文章编号：2096-1618(2019)06-0606-04

基于神经网络的静态手势识别算法实现

包兆华， 高瑜翔， 夏朝禹， 郭春妮
(成都信息工程大学通信工程学院 微电子学院，四川 成都 610225)

摘要：随着物联网技术的发展，手势识别在当今的人机交互中起着至关重要的作用。针对复杂背景下手势识别率低、算法鲁棒性差的问题，提出了一种基于神经网络手势识别方法对26个英文字母实现静态手势识别，该算法由手势检测和特征提取及识别3部分构成。在手势检测部分，解决手势区域提取困难的问题；在手势特征提取部分，通过肤色检测提取出手的轮廓信息的二值图像；在识别阶段，使用从LeNet-5改进的CNN来识别手势。在自己制作的数据集下对神经网络进行训练，最终获得较高的识别率；并在NUS-II和Marcel两个复杂背景的公共数据集上进行了验证实验，识别率分别达到95.31%和98.10%。结果表明，该方法可以在复杂环境下对手势进行精确识别具有较高的稳定性。

关 键 词：手势识别；神经网络；腐蚀；膨胀；特征提取
中图分类号：TP391.4 **文献标志码**：A
doi：10.16836/j.cnki.jcuit.2019.06.008

0 引言

手是人与世界交互的主要工具之一，手势识别在当今的人机交互中起着至关重要的作用，不仅如此，还可以运用于听力丧失或受损人群间的交流。因此手势识别技术对科学技术的发展乃至人类发展的重要性不言而喻。

目前主要有传统方法和神经网络的方法运用于神经网络的识别。传统方法有基于模板匹配、基于数据手套、基于隐马尔科夫模型、基于贝叶斯模型等。Liu等[1]提出了一种基于多特征融合和模板匹配的手势识别方法，Dai等[2]提出了一种基于隐马尔科夫模型的手势识别方法。以上方法达到了很好的效果，但实时性和准确率都有待提高。Lu等[3]提出了一种基于数据手套的手势识别方法，但是数据手套需要使用数据手套对数据进行采集，最终的识别率很容易受到设备的影响。Pisharady等[4]利用贝叶斯注意力模型进行手势检测，再利用支持向量机(SVM)完成手势识别，该模型计算过于复杂，实用性较差。神经网络的方法有Sangi等[5]通过提取梯度直方图(HOG)来描述手的特征并用神经网络判别手势，但如果图像有人脸等肤色背景时，梯度直方图特征会受到较大的影响，导致算法识别率较低。王龙等[6]利用肤色模型进行手势检测，但由于手部肤色和脸部区域肤色相似，故不能有效地分离出手部区域。文献[7]利用神经网络对原始图像直接进行手势识别，当图像中存在人体肤色等复杂背景干扰时，神经网络不能学习到有效的识别模式，识别率显著下降。R-CNN[8]、Faster R-CNN[9]、YOLO[10]、SSD[11]等大型网络框架可以很好地实现手势的识别，但大型网络计算量太大不适合运用手势识别。以上方法存在识别率低，实时性不满足要求，算法复杂度大的问题。因此通过改进网络结构，提出了基于联合小型卷积神经网络手势识别算法以解决上述问题。

收稿日期：2019-06-09

1 算法

1.1 数据集的制作及准备

为了更加贴近于实际应用，使训练的网络拥有更强的有泛化能力，采集多个人多个场景的26个英文字母，数据集中共有26000张图片，每个英文字母各有1000张。由于采集到的图片大小不一，为了使数据集便于训练，将图片统一调整为224×224的彩色图片。将数据集中的图片分为20000张训练集，4000张验证集和2000张测试集，选取方式是从26000张图片中随机选取。将其分为3个部分使训练出的模型拥有更好的性能。数据集中的图片部分如图1所示。

图1 数据集中的部分图片

1.2 整体系统框架

如图2所示，系统整体框架分为手部区域分割，手部区域处理，手部识别3个部分。手部区域分割主要作用是将手部区域从背景中分割出来，该部分的性能很大一部分影响系统的主要性能；手部区域处理为减少手部

识别网络的深度手动提取的手部轮廓特征,这样精简了后续的网络,同时减小了训练的时长;手部识别通过手动提取的特征识别出是何种手势。

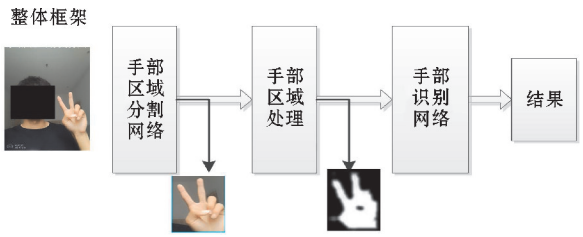


图 2　系统框架

1.3　手部区域分割网络

手部区域分割网络是将手部区域从复杂的背景中分割出来,目前有不同的方法运用于手部区域分割,如 R-CNN、Faster R-CNN、YOLO、SSD 等网络可以很好地实现手部区域的分割。但是它们是通用的物体检测框架,如果运用于手部区域检测,虽然可以实现其功能,但是计算量太大不适合运用于手势分割网络;借鉴 Convolutional Pose Machines[12] 一文中使用神经网络提取人体的关键部位,文中使用该神经网络来进行手部关键点的提取,通过对关键点提取从而判别出手部的位置,如图 3 所示。图 3 左是通过神经网络提取出手部区域的关键点,图 3 右是通过手部区域关键点的提取出完整的手部区域。

图 3　手部区域分割

1.4　手部区域分割网络

手部区域处理部分是通过将手部区域分割出来的手部区域人工提取特征。因为手势识别部分卷积神经网络的卷积层的作用是提取出图片数据中手势的特征,如果在将图片送入神经网络判别之前对图片进行处理以便使得神经网络的特征提取需要较少的卷积核,从而达到提高识别率的同时优化网络结构。手部区域特征提取的流程如图 4 所示。

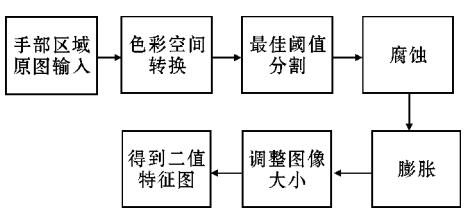


图 4　特征提取流程图

在 Ycbcr 颜色空间中进行手势分割是由于在 Ycbcr 颜色空间肤色范围紧密,受光照干扰较小且肤色在 YCbcr 颜色空间内不是随机分布,而是分布在固定的范围内且范围较小,皮肤的 Cr 分量分布在 133 ~ 173,Cb 分量大约在 77 ~ 127。根据 Cr 和 Cb 两个分量分布范围可以分割手部区域,如图 5(b)所示。分割后的图像依然会受到噪声等其他因素的影响,因此在分割后对图像分别进行腐蚀膨胀算法去除其他因素的影响,如图 5(d)所示。最后将图像调整为 28×28 大小的二值图像,一是因为这是人为采集的手部特征,只需要将手部的轮廓采集到就满足要求了,如果过大会增大计算量并且对最终的计算结果没有影响;二是因为手部识别网络的输入图像为 28×28 大小的,图像处理的情况如图 5 所示。

(a)手部图像　(b)最佳阈值分割　(c)腐蚀膨胀　(d)大小调整

图 5　手部区域处理过程

1.5　手部识别网络

使用修改神经网络 LeNet-5[13] 来识别手势。修改 LeNet-5 的结构如图 6 所示,CNN 以 28×28 大小的二进制图像为输入。卷积神经网络的主要组成:(1)卷积层:卷积运算的目的是提取输入图像的不同特征,第 1 层卷积层可能只能提取一些低级的特征,如边缘、线条和角等特征,第 2 层的卷积网络能从低级特征中提取更复杂的高级特征。(2)池化层:实际上是一种形式的向下采样。有多种不同形式的非线性池化函数,而其中最大池化和平均采样最为常见,相当于把一张分辨率较高的图片转化为分辨率较低的图片,可进一步缩小最后全连接层中节点的个数,从而达到减少整个神经网络中参数的目的。(3)全连接层:一般都在最后几层,负责根据卷积提取的特征来判别具体的手势。(4)最后 1 层 Softmax 层负责最终判别手势的类别,文中需要判别 26 个手势,因此共有 26 个神经元。改进的 LenNet-5 共有 7 层(不包括输入层),每层都包

含不同数量的训练参数,如图 6 所示。

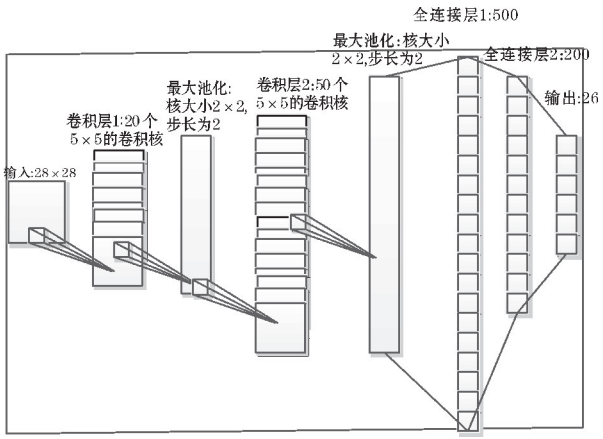


图 6 改进 LeNet-5 结构图

2 神经网络的训练及最终的结果

2.1 神经网络的训练及在自制数据集下的准确率

数据集中准备多个人多个场景的 26 个英文字母,共有 26000 张图片,每个英文字母各有 1000 张。将采集到的图片经过手部区域分割网络和手部区域处理后得到 26000 张 28×28 大小的二值图像。根据手势的类型使用不同的数字对其进行标记,如图 7 所示。

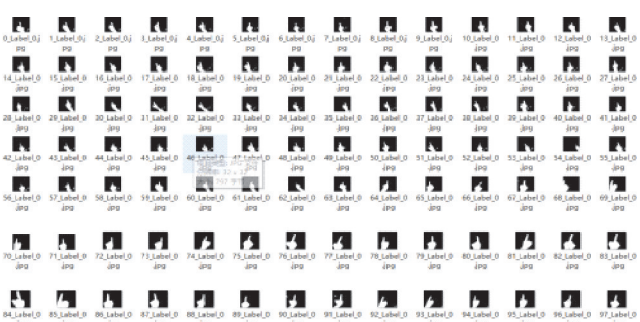
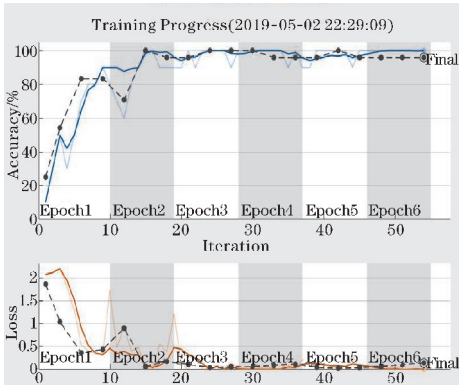


图 7 标记好的部分数据集数据

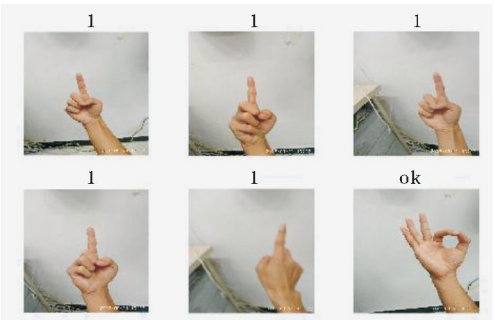
将标记好的图片从 26000 张图片中随机选取 20000 张送入卷积神经网络中进行训练。最终训练集准确率达到 98%,测试集准确率达到 95%,见图 8。

2.2 实验结果与分析

文中实验环境为一台 CPU 为 Intel(R) Core(TM) i5-3230M CPU @2.60 GHz、内存为8 GB的计算机,实验软件为 Matlab,除了在自制的数据集下进行实验,还在 Marcel 和 NUS-II 两种复杂背景手势数据集进行实验(表 1、表 2)。对算法的时间复杂度进行分析(表 3),并在计算机上测试单个算法执行需要的时间(表 4),文中算法执行时间最少。结合识别率和时间复杂度及单机执行的时间可以看出,文中算法有较高的效率。



(a)训练过程



(b)部分测试图片结果

图 8 最终结果

表 1 Marcel 数据集不同算法的识别率

方法	识别率/%
肤色+ CNN[6]	97.20
CNN+ SVM[13]	98.00
高斯模型[14]	94.00
文中算法	98.10

表 2 NUS-II 数据集不同算法的识别率

方法	识别率/%
肤色+CNN[6]	92.00
CNN+ SVM[13]	93.01
高斯模型[14]	89.12
文中算法	95.31

表 3 不同算法的时间复杂度

方法	时间复杂度
肤色+CNN[6]	$O(n^2)$
CNN+ SVM[13]	$O(n^3)$
高斯模型[14]	$O(n^2)$
文中算法	$O(n)$

表 4 单机执行时间

方法	时间/s
肤色+CNN[6]	0.09
CNN+ SVM[13]	0.19
高斯模型[14]	0.075
文中算法	0.04

3　结束语

针对应用中复杂背景严重影响手势识别性能的问题及现有的网络过于庞大的问题,提出一种将两个小型神经网络和人工提取图像特征结合的算法。在 Matlab 对神经网络进行训练,最终在自制数据集和 Marcel 及 NUS-II 数据集上获得较高的识别率;实验结果表明,方法在普通环境和复杂环境都能完成对静态手势的精确识别且相比于其他算法有较高的实时性。

参考文献:

[1] Liu Y, Zhang L, Zhang S. A hand gesture recognition method based on multi-feature fusion and template matching [J]. Procedia Engineering, 2012, 29(4):1678-1684.

[2] Dai Y K, Zhou Z H, Chen X, et al. A novel method for simultaneous gesture segmentation an recognition based on HMM[C]. Proceedings of the 2017 International Symposium on Intelligent Signal Processing and Communication Systems. Piscataway, NJ: IEEE, November 6-9, 2017:684-688.

[3] Lü N, Yang Y J, Xu T. Sparse decomposition for data glove gesture recognition[C]. Proceedings of the 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). Piscataway, NJ: IEEE, 2017:1-5.

[4] 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.

[5] Sangi P, Matilainen M, Silven O. Rotation tolerant hand pose recognition using aggregation of gradient orientations[M]. Lecture Notes in Computer Science, Berlin, Germany: Springer, 2016:257-267.

[6] 王龙,刘辉,王彬,等. 结合肤色模型和卷积神经网络的手势识别方法[J]. 计算机工程与应用, 2017,53(6):209-214.

[7] Mohanty A, Rambhatla S S, Sahay R R. Deep gesture: Static hand gesture recognition using CNN [M]. Advances in Intelligent Systems and Computing, Berlin, Germany: Springer, 2017:449-461.

[8] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014:580-587.

[9] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]. International Conference on Neural Information Processing Systems. [S. l.]: MIT Press, 2015:91-99.

[10] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016:779-788.

[11] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]. European Conference on Computer Vision. Springer International Publishing, 2016:21-37.

[12] Wei, Shih-En, et al. Convolutional Pose Machines [C]. The Robotics Institute Carnegie Mellon University. CVPR, 2016.

[13] Y. LeCun, L. Bottou L, Y. Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[14] 吴晴. 基于改进的 CNN 和 SVM 手势识别算法研究[D]. 南昌:江西农业大学, 2018.

[15] Jia J, Jiang J M, Wang D. Recognition of hand gesture based on Gaussian mixture model[C]. International Workshop on Content-Based Multimedia Indexing. Piscataway, USA: IEEE, 2008: 353-356.

Implementation of Static Gesture Recognition Algorithm based on Neural Network

BAO Zhaohua,　GAO Yuxiang,　XIA Chaoyu,　GUO Chunni

(College of Communication Engineering College of Microelectronics, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: With the development of Internet of things technology, gesture recognition plays a vital role in today's human-computer interaction. Aiming at the problem of low recognition rate and poor algorithm robustness in complex background, this paper proposes a gesture recognition method based on neural network to achieve static gesture recognition of 26 English letters. The algorithm consists of gesture detection and feature extraction and recognition. In the gesture detection section, the problem of difficulty in extracting the gesture area is solved. In the gesture feature extraction section, the binary image of the contour information of the hand is extracted by the skin color detection. In the recognition phase, the gesture is recognized using the CNN improved from LeNet-5. The neural network was trained under the data set produced by ourselves, and finally the comparatively higher recognition rate was obtained. The verification experiments were carried out on common datasets of NUS-II and Marcel, which have complex background, and the recognition rates reached 95.31% and 98.10% respectively. The results show that the method can achieve high stability in the accurate recognition of gestures in complex environments.

Keywords: gesture recognition; neural network; corrosion; expansion; feature extraction