

基于多特征加权融合的静态手势识别^①



陈影柔, 田秋红, 杨慧敏, 梁庆龙, 包嘉欣

(浙江理工大学 信息学院, 杭州 310018)

通讯作者: 田秋红, E-mail: tianqihong@zstu.edu.cn

摘要: 针对人工提取特征的单一性及卷积神经网络提取特征的遗漏性问题, 提出了一种基于多特征加权融合的静态手势识别方法. 首先, 提取分割后的手势图像的傅里叶和 Hu 矩等形状特征, 将两者融合作为手势图像的局部特征; 设计双通道卷积神经网络提取手势图像的深层次特征, 采用主成分分析方法对提取的特征进行降维; 然后, 将提取的局部特征和深层次特征进行加权融合作为手势识别的有效特征描述; 最后, 使用 Softmax 分类器进行手势图像分类. 实验结果验证了提出方法的有效性, 在手势图像数据集上的识别准确率达到 99% 以上.

关键词: 手势识别; 多特征加权融合; 局部特征; 双通道; 深层次特征

引用格式: 陈影柔, 田秋红, 杨慧敏, 梁庆龙, 包嘉欣. 基于多特征加权融合的静态手势识别. 计算机系统应用, 2021, 30(2): 20-27. <http://www.c-s-a.org.cn/1003-3254/7748.html>

Static Hand Gesture Recognition Based on Multi-Feature Weighted Fusion

CHEN Ying-Rou, TIAN Qiu-Hong, YANG Hui-Min, LIANG Qing-Long, BAO Jia-Xin

(School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: Static hand gesture recognition based on multi-feature weighted fusion is proposed to solve the problems of singularity and omission in convolutional neural network for feature extraction. Firstly, the Fourier and Hu moments of the segmented gesture image are extracted and fused as the local features. Besides, a dual-channel convolutional neural network is designed to extract the deep features of the gesture image, which are further treated by dimensionality reduction by principal component analysis. Secondly, the extracted local and deep features are weighted and fused as effective description for hand gesture recognition. Finally, gesture images are classified with Softmax classifier. Experimental results verify the proposed method, and the recognition accuracy reaches over 99% on the image dataset.

Key words: hand gesture recognition; multi-feature weighted fusion; local feature; dual-channel; deep hierarchical feature

手势识别作为一种新兴的人机交互技术, 被广泛应用到虚拟现实系统、互动游戏平台、手语识别和控制机器人等领域^[1-4], 其最终目的是实现人机自然交互. 最初的手势识别通过佩戴内置传感器的物理设备来获取手势的空间信息, 从而识别出手势语义信息. 佩戴物

理设备的方式虽能提高手势识别的准确度, 但在一定程度上影响了人机交互的自然性. 因此, 目前众多学者转向基于视觉的手势识别研究. 基于视觉的手势识别利用图像采集设备来捕捉手势图像, 通过对手势图像分析以提取出手势特征, 进而识别出手势语义信息^[5].

① 基金项目: 国家自然科学基金 (51405448); 浙江理工大学博士科研启动项目 (11122932611817); 浙江省大学生科技成果推广项目 (14530031661961); 国家级大学生创新创业训练计划 (201910338012); 浙江理工大学大学生科创项目 (11120032662029, 11120132662028)

Foundation item: National Natural Science Foundation of China (51405448); Scientific Research Start-up Fund for Doctorate of Zhejiang Sci-Tech University (11122932611817); Graduates Scientific Research Achievement Promotion Program of Zhejiang Province (14530031661961); National College Student Innovation and Entrepreneurship Training Program (201910338012); Science Innovation Program for College Student of Zhejiang Sci-Tech University (11120032662029, 11120132662028)

收稿时间: 2020-05-28; 修改时间: 2020-06-19, 2020-06-28; 采用时间: 2020-07-10; csa 在线出版时间: 2021-01-27

手势特征的选择直接影响到手势识别的准确率。目前常见的手势特征提取方法主要包括两类:基于人工设计的手势特征提取和基于卷积神经网络的手势特征提取。基于人工设计的手势特征提取通常针对特定的数据集来设计有效的手势特征,从而进行手势建模。

Hu 矩特征和傅里叶描述子是常见的图像特征,常用来描述图像的轮廓信息,因其具有平移、旋转和尺度不变性,常用来进行手势建模。Paulraj 等^[6]利用从左右手势图像中提取的 Hu 矩特征建立手势图像识别模型,该方法对手势图像的平均识别率为 92.85%。Fernando 等^[7]提出基于手部高宽度比与 Hu 矩特征融合的手势识别方法,该方法不受图像背景和小范围亮度变化的干扰。李丹娇等^[8]提出一种结合 CSS 和傅里叶描述子的手势特征提取方法。基于人工设计特征的手势特征提取能够提取出具有针对性的手势特征,且对于具有相似背景的手势图像集识别效果较好。但该方法学习能力不强,提取的手势特征不能适应多变的手势背景环境。此外,在样本量不断增大的情况下,识别率不会显著提高。另外,人手是非刚性物体,手势的变换具有多样性和复杂性,人工设计的手势特征往往会忽略手势的细节特征,导致对图像的描述不够完整,从而加大手势识别的难度。

基于卷积神经网络的手势识别方法可以从大量数据中自主学习手势图像的深层次特征。Pigou 等^[9]构建了基于卷积神经网络的手势图像识别系统,该系统对 20 种意大利手势图像的识别准确率为 91.7%。Bobić 等^[10]将提取的手势图像的梯度方向直方图作为网络的输入,利用多层前馈神经网络和反向传播算法进行训练,实现了手势的准确识别。Bheda 等^[11]提出了一种利用深度卷积网络对美国手语字母和数字图像进行分类的方法。Rao 等^[12]对手语视频帧进行预处理、分割和特征提取,形成手语特征空间,并利用前馈网络对提取的特征训练并进行测试,该方法对手势视频的识别率为 90%。Chai 等^[13]利用 Faster R-CNN 对手部区域进行检测和分割,并将手势的彩色图像和深度图像作为卷积神经网络的输入进行训练,有效地提高了手势识别率,但该方法在训练网络时产生的参数量过多,降低了手势识别效率。基于卷积神经网络的手势识别方法虽能够提取深层次的手势图像特征,有效地表征手势图像的全局语义信息,但该方法忽略了底层网络提取的手势局部特征。另外,目标手势在原图中占比较小,手势特征

信息经过层层卷积、池化等操作后,目标手势检测的敏感度有所降低,而为了丰富手势的特征信息,实现较高的识别精度,往往需要大量的数据集和较深的网络层进行训练,消耗了大量的运行时间,难以保证手势识别的实时性。另外,这两类方法的研究手势图像中往往忽略脖子、手臂等干扰情况。

综上所述,单一特征的手势识别在复杂背景下的识别率有待提高,因此,本文提出了基于多特征加权融合的静态手势识别方法。该方法提取分割后的手势图像的傅里叶和 Hu 矩等形状特征,将两者融合作为手势图像的局部特征,利用优化的 VGG 网络和轻量级 AlexNet 网络设计双通道卷积网络模型,用来提取手势图像的深层次特征,最后将分割后的手势图像的局部特征与手势灰度图像的深层次特征进行加权融合,使融合后的特征更有效。手势识别算法流程如图 1 所示。

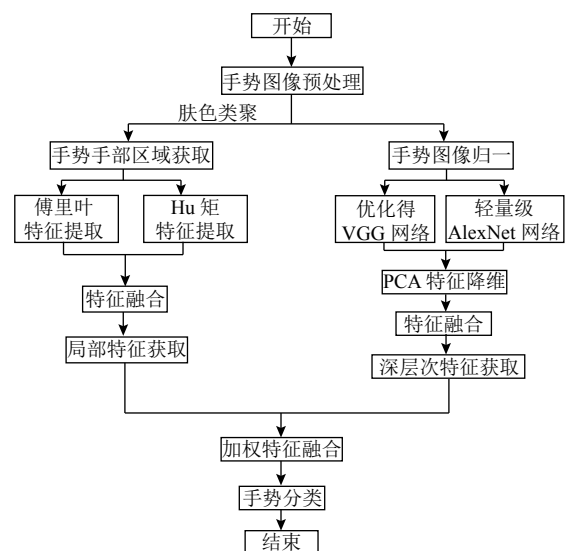


图 1 手势识别算法总体框架

1 局部特征提取

本文采取肤色聚类及腐蚀、膨胀等一系列形态学操作对手部区域进行预处理^[14],得到目标手部区域^[15],预处理流程如图 2 所示。由于不同手势的手部区域面积、手势方向、手部形状等均不相同,因此在手势识别中通过形状特征来描述手势图像的特征具有很大的优势。傅里叶描述子可以提取手部形状信息, Hu 矩则可以描述手部区域面积、重心与对称性等细节特征。因此,将两者融合能够有效表达手势图像的局部特征,更好地区分不同的手势图像。

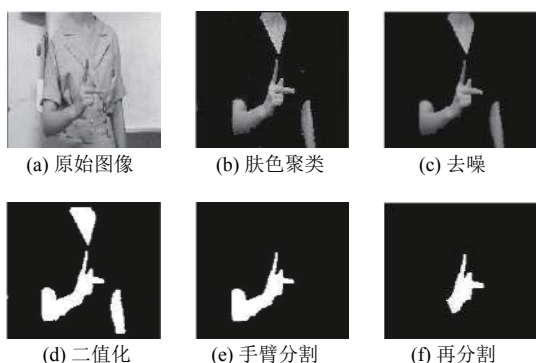


图2 手势预处理

1.1 傅里叶描述子特征

傅里叶描述子 (Fourier Descriptor, FD) 是一种计算简单的手势图像形状和边界特征的表示方法^[16]. 每个傅里叶描述符都具有物理意义, 能够兼顾描述手势的局部特征和全局特征. 因此, 本文选择傅里叶描述子作为手势图像局部特征的描绘子之一. 首先提取手势闭合轮廓信息, 由轮廓图像上各像素点的坐标可得到手势轮廓的傅里叶描述序列, 如式 (1) 所示.

$$T_u = \sum_{k=0}^{n-1} t(k)e^{(2j\pi uk)/n} \quad (1)$$

其中, T_u 为边界的傅里叶描述子, $u = 0, 1, \dots, n-1$. $t(k)$ 为轮廓像素点的复数坐标序列.

为了保证手势图像在空间域中旋转、平移以及尺度变换的稳定性, 需要对傅里叶描述子进行归一化. 归一化傅里叶描述子 d_v 的定义如式 (2) 所示.

$$d_v = \frac{|T_u|}{|T_1|} \quad (2)$$

其中, $u = 2, 3, \dots, n-1$; $v = 1, 2, \dots, n-2$. 将 d_v 写成向量形式, 则为手势轮廓的傅里叶描述子特征向量.

1.2 Hu 矩特征

Hu 矩能够全面地描述手势图像的形状特征, 且不受手势旋转和缩放等因素的影响, 因此本文采用 Hu 矩特征提取手势图像的形状及轮廓特征, 并将其作为手势图像的局部特征.

手势图像的 $(r+s)$ 阶几何矩可以定义为:

$$m_{rs} = \sum_{i=1}^I \sum_{j=1}^J i^r \cdot j^s \cdot V(i, j) \quad (3)$$

其中, $V(i, j)$ 是点 (i, j) 处的灰度值, I 和 J 分别是手势图像的宽度和高度, $r, s = 0, 1, 2, \dots$.

则对应的图像的 $(r+s)$ 阶中心矩定义为:

$$\mu_{rs} = \sum_{i=1}^I \sum_{j=1}^J (i - \bar{x}_0)^r \cdot (j - \bar{y}_0)^s \cdot V(i, j) \quad (4)$$

其中, $r, s = 0, 1, 2, \dots$, $\bar{x}_0 = \frac{m_{10}}{m_{00}}$, $\bar{y}_0 = \frac{m_{01}}{m_{00}}$, (\bar{x}_0, \bar{y}_0) 为手势图像的重心坐标. 则相应图像的归一化中心矩定义为:

$$\eta_{rs} = \frac{\mu_{rs}}{\mu_{00}^t} \quad (5)$$

其中, $t = \frac{r+s+2}{2}$, $r+s = 2, 3, \dots$.

Hu 矩特征包含手部区域的重心、手部区域的面积以及对称性等局部特征, 通过手势图像的二阶和三阶归一化中心矩的线性组合可以构造 7 个不变矩, 具体计算公式如下:

$$\begin{cases} \phi_1 = \eta_{02} + \eta_{20} \\ \phi_2 = (\eta_{02} - \eta_{20})^2 + 4\eta_{11}^2 \\ \phi_3 = (\eta_{30} - \eta_{03})^2 + (3\eta_{21} - \eta_{03})^2 \\ \phi_4 = (\eta_{21} + \eta_{03})^2 + (\eta_{12} + \eta_{30})^2 \\ \phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) \left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \right] \\ \quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right] \\ \phi_6 = (\eta_{20} - \eta_{02}) \left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right] \\ \quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ \phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) \left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2 \right] \\ \quad + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{30}) \left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 \right] \end{cases} \quad (6)$$

根据 7 个不变矩的具体计算公式对手势图像进行特征提取, 形成的特征向量为 $H = (\phi_1, \phi_2, \phi_3, \phi_4, \phi_5, \phi_6, \phi_7)$.

2 深层次特征提取及加权特征融合

使用 Hu 矩和傅里叶描述子不能完整的描述手势图像的局部表征信息, 导致对手势图像的描述不够完整. 另外, 使用通用的特征描述子提取的特征不能适应多变的手势背景环境. 此外, 在样本量不断增大的情况下, 人工设计手势特征的方法在特征学习方面和识别率方面表现出了局限性. 目前, 卷积神经网络 (Convolutional Neural Networks, CNN) 已经成为模式识别领域的研究热点之一, 其在特征提取阶段表现出优越性, 因此普遍应用在大规模图像分类中. 然而, 由于目标手势在手势图像中占比较小, 往往需要训练较深的网络层来提高手势识别的精度, 但设计深层网络容易产生冗余参数, 不仅容易产生过拟合问题, 而且影响手势识别的实时性, 而参数量较少的浅层网络不足以满足手势特征表示. 故本文设计了一种新的双通道卷积神经网络特征提取方法来提取手势图像的深层次特征, 以提高手势

识别精度和实时性。

2.1 双通道卷积神经网络

本文采用轻量级 AlexNet 网络和优化的 VGG 网络作为网络模型的子结构来设计双通道卷积神经网络模型,其结构如图 3 所示。

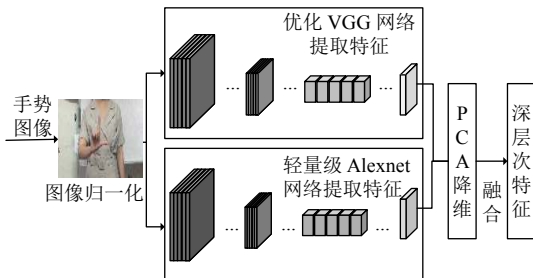


图3 基于双通道卷积神经网络的手势图像深层次特征提取

2.1.1 轻量级 AlexNet 网络

AlexNet 网络是 2012 年提出的一种深度神经网络,其网络结构的结构图如图 2(a) 所示。该网络包含 8 个用于特征学习的网络层,前 5 个网络层是用于手势图像特征提取的卷积层,后 3 个是用于特征整合的全连接层。AlexNet 网络采用随机失活、ReLU 激活函数和截取图像等方法有效地降低了网络的过拟合概率,且通过重叠的最大池化丰富图像特征。综合上述优点,选取 AlexNet 网络作为手势图像的特征提取网络之一,为了降低 AlexNet 网络的参数量,对 AlexNet 网络进行压缩,获得轻量级 AlexNet 网络。参数量对比如表 1 所示。

表 1 AlexNet 网络模型改进及参数量变化

AlexNet网络	参数量	轻量级AlexNet网络	参数量
11×11 Conv, 96, pool/2	34944	11×11 Conv, 96, pool/2	34944
5×5 Conv, 256, pool/2	614656	5×5 Conv, 256, pool/2	614656
3×3 Conv, 384	885120	3×3 Conv, 384	885120
3×3 Conv, 384	1327488	3×3 Conv, 48	165936
3×3 Conv, 256, pool/2	884992	3×3 Conv, 32, pool/2	13856
Fc, 4096	37752832		
Fc, 4096	16781312	Fc, 128	147456
Fc, 4096	4097000		

如表 1 所示,轻量级 AlexNet 网络相比于 AlexNet 网络,Conv4 和 Conv5 的卷积核个数分别变为 48 和 32,全连接层的神经元个数变为 128,大大减少了参数量。

2.1.2 优化的 VGG 网络

VGG 是 AlexNet 网络的进阶版,共有 A、B、C、D 和 E 等 5 种网络结构^[17],本文采用性能最好的 VGG16 网络作为手势图像特征提取网络之一,该网络包含 5 个块结构,共有 13 个卷积层,3 个全连接层,且卷积

层之间用池化层分开。VGG 网络采用多个 3×3 卷积核,模仿出了更大的感受野,并减少了手势识别模型训练产生的参数量。故本文选取 VGG 网络作为手势图像的特征提取网络之一并对其进行优化,以提高网络模型的性能和稳定性。优化后的 VGG 网络块结构如图 4 所示,将批量归一化 (Batch Normalization, BN) 层加到网络模型中,在特征提取网络中的每一层卷积层后加入 BN 层,使得下一层卷积层的输入特征与当前层的特征服从相同的正态分布,实现手势图像的准确分类。

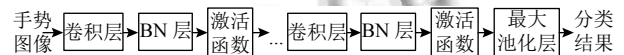


图4 VGG 网络结构图

2.2 PCA 降维

考虑到双通道卷积神经网络生成的特征维度较大,易造成特征表示的繁冗以及计算效率的降低,因此,需要对提取的特征进行有效的降维。使用 PCA 网络进行主成分分析是对多维数据进行统计分析的方法之一^[18]。利用 PCA 变换不仅减少了输入信息量,而且将输入集转换为的一组有序的独立成分。因此,本文选用 PCA 方法对双通道卷积神经网络生成的高维特征进行降维。PCA 降维算法如算法 1。

算法 1. PCA 降维算法

- 1 给定输入向量 $X=(x_1, x_2, \dots, x_m)$, 计算其协方差矩阵 $C=\frac{1}{m}XX^T$ 。
- 2 求出协方差矩阵的特征值及对应的特征向量。
- 3 将特征值按照重要性排序 $K_1 > K_2 > \dots > K_m > 0$ 。
- 4 将得到的特征向量按对应特征值大小从上到下按行排列成矩阵,取前 k 行组成矩阵 P 。
- 5 $Z=PX$ 即为降维到 k 维后的数据。

2.3 加权特征融合

由于手形的复杂性,只采取单一手势特征无法实现手势精准分类,因此本文将提取的手势局部特征和降维后的深层次特征进行融合。由于双通道卷积神经网络得到的深层次手势特征识别结果优于局部特征,因此将提取的手势局部特征与手势深层次特征进行权重分配,以得到较高的识别率。

3 数据集制作与实验结果分析

3.1 手势图像数据集

实验数据集共包含 26 种手势,由 30 位手语者采集而成,单张手势图像存在脖子、干扰手臂和复杂的背景区域。对于每位手语者,每种手势采集 50 张手势

图像,包括不同背景、不同方向、不同光照的手势图像,该数据集的手势图像总数为 39 000(=30×50×26)张。

本实验将手势图像按照 3:1 的比例设置训练集和测试集。部分手势图如图 5 所示。



图 5 部分手势图像数据集

3.2 实验结果分析

3.2.1 批处理大小及学习率的设置

为了选择最优的 $batch_size$, 本文对批处理大小 ($batch_size$) 进行对比实验, 实验结果如图 6 所示。由图 6 可知, 当 $batch_size$ 为 32 时, 损失值波动幅度大且准确率低, 而 $batch_size$ 64 和 128 时, 损失值波动均较小, 准确率提高均较快, 且在一定迭代次数后 $batch_size$ 为 64 准确率与 $batch_size$ 为 128 的准确率基本相同。故在兼顾速度和准确率两方面因素, 本文选用 $batch_size$ 为 64。

为了防止手势图像识别模型出现过拟合问题, 本文网络的最大迭代次数设为 60 000 次, 通过指数衰减策略来更新训练过程中的学习率, 学习率更新公式如式 (10)。

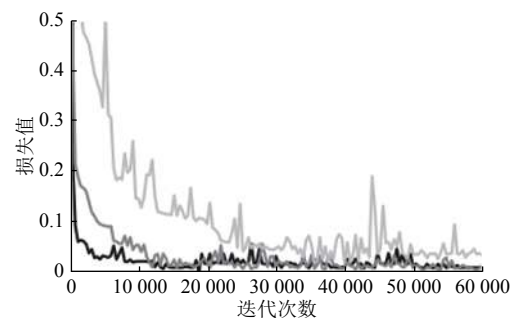
$$learn_rate = lr \times decay_rate^{\frac{global_steps}{decay_steps}} \quad (7)$$

其中, $global_steps$ 为当前的迭代次数, lr 为初始学习率, 初值为 0.001, $decay_rate$ 为学习率衰减系数, 初值为 0.9, $decay_steps$ 为学习率衰减速度, 初值为 5000。

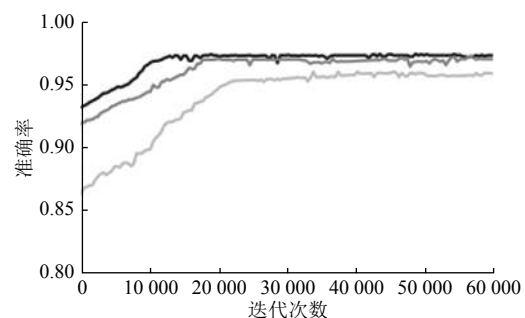
3.2.2 网络结构对比分析

为了验证本文所构建的双通道卷积神经网络提取手势图像的可行性, 对自建数据库中手势图像构建四组网络模型进行实验研究, 分别为: VGG 网络模型、AlexNet 网络模型、基于 VGG 和 AlexNet 的双通道卷积神经网络模型、基于优化的 VGG 和轻量级 AlexNet

的双通道卷积神经网络模型。实验结果如图 7 所示。本文提出的双通道卷积神经网络模型减少了模型参数量, 同时有效提高手势图像识别准确率。



(a) 损失值随迭代次数的变化



(b) 准确率随迭代次数的变化

— $batch_size=128$ — $batch_size=64$ — $batch_size=32$

图 6 使用不同 $batch_size$ 训练网络的实验结果比较

为了进一步验证本文提出方法的有效性, 本文在 ASL 公共手势图像数据集上进行了实验研究, 得到识

别准确率为 96.42%。与 MPC 方法^[19]、深度学习的 CNN 方法、DBN 方法^[20] 和 HSF-RDF^[21] 相比, 识别准确率有明显提高, 结果如表 2 所示。由此充分验证了提出方法的有效性。

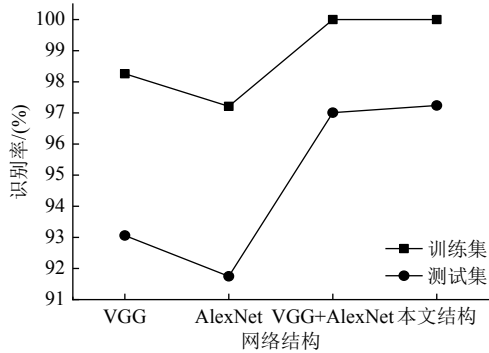


图7 不同网络结构对手势图像的识别效果

表2 不同方法对不同数据集的识别结果

识别方法	MPC	CNN	DBN	HSF-RDF	本文方法
识别率(%)	90.19	88.22	96.14	75	96.42

3.2.3 VGG 网络优化

为了提高网络稳定性, 本文在 VGG 网络模型中增加 BN 层, 其变化结果如图 8 所示。由图 8 可知: BN 层对网络损失的稳定性具有重要作用, 添加 BN 层后随着迭代次数增加损失值迅速下降且稳定, 同时准确率得到明显提高。

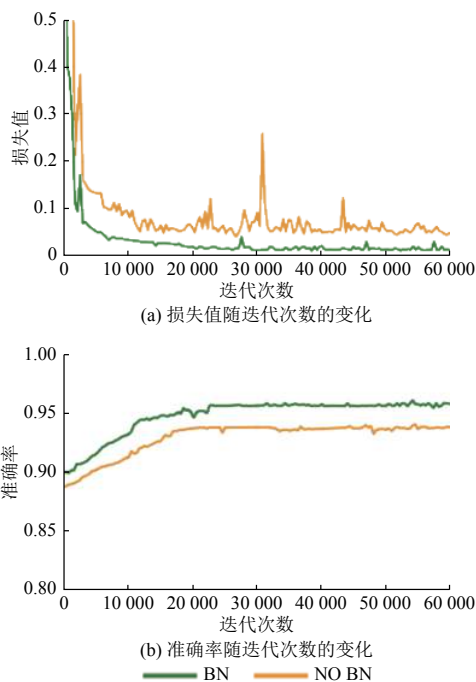


图8 添加BN层的网络模型对比

3.2.4 特征对比实验

基于双通道卷积神经网络的特征提取方式虽然能提取手势灰度图像的深层次特征, 但会引起前层网络提取的手势局部特征丢失。而手势局部特征在相似背景下手势图像识别中起到了非常关键的作用^[22]。本文基于傅里叶描述子 (Fourier Descriptor, FD)^[16] 和具有旋 Hu 矩特征, 构建手势识别方法进行对比实验。本文分别提取原始手势图像的傅里叶描述子、Hu 矩和傅里叶描述子与 Hu 矩融合作为手势图像特征输入到双通道卷积神经网络进行分类识别, 如图 9 所示。由图 9 可知, 本文方法比其他方法有较高准确率, 验证了本文提出方法的有效性。

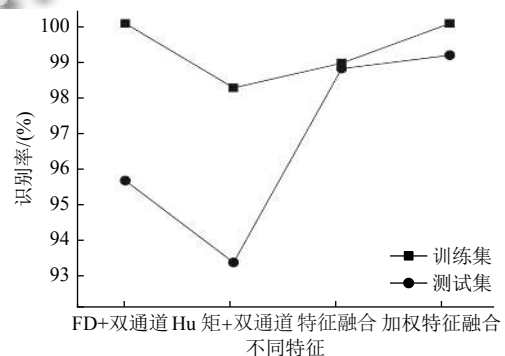


图9 不同特征组合的实验结果

3.2.5 不同权重比的设置

通过对 ASL 公共手势图像数据集进行对比实验, 本文设计的双通道网络的识别准确率已达 96.4%。但由于自建数据库中每张手势图像均存在脖子、干扰手臂等复杂背景, 现有方法的识别准确率不高。故本文提出融合局部特征和深层特征方法进行手势识别提高复杂背景下手势图像识别率。为了验证局部特征和深层次特征在手势特征表示中所起作用, 赋予手势局部特征不同的权重对图 5 所示数据库进行实验。通过调整局部特征与深层次特征所占权重比不同, 统计手势识别率, 结果如图 10 所示。由图 10 可知: 当手势局部特征所占权重因子为 0.3 时, 手势识别率是一个拐点, 在其之前手势识别率随着局部特征权重比增加而提高, 在其之后手势识别率随着局部特征权重比增加而降低, 故由实验结果可知手势局部特征所占权重因子为 0.3 最佳。

3.2.6 实验验证

为了验证本文提出方法对自建数据库中手势图像

的有效性,对自建数据库进行实验测试.与现有手势数据库(ASL)相比,自建数据集共包含26种英文字母手势,且单张手势图像存在脖子、干扰手臂和复杂背景区域等,故现有方法难以实现对自建数据库的准确识别.为了减少手势图像中脖子、干扰手臂和复杂的背景区域等干扰信息,本文设计基于手臂分割获取局部特征方法,解决深层次特征在卷积过程中局部信息丢失问题.本文提出方法对图5中每个手势图像分别提取FD特征和Hu矩特征两局部特征,基于双通道卷积神经网络提取手势图像的深层次特征;然后按照3:10的权重将局部特征与深层次特征进行融合,并将融合的特征作为Softmax分类器的输入训练静态手势图像的识别模型进行识别,每种字母手势对应的识别率如表3所示.由表3可知,对自建数据库的26个复杂背景下的静态手势图像的识别准确率均大于96%,识别准确率高,26个静态手势图像的平均识别率为99.13%.

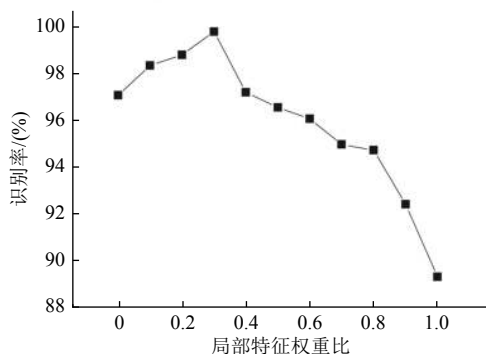


图10 不同局部特征权重比下的手势识别率

表3 字母手势识别结果

字母	识别率(%)	字母	识别率(%)	字母	识别率(%)
A	100	J	98.91	S	98.93
B	98.21	K	100	T	100
C	99.42	L	98.52	U	97.93
D	98.93	M	96.67	V	98.75
E	97.98	N	96.86	W	98.81
F	99.82	O	96.73	X	100
G	98.85	P	98.87	Y	99.95
H	97.89	Q	98.65	Z	99.53
I	97.63	R	99.79		

4 结论与展望

本文提出了基于多特征加权融合的手势识别方法.该方法使用傅里叶描述子和Hu矩提取分割后的手势图像的形状特征,将融合后的特征作为手势图像的局

部特征.其次,对VGG网络模型和AlexNet网络模型进行优化,并利用优化的VGG网络和轻量级AlexNet网络构建双通道卷积神经网络结构,提取手势图像的深层次特征,采用主成分分析方法对深层次特征进行降维,并将降维后的特征与局部特征进行加权融合,以充分利用手势的各部分特征;最后,采用Softmax分类器对输入的特征进行识别,实验结果表明,提出的方法对手势图像的识别准确率较高.虽然本文方法在进行手势图像识别时取得了较好的效果,但是训练网络模型仍需要较长的学习时间,未来将致力于减少算法运行时间,并将其应用在大型数据集中.

参考文献

- 1 Wang C, Liu Z, Chan SC. Superpixel-based hand gesture recognition with Kinect depth camera. *IEEE Transactions on Multimedia*, 2015, 17(1): 29–39. [doi: 10.1109/TMM.2014.2374357]
- 2 Ren Z, Yuan JS, Meng JJ, *et al.* Robust part-based hand gesture recognition using Kinect sensor. *IEEE Transactions on Multimedia*, 2013, 15(5): 1110–1120. [doi: 10.1109/TMM.2013.2246148]
- 3 Hikawa H, Kaida K. Novel FPGA implementation of hand sign recognition system with SOM-Hebb classifier. *IEEE Transactions on Circuits and Systems for Video Technology*, 2015, 25(1): 153–166. [doi: 10.1109/TCSVT.2014.2335831]
- 4 Neto P, Pereira D, Pires JN, *et al.* Real-time and continuous hand gesture spotting: An approach based on artificial neural networks. *Proceedings of 2013 IEEE International Conference on Robotics and Automation*. Karlsruhe, Germany. 2013. 178–183.
- 5 田秋红, 杨慧敏, 梁庆龙, 等. 视觉动态手势识别综述. *浙江理工大学学报(自然科学版)*, 2020, 43(4): 557–569.
- 6 Paulraj MP, Yaacob S, Bin Zanar Azalan MS, *et al.* A phoneme based sign language recognition system using skin color segmentation. *Proceedings of 2010 6th International Colloquium on Signal Processing & its Applications*. Malacca City, Malaysia. 2010. 1–5.
- 7 Fernando M, Wijjayanayake J. Novel approach to use HU moments with image processing techniques for real time sign language communication. *International Journal of Image Process*, 2015, 9(6): 335–345.
- 8 李丹娇, 彭进业, 冯晓毅, 等. 结合CSS与傅里叶描述子的手势特征提取. *计算机工程*, 2012, 38(6): 178–180. [doi: 10.3969/j.issn.1000-3428.2012.06.058]
- 9 Pigou L, Dieleman S, Kindermans PJ, *et al.* Sign language

- recognition using convolutional neural networks. Proceedings of the European Conference on Computer Vision. Zurich, Switzerland. 2015. 572–578.
- 10 Bobić V, Tadić P, Kvašček G. Hand gesture recognition using neural network based techniques. Proceedings of 2016 13th Symposium on Neural Networks and Applications. Belgrade, Serbia. 2016. 1–4.
 - 11 Bheda V, Radpour D. Using deep convolutional networks for gesture recognition in American sign language. arXiv preprint arXiv: 1710.06836, 2017.
 - 12 Rao GA, Kishore PVV, Kumar DA, *et al.* Neural network classifier for continuous sign language recognition with selfie video. Far East Journal of Electronics and Communications, 2017, 17(1): 49–71. [doi: [10.17654/EC017010049](https://doi.org/10.17654/EC017010049)]
 - 13 Chai XJ, Liu ZP, Yin F, *et al.* Two streams recurrent neural networks for large-scale Continuous gesture recognition. Proceedings of the 23rd International Conference on Pattern Recognition. Cancun, Mexico. 2016. 31–36.
 - 14 包嘉欣. 基于多特征融合的静态手势识别研究 [硕士学位论文]. 浙江理工大学, 2019.
 - 15 武红娇. 复杂背景下的静态手语识别技术研究 [硕士学位论文]. 沈阳: 沈阳工业大学, 2017.
 - 16 Rahmany I, Nemmla MEA, Khelifa N, *et al.* Automatic detection of intracranial aneurysm using LBP and Fourier descriptor in angiographic images. International Journal of Computer Assisted Radiology and Surgery, 2019, 14(8): 1353–1364. [doi: [10.1007/s11548-019-01996-0](https://doi.org/10.1007/s11548-019-01996-0)]
 - 17 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. <https://arxiv.org/abs/1409.1556>. (2014-09-04)[2015-04-10].
 - 18 Mrówczyńska M, Sztubecki J, Greinert A. Compression of results of geodetic displacement measurements using the PCA method and neural networks. Measurement, 2020, 158: 107693. [doi: [10.1016/j.measurement.2020.107693](https://doi.org/10.1016/j.measurement.2020.107693)]
 - 19 Pansare JR, Gawande SH, Ingle M. Real-time static hand gesture recognition for American Sign Language (ASL) in complex background. Journal of Signal and Information Processing, 2012, 3(3): 364–367. [doi: [10.4236/jsip.2012.33047](https://doi.org/10.4236/jsip.2012.33047)]
 - 20 许西丹. 基于耦合深度信念网络的图像识别方法 [硕士学位论文]. 西安: 陕西师范大学, 2018.
 - 21 Pugeault N, Bowden R. Spelling it out: Real-time ASL fingerspelling recognition. Proceedings of 2011 IEEE International Conference on Computer Vision Workshops. Barcelona, Spain. 2011. 1114–1119.
 - 22 Guo S, Huang WL, Wang LM, *et al.* Locally supervised deep hybrid model for scene recognition. IEEE Transactions on Image Processing, 2017, 26(2): 808–820. [doi: [10.1109/TIP.2016.2629443](https://doi.org/10.1109/TIP.2016.2629443)]