

# 基于序列特征的2D CNN的动态手势识别

李振宇,张立民,邓向阳,王彦哲

(海军航空大学, 山东 烟台 264001)

**摘要:**为了兼顾识别准确度和运行速度,改进了2D卷积神经网络提取多帧特征并使用长短期记忆网络进行处理特征序列,使用Softmax分类器输出分类结果;实验结果表明:基于序列特征的2D CNN网络在CHGDs数据集上的识别准确率达86.97%,比CNN卷积神经网络提高了11.99%,与3D CNN性能基本相当的同时,速度是3D CNN的6.98倍。

**关键词:**深度学习;动态手势识别;2D卷积神经网络;3D卷积神经网络;长短期记忆网络

**本文引用格式:**李振宇,张立民,邓向阳,等.基于序列特征的2D CNN的动态手势识别[J].兵器装备工程学报,2019,40(2):141-146.

**Citation format:** LI Zhenyu, ZHANG Liming, DENG Xiangyang, et al. Dynamic Desture Recognition Based on Sequence Feature 2D CNN[J]. Journal of Ordnance Equipment Engineering, 2019, 40(2): 141-146.

中图分类号: TP39

文献标识码: A

文章编号: 2096-2304(2019)02-0141-06

## Dynamic Desture Recognition Based on Sequence Feature 2D CNN

LI Zhenyu, ZHANG Liming, DENG Xiangyang, WANG Yanzhe

(Naval Aviation University, Yantai 264001, China)

**Abstract:** In order to balance the recognition accuracy and running speed, this paper improved 2D convolutional neural network to extract multi-frame features and used long short-term memory (LSTM) to process the feature sequences, and used the softmax classifier to output the classification results. The experimental results show that the recognition accuracy of the 2D CNN network based on sequence features on the CHGDs dataset is 86.97%, which is 11.99% higher than that of the CNN convolutional neural network. Compared with the performance of 3D CNN, the speed is 6.98 of 3D CNN Times.

**Key words:** deep learning; dynamic gesture recognition; 2D convolutional neural network; 3D convolutional neural network; long short-term memory (LSTM)

手势是一种表示信息的人体行为动作,因其具有含义丰富、通用性高和蕴含感情的优点<sup>[1]</sup>,被广泛应用于智能家居<sup>[2]</sup>、虚拟现实等<sup>[3]</sup>领域。现阶段基于视觉的手势识别所需设备简单,使用也比较方便,因此对其进行研究具有非常重要的实际意义。由于手势随个体的不同具有多样性、多义性和随机性等特点<sup>[4]</sup>,而手势识别的应用领域通常涉及人机交互,这就需要相关算法在具有较好的鲁棒性的同时,具有较

高的实时性<sup>[5]</sup>。

在静态手势识别方面,常用方法有基于模板匹配的方法<sup>[6]</sup>、基于图像特征方法<sup>[7]</sup>、支持向量机(SVM)<sup>[8]</sup>等。这些方法进行静态识别时对图像本身具有较高的要求,无法有效应对动态手势中容易出现的光线变化、角度变化、形状变化等问题。为此,熵分析法<sup>[9]</sup>、隐马尔科夫模型<sup>[10]</sup>、光流法<sup>[11]</sup>等用于动态手势识别的方法也相继被提出。但熵分析法稳

收稿日期:2018-09-08;修回日期:2018-10-20

基金项目:国家自然科学基金重大研究计划资助项目(91538201);泰山学者工程专项经费资助项目(TS201511020)

作者简介:李振宇(1993—),男,硕士研究生,主要从事深度学习研究。

通讯作者:张立民(1966—),男,教授,博士生导师,主要从事电子系统仿真、人工智能研究, E-mail: iamwyz@foxmail.com。

定性较差,隐马尔科夫模型在处理复杂手势时数学模型极为复杂,稀疏光流法存在角点检测不稳定的缺点,稠密光流法运行速率慢。从2012年 AlexNet<sup>[12]</sup> 获得 ILSVRC2012 的冠军开始,基于深度学习的方法在视觉图像方面获得了巨大的优势,在民用领域和军用领域<sup>[13]</sup> 都得到应用。但无论是 GoogLeNet<sup>[14]</sup> 还是 ResNet<sup>[15]</sup>, 都是对单张图片进行识别,缺乏时间上的特征提取。文献[16]提出和使用3D卷积神经网络(Convolutional Neural Network, CNN)对时间序列上的特征进行提取,取得了一定效果。但3D卷积神经网络使用的3D卷积核引入了大量的参数,使得网络复杂度增高,计算开销增大,极大地增加了训练识别和工程化难度。

本文提出了一种基于2D CNN的动态手势识别方法,针对传统2D CNN无法对时间序列特征进行提取的缺点,对其网络结构进行改进,使其能够用于视频序列的识别,然后将该模型用于动态手势识别,验证该模型的有效性。

## 1 经典卷积神经网络

卷积神经网络(Convolutional Neural Network, CNN)<sup>[17]</sup> 在视觉处理领域具有独特的优势,通过引入感受野这一生物学概念采取了信号—低级抽象—高级抽象的识别。通过逐步提取局部特征,最后到全局特征的方法,使得卷积神经网络对图像在平移变换,缩放,旋转或其他方式的形变上具有高度不变性的。卷积神经网络的组成结构主要包括卷积层(Convolutional layer),池化层(下采样层 Down-pooling),全连接层(Fully Connected Layer),激活函数(Activation Function)和分类器(Classifier)。经典2D卷积神经网络结构如图1所示。

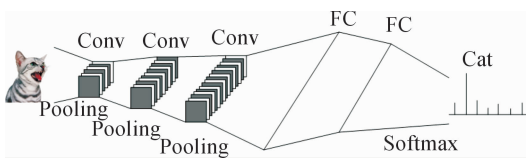


图1 卷积神经网络结构

在卷积神经网络中,最核心的特征提取方式是由卷积核对输入进行卷积计算。不同的卷积核不同的特征具有不同的兴奋特性,通过使用不同的卷积核能够提取不同的网络感兴趣特征。同时,在卷积层使用权值共享的方式大大减少了神经元个数,加快了网络的参数更新和计算。卷积过程示意图如图2。

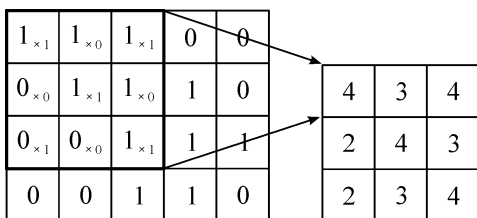


图2 卷积过程示意图

$$x_j^n = f\left(\sum_{i \in M_i} x_i^{n-1} * k_{ij} + b_j^n\right) \quad (1)$$

式(1)中, $x_j^n$ 表示的是第 $n$ 个卷积层的第 $j$ 张特征映射图, $x_i^{n-1}$ 表示的是第 $n-1$ 层的第 $i$ 个输出特征映射图, $f(\cdot)$ 表示激活函数, $M_i$ 是所有输入图像的总集合, $k_{ij}$ 表示上一个层输出的第 $i$ 个特征映射图与目前层的第 $j$ 个特征映射图之间的卷积核, $b_j^n$ 则表示第 $n$ 层卷积层的偏置。为了使提取的特征经过逐层提取后趋于全局化,卷积神经网络通常使用多层卷积层反复提取。

为了有效地降低特征分辨率,获得空间不变性和避免过拟合等问题,CNN使用了池化处理方式,池化过程由式(2)表示。

$$x_j^n = f\left(\beta_j^{n-1} \text{down}(x_j^{n-1}) + b_j^n\right) \quad (2)$$

其中, $x_j^n$ 表示第 $n$ 个池化层的第 $j$ 个特性映射图,函数 $f(\cdot)$ 表示激活函数,down表示下采样,即池化的方法,常用的池化方法有最大值池化和均值池化, $\beta_j^n$ 表示乘性偏置, $b_j$ 表示加性偏置。

当图像经过卷积层和池化层逐层提取特征后,将获得最终的特征映射图,将这些特征映射图与全连接层相连,使每个输入与输出的神经元产生联系,最大程度的表示全局特征。全连接层可以由式(3)表示:

$$h(x) = f(W^T x + b) \quad (3)$$

其中, $x$ 为全连接层的输入特征, $h(x)$ 为全连接层的输出, $W$ 表示全连接层每个神经元的权值, $b$ 表示偏置,函数 $f(\cdot)$ 表示激活函数。

3D卷积神经网络与经典2D卷积神经网络在结构上基本一样,其差别主要体现在卷积核。3D卷积神经网络采用3个维度描述的卷积核,使得3D卷积神经网络能同时输入多帧图片进行特征提取,从而提取到2D网络无法提取到的时间序列信息。3D卷积计算过程如式(4)表示:

$$v_{ij}^{xy} = \tanh\left(b_{ij} + \sum_{m=0}^{P_i-1} \sum_{p=0}^{Q_i-1} \sum_{q=0}^{R_i-1} w_{ijm}^{pq} v_{(i-1)m}^{((x+p)(y+q))}\right) \quad (4)$$

## 2 基于序列特征的2D CNN结构

为了能兼顾性能和速度,本文提出由经典的CNN网络提取单帧图片的特征,再将这些静态特征序列进行时序上的融合处理,最后由softmax分类器进行分类的基于序列特征的2D CNN结构(Sequence Feature CNN, SFCNN),网络结构如图3所示。

该结构由前后两部分组成,前半部分为CNN的卷积池化层部分,用于单帧空间特征提取,后半部分使用长短期记忆网络(Long Short-Term Memory, LSTM)对序列特征进行提取。当输入多帧图片时,2D卷积神经网络分别对每张图片进行特征提取,获得一个固定大小的特征向量,然后将这些特征按照实际产生的顺序,依次作为LSTM的输入。每一次特征向量经过LSTM都会对LSTM的细胞状态进行更新,将

从初始到结束的最重要信息保留下来。

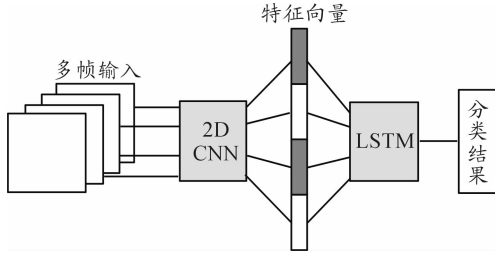


图3 网络结构

LSTM 作为一种循环神经网络,继承了循环神经网络在时间关联上的特征提取能力,同时解决了循环神经网络在长时间跨度上梯度消失和梯度爆炸的问题,同时相较于全连接网络,能够保留在时间上的时序特征。LSTM 的基本结构为块(block),如图4所示,在每一个 block 中都有4个输入和1个输出,其中这四个输入都是相同的。在 LSTM 中,设置一个细胞状态(cell state)作为信息的寄存器,它贯穿于整个网络状态,用以储存时间维度上有关联的关键信息,只接受少量的来自网络的线性交互。4个输入分别对应1个输入和3个门(gate),这3个门分别是输入门(input gate)、遗忘门(forget gate)、输出门(output gate)。这里的门实际上是一个输入为向量的全连接层,输出为0~1之间的向量。细胞是一个只继承自己上一个状态的量,在本时刻状态,它能由遗忘门来对自己进行选择性遗忘,遗忘内容由上一时刻输出和本时刻输入共同决定的,遗忘程度是由遗忘门输出值为0、1之间的向量来控制的。同时,细胞状态也能通过输入门向自己添加部分新的内容,添加的内容也是由上一时刻的输出和本时刻的输入共同决定的,而添加的权重则是由输入门输出值为0、1之间的向量来控制的。输出门则是用来控制细胞状态向隐藏节点的输出量。为了更好的利用上一时刻的细胞状态,通过在块内添加窥视孔(peephole)结构,能够更好的将输入门、遗忘门、输出门和上一时刻的细胞状态充分联系起来。

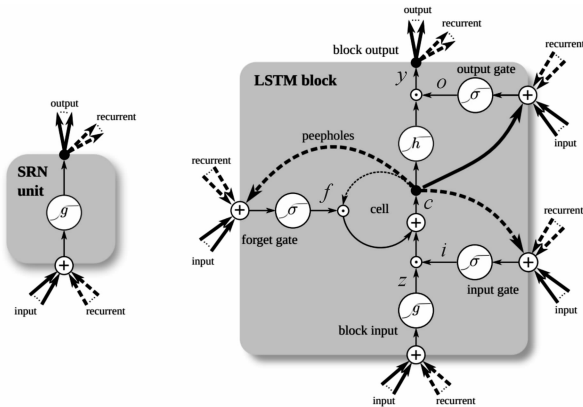


图4 LSTM 结构

假设在  $t$  时刻,输入图片经过卷积池化后得到向量为  $x^t$ ,

$N$  表示第  $N$  个块, $M$  表示第  $M$  个输出。那么该块的细胞状态输入为

$$\bar{z}^t = W_z x^t + R_z y^{t-1} + b_z \quad (5)$$

$$z^t = g(\bar{z}^t)$$

输入门矩阵为

$$\bar{i}^t = W_i x^t + R_i y^{t-1} + p_i \odot c^{t-1} + b_i \quad (6)$$

$$i^t = \sigma(\bar{i}^t)$$

遗忘门矩阵为:

$$\bar{f}^t = W_f x^t + R_f y^{t-1} + p_f \odot c^{t-1} + b_f \quad (7)$$

$$f^t = \sigma(\bar{f}^t)$$

将上一时刻的细胞状态进行更新到当前时刻:

$$c^t = z_t \odot i^t + c_{t-1} \odot f^t \quad (8)$$

得到输出门矩阵:

$$\bar{o}^t = W_o x^t + R_o y^{t-1} + p_o \odot c^{t-1} + b_o \quad (9)$$

$$o^t = \sigma(\bar{o}^t)$$

最后该块输出为

$$y^t = h(c^t) \odot o^t \quad (10)$$

输入权值为  $W_z, W_i, W_f, W_o \in \mathbb{R}^{N \times M}$ , 反馈权值为  $R_z, R_i, R_f, R_o \in \mathbb{R}^{N \times M}$ , 窥视孔权值为  $p_i, p_f, p_o \in \mathbb{R}^N$ , 偏置为  $b_z, b_i, b_f, b_o \in \mathbb{R}^N$ ,  $g, \sigma$  和  $h$  都是逐点非线性激活函数,  $\sigma$  是 sigmoid 激活函数 ( $\sigma(x) = \frac{1}{1 + e^{-x}}$ ),  $g$  和  $h$  均为 tangent 激活函数 ( $g(x) = h(x) = \tanh(x)$ )。

最后对输出的结果  $y^t$  使用 softmax 分类器进行分类,得到预测结果:

$$S_i = \frac{e^{y_i^t}}{\sum_j e^{y_j^t}} \quad (11)$$

假设有  $N$  张大小为  $W \times H$  图片作为输入,对  $M$  种类别进行分类,SFCNN 网络结构的时间复杂度可以由式(12)来描述:

$$K = N \sum_s \sum_a^W \sum_b^H \left( \sum_{c=1}^m \sum_{d=1}^n k_{cd} X_{ab} \right) + \frac{WHNL'M}{s^2} \quad (12)$$

式中, $s$  表示 2D CNN 的卷积层,卷积核大小为  $i \times j$ ,  $k$  为卷积核,  $X$  表示当前输入,  $t$  表示 LSTM 网络的隐藏层层数,每层由  $l$  个神经元组成。

当 2D CNN 最后的全连接层与 SFCNN 结构的 LSTM 结构具有相同尺寸时,2D CNN 有着和式(12)一样的时间复杂度。

3D CNN 的时间复杂度可以由式(13)表示:

$$K = \sum_a^W \sum_b^H \left( \sum_{c=1}^m \sum_{d=1}^n k_{cd} X_{ab} \right) \quad (13)$$

$$C = \sum_{t=1}^T \sum_{p=1}^P \sum_{q=1}^Q K_{pqt} + \frac{WHNL'M}{s^2}$$

$P, Q$  分别表示 3D 卷积核的个数和每个卷积核第 3 个维度大小,  $T$  为 3D CNN 的卷积层数。

从式(12)和式(13)可以看出本文提出的结构 SFCNN 计算复杂度与经典 2D CNN 基本相当,远低于 3D CNN。

### 3 实验结果与分析

本文实验采用的硬件为 Intel i5-8400 CPU, 8GB 内存, 软件平台为 Windows 系统下的 TensorFlow 平台。实验使用的数据集为 Cambridge-Gesture data base<sup>[18]</sup>, 该数据集包含 9 种不同的运动手势: Flat/Leftward (FL)、Spread/Leftward (SL)、V-shape/Leftward (VL)、Flat/Rightward (FR)、Spread/Rightward (SR)、V-shape/Rightward (VR)、Flat/Contract (FC)、Spread/Contract (SC)、V-shape/Contract (VC), 具体的手势示意图如图 5。该数据集中, 每个类别手势包含 100 个视频片段总共约 64 000 张图片, 每张图片均为  $320 \times 240$  的彩色图像。

在本实验中, 使用的 SFCNN 结构采用 2 层卷积的 2D CNN 和 2 层隐藏层的 LSTM 结构, 作为基本模型, 其主要构成网络层如表 1 所示。

表 1 基于序列特征的 2D 卷积神经网络

层数	类型	尺寸	层数	类型	尺寸
1	输入层	-	4	隐藏层 H1	64
2	卷积层 C1	$7 \times 7$	5	隐藏层 H1	64
3	卷积层 C1	$5 \times 5$	6	输出层	9

将 CGdb 数据集中的图片进行灰度化、高斯模糊处理和直方图均衡化处理, 送入网络进行训练。不同的超参数对网络的识别效果会产生很大的影响, 其准确率变化训练损失函数下降如图 6 所示, 实验中采用了 4 组超参数进行实验, 其中学习率  $L=0.005$ 、隐层节点个数  $N=32$  时效果最好, 准确率达到 86.97%。当准确率达到峰值后, 网络的识别准确率随着迭代次数的增加而降低, 说明此时网络开始过拟合, 从损失函数急剧上升也可以得到过拟合的结论。

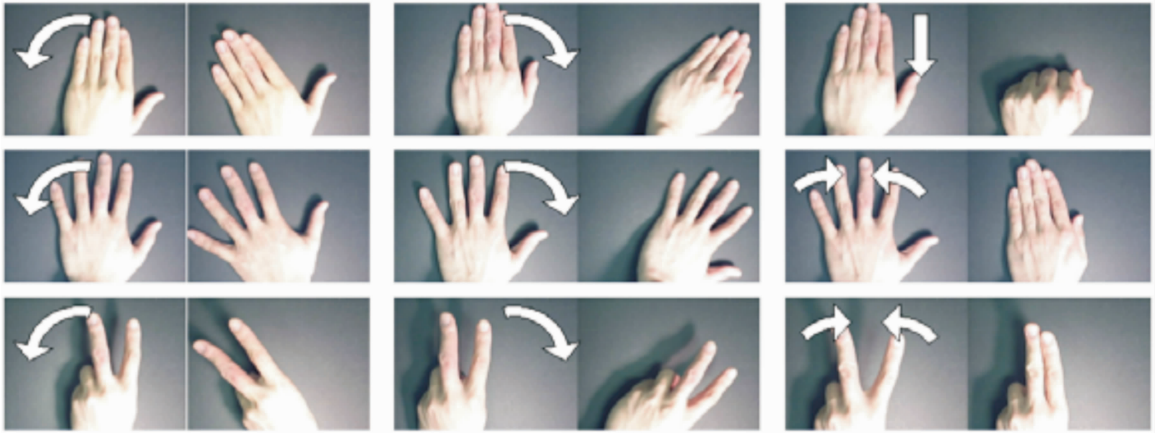


图 5 Cambridge-Gesture data base 9 种示意图

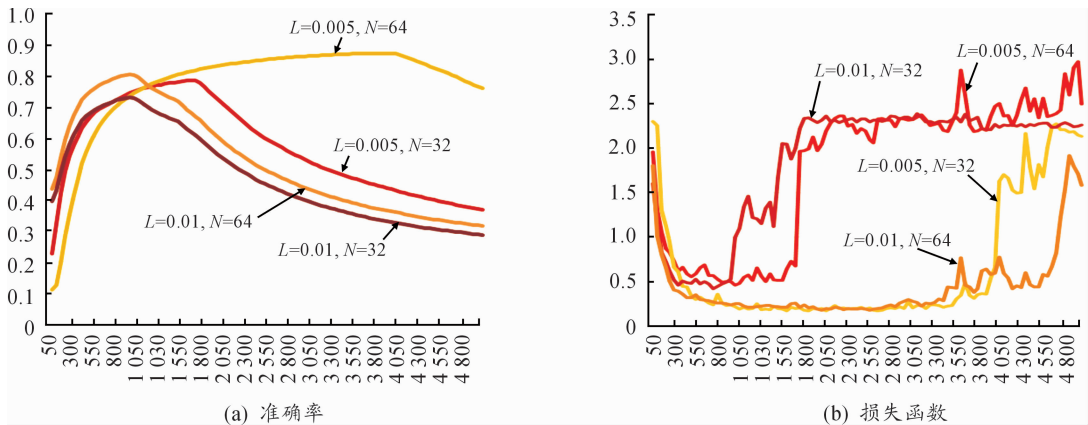


图 6 SFCNN 实验结果

选取性能最佳的超参数组  $L=0.005$ 、 $N=32$ , 画出混淆矩阵, 如图 7 所示, 横向标签代表真实标签, 纵向标签代表分类标签。

从混淆矩阵可以发现, SFCNN 在对 FC 和 SC 两种分类

上产生了大量的误判, 而其他几种类别相对判断比较准确。从网络结构上分析, 当其他具有相同静态姿势的手势具有较好分辨率的情况下, 说明网络的 CNN 部分对特征提取是准确的, 而 FC 和 SC 两种手势具有相同的静态姿势, 说明 LSTM

网络在细胞状态更新上出现了误判,将细胞状态更新到了错误状态或保留了错误状态。

Flat/Leftward (FL)	0.90	0	0	0.04	0	0.01	0.02	0	0.01
Spread/Leftward (SL)	0.01	0.93	0.01	0	0.04	0	0	0	0.01
V-shape/Leftward (VL)	0.01	0	0.96	0	0	0.01	0	0	0.02
Flat/Rightward (FR)	0.02	0	0	0.89	0.07	0	0.02	0	0
Spread/Rightward (SR)	0	0.07	0	0	0.91	0	0.01	0	0.01
V-shape/Rightward (VR)	0	0.02	0	0	0.01	0.95	0	0.02	0
Flat/Contract (FC)	0.02	0	0	0.01	0	0	0.72	0.22	0.01
Spread/Contract (SC)	0	0.01	0	0	0.02	0	0.19	0.71	0.04
V-shape/Contract (VC)	0.02	0.01	0.07	0.01	0.01	0.04	0.02	0.01	0.81
Flat/Leftward (FL)									
Spread/Leftward (SL)									
V-shape/Leftward (VL)									
Flat/Rightward (FR)									
Spread/Rightward (SR)									
V-shape/Rightward (VR)									
Flat/Contract (FC)									
Spread/Contract (SC)									
V-shape/Contract (VC)									

图7 SFCNN 混淆矩阵

用于与 SFCNN 进行识别准确率比较的 3D CNN 采用 3 层卷积和 3 层池化交替结构,每个卷积核均为  $5 \times 5 \times 3$  大小。表 2、表 3 将 SFCNN 与 3D CNN 对数据库每个类别的识别准确率进行了列举,从最终的平均识别率来看,SFCNN 与 3D CNN 的性能相当,从各个子类别来看,去除 FC 和 SC 两种类别后,SFCNN 性能甚至优于 3D CNN。由于在准确率比较中 3D CNN 使用了 3 个卷积池化层结构,直接用来与 2 个卷积池化层的 SFCNN 进行速度上的比较是不公正的,因此在进行速度比较时,为了体现相同深度下二者速度差异,削减了一个 3D CNN 的卷积池化层,其他参数设置均参照 SFCNN。表 4 比较了两种网络的运行速度,以每个网络完成一个 Epoch 作为标准,若以 3D CNN 的速度作为 1,那么 SFCNN 的速度为 6.98,远远优于 3D CNN 的速度。

表 2 SFCNN 每个类别识别结果

类别	识别率/%
Flat/Leftward (FL)	92.14
Spread/Leftward (SL)	93.27
V-shape/Leftward (VL)	95.57
Flat/Rightward (FR)	89.03
Spread/Rightward (SR)	90.71
V-shape/Rightward (VR)	94.63
Flat/Contract (FC)	72.02
Spread/Contract (SC)	74.02
V-shape/Contract (VC)	81.38
平均值	86.97

表 5 中列举了传统方法小波变换 + 支持向量机 (Gabor + SVM)、深度学习方法 CNN、3D CNN 和 SFCNN 在 CGdb 数据集上的识别效果,其中 Gabor 采用 4 尺度 8 方向的小波核。SFCNN 虽然比 3D CNN 略低 1.34%,但相比于

Gabor + SVM 和 CNN 提高了 32.28% 和 11.99%,与 3D CNN 识别效果处于同一级别。

表 3 3D CNN 每个类别识别结果<sup>[19]</sup>

类别	识别率/%
Flat/Leftward (FL)	87.24
Spread/Leftward (SL)	84.15
V-shape/Leftward (VL)	91.63
Flat/Rightward (FR)	88.71
Spread/Rightward (SR)	89.32
V-shape/Rightward (VR)	92.04
Flat/Contract (FC)	90.40
Spread/Contract (SC)	84.44
V-shape/Contract (VC)	86.90
平均值	88.31

表 4 网络运行速度

	3D CNN	SFCNN
运行速度(每个 Epoch 时间)	426 s	61 s
性能提升	1	6.98

表 5 多种方法在 CGdb 上性能比较

方法	准确率/%
Gabor + SVM	64.69
CNN	74.98
3D CNN	88.31
SFCNN	86.97

## 4 结论

1) 本文提出的 SFCNN 结构能够基本达到 3D CNN 识别效果的同时,运行速度远远优于 3D CNN。

2) 虽然 SFCNN 能够同时兼顾性能和速度,但是受到网络结构参数和超参数影响较大,需要大量时间对网络调参。

## 参考文献:

- [1] 王海鹏,龚岩,刘武,等.一种时空多尺度适应的手势识别方法研究[J].计算机科学,2017,44(12):287-291.
- [2] 肖寒君.基于深度传感器的手势追踪系统设计与实现[D].哈尔滨:哈尔滨工业大学,2017.
- [3] 赵小川.手势识别技术研究综述与应用展望[C]//第十一届全国信号和智能信息处理与应用学术会议专刊.中

- 国高科技产业化研究会智能信息处理产业化分会、中国高科技产业化研究会信号处理专家委员会,2017.
- [4] 刘蓉,刘明.基于三轴加速度传感器的手势识别[J].计算机工程,2011,37(24):141-143.
- [5] BARROS P,PARISI G I,JIRAK D,et al. Real-time Gesture Recognition Using a Humanoid Robot with a Deep Neural Architecture[C]//Humanoid Robots (Humanoids),2014 14th IEEE-RAS International Conference on. IEEE,2014:646-651.
- [6] 王浩宇,漆晶,方天恩,等.基于轨迹模板匹配的动态手势识别方法[J].单片机与嵌入式系统应用,2017,17(7):39-43.
- [7] 郭兴伟,葛元,王林泉.基于形状特征的字母手势的分类及识别算法[J].计算机工程,2004,30(18):130-132.
- [8] 刘小建,张元.基于多特征提取和 SVM 分类的手势识别[J].计算机工程与设计,2017,38(4):953-958.
- [9] LEE J S,LEE Y J,LEE E H,et al. Hand Region Extraction and Gesture Recognition from Video Stream with Complex Background through Entropy Analysis[C]//The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE,2004:1513-1516.
- [10] LIU N,LOVELL B C,KOOTSOOKOS P J,et al. Model Structure Selection & Training Algorithms for an HMM Gesture Recognition System[C]//Frontiers in Handwriting Recognition,2004. IWFHR-9 2004. Ninth International Workshop on. IEEE,2004:100-105.
- [11] 刘江华,陈佳品,程君实.基于光流及耦合隐马尔可夫模型的动态手势识别[J].上海交通大学学报,2003,37(5):720-723.
- [12] KRIZHEVSKY A,SUTSKEVER I,HINTON G E. Imagenet Classification with Deep Convolutional Neural Networks[C]//Advances in neural information processing systems. 2012:1097-1105.
- [13] 万鸾飞.基于多机器人离线编程系统设计[J].重庆工商大学学报(自然科学版),2016,33(6):94-97.
- [14] SZEGEDY C,LIU W,JIA Y,et al. Going Deeper with Convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition,2015:1-9.
- [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,2016:770-778.
- [16] JI S,XU W,YANG M,et al. 3D Convolutional Neural Networks for Human Action Recognition[J]. IEEE transactions on pattern analysis and machine intelligence,2013,35(1):221-231.
- [17] LECUN Y,BOTTOU L,BENGIO Y,et al. Gradient-based Learning Applied to Document Recognition[J]. Proceedings of the IEEE,1998,86(11):2278-2324.
- [18] KIM T K,CIPOLLA R. Canonical Correlation Analysis of Video Volume Tensors for Action Categorization and Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2009,31(8):1415-1428.
- [19] 易生,梁华刚,茹锋.基于多列深度 3D 卷积神经网络的手势识别[J].计算机工程,2017,34(8):243-248.

(责任编辑 杨继森)