基于多尺度增量学习的单人体操动作中关键点检测方法

江佳鸿,夏 楠*,李长吾,周思瑶,于鑫淼
(大连工业大学信息科学与工程学院,辽宁大连 116034)

摘 要: 人体关键点检测是计算机视觉的热点研究领域. 目前,对于体操动作关键点检测,仍存在检测精度不足及缺乏细节部位检测能力等问题. 为了提升检测精度,本文设计了一种多分辨率网络,该网络在浅层具备较大感受野,同时能够利用高分辨率通道增强细节特征的提取能力. 为实现对手部及脚部关键点的检测,设计了一种增量学习网络. 该网络融合了多分辨率网络的浅层特征并利用自建数据集计算深层特征以提升网络对手部及脚部关键点的检测能力. 最后对两个网络输出结果进行合并. 计算机仿真表明,多分辨率网络在COCO2017关键点检测数据集上达到了94.4%的准确率,并且增量学习网络能够在训练数据较少的情况下实现对细节部位关键点的准确检测.

关键词: 人体关键点检测;体操动作;多分辨率网络;增量学习;权重融合
基金项目: 教育部产学合作协同育人项目(No.220603231024713)
中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112(2024)05-1730-13
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20230729

Keypoint Detection Method for Single Person Gymnastics Actions Based on Multi-Scale Incremental Learning

JIANG Jia-hong, XIA Nan*, LI Chang-wu, ZHOU Si-yao, YU Xin-miao
(*School of Information Science and Engineering*, *Dalian Polytecnic University*, *Dalian*, *Liaoning* 116034, *China*)

Abstract: Keypoint detection of human body is a hot research area in computer vision. At present there exist some problems for keypoint detection in gymnastics actions, such as insufficient detection accuracy and lack of capability to detect detailed body parts. In order to improve the detection accuracy, this paper proposes a multi-resolution network that has a larger receptive field in the shallow layers and can utilize high-resolution channel to enhance the extraction of detailed features. To achieve the detection of keypoints of hands and feet, an incremental learning network is designed. The network fuses the shallow features of the multi-resolution network and computes deep features using a gymnastics actions self-built dataset, so that the detection ability of keypoints on hands and feet is improved. Finally, the output results of the two sub-networks are concated. Computer simulations demonstrate that the multi-resolution network achieves an accuracy rate of 94.4% on the COCO2017 keypoint detection dataset, and the incremental learning network can accurately detect keypoints of detailed body parts with fewer training data.

Key words: human keypoint detection; gymnastics actions; multi-resolution network; incremental learning; weight fusion

Foundation Item(s): Ministry of Education Industry-University Cooperation and Collaborative Education Project (No.220603231024713)

1 引言

单人体操运动是诸多体育赛事的主流比赛项目. 将体操动作姿态评估技术作为辅助手段引入到比赛当中,对于提升评判准确性具有重要的意义[1~4]. 关键点检测作为姿态估计中的一项核心技术,虽然目前已经取得了很大的进展,但将其应用于体操动作姿态估计

收稿日期:2023-08-03;修回日期:2024-02-24;责任编辑:朱梅玉
*通讯作者:夏楠

中仍存在较多具有挑战性的问题有待解决,主要有以下两方面:其一是关键点检测精度有待提升,其二是对手部及脚部关键点检测不充分.

在人体关键点检测的经典算法中,文献[5]设计了一种双阶段网络,在一阶段网络中优先检测特征明显的关键点,二阶段网络通过整合来自一阶段网络的多级别特征来识别其余关键点. 文献[6]介绍的参数化姿态非最大抑制的方法,一定程度上解决姿态冗余检测问题. 文献[7]提出的Hourgalss网络通过结合了不同尺度关键点的特征提高了检测精度. 文献[8]提出一种高分辨率网络,该网络结合多尺度融合方法提高了不同尺度特征图的特征信息相互传递的检测精度. 文献[9~12]均在不同程度上实现了人体关键点检测. 近年来,有许多学者对上述经典算法进行了一定程度上的改进并在此基础上提出了新的关键点检测算法. 如文献[13]提出了一种在全局特征中提取出实例特征的方法,提高了单人关键点检测精度. 文献[14]提出自适应权重热力图回归的损失函数以解决不同尺度对象所对应的感受野不同产生的问题. 文献[15]提出了一种基于序列多尺度特征融合的方法对High-Resolution Net(HRNet)进行改进提高了关键点检测精度. 文献[16]提出了一种使用三维卷积模块,以类似沙漏结构的形式组合设计并且使用残差模块并联的方式对特征图进行监控融合,对人体关键点检测精度有所提升. 文献[17]于网络浅层设计了一个特征共享模块来融合不同尺度图片的信息进而提升检测精度. 文献[18]提出了一种通过粗略姿态过滤来改善多人体姿态估计的方法,通过首先估计每个候选区域的粗略姿态,然后使用这些估计值来过滤和优化最终的姿态估计结果,从而提高了多人体姿态估计的准确性. 文献[19]提出了一种通过学习来获取人体姿态估计质量的方法,通过将姿态估计任务视为一个多任务学习问题,同时学习估计准确性和姿态质量,通过多任务之间的相互促进从而提高了人体姿态估计的准确性. 上述文献虽然在一定程度上实现了对人体关键点的检测,但由于关键点检测本质上为像素级坐标回归及关键点分类任务,需要充分结合全局语义信息以及关键点细节特征信息[20~22]. 上述算法存在的不足之处为网络特征图分辨率较低,关键点的细节特征不易于被网络提取,且网络输入单一,浅层无法在不同尺度提取全局语义信息.

对体操动作的关键点检测中,在提高检测精度的同时,实现对关键点的全面检测同样很重要. 现有关键点检测算法只具备人体躯干及四肢关键点的检测能力,无法准确地对体操动作的规范性进行评估. 因此许多学者逐渐把研究重心转向了对细节部位关键点的补充检测. 针对这一问题,文献[23]为更好地分析拉丁舞动作,通过人工标注脚掌及手掌数据集实现了对上述部位的补充检测,提高了舞蹈姿态估计效果. 但这种人工标注的方法不但误差较大,而且训练效率低下并耗费人力及时间成本. 增量学习是一种利用现有知识扩展模型对新知识的学习能力的方法,该方法通过添加新的样本或类别对现有模型进行更新,使其掌握新知识[24]. 这一方法可以很好地应用到关键点检测算法中实现对新关键点的检测. 因此,文献[25]在对脸部关键点进行聚类时结合增量学习的方法,在原有网络中新增加了关键点,提升了聚类效果并且具备对新知识的学习能力,但其新增关键点与原始关键点特征相似度较低,降低了网络对原始关键点的检测能力.

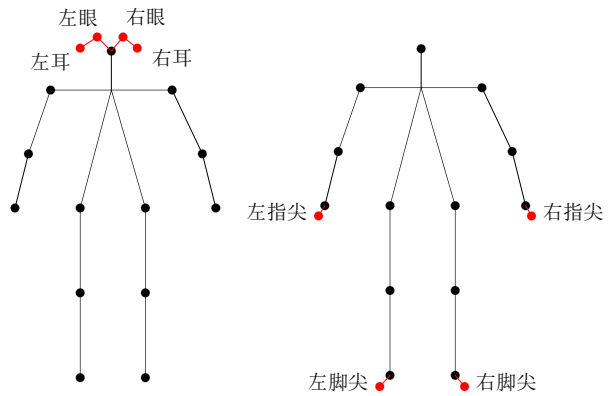
本文旨在于体操动作中提高关键点检测精度以及在训练数据不足的情况下实现对手部及脚部关键点的补充检测. 首先为了提升检测精度,本文以HRNet网络为基础提出了一种多分辨率网络. 该网络能在两种不同分辨率的输入上分别进行不同尺度特征提取,并通过多尺度融合方法将不同尺度特征进行融合,以此使网络浅层兼具全局信息及细节信息的提取能力. 同时于该网络中,提出了一种高分辨率通道,该通道能够生成高分辨率特征图并在此基础上充分提取关键点细节信息,以此提升网络关键点的检测能力. 其次为了利用有限的自建训练数据集实现对手部及脚部关键点的高精度补充检测,本文设计了一个双阶段网络并基于增量学习的思想提出了一种权重融合方法. 其中,一阶段网络为人体躯干及四肢检测网络,且有充足训练数据训练该网络权重;二阶段网络为手部及脚部关键点检测网络,仅需利用较少的训练数据. 权重融合方法旨将一阶段网络中浅层权重融入到二阶段网络中作为先验知识. 由于神经网络浅层提取宽泛特征,深层提取细节特征[26],而躯干四肢以及手部和脚部关键点均为人体关键点,它们之间宽泛特征具有很强的相似性[27,28]. 因此通过权重融合能够使二阶段网络在无需训练的前提下具备手部及脚部关键点宽泛特征的提取能力. 进而在此基础上利用少量数据训练实现对手部及脚部关键点的检测. 为验证提出方法的有效性,本文分别在COCO2017关键点检测数据集[29]、Halpe-FullBody数据集[30]以及自建体操动作数据集上进行了对比实验、消融实验以及一系列可视化效果展示,均取得了良好的效果,验证了本文方法的有效性.

2　相关工作

2.1　体操动作关键点检测

人体关键点检测技术被广泛用于评估人体姿态动作的规范性. 但传统的人体关键点检测方法通常只关注躯干及四肢关键点,其人体关键点几何关系如图1(a)

所示，这些方法在评估体操动作时无法检测更加细节的关键点，进而无法更详细地评估体操动作的规范性．为解决这一问题，文献[31]在进行人体行为分析与步态识别研究中，增加了脚掌及躯干若干个关键点的检测，实现了更详细的步态估计效果．文献[32]在分析运动员运动姿态时补充了脚部关键点以便教练能够更好地对运动员进行姿态纠正．文献[33]在设计用于人机交互的姿态估计系统时加入了手部对系统的控制，实现了更细节的互动效果．文献[34]设计了一个三通道网络，实现了对脸部、身体、手部关键点的联合检测．上述算法虽然通过对细节部位关键点的补充检测实现了更好的姿态估计效果，但均通过手工标注的方法获得训练数据，因此效率低下．除此之外，不同的体操运动所关注的细节人体关键点也不同，因此该领域缺少统一并充足的训练数据．为了改善训练数据不足的问题，文献[35]提出了一种基于局部深度一致性的自监督姿态估计方法，通过在虚拟数据集上预训练网络，并在无标注的真实数据集上进行模型拟合，一定程度上解决了数据不足的问题．文献[36]为解决训练数据不足的问题引入增量学习思想，在大型数据集上进行预训练并在此基础上对新知识进行学习．但这两种方法在虚拟数据集或大型数据集上获得的先验知识，与自身任务的关联性较低，因此有待改进．结合上述算法和实际体操姿态估计的需求，在体操动作中的关键点检测需要在检测躯干及四肢关键点的同时还应准确地检测出手部及脚部的关键点，以便更全面地实现对体操动作规范性的评估．同时应该相对弱化脸部关键点的检测以避免姿态的冗余估计，重点关注身体、四肢、手部和脚部的姿态．因此在体操动作中需要被关注的关键点如图1(*b*)所示．



(*a*) 传统动作 (*b*) 体操运动

图1 姿态估计关键点几何关系图

2.2 多尺度融合

HRNet[8]由4个并行的多分辨率子网络构成，每个子网络采用ResNet模块设计原则，由4个残差单元组成．由于其具备较好的提取输入图像的多分辨率特征以及特征表示能力，在目标检测、识别、图像分割以及人体关键点检测中可获得较好的结果．在HRNet中有多种不同分辨率大小特征图，其中高分辨率特征图蕴含细节信息较多，低分辨率特征图可提供丰富全局信息[8]．多尺度融合方法可以充分地将不同尺寸特征图进行融合，并于不同尺度生成融合后兼具全局信息与细节信息的特征图[37~39]，其表达式为[8]

$$\begin{array}{l}\boldsymbol{C}_{3,1}^{1}\searrow\quad\nearrow\boldsymbol{C}_{3,1}^{2}\searrow\quad\nearrow\boldsymbol{C}_{3,1}^{3}\searrow\\ \boldsymbol{C}_{3,2}^{1}\rightarrow\boldsymbol{\varepsilon}_{3}^{1}\rightarrow\boldsymbol{C}_{3,2}^{2}\rightarrow\boldsymbol{\varepsilon}_{3}^{2}\rightarrow\boldsymbol{C}_{3,2}^{2}\rightarrow\boldsymbol{\varepsilon}_{3}^{3}\\ \boldsymbol{C}_{3,3}^{1}\nearrow\quad\searrow\boldsymbol{C}_{3,3}^{3}\nearrow\quad\searrow\boldsymbol{C}_{3,3}^{3}\nearrow\end{array}\tag{1}$$

其中，矩阵$\boldsymbol{C}_{s,r}^{b}$为在HRNet网络的阶段s中第b块分辨率为r的特征图；$\boldsymbol{\varepsilon}_{s}^{b}$为相对应的融合单元．多尺度融合示意图如图2所示，不同分辨率的特征图为分别通过不同的上采样及下采样的组合于不同分辨率生成融合后的特征图．该特征图融合不同分辨率特征图中的特征信息，提升网络的多尺度感知能力．多尺度融合模块中并行处理不同分辨率下的特征图融合操作．多尺度融合后的融合输出为

$$\boldsymbol{Y}_{k}=\sum_{i=1}^{s}a\left(\boldsymbol{X}_{i},k,\text{up/down/none}\right)\tag{2}$$

其中，a为常数系数；$\boldsymbol{X}_{i}$为输入特征矩阵；s为并行特征图数目；up为采用最邻近上采样，代表将输入矩阵$\boldsymbol{X}_{i}$的分辨率从i提高到k；down为采用步长为2的3×3卷积进行下采样，代表将输入矩阵$\boldsymbol{X}_{i}$的分辨率从i降低到k；none为不采取任何操作用于相同尺寸特征图的整合，上下采样均可以通过连续卷积增加或降低特征图的尺寸；$\boldsymbol{Y}_{k}$代表输出特征图．

3 本文方法

3.1 基于增量学习的双阶段网络

为在不影响躯干及四肢关键点检测精度前提下，仅利用少量训练数据实现对手部及脚部关键点的补充检测，本文基于增量学习的思想，提出了一种双阶段网络，其示意图如图3所示．其中，一阶段网络为人体躯干及四肢关键点检测网络，且具备充足训练数据；二阶段网络为手部及脚部关键点补充网络，且训练数据不足．一阶段网络经过充足数据训练后，将其浅层网络权重通过权重融合方法融入二阶段网络对应部分，使一阶段网络对躯干及四肢关键点宽泛特征的提取能力作为二阶段网络的先验知识．之后在二阶段网络中，利用自建体操动作数据集对该部分权重进行微调，使网络具备对手部及脚部关键点的检测能力，并在此基础上对深层权重进行训练．最后将两个子网络在通道维度上进行Concat拼接实现人体躯干及四肢、手部及脚部关键点的联合检测．为提升关键点检测精度，本文设计

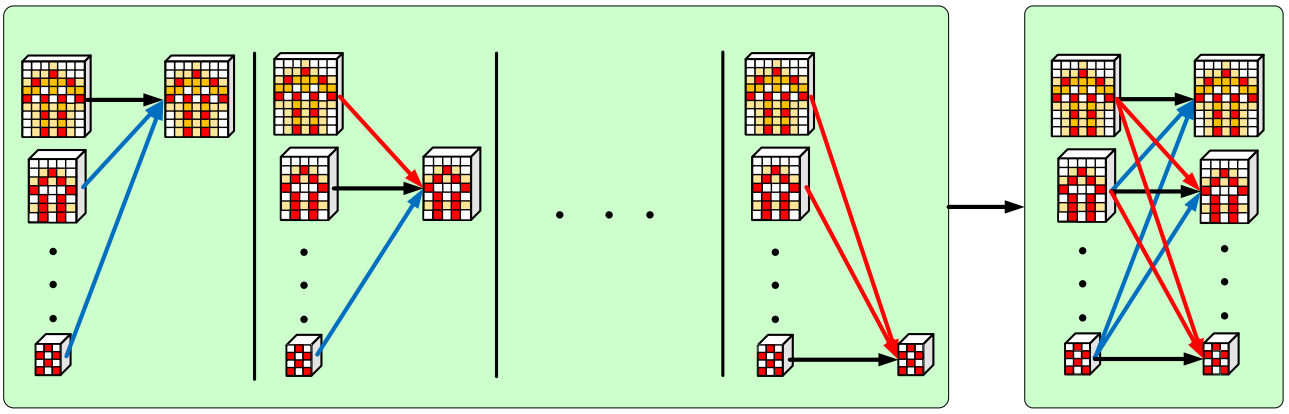


图2　多尺度融合示意图

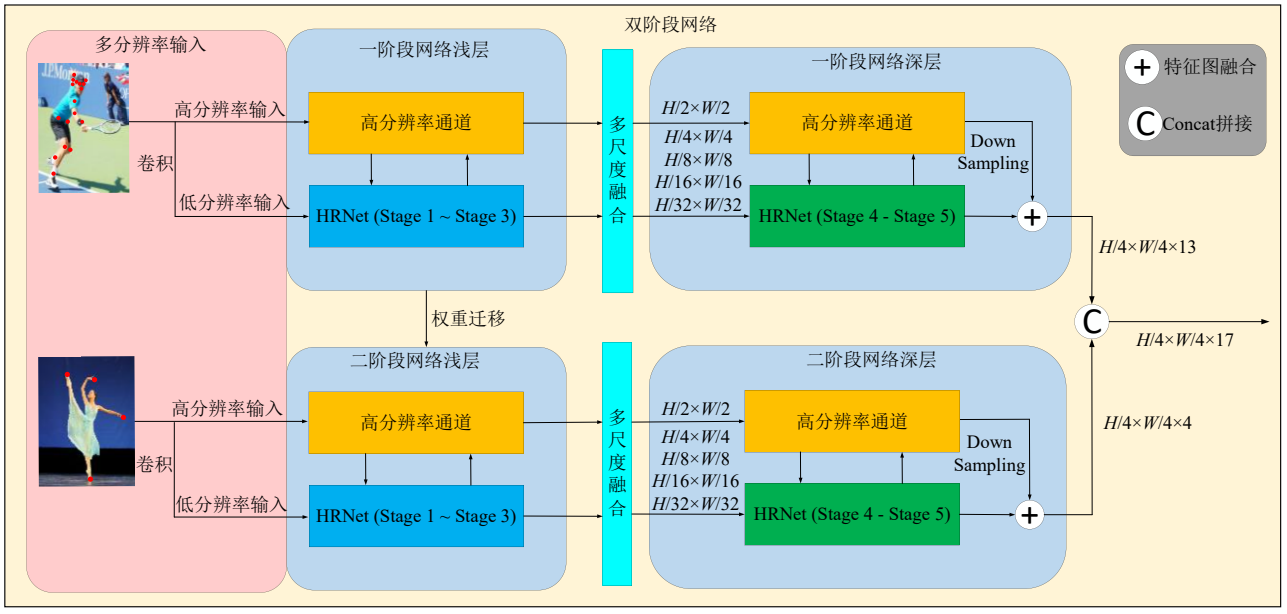


图3　基于增量学习的双阶段网络

了一种多分辨率网络并将其应用于一阶段网络与二阶段网络中. 多分辨率网络与权重融合方法将在第3.2节及第3.3节详细叙述.

3.2　多分辨率网络

为提升关键点的检测精度,本文基于HRNet设计了一种多分辨率网络. 该网络有两点改进,分别为多分辨率输入及高分辨率通道. 其示意图如图4所示,其中黄色网络结构代表高分辨率通道,用于接收高分辨率图片输入,蓝色部分为HRNet网络[8],用于接收低分辨率输入.

传统的关键点检测网络通过逐层堆叠卷积层的操作来获取全局特征信息,因此在网络深层才具备较好的全局特征提取能力,而浅层只能提取像素级别细节特征. 良好的关键点检测结果需要细节特征与全局特征共同参与预测[8]. 为使网络在浅层具有全局特征的提取能力,本文对HRNet进行改进设计了一种多分辨率输入方法,即将传统网络的单输入改为两种不同分辨率的双输入. 其中,高分辨率输入为分辨率大小为256×192的输入图片,低分辨率输入为高分辨率输入经过卷积后形成的分辨率为128×96的特征图. 该方法可以使网络浅层于高分辨率输入中提取细节信息并且于低分辨率输入中提取全局信息,提升网络的跨分辨率学习能力. 具体地,由于低分辨率输入尺寸较小,虽然大量细节特征被减弱,但更多展现关键点之间的拓扑结构关系,即网络在相同卷积核大小下,可以在低分辨率特征图中提取较多结构关系特征. 因此增加低分辨率输入使网络在浅层具有良好的全局信息提取能力,可以充分推理各个关键点之间的关系,从而网络的检测能力.

随着卷积层的堆叠特征图分辨率逐渐降低,虽然会增大感受野提升全局信息的提取能力,但会使输入图片中的细节信息逐渐丢失进而影响网络对细节信息

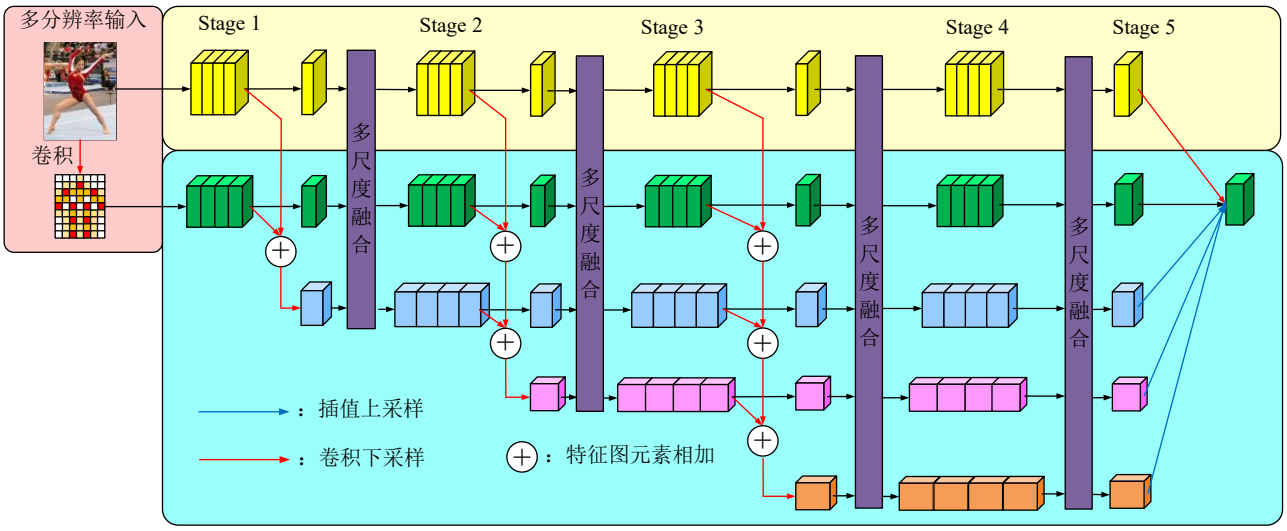


图4 多分辨率网络结构图

的提取，关键点检测任务要求网络对细节信息的提取能力较高．因此本文设计了一组高分辨率通道，该通道通过生成高分辨率特征图以增强网络对细节特征的捕获能力．在HRNet中有以下四种不同尺度大小特征图，$\boldsymbol{F}_1 \in \mathbb{R}^{C\times(H/4)\times(W/4)}$，$\boldsymbol{F}_2 \in \mathbb{R}^{2C\times(H/8)\times(W/8)}$，$\boldsymbol{F}_3 \in \mathbb{R}^{4C\times(H/16)\times(W/16)}$，$\boldsymbol{F}_4 \in \mathbb{R}^{8C\times(H/32)\times(W/32)}$．$H\times W$代表输入图片的分辨率，$C$=32代表特征图通道数．由于关键点检测为像素级坐标回归任务，其对关键点细节信息的要求较高．本文高分辨率通道能够生成特征图$\boldsymbol{F}_{\text{high}} \in \mathbb{R}^{(C/2)\times(H/2)\times(W/2)}$．该分辨率特征图中每个像素点对应的特征向量表达的输入图像中信息更加详细，因此网络能够提取更多、更详细的特征信息．同时在多尺度融合模块中，高分辨率通道与HRNet原始分辨率特征图共同进行信息融合，使网络在更加丰富的尺度上感知特征信息，进而提升网络检测精度，多尺度融合操作示意图如图2所示．

准确的关键点检测结果需要网络具备良好的全局信息与细节信息的提取能力．本文网络设计的高分辨率通道以及多分辨率输入提升了网络细节信息及全局信息提取能力．网络结构参数及后处理操作如表1所示，在图4中同一尺寸大小的特征图（相同颜色）为1路，同1路中每一个特征图为1层，例如第1路第（1，2）层为第一路的第一层及第二层．Down Sample为下采样，Up Sample为上采样，例如，Down Sample×m代表下采样m倍，Up Sample×k代表上采样k倍．

3.3 双阶段网络中权重融合方法

与传统人体关键点检测任务不同，体操动作关键点检测在关注躯干及四肢部位的同时更加关注手部和脚部的关键点．但针对这些部位的检测缺少充足的训练数据，大多算法采用人工自制数据集的方法生成数据[31~34]．这种方法不仅耗费时间成本及人力成本且易产生误差．一些基于增量学习的方法[35,36]在一定程度上弥补了训练数据不足的问题，但其融合的权重与自身任务相似度较低，因此需要很大程度的调整．为改善这一问题，本文提出了一种权重迁移的方法，即将一阶段网络中与二阶段网络高度适配的网络权重迁移到二阶段网络中来代替原始初始化的权重，使其作为二阶段网络的先验知识．具体地，网络实现精确的关键点检测需通过训练最小化其损失函数$L(\boldsymbol{w})$使网络达到收敛状态．损失函数$L(\boldsymbol{w})$如下式所示

$$L(\boldsymbol{w}) = \sum_{i=1}^{n} \left\| \boldsymbol{f}\left(\boldsymbol{w}, \boldsymbol{x}_i\right) - \boldsymbol{y}_i \right\|^2 \tag{3}$$

其中，$\boldsymbol{f}\left(\boldsymbol{w}, \boldsymbol{x}_i\right)$代表网络利用整体网络的权重$\boldsymbol{w}$对输入图片$\boldsymbol{x}_i$经过正向传播推理关键点坐标预测值；$\boldsymbol{y}_i$代表关键点坐标真实值；$n$代表图片数量．在一阶段网络中，降低$L\left(\boldsymbol{w}_1\right)$的过程需要大量训练数据对网络多次迭代训练，$\boldsymbol{w}_1$代表一阶段网络的整体网络权重．显然一阶段网络满足该条件，因此$L\left(\boldsymbol{w}_1\right)$已达到最小且网络浅层权重具备检测人体关键点宽泛特征的能力．因训练数据的不足，二阶段网络权重$\boldsymbol{w}_2$没有达到最优，其损失函数$L\left(\boldsymbol{w}_2\right)$无法达到其收敛值．在神经网络中，浅层网络能够提取物体宽泛特征，深层网络提取细节特征[40]．由于人体部位关键点宽泛特征具有很强的相似性，因此一阶段网络浅层权重同样可以被用于二阶段网络来提取手部及脚部关键点的宽泛特征．因此将一阶段网络Stage 1~Stage 3权重融合到二阶段网络中对应阶段可使二阶段网络具备检测手部及脚部关键点宽泛特征的先验能力，即使其损失函数初始值在更接近最小值处．

为证明双阶段网络间权重融合方法的有效性，给出以下证明．$\boldsymbol{w}_1$经过大量数据训练已达到最优即$L\left(\boldsymbol{w}_1\right)$达到最小．为证明$\boldsymbol{w}_1$可以高度适配二阶段网络，

表1　多分辨率网络参数及后处理操作

阶段	网络层级	特征图尺寸(长×宽×通道)	后处理
Stage 1	第1路 第(1,2)层	128×96×16	无
	第1路 第3层	128×96×16	Down Sample×4
	第1路 第4层	128×96×16	多尺度融合
	第2路 第(1,2)层	64×48×32	无
	第2路 第3层	64×48×32	Down Sample×2
	第2路 第4层	64×48×32	多尺度融合
	第3路 第1层	32×24×64	多尺度融合
Stage 2	第1路 第(1,2,3)层	128×96×16	无
	第1路 第4层	128×96×16	Down Sample×8
	第1路 第5层	128×96×16	多尺度融合
	第2路 第(1,2,3)层	64×48×32	无
	第2路 第4层	64×48×32	Down Sample×4
	第2路 第5层	64×48×32	多尺度融合
	第3路 第(1,2,3)层	32×24×64	无
	第3路 第4层	32×24×64	Down Sample×2
	第3路 第5层	32×24×64	多尺度融合
	第4路 第1层	16×12×128	多尺度融合
Stage 3	第1路 第(1,2,3)层	128×96×16	无
	第1路 第4层	128×96×16	Down Sample×16
	第1路 第5层	128×96×16	多尺度融合
	第2路 第(1,2,3)层	64×48×32	无
	第2路 第4层	64×48×32	Down Sample×8
	第2路 第5层	64×48×32	多尺度融合
	第3路 第(1,2,3)层	32×24×64	无
	第3路 第4层	32×24×64	Down Sample×4
	第3路 第5层	32×24×64	多尺度融合
	第4路 第(1,2,3)层	16×12×128	无
	第4路 第4层	16×12×128	Down Sample×2
	第4路 第5层	16×12×128	多尺度融合
	第5路 第1层	8×6×256	多尺度融合
Stage 4	第1路 第(1,2,3)层	128×96×16	无
	第1路 第4层	128×96×16	多尺度融合
	第2路 第(1,2,3)层	64×48×32	无
	第2路 第4层	64×48×32	多尺度融合
	第3路 第(1,2,3)层	32×24×64	无
	第3路 第4层	32×24×64	多尺度融合
	第4路 第(1,2,3)层	16×12×128	无
	第4路 第4层	16×12×128	多尺度融合
	第5路 第(1,2,3)层	8×6×256	无
	第5路 第4层	8×16×256	多尺度融合
Stage 5	第1路 第1层	128×96×16	Down Sample×2
	第2路 第1层	64×48×32	无
	第2路 第2层	64×48×32	无
	第3路 第1层	32×24×64	Up Sample×2
	第4路 第1层	16×12×128	Up Sample×4
	第5路 第1层	8×6×256	Up Sample×8

将二阶段网络损失函数$L(\boldsymbol{w}_2)$在$\boldsymbol{w}_1$处进行泰勒级数展开，即

$$L(\boldsymbol{w}_2)\approx L(\boldsymbol{w}_1)+\Delta\boldsymbol{w}\cdot\nabla L(\boldsymbol{w}_1) \tag{4}$$

其中，$\Delta\boldsymbol{w}=(\boldsymbol{w}_2-\boldsymbol{w}_1)$；$\nabla$代表求梯度．对于二阶段网络Stage 1~Stage 3，网络权重$\boldsymbol{w}_2$融入一阶段权重$\boldsymbol{w}_1$，即$\Delta\boldsymbol{w}\approx 0$，由式(4)可得$L(\boldsymbol{w}_2)\approx L(\boldsymbol{w}_1)$．又因为两个子网络任务的相似程度较高，因此二阶段网络Stage 1~Stage 3接近收敛，即具备人体关键点宽泛特征的提取能力．对于Stage 4、Stage 5，该部分网络权重$\boldsymbol{w}_2$为随机初始化，由式(4)有$L(\boldsymbol{w}_2)\neq L(\boldsymbol{w}_1)$，因此$\boldsymbol{w}_2$无法使该部分网络收敛．但由于反向传播理论中权重更新是逐层传递的，因此相比直接训练的方法，二阶段网络Stage 1~Stage 3由于融入了$\boldsymbol{w}_1$，能够使Stage 4-Stage 5中权重$\boldsymbol{w}_2$在此基础上进行更新，因此该部分网络检测能力及收敛速度均优于直接训练的方法．综上所述，本文权重融合方法将一阶段网络Stage 1~Stage 3权重融入二阶段网络对应部分，使后者具备良好的人体关键点宽泛特征的提取能力．在此之后，使用少量手部及脚部训练数据对二阶段网络Stage 1~Stage 3权重进行微调，提升对手部及脚部关键点特征的专注力，并在此基础上对Stage 4、Stage 5权重进行训练，使网络达到收敛．最后将两个子网络的特征图输出在通道维度上进行Concat拼接，整体双阶段网络实现人体躯干、四肢以及手部和脚部关键点的联合检测．其中Concat拼接公式为

$$\boldsymbol{Z}_c=\sum_{i=1}^{c_1}\boldsymbol{y}_1^i*\boldsymbol{w}_1^i+\sum_{j=1}^{c_2}\boldsymbol{y}_2^j*\boldsymbol{w}_2^j \tag{5}$$

其中，$\boldsymbol{y}_1^i$代表一阶段网络输出；$\boldsymbol{y}_2^j$代表二阶段网络输出；c_1、c_2分别代表一阶段网络及二阶段网络输出特征图的通道数；$\boldsymbol{w}_1^i$、$\boldsymbol{w}_2^j$分别代表一阶段网络与二阶段网络输出层卷积核；*代表卷积运算；$\boldsymbol{Z}_c$代表拼接后的输出．经过Concat拼接后整个网络能够检测人体躯干和四肢关键点以及手部和脚部关键点．

3.4 双阶段网络训练及测试流程

双阶段网络的训练过程分为两个阶段，其流程图如图5所示．首先利用COCO2017关键点检测数据集对一阶段网络进行训练，该数据集标注了人体躯干及四肢部位共17个关键点，但由于体操动作不关注脸部细节关键点，因此将眼睛及耳朵部位关键点训练权重置为0，只保留鼻子部位关键点作为脸部关键点．一阶段网络训练结束后，加载其浅层网络权重并融入二阶段网络浅层．之后使用自建体操动作数据集对二阶段网络浅层权重进行微调并在此基础上对深层网络进行训练．最后将两个子网络于通道维度上进行拼接．

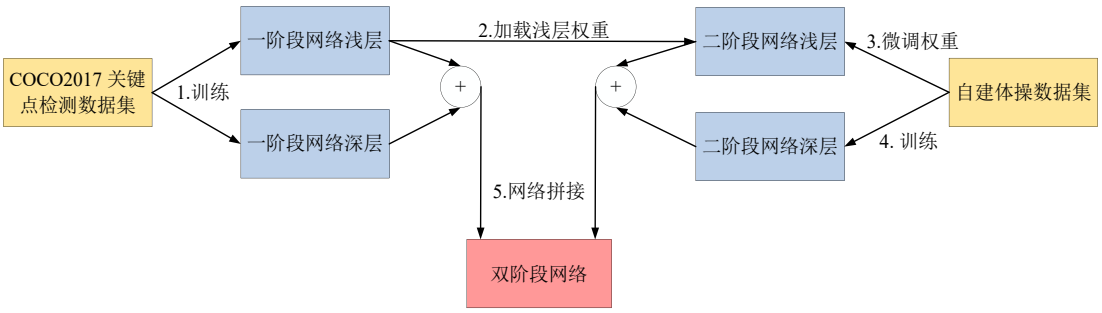


图5 双阶段网络训练流程图

该网络测试过程为一个端到端的阶段．输入为两张不同分辨率单人体操动作图片，其中低分辨率输入为高分辨率输入经过3×3卷积下采样获得．双输入被同时输入到一阶段网络与二阶段网络中分别进行不同位置关键点的检测．最后将两个子网络的特征图输出在通道维度上进行Concat拼接，实现对人体躯干四肢及手部和脚部关键点的联合检测．

4 实验结果分析与可视化展示

4.1 实验数据、对比算法及评价指标

为验证本文算法的有效性，在COCO2017关键点检测数据集、Halpe-FullBody数据集及自建体操动作数据集上对所提出网络进行对比分析．COCO2017关键点检测数据集共标注了人体躯干及四肢共17个关键点，其中包含57 000训练数据，5 000验证数据以及20 000的测试数据．Halpe-FullBody数据集共标注了人体躯干及四肢、脸部、手部及脚部共136个关键点，其中，68个人脸关键点，42个手部关键点，20个人体躯干关键点以及6个脚部关键点．其中包含50 000训练数据，5 000验证数据．本文自建体操动作数据集共标注了4个关键点，包括左右指尖2个关键点及左右脚尖2个关键点，共标注了20 000张图片，其中18 000张用于训练，2 000用于验证．将本文提出的算法与其他主流的算法进行对比分析，采用OKS(Object Keypoint Similarity)作

为定量评价指标[8]. OKS是一种通过比较预测关键点和真实关键点之间的距离来评估目标检测和关键点定位性能的指标. 其计算公式为

$$\mathrm{OKS}=\frac{\sum_i \exp\left(-d_i^2/2s^2k_i^2\right)\delta\left(v_i>0\right)}{\sum_i \delta\left(v_i>0\right)} \tag{6}$$

其中,i代表关键点下标;d_i代表预测关键点与真实关键点之间的欧氏距离,该距离衡量预测关键点位置与真实关键点位置的差异程度;v_i代表关键点可见性标志,其值为0、1或2,分别表示未标记、已标记但不可见或可见且已标记;$\delta(v_i>0)$用于判断关键点是否可见;s表示尺度参数,用于控制欧氏距离在计算中的影响程度;k_i为关键点衰减常数.

4.2　实验设置

本文在python 3.8的Pytorch框架上构建软件仿真平台. 硬件平台为一台Windows 11系统计算机,一块NVIDIA GeForce RTX 3060显卡. 网络共有两种不同分辨率输入,其中高分辨率输入大小为256×192,低分辨率输入为高分辨率输入经过一个卷积核大小为3×3、步长为2的卷积操作对其进行下采样获得,其分辨率大小为128×96. 同时对输入图像进行图像增强操作,包括随机翻转、±45°的旋转、±35%的尺度缩放. 本文实验使用Adam作为优化器,初始学习率设置为0.001,学习率衰减倍数为0.1,分别在第170轮及第200轮进行学习率衰减. 网络共训练210轮.

4.3　对比实验

为验证本文设计的多分辨率网络在人体关键点检测及体操动作关键点检测中是否具备良好的性能,在COCO2017关键点检测数据集、Halpe-FullBody数据集及自建体操动作数据集上按照OKS指标与其他对比算法进行对比实验. 表2为在COCO2017关键点检测数据集上各项指标的对比结果. 通过表2可以得出本文网络的AP高于所有对比算法. 仅仅AP.L低于文献[18]的1.9%,其余指标均高于该算法. 为了更加全面地验证本文网络的性能,在Halpe-FullBody数据集上分别对全身、脸部、手部、脚部和肢体分别进行了效果评估并与其他算法进行了对比,表3为网络指标即对比结果. 通过表3可得出本文网络在全身、脚部、手部及脸部均取得了最好的效果. 综上所述,本文网络在COCO2017关键点检测数据集、Halpe-FullBody数据集上均取得了具有竞争力的效果,即具有更高的检测精度. 其原因是本文的多分辨率输入提升了网络浅层的感受野,进而使其可以获得人体关键点之间的结构关系. 除此之外,本文高分辨率通道生成的$\boldsymbol{F}_{\mathrm{high}}$特征图中蕴含更丰富的细节信息,使关键点特征更加详细,因此能够更加准确地描述关键点的特征.

表2　本文网络与其他算法在COCO2017关键点检测数据集的对比结果　　单位:%

对比算法	AP	AP.5	AP.75	AP.M	AP.L	AR
文献[5]	72.1	91.4	80.0	68.7	77.2	78.5
文献[6]	72.3	89.2	79.1	68.0	78.6	—
文献[7]	65.5	86.8	72.3	60.6	72.6	70.2
文献[8]	74.4	90.5	81.9	70.8	81.0	79.8
文献[11]	70.5	89.3	77.2	66.6	75.8	74.9
文献[13]	68.9	89.9	76.0	63.2	77.7	74.5
文献[14]	72.0	90.7	78.8	67.8	77.7	—
文献[15]	75.7	92.8	82.8	71.9	81.4	79.7
文献[18]	77.3	92.1	83.8	73.6	83.3	—
文献[19]	73.9	92.4	82.3	70.7	79.7	79.3
文献[25]	74.5	90.9	82.1	71.6	80.6	80.3
文献[33]	72.5	89.9	80.3	70.4	80.1	78.6
本文网络	77.6	94.4	84.8	75.0	81.4	80.5

表3　本文网络与其他算法在Halpe-FullBody数据集的对比结果　　单位:%

对比算法	Full-Body						Foot		Face		Hand		Body	
	AP	AP.5	AP.75	AP.L	AP.M	AR	AP	AR	AP	AR	AP	AR	AP	AR
文献[8]	38.7	78.2	34.6	39.3	43.2	52.2	58.1	74.9	42.9	55.8	10.4	20.4	60.5	71.3
文献[13]	35.4	74.2	33.1	36.6	40.2	48.7	55.2	70.9	42.6	53.1	11.4	19.7	57.8	69.4
文献[14]	36.3	76.9	33.6	38.1	42.5	50.2	56.6	72.1	43.7	55.2	10.1	19.6	61.2	70.7
文献[18]	46.1	80.2	46.3	48.2	45.8	53.2	70.3	78.8	51.8	56.2	20.1	29.6	66.2	71.3
文献[19]	45.1	78.3	45.3	46.7	43.4	51.9	70.0	75.4	49.3	54.1	18.8	27.1	65.2	68.8
文献[25]	42.7	80.3	41.2	44.6	43.3	51.3	70.2	77.8	50.5	56.9	13.6	21.0	64.8	69.9
文献[33]	44.1	77.2	44.4	47.0	44.6	53.2	70.6	78.1	49.1	58.0	20.7	29.4	65.0	69.9
本文网络	46.2	80.5	47.7	49.1	46.4	54.8	72.2	79.3	52.6	57.7	22.5	29.8	67.6	71.9

为验证融合了一阶段网络权重后的二阶段网络能够在训练数据不足的前提下实现对手部及脚部关键点的检测,本文在自建体操动作数据集上对二阶段网络与其他对比算法进行定量分析. 对比结果如表4所示,本文二阶段网络融合了一阶段网络权重后,仅使用其他算法1/4的训练数据即可达到与之相当的检测精度. 当使用相等的训练数据时,二阶段网络检测精度较比其他对比算法提升了25%左右. 因此证明结合了权重

融合方法的二阶段网络能够实现在训练数据不足的前提下依然能够达到与其他对比算法相当的精度,在训练数据与其他对比算法相同时,其准确率明显高于其他对比算法. 其原因是二阶段网络浅层权重的初始值与收敛状态的一阶段网络浅层相同,又因为该部分权重在两个子网络浅层间具有高度适配性,因此由式(4)可得$L(\boldsymbol{w}_2)\approx L(\boldsymbol{w}_1)$,即二阶段网络浅层未经训练便接近收敛. 因此该部分网络仅利用少量训练数据即可达到收敛状态,进而减少了所使用训练数据的数量.

表4 二阶段网络与其他算法使用数据量及检测精度对比结果

对比算法	训练数据/张	AP/%	AR/%
文献[8]	≈20 000	51.6	61.4
文献[9]	≈20 000	54.6	66.3
文献[25]	≈20 000	57.9	73.4
文献[33]	≈20 000	55.3	70.6
二阶段网络	5 000	55.8	61.8
	10 000	67.3	69.2
	≈20 000	80.4	76.7

4.4 消融实验分析

在训练数据不足的前提下,网络无法达到良好的关键点检测精度. 为更好地验证本文权重融合的方法能够解决训练数据不足的问题,本文利用自建体操动作数据集对是否使用权重融合方法的二阶段网络进行对比分析. 其结果如表5所示,融合权重的二阶段网络在相同训练数据的前提下比未融合权重的二阶段网络精度提升了23.8%,同时比对比算法精度提升22.5%. 这验证了双阶段网络之间的权重迁移方法是有效的,能够提升关键点检测精度.

表5 二阶段网络是否融合一阶段网络权重检测精度对比结果

对比算法	训练数据/张	AP/%	AR/%
无融合权重的文献[25]	≈20 000	57.9	73.4
无融合权重的二阶段网络	≈20 000	56.6	62.5
融合权重的二阶段网络	≈20 000	80.4	76.7

表5中三种算法对肢体末端的关键点检测效果如图6所示,其中无权重融合的对比算法与二阶段网络产生了漏检以及偏差较大的错检现象. 融合了权重的二阶段网络无漏检现象发生,并且检测结果接近真实位置. 综上所述,本文二阶段网络能够很好地实现对肢体末端关键点的检测. 分析其原因是融合权重的二阶段网络相当于获得了由一阶段网络提供的先验知识,该先验知识由充足训练数据获得,因此具有良好的特征表达能力并且能够高度适配于二阶段网络. 因此在相同训练数据的前提下,其效果要优于无融合权重的网络.

(*a*) 无融合权重的文献[25]

(*b*) 无融合权重的二阶段网络

(*c*) 融合权重的二阶段网络

图6 肢体末端关键点检测效果图

为了验证本文设计的多分辨率输入方法的有效性,对其进行消融实验对比分析. 本次实验将多分辨率网络的双输入分别改为高分辨率或低分辨率的单输入,当输入为高分辨率时,低分辨率输入由高分辨率输入下采样生成的特征图代替. 同理输入为低分辨率时,高分辨率输入为低分辨率的临近上采样特征图. 在COCO2017关键点检测数据集中的实验结果如表6所示,可知多分辨率输入较比其他两种单输入效果更好. 其原因是当输入仅为高分辨率图片时,网络浅层无法获得足够大的感受野,因此不能很好地提取关键点之间的关联性及全局特征信息. 反之,当输入仅为低分辨率输入时,其分辨率较低,大量细节信息被模糊化进而

无法被网络提取,因此其精度较差.

表6　多分辨率改进实验对比结果　　单位:%

对比方法	AP	AR
高分辨率输入	75.1	76.7
低分辨率输入	70.2	67.6
多分辨率输入	77.6	80.5

算法复杂度分析如表7所示,使用GFLOPs(Giga Floating Point Operations Per second)作为神经网络模型计算复杂度指标[32,33]. 由于多分辨率网络中设计了高分辨率通道来生成高分辨率特征图,因此于该特征图上进行的卷积运算生成的参数量高于其余分辨率特征图. 并且本文网络的多尺度融合模块因高分辨率通道的参与计算,其参数量也随之增加. 因此本文多分辨率网络复杂度略高于部分多分辨率网络的对比算法,但精度均优于其他算法.

4.5　可视化展示

本文双阶段网络关键点检测效果图如图7所示,其

表7　算法复杂度

对比算法	GFLOPs
文献[8]	16.04
文献[9]	451.09
文献[25]	28.49
文献[33]	49.81
多分辨率网络	53.61
双阶段网络	103.61

中被红色圆圈标记的点为二阶段网络对手部及脚部关键点的检测结果. 为了证明在体操动作中,增加对这些部位关键点的检测能够关注到更加细节的动作,本文将二阶段网络姿态估计效果与HRNet及Openpose的姿态估计效果进行对比,效果图如图8所示. 本文双阶段网络相比其他两种算法的姿态估计效果,实现了躯干、四肢、手部和脚部姿态的联合估计,使体操动作姿态更加清晰. 因此在体操姿态估计任务中有更好的应用价值.

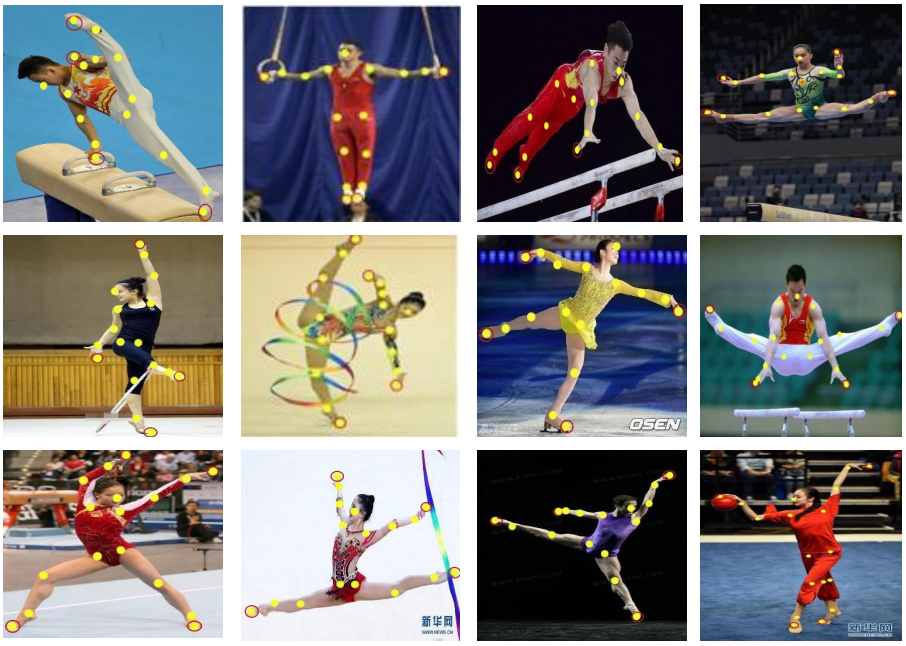


图7　双阶段网络体操动作关键点检测效果图

5　结论

为提升关键点检测精度,本文设计了一种多分辨率网络. 该网络在两种不同分辨率输入图片中提取不同尺度信息,使网络在浅层即具有较大感受野,提升网络检测精度. 同时高分辨率通道可提高网络提取关键点细节特征的能力. 本文在COCO2017关键点检测数据集达到了94.4%的精度,在Halpe-FullBody数据集上的实验结果均高于所有对比算法,进而验证了该网络性能的优越性. 为了实现对手部及脚部关键点的补充检测,本文设计了一种基于增量学习的双阶段网络. 同时为解决训练数据不足的问题提出了一种权重融合的方法,将一阶段网络浅层权重融入二阶段网络浅层,使其无需训练即具备关键点宽泛特征的提取能力,大大减少了训练数据的使用量. 本文在自建体操动作数据集中证明了权重融合方法的有效性. 在消融实验中进一步验证了多分辨率网络以及权重融合方法的有效性. 并从算法复杂度角度对算法进行分析,虽然算法复

(*a*) Openpose体操姿态估计效果图

(*b*) HRNet体操姿态估计效果图

(*c*) 本文双阶段网络体操姿态估计效果图

图8 体操姿态估计效果图

杂度较比其他算法略微提高,但本文网络精度均高于其余算法. 最后在效果图可视化展示中,证明了本文双阶段网络通过对手部及脚部关键点的补充检测,在体操动作中实现更细节的姿态估计. 同时可根据实际任务的需要更改二阶段网络所检测关键点,因此本文算法具有广泛的适用范围,可以更好地应用于各种实际体操姿态估计场景中.

参考文献

[1] ZHANG S Q, WANG C F, DONG W L, et al. A survey on depth ambiguity of 3D human pose estimation[J]. Applied Sciences, 2022, 12(20): 10591.

[2] 罗会兰, 童康, 孔繁胜. 基于深度学习的视频中人体动作识别进展综述[J]. 电子学报, 2019, 47(5): 1162-1173.
LUO H L, TONG K, KONG F S. The progress of human action recognition in videos based on deep learning: A review[J]. Acta Electronica Sinica, 2019, 47(5): 1162-1173. (in Chinese)

[3] 刘世林. 基于深度学习的舞蹈动作识别研究[D]. 成都: 电子科技大学, 2022.
LIU S L. Research on Dance Action Recognition Based on Deep Learning[D]. Chengdu: University of Electronic Science and Technology of China, 2022. (in Chinese)

[4] 任笑圆, 蒋李兵, 钟卫军, 等. 基于视觉的非合作空间目标三维姿态估计方法[J]. 电子与信息学报, 2021, 43(12): 3476-3485.
REN X Y, JIANG L B, ZHONG W J, et al. A vision-based method for 3D pose estimation of non-cooperative space target[J]. Journal of Electronics & Information Technology, 2021, 43(12): 3476-3485. (in Chinese)

[5] CHEN Y L, WANG Z C, PENG Y X, et al. Cascaded pyramid network for multi-person pose estimation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7103-7112.

[6] FANG H S, XIE S Q, TAI Y W, et al. RMPE: Regional multi-person pose estimation[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 2353-2362.

[7] NEWELL A, YANG K Y, DENG J. Stacked hourglass networks for human pose estimation[C]//European Conference on Computer Vision. Cham: Springer, 2016: 483-499.

[8] 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). Piscataway: IEEE, 2019: 5686-5696.

[9] CAO 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.

[10] INSAFUTDINOV E, PISHCHULIN L, ANDRES B, et al. DeeperCut: A deeper, stronger, and faster multi-person pose estimation model[C]//European Conference on Com-

puter Vision. Cham: Springer, 2016: 34-50.

[11] CHENG B W, XIAO B, WANG J D, et al. Bottom-up higher-resolution networks for multi-person pose estimation[C]//2020 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 1-10.

[12] KREISS S, BERTONI L, ALAHI A. PifPaf: Composite fields for human pose estimation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 11969-11978.

[13] WANG D, ZHANG S, HUANG W W, et al. Contextual instance decoupling for robust multiperson pose estimation[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 11050-11058.

[14] LUO Z X, WANG Z C, HUANG Y, et al. Rethinking the heatmap regression for bottom-up human pose estimation [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 13259-13268.

[15] 杨红红, 王刘丽, 张玉梅, 等. 基于序列多尺度特征融合表示的层级舞蹈动作姿态估计方法[J]. 电子学报, 2021, 49(12): 2428-2436.
YANG H H, WANG L L, ZHANG Y M, et al. Hierarchical dance pose estimation algorithm based on sequential multi-scale feature fusion[J]. Acta Electronica Sinica, 2021, 49(12): 2428-2436. (in Chinese)

[16] 沈栎, 陈莹. 带特征监控的高维信息编解码端到端无标记人体姿态估计网络[J]. 电子学报, 2020, 48(8): 1528-1537.
SHEN L, CHEN Y. Feature monitored high-dimension endecoder net for end to end markless human pose estimation[J]. Acta Electronica Sinica, 2020, 48(8): 1528-1537. (in Chinese)

[17] WANG Y J, LUO Y M, BAI G H, et al. UformPose: A U-shaped hierarchical multi-scale keypoint-aware framework for human pose estimation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(4): 1697-1709.

[18] KE L P, CHANG M C, QI H G, et al. DetPoseNet: Improving multi-person pose estimation via coarse-pose filtering[J]. IEEE Transactions on Image Processing, 2022, 31: 2782-2795.

[19] ZHAO L, XU J, GONG C, et al. Learning to acquire the quality of human pose estimation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(4): 1555-1568.

[20] LIN J L, ZHENG Z D, ZHONG Z, et al. Joint representation learning and keypoint detection for cross-view geo-localization[J]. IEEE Transactions on Image Processing, 2022, 31: 3780-3792.

[21] 储珺, 束雯, 周子博, 等. 结合语义和多层特征融合的行人检测[J]. 自动化学报, 2022, 48(1): 282-291.
CHU J, SHU W, ZHOU Z B, et al. Combining semantics with multi-level feature fusion for pedestrian detection[J]. Acta Automatica Sinica, 2022, 48(1): 282-291. (in Chinese)

[22] SUN H M, YANG F S, MA J W. Seismic random noise attenuation via self-supervised transfer learning[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 3146173.

[23] 许啸. 基于深度学习的舞蹈动作分析与生成[D]. 北京: 北京工业大学, 2021.
XU X. Analysis and Generation of Dance Movements Based on Deep Learning[D]. Beijing: Beijing University of Technology, 2021. (in Chinese)

[24] 赵海燕, 马权益, 曹健, 等. 面向任务扩展的增量学习动态神经网络: 研究进展与展望[J]. 电子学报, 2023, 51(6): 1710-1724.
ZHAO H Y, MA Q Y, CAO J, et al. Dynamic neural network for incremental learning with task extended: Research progress and prospect[J]. Acta Electronica Sinica, 2023, 51(6): 1710-1724. (in Chinese)

[25] ZHAO X, WANG Z D, GAO L, et al. Incremental face clustering with optimal summary learning via graph convolutional network[J]. Tsinghua Science and Technology, 2021, 26(4): 536-547.

[26] ZHANG T, LIAN J X, WEN J T, et al. Multi-person pose estimation in the wild: Using adversarial method to train a top-down pose estimation network[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(7): 3919-3929.

[27] WEN B, ZHU Q Y. Class-incremental learning based on big dataset pre-trained models[J]. IEEE Access, 2023, 11: 62028-62038.

[28] FANG H S, LU G S, FANG X L, et al. Weakly and semi supervised human body part parsing via pose-guided knowledge transfer[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 70-78.

[29] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context[C]//European Confer-

ence on Computer Vision. Cham: Springer, 2014: 740-755.

[30] 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, 2023, 45(6): 7157-7173.

[31] 李超. 基于卷积神经网络的人体行为分析与步态识别研究[D]. 杭州: 浙江大学, 2019.
LI C. Human Motion Analysis and Gait Recognition Based on Deep Convolutional Neural Network[D]. Hangzhou: Zhejiang University, 2019. (in Chinese)

[32] PANDUREVIC D, DRAGA P, SUTOR A, et al. Analysis of competition and training videos of speed climbing athletes using feature and human body keypoint detection algorithms[J]. Sensors, 2022, 22(6): 2251.

[33] LIU K, CHEN L L, XIE L, et al. Auto calibration of multi-camera system for human pose estimation[J]. IET Computer Vision, 2022, 16(7): 607-618.

[34] XU L M, JIN S, LIU W T, et al. ZoomNAS: Searching for whole-body human pose estimation in the wild[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(4): 5296-5313.

[35] 王敬宇, 黄伟亭, 刘聪, 等. 基于局部深度一致性的自监督手部姿态估计[J]. 电子学报, 2023, 51(6): 1644-1653.
WANG J Y, HUANG W T, LIU C, et al. Self-supervised hand pose estimation with regional depth correspondence [J]. Acta Electronica Sinica, 2023, 51(6): 1644-1653. (in Chinese)

[36] YAN Q, XU Y, YANG X K. A robust homography estimation method based on keypoint consensus and appearance similarity[C]//2012 IEEE International Conference on Multimedia and Expo. Piscataway: IEEE, 2012: 586-591.

[37] ZHANG T L, JIA S C, CHENG X, et al. Tuning convolutional spiking neural network with biologically plausible reward propagation[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(12): 7621-7631.

[38] LANG Y Z, QIAN Y S, KONG X Y, et al. Effective enhancement method of low-light-level images based on the guided filter and multi-scale fusion[J]. Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 2023, 40(1): 1-9.

[39] CHEN M Z, WANG X, WANG M Z, et al. Estimating rainfall from surveillance audio based on parallel network with multi-scale fusion and attention mechanism[J]. Remote Sensing, 2022, 14(22): 5750.

[40] 李超, 黄新宇, 王凯. 基于特征融合和自学习锚框的高分辨率图像小目标检测算法[J]. 电子学报, 2022, 50(7): 1684-1695.
LI C, HUANG X Y, WANG K. Small object detection of high-resolution images based on feature fusion and learnable anchor[J]. Acta Electronica Sinica, 2022, 50(7): 1684-1695. (in Chinese)

作者简介

江佳鸿 男,1999年生,辽宁辽阳人. 现为大连工业大学研究生. 主要研究方向为图像处理,人体姿态估计.
E-mail: jjh19990901@163.com

夏 楠 男,1983年生,辽宁大连人. 博士. 大连工业大学信息科学与工程学院副教授. 主要研究方向为图像处理、人体姿态估计等.
E-mail: xia_nan0520@aliyun.com

李长吾 男,1966年生,辽宁大连人. 博士. 大连工业大学校长,教授. 主要研究方向为测试计量技术及仪器、数字信号处理,图像处理.
E-mail: lichangwu123456@163.com

周思瑶 女,2000年生,辽宁锦州人. 大连工业大学研究生. 主要研究方向为计算机视觉.
E-mail: 12916004471@qq.com

于鑫淼 男,2000年生,辽宁辽阳人. 大连工业大学研究生. 主要研究方向为目标检测.
E-mail: 1179308761@qq.com