湖南农业大学学报(自然科学版)　2021，47(5)：603–608．DOI：10.13331/j.cnki.jhau.2021.05.018
Journal of Hunan Agricultural University(Natural Sciences)

引用格式：
李子茂，余慧，夏梦，郑禄，徐杰．基于图像特征融合的农事活动行为的识别[J]．湖南农业大学学报(自然科学版)，2021，47(5)：603–608．
LI Z M，YU H，XIA M，ZHENG L，XU J．Recognition of the agricultural activities based on image feature fusion[J]．Journal of Hunan Agricultural University(Natural Sciences)，2021，47(5)：603–608．
投稿网址：http://xb.hunau.edu.cn

基于图像特征融合的农事活动行为的识别

李子茂[1,2]，余慧[1,2]，夏梦[1,2]，郑禄[1,2]，徐杰[1,2]

(1.中南民族大学计算机科学学院，湖北 武汉 430074；2.湖北省制造企业智能管理工程技术研究中心，湖北 武汉 430074)

摘　要：针对农事活动图像中人体姿态所隐含的行为信息以及人与农具所隐含的关联信息，提出了一种基于图像特征融合的农事活动行为的识别方法：利用人体姿态估计技术 OpenPose 提取农事行为关节点位置信息，利用目标检测 YOLOv3 提取农事行为中农具的位置和分类信息，用以构建农事行为的距离空间特征矩阵和角度空间特征矩阵，并将这些特征进行图像特征融合，建立基于图像显式特征和隐式特征融合的农事活动行为识别方法 EI–SVM，实现农事活动行为的识别。试验结果表明，EI–SVM 方法对农事活动行为识别的准确率可达 94.87%，在公用数据集上准确率达到 92.39%。

关　键　词：农事活动；人体姿态；行为识别；目标检测；图像特征融合

中图分类号：TP391　　文献标志码：A　　文章编号：1007–1032(2021)05–0603–06

Recognition of the agricultural activities based on image feature fusion

LI Zimao[1,2]，YU Hui[1,2]，XIA Meng[1,2]，ZHENG Lu[1,2]，XU Jie[1,2]

(1.College of Computer Science, South-Central University for Nationalities, Wuhan, Hubei 430074, China; 2.Hubei Provincial Engineering Research Center for Intelligent Management of Manufacturing Enterprises, Wuhan, Hubei 430074, China)

Abstract: Aiming at the behavior information implied by human posture in agricultural activity images and the association information implied by human and agricultural tools, a recognition method of agricultural activity behavior based on image feature fusion is proposed. Human posture estimation technology OpenPose is used to extract the joint position information of agricultural behavior, and target detection YOLOv3 is used to extract the position and classification information of agricultural tools in agricultural behavior. These information is used to construct the distance space feature matrix and angle space feature matrix of agricultural behavior, to fuse the above image features. Based on explicit and implicit features, the recognition method EI-SVM was establish to realize the recognition of agricultural activity behavior. The experimental results show that the accuracy of EI-SVM method for agricultural activity behavior recognition is 94.87%, and the accuracy on public data set is 92.39%.

Keywords: agricultural activities; body posture; behavior identification; target detection; image feature fusion

计算机视觉技术在行为识别上的应用，传统方法是手工提取图像特征[1]，近几年利用机器学习或深度学习[2]提取图像特征并进行特征融合成为常用的方法[3–4]。对静态图像行为的特征挖掘发现，人体姿态可作为行为识别的一个较好的出发点。利用人体姿态估计[5]对于农事活动行为识别，可以利用图像数据中的表象特征，如农具类别、肢体姿态来区分单张静态图像的动作[6]。人类农事活动行为有

收稿日期：2020–11–01　　修回日期：2021–08–26
基金项目：湖北省科学技术厅创新专项(2019ABA101)
作者简介：李子茂(1974—)，男，山东临清人，博士，教授，主要从事算法设计与分析和计算复杂性研究，lizm@mail.scuec.edu.cn

一些比较显著的特征，如除草时人的腿部呈交叉状、手部与身躯存在一定角度等。如仅仅将人体姿态估计运用到行为识别任务中还不足以描述行为动作，还应考虑所使用农具在行为动作交互时的特征，如除草行为以农具锄头作为行为的交互工具；因此，利用目标检测[7]对农事图像中的农具进行特征提取，可最大化地利用图像中的特征。

在基于特征提取的行为识别中，TANG 等[8]提出基于图像的神经网络来捕捉关节连接点之间的依赖关系，进行行为识别，但该方法仅提取了人体信息，丢失了与人交互的物体信息；CHOUTAS 等[9]使用人体关节作为关键点，编码后将获取的特征图输送到简单的 CNN 中，即可用来进行行为识别分类，但该方法对于相似行为的动作识别度不高；LIU 等[10]设计了一种基于骨架的人类动作 HDS–SP 描述符，利用空间和时间信息基于骨架实现行为识别，但由于骨架图仅能提取关节的局部物理依赖关系，难以捕捉隐藏的隐式关节相关性；LI 等[11]提出了 A–link 推理模块的编码解码器结构，直接从动作中捕获潜在依赖性，扩展骨架图以表示更高阶的依赖，提出了行动结构图卷积网络，该方法准确度较高，但网络也较为复杂。针对 GCN 网络仅捕获关节之间的局部物理依赖关系，并且不加选择地使用所有骨骼数据的问题，DING 等[12]提出了基于语义的图形卷积网络，实现骨架特征的灵活图形表示，用于基于骨架的行为识别，但该网络复杂度也较高。

将人体姿态估计和目标检测技术运用到人、物交互的图像行为识别中具有更好的识别效果，在农事活动行为的人与农具交互活动中，对农具和人体姿势的相互关系进行建模[13–14]，基于此，笔者首先利用农事活动的人体姿态估计 OpenPose 模型[15]得到的关节点坐标作为显式特征，构建农事活动行为空间关系矩阵，实现农事活动行为的隐式行为特征构建；再利用目标检测 YOLOv3 模型[16]获取农具的分类和位置信息作为显式特征，构建农具与对应农事活动行为的人体关节点的空间关系矩阵，实现农事活动行为的隐式关系特征构建；融合显式特征与隐式特征作为特征向量，作为分类器支持向量机(SVM)[17]的输入，训练模型获取分类结果，以期得到更准确的农事活动行为识别方法。

1 农事活动行为特征的构建

不同类别的农事活动图像中可能存在相似的行为(例如除草和浇水等)，即使是同一行为类别的图像，也可根据它们之间的细小差别，来区分不同行为类型(例如除草动作有正在除草或者拿着农具在休息等)，仅通过人体姿态的特征并不能较好地识别农事活动的行为。对于农事活动行为，人体姿态和使用工具都是影响行为动作识别的因素。为解决 2 种行为可能存在类似行为动作的问题，区分农具相同但行为动作差异较大的图像，拟通过人体姿态估计获取人体关节点信息，利用目标检测获取农事行为中的农具特征信息，结合两者提取出图像显式特征，并且利用显式特征手工构建隐式特征，增强静态图像表示能力，实现对农事行为特征的最大化提取。

1.1 显式特征的提取

利用 OpenPose 模型[15]的第 1 个卷积层分支获取人体肢体部位的置信图信息，对关节点的坐标进行预测；再在下一个分支对输入的特征进行卷积训练，输出关节点亲和度向量场，输入关节点亲和度向量场和原始图像特征，获取人体身体部位的亲和域信息，对关节点的连接组成集合，提取静态图像的姿态特征，进行人体姿态特征构建。通过 OpenPose 获取人体骨架图及 18 个关节点坐标，作为图像中行为动作的显式信息。

采用 YOLOv3 模型[16]对农事活动图像中不同大小的农具进行检测。由于农事行为存在人和农具的交互，通过 YOLOv3 获取农具的类别和位置信息，作为图像中行为动作的显式信息。

考虑到农事活动图像中人与农具之间及人体姿态本身所隐含的关联结构信息，从显式特征中提取隐式特征，引入关键关节点与农具之间的距离作为隐式特征中的关系特征，并且引入行为动作人体姿态的关节角度作为隐式特征中的行为特征，将 2 种特征作为静态图像行为动作的隐式特征。

1.2 隐式特征的构建

根据不同类型农事活动行为整体所具备的特征，手工选取与农事活动行为动作关联较为密切的 8 个关节点，通过 OpenPose 提取了 18 个关节点坐标，通过归一化处理已知坐标序列中的(3,4,6,7,9,10,

12,13)，即右肘、右手首、左肘、左手首、右膝、右足首、左膝、左足首的 8 个坐标以及农具的位置坐标，计算 8 个关节点与农具两点连线后的距离 $D(i)$，构建农事活动行为关键关节点的距离空间矩阵(DSM)。

为了加强图像特征的表示，在获取图像人体姿态显式特征的基础上，提取图像中人体姿态所特有的结构作为隐式特征中的行为特征。根据不同类型农事活动行为整体所具备的特征，手工选取 6 个角度变化较为明显的关节点对。通过人体姿态估计坐标图的已知坐标序列中的(2,3,4)，(5,6,7)，(1,8,9)，(8,9,10)，(1,11,12)，(11,12,13)，即右肘角度、左肘角度、右腰角度、右膝角度、左腰角度、左膝角度，计算每一个数据中两点连线后形成的角度 $A(j)$，构建农事活动行为关键关节点的角度空间矩阵(ASM)。

2　基于特征融合的 EI–SVM 方法

通过 OpenPose 模型提取图像中 18 个人体骨骼点坐标，将坐标点归一化处理后与关节点次序一一对应进行存储。通过 YOLOv3 模型提取图像中农具的位置和分类信息，将检测框的中心点坐标归一化处理后记为农具的位置信息，分类信息由模型输出即可获得。此时显式特征向量由 18 个[x,y]关节坐标点数据，以及 1 个[x,y,$label$]农具数据组成，最终，图像的显式特征向量的维度为[1,39]。

隐式特征中包含关系隐式特征和行为隐式特征。2 个特征分别由距离空间矩阵(DSM)与角度空间矩阵(ASM)描述。为了提取 DSM，在训练阶段提取 8 个关节点与农具的距离，提取所有训练集中距离的最大值与最小值形成 DSM，此时图像的关系隐式特征向量由 8 个[$d_{\min}$, $d_{\max}$]数据组成。为了提取 ASM，在训练阶段提取手工选择的关键关节点形成的角度，提取所有训练集中角度的最大值与最小值形成 ASM，此时图像的行为隐式特征向量由 6 个[$a_{\min}$, $a_{\max}$]数据组成。最终融合关系和行为隐式特征，隐式特征向量的维度为[1,28]。其中，$d_{\min}$ 和 $d_{\max}$ 分别表示关系隐式特征向量中的最小距离和最大距离，$a_{\min}$ 和 $a_{\max}$ 分别表示行为隐式特征向量中的最小角度和最大角度。

将图像的显式特征向量和隐式特征向量组成总特征向量，共同输入到分类器 SVM 中。通过寻优找到模型合适的核函数以及相关参数，利用训练好的模型对预测集进行评判。

根据显式和隐式图像特征，提出基于图像显式特征(explicit features)和隐式特征(implicit features)融合的农事活动行为方法——EI–SVM。结构上分为两部分，第一部分为特征提取，第二部分为特征构建。特征提取即利用目标检测和人体姿态估计提取已有的显式特征，特征构建即利用显式特征构建隐式特征。

训练模型：

a) 利用 OpenPose 计算人体关节点坐标并将其归一化；

b) 利用坐标点计算关键关节点的角度；

c) 提取训练集中每个角度对应的最大值与最小值，构建 ASM；

d) 标注图像中农具，训练 YOLOv3 模型，对数据集构建合适的锚框(AnchorBox)；

e) 使用训练效果最佳的 YOLOv3 模型提取农具的位置和分类信息；

f) 利用 YOLOv3 提取的农具检测框中心点坐标，计算中心点到关节点的距离；

g) 提取训练集中所有 2 个坐标点之间的最大值与最小值，构建 DSM；

h) 融合以上特征作为特征向量，输入到 SVM 分类器，对模型进行训练。

预测模型：

i) 分别利用 OpenPose 模型与 YOLOv3 模型提取图像中的人体关节点坐标和农具位置以及分类信息；

j) 构建 DSM，此时最大值 $d_{\max}$ 为目标检测得到的检测框的中心点坐标到关节点的距离，最小值 $d_{\min}$ 为检测框的角点到关节点的距离；

k) 构建 ASM，此时最大值 $a_{\max}$ 为 3 个关节点形成的角度，最小值 $a_{\min}$ 为训练集中对应关节点位置的最小值；

l) 融合以上特征作为待识别图像的特征向量，输入到训练好的 SVM 分类模型中，分类输出结果。

3　结果与分析

3.1　数据集

采用自建数据集对提出的行为识别方法进行评价，拍摄 3 类农事活动图像，大小 3456 像素×4608 像素，每一类农事行为都有休息状态，而休息状态

的图像在行为识别分类时将与工作状态的图像有类别区分，如图 1–4 除草休息状态在行为分类时将不会作为图 1–1 除草类别。调整图像大小后，输入到神经网络中提取特征，其中训练图像和测试图像按 8∶2 随机分配。为增强模型的泛化能力，对原始图像进行数据增强处理，如图像翻转、缩放，对比度、锐度、色调、色彩饱和度调节等。

1 除草；2 浇水；3 喷药；4 除草休息状态；5 浇水休息状态；6 喷药休息状态。

图 1 农事行为原始图像

Fig.1 Image samples of agricultural behavior

为验证模型的有效性，选取 PPMI 公开数据集(为人与乐器交互行为的数据集，与人与农具交互行为存在类似性)进行验证。该数据集包含 2 种场景且每种场景对应 12 类图像，其中 2 种场景为人与乐器进行交互的行为动作和手持乐器非交互动作，用以对应农事活动数据集中的工作状态和休息状态。12 类图像分别对应 12 种不同的乐器，每一类约 200 张，对该数据集进行图像增强处理。

3.2 试验环境设置

试验采用 PyCharm 集成开发环境，在 Python 3.7.2 环境下进行，计算机配置为主频 3.40GHz4，核 CPU，8G 内存。

3.3 行为识别结果

经特征提取后，由 OpenPose 提取的显式和隐式特征的维度共为[1,48]。对 YOLOv3 进行模型训练，其中每批次图像 50 帧，学习率调节参数步长 10 000，速率 0.001，提取出 9 个最佳锚框(AnchorBox)，提取出农具显式特征并与人体关节点构建的距离隐式特征的维度共为[1,19]。将大小为[*N*,67]的特征向量作为 SVM 的输入进行类别分类，其中 *N* 为训练集个数。选取了 SVM 中 4 个适用于分类的核函数(linear 线性核函数、poly 多项核函数、rbf 高斯核函数和 sigmoid 核函数)进行对比。由于试验数据集在数据增强后约有 10 000 张，通过 GridSearchCV 进行参数寻优，实现自动调参，并用 SVM 中用于分类的 SVC 方法进行分类。结果表明，rbf 核函数的效果最佳，支持向量的维度为[168,74,37,69]。利用 rbf 核函数处理非线性分类问题，不同的惩罚系数 *C* 和参数 gamma 对模型有不同的评价。*C* 越小，容易欠拟合；*C* 越大，容易过拟合。gamma 值越大，支持向量越少；gamma 值越小，支持向量越多。通过十叠交叉验证对网格形式的参数(*C*，gamma)进行搜索，对比每一组参数的交叉验证精度，得到最优参数组合：核函数 rbf，*C* 为 35，gamma 值为 0.6，农事行为识别准确率为 94.87%。

图 2 和图 3 分别为经过交叉验证后，在最优 gamma 参数下不同 *C* 的识别准确率，以及在最优参数 *C* 下不同 gamma 参数的识别准确率。对于模型

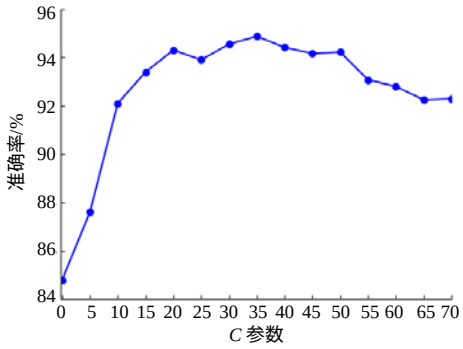

图 2 不同 *C* 的农事行为分类准确率

Fig.2 Dependence of the classification accuracy on C

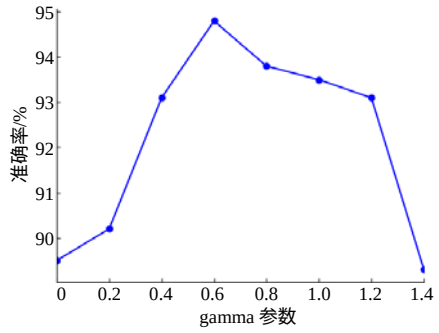

图 3 不同 gamma 参数的农事行为分类准确率

Fig.3 Dependence of the classification accuracy on gamma

训练的结果以及 C 和 gamma 值的最佳取值可见，在最佳惩罚系数 C 下，模型拟合效果尚佳；在最佳 gamma 参数范围的平均值下，支持向量较多，说明了数据集中存在类别相似的行为动作。

3.4　行为识别方法的有效性

为验证图像多特征在行为识别任务中的有效性，分别对图像的单个特征和多特征以及多特征中的显式特征与隐式特征融合进行行为识别。其中，人体姿态单特征表示仅利用人体姿态估计对图像进行行为识别；农具单特征表示仅利用目标检测对图像进行行为识别；人体姿态+农具显式特征表示仅利用显式特征进行行为识别；人体姿态+农具+距离特征表示利用显式特征和隐式特征中的距离空间矩阵作为总特征进行行为识别；人体姿态+农具+角度特征表示利用显式特征和隐式特征中的角度空间矩阵作为总特征进行行为识别；人体姿态+农具+距离和角度特征表示利用显式特征和隐式特征作为总特征进行行为识别。从表 1 可以看出，多特征比单特征表现出更好的识别能力，最终基于显式特征和隐式特征的融合特征，在农事活动数据集上表现出最好的识别效果。

表 1　不同特征农事行为的分类识别准确率

Table 1　Classification accuracy of farming behavior with different characteristics

分类方法	识别准确率/%			
	除草	浇水	喷药	休息状态
农具单显式特征	72.40	71.06	78.61	53.23
人体姿态单显式特征	79.82	79.25	77.43	73.17
人体姿态+农具显式特征	90.48	89.77	87.52	85.71
人体姿态+农具+距离特征	92.87	91.32	93.83	83.27
人体姿态+农具+角度特征	93.05	90.16	91.78	87.81
人体姿态+农具+距离和角度特征	96.31	92.73	96.65	92.88

为了验证 EI–SVM 方法的有效性，选用试验寻优参数为 rbf 核函数来进行验证，在 rbf 核函数下取不同的 C 和 gamma 值分别对人体姿态单特征、农具单特征以及多特征进行模型准确度的比较。

如图 4 和图 5 所示，在利用多特征进行农事活动行为识别时，其识别准确率高于单特征的。在人体姿态单特征和农具单特征中，人体姿态单特征表现效果较好。说明人体姿态特征是农事活动行为识别的有效特征。通过这 2 种特征可最大化利用图像的特征，从而实现对农事活动行为的识别。

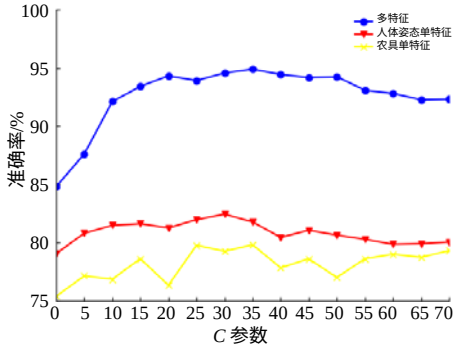

图 4　相同 C 的农事行为不同特征的识别准确率

Fig.4　Dependence of the classification accuracy on C for different features

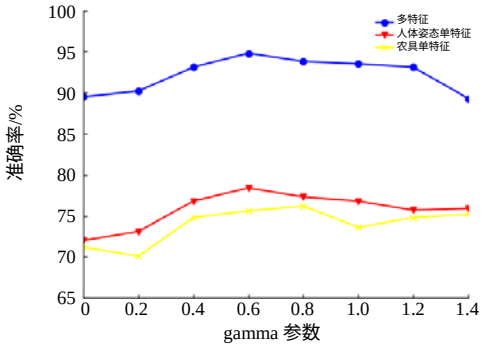

图 5　相同 gamma 参数的农事行为不同特征的识别准确率

Fig.5　Dependence of the classification accuracy on gamma for different features

将 EI–SVM 方法对 PPMI 公开数据集进行行为估计，与农事行为的特征类似，每一类分为 play 和 with 2 种行为，play 表示行为动作的正式状态，with 表示行为动作的休息状态。经试验验证其准确率为 92.39%，表明所提出的基于图像多特征进行行为识别的方法可行并有效。

参考文献：

[1]　赵朵朵，章坚武，郭春生，等．基于深度学习的视频行为识别方法综述[J]．电信科学，2019，35(12)：99–111．
ZHAO D D，ZHANG J W，GUO C S，et al．A survey of video behavior recognition based on deep learning[J]．Telecommunications Science，2019，35(12)：99–111．

[2]　蔡强，邓毅彪，李海生，等．基于深度学习的人体行为识别方法综述[J]．计算机科学，2020，47(4)：85–93．
CAI Q，DENG Y B，LI H S，et al．Survey on human action recognition based on deep learning[J]．Computer Science，2020，47(4)：85–93．

[3]　李衡霞，龙陈锋，曾蒙，等．一种基于深度卷积神经网络的油菜虫害检测方法[J]．湖南农业大学学报(自然科学版)，2019，45(5)：560–564．
LI H X，LONG C F，ZENG M，et al．A rape pest

detection method based on deep convolution neural network [J]. Journal of Hunan Agricultural University (Natural Sciences)，2019，45(5)：560–564.

[4] 张重阳，陈明，冯国富，等. 基于多特征融合与机器学习的鱼类摄食行为的检测[J]. 湖南农业大学学报(自然科学版)，2019，45(1)：97–102.
ZHANG C Y，CHENG M，FENG G F，et al. Detection of fish feeding behavior based on multi feature fusion and machine learning [J]. Journal of Hunan Agricultural University(Natural Sciences)，2019，45(1)：97–102.

[5] 邓益侬，罗健欣，金凤林. 基于深度学习的人体姿态估计方法综述[J]. 计算机工程与应用，2019，55(19)：22–42.
DENG Y N，LUO J X，JIN F L. Overview of human pose estimation methods based on deep learning[J]. Computer Engineering and Applications，2019，55(19)：22–42.

[6] 姜夕凯，苏松志，李绍滋，等. 基于单张静态图像的人体行为识别方法综述[J]. 漳州师范学院学报(自然科学版)，2011，24(4)：23–26.
JIANG X K，SU S Z，LI S Z，et al. A survey of recognizing action from single still images[J]. Journal of Zhangzhou Normal University(Natural Science)，2011，24(4)：23–26.

[7] 曹燕，李欢，王天宝. 基于深度学习的目标检测算法研究综述[J]. 计算机与现代化，2020(5)：63–69.
CAO Y，LI H，WANG T B. A survey of research on target detection algorithms based on deep learning[J]. Computer and Modernization，2020(5)：63–69.

[8] TANG Y，TIAN Y，LU J，et al. Deep progressive reinforcement learning for skeleton-based action recognition[C]//IEEE/CVF . Conference on Computer Vision and Pattern Recognition. Salt Lake City：CVPR，2018：5323–5332.

[9] CHOUTAS V，WEINZAEPFEL P，REVAUD J，et al. PoTion ： pose motion representation for action recognition[C]//IEEE/CVF . Conference on Computer Vision and Pattern Recognition. Salt Lake City：CVPR，2018：7024–7033.

[10] LIU J，WANG Z，LIU H. HDS-SP：a novel descriptor for skeleton-based human action recognition[J]. Neurocomputing，2020，385：22–32.

[11] LI M，CHEN S，CHEN X，et al. Actional-Structural graph convolutional networks for skeleton-based action recognition[C]//IEEE/CVF. Computer Vision and Pattern Recognition. Vancouver：CVPR，2019：3595–3603.

[12] DING X，YANG K，CHEN W. A semantics-guided graph convolutional network for skeleton-based action recognition[C]//ICIAI . 2020 the 4th International Conference on Innovation in Artificial Intelligence. Overseas Chinese University：ICIAI，2020，2138–2206.

[13] XU B，LI J，WONG Y，et al. Interact as you intend：intention-driven human-object interaction detection[C]. //IEEE/CVF. 2019 IEEE Transactions on Multimedia. Vancouver：CVPR，2019：1423–1432.

[14] BOUALIA S N，ESSOUKRI BEN AMARA N. Pose-based human activity recognition：a review[C]// IEEE/CVF. 2019 15th International Wireless Communica- tions & Mobile Computing Conference. Tangier，Morocco，IWCMC. 2019：1468–1475.

[15] 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(5)：7291–7299.

[16] REDMON J，FARHADI A. YOLOv3：An Incremental Improvement[EB/OL]. arXiv e-prints，2018，128(4)：68–76.

[17] CORTES C，VAPNIK V N . Support-vector networks[J]. Machine Learning，1995，20(3)：273–297.

责任编辑：罗慧敏
英文编辑：吴志立