

基于自学习特征金字塔网络的人体关键点检测算法

孔令军^{1,2}, 赵子昂², 刘伟光¹, 周耀威¹

(1. 金陵科技学院网络与通信工程学院, 江苏 南京 211169; 2. 南京邮电大学通信与信息工程学院, 江苏 南京 210094)

摘要:针对现有卷积神经网络结构不能充分挖掘高低级语义特征的缺点,提出一种联合自学习的属性金字塔模块(JSLAPM),其由多维自学习模块(MSLM)和特征金字塔模块(FPM)构成,可应用到任意主干网络结构中。MSLM通过学习空间维度以及通道维度上的特征矩阵重要性来调整特征矩阵,而FPM通过融合不同深度的特征矩阵来增强特征属性的表达能力。此外,结合通道分离提取模块(CSEM),提出了一种特征金字塔注意力网络(FPANet)。实验结果表明,所提出的网络模型可将VGGNet和ResNet的主干网络精度分别提升了近3%。

关键词:卷积神经网络;人体关键点检测;特征金字塔;注意力机制

中图分类号:TP391.41

文献标识码:A

文章编号:1672-755X(2021)03-0013-09

Human Keypoint Detection Algorithm Based on Self-learning Feature Pyramid Network

KONG Ling-jun^{1,2}, ZHAO Zi-ang², LIU Wei-guang¹, ZHOU Yao-wei¹

(1. Jinling Institute of Technology, Nanjing 211169, China;

2. Nanjing University of Posts and Telecommunications, Nanjing 210094, China)

Abstract: In view of the shortcomings of the existing convolutional neural network structure that cannot fully excavate high and low-level semantic features, a new joint self-learning attribute pyramid module(JSLAPM) is proposed in this paper, composed of multidimensional self-learning module(MSLM) and feature pyramid modul(FPM), which can be applied to any backbone network structure. MSLM adjusts the feature matrix by learning the importance of feature matrix in spatial dimension and channel dimension, while FPM enhances the expression ability of feature attributes by fusing feature matrix of different depths. In addition, a self-learning feature pyramid attention network(FPANet) is also proposed based on the channel separation and extraction module(CSEM). Experimental results show that the proposed network model can improve the accuracy of VGGNet and ResNet backbone networks by nearly 3% respectively.

Key words: convolutional neural networks; human keypoint detection; feature pyramid; attention mechanism

随着人工智能产业实践的深入发展^[1-2],人工智能已经成为数字时代的重要标志,以人工智能为代表的

收稿日期:2021-06-13

基金项目:中国博士后科学基金资助项目(2020M671595);江苏省博士后科研资助计划资助项目(2020Z198);金陵科技学院高层次人才科研启动基金(jit-b-202110)

作者简介:孔令军(1982—),男,黑龙江鸡西人,副教授,博士,主要从事人工智能与信息科学交叉研究、通信和存储系统中新型信号处理与纠错码技术研究。

数字经济已经成为经济发展的新引擎。近年来计算机超大规模并行计算能力的提升以及各种传感技术的研发,极大地推动了深度学习技术在图像识别、自然语言处理、语音识别等多个领域中的发展。作为图像识别领域的典型代表,人体关键点检测的重要性日益凸显^[3]。人体关键点检测在人机交互^[4]、病人监护系统^[5]、智能视频监控^[6-7]、虚拟现实、运动员辅助训练、智能家居^[8]、人体动画、智能安防等领域都有着重要的意义和广泛的应用前景,是诸多计算机视觉任务,例如动作分类、异常行为检测以及自动驾驶等的基础^[9]。

人体关键点检测是指基于图像来重建人的关节和肢干,找出人体骨骼的关键点并将其组合^[10-11]。人体关键点(通常是特殊部位或骨骼关节点)的定位至今依旧是计算机视觉(computer vision, CV)领域的热门研究方向。近年来,随着卷积网络的发展^[12-13],在深度学习领域内人体关键点检测算法研究取得了巨大的进步。在这些深度学习算法中,结合高低级语义的特征金字塔模块^[14](feature pyramid module, FPM)却很少提及,但在目标检测、像素切割等任务中,已经验证了 FPM 能够显著地增强网络的提取效果。受到启发,本文研究 FPM 并将其应用到图片特征提取网络中,来提升人体关键点检测的效果。在图片特征提取网络中流通的特征矩阵包含重要或者不重要的信息,在像素分类等类似任务中,已经证明了注意力模块(attention module, AM)能够通过自学习的方式增强特征矩阵中 useful 信息的表达。因此,为了提升人体关键点任务中图片特征提取网络对有用信息的提取能力,本文也将 AM 应用到图片特征提取网络中,并提出了多维自学习模块(multidimensional self-learning module, MSLM)。

基于上述研究,本文提出了一种全新的并能够应用于人体关键点检测的图片特征提取网络,即特征金字塔注意力网络(feature pyramid attention network, FPANet)。FPANet 能够解决现有算法不能充分利用不同分辨率的特征矩阵内信息以及不能够自动增强特征矩阵中 useful 信息表达能力的问题;同时,在预测阶段还将使用通道分离提取模块(channel separation and extraction module, CSEM)降低关键点之间的相互影响,本文将通过实验展现研究成果。

1 算法框架设计

为了让图片特征提取网络提取到的信息更加有用,本文研究 FPM 和 MSLM 并将其应用到 Res-Net^[15]、VGGNet^[16] 网络结构中,尝试使用 CSEM 来降低不同人体关键点之间的相互影响。

本文设计的 FPANet 框架如图 1 所示。

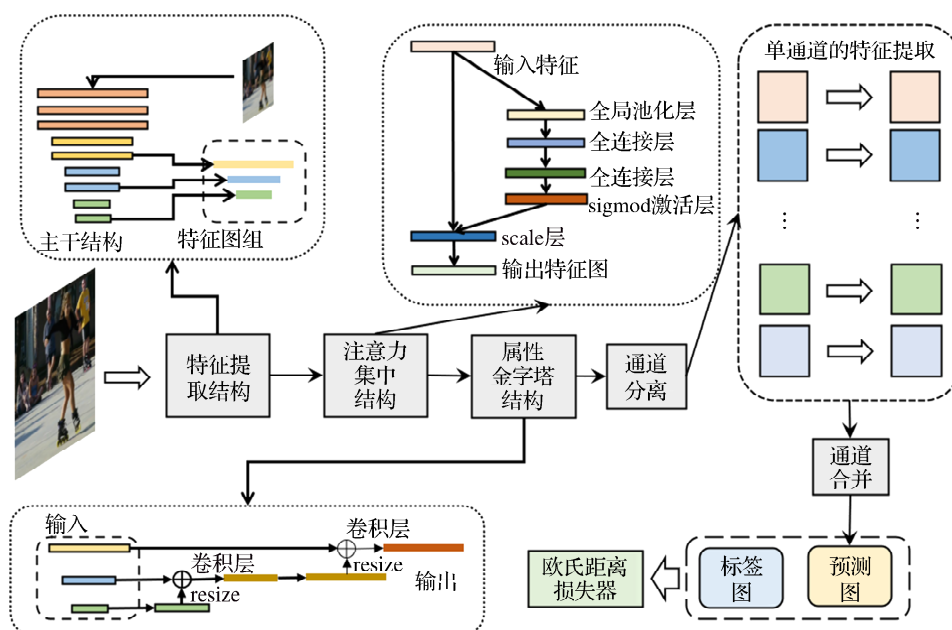


图 1 FPANet 整体架构

FPANet 框架使用了 FPM、MSLM、CSEM 3 个模块,框架设计过程为:1)图片通过主干结构提取到不同分辨率的特征矩阵组后,MSLM 通过组内的特征矩阵来调节特征中有效信息的表达;2)特征矩阵通过 FPM 来结合高低级语义特征得到最终的图片特征;3)使用 CSEM 来预测人体关键点热点图以及人体关键点间关系的热点图,并使用标签对网络进行训练。

1.1 多维自学习模块(MSLM)

本文提出了以空间自学习模块以及通道自学习模块为基础设计的全新 MSLM,解决了模块过小或者过大的问题。模块在经过空间注意力自学习模块调整后,并不直接进行第二次调整,而是先使用 3×3 的卷积进行调整、缓冲后,再使用空间自学习模块进行调整。调整是以浅层特征中更加直观的特征为依据进行的。

1.2 联合自学习的属性金字塔模块(JSLAPM)

受到目标检测、分类等任务使用金字塔结构充分融合高低级特征矩阵的启发,本文提出了如图 2 所示的联合自学习属性金字塔模块(joint self-learning attribute pyramid module, JSLAPM),将金字塔结构应用到了主干网络结构中,增加了对不同分辨率的特征矩阵融合能力,在进行特征融合之前,使用了多维自学习模块(MSLM)进行预处理,然后采用了特征金字塔模块(FPM)将提取到的特征通过属性金字塔结构,按不同层次、分步骤地融合到一个特征中,以增强模型的表达能力。

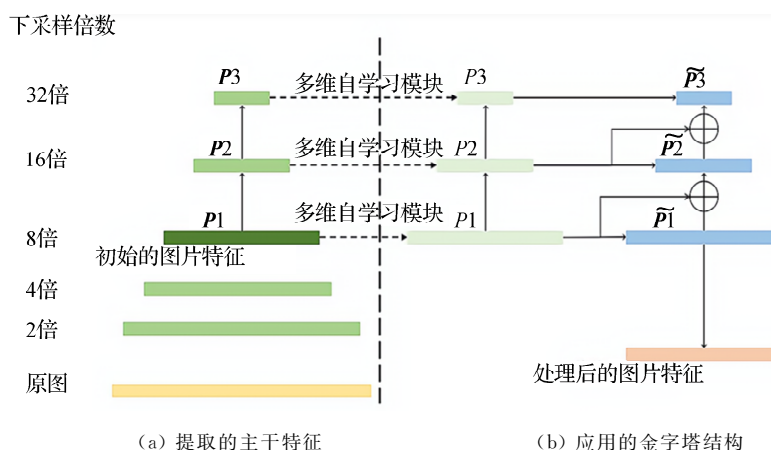


图2 联合自学习的属性金字塔模块(JSLAPM)

图 2(a)为主干网络提取的不同分辨率的特征矩阵,在不使用 JSLAPM 时,提取得到的图片特征矩阵正是图中的 P_1 。图 2(b)展示了使用 JSLAPM 的过程,过程描述如下:

1)通过主干网络提取得到不同分辨率的由特征矩阵组成的特征矩阵组,特征矩阵组中的特征矩阵分辨率分别是原图下采样的 8 倍、16 倍、32 倍,然后使用 MSLM 分别对这些特征矩阵进行自适应学习;

2)通过卷积对 P_3 操作,但并不改变特征图的维度,得到了特征图组 $\tilde{P}_3$,然后对此特征图进行下采样得到特征图,并与 P_2 进行通道合并得到了 $\tilde{P}_2$,利用相同的过程得到最终的特征图 $\tilde{P}_1$,执行了融合不同高低级语义的功能;

3)以步骤(2)提取得到的特征矩阵为基础进行一次普通卷积,得到了处理后的图片特征矩阵,为以后的预测等操作提供原始的输入。

1.3 通道分离提取模块(CSEM)

CSEM 提取了一种支路,是由预测前对特征矩阵分离所得。通过训练不同的权重,不同的支路决定着不同的人体关键点预测,或者不同的关键点之间的关系预测。

通道分离的实现过程如图 3 所示,图中的单通道预测模块由卷积层和池化层组成。

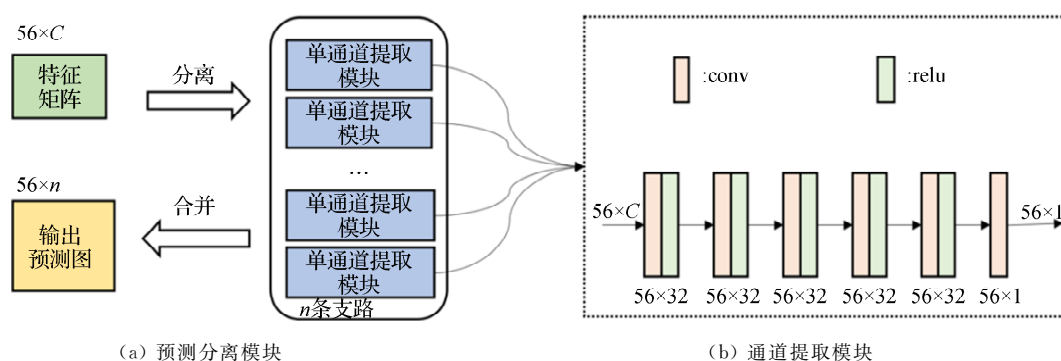


图 3 通道分离提取模块

1)单通道提取模块。其为基本的组成单元,数据通过此模块的具体步骤如下:首先 $56 \times 56 \times C$ 的特征矩阵通过卷积池化层不断地提取特征,在不改变原特征矩阵维度的情况下,连续提取 5 次,最终使用卷积核提取得到一个 56×1 的特征图。

2)应用过程。将提取图片的特征矩阵通过单通道来提取模块,重复上述操作 n 次得到 n 个 56×1 的特征图,其中 1 代表预测中某个关键点或者关键点间关系的预测图。

3)通过通道合并的方式对这 n 个特征图进行通道合并,得到 $56 \times n$ 维的特征图,特征图通过 CSEM 得到了预测图,并对其预测图进行损失训练。

4)将图片的特征与此预测图进行通道合并后,再进行相应的操作得到第二阶段的预测图,预测结构与图片的特征保持相同,并对第二阶段的预测图进行损失训练,不断重复此过程 5 次,起到多损失训练的作用(图 4)。

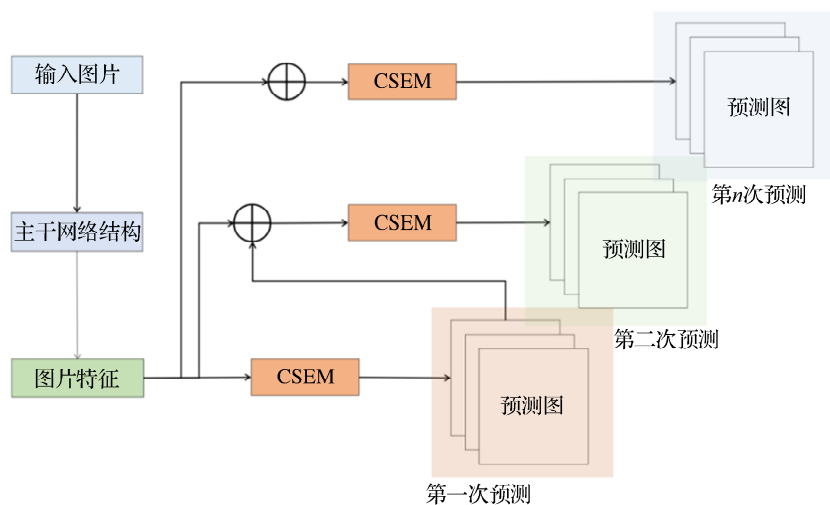


图 4 多层损失预测以及 CSEM

输入图片通过主干网络结构(VGGNet、ResNet 等)进行特征提取,得到了图片的主干特征,再应用 FPM 和 MSLM,进一步提取得到图片的主干特征,最后将 CSEM 应用到网络中。多层损失预测的过程为:首先特征图通过 CSEM 得到了预测图,此预测图预测人体关键点热点图或者人体关键点间的关系热点图,并对其预测图进行损失训练;将图片的特征与此预测图进行通道合并后,再进行相应的操作得到第二阶段的预测图,预测结构保持与图片的特征相同,依旧对此预测图进行损失训练,不断重复此过程 5 次,起到多损失训练的作用(图 4)。

2 网络训练

2.1 标签与损失函数

为了训练出能够检测人体关键点的模型,本文将人体关键点的热点图作为标签,图中的每个像素位置都给定一个概率,表示该点属于对应类别关键点的概率,距离关键点位置越近的像素点的概率越接近1,距离关键点位置越远的像素点的概率越接近0。训练过程中标签转换流程如图5所示。

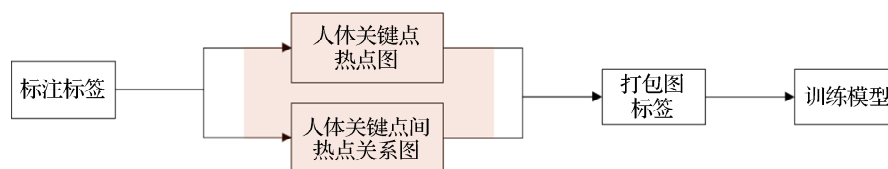


图5 训练过程中标签转换流程

首先通过标注标签创建标签图,标签图包括人体关键点标签图以及关键点间关系图,本文定义关于人体关键点的热点图 $S_{j,k}^*(p)$,如式(1)所示:

$$S_{j,k}^*(p) = \exp\left(-\frac{\|D(p, l_{j,k})\|_2^2}{2\sigma^2}\right) \quad (1)$$

其中, j 代表关于某个类型的人体关键点的热点图; p 代表热点图上的某个点; $l_{j,k}$ 代表某个关键点 j 里面的第 k 个类型的关键点; D 代表某个点与离其最近的关键点的距离; σ 代表控制热点图上值的衰减速度。由于图中不仅仅包含一种类型的人体关键点,热点图中某些点所处的概率值将会受到多个关键点的影响,在这种情况下,本文取该点所受影响最大的点作为最后的值。最终热点图上的点能够表示成式(2)的形式:

$$S_j^*(p) = \max_k S_{j,k}^*(p) \quad (2)$$

本文使用对当前点影响最大的人体关键点作为中心点得到最终的热点图,在预测阶段,预测结果也将预测成类似原标签的热点图。

为了标注人体关键点间关系的热点图,本文定义的 $L_c^*(p)$ 表示使用局部亲和域(part affinity field, PAF)建模骨骼区域,对于骨骼区域的每一个像素,使用2D向量同时表征位置 and 方向信息。假定 $x_{j1,k}$ 和 $x_{j2,k}$ 分别代表关键点 $j1, j2$ 的位置坐标, k, c 分别表示人物索引与关键点索引,则有式(3):

$$L_c^*(p) = \begin{cases} v, & \text{if } p \text{ on limb } c, k \\ 0, & \text{others} \end{cases} \quad (3)$$

其中 v 为位置 $j1$ 到位置 $j2$ 的单位向量,其表达式如下:

$$v = \frac{x_{j2,k} - x_{j1,k}}{\|x_{j2,k} - x_{j1,k}\|} \quad (4)$$

位置 $j1$ 到位置 $j2$ 之间的区域部分以点集 p 表示,数学公式如下:

$$0 \leq v \cdot (p - x_{j1,k}) \leq l_{c,k} \text{ and } |v \cdot (p - x_{j1,k})| \leq \sigma_l \quad (5)$$

其中, $l_{c,k} = \|x_{j2,k} - x_{j1,k}\|_2$ 表示骨骼长度; σ_l 表示骨骼粗细。

在全卷积网络(fully convolutional networks, FCN)中,模型使用ReLU激活函数作为最后一层,本文的模型预测结果为人体的热点图,模型输出的最佳范围为 $0 \sim 1$,在该范围内具有线性性质的激活函数。训练过程中,采用了Tanh函数作为模型的激活函数,Tanh的解析函数为:

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (6)$$

预测结果是关于人体几个关键点的热点图以及人体关键点间的关系图,标注图也是热点图,在这种情况下本文使用欧氏距离 E_{err} 作为损失函数进行训练,如式(7)所示:

$$E_{\text{err}} = \frac{1}{2} \sum_{j=1}^l \sum_{x=1}^w \sum_{y=1}^h (\hat{y}_{x,y}^j - y_{x,y}^j)^2 \quad (7)$$

其中 j 表示人体某个结构的关键点的热点图,取值范围为 $(0,1)$; x, y 分别表示热点图上的横纵坐标; $y_{x,y}^j$ 表示某个类型人体关键点的热点图上横纵坐标为 x, y 的值; $y_{x,y}^j$ 表示与标签热点图相对应的标签点^[17]。

2.2 导数计算

1) 卷积层。假设已知卷积层为 δ^l , 上一层隐藏层为 δ^{l-1} 。在深度学习动态神经网络(dynamic neural network, DNN) 中, 卷积层的前向传播公式为:

$$a^l = \sigma(z^l) = \sigma(W^l a^{l-1} + b^l) \quad (8)$$

其中 W 是权重; a^l 为前向传播算法计算出来的输出值; b 是偏置; l 是输出层数; z 是卷积层; σ 是激活函数。

δ^{l-1} 与 δ^l 的递推关系为:

$$\delta^l = \frac{\partial J(W, b)}{\partial z^l} = \frac{\partial J(W, b)}{\partial z^{l+1}} \frac{\partial z^{l+1}}{\partial z^l} = \delta^{l+1} \frac{\partial z^{l+1}}{\partial z^l} \quad (9)$$

因此要推导出 δ^{l-1} 与 δ^l 的关系, 就必须计算出 $\frac{\partial z^l}{\partial z^{l-1}}$ 的梯度公式。 z^l 与 z^{l-1} 的关系为:

$$z^l = W^l a^{l-1} + b^l = W^l \sigma(z^{l-1}) + b^l \quad (10)$$

通过式(8)、(9)、(10) 能够得到, 反向传播过程中, 数据经过卷积层上每个点中的导数计算公式为:

$$\delta^{l-1} = \delta^l \frac{\partial z^l}{\partial z^{l-1}} = \delta^l \times \text{rot180}(W^l) \odot \sigma'(z^{l-1}) \quad (11)$$

其中, $\odot$ 是哈达玛积, 这个式子和 DNN 类似, 区别在于对含卷积的式子求导时, 卷积核被旋转了 180° , 即式子中的 $\text{rot180}(\cdot)$; 翻转 180° 的意思是上下翻转一次, 再左右翻转一次, 在 DNN 中只是矩阵的转置。求出 δ^{l-1} 的表达式, 就能求出每一层 W^{l-1}, b^{l-1} 的梯度, 对每层进行梯度更新。

2) 池化层。假设已知池化层为 δ^l , 推导出上一隐藏层 δ^{l-1} 的过程如下: 在前向传播过程中, 一般会用最大池化层或者平均池化层进行池化, 池化的区域大小是已知的。反过来则需要缩小误差 δ^l , 还原前一次较大区域对应的误差。

在反向传播时, 首先把 δ^l 的所有子矩阵大小还原成池化之前的大小, 如果是最大池化层, 则把 δ^l 的所有子矩阵各个池化局域的值放到之前对应的位置上, 如果为平均池化层, 则把 δ^l 所有子矩阵的各个池化局域的值取平均后放在还原后的子矩阵位置上, 此过程被称作下采样^[18]。

这样就得到了上一层 $\frac{\partial J(W, b)}{\partial a_k^{l-1}}$ 的值, 要得到 δ_k^{l-1} , 公式为:

$$\delta_k^{l-1} = \left(\frac{\partial a_k^{l-1}}{\partial z_k^{l-1}} \right)^T \frac{\partial J(W, b)}{\partial a_k^{l-1}} = \text{upsample}(\delta_k^l) \odot \delta'(z_k^{l-1}) \quad (12)$$

其中, upsample 函数完成了池化误差矩阵放大与误差重新分配的逻辑。

对于 δ^{l-1} , 计算公式为:

$$\delta^{l-1} = \text{upsample}(\delta^l) \odot \delta'(z^{l-1}) \quad (13)$$

2.3 学习算法

本文在全新框架训练过程中使用了自适应矩估计(adaptive moment estimation, ADAM) 优化器。除了计算过去梯度平方和的指数衰减之外, ADAM 还增加了一项过去梯度平均数的指数衰减, 类似于动量(momentum), 即:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (14)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (15)$$

式中 m_t 和 v_t 是梯度的一阶(均值)和二阶(方差)的估计值, 这些值一般初始化设置为 0, 但是受到 β 的影响变得有偏差, 一般修正如下:

$$g_t^2 \hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (16)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (17)$$

之后的参数更新类似于 AdaDelta 算法^[19] 和 RMSprop 算法^[19] :

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_{t+\epsilon}}} \hat{m}_t$$

(18)

3 实验结果与分析

3.1 实验环境

本文的实验在 MS COCO 测试集^[20]上测试,得到的结果记录到 Json 文件中,并调用官方接口进行测试,实现方式均在 Tensorflow 框架上进行实验。环境处理器为 AMD2700X,显卡为 RTX2080Ti,内存 64 G,显存 11 G。训练设置的初始学习率为 0.001,Batch_size 设置为 32,在数据集上轮训了 1 000 多次,使用随机梯度下降算法进行优化;其次分别用模型进行判断得到不同分辨率的热点图;再次将这些热点图调整到原图相同的大小,进行相加操作得到最终预测图,对这个预测图进行后续的处理得到最终的人体评估结果。图 6 是预测的人体关键点热点图。

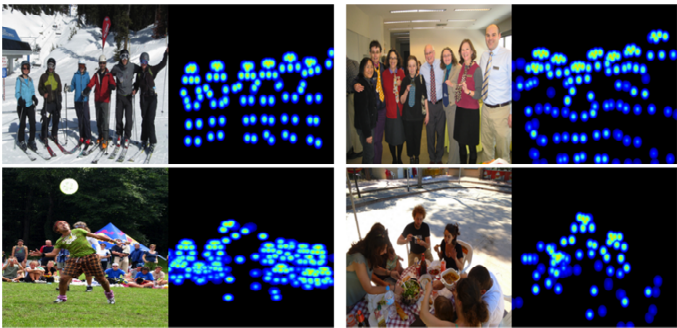


图 6 人体关键点预测热点图

3.2 结果对比与分析

首先使用对象关键点相似性(object keypoint similarity,OKS)判别人体对应关键点的检测效果,最终通过设置某个值,将检测的各种大于此值的 OKS 作为结果,并计算平均精度(average precision,AP)作为评价模型好坏的参数,其中 AP50 和 AP75 分别表示使用 OKS 大于 0.5 和 0.75 时,分别计算出来的平均精确度,而 APM、APL 分别表示中等目标以及大目标的平均精确度。测试结果见表 1。

表 1 多维自学习模块的对比结果

主干模型	指标值/%				
	AP	AP50	AP75	APM	APL
VGGNet	58.4	81.5	62.6	54.5	65.1
ResNet	56.3	80.1	61.4	53.1	64.2
VGGNet+MSLM	58.6	82.8	62.0	55.9	65.9
ResNet+MSLM	56.7	80.3	61.5	53.3	64.8

通过表 1 中的数据可知,主干网络添加 MSLM 后,提升了检测的 AP 指标值,VGGNet 的 AP 指标值从 58.4%提升到 58.6%,提高了 0.3%;ResNet 的 AP 指标值从 56.3%提升到 56.7%,提高了 0.7%。因此可以得到,主干网络添加 MSLM 能够略微地提升模型的效果。

金字塔结构的目的是结合高低级的语义特征,增强模型的表达能力。表 2 为包含了 FPM 的前后对比效果。

从表 2 可以发现,主干网络添加 FPM 后,AP 指标值都发生了改善。VGGNet 的 AP 指标值从 58.4%提升到 58.7%,提高了 0.5%;ResNet 的 AP 指标值从 56.3%提升到 56.8%,提高了 0.8%,由此可得,主干网络添加 FPM 能够提升模型的效果。

表 3 证明,使用了单通道分离模块的网络效果比未使用的效果好。VGGNet 的 AP 指标值从

58.4%提升到 58.7%，提高了 0.5%；ResNet 的 AP 指标值从 56.3%提升到 56.6%，提高了 0.5%，证明 CSEM 能够有效地增强网络效果。

表 2 特征金字塔模块的对比结果

主干模型	指标值/%				
	AP	AP50	AP75	APM	APL
VGGNet	58.4	81.5	62.6	54.5	65.1
ResNet	56.3	80.1	61.4	53.1	64.2
VGGNet+FPM	58.7	81.4	62.8	54.9	66.2
ResNet+FPM	56.8	80.4	61.6	53.3	64.3

表 3 通道分离提取模块的对比结果

主干模型	指标值/%				
	AP	AP50	AP75	APM	APL
VGGNet	58.4	81.5	62.6	54.5	65.1
ResNet	56.3	80.1	61.4	53.1	64.2
VGGNet +CSEM	58.7	83.6	63.2	54.8	65.9
ResNet+CSEM	56.6	82.1	63.3	55.0	66.2

综合上述实验结果并进行整合,表 4 描述了控制不同模块进行实验的结果,由表中的结果可知,单独使用不同的模型应用到网络中都能提升 AP 指标值。

1)将 MSLM 分别应用在以 VGGNet 为主干和以 ResNet 为主干的 CSEM 网络模块中,能够提升模块的 AP 指标值；

2)将 FPM 分别应用到以 VGGNet 为主干和以 ResNet 为主干的 CSEM 网络模块中,AP 指标值均得到提升,证明了此模块的通用性；

3)将这两个模块全都应用到 CSEM 的网络模块中,分别得到了这两个模块的最高正确率,VGGNet 的 AP 指标从 58.4%上升到 60.1%，提高了 2.9%；ResNet 的 AP 指标从 56.3%上升到 57.9%，提高了 2.8%，表明了 CSEM、JSLAPM 作为应用模块添加到主干网络后提高了一定的精确度。

表 4 综合对比结果

主干模型	指标值/%				
	AP	AP50	AP75	APM	APL
VGGNet	58.4	81.5	62.6	54.5	65.1
VGGNet +CSEM	58.7	83.6	63.2	54.8	65.9
VGGNet +CSEM+MSLM	58.8	82.4	63.9	54.8	66.5
VGGNet+CSEM+FPM	59.0	83.1	63.5	55.6	67.1
VGGNet+CSEM+JSLAPM	60.1	84.0	64.1	56.1	67.3
ResNet	56.3	80.1	61.4	53.1	64.2
ResNet+CSEM	56.6	82.1	63.3	55.0	66.2
ResNet+CSEM+MSLM	57.1	82.1	60.7	54.3	66.2
ResNet+CSEM+FPM	57.2	81.8	61.9	54.6	67.0
ResNet+CSEM+JSLAPM	57.9	82.4	62.3	54.7	67.1

4 结 语

本文将 MSLM 应用到主干模块提取得到的特征图上,让其自动地学习特征矩阵中不同通道数的重要性,自动调整不同通道之间特征值的权重;同时,将金字塔结构结合高低级语义特征应用到调整后的特征图上,组成联合自学习的属性金字塔模块(JSLAPM),并以此为新的主干特征。然后,基于上述研究提出

了一种单通道分离提取模块,并应用到人体关键点检测识别任务中。最后,详细地介绍了上述3种模块是如何应用到现有的网络中,并在人体关键点检测数据集上以VGGNet和ResNet为主干网络进行对比实验,证明了所提网络结构模块的有效性。

参考文献:

- [1] 吴刚,朱勇,苏守宝,等. 嵌入DCNN深度特征与半监督学习的梯度提升决策算法[J]. 金陵科技学院学报,2020,36(3): 7-13
- [2] 王宇,方群. 人工智能在微流控系统中的应用[J]. 分析化学,2020,48(4):439-448
- [3] 杨连平,孙玉波,张红良,等. 基于编解码残差的人体关键点匹配网络[J]. 计算机科学,2020,47(6):114-120
- [4] 唐晖,王庆,陈洪,等. 基于深度学习的体感交互方法[J]. 计算机与现代化,2019(2):7-14
- [5] 王睿,朱业安,卢巍. 骨骼关键点检测技术在康复评估中的应用进展[J]. 中国康复医学杂志,2020,35(7):886-890
- [6] 夏瀚笙,沈岨,胡委. 基于人体关键点的分心驾驶行为识别[J]. 计算机技术与发展,2019(7):1-5
- [7] 姚巍巍,张洁. 融合目标检测和人体关键点检测的铁路司机行为识别[J]. 计算机测量与控制,2020,28(6):212-216
- [8] 季晨颖,赵鸣博,李潮. 姿态检测网络在服装关键点检测中的应用[J]. 中国科技论文,2020,15(3):255-259
- [9] 胡保林. 基于深度学习的人体关键点检测[D]. 成都:电子科技大学,2019
- [10] 韩锐,黄泽帆. 基于人体姿态动态特征的跌倒行为识别方法[J]. 湖南大学学报(自然科学版),2020,47(12):69-76
- [11] 周燕,刘紫琴,曾凡智,等. 深度学习的二维人体姿态估计综述[J]. 计算机科学与探索,2020,15(4):1-17
- [12] 丁圣勇,樊勇兵,陈楠. 基于UNet结构的人体骨骼点检测[J]. 广东通信技术,2018,38(11):64-69,77
- [13] 强保华,翟艺杰,陈金龙,等. 基于改进CPMs和SqueezeNet的轻量级人体骨骼关键点检测模型[J]. 计算机应用,2020,40(6):1806-1811
- [14] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii:IEEE Xplore,2017:2117-2125
- [15] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas:Public Access Policy,2016:770-778
- [16] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-10-12)[2021-04-05]. <https://export.arxiv.org/abs/1409.1556>
- [17] 刘杰. 基于深度学习的人体位置与关键点检测[D]. 南京:南京大学,2018
- [18] YAZAN E, TALU M F. Comparison of the stochastic gradient descent based optimization techniques[C]//2017 International Artificial Intelligence and Data Processing Symposium, Texas:IEEE,2017:1-5
- [19] 王雷. 面向人体姿态估计的轻量化神经网络研究[D]. 杭州:浙江大学,2020
- [20] CHEN X, FANG H, LIN T Y, et al. Microsoft COCO captions: data collection and evaluation server[EB/OL]. (2015-03-23)[2021-04-05]. <https://arxiv.org/abs/1504.00325>

(责任编辑:谭彩霞)