

Deep Neural Network Combined with GHS for Static Gesture Recognition^{*}

ZHANG Dongping^{1*}, SHU Yuan¹, ZHOU Zhihong²

(1. Key Laboratory of Electromagnetic Wave Information Technology and Metrology of Zhejiang Province,

College of Information, China Jiliang University, Hangzhou Zhejiang 310018, China;

2. Shanghai Key Laboratory of Integrated Administration Technologies for Information Security, Shanghai 200240, China)

Abstract: In order to solve the problem of low accuracy of static gestures in the recognition process under unconstrained environment, this paper proposes a deep neural network combined with Grayscale Image of Hand Skeleton (GHS), using the key points of the hand and their correlations to build a grayscale image of the hand skeleton. The input of the network is GHS image and RGB image, the backbone network is yolov3, and the extended convolution residual module is added. After the feature fusion of the GHS image and the RGB image, the feature on each channel is scaled by the SE module, using RReLU Activation function to replace Leaky ReLU activation function. Through the key points of the hand and the connection information between them, the image features of the hand are enhanced, the difference between gestures is increased, and the influence of the unconstrained environment on gesture recognition is reduced to improve the accuracy of gesture recognition. The experimental results show that compared with other methods on the Microsoft Kinect & Leap Motion data set, the average accuracy rate of the method in this paper reaches the highest, 99.68%; on the Creative Senz3D data set, the average accuracy rate of the method in this paper reaches the highest, 99.8%.

Key words: deep learning; gesture recognition; GHS; unconstrained environment

EEACC: 6135; 6135E

doi: 10.3969/j.issn.1004-1699.2021.02.010

融合手部骨架灰度图的深度神经网络静态手势识别^{*}

章东平^{1*}, 束元¹, 周志洪²

(1. 中国计量大学信息工程学院, 浙江省电磁波信息技术与计量检测重点实验室, 浙江 杭州 310018;

2. 上海市信息安全综合管理技术研究重点实验室, 上海 200240)

摘要: 针对在无约束环境下静态手势在识别过程中准确率不高的问题, 本文提出了一种融合手部骨架灰度图 (Grayscale Image of Hand Skeleton, GHS) 的深度神经网络, 使用手部关键点及其相互关联性构建手部骨架灰度图。网络的输入为 GHS 图像和 RGB 图像, 主干网络为 yolov3, 添加了扩展卷积残差模块, 在 GHS 图像和 RGB 图像进行特征融合后, 通过 SE 模块对每个通道上的特征进行缩放, 采用 RReLU 激活函数来代替 Leaky ReLU 激活函数。通过手部关键点及其相互间的连接信息增强手部图像特征, 增大手势的类间差异, 同时降低无约束环境对手势识别的影响, 以提高手势识别的准确率。实验结果表明, 在 Microsoft Kinect & Leap Motion 数据集上相比其他方法, 本文方法的平均准确率达到最高, 为 99.68%; 在 Creative Senz3D 数据集上相比其他方法, 本文方法平均准确率达到最高, 为 99.8%。

关键词: 深度学习; 手势识别; 手部骨架灰度图; 无约束环境

中图分类号: TP391.4

文献标识码: A

文章编号: 1004-1699(2021)02-0203-08

近年来, 手势识别受到了很多学者的广泛关注^[1-5], 但基于机器视觉的手势识别, 尤其是在无约束环境下通过手部动作来控制智能机器^[6]方面, 仍

然是一项具有挑战性的任务。在无约束环境下进行手势识别时涉及的挑战主要包括: 手势的自由度值较大^[7]、自我遮挡、手部定位不准确和背景环境复

项目来源: 浙江省重点研发计划项目 (2020C03104); 上海市信息安全综合管理技术研究重点实验室开放课题 (AGK201900X)

收稿日期: 2020-06-01 修改日期: 2020-12-11

杂等方面。基于上述众多因素的影响,无约束环境下手势识别算法的综合性能普遍不高,尤其是识别准确率方面,具有较大的提升空间。

手势识别基于运动状态可分为静态手势识别和动态手势识别。静态手势,是指每一种静态形式的手部姿势都分别表示某种特定的意义,即一张图片代表一个手势。动态手势是在一定的时间段内,由一组静态手部姿势形成的一个更为复杂的手势来表示某种特定的意义,即一段视频代表一个手势。

基于传统方法的手势识别算法主要有:一种贝叶斯视觉注意模型^[8]在检测和识别静态手势时,将选择性注意模型引入,且模拟了人体大脑视觉皮层的部分功能,然后使用支持向量机(SVM)分类器对手势进行分类,该方法在 NUSHP-II 手势数据集上的准确率为 94.36%;一种有效结合机器学习模型与模板匹配的方法^[9],在作者的数据集上获得了 99.75%的基本手势识别率以及 100%的复杂手势识别率;一组深度混合分类器^[10],用于以自我为中心的手势识别,在 NUSHP-II 数据集的识别精度为 97.72%;一种朴素贝叶斯分类器系统^[11],在基于静态手势的协作机器人方面表现良好,其结果在不同的工作环境、距离、角度、不同大小的人等条件下产生了较好的鲁棒性;一种融合表面肌电和加速度传感信息的识别方法^[12],对四路表面肌电与三轴加速度信号进行数据采集,通过支持向量机完成分类,在作者的数据集上达到了 91.2%的识别率。

基于神经网络的动态手势识别算法主要有:文献[13]提出了一种 ResNet-10 轻型 CNN 架构用于手势检测,提出了一种 D-CNN ResNeXt-101 模型来对手势进行分类,在 EgoGesture 数据集和 NVIDIA Dynamic Hand Gesture 数据集进行了手势分类测试,准确度分别为 94.04% 和 83.82%;一种融合了 Masked Res-C3D 网络和骨架 LSTM 模型^[14],并提出了一种通过权重融合层融合异常结果和分类模型缺陷的方案,用于 RGB-D 视频中的异常手势识别,在 IsoGD 数据集上识别率为 68.42%。

基于神经网络的静态手势识别算法主要有:一种使用 CNN 的手势识别^[15],其背景减少技术用于去除建议手部区域的敏感背景,并获得了 95.96%的平均识别率;一种基于 R-CNN 的手势识别^[16]产生了识别精度为 99.4%,但平均一张图像的识别计算时间为 40 s,这在计算上是较为昂贵的;一种使用 Faster R-CNN 对象检测器为协作机器人建立的智能手势识别^[17],在作者自己的数据集上测试的平均准确率为 91.94%,且平均一张图像的识别计算时间为

230 ms;一种通过迁移学习利用两个天线的阻抗变化和深度卷积神经网络对频谱图图像对手势进行分类^[18],该方法在作者自己的数据集上准确度达到 94.6%;一种带有 CPF 的 Faster R-CNN 对象检测器来找到 RGB 图像中手的准确位置^[19],在 Gloves 数据集上进行了测试,平均精度为 95.51%。

为解决在无约束环境下静态手势识别准确率不高的问题,本文提出了一种融合手部骨架灰度图的深度神经网络。首先,构建了一种基于手部关键点及其相关性的手部骨架灰度图作为网络输入。当训练网络时仅输入 RGB 图像,因图像中的手容易受到无约束环境下灯光、自身手指间的遮挡以及背景颜色,尤其是自身肤色等方面的影响,导致网络仅依靠通过 RGB 图像获得的纹理、颜色和轮廓等特征信息无法实现更精准的手部检测定位,从而进一步的导致静态手势识别在无约束环境下手部定位不准确且准确率不高的情况;而在输入 RGB 图像的基础上同时输入手部骨骼灰度图,并将两者以增加通道的形式进行特征融合后,通过手部关键点及其相互间的连接信息增强手部图像特征,增大手势的类间差异,同时降低无约束环境对手势识别的影响,以提高手部识别的准确率。其次,在网络结构方面,以 yolov3^[20]神经网络作为主干网络,将残差模块中的卷积替换为扩展卷积,通过增大卷积视野的方式来提高对手势占画面比较大情况下的准确率。然后,在手部骨架灰度图和 RGB 图像特征融合后,通过 SE 模块学习通道之间的相关性,自动学习每个特征通道的重要程度,对每个通道上的特征进行缩放,以增强两张图像之间的特征融合。最后,采用 RReLU 激活函数来代替 Leaky ReLU 激活函数,以提升网络的鲁棒性,从而降低无约束环境对静态手势识别的影响。

1 网络结构

一种融合手部骨架灰度图的深度神经网络如图 1 所示:以 yolov3 神经网络为主干网络,首先通过 RGB 图像经 openpose 提取的手部数据来构建手部骨架灰度图,并连同 RGB 图像一起输入到网络中;其次在多尺度模块前以增加通道的形式将 RGB 图像的特征与手部骨架灰度图进行特征融合,并一个通过 SE 模块学习通道之间的相关性,对每个通道上的特征进行缩放;然后将残差模块中的 1×1 的卷积改成膨胀系数为 2 的扩展卷积, 3×3 的卷积改成膨胀系数为 3 的扩展卷积;最后将激活函数由 Leaky ReLU 函数改为 RReLU 函数。

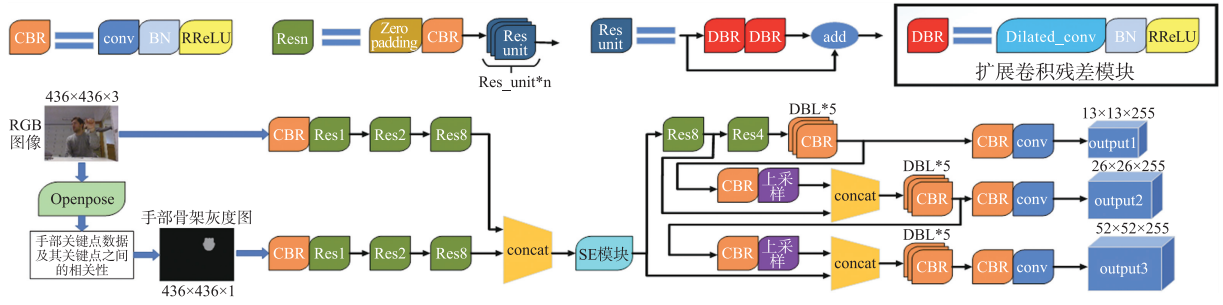


图 1 融合手部骨架灰度图的深度神经网络

2 手部骨架灰度图构建

2.1 基于 openpose 获取手部骨架信息

卡内基梅隆大学在 2017 年 CVPR 会议上提出一种提取手部关键点的开源代码^[21], 简称 openpose。手部关键点名称如表 1 所示, 手部骨架示意图如图 2 所示, 表 1 中的序号分别对应图 2 中的关键点序号。该算法的原理主要是利用立体几何, 以多视图作为监督信号源, 生成一致的手部关键点标签, 引导训练手部关键点检测器, 输出手部 21 个关键点归一化后的二维坐标值及其对应的得分。

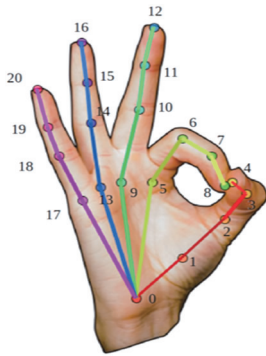


图 2 手部骨架图

表 1 openpose 手部关键点名称

序号	关键点名称	序号	关键点名称
0	手掌	11	中指上节点
1	大拇指根节点	12	中指指尖
2	大拇指下节点	13	无名指根节点
3	大拇指上节点	14	无名指下节点
4	大拇指指尖	15	无名指上节点
5	食指根节点	16	无名指指尖
6	食指下节点	17	小拇指根节点
7	食指上节点	18	小拇指下节点
8	食指指尖	19	小拇指上节点
9	中指根节点	20	小拇指指尖
10	中指下节点		

2.2 手部骨架灰度图构建

提出的一种基于手部关键点信息和关键点之间

相关性的手部骨架灰度图, 如图 3 所示, 是在 openpose 输出的基础上, 将归一化后 21 个手部关键点位置坐标值 (x, y) 及其对应关键点得分的输出文件进行如下计算和处理所得。

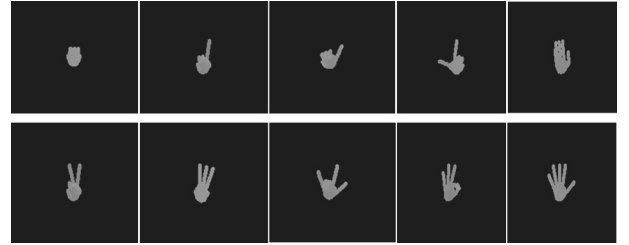


图 3 部分手势的 GHS 图像

首先, 按照原图的像素级宽高进行还原, 以便在原图像上定位。然后以关键点为圆心, r 个像素点为半径画圆, 所得的圆内所有像素点的灰度值, 均为关键点得分乘以 255。最后, 关键点之间的关联性表现在以上述圆的直径为宽, 关键点之间欧式距离为长, 构成的长方形内所有像素点的灰度值为两个关键点的平均得分乘以 255。手部骨架灰度图的其他像素点的灰度值均置为 0。上述计算如式(1)~式(3)所示:

$$\begin{cases} gv_k = 255 \times score_i \\ \text{s.t. } \|I - K\|_2 \leq r \end{cases} \quad (1)$$

$$\begin{cases} gv_k = 255 \times score_j \\ \text{s.t. } \|J - K\|_2 \leq r \end{cases} \quad (2)$$

$$\begin{cases} gv_k = \frac{255 \times (score_i + score_j)}{2} \\ \text{s.t. } \begin{cases} d(K, l_i) + d(K, l_j) = \|I - J\|_2 \\ d(K, l_{ij}) \leq r \end{cases} \end{cases} \quad (3)$$

式中: gv 表示灰度值, k 表示第 k 个像素点, $score$ 表示关键点得分, i 和 j 分别表示第 i 和 j 个关键点, I 、 J 和 K 分别表示对应关键点在原图像上的坐标值 (x, y) , $I - K_2$ 表示求两个坐标点之间的欧式距离; $d(K, l_i)$ 表示求点 K 到直线 l_i 之间的距离, l_{ij} 表示过点 I 和 J 的直线, l_i 表示过点 I 且与直线 l_{ij} 垂直的直线, l_j 表示过点 J 且与直线 l_{ij} 垂直的直线。

2 SE 模块

SE 模块^[22]是 2018 年由来自 Momenta 的高级研发工程师胡杰及其所在 WMW 团队在 CVPR 会议上提出的,模块的网络结构如图 4 所示,其中 e 是一个缩放参数,目的是为了减少通道数从而降低计算量。该模块的组成有 3 个部分,分别是 Squeeze 操作、Excitation 操作和 Scale 操作。

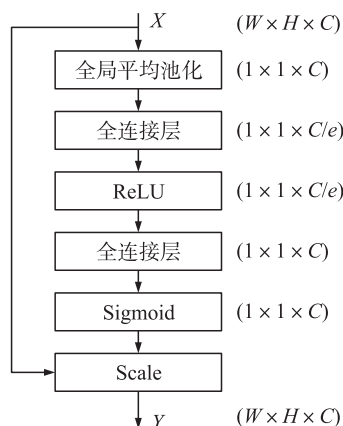


图 4 SE 模块网络结构

Squeeze 操作如式(4)所示,利用全局的池化,将大小为 $C \times H \times W$ 的输入特征综合为 $C \times 1 \times 1$ 的特征描述,其中 C 为通道数, H 和 W 分别为图像的高宽。

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (4)$$

Excitation 操作如式(5)所示,先利用一个全连接层操作将 W_1 和 z_c 相乘,其中 W_1 的维度是 $(C/r) \times C$,再经过一个 ReLU 层,输出的维度不变;然后再将结果和 W_2 相乘,且和 W_2 相乘也是一个全连接层的过程, W_2 的维度是 $C \times (C/r)$,因此输出的维度就是 $1 \times 1 \times C$;最后再经过 sigmoid 函数,得到 s 。

$$s_c = F_{ex}(z_c, W) = \sigma[W_2 \delta(W_1 z_c)] \quad (5)$$

式中: δ 和 σ 分别表示第一次全连接操作和第二次全连接操作。

Scale 操作如式(6)所示,即把 u_c 矩阵中的每个值都乘以 s_c ,其中 u_c 是一个二维矩阵特征图, s_c 是权重系数, Y_c 表示结合了权重之后 SE 模块的最终输出。

$$Y_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \quad (6)$$

3 改进 yolov3 原激活函数

3.1 Leaky ReLU 激活函数

Yolov3 中使用的激活函数是 Leaky ReLU 激活函数,作用在于对网络的上一层节点输出做非线性

转换,同时解决神经网络的梯度消失问题。相较于 ReLU 激活函数,Leaky ReLU 是给所有负值赋予一个非零斜率,其计算如式(7)所示:

$$y_i = \begin{cases} x_i & \text{if } x_i \geq 0 \\ a_i x_i & \text{if } x_i < 0 \end{cases} \quad (7)$$

式中: x_i 表示函数输入, y_i 表示函数输出, a_i 表示一个在 $(0, 1)$ 区间内的一个固定参数。由于在训练过程中 a_i 是固定不变的,从而限制了网络在无约束环境下的鲁棒性。

3.2 RReLU 激活函数

RReLU 激活函数,是在 Leaky ReLU 激活函数的基础上,将负值的斜率在训练中变为随机的值,即 a_i 是从一个均匀的分布中随机抽取的数值,但在测试环节就会固定下来。RReLU 激活函数如式(8)所示:

$$y_i = \begin{cases} x_i & \text{if } x_i \geq 0 \\ a_i x_i & \text{if } x_i < 0 \end{cases} \quad (8)$$

式中: $a_i \in U(l, u)$, $l < u$ 且 $u \in [0, 1)$ 。将 Yolov3 中的激活函数由 Leaky ReLU 函数改为 RReLU 函数后,相比之下,对测试数据集的鲁棒性得到了提升,从而降低了无约束环境对静态手势识别的影响。

3.3 Yolov3 为主干网络

Yolov3 网络引入了 FPN 结构,同时它的检测层由三级特征层的融合,直接将浅层网络所得的特征与深度卷积后的特征进行融合,这样的结构解决了其他部分网络(例如 SSD 模型)最后一层自动筛掉小物体的问题,提高了小物体的识别能力,从而提高了在手部占画面比较小情况下的手势识别精度。

且 Yolov3 网络模型的识别速度达到了较高的实时水平,而本文因手部骨架灰度图的提取需要一定时间,所以以 Yolov3 为主干网络后可一定程度上保证了本文提出的手势识别网络模型的识别速度。

4 实验

4.1 实验数据集介绍

4.1.1 Microsoft Kinect & Leap Motion 数据集

Microsoft Kinect & Leap Motion 数据集^[23]包含 RGB 数据、深度数据和 Leap Motion 数据。该数据集包含由 14 个不同的人分别做 10 种不同的手势,并且每个人对每个手势重复做 10 次,总共有 1 400 个不同的数据样本。对应 10 个手势的 RGB 图像图 5 所示,例如拳头、“Yes”、数字以及小拇指等。



图5 Microsoft Kinect & Leap Motion 数据集
RGB 手势图像示意图

4.1.2 Creative Senz3D 数据集

Creative Senz3D 数据集^[24]是通过 Creative Senz3D 相机获取的。该数据集包含 4 个不同的人执行的 11 个不同的手势,同时每个人将每个手势分别重复 30 次,总共进行了 1 320 次采集。对应 11 个手势的 RGB 图像图 6 所示,例如拳头,数字 1 至 5 等。



图6 Creative Senz3D 数据集 RGB 手势图像示意图

4.2 实验参数设置

所有的训练和测试均在台式电脑上进行,电脑的主要配置为 Intel(R) Core(TM) i5-8400 CPU @ 2.80 GHz、6GB 的 GPU NVIDIA GeForce GTX 1060 显卡和 8GB 的运行内存。程序在 Windows10 系统下运行,同时调用并行计算架构(CUDA)、英伟达神经网络库(Cudnn)、开源计算机视觉库(OPENCV)。模型训练时使用官网提供预训练权重文件来初始化本文的网络权重参数,以便训练时加快模型的收敛速度。在 yolov3 中,由于适合的先验尺寸将会直接影响到静态手势识别时的速度以及目标框在定位时的精度,因此对所有训练样本数据集上标注出的边界框数据采用 K-means 聚类来获得最优的 9 组先验框尺寸。具体训练参数值如表 2 和表 3 所示。其中,参数半径 r 用于构建手部骨架灰度图,缩放参数 e 用于 SE 模块的通道数缩放。

表2 Microsoft Kinect & Leap Motion 数据集的训练参数设置

参数名	参数值
批样本大小	16 个
批分割	8 个
迭代	20 000 次
学习率衰减步长	16 000 次, 18 000 次
学习率衰减因子	0.1, 0.1
动量	0.9
权重衰减	0.000 5
9 组先验框尺寸 (anchor)	(58, 75), (44, 106), (52, 96), (50, 121), (67, 91), (79, 81), (61, 115), (77, 108), (85, 123)
半径 r	15
缩放参数 e	16

表3 Creative Senz3D 数据集的训练参数设置

参数名	参数值
批样本大小	16 个
批分割	8 个
迭代	22 000 次
学习率衰减步长	17 000 次, 19 000 次
学习率衰减因子	0.1, 0.1
动量	0.9
权重衰减	0.0005
9 组先验框尺寸 (anchor)	(99, 159), (134, 144), (100, 221), (167, 164), (129, 214), (116, 252), (143, 249), (169, 232), (187, 285)
半径 r	15
缩放参数 e	16

在 Microsoft Kinect & Leap Motion 数据集上,最终模型共训练 20 000 次,耗时 29 h。训练模型时,从 14 个人中随机抽取 10 个人作为训练数据,即训练数据集共 1 000 张 RGB 图片和 1 000 张对应的手部骨架灰度图,剩余的 4 个人的 400 张 RGB 图像和 400 张对应的手部骨架灰度图作为测试数据集。在训练过程中,通过绘制模型损失曲线观察训练动态过程,其对应的损失值变化曲线如图 7 所示。可看出:模型在前期迭代中损失值缩减迅速,模型快速拟合;当迭代训练 4 000 次后,损失值下降放缓;当迭代至 20 000

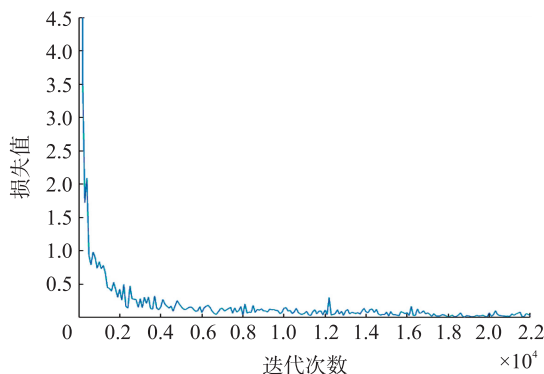


图7 Microsoft Kinect & Leap Motion
数据集训练时损失曲线图

次时,损失值收敛至 0.010 3,结束训练在 Creative Senz3D 数据集上,最终模型共训练 22 000 次,耗时 25 h,训练模型时,按照不同的手势执行人采用留一法将数据分为训练数据集和测试数据集。训练数据集由 3 个人共 990 张 RGB 图片和 990 张对应的手部骨架灰度图,剩余的 1 个人的 330 张 RGB 图像和 330 张对应的手部骨架灰度图作为测试数据集。在训练过程中,通过绘制模型损失曲线观察训练动态过程,其对应的损失值变化曲线如图 8 所示。可看出:模型在前期迭代中损失值快速缩减,模型快速拟合;当迭代训练 4 400 次后,损失值下降放缓;当迭代至 22 000 次时,损失值收敛至 0.021 5,结束训练。

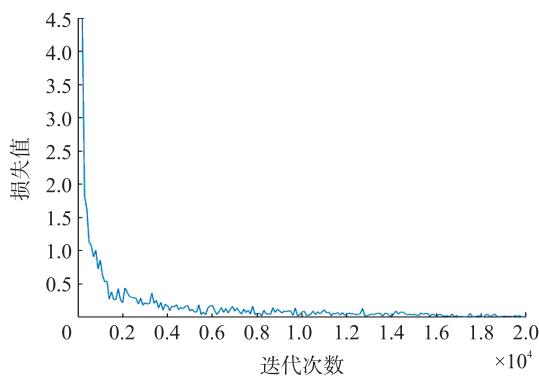


图 8 Creative Senz3D 数据集训练时损失曲线图

4.4 在 Microsoft Kinect & Leap Motion 数据集和 Creative Senz3D 数据集上实验结果分析

本文提出的一种融合手部骨架灰度图的深度神经网络,每一步改进分别在 Microsoft Kinect & Leap Motion 数据集上测试结果如表 4 所示,在 Creative Senz3D 数据集上测试结果如表 5 所示。

该网络算法在 Microsoft Kinect & Leap Motion 数据集上每一类的准确率情况,具体如表 6 所示,其中除 G7 和 G8 的准确率分别为 98.1% 和 98.7%,其他类别准确率均达到了 100%。在 Creative Senz3D 数据集上每一类的准确率情况,具体如表 7 所示,其中除 G1、G3 和 G4 的准确率分别为 99.31%、99.17% 和 99.32%,其他类别准确率均达到了 100%。

表 4 每一步改进在 Microsoft Kinect & Leap Motion 数据集上的测试结果

方法	平均准确率/%
Yolov3 ^[20] (RGB)	98
本文方法 (RGB+GHS)	99.2
本文方法 (RGB+GHS+RRReLU)	99.5
本文方法 (RGB+GHS+RRReLU+扩展卷积残差模块)	99.58
本文方法 (RGB+GHS+RRReLU+扩展卷积残差模块+SE 模块)	99.68

表 5 每一步改进在 Creative Senz3D 数据集上测试结果

方法	平均准确率/%
Yolov3 (RGB) ^[20]	98.3
本文方法 (RGB+GHS)	98.6
本文方法 (RGB+GHS+RRReLU)	98.7
本文方法 (RGB+GHS+RRReLU+扩展卷积残差模块)	99.6
RGB+GHS+RRReLU+扩展卷积残差模块+SE 模块)	99.8

表 6 Microsoft Kinect & Leap Motion 数据集上 10 类的测试结果

类别	平均准确率/%	类别	平均准确率/%
G1	100	G6	100
G2	100	G7	98.1
G3	100	G8	98.7
G4	100	G9	100
G5	100	G10	100

表 7 Creative Senz3D 数据集上 11 类的测试结果

类别	平均准确率/%	类别	平均准确率/%
G1	99.31	G7	100
G2	100	G8	100
G3	99.17	G9	100
G4	99.32	G10	100
G5	100	G11	100
G6	100		

不同的手势识别方法以及 yolov3 在 Microsoft Kinect & Leap Motion 数据集上测试结果如表 8 所示。通过对比可以清晰的看出,本文提出的方法比其他方法在平均准确率上至少提高了 2.88%,且基于 GHS 图像和 RGB 图像的手势识别速度为 26 fps。

表 8 不同算法在 Microsoft Kinect & Leap Motion 数据集上的测试结果

方法	平均准确率/%
3LM+2K ^[25] (Kinect)	89.7
3LM+2K ^[25] (Kinect+Leap Motion)	91.3
SFPS ^[26]	95
SP-EMD ^[18] (Kinect)	95.8
4LM+4K ^[23] (Kinect)	96.3
4LM+4K ^[23] (Kinect+Leap Motion)	96.5
CSG-EMD ^[27] (Kinect)	96.6
3D Shape Context ^[28]	96.8
本文方法	99.68

不同的手势识别方法以及 yolov3 在 Creative Senz3D 数据集上测试结果如表 9 所示。通过对比可以清晰的看出,本文提出的方法比其他方法在平均准确率上至少提高了 0.2%,且基于 GHS 图像和

RGB 图像的手势识别速度为 31 fps。

表 9 不同算法在 Creative Senz3D 数据集上的测试结果

方法	平均准确率/%
Convex Hull concavities ^[29]	65
Distance features ^[30]	75
HandSegNet ^[31]	77
Δ -GAN-no rolling ^[32]	88
3DFA ^[24]	90
Memo et al. ^[33]	90
Unsupervised Refinable Nets ^[34]	94
GestureGAN ^[35]	96
SP-EMD ^[36]	96.6
CSG-EMD ^[27]	97.4
SFPS ^[26]	97.4
PG ^[37]	98.7
Δ -GAN ^[32]	99
Ma et al. ^[38]	99
Yan et al. ^[39]	99.5
PoseGAN ^[40]	99.5
3D Shape Context ^[28]	99.6
本文方法	99.8

5 结论

通过上述实验结果表格的对比,说明无约束环境下静态手势在识别过程中,使用手部关键点及其相互关联性构建手部骨架灰度图,以 yolov3 为主干网络,添加扩展卷积残差模块和 SE 模块,并用 RReLU 函数作为激活函数的方法在识别准确率上有着较大的提升。该方法在 Microsoft Kinect & Leap Motion 数据集和 Creative Senz3D 数据集上均进行了测试对比实验,相比其他静态手势识别方法,本文提出的识别方法在这两个无约束环境下的静态手势数据集有着优秀的表现。在 Microsoft Kinect & Leap Motion 数据集和 Creative Senz3D 数据集上准确率均达到了最高,分别为 99.68% 和 99.8%,且基于 GHS 图像和 RGB 图像的手势识别速度分别为 26 fps 和 31 fps。

手势识别在空间定位方面依旧有着不小问题。如今的手势识别算法距离理想的空间定位能力,即得到手部的精准位置还有着不小的差距,而其也正是未来人机交互的重要基础。同时,本文提出的解决方法仍未能完全解决对应的挑战和难点,尤其在网络延迟和实时性方面依旧有着较大的提升空间。

参考文献:

[1] Wu G, Kang W. Robust Fingertip Detection in a Complex Environment[J]. IEEE Transactions on Multimedia, 2016, 18(6): 978-987.

[2] Mitra S, Acharya T. Gesture Recognition: A Survey[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2007, 37(3): 311-324.

[3] Pisharady P K, Saerbeck M. Recent Methods and Databases in Vision-Based Hand Gesture Recognition: A Review[J]. Computer Vision and Image Understanding, 2015, 141: 152-165.

[4] Lee H, Lim S Y, Lee I, et al. Multi-Modal User Interaction Method Based on Gaze Tracking and Gesture Recognition[J]. Signal Processing: Image Communication, 2013, 28(2): 114-126.

[5] Nyirarugira C, Kim T Y. Stratified Gesture Recognition Using the Normalized Longest Common Subsequence with Rough Sets[J]. Signal Processing: Image Communication, 2015, 30: 178-189.

[6] Nuzzi C, Pasinetti S, Lancini M, et al. Deep Learning Based Machine Vision: First Steps Towards a Hand Gesture Recognition Set Up for Collaborative Robots[C]//2018 Workshop on Metrology for Industry 4.0 and IoT. IEEE, 2018: 28-33.

[7] Oikonomidis I, Lourakis M I A, Argyros A A. Evolutionary Quasi-Random Search for Hand Articulations Tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 3422-3429.

[8] Pisharady P K, Vadakkepat P, Loh A P. Attention Based Detection and Recognition of Hand Postures Against Complex Backgrounds[J]. International Journal of Computer Vision, 2013, 101(3): 403-419.

[9] 谢仁强, 曹俊诚. 基于加速度传感器的可扩展手势识别[J]. 传感技术学报, 2016, 29(5): 659-664.

[10] Ji P, Song A, Xiong P, et al. Egocentric-Vision Based Hand Posture Control System for Reconnaissance Robots[J]. Journal of Intelligent & Robotic Systems, 2017, 87(3-4): 583-599.

[11] Tellaeche A, Kildal J, Maurtua I. A Flexible System for Gesture Based Human-Robot Interaction[J]. Procedia CIRP, 2018, 72: 57-62.

[12] 鲍磊, 罗志增, 席旭刚, 等. 融合表面肌电和加速度的手势动作识别[J]. 传感技术学报, 2019, 32(12): 85-90.

[13] Köpüklü O, Gunduz A, Kose N, et al. Real-Time Hand Gesture Detection and Classification Using Convolutional Neural Networks[C]//2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE, 2019: 1-8.

[14] Lin C, Lin X, Xie Y, et al. Abnormal Gesture Recognition Based on Multi-Model Fusion Strategy[J]. Machine Vision and Applications, 2019, 30(5): 889-900.

[15] Lin H I, Hsu M H, Chen W K. Human Hand Gesture Recognition Using a Convolution Neural Network[C]//2014 IEEE International Conference on Automation Science and Engineering (CASE). IEEE, 2014: 1038-1043.

[16] Arenas J O P, Moreno R J, Murillo P C U. Hand Gesture Recognition by Means of Region-Based Convolutional Neural Networks[J]. Contemporary Engineering Sciences, 2017, 10(27): 1329-1342.

[17] Nuzzi C, Pasinetti S, Lancini M, et al. Deep Learning Based Machine Vision: First Steps Towards a Hand Gesture Recognition Set Up for Collaborative Robots[C]//2018 Workshop on Metrology for Industry 4.0 and IoT. IEEE, 2018: 28-33.

[18] Alnujaim I, Alali H, Khan F, et al. Hand Gesture Recognition

- Using Input Impedance Variation of Two Antennas with Transfer Learning[J]. IEEE Sensors Journal, 2018, 18(10): 4129–4135.
- [19] Nuzzi C, Pasinetti S, Lancini M, et al. Deep Learning-Based Hand Gesture Recognition for Collaborative Robots[J]. IEEE Instrumentation & Measurement Magazine, 2019, 22(2): 44–51.
- [20] Redmon J, Farhadi A. Yolov3: An Incremental Improvement [J/OL]. arXiv preprint arXiv:1804.02767, 2018. <https://arxiv.org/pdf/1804.02767.pdf>.
- [21] Simon T, Joo H, Matthews I, et al. Hand Keypoint Detection in Single Images Using Multiview Bootstrapping[C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2017: 1145–1153.
- [22] Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132–7141.
- [23] Marin G, Dominio F, Zanuttigh P. Hand Gesture Recognition with Jointly Calibrated Leap Motion and Depth Sensor[J]. Multimedia Tools and Applications, 2016, 75(22): 14991–15015.
- [24] Memo A, Zanuttigh P. Head-Mounted Gesture Controlled Interface for Human-Computer Interaction[J]. Multimedia Tools and Applications, 2018, 77(1): 27–53.
- [25] Marin G, Dominio F, Zanuttigh P. Hand Gesture Recognition with Leap Motion and Kinect Devices[C]//2014 IEEE International conference on image processing (ICIP). IEEE, 2014: 1565–1569.
- [26] He Y, Yang J, Shao Z, et al. Salient Feature Point Selection for Real Time RGB-D Hand Gesture Recognition[C]//2017 IEEE International Conference on Real-Time Computing and Robotics (RCAR). IEEE, 2017: 103–108.
- [27] Wang C, Liu Z, Zhu M, et al. A Hand Gesture Recognition System Based on Canonical Superpixel-Graph[J]. Signal Processing: Image Communication, 2017, 58: 87–98.
- [28] Zhu C, Yang J, Shao Z, et al. Vision Based Hand Gesture Recognition Using 3d Shape Context[J]. IEEE/CAA Journal of Automatica Sinica, 2019(99): 1–14.
- [29] Dominio F, Marin G, Piazza M, et al. Feature Descriptors for Depth-Based Hand Gesture Recognition[M]//Computer Vision and Machine Learning with RGB-D Sensors. Springer, Cham, 2014: 215–237.
- [30] Dominio F, Donadeo M, Zanuttigh P. Combining Multiple Depth-Based Descriptors for Hand Gesture Recognition[J]. Pattern Recognition Letters, 2014, 50: 101–111.
- [31] Zimmermann C, Brox T. Learning to Estimate 3d Hand Pose from Single RGB Images[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 4903–4911.
- [32] Liu Y, De Nadai M, Zen G, et al. Gesture-to-Gesture Translation in the Wild via Category-Independent Conditional Maps[C]//Proceedings of the 27th ACM International Conference on Multimedia. 2019: 1916–1924.
- [33] Memo A, Zanuttigh P. Head-Mounted Gesture Controlled Interface for Human-Computer Interaction[J]. Multimedia Tools and Applications, 2018, 77(1): 27–53.
- [34] Dibra E, Melchior S, Balkis A, et al. Monocular RGB Hand Pose Inference from Unsupervised Refinable Nets[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018: 1075–1085.
- [35] Tang H, Wang W, Xu D, et al. Gesturegan for Hand Gesture-to-Gesture Translation in the Wild[C]//Proceedings of the 26th ACM international conference on Multimedia. 2018: 774–782.
- [36] Wang C, Liu Z, Chan S C. Superpixel-Based Hand Gesture Recognition with Kinect Depth Camera[J]. IEEE Transactions on Multimedia, 2014, 17(1): 29–39.
- [37] Ma L, Jia X, Sun Q, et al. Pose Guided Person Image Generation[C]//Advances in Neural Information Processing Systems. 2017: 406–416.
- [38] Ma L, Sun Q, Georgoulis S, et al. Disentangled Person Image Generation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 99–108.
- [39] Yan Y, Xu J, Ni B, et al. Skeleton-Aided Articulated Motion Generation[C]//Proceedings of the 25th ACM International Conference on Multimedia. 2017: 199–207.
- [40] Siarohin A, Sangineto E, Lathuilière S, et al. Deformable Gans for Pose-Based Human Image Generation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 3408–3416.



章东平(1970—),男,博士,教授,主要研究方向为计算机视觉、人工智能, 06a0303103@cjlu.edu.cn;



束元(1995—),男,硕士研究生,主要研究方向为计算机视觉、手势识别, 1316009369@qq.com。