

吊装机器人肢体动作指令识别技术研究

倪涛¹ 邹少元¹ 刘海强¹ 黄玲涛¹ 陈宁² 张红彦¹
(1. 吉林大学机械与航空航天工程学院, 长春 130022; 2. 集美大学机械与能源工程学院, 厦门 361021)

摘要: 鉴于 Kinect 相机进行肢体识别监控距离有限,提出使用网络大变焦摄像头、构建 CNN-BP 融合网络进行肢体动作识别,并以 9 组机器人吊装指令为例进行训练和识别。首先,基于 OpenPose 提取 18 个骨架节点坐标,生成 RGB 骨架图和骨架向量;然后,采用迁移学习方法对 RGB 骨架图使用 InceptionV3 网络提取图像深层抽象特征,并对训练数据集采用旋转、平移、缩放和仿射多种数据增强方式,以扩充训练数据,防止过拟合;再将提取的骨架向量使用 BP 神经网络提取点线面等浅层特征;最后对 InceptionV3 网络和 BP 神经网络输出进行融合,并使用 Softmax 求解器得到肢体识别结果。将肢体识别结果输入机器人辅助吊装控制系统,建立双重验证控制方法,完成机器人辅助吊装操作。实验结果表明,该方法保证了模型运行的精度和时效性,实时识别精度达 0.99 以上,大大提升了远距离人机交互能力。

关键词: 吊装机器人; BP 神经网络; 肢体识别; OpenPose; InceptionV3
中图分类号: TP242.2; TP391.4 **文献标识码:** A **文章编号:** 1000-1298(2019)06-0405-07

Research on Limb Motion Command Recognition Technology of Lifting Robot

NI Tao¹ ZOU Shaoyuan¹ LIU Haiqiang¹ HUANG Lingtao¹ CHEN Ning² ZHANG Hongyan¹
(1. College of Mechanical and Aerospace Engineering, Jilin University, Changchun 130022, China
2. School of Mechanical and Energy Engineering, Jimei University, Xiamen 361021, China)

Abstract: In view of the limited monitoring distance of Kinect for limb recognition, the large zoom network camera was used and CNN-BP fusion network for human behavior recognition was constructed, and the nine groups of robot lifting instructions were trained and identified. Firstly, totally 18 skeleton nodes were extracted based on OpenPose to generate RGB skeleton map and skeleton vector. Then, using the migration learning method, the InceptionV3 network was used to extract the deep abstract features of the image, and the training data set was rotated, translated, scaled and affine. A variety of data enhancement methods were used to extend the training data to prevent overfitting; and then the extracted skeleton vector was extracted from the shallow layer features such as the point line surface using BP neural network; the InceptionV3 network and the BP neural network output were merged and obtained by using the Softmax solver to obtain limb classification results. Finally, the result of limb recognition was input into the robot auxiliary hoisting control system, and the double verification control mode was established to complete the robot auxiliary hoisting operation. The test results showed that the method ensured the timeliness of the model operation, and the real-time recognition accuracy reached 0.99, which greatly improved the long-distance human-computer interaction capability.

Key words: lifting robot; back propagation neural network; limb recognition; OpenPose; InceptionV3

0 引言

随着人体行为识别技术和机器人技术的快速发展,人体行为识别在虚拟现实^[1]、视频监控^[2-8]和人

机交互^[9-11]等领域被广泛应用。在港口、码头、矿山等大型货物的远距离吊运场合,驾驶员与现场指挥员之间距离较远,难以准确把握现场操作员的指令信号。

收稿日期: 2018-11-16 修回日期: 2018-12-16
基金项目: 国家自然科学基金项目(51575219)和福建省海洋经济创新发展区域示范项目(2014FJPT03)
作者简介: 倪涛(1978—),男,教授,博士生导师,主要从事机器人技术和机器视觉研究,E-mail: nitao@jlu.edu.cn
通信作者: 陈宁(1972—),男,副教授,主要从事机器视觉及运动控制技术研究,E-mail: cn1972@jmu.edu.cn

近年来,肢体识别技术主要使用卷积神经网络(Convolutional neural network, CNN)和循环神经网络(Recurrent neural network, RNN)两种网络进行训练和识别。文献[12]构建多级 CNN 结构,提取图像多种时空信息特征,该方法使用图像序列以及对应的光流图像序列作为卷积神经网络输入进行训练,但网络非常复杂,训练难度大且容易过拟合。文献[13-16]基于 Kinect 相机提取的骨架,使用精细构造的多层 RNN 网络进行肢体行为识别,但是该方法网络输入只局限于骨架节点的坐标,提取的特征信息有限。文献[17-18]构建多级 CNN-RNN 网络,该方法综合了 CNN 训练简单和 RNN 能够获取更多的上下文信息的优势,但增大了网络复杂度,难以达到实时识别的应用需求。

为了克服上述方法存在的问题,本文提出一种 CNN-BP 融合网络,采用 InceptionV3 和 BP 网络融合方法对机器人吊装指挥姿势进行识别。首先基于

OpenPose(卡内基梅隆大学计算机科学学院 CMU AI 计划)提取图像骨架节点坐标,并由此生成 RGB 骨架图作为 InceptionV3 网络的输入层,同时将骨架节点坐标生成骨架向量作为 BP 神经网络输入层进行训练,再将两个网络的输出层进行融合,并使用 Softmax 求解器得到识别结果,最后使用双重验证控制方法实现机器人辅助吊装工作。

1 系统组成

如图 1 所示,系统包含人体肢体识别模块和机器人辅助吊装控制模块两部分。鉴于 Kinect 相机进行肢体识别监控距离有限,不适合吊运机器人的远距离操控和大作业空间的应用场合,本文采用网络大变焦摄像头作为图像采集设备,对采集图像基于 OpenPose 提取人体骨架节点坐标信息。该网络大变焦摄像头监控距离最大为 80 m,满足远距离人机交互作业。

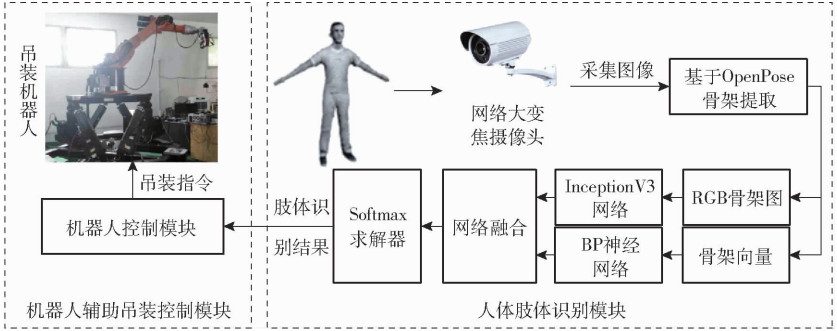


图 1 系统总体方案

Fig. 1 System overall plan

人体肢体识别模块中,构建 CNN-BP 融合网络进行训练和识别。依据 OpenPose 提取的骨架节点的坐标生成 RGB 骨架图作为 InceptionV3 网络输入,采用迁移学习方法提取 RGB 骨架图深层抽象特征。同时将 18 个骨架坐标生成一个骨架向量输入 BP 神经网络提取其骨架坐标间点、线以及面等浅层特征。最后将两个网络输出层进行融合并用 Softmax 求解器得到识别结果。本文选取 9 组机器人吊装指令(参考机器人吊装指挥信号 GB 5082—1985),使用 CNN-BP 网络对采集训练数据进行训练和识别,将识别结果输入机器人辅助吊装控制模块,采用双重验证控制方法完成机器人辅助吊装操作。

2 数据处理

图像采集设备为由安迅士网络摄像机 (AXIS-Q1635 型) 和海康威视大变焦镜头 (HV1140D-8MPIR 型) 组成的网络大变焦摄像头,如图 2 所示。为保证在摄像机振动情况下成像稳定,AXIS-

Q1635 型摄像机配备一个 1/2 英寸传感器。HV1140D-8MPIR 型镜头是一款自动光圈手动变焦 800 万像素大变焦镜头,提供焦距变化范围为 11~40 mm,最大监控距离为 80 m。

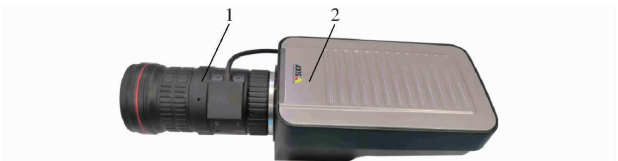


图 2 网络大变焦摄像头示意图

Fig. 2 Large-zoom webcam schematic

1. HV1140D-8MPIR 型镜头 2. AXIS-Q1635 型摄像机

数据预处理的关键是获得 RGB 骨架图和骨架向量,RGB 骨架图和骨架向量获取操作流程如图 3 所示,其操作步骤为:①首先通过网络大变焦摄像头采集一帧图像,经 OpenPose 提取人体 18 个骨架节点坐标。②将 18 个骨架坐标展成一行,生成骨架向量。③对骨架节点间使用不同颜色的椭圆进行连接,并使用不同颜色进行着色。为强化肢体特征信息,剔除背景信息获得 RGB 骨架图。

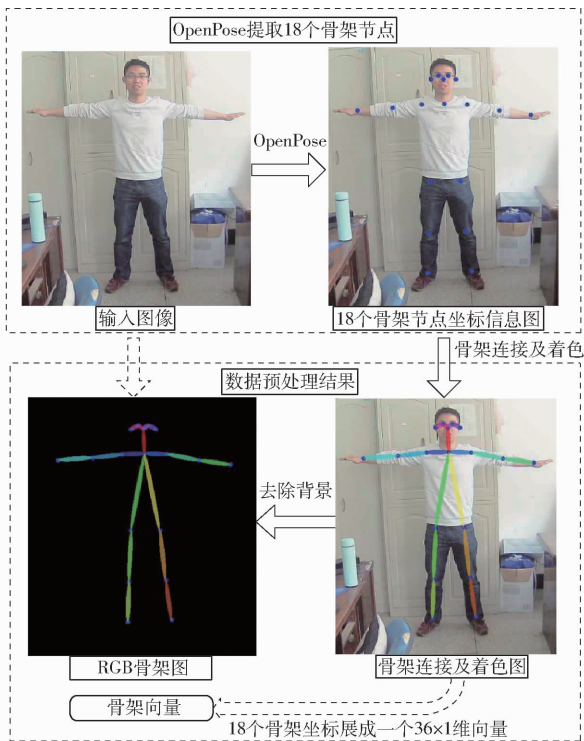


图 3 RGB 骨架图和骨架向量获取示意图
Fig.3 Schematic of RGB skeleton diagram and skeleton vector acquisition

为验证 CNN - BP 分类性能,参考 GB 5082—1985《起重吊运指挥信号》选取的 9 组待分类指挥信号分别为:预备、要主钩、要副钩、吊钩上升、吊钩下降、水平向左移动、水平向右移动、吊钩停止、紧急结束。考虑到现实环境下人的姿态多变,为了提升 CNN - BP 网络模型的性能和泛化能力,对提取的 RGB 骨架图采用 $-15^{\circ} \sim 15^{\circ}$ 随机旋转、水平垂直两个方向 $-20 \sim 20$ 像素点随机平移、 $0.8 \sim 1.2$ 倍随机缩放和随机初始化仿射变换矩阵进行仿射变换多种数据增强方式,数据增强效果如图 4 所示。分别对 9 组机器人吊装指令各采集 1 000 幅图像,共计 9 000 幅初始图像,每幅图像每种数据增强方式随机产生 1 幅图像共生成 45 000 幅训练数据。使用数据增强方式大大扩充了训练数据的数量,可以有效防止模型过拟合。

3 肢体动作识别分类器构建

3.1 CNN - BP 网络架构

卷积神经网络是一种深度前馈神经网络,可以很好地提取图像的局部和更深层次的抽象特征。BP 神经网络是一种前向传播,误差逆向传播训练的多层感知神经网络,能够很好地提取数据浅层特征。参考文献[19 - 20]构建 CNN - BP 融合网络架构如图 5 所示,分类过程主要包括网络输入预处理模块和 CNN - BP 网络模块两部分。其中网络输入预处

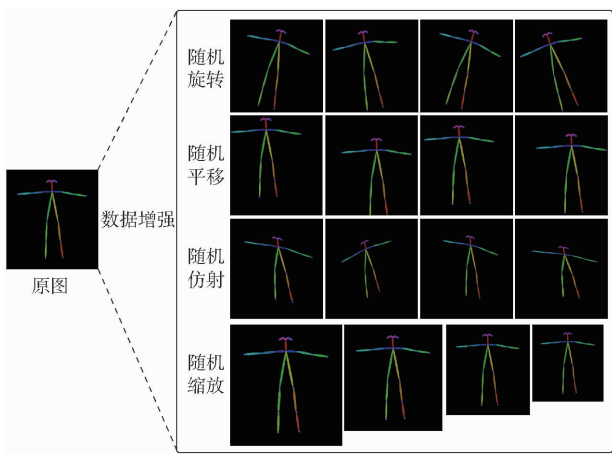


图 4 4 种数据增强方式示意图
Fig.4 Sketch of four data enhancement methods

理模块用于提取 RGB 骨架图和骨架向量;CNN - BP 网络模块包含 CNN 网络迁移学习模块、BP 神经网络模块和网络融合模块。

3.2 基于 InceptionV3 网络特征提取

经典 CNN 网络在 ImageNet 数据集效果如表 1 所示。由表 1 可知,InceptionV3 和 IncetionResNetV2 在 ImageNet 数据集效果较优,但考虑模型复杂度,选择 InceptionV3 网络提取 RGB 骨架图特征。由于训练数据有限,文献[21 - 23]对深度 CNN 网络 ImageNet 数据集上使用迁移学习方法进行分类任务,得到了较优的性能。迁移学习是利用源域解决任务获得一些知识来提升目标分类性能的一种算法,文献[21 - 23]先将 InceptionV3 网络在大型图像数据集 ImageNet 上进行预训练,训练后保留卷积层训练参数用来初始化应用于本文 RGB 骨架图分类任务 InceptionV3 网络卷积层参数,具体实现过程如图 6 所示。

采用迁移学习方法,使用随机梯度下降算法 (Stochastic gradient descent optimizer,SGD),设置批大小为 32、低学习率 0.001、动量为 0.9 进行训练,网络终止迭代轮数为 1 000。对底层和顶层卷积层激活后的特征图可视化如图 7 所示(限于篇幅仅可视化每个特征图的前 32 个通道)。由图 7 可知,靠近输入层的卷积层(conv2d_1、conv2d_3、conv2d_5)激活保留图像的大量边缘、颜色和形状等信息,随着层数的加深靠近输出层的卷积层(mixed8、mixed9 和 mixed10)激活后变得越来越抽象并且难以直观理解,其表示图像更高层次的特征信息。随着深度网络的加深,可以将无关信息过滤并放大和细化用于分类的特征信息。其中 InceptionV3 卷积层命名参考 Keras 深度学习框架。

3.3 BP 神经网络特征提取

OpenPose 提取的 18 个骨架节点坐标记为 $(P_1,$

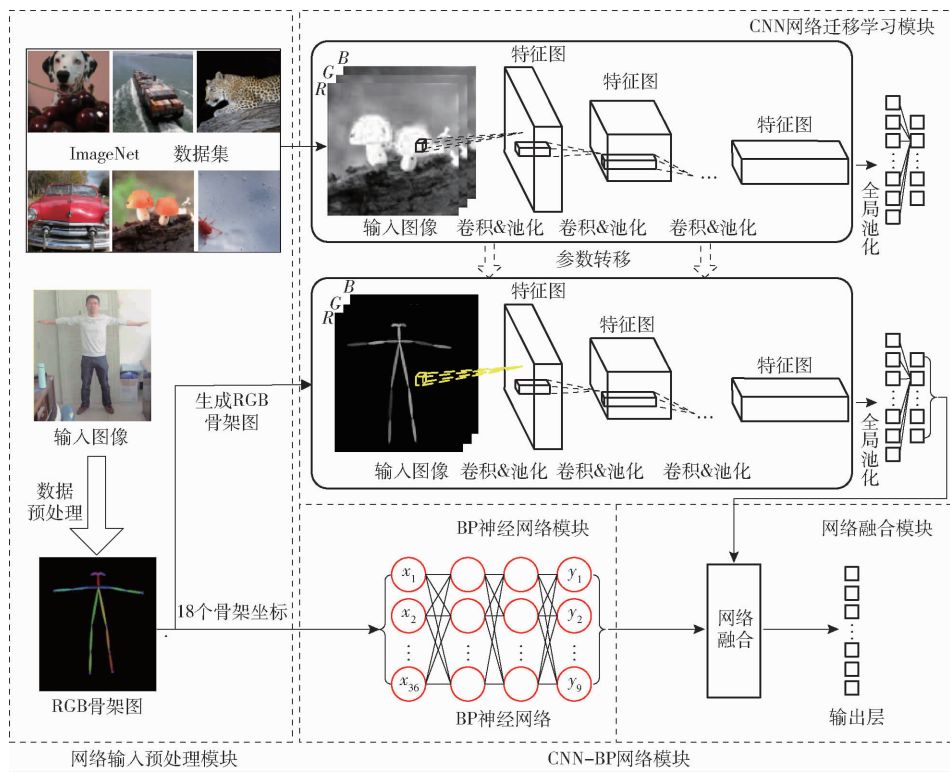


图 5 CNN - BP 网络架构

Fig. 5 CNN - BP network architecture

表 1 经典 CNN 网络在 ImageNet 上效果

Tab.1 Effect of classic CNN network on ImageNet

模型	Top1 准确率	Top5 准确率	参数数目	深度
Xception	0. 790	0. 945	22 910 480	126
VGG16	0. 715	0. 901	138 357 544	23
VGG19	0. 727	0. 910	143 667 240	26
ResNet50	0. 759	0. 929	25 636 712	168
InceptionV3	0. 788	0. 944	23 851 784	159
IncetionResNetV2	0. 804	0. 953	55 873 736	572
MobileNet	0. 665	0. 871	4 253 864	88

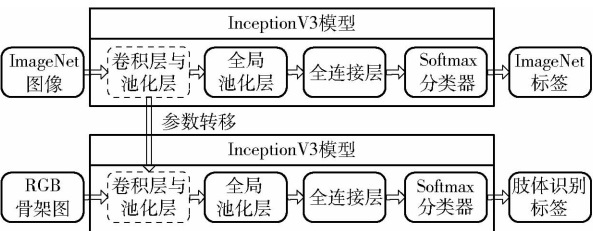


图 6 迁移学习示意图

Fig. 6 Migration learning schematic

$P_2, \dots, P_{18}$), 将其展成一个骨架向量 ($P_{x1}, P_{y1}, P_{x2}, P_{y2}, \dots, P_{x18}, P_{y18}$) 作为 BP 神经网络输入层。点与点特征之间构成线特征, 线与线特征之间构成面特征, 本文 BP 神经网络设置 2 个隐含层, 网络结构示意图如图 8 所示。网络训练参数为: 学习率为 0.005, 第 1 隐含层节点个数为 100, 第 2 隐含层节点个数为 50, dropout 概率为 0.5, 正则化系数 1.1, 终

止迭代次数为 1 000。

3.4 CNN - BP 网络融合

对 InceptionV3 网络输出 $y_i^a (y_1^a, y_2^a, \dots, y_n^a)$ 与 BP 神经网络输出 $y_i^b (y_1^b, y_2^b, \dots, y_n^b)$ 采用 3 种方式进行融合, 计算式为

$$\begin{cases} y_i^{\text{sum}} = (y_1^a + y_1^b, y_2^a + y_2^b, \dots, y_n^a + y_n^b) \\ y_i^{\text{max}} = (\max(y_1^a, y_1^b), \max(y_2^a, y_2^b), \dots, \max(y_n^a, y_n^b)) \\ y_i^{\text{cat}} = (y_1^a \times y_1^b, y_2^a \times y_2^b, \dots, y_n^a \times y_n^b) \end{cases} \quad (1)$$

式中 y_i^{sum} ——求和方式进行融合的结果
 y_i^{max} ——Max 融合方式进行融合的结果
 y_i^{cat} ——乘积方式进行融合结果
 n ——分类类别数目

将获得的输出结果 $y_i^{\text{sum}}, y_i^{\text{max}}$ 和 y_i^{cat} 融合方法分别为 Sum 融合、Max 融合和 Cat 融合。对融合结果使用 Softmax 分类器进行分类, 构建损失函数 $L(\theta)$ 为

$$L(\theta) = - \left[\sum_{i=1}^N \sum_{k=1}^K l\{y^{(i)} = k\} \lg \frac{\exp(\theta^{(k)\text{T}} x^{(i)})}{\sum_{j=1}^K \exp(\theta^{(j)\text{T}} x^{(i)})} \right] \quad (2)$$

式中 N ——训练样本数
 K ——样本类别数
 $y^{(i)}$ ——第 i 个样本预测类别
 θ ——权重

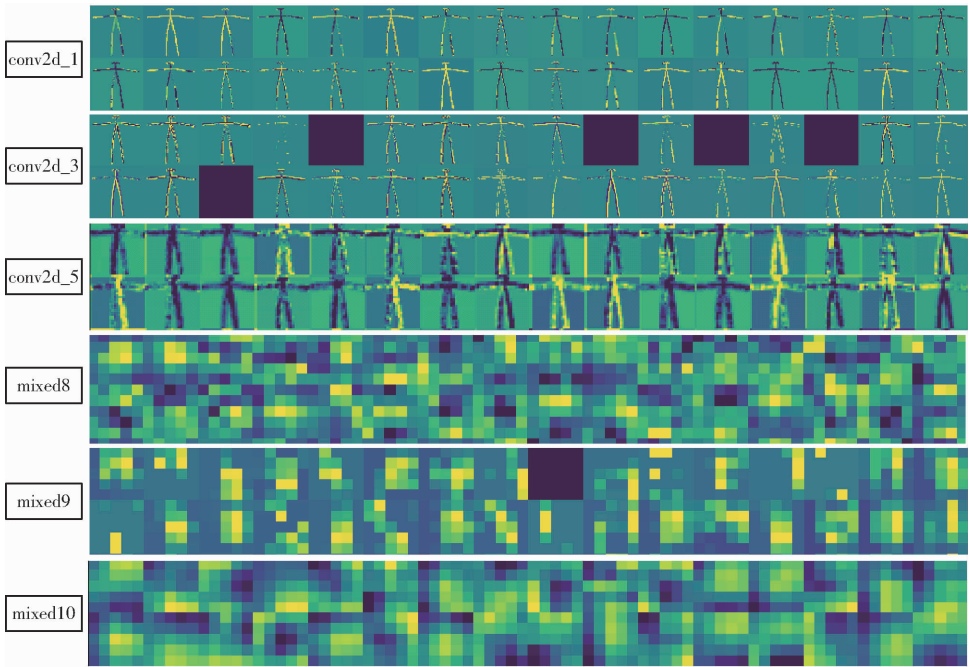


图 7 卷积层激活后的特征图可视化图

Fig. 7 Feature map visualization after convolutional layer activation

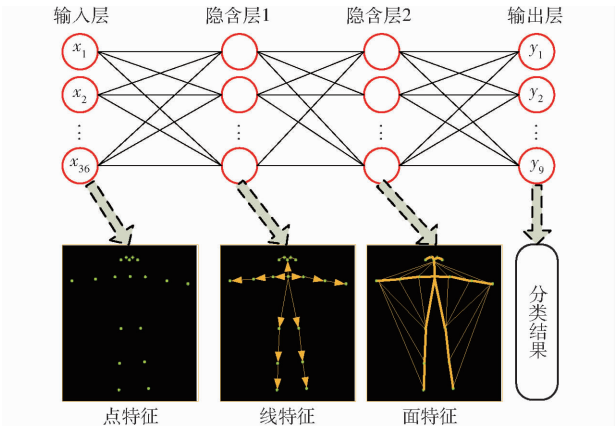


图 8 BP 神经网络结构示意图

Fig. 8 Schematic of BP neural network structure

- $x^{(i)}$ ——第 i 个样本的特征
- $\exp(\boldsymbol{\theta}^{(k)\text{T}}\boldsymbol{x}^{(i)})$ ——样本 i 被分类成类别 k 的概率
- $l\{y^{(i)} = k\}$ ——正样本

4 实验及结果分析

4.1 实验环境

实验控制设备为计算机,配置为 Intel i7 处理器,内存 8 GB,显卡 GTX980,Win10 环境使用 PyQt 编写软件控制界面。鉴于 OpenPose 与 InceptionV3 复杂度过高,为了确保实时性采用双 GTX980 显卡(显存 8 GB),模型在 GPU 进行运算。为了降低计算量,设置 OpenPose 的 scale_search 参数为“0.5, 1”,并将输入图像压缩至 320 像素 × 240 像素,得到单帧图像识别时间为 386 ms,为了实时识别采用多

线程编程将视频预览与算法运算分离。其中吊装机器人实物如图 9 所示。



图 9 吊装机器人实物图

Fig. 9 Physical diagram of lifting robot

4.2 网络训练结果

为验证提出的 CNN-BP 网络模型的准确性,分别使用 InceptionV3 网络、BP 神经网络和 CNN-BP 网络,对吊装机器人 9 种吊装指挥姿势进行训练,分类精度如表 2 所示。从表 2 可以看出,CNN-BP 采用 Sum 融合方式识别精度最高,其分类结果的混淆矩阵和 ROC 曲线如图 10、11 所示,可以看出该方法对 9 组机器人吊装姿势都能够准确分类,平均识别精度达到 0.953。同时 9 组动作识别的 ROC 曲线

表 2 不同网络分类结果

Tab. 2 Different network classification results			
模型	BP 神经网络		InceptionV3 网络
精度	0.906		0.926
模型	CNN – BP 融合网络		
	Sum 融合	Max 融合	Cat 融合
	精度	0.953	0.915

AUC 值不低于 0.92, 平均 AUC 达到 0.96。其中 RUC 为 ROC 曲线和横轴围成的面积, 取值范围为 0~1, 越接近 1 表示分类器分类性能越好。

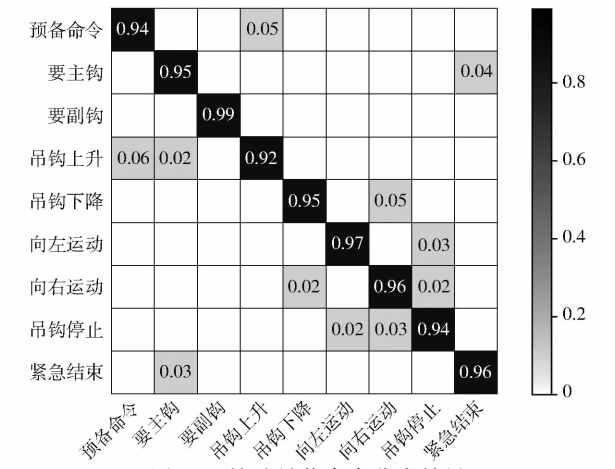


图 10 辅助吊装命令分类结果

Fig. 10 Auxiliary lifting command classification result

4.3 机器人辅助吊装系统控制方法

基于双重验证控制方法的机器人辅助吊装控制过程和 9 组控制信号如图 12 所示, 机器人辅助

控制系统初始状态处于待命状态, 在 CNN-BP 网络识别出“预备命令”控制系统才会进入工作状态。而肢体识别过程采取间隔 50 ms 采集两帧图像先后输入 CNN-BP 网络分类模块, 对两次识别结果进行判断, 如果相等且识别结果为“预备指令”则开启机器人吊装控制, 在开启机器人辅助吊装系统后将识别的控制信号发送给机器人吊装控制器完成吊装操作, 如果两次识别结果不相等则丢弃本次识别结果, 该方法实时识别精度达 0.99 以上。

5 结束语

鉴于 Kinect 进行肢体识别监控距离有限, 搭建了使用网络大变焦摄像头并基于 OpenPose 的 CNN-BP 融合网络, 使用迁移学习方法在 ImageNet 数据集上对 InceptionV3 网络预训练。采用旋转、平移、缩放和仿射方法对采集的训练数据集进行数据增强, 有效防止了过拟合。采用 3 种方式对 InceptionV3 网络和 BP 神经网络两种网络输出进行

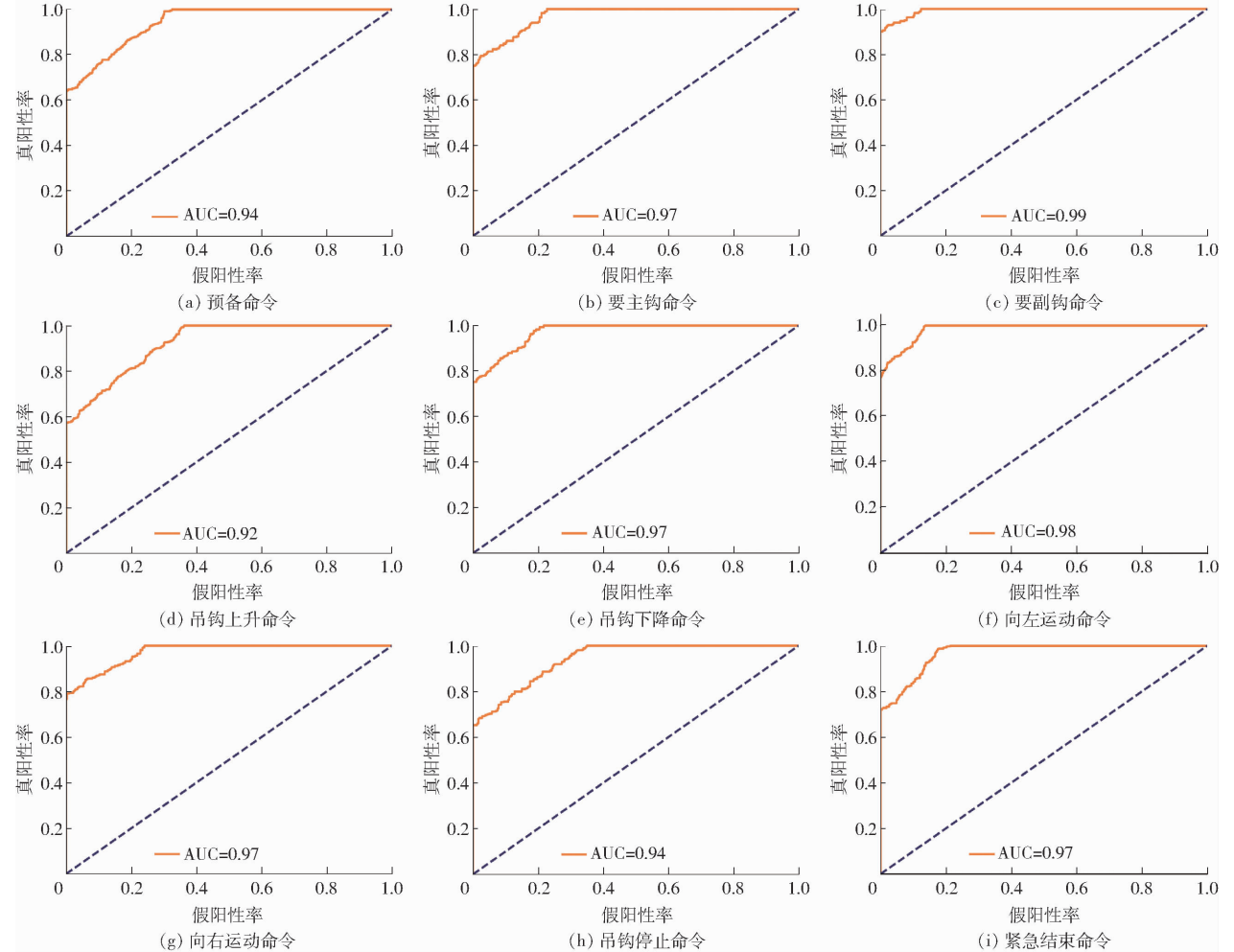


图 11 9 组辅助吊装命令识别 ROC 曲线

Fig. 11 Nine groups of auxiliary lifting commands identified ROC curves

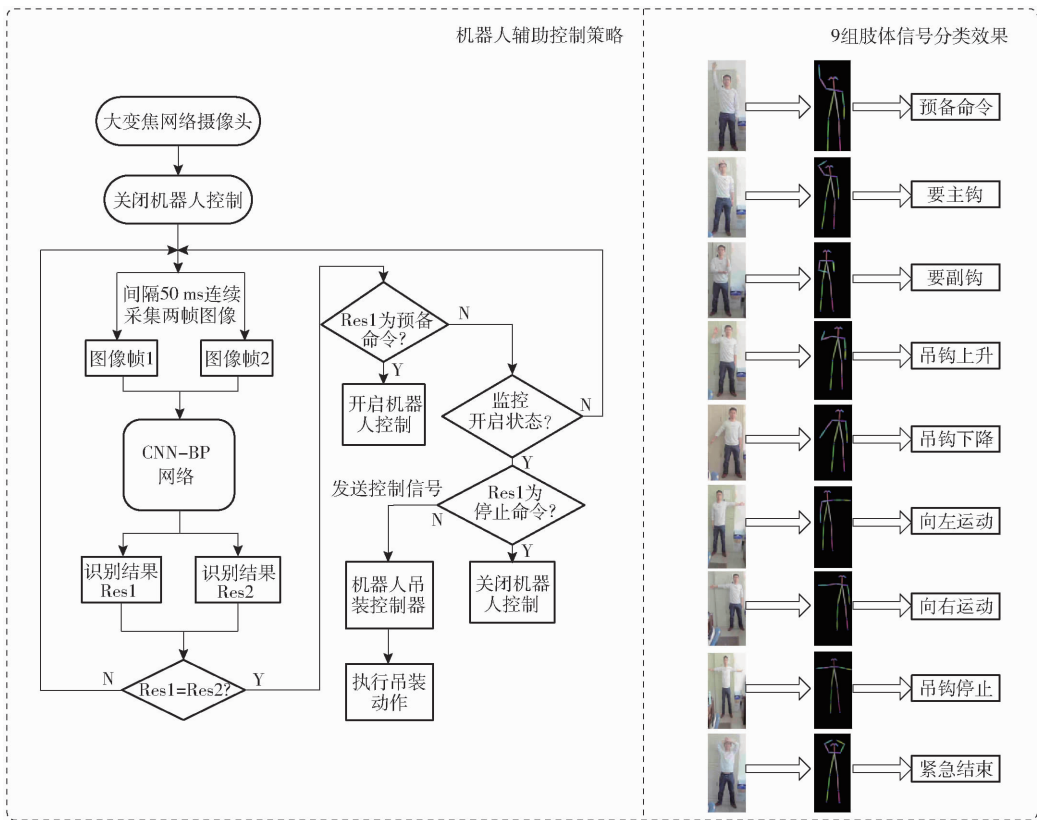


图 12 机器人辅助吊装系统
Fig. 12 Robotic assisted lifting system

融合,结果表明,Sum 融合方式可以得到更高的模型分类精度。对机器人辅助吊装系统建立了双重验证控制方法。实验结果表明,采用 CNN - BP 融合网络

和双重验证方法保证了模型运行的精度和时效性,识别精度达 0.99 以上,大大提升了远距离人机交互能力。

参 考 文 献

[1] PATRICK R, WOHLFROMM L, TRACHT K. Implementation of virtual reality systems for simulation of human-robot collaboration[J]. Procedia Manufacturing, 2018, 19:164 - 170.

[2] GHABRI S, OUARDA W, ALIM I A M. Towards human behavior recognition based on spatio temporal features and support vector machines[C]// International Conference on Machine Vision, 2017.

[3] 黄凯. 人体行为识别研究及其在水电站视频监控中的应用[D]. 杭州: 浙江工业大学, 2013.

HUANG Kai. Research on human behavior recognition and application in video monitoring of hydropower station [D]. Hangzhou: Zhejiang University of Technology, 2013. (in Chinese)

[4] 满君丰, 李倩倩, 温向兵. 视频监控中可变更人体行为的识别[J]. 东南大学学报(自然科学版), 2011, 41(3):492 - 497.

MAN Junfeng, LI Qianqian, WEN Xiangbing. Recognition for changeable human behaviors in video surveillance[J]. Journal of Southeast University(Natural Science Edition), 2011, 41(3):492 - 497. (in Chinese)

[5] 何杰. 视频监控中的人体异常行为识别方法研究[D]. 重庆: 重庆大学, 2014.

HE Jie. Research on human abnormal activity recognition in video surveillance[D]. Chongqing: Chongqing University, 2014. (in Chinese)

[6] GOWSIKHAA D, ABIRAMI S, BASKARAN R. Automated human behavior analysis from surveillance videos: a survey[J]. Artificial Intelligence Review, 2014, 42(4):747 - 765.

[7] BAXTER R H, ROBERTSON N M, LANE D M. Human behaviour recognition in data-scarce domains [J]. Pattern Recognition, 2015, 48(8):2377 - 2393.

[8] LAO W W, HAN J J, WITH D P P. Automatic video-based human motion analyzer for consumer surveillance system[J]. IEEE Transactions on Consumer Electronics, 2009, 55(2):591 - 598.

[9] 倪涛,赵泳嘉,张红彦,等. 基于 Kinect 动态手势识别的机械臂实时位姿控制系统[J/OL]. 农业机械学报, 2017, 48(10):417 - 423.

NI Tao, ZHAO Yongjia, ZHANG Hongyan, et al. Real-time mechanical arm position and pose control system by dynamic hand gesture recognition based on Kinect device[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(10):417 - 423. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20171053&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2017.10.053. (in Chinese)

- WANG Manxin, SHEN Qiusheng, ZU Li, et al. Stiffness analysis of 3 - RRS parallel mechanism with consideration of gravity [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(11):392 - 402. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20181148&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2018.11.048. (in Chinese)
- [17] 胡波,路懿. 求解3-RPS并联刚度的新方法[J]. 机械工程学报,2010,46(1):24-29.
HU Bo, LU Yi. New approach for analyzing the stiffness of 3 - RPS parallel manipulator [J]. Journal of Mechanical Engineering, 2010, 46(1): 24 - 29. (in Chinese)
- [18] TAT J T, YANG G L, CHEN I M. A large deflection and high payload flexure-based parallel manipulator for UV nanoimprint lithography:Part I. Modeling and analyses[J]. Precision Engineering,2014,38(4):861 - 871.
- [19] LU Y, CHEN L W, WANG P. Statics and stiffness analysis of a novel six-component force/torque sensor with 3 - RPPS compliant parallel structure[J]. Mechanism and Machine Theory,2013,62:99 - 111.
- [20] 熊有伦. 机器人力传感器的各项同性[J]. 自动化学报,1996,22(1):10-16.
XIONG Youlun. On isotropy of robot's force sensors[J]. Acta Automatica Sinica, 1996, 22(1): 10 - 16. (in Chinese)
- [21] YAO J T, HOU Y L, WANG H, et al. Spatially isotropic configuration of Stewart platform-based force sensor[J]. Mechanism and Machine Theory, 2011, 46(2): 142 - 155.
- ~~~~~
- (上接第411页)
- [10] 权龙哲,李成林,冯正阳,等. 体感操控多臂栅室机器人作业决策规划算法研究[J/OL]. 农业机械学报, 2017,48(3): 14 - 23.
QUAN Longzhe, LI Chenglin, FENG Zhengyang, et al. Algorithm of works' decision for three arms robot in greenhouse based on control with motion sensing technology [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(3):14 - 23. http://www.jcsam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20170302&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2017.03.002. (in Chinese)
- [11] 唐新星,倪涛,何丽鹏,等. 基于立体视觉的遥操作工程机器人自主作业系统[J/OL]. 农业机械学报,2012,43(10): 224 - 228.
TANG Xinxing, NI Tao, HE Lipeng, et al. Autonomous task control system of construction tele-robot based on stereo vision [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2012,43(10):224 - 228. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20121040&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2012.10.040. (in Chinese)
- [12] 马森,李贻斌. 基于多级图像序列和卷积神经网络的人体行为识别[J]. 吉林大学学报(工学版), 2017, 47(4):1244 - 1252.
MA Miao, LI Yibin. Multi-level image sequences and convolutional neural networks based human action recognition method [J]. Journal of Jilin University(Engineering and Technology Edition), 2017, 47(4):1244 - 1252. (in Chinese)
- [13] DU Y, WANG W, WANG L. Hierarchical recurrent neural network for skeleton based action recognition[C]//CVPR, 2015.
- [14] ZHU W, LAN C, XING J, et al. Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks[C]//The 30th AAAI Conference on Artificial Intelligence (AAAI-16), 2016.
- [15] SHAHROUDY A, LIU J, NG T T, et al. NTU RGB + D: a large scale dataset for 3D human activity analysis[C]//CVPR, 2016.
- [16] LIU J, SHAHROUDY A, XU D, et al. Spatio-temporal LSTM with trust gates for 3D human action recognition[C]//European Conference on Computer Vision (ECCV), 2016.
- [17] WU Z, JIANG Y G, WANG X, et al. Multi-stream multi-class fusion of deep networks for video classification[C]//ACM on Multimedia Conference, ACM, 2016.
- [18] SHI Y, TIAN Y, WANG Y, et al. Sequential deep trajectory descriptor for action recognition with three-stream CNN[J]. IEEE Transactions on Multimedia, 2016, 19(7):1510 - 1520.
- [19] WANG J, YANG Y, MAO J, et al. CNN - RNN: a unified framework for multi-label image classification[C]//CVPR, 2016.
- [20] FAN Y, LU X, LI D, et al. Video-based emotion recognition using CNN - RNN and C3D hybrid networks[C]//International Conference on Multimodal Interaction, ACM, 2016.
- [21] LIANG G, HONG H, XIE W, et al. Combining convolutional neural network with recursive neural network for blood cell image classification[J]. IEEE Access, 2018, 99(6): 36188 - 36197.
- [22] VESAL S, RAVIKUMAR N, DAVARI A A, et al. Classification of breast cancer histology images using transfer learning[C]//International Conference Image Analysis & Recognition. Springer, Cham, 2018.
- [23] RAJAGOPAL A K, SUBRAMANIAN R, RICCI E, et al. Exploring transfer learning approaches for head pose classification from multi-view surveillance images[J]. International Journal of Computer Vision, 2014, 109(1 - 2):146 - 167.