

一种结合轻量级注意力机制的人体姿态估计算法

李文星¹ 喻明毫² 王子牛^{1*} 高建瓴² 林宁³

¹(贵州大学计算机科学与技术学院 贵州 贵阳 550025)

²(贵州大学大数据与信息工程学院 贵州 贵阳 550025)

³(中国科学院计算技术研究所计算机体系结构国家重点实验室 北京 100190)

摘要 针对现有的人体姿态估计模型存在的模型参数量和计算量大、冗余度高、耗时长等问题,提出一种基于轻量级注意力机制的网络框架。使用轻量级网络 MobilenetV3 替代了原 OpenPose 的主干网络 VGG-19;对 OpenPose 的二分支多阶段的卷积神经网络框架进行压缩;引入空间和通道相结合的注意力机制模块 CBAM 对模型的速度和精度进行权衡。实验结果表明,该方法下的网络模型大小和浮点计算量分别为 10.51 MB 和 22.65 GFlops,相对于原 OpenPose 减少了 79.91% 和 83.35%;在 COCO2017 测试集下,能够在保持较高的检测精度和召回率的基础上显著提升检测速度。

关键词 人体姿态估计 计算机视觉 OpenPose 轻量级网络 注意力机制

中图分类号 TP391.41

文献标志码 A

DOI:10.3969/j.issn.1000-386x.2024.05.021

A HUMAN POSE ESTIMATION ALGORITHM COMBINED WITH LIGHTWEIGHT ATTENTION MECHANISM

Li Wenxing¹ Yu Minghao² Wang Ziniu^{1*} Gao Jianling² Lin Ning³

¹(School of Computer Science and Technology, Guizhou University, Guiyang 550025, Guizhou, China)

²(School of Big Data and Information Engineering, Guizhou University, Guiyang 550025, Guizhou, China)

³(State Key Laboratory of Computer Architecture, Institute of Computing Technology Chinese Academy of Sciences, Beijing 100190, China)

Abstract Aimed at the problems of the large amount of model parameters and calculation, high redundancy, and long time-consuming in existing human pose estimation models, a network framework of lightweight attention mechanism is proposed. The lightweight network MobilenetV3 was used to replace the original OpenPose backbone network VGG-1. The two-branch multi-stage convolution neural network framework of OpenPose was compressed. The attention mechanism module CBAM that combined space and channel was introduced, and the speed and accuracy of the model were weighed. Experimental results show that the network model size and floating-point calculation amount under this method are 10.51 MB and 22.65 GFlops, respectively, which are reduced by 79.91% and 83.35% compared with the original OpenPose. Under the COCO2017 test set, this algorithm significantly improves the detection speed on the basis of maintaining high detection accuracy and recall rate.

Keywords Human pose estimation Computer vision OpenPose Lightweight network Attention mechanism

0 引言

随着深度学习和人工智能的迅速发展,计算机视

觉技术也日趋成熟^[1]。人体姿态估计作为计算机视觉领域的重要分支,近年来也扮演着不可或缺的角色,广泛应用于动作识别与捕捉、运动追踪、3D 试衣、监控安防等众多领域^[2-3]。所谓人体姿态估计,指的是给定

一幅图片或一段视频流,对画面中的人物进行身体部位关键点的检测、标注和连接。

人体姿态估计任务一直以来都是 CV 领域里既基础又重要的问题,发展至今,已经有了很多优秀的算法模型。例如,早期由 Toshev 等^[4]提出的 DeepPose,首次使用深度神经网络(DNN)进行姿态估计,它将 2D 姿态估计由原先的图像处理 and 模板匹配问题转化为卷积神经网络中的图像特征提取和关键点位置坐标回归问题,同时应用了相关的回归准则来估计被遮挡或未出现的人体关键点;2016 年的 Hourglass^[5]继承了 DeepPose 所提出的多分辨率特征思想,采用沙漏网络结构融合了高低层特征且在模型中加入中继监督(Intermediate Supervision)进行关键点预测,这让它在 FLIC 和 MPII 数据集上取得了当时史无前例的成就;同年卡耐基梅隆大学的 Wei 等^[6]提出了 CPM(Convolutional Pose Machine)算法,该算法使用多阶段 CNN 的序列组合来提取关键点的位置信息,并将前一阶段的预测输出作为下一阶段的预测输入,从而随着阶段数目的增加,检测结果也更加准确;COCO 人体关键点检测冠军 OpenPose^[7]就是基于 CPM 来构造的,它采用自下而上的检测方法,可以实现人体动作、面部表情、手指运动等姿态估计,并且适用于单人和多人任务,有着较好的鲁棒性;2017 年,旷视科技 face++ 团队提出了 CPN(Cascaded Pyramid Network)^[8]算法,CPN 级联了全局网络和精炼网络两个部分,其创新性地提出一种新的检测方式,即先使用全局网络预测容易的关键点,对难以检测的关键点精炼网络会采取在线难例挖掘(Online Hard Negative Mining)的策略进行提取;2019 年,Sun 等^[9]将高分辨率与低分辨率网络并行连接取代先前的串行连接,使得模型在运算过程中特征图能够始终保持高分辨率,该网络可以进行重复的多尺度特征融合,再次提高了关键点检测性能。计算机视觉的发展趋势是增加网络的宽度和深度以获得更好的性能表现。然而,更好的性能表现往往需要消耗巨大的 GPU 算力和占用较大的存储空间,这对于计算能力和存储空间有限的端设备来说并非易事。为了让检测模型适应端设备,众多学者开始致力于轻量级主干网络的研究^[10-12]。

轻量级网络架构的出现很好地改善了模型存在的参数量和计算量大、冗余度高、耗时长等不足,它能在保证模型准确率的同时体积更小、速度更快。自 2016 年提出至今,受到了学术界广泛的关注,尤其是在计算机视觉领域,已经成为主流的研究热点。Iandola 等^[13]提出了 SqueezeNet,网络结构采用了模块化卷积—Fire module,对输入的通道数目进行压缩和扩张,减少了整

个模型的参数量;Zhang 等^[14]提出了 ShuffleNet,使用逐点分组卷积(Pointwise Group Convolutions)来减少 11 卷积的计算复杂度,利用通道重排(Channel Shuffle)获取通道间的流动信息;Mehta 等^[15]提出了 ESPNet,核心在于提出的 ESP 模块能将标准卷积分解为逐点卷积(Point-wise Convolution)和空间金字塔膨胀卷积(Spatial Pyramid of Dilated Convolutions),增大了感受野,减少了占用内存和计算量。2017 年 Howard 等^[16]提出了轻量级深度神经网络 MobileNetV1,用深度可分离卷积(Depthwise Separable Convolution)替代标准卷积,目的是将网络小型化,方便运用在移动平台。MobileNet 系列网络以更小的体积、更少的计算量和更高的精度使它在轻量级网络中脱颖而出。轻量级算法也使得某些真实的应用场景(如目标检测)在移动端和其他嵌入式平台上部署成为可能^[17-19]。

对于 Hourglass 之类的多阶段预测算法,虽然通过对关键点的不断优化提高了检测精度,但是重复的编解码过程也会带来巨大的参数量和计算量。并且,目前先进的关键点检测算法也需要较大的 GPU 算力。因此,为了实现准确且快速的人体关键点检测,同时减少模型的计算量和占用空间,我们结合 OpenPose 检测模型和 MobileNetV3^[20]轻量级网络,提出一种基于轻量级注意力机制的人体姿态估计算法。首先,使用轻量级网络 MobilenetV3 作为 OpenPose 的主干网络,替代之前的 VGG-19^[21];其次,对 OpenPose 的二分支多阶段卷积神经网络结构进行修改,合并了除输出阶段以外的其他网络层数;最后,在此基础上引入了空间和通道相结合的注意力机制模块 CBAM。结果表明,轻量级网络和注意力机制的结合,能够在保持较高的检测精度和召回率的基础上有效缩减模型的规模,提升模型检测速度。

1 算法模型设计分析

1.1 轻量级网络

MobilenetV3^[20]是谷歌公司 2019 年提出来的轻量级网络架构,它引入了 MobilenetV1^[16]的深度可分离卷积、MobilenetV2^[22]的线形瓶颈层(Linear Bottlenecks)和倒残差结构(Inverted Residual),并提出一种新的激活函数 h-swish 替代之前的 swish 函数来减少计算量。h-swish $[x]$ 和 swish 的表达式如式(1)和式(2)所示。

$$h-swish[x] = x \frac{\text{ReLU6}(x+3)}{6} \quad (1)$$

$$swish[x] = x \cdot \text{sigmoid}(\beta x) \quad (2)$$

式中:ReLU6 为 ReLU 的一种激活函数,使用 ReLU6 可以让模型更早地学习到稀疏特征以及防止训练过程中的数值爆炸; β 为常数。

当我们输入一幅大小为 $D_F \times D_F \times M$ 的特征图,经过 $D_K \times D_K \times M \times N$ 卷积核的作用得到它的输出特征图尺寸为 $D_F \times D_F \times N$ 。其中: D_F 表示输入的特征图尺寸; D_K 为卷积核的尺寸; M 和 N 分别表示输入输出的通道数。此时,模型卷积核的计算量可表示为:

$$D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F \quad (3)$$

如果我们使用深度可分离卷积替代传统的标准卷积,输入的特征图尺寸会被分解为深度卷积 $D_K \times D_K \times 1 \times M$ 和逐点卷积 $1 \times 1 \times M \times N$,对应输出特征图的尺寸为 $D_K \times D_K \times N$ 。模型卷积核的计算量为:

$$D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F \quad (4)$$

相比较传统的标准卷积,深度可分离卷积能将卷积核的计算量压缩为:

$$\frac{D_K \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_K \cdot D_K \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_K^2} \quad (5)$$

1.2 注意力机制

注意力机制(Attention Mechanism)是深度学习任务中的关键技术,它让模型能够更加关注某些局部的有用信息。从自然语言处理中的机器翻译、情感分类与自动问答到计算机视觉中的图像识别、图像分类与语义分割中均有涉及^[23]。注意力机制模型架构从原理上可以分为空间注意力机制、通道注意力机制、空间和通道混合注意力机制。

MobilenetV3 在其核心架构中加入通过压缩操作(Squeeze)和激励操作(Excitation)的通道注意力机制

SE-Block,用来加强网络的学习能力。如图 1 所示,对于一幅输入大小为 $H \times W \times C$ 的特征图,首先进入压缩操作,它是一个全局平均池化(Global Average Pooling, GAP)的过程,将特征图压缩为 $1 \times 1 \times C$ 的向量;激励操作的目的是为了获取前一步压缩操作中聚合的信息,学习各个通道之间的关系;最后进行 Scale 操作,二维矩阵相乘,得到最后输出的特征图。

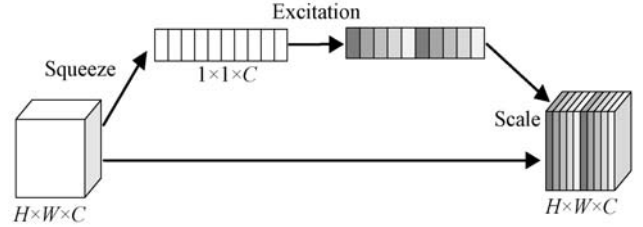


图 1 SE 模块结构示意图

1.3 OpenPose 检测算法

人体姿态识别项目 OpenPose 是由卡耐基梅隆大学(CMU)的 Cao 等^[7]于 2016 年提出,它是以 Caffe 为深度学习开发框架、基于卷积神经网络 VGG-19 和监督学习的人体姿态估计模型。OpenPose 的算法流程如图 2 所示,首先主干网络 VGG-19 会对图像进行特征提取,然后对关键点的热力图(Heatmaps)和表示配对关系的部分亲和域(Part Affinity Fields, PAFs)作初始估计,再进入 5 个增强阶段来提升效果。在 COCO 数据集下,它一共可以找到身体部位 18 个关键点(左右耳、左右眼、鼻子、脖子、左右肩、左右肘、左右腕、左右髌、左右膝和左右踝),然后从预定义的配对列表中搜索最佳配对关系进行连接。最后调整全局特征图的大小,对所有关键点进行提取和分组拼接后输出检测结果。

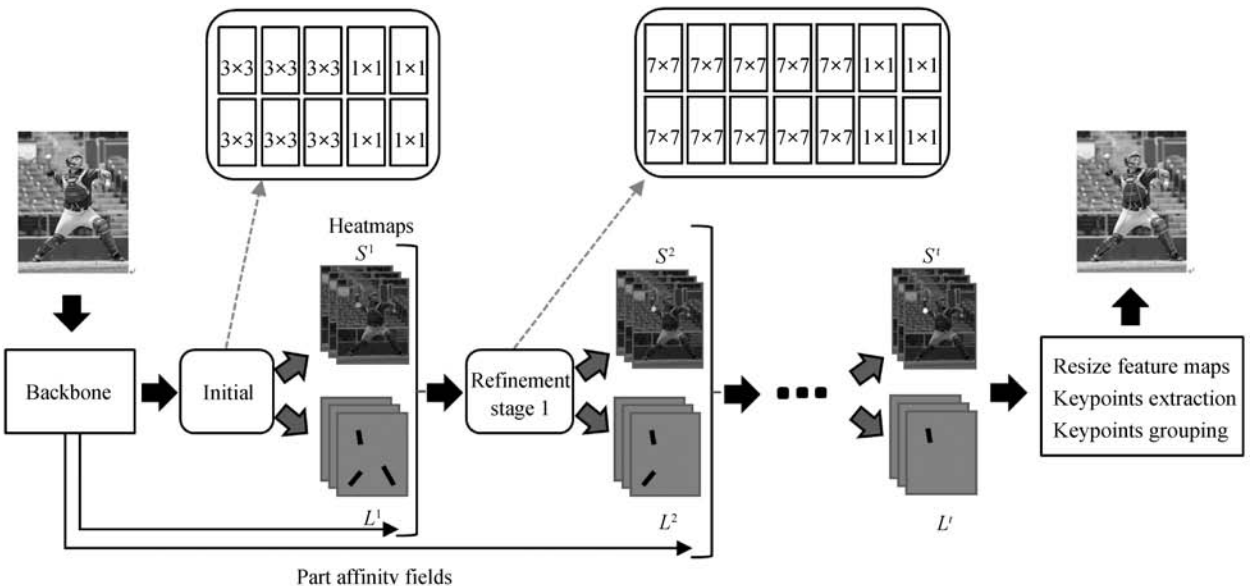


图 2 OpenPose 算法流程

2 改进的轻量级注意力机制模型

2.1 改进的网络结构

作为 ImageNet^[20] 大规模视觉挑战赛 (ILSVRC) 中定位任务冠军和分类任务亚军的 VGGNet^[21], 自提出以来广泛应用于图像的特征提取。然而, 虽然 VGGNet 在所有层中使用 3×3 的卷积核替代传统的标准卷积核 5×5 和 7×7 , 但是相比于主流的轻量级网络它仍然存在网络规模大的不足。

本文将原 OpenPose 的主干网络 VGG-19 替换成轻量级网络 MobileNetV3, 用深度可分离卷积替代标准卷积可以大幅度地减小模型的参数量和计算量。每个阶段都会产生新的关键点热力图和部分亲和域的估计值, 并结合经主干网络提取的特征值一起进入到下一个阶段。基于这一事实, 我们决定在网络中共享大部分计算, 除了输出结果的最后两层。初始阶段网络结构改进如图 3 所示, 增强阶段改进方式与初始阶段相同。同时将网络中的 7×7 卷积核用三个 3×3 卷积核替代, 在感受野不变的前提下不仅减少了参数, 而且使得网络更深, 非线性表达能力也更好。

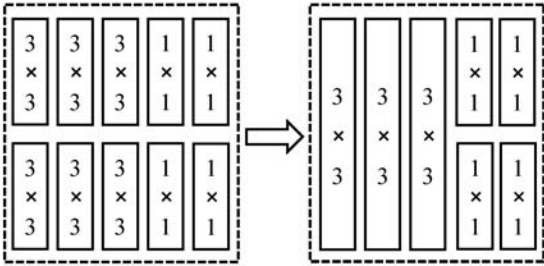


图3 初始阶段网络结构改进

2.2 注意力机制模块 CBAM

Convolutional Block Attention Module (CBAM)^[24] 是一种空间 (Spatial) 和通道 (Channel) 相结合的注意力机制模块。它相对于 SE-Block 多了一个空间注意力机制, 性能上能够得到一定的提升。通过在 ImageNet、COCO 和 VOC 数据集上的大量实验验证表明, 引入 CBAM 的模型在分类和检测结果上均有所改进, 充分说明了该模块的有效性^[21]。

本文中, 我们使用 CBAM 模块取代 MobileNetV3 自身的 SE 模块来实现 attention 机制。图 4 为 CBAM 的网络结构, Channel Attention Model 用来提取重要的数据, Spatial Attention Module 会对其做一个关键位置信息的补充, 两模块的表达式分别如式 (6) 和式 (7) 所示。因此, CBAM 可以同时捕获空间和通道两个维度

的特征图权重信息, 从而增强卷积神经网络的表达能力。

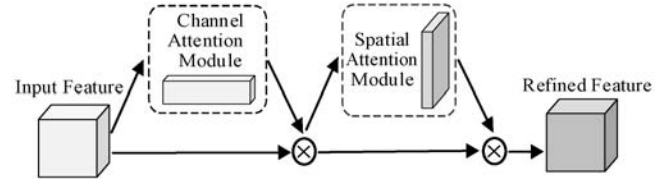


图4 CBAM 模块结构示意图

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) = \sigma(W_1(W_0(F_{\text{avg}}^c)) + W_1(W_0(F_{\text{max}}^c)))W_0 \quad (6)$$

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) = \sigma(f^{7 \times 7}(F_{\text{avg}}^s; F_{\text{max}}^s)) \quad (7)$$

式中: σ 是 sigmoid 函数; W_0 和 W_1 是多层感知机的权重, $W_0 \in R^{c/r \times c}$, $W_1 \in R^{c/r \times c}$, R 和 r 代表降维比例和升维比例。 7×7 表示卷积核的大小, 在这里使用 7×7 卷积核是因为它会比 3×3 卷积核获得更好的效果。

2.3 损失函数

本文的损失函数主要是参考 OpenPose 的源码来确定的, 在 OpenPose 中, 损失函数的计算方式定义为生成关键点 Heatmaps 和 PAFs (图 2) 两阶段的损失之和。

$$f_s = \sum_{j=1}^J \sum_p W(p) \cdot \|S_j^s(p) - S_j^*(p)\|_2^2 \quad (8)$$

$$f_L = \sum_{c=1}^C \sum_p W(p) \cdot \|L_c^L(p) - L_c^*(p)\|_2^2 \quad (9)$$

$$f = \sum_{i=1}^T (f_s + f_L) \quad (10)$$

式中: f_s 是生成 Heatmaps 阶段的损失; f_L 是生成 PAFs 阶段的损失; W 是掩膜 (mask) 操作, 它是由 0 和 1 组成的二进制图像且当无法检测到关键点或者标签丢失时 $W(p) = 0$, 目的是为了防止在训练中影响正确的预测; S_j^s 和 L_c^L 表示关键点热力图和部分亲和域的真实值 (实际位置); S_j^* 和 L_c^* 表示对应的预测值 (预测位置); 模型的每个阶段都设置了中继监督 (Intermediate Supervision) 优化, 它可以通过周期性地补充梯度来解决梯度消失问题; f 代表最终的损失。

本文的改进方案里共享了初始阶段和增强阶段的部分网络层数, 但是不影响损失函数的计算方式, 这是因为损失是在生成 Heatmaps 和 PAFs 的两个特征图部分进行统计的, 即预测 loss 时只使用网络最后一层输出值。初始阶段 loss 计算位置如图 5 所示, F 是图像经主干网络提取到的特征图, 增强阶段损失计算位置与初始阶段相同。

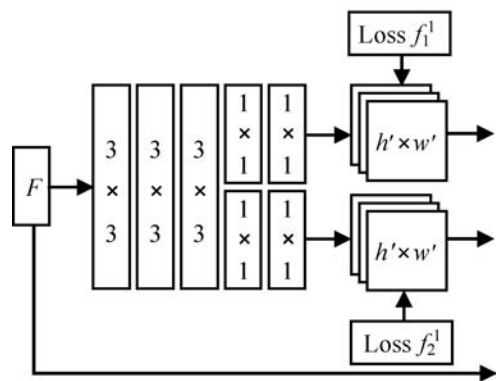


图 5 初始阶段损失计算位置

3 实验

为了评估本文改进方案的可行性,本次实验设计了与原 OpenPose 模型以及分别使用 SqueezeNet、ShuffleNet、MobileNet 系列作为主干网络的 OpenPose 进行对比实验。

实验的训练环境是基于 Linux 内核的 Ubuntu 操作系统,CPU 为 Intel(R) Xeon(R) Silver 4210, GPU 为 GeForce RTX 2080Ti,深度学习框架为 PyTorch。为了体现轻量级网络的优势,测试环境选用的是个人 PC 设备,CPU 为 Intel(R) Core(TM) i7-4770HQ, GPU 为集显 Intel Iris Pro 1536。

3.1 数据集

实验的数据集采用 MS COCO 2017 数据集。COCO (Common Objects in Context) 数据集是由微软团队创作,可以进行目标检测、人体关键点检测、语义分割、字幕生成等任务。MS COCO 2017 数据集中的图片分为训练集、验证集和测试集三类,包含 80 种目标类别,其中:训练集 118 287 幅,验证集 5 000 幅,测试集 40 670 幅。实验中我们对数据集进行了相关的预处理,保证了数据的有效可靠性。

3.2 验证标准及评价指标

我们将网络的输入分辨率设置为 368×368 ,进行单尺度 (Single-Scale) 测试。实验的验证集为 COCO 2017 VAL,验证方法是 MS COCO 官方给定的评估标准 OKS,即目标关键点相似性 (Object Keypoint Similarity)。OKS 的公式定义为:

$$O_{ks} = \frac{\sum_i \exp(-d_i^2/2s^2k_i^2)\delta(v_i > 0)}{\sum_i \delta(v_i > 0)} \quad (11)$$

式中: d_i 表示每个关键点实际位置和预测位置间的欧

氏距离, v_i 为实际真值的可见性标记($v_i > 0$ 表示能够观察到关键点, $v_i < 0$ 表示不能观察到关键点);目标尺寸记为 s ;每个关键点的控制衰减系数记为 k_i 。

选择合适的指标去衡量实验结果至关重要,本文评估方法综合了模型的参数量 (Params)、浮点计算量 (FLOPs)、乘法操作总量 (MADD)、读写内存 (MemR + W) 大小。采用 AP 和 AR 表示模型的检测结果,这里 AP 和 AR 指 $O_{ks} = 0.50, 0.55, 0.60, \dots, 0.95$ 时的平均准确度和平均召回率。同时使用视频流的每秒传输帧数 (FPS) 和图片的平均检测时间 (Avg_time) 来测试模型的检测速度。

3.3 实验结果分析

3.3.1 训练结果

图 6 是本文算法下 Heatmaps 和 PAFs 阶段的平均损失曲线。实验采用 Adam 优化算法,对模型进行三个阶段训练。第一阶段 Epoch 设置为 0 ~ 100,损失在这一阶段迅速下降;第二阶段 Epoch 设置为 101 ~ 200,这一阶段损失缓慢下降并开始趋于平稳;第三阶段从 201 ~ 280,随着训练过程中 Epoch 数量的增加,模型的损失已基本保持稳定。

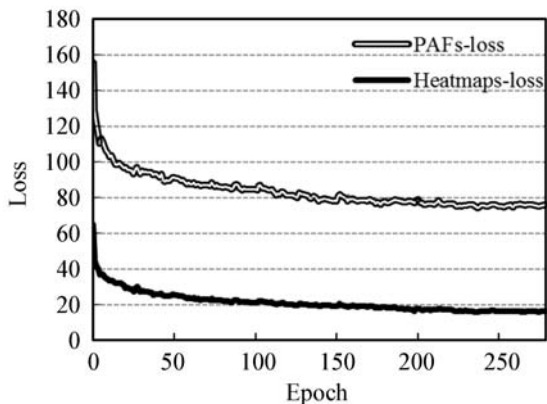


图 6 二分支阶段平均损失曲线

3.3.2 模型复杂度分析

实验中我们参考了 OpenPose 在不同主干网络和本文改进方案下的各指标表现情况,结果如表 1 所示。除了原 OpenPose 的主干网络使用的是 VGG-19 前十层,轻量级网络 SqueezeNet、ShuffleNet、MobileNet 系列的网络层数是通过多次实验反复对比后确定的。其中 SqueezeNet 系列、ShuffleNet 系列、MobileNet 系列中的 MobileNetV1 和 MobileNetV2 结构均没有改动,对于 MobileNetV3,我们在权衡模型参数和性能的表现情况后,削减掉了网络后两层。因为使用更深的网络结构会提高模型的精度,但同时也会增加模型的参数、计算量。

表 1 不同主干网络下模型复杂度对比

主干网络	Params /MB	MADD /GMAdd	FLOPs /GFlops	MemR + W /GB
VGG-19	52.31	271.94	136.07	1.10
SqueezeNet	18.43	88.99	44.67	2.23
ShuffleNetV1	21.15	119.40	59.93	5.30
ShuffleNetV2	19.99	99.64	49.99	3.36
MobileNetV1	20.63	87.82	44.02	1.74
MobileNetV2	19.76	85.35	42.82	2.55
MobileNetV3	18.84	80.20	40.20	1.64
MobileNetV3-CBAM	10.51	45.17	22.65	1.20

通过表 1 中数据可以发现,使用轻量级网络作为 Backbone 的 OpenPose 模型各参数指标均有所下降,并且在本文提出的 MV3-CBAM-OpenPose 算法下,模型大小仅为 10.51 MB,乘加操作总量仅为 45.17 GMAdd,浮点计算量仅为 22.65 GFlops,分别较原 OpenPose 降低了 79.91%、83.39%、83.35%。读写内存大小为 1.20 GB,稍有所增加。

3.3.3 模型性能分析

表 2 是不同算法模型的检测性能对比。MV3-CBAM-OpenPose 算法能够在现有的 OpenPose 基础上,将 FPS 从 1.08 提升至 22.76,完成每幅图片的平均检测时间只需 2.53 s,并且精度和召回率相差无几(AP 和 AR 相差 1% 以内)。根据表 2 可知,将 OpenPose 的主干网络替换为轻量级网络会明显提升模型的检测速度 FPS,并且使用 MobileNet 系列作为 Backbone 的 OpenPose 性能更佳。同时,相比较只更改 Backbone 为 MobileNetV3 的 OpenPose,本文算法各项性能指标均更优。

表 2 不同算法模型性能对比

算法	AP /%	AR /%	FPS/(帧· s ⁻¹)	Avg_ time/s
OpenPose	48.52	52.77	1.08	7.29
SqueezeNet-OpenPose	43.81	47.93	13.81	6.84
ShuffleNetV1-OpenPose	45.06	48.85	16.23	6.01
ShuffleNetV2-OpenPose	45.43	49.13	16.50	5.77
MobileNetV1-OpenPose	46.33	50.02	20.93	7.32
MobileNetV2-OpenPose	45.73	49.30	22.51	6.58
MobileNetV3-OpenPose	47.61	51.58	21.70	3.47
MV3-CBAM-OpenPose	48.26	52.15	22.76	2.53

3.3.4 检测结果分析

使用本文算法的模型检测效果如图 7 所示,模型能对图片中的人物进行有效的姿态估计,分别识别出身体部位对应的关键点并进行正确的连接,对于不同的光线和拍摄角度,也都能得到较好的预测结果。甚至对于图 7(b)和图 7(c)的图片,人物身体部位被部分遮挡,但是模型却对其作出了正确的预测输出。

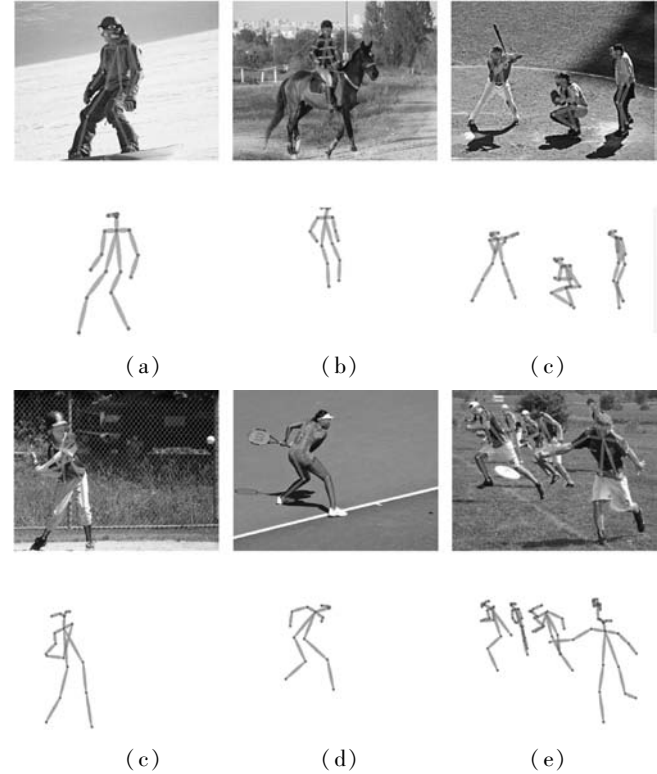


图 7 人体姿态估计效果

模型也有一定的不足,存在严重遮挡的图片可能会产生不准确的预测,比如图 7(d)的人物左右臂出现连接错误,图 7(e)的人物被遮挡的右臂连接缺失,图 7(f)的相邻人物预测结果出现部分混乱。

4 结 语

本文提出一种结合轻量级注意力机制的人体姿态估计算法: MV3-CBAM-OpenPose。通过使用轻量级网络 MobilenetV3 替代原 OpenPose 的主干网络 VGG-19,并在网络中共享了大部分计算,降低了模型的复杂度;使用 3×3 卷积核替代网络中的 7×7 卷积核,增强了非线性表达能力;最后引入空间和通道相结合的注意力机制模块 CBAM,提升模型的检测性能。实验结果表明,本文算法能够显著提升人体姿态估计的速度,并且能够保持较高的精度和召回率。该模型对现实生活中的人体动作分析、人体行为检测等研究具有积极的作用,针对算法存在的检测问题,下一步的工作我们将

进一步改进或使用其他算法来提高模型的检测能力。

参 考 文 献

- [1] 张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用[J]. 计算机学报, 2019, 42(3): 453-482.
- [2] 冯晓月, 宋杰. 二维人体姿态估计研究进展[J]. 计算机科学, 2020, 47(11): 128-136.
- [3] 梁锦明. 人体姿态估计与识别技术的研究[D]. 长春: 长春工业大学, 2020.
- [4] Toshev A, Szegedy C. DeepPose: Human pose estimation via deep neural networks[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1653-1660.
- [5] Newell A, Yang K, Deng J. Stacked hourglass networks for human pose estimation[C]//European Conference on Computer Vision, 2016: 483-499.
- [6] Wei S, Ramakrishna V, Kanade T, et al. Convolutional pose machines[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4724-4732.
- [7] 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, 2018, 43(1): 172-186.
- [8] Chen Y, Wang Z, Peng Y, et al. Cascaded pyramid network for multi-person pose estimation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 7103-7112.
- [9] Sun K, Xiao B, Liu D, et al. Deep high-resolution representation learning for human pose estimation[EB]. arXiv: 1902.09212, 2019.
- [10] Osokin D. Real-time 2D multi-person pose estimation on CPU: Lightweight OpenPose[C]//8th International Conference on Pattern Recognition Applications and Methods, 2018: 744-748.
- [11] Santavas N, Kansizoglou I, Bampis L, et al. Attention! A lightweight 2D hand pose estimation approach[J]. IEEE Sensors Journal, 2020, 21(10): 11488-11496.
- [12] Marvasti-Zadeh S, Ghanei-Yakhdan H, Kasaei S. Efficient scale estimation methods using lightweight deep convolutional neural networks for visual tracking[J]. Neural Computing and Applications, 2021, 33(14): 8319-8334.
- [13] Iandola F, Han S, Moskewicz M, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size[EB]. arXiv:1602.07360, 2016.
- [14] Zhang X, Zhou X, Lin M, et al. ShuffleNet: An Extremely efficient convolutional neural network for mobile devices[EB]. arXiv:1707.01083, 2017.
- [15] Mehta S, Rastegari M, Caspi A, et al. ESPNet: Efficient spatial pyramid of dilated convolutions for semantic segmentation[C]//European Conference on Computer Vision, 2018: 561-580.
- [16] Howard A, Zhu M, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB]. arXiv:1704.04861, 2017.
- [17] 蒋洋涛. 基于轻量级神经网络的交通场景的目标检测方法研究[D]. 杭州: 杭州电子科技大学, 2020.
- [18] 高刘雅, 孙冬, 卢一相. 基于轻量级注意力机制的人脸检测算法[J]. 激光与光电子学进展, 2021, 58(2): 130-138.
- [19] Niu Y, Lyu Z, Liu X, et al. A lightweight target detection network for embedded platforms[C]//2020 39th Chinese Control Conference, 2020: 7617-7621.
- [20] Howard A, Sandler M, Chen B, et al. Searching for mobilenet3[C]//2019 IEEE/CVF International Conference on Computer Vision, 2019: 1314-1324.
- [21] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB]. arXiv:1409.1556, 2014.
- [22] Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [23] 朱张莉, 饶元, 吴渊, 等. 注意力机制在深度学习中的研究进展[J]. 中文信息学报, 2019, 33(6): 1-11.
- [24] Woo S, Park J, Lee J, et al. CBAM: Convolutional block attention module[EB]. arXiv:1807.06521, 2018.

(上接第100页)

- [9] 吴文宣, 唐志军, 晁武杰, 等. 基于智能变电站组合设备仿真测试与集成联调技术方案研究[J]. 电网与清洁能源, 2018, 34(4): 42-48.
- [10] 习伟, 姚浩, 陈波, 等. 基于零延时切换的可测交换网在智能变电站的应用[J]. 电力科学与技术学报, 2019, 34(4): 42-47.
- [11] 李金, 胡荣, 王丽华, 等. 智能变电站 IEC 61850 Ed 2.0 工程配置应用方案[J]. 电力系统自动化, 2018, 42(2): 154-159.
- [12] 王莹, 孙海涛, 薛彦登, 等. 750 kV 变电站噪声测量与特性分析[J]. 西安工程大学学报, 2019, 33(5): 524-530.
- [13] 李勇, 陈雨, 蔡晔, 等. 基于信息物理接口矩阵的 IEC61850 变电站自动化系统可靠性分析[J]. 电力自动化设备, 2019, 39(1): 84-90.
- [14] 高磊, 宋亮亮, 杨毅, 等. 基于多参量模型的智能变电站二次设备状态评估方法及应用[J]. 电力自动化设备, 2018, 38(10): 210-215.