Journal of Computer Applications
计算机应用, 2020, 40(6): 1806 - 1811
ISSN 1001-9081
CODEN JYIIDU
2020-06-10
http://www.joca.cn

文章编号:1001-9081(2020)06-1806-06
DOI:10.11772/j.issn.1001-9081.2019101866

基于改进CPMs和SqueezeNet的轻量级人体骨骼关键点检测模型

强保华[1,2],翟艺杰[1],陈金龙[1*],谢　武[1],郑　虹[1],
王学文[2],张世豪[1]

(1. 广西可信软件重点实验室(桂林电子科技大学),广西 桂林 541004;
2. 广西图像图形与智能处理重点实验室(桂林电子科技大学),广西 桂林 541004)

(*通信作者电子邮箱chengjl@guet.edu.cn)

摘　要:针对目前的人体骨骼关键点检测模型参数多、训练时间长和检测速度慢的问题,提出了一种将人体骨骼关键点检测模型CPMs与小型卷积神经网络模型SqueezeNet相结合的检测方法。首先,采用4个Stage的CPMs(CPMs-Stage4)对人物图像进行关键点检测;然后,在CPMs-Stage4中引入SqueezeNet的Fire Module网络结构,利用Fire Module结构大大压缩模型参数,得到一种新的轻量级人体骨骼关键点检测模型SqueezeNet15-CPMs-Stage4。在扩展的LSP数据集上的验证结果显示,与CPMs相比,SqueezeNet15-CPMs-Stage4模型在训练时间上减少86.68%,在单张图像检测时间上减少44.27%,准确率达到90.4%;与改进的VGG-16、DeepCut和DeeperCut三种参照模型相比,SqueezeNet15-CPMs-Stage4模型在训练时间、检测速度和准确率方面均是最优的。实验结果表明,所提模型不仅检测准确率高,而且训练时间短、检测速度快,能够有效降低人体骨骼关键点检测模型的训练成本。

关键词:人体骨骼关键点检测;人体姿态估计;深度学习;卷积神经网络;轻量级;CPMs;SqueezeNet

中图分类号:TP391.4　　**文献标志码**:A

Lightweight human skeleton key point detection model based on improved convolutional pose machines and SqueezeNet

QIANG Baohua[1,2], ZHAI Yijie[1], CHEN Jinlong[1*], XIE Wu[1], ZHENG Hong[1],
WANG Xuewen[2], ZHANG Shihao[1]

(1. *Guangxi Key Laboratory of Trusted Software* (*Guilin University of Electronic Technology*), *Guilin Guangxi* 541004, *China*;
2. *Guangxi Key Laboratory of Image and Graphics Intelligent Processing* (*Guilin University of Electronic Technology*), *Guilin Guangxi* 541004, *China*)

Abstract: In order to solve the problems of too many parameters, long training time and slow detection speed of the existing human skeleton key point detection models, a detection method combining the human skeleton key point detection model called Convolutional Pose Machines (CPMs) and the lightweight convolutional neural network model called SqueezeNet was proposed. Firstly, the CPMs with 4 stages (CPMs-Stage4) was used to detect the key points of the human images. Then, the Fire Module network structure of SqueezeNet was introduced into CPMs-Stage4 to reduce the model parameters greatly, and thus to obtain a new lightweight human skeleton key point detection model called SqueezeNet15-CPMs-Stage4. The verification results on the extended Leeds Sports Pose (LSP) dataset show that, compared with CPMs, SqueezeNet15-CPMs-Stage4 model has the training time reduced by 86.68%, the detection time of single image reduced by 44.27%, and the detection accuracy of 90.4%; and the proposed model performs the best in training time, detection speed and accuracy compared with three reference models improved VGG-16, DeepCut and DeeperCut. The experimental results show that the proposed model achieves high detection accuracy with short training time and fast detection speed, and can effectively reduce the training cost of the human skeleton key point detection model.

Key words: human skeleton key point detection; human pose estimate; deep learning; Convolutional Neural Network (CNN); lightweight; Convolutional Pose Machines (CPMs); SqueezeNet

收稿日期:2019-11-01;**修回日期**:2019-12-20;**录用日期**:2020-01-02。　**基金项目**:国家自然科学基金资助项目(61762025);广西重点研究发展计划项目(AB17195053,AB18126063);广西自然科学基金资助项目(2017GXNSFAA198226);桂林科技发展计划项目(20180107-4)。

作者简介:强保华(1972—),男,河南南阳人,教授,博士,CCF会员,主要研究方向:大数据分析、图像处理;　翟艺杰(1995—),女,河南周口人,硕士研究生,主要研究方向:人体骨骼关键点检测、深度学习;　陈金龙(1979—),男,江西高安人,高级实验师,硕士,主要研究方向:图像处理、机器学习;　谢武(1979—),男,江西宜春人,副教授,博士,CCF会员,主要研究方向:数据挖掘、信息处理;　郑虹(1975—),女,江西吉安人,讲师,博士,主要研究方向:图像处理、机器学习;　王学文(1979—),男,湖北黄冈人,讲师,硕士,主要研究方向:机器学习、机器视觉;　张世豪(1991—),男,河南许昌人,硕士,主要研究方向:人体骨骼关键点检测、图像处理。

0　引言

随着计算机视觉技术的发展，人体姿态估计已经成为众多领域的研究热点，并且得到更普遍的应用，如步态分析[1]、动作捕捉[2]、行为识别[3]和人机交互[4]等。人体骨骼关键点检测是用于人体姿态估计的一类算法，近几年，卷积神经网络的兴起，让人体骨骼关键点检测技术有了很大提升，然而，如何简化模型、提高检测模型的准确率和检测速度仍是目前面临的一个问题。

基于深度学习的人体骨骼关键点检测算法，可以通过一系列深层网络自动学习图像数据中的隐含关系，提取出更抽象的图像特征，具有比传统方法更强的特征表达能力[5]。近年来众多学者对此问题进行了研究，Lifshitz等[6]提出基于16层的VGG（Visual Geometry Group based on 16 layers，VGG-16）网络模型预测人体各关键点位置，但是准确率有待提升。Pishchulin等[7]提出DeepCut，结合Fast R-CNN（Regions with Convolutional Neural Network features）检测人体骨骼关键点，提升了准确率，但是检测速度较慢。之后，Insafutdinov等[8]提出DeeperCut，结合ResNet（Residual Network）进行检测，进一步提高检测精度和速度。2016年，Wei等[9]提出的CPMs（Convolutional Pose Machines）模型在人体骨骼关键点检测的标准数据集MPII（Max Planck Institut Informatik）人体姿态数据集[10]和LSP（Leeds Sports Pose）数据集[11]上都取得不错的检测效果，具有较好的鲁棒性。然而，这种方法仍然具有参数多、训练时间长和检测速度不理想的问题。因此，本文主要研究如何改进人体骨骼关键点检测模型CPMs，以减少模型参数和训练时间、提高检测速度。

2016年出现的轻量级卷积神经网络模型SqueezeNet[12]有效地解决了网络模型参数多的问题。SqueezeNet能达到很好的识别精度，且与其他模型相比参数更少。因此，本文结合CPMs和SqueezeNet的优势，设计了一种基于CPMs和SqueezeNet的轻量级人体骨骼关键点检测模型。本文主要工作如下：

1）针对CPMs模型训练时间长、检测速度慢的问题，采用CPMs-Stage4模型。CPMs-Stage4通过减少两个预测阶段缩短训练时间、提高检测速度。但由于预测阶段较少，CPMs-Stage4的检测准确率有待提升。

2）针对CPMs-Stage4模型检测准确率降低、模型参数多的问题，结合SqueezeNet与CPMs-Stage4设计SqueezeNet15-CPMs-Stage4模型。新模型利用SqueezeNet的网络结构重新设计CPMs-Stage4的第一阶段，一方面，改进后的模型具有更深的网络层数，进而增强新模型的特征提取能力，提高准确率；另一方面，利用SqueezeNet压缩模型权值参数，使新模型具有更少的参数和更快的检测速度。此外，模型训练中引入初始结构，显著降低训练时间。

1　网络模型

1.1　CPMs模型

CPMs是2016年由卡内基梅隆大学（Carnegie Mellon University，CMU）机器人研究所的Wei等[9]提出的使用卷积神经网络进行单人人体骨骼关键点检测的模型，具有鲁棒性好、准确率高的优点。CPMs采用一系列顺序化卷积架构来表达空间和纹理信息，逐步预测使最终结果更精确[9]，CPMs框架如图1所示。每个Stage均是一个预测阶段，Stage1是一个基本的卷积神经网络，Stage>1部分是相同的卷积结构，每个Stage的输出均添加一个“Loss”损失函数，最小化关键点的预测坐标与真实标注坐标之间的距离，“Center map”是一个高斯函数模板，把预测图中的关键点显示在各自的中心区域，最终生成包含各人体骨骼关键点的预测图。

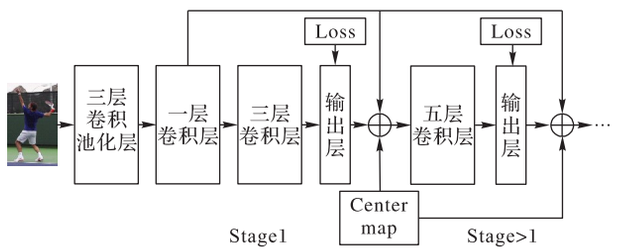


图1　CPMs框架

Fig. 1　Framework of CPMs

CPMs中Stage1对人体骨骼关键点进行粗略预测，从输入图像中直接生成关键点的响应图。Stage>1部分，将特征提取层提取的深度为128的特征图作为纹理信息[13]，前一个Stage输出的响应图作为上下文空间信息，将两者与中心约束三方面信息融合使下一个Stage输出的预测图更准确[14]。每个Stage共输出15个响应图，包括14个关键点和1个背景响应图。各个Stage输入的特征图均进行多尺度处理[9]，将不同尺度的特征图和响应图作为输入，可以避免关键点之间的远近距离对预测图的影响过大，保证预测精度。

CPMs根据输入图像确定一系列缩放尺度，然后预测不同尺度下每个关键点的置信值，最后对不同Stage下各关键点所有尺度的置信值相加求和，将得分最高的置信值作为该关键点的最终预测结果，并将结果在图像中可视化。

CPMs算法的伪代码见算法1，$pd_s()$表示预测因子[15]，预测不同Stage下人体骨骼关键点的位置。定义第n个人体骨骼关键点在图像中的像素位置为$W_n \in Z$，其中Z是图像中所有像素位置(u,v)的集合，要预测的所有N个关键点位置为$W=(W_1, W_2, \cdots, W_N)$。用$\boldsymbol{X}_z$表示图像$z$处的特征图，每个预测因子$pd_s()$根据$\boldsymbol{X}_z$预测各个关键点的位置，生成一个响应图$W_n = z, \forall z \in Z$。预测因子$pd_s()$在$s$阶段预测的第$n$个人体骨骼关键点在图像位置$z$处的得分表示为$\boldsymbol{h}_s^n(W_n = z)$。$\boldsymbol{\varphi}_s()$表示$\boldsymbol{h}_{s-1}$得到的特征图，$pd_s()$表示前一Stage对上下文特征的映射，融合上下文空间信息。

算法1　CPMs。

输入　图像；

输出　标记关键点后的图像。

*multiplier*为缩放尺度，S为总阶段数，$\boldsymbol{X}_z$为Stage=1时输入的特征图，$\boldsymbol{X}'_z$为Stage>1时输入的特征图，允许$\boldsymbol{X}_z$与$\boldsymbol{X}'_z$不同。

1）　for $m=1:$ length(*multiplier*)

2）　　for $s=1:S$

3）　　$s=1$时，$pd_1(\boldsymbol{X}_z) \rightarrow \{\boldsymbol{h}_1^n(W_n=z)\}_{n\in\{0,1,\cdots,N\}}$

4）　　$s>1$时，$pd_s(\boldsymbol{X}'_z, \boldsymbol{\varphi}_s(z, \boldsymbol{h}_{s-1})) \rightarrow \{\boldsymbol{h}_s^n(W_n=z)\}_{n\in\{0,1,\cdots,N+1\}}$

5）　for $s=1:S$

6）　$\sum_{m}^{\text{length}(multiplier)} \{\boldsymbol{h}_s^n(W_n=z)\}_{n\in\{0,1,\cdots,N+1\}}$

7）　得分最高的点作为该关键点的最终预测位置(u,v)。

1.2　CPMs-Stage4模型

Wei等[9]在其研究中将1-Stage至6-Stage对应的CPMs在数据集上的检测准确率做了对比，指出6-Stage的CPMs效果最佳。

虽然6-Stage的CPMs图像提取能力更强，但是模型参数

多和训练时间长等问题影响了模型的检测速度。而 4-Stage 相比 6-Stage 对应的模型减少 2 个预测阶段，在模型参数数量和训练时间上更有优势，而且检测效果较好，因此本文采用论文中 4 个 Stage 的 CPMs-Stage4 进行人体骨骼关键点检测。

然而 CPMs-Stage4 不仅在模型参数和训练时间上提升较小，还存在检测速度不够快的问题。此外，由于模型中预测阶段较少以及网络层数不够深，CPMs-Stage4 的检测准确率也有所下降。为提高准确率、减少模型参数、加快检测速度，一种有效的方法就是增加网络结构深度和卷积层数并且减少权值参数。

1.3 SqueezeNet15-CPMs-Stage4 模型

SqueezeNet 是 2017 年由 Iandola 等[12]提出的一个轻量型的网络模型，该网络模型能保证识别精度，同时将原始 AlexNet 参数压缩至原来的约 1/50，使模型大小只有 4.8 MB。SqueezeNet 模型的核心构件是 Fire Module，Fire Module 将一个卷积层分解为一个 squeeze 层和一个 expand 层，并各自带上 ReLU 激活层，增加网络结构的深度。squeeze 层包含的全部是 1×1 的卷积核，expand 层包含 1×1 和 3×3 的卷积核，每一个 Fire Module 的最后一层用 Average Pooling 层替换全连接层，大幅减少了模型权值参数。SqueezeNet 证明了小的神经网络也能达到很好的识别精度。

为解决 CPMs-Stage4 在准确率、模型参数和检测速度上的问题，本文将 SqueezeNet 的 Fire Module 结构引入 CPMs-Stage4 的 Stage1 中，用 SqueezeNet 的前 15 层 Fire8 替换 Stage1 中一个卷积池化层，并对每个 Stage 新增两个卷积层，提出 SqueezeNet15-CPMs-Stage4 模型。该模型 Stage2-Stage4 的网络结构均相同，每个 Stage 的输出都作为下个 Stage 的融合内容之一。在 Stage1 中，输入图像经过 Fire8 和五层卷积后提取的特征均作为后续每个 Stage 的输入之一，SqueezeNet15-CPMs-Stage4 框架如图 2 所示。

新模型在 Stage1 中引入 Fire Module 结构，同时使用 4 个卷积架构，不仅增加网络结构的深度和卷积层数，而且大量减少权值参数，从而使准确率、模型参数以及检测速度都有很大的提升。

本文模型中 Stage1 的网络结构如表 1 所示，Stage2~Stage4 的网络结构如表 2 所示。表 1 和表 2 表示模型中各 Stage 的网络深度和特征图的变化。

表 1 Stage1 的网络结构

Tab. 1 Network structure of Stage1

名称	层	卷积核	输出尺寸
卷积层 1	1	7×7/2	184×184×64
最大池化层	0	3×3/2	92×92×64
卷积层 2	2	3×3/1	92×92×192
最大池化层	0	3×3/2	46×46×192
Fire8	15	—	46×46×512
卷积层 3	1	3×3/1	46×46×256
卷积层 4	1	3×3/1	46×46×256
卷积层 5	1	3×3/1	46×46×256
卷积层 6	1	3×3/1	46×46×256
卷积层 7	1	3×3/1	46×46×128
新增卷积层 1	1	1×1/1	46×46×512
新增卷积层 2	1	1×1/1	46×46×15

表 2 Stage2~Stage4 的网络结构

Tab. 2 Network structure of Stage2-Stage4

名称	层	卷积核	输出尺寸
卷积层 7_Stage1	1	3×3/1	46×46×128
中心池化层	0	9×9/8	46×46×1
新增卷积层 2_Stage1	1	1×1/1	46×46×15
融合	0	—	46×46×144
卷积层 3	1	7×7/1	46×46×128
卷积层 4	1	7×7/1	46×46×128
卷积层 5	1	7×7/1	46×46×128
卷积层 6	1	7×7/1	46×46×128
卷积层 7	1	7×7/1	46×46×128
新增卷积层 1	1	1×1/1	46×46×128
新增卷积层 2	1	1×1/1	46×46×15

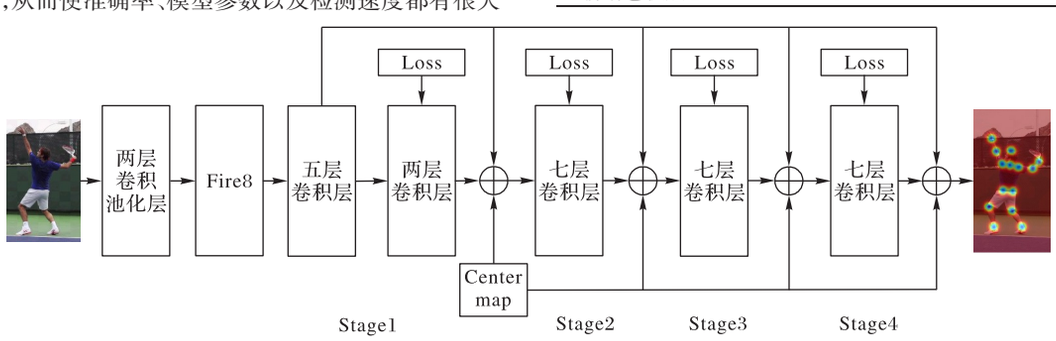


图 2 SqueezeNet15-CPMs-Stage4 框架

Fig. 2 Framework of SqueezeNet15-CPMs-Stage4

2 实验与结果分析

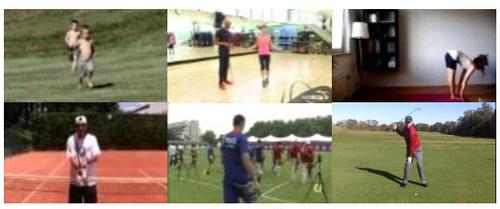
为了评价算法性能，本文使用 MPII 人体姿态数据集、LSP 数据集和扩展的 LSP（LSP extended，LSPet）[16]三个基准数据集进行评估，这三个数据集都是来自于真实的人类日常活动的图像。本文使用关键点正确估计的比例（Percentage of Correct Keypoints，PCK）来评估所有算法的准确率，2.2 节中对 PCK 有详细定义，使用单张图像检测时间反映检测速度，单张图像检测时间越短，检测速度越快。本文从两个方面来验证本文模型算法的性能：第一方面实验展示了在两个不同数据集上训练的 SqueezeNet15-CPMs-Stage4 与 CPMs、CPMs-Stage4 的对比，验证 SqueezeNet15-CPMs-Stage4 模型在提高准确率和加快检测速度方面的有效性；第二方面实验则展示了在扩展的 LSP 上训练的 SqueezeNet15-CPMs-Stage4 与目前主流模型算法性能的对比。

2.1 数据集

MPII 人体姿态数据集是目前评价人体姿态估计的一种最先进的基准数据集。该数据集大约包括 25 000 张图片，其中包含 40 000 多个带有人体关节注释的人，每个人的各个关

节点的位置坐标和可见性都被标注了。将MPII人体姿态数据集中25 000个人作为训练样本，3 000个人作为验证样本，检测范围为全身的14个人体骨骼关键点。

LSP数据集包含2 000张照片，扩展的LSP数据集包含10 000张图片，LSP和扩展的LSP数据集中每幅图片都标注了14个关键点的具体位置。从LSP和扩展的LSP数据集中随机抽取11 000张图片作为训练样本，剩余1 000张图片作为验证样本。从数据集中随机选取的数据集样本示例如图3所示。



(a) MPII样本

(b) LSP样本

(c) 扩展的LSP样本

图3　三个数据集中部分样本示例

Fig. 3　Some samples from three datasets

2.2　实验环境和评估指标

本文采用Caffe深度学习框架作为项目的支持框架，实验中硬件环境采用的CPU是20核的Intel Xeon E5-2698v4，内存为50 GB，GPU是NIVDIA Tesla P100；软件环境中使用LINUX 64 Ubuntu14.04的操作系统，使用Python 2.7作为编程语言，使用Pycharm 2017.1.2作为开发工具。

为验证本文模型在人体骨骼关键点检测中的泛化能力[17]和在准确率和检测速度方面的有效性，本文设置了两组实验，分别在MPII人体姿态训练样本和扩展的LSP训练样本上进行模型训练。

第一组实验是在MPII训练样本上训练的新模型，与Wei等[9]在MPII上训练的CPMs、CPMs-Stage4作对比；第二组实验是在扩展的LSP训练样本上训练的新模型，与Wei等[9]在扩展的LSP上训练的CPMs、CPMs-Stage4作对比。两组实验均在扩展的LSP验证集上进行模型验证。

本文使用目前通用的准确率评估指标PCK作为模型评估的度量标准。PCK定义为模型检测的关键点与正确标注关键点之间的归一化距离，小于某一设定阈值的一定比例p，又称为PCK@p评估方法[14]，常用的PCK评估有PCK@0.5、PCK@0.2。本文选用PCK@0.2作为本文模型在扩展的LSP验证集上的准确率评估标准。

根据PCK@0.2的评估标准，若模型检测的关键点与正确标注关键点之间的像素坐标距离小于人体躯干长度的一定比例0.2时，表示对该关键点检测正确[14]。记人体骨骼关键点检测正确的个数为RD，总检测的人体骨骼关键点个数为AD，则检测准确率的表达式如式(1)所示：

$$accuracy_{\mathrm{PCK@}p}=\frac{RD}{AD} \tag{1}$$

2.3　SqueezeNet15-CPMs-Stage4模型损失函数及训练方法

CPMs的顺序框架提供了一种训练深层网络的方法，通过在每个Stage的输出位置定义一个“Loss”函数，最小化每一个人体骨骼关键点的预测响应图与它的真实标注图之间的距离，从而引导网络模型达到一个预期的检测效果[15]。SqueezeNet15-CPMs-Stage4中每个Stage都会输出第n个人体骨骼关键点的预测响应图，而第n个人体骨骼关键点的真实标注图被记作$\boldsymbol{h}_*^n(W_n=z)$，通过在每个人体骨骼关键点$n$的真实坐标位置放置一个高斯响应，来构造真实标注响应图，定义式(2)为各Stage最小化输出中的代价函数：

$$g_s=\sum_{n=1}^{N+1}\sum_{z\in Z}\left\|\boldsymbol{h}_s^n(z)-\boldsymbol{h}_*^n(z)\right\|_2^2 \tag{2}$$

其中：n遍历每一个人体骨骼关键点；z表示图像位置。取所有Stage代价函数g_s的总和G为最终代价函数，用式(3)表示：

$$G=\sum_{s=1}^{S}g_s \tag{3}$$

采用带动量的随机梯度下降法联合训练所有网络。为了在所有后续阶段共享图像特征，本文网络模型在Stage>1阶段共享相应的卷积层权重(如图2所示)。

模型训练时，由于刚开始的模型还不能学习到很好的特征，检测效果与真实标注点相差较大，导致损失函数值变化较大，容易引起梯度分散。如果可以在原来模型的基础上继续训练，不仅可以减少模型的训练时间，而且可以提高准确率。因此，本文在模型训练时引入SqueezeNet的预训练模型初始化网络，使用微调的方法[18]在SqueezeNet权重的基础上训练模型，以达到提高准确率、缩短训练时间的目的，然后在扩展的LSP验证集上进行验证，本文网络模型的训练方法如表3所示，各参数的设置值来自深度学习的训练经验。

表3　SqueezeNet15-CPMs-Stage4训练方法

Tab. 3　Training method of SqueezeNet15-CPMs-Stage4

参数名称	设置值
学习策略	Step
初始学习率	0.000 080
批处理大小	16
固定迭代次数	120 000
动量	0.9
数据集格式	LMDB
权重衰减	0.000 5
最大迭代次数	300 000

2.4　在数据集上验证

将第一组实验在扩展的LSP验证集上进行验证对比，结果如表4所示。

从表4可以看出，本文提出的模型不仅准确率最高，而且训练时间最少、单张图像检测时间最短，也即检测速度最快。

CPMs-Stage4 虽然比 CPMs 的训练时间要少得多，但是其准确率也低于 CPMs，其主要原因在于 CPMs-Stage4 采用的是 4 个 Stage 的卷积架构，相比 CPMs 少了 2 个卷积架构，因此在提取特征方面要比 CPMs 稍差一些。

表 4 MPII 数据集上训练的模型的验证结果对比

Tab. 4 Comparison of validation results of models trained on MPII dataset

网络模型	训练时间/h	单张图像检测时间/ms	准确率(PCK)/%
CPMs	180.0	260.7	85.54
CPMs-Stage4	77.4	216.3	84.47
SqueezeNet15-CPMs-Stage4	20.8	148.2	86.51

将第二组实验在扩展的 LSP 验证集上进行验证对比，结果如表 5 所示，其中：带“*”表示在模型训练过程中将 MPII 训练样本加入到 LSP 和扩展的 LSP 训练样本中。

从表 5 可以看出，本文提出的模型准确率与 CPMs 相匹敌达到 90.4%，且训练时间和单张图像检测时间最佳，其主要原因在于，与另外两个模型相比，本文模型在增加网络层数和卷积层数的基础上压缩模型权值参数，因此具有较高的准确率、更少的模型参数和更快的检测速度。与 CPMs 相比，本文模型在训练时间上减少 86.68%，在单张图像检测时间上减少 44.27%。

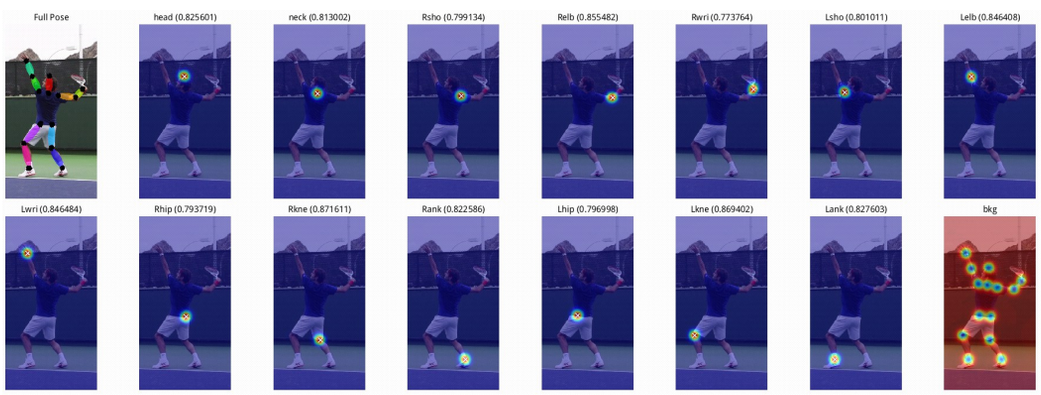
表 5 扩展 LSP 数据集上训练的模型的验证结果对比

Tab. 5 Comparison of validation results of models trained on extend LSP dataset

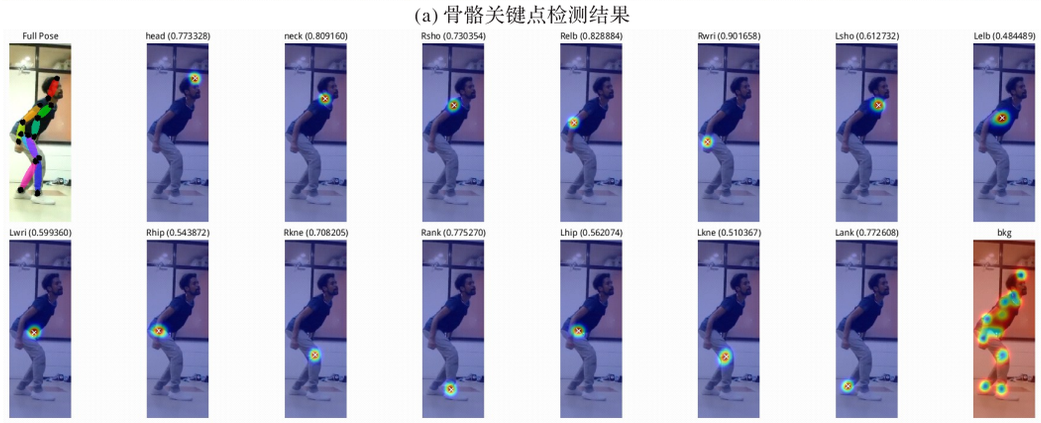
网络模型	训练时间/h	单张图像检测时间/ms	准确率(PCK)/%
CPMs*	280.0	260.7	90.5
CPMs-Stage4*	75.2	215.9	87.3
SqueezeNet15-CPMs-Stage4*	37.3	145.3	90.4

在相同的实验环境下进行的第一组实验和第二组实验的实验数据对比显示，在同等配置下，本文提出的模型训练时间更少，运行速度更快、更稳定，对资源的消耗更少。

本文模型可以检测全身范围内的 14 个骨骼关键点，包括头部、颈部、左肩、右肩、左肘、右肘、左腕、右腕、左髋、右髋、左膝、右膝、左脚踝、右脚踝。随机从验证集中挑选一张图像和部分关键点被遮挡的图像进行检测，两幅图像中 14 个关键点的检测详情如图 4 所示，图中的“Full Pose”为真实关键点标注图，“bkg”为 SqueezeNet15-CPMs-Stage4 模型检测的关键点标注图。从图 4 中两幅图像各自的检测结果可以看出，无论图中关键点有无遮挡，各关键点的检测结果接近真实标注关键点。



(a) 骨骼关键点检测结果



(b) 有遮挡骨骼关键点检测结果

图 4 SqueezeNet15-CPMs-Stage4 关键点检测结果

Fig. 4 Key point detection results of SqueezeNet15-CPMs-Stage4

将在扩展的 LSP 训练集上训练好的本文模型(SqueezeNet15-CPMs-Stage4*)从训练时间、单张图像检测时间和准确率三方面与改进 VGG-16[6]、DeepCut[7]、DeeperCut[8]和 CPMs[9]等人体骨骼关键点检测模型作对比，结果如表 6 所示。

从表 6 可以看出，本文模型在单张图像检测时间上，相较改进 VGG-16 模型减少 79.24%，相较 DeeperCut 模型减少 36.83%。与上述主流的人体骨骼关键点检测模型相比，本文模型的检测准确率不仅与 CPMs 公开的准确率相匹敌，且具有最快的训练时间和检测速度，明显优于其他几种参照模型。

表6　本文模型与参照模型的对比

Tab. 6　Comparison between proposed model and reference models

网络模型	训练时间/h	单张图像检测时间/ms	准确率(PCK)/%
改进VGG-16	—	700.0	86.7
DeepCut	—	57 995.0	87.1
DeeperCut	120.0	230.0	90.1
CPMs	280.0	260.7	90.5
SqueezeNet15-CPMs-Stage4*	37.3	145.3	90.4

3　结语

本文提出了一种基于CPMs和SqueezeNet的单人人体骨骼关键点检测模型，该模型在训练时间和检测速度方面均优于主流参照模型。CPMs是一种鲁棒性好、准确率高的人体骨骼关键点检测模型，SqueezeNet是一种识别精度高，模型参数小于0.5 MB的轻量级卷积神经网络模型。实验结果表明，准确率较高的人体骨骼关键点检测模型与识别精度高的轻量级网络模型相结合设计新模型的方法是有效的，改进后的模型，不仅具有较高的准确率，而且大大缩短了模型训练时间、提高了检测速度。后期将继续研究如何改进本文模型，使模型参数更少，以及如何将其他人体骨骼关键点检测经典模型与识别精度高的模型相结合，设计新的网络模型。

参考文献（References）

[1] YANG X J. Human recognition using multi-frame gait silhouette matching [J]. International Journal of Advancements in Computing Technology, 2012, 4(22): 788-795.

[2] MOESLUND T B, HILTON A, KRÜGER V. A survey of advances in vision-based human motion capture and analysis [J]. Computer Vision and Image Understanding, 2006, 104(2/3): 90-126.

[3] WU X, LIANG W, JIA Y. Action recognition feedback-based framework for human pose reconstruction from monocular images [J]. Pattern Recognition Letters, 2009, 30(12): 1077-1085.

[4] HUANG C L, CHUNG C Y. A real-time model-based human motion tracking and analysis for human-computer interface systems [J]. EURASIP Journal on Advances in Signal Processing, 2004, 2004(11): Article No. 616891.

[5] 郑远攀，李广阳，李晔. 深度学习在图像识别中的应用研究综述[J]. 计算机工程与应用，2019，55(12)：20-36.（ZHENG Y P, LI G Y, LI Y. Survey of application of deep learning in image recognition [J]. Computer Engineering and Applications, 2019, 55(12): 20-36.）

[6] LIFSHITZ I, FETAYA E, ULLMAN S. Human pose estimation using deep consensus voting [C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9906. Cham: Springer, 2016: 246-260.

[7] PISHCHULIN L, INSAFUTDINOV E, TANG S, et al. DeepCut: joint subset partition and labeling for multi person pose estimation [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 4929-4937.

[8] INSAFUTDINOV E, PISHCHULIN L, ANDRES B, et al. DeeperCut: a deeper, stronger, and faster multi-person pose estimation model [C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9910. Cham: Springer, 2016: 34-50.

[9] WEI S E, RAMAKRISHNA V, KANADE T, et al. Convolutional pose machines [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 4724-4732.

[10] ANDRILUKA M, PISHCHULIN L, GEHLER P, et al. MPII human pose dataset [DB/OL]. [2019-04-07]. http://human-pose.mpi-inf.mpg.de.

[11] JOHNSON S, EVERINGHAM M. Leeds sports pose dataset [DB/OL]. [2019-04-07]. http://sam.johnson.io/research/lsp.html.

[12] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size [EB/OL]. [2019-02-21]. https://arxiv.org/pdf/1602.07360.pdf.

[13] RAMAKRISHNA V, MUNOZ D, HEBERT M, et al. Pose machines: articulated pose estimation via inference machines [C]// Proceedings of the 2014 European Conference on Computer Vision, LNCS 8690. Cham: Springer, 2014: 33-47.

[14] 张世豪. 基于深度学习的人体骨骼关键点检测方法研究[D]. 桂林：桂林电子科技大学，2019：38-42.（ZHANG S H. Research on human pose estimation method based on deep learning [D]. Guilin: Guilin University of Electronic Technology, 2019: 38-42.）

[15] QIANG B, ZHANG S, ZHAN Y, et al. Improved convolutional pose machines for human pose estimation using image sensor data [J]. Sensors, 2019, 19(3): Article No. 718.

[16] JOHNSON S, EVERINGHAM M. Leeds sports pose extended training dataset [DB/OL]. [2019-04-07]. http://sam.johnson.io/research/lspet.html.

[17] 武妍，张立明. 神经网络的泛化能力与结构优化算法研究[J]. 计算机应用研究，2002，19(6)：21-25，84.（WU Y, ZHANG L M. A survey of research work on neural network generalization and structure optimization algorithms [J]. Application Research of Computers, 2002, 19(6): 21-25, 84.）

[18] MOHAMED A, HINTON G, PENN G. Understanding how deep belief networks perform acoustic modeling [C]// Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2012: 4273-4276.

This work is partially supported by the National Natural Science Foundation of China(61762025), the Guangxi Key Research and Development Program (AB17195053, AB18126063), the Natural Science Foundation of Guangxi(2017GXNSFAA198226), the Guilin Science and Technology Development Program(20180107-4).

QIANG Baohua, born in 1972, Ph. D., professor. His research interests include big data analysis, image processing.

ZHAI Yijie, born in 1995, M. S. candidate. Her research interests include human skeleton key point detection, deep learning.

CHEN Jinlong, born in 1979, M. S., senior experimentalist. His research interests include image processing, machine learning.

XIE Wu, born in 1979, Ph. D., associate professor. His research interests include data mining, information processing.

ZHENG Hong, born in 1975, Ph. D., lecturer. Her research interests include image processing, machine learning.

WANG Xuewen, born in 1979, M. S., lecturer. His research interests include machine learning, machine vision.

ZHANG Shihao, born in 1991, M. S. His research interests include human skeleton key point detection, image processing.