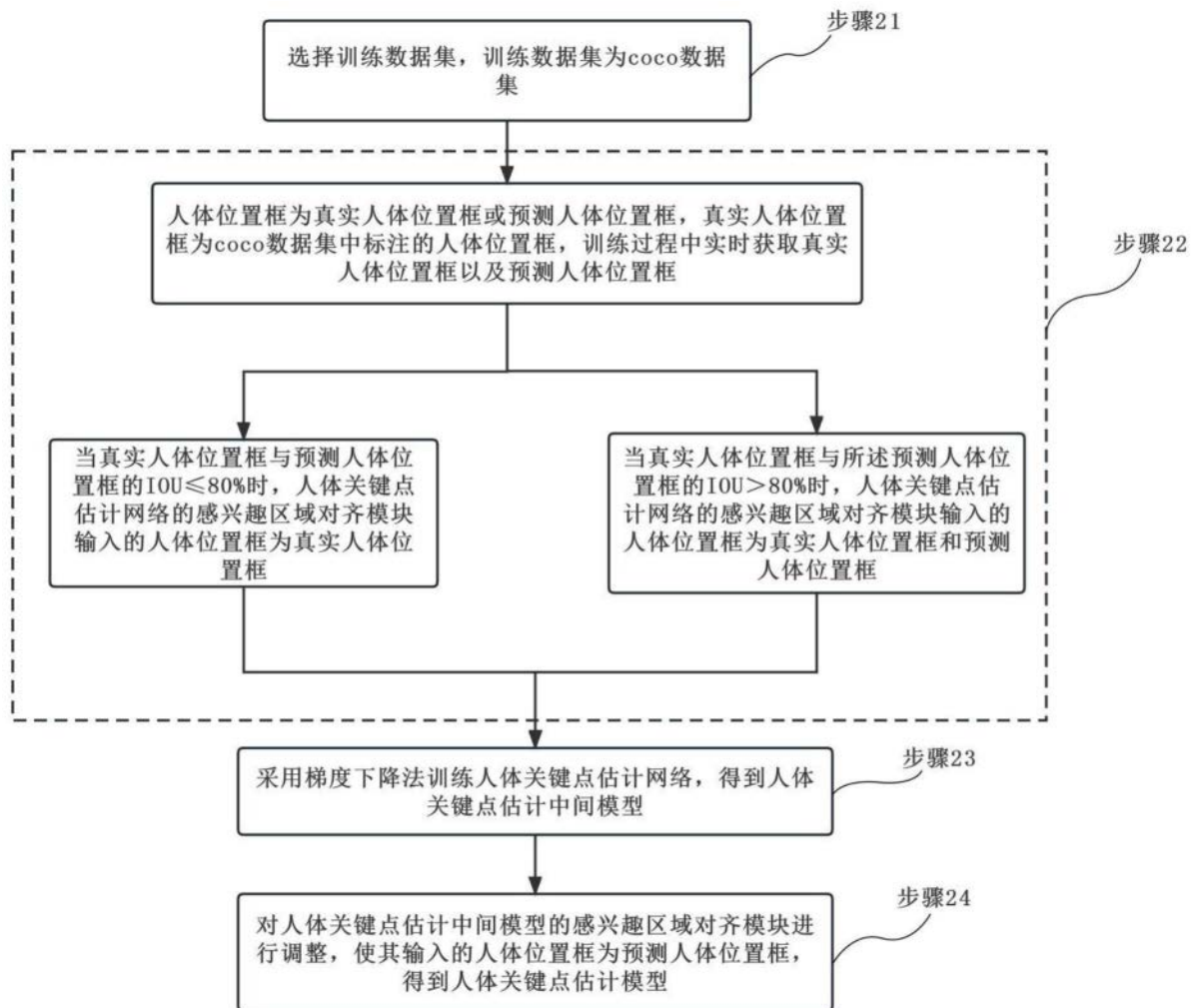


图3

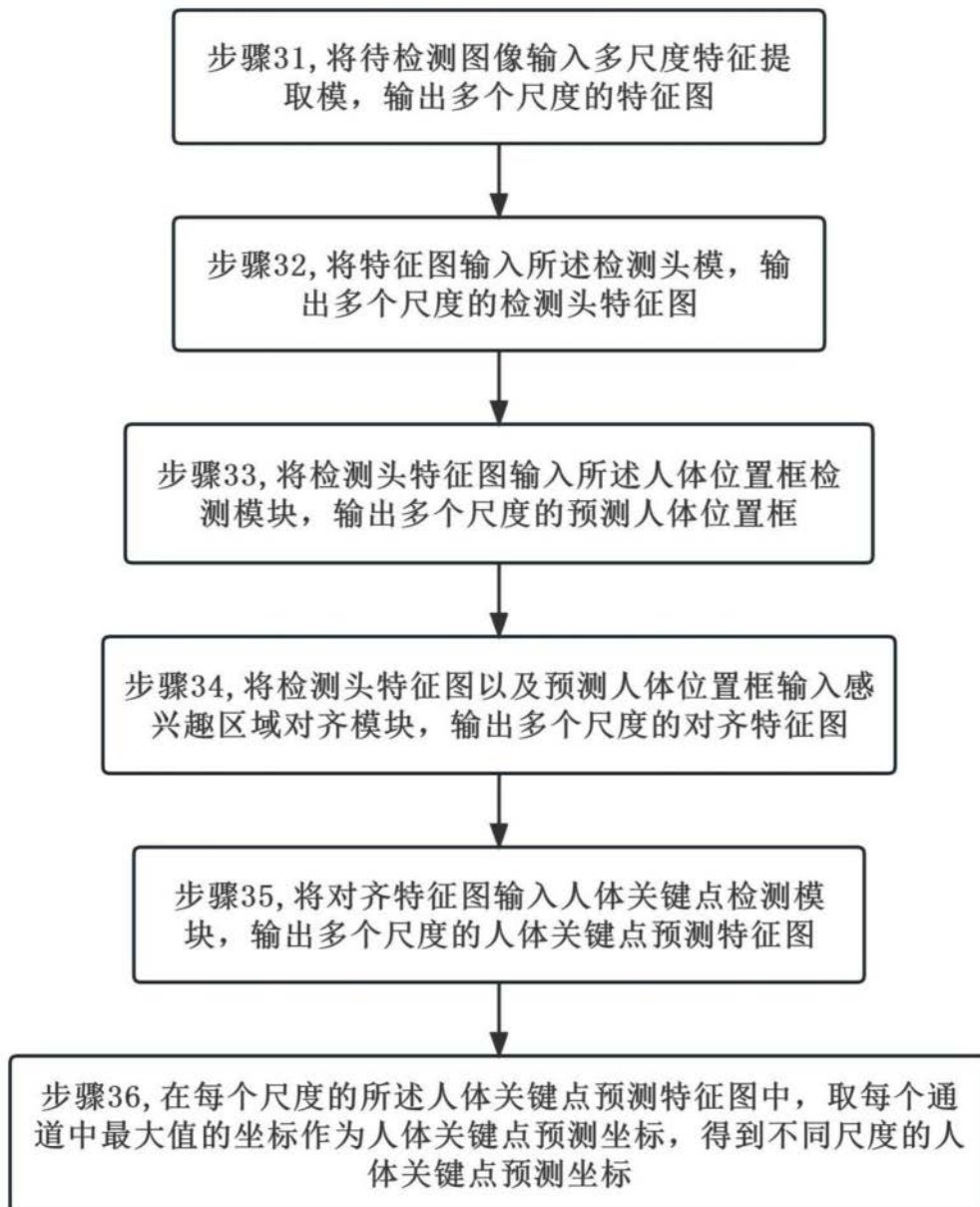


图4

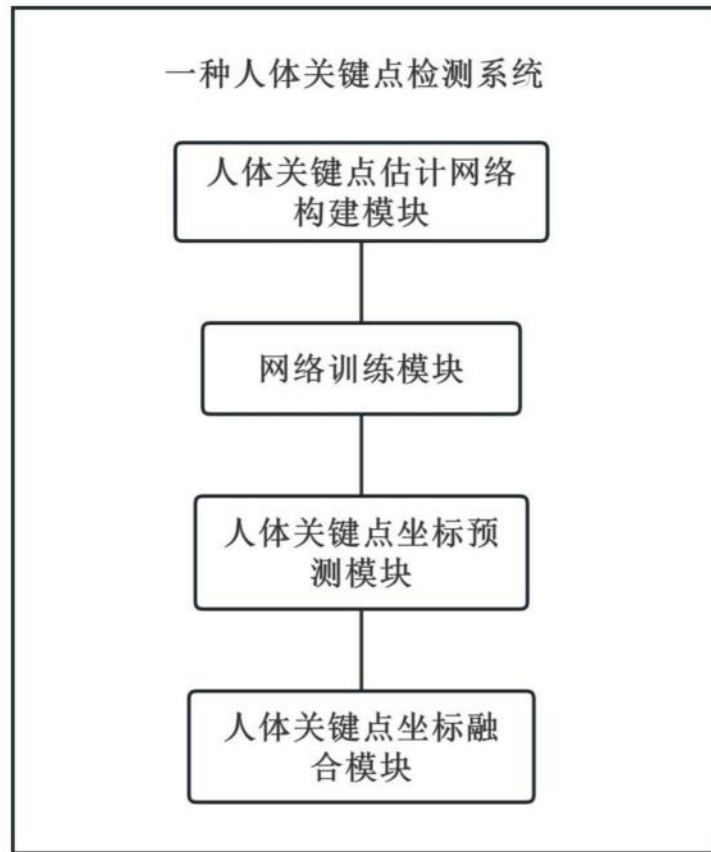


图5