

文章编号: 2095-2163(2021)06-0157-04

中图分类号: TP391.41;TP183

文献标志码: A

基于视觉的手势识别算法及应用的研究

黄彦铭, 宁媛

(贵州大学 电气工程学院, 贵阳 550025)

摘要: 随着计算机快速的发展以及广泛的应用,人机交互技术已成为人们日常生活中不可或缺的部分,而手势识别则顺应了发展需要。原始手势图像需要做一些预处理操作才能更好的进行下一步操作分析。本文主要针对基于视觉的手势图像,设计了 Alex CNN 卷积神经网络模型,研究了图像与处理算法,包括数据增广、肤色提取、形态学处理等操作,对所设计的神经网络模型准确率进行测试。结果表明用经过图像预处理后的图像所训练得到的模型准确率相比用原始图像进行训练所得到的模型准确率大为提升。

关键词: 手势图像; 图像预处理; 卷积神经网络

Research on gesture recognition algorithm and application based on vision

HUANG Yanming, NING Yuan

(The Electrical Engineering College, Guizhou University, Guiyang 550025, China)

[Abstract] With the wide application of computer, human-computer interaction has become an important part of people's daily life, and gesture recognition meets the development needs. The original gesture image needs to do some preprocessing operations in order to better analyze the next operation. In this paper, Alex CNN is designed for gesture images based on vision Convolution neural network model, studied the image and processing algorithm, including data augmentation, skin color extraction, morphological processing and other operations, the impact on the accuracy of the designed neural network model. The results show that the accuracy of the model trained by the image preprocessing is greatly improved compared with the accuracy of the model trained by the original image.

[Key words] gesture image; image preprocessing; convolutional neural network

0 引言

人的手势自身就具有广泛的多义性、多样性、多元化以及空间和时间上的差异性等多种独特的特点,再加上人手形体的复杂性以及人体视觉的不确定性,所以基于视觉的手势识别成为了多学科、多领域交叉的研究难题。

手势识别主要是利用机器设备对人的手、胳膊等各关节的空间角度位置进行检测。这些机器设备大多是通过有线通信技术将计算机系统与人体互相结合连接,让使用者的手势信息准确地传送到识别系统内,其中典型设备以数据手套等机器设备为代表^[1]。数据手套是由各种传感器组成的,使用者利用这些传感器将自身手的位置、角度以及手指的角度等信息录入到计算机系统内,达到手势识别的目的。尽管数据手套能够对人手识别达到良好的识别检测效果,但由于其价格昂贵,难以普遍应用。在此之后,光学标记方法的出现淘汰了数据手套,使用者

的手上戴上光学标记,通过红外传感器将人手位置、形态等变化传送到系统内,达到良好的识别效果,但此方法约束了人手的活动范围,也避免不了复杂的设备使用^[2]。

虽然引入外部设备能够提升手势识别的快速性、稳定性以及准确性,但却会掩盖手势的自然表达方式,于是基于视觉的手势识别方式应运而生^[3]。

1 总体设计方案

本研究的设计流程如图 1 所示。利用 OpenCV 调用电脑摄像头采集人体手势图像,建立手势库,以此为基础,对手势库中所采集到的手势图像进行预处理,将图像中的手势部分提取出来,分为测试集和训练集 2 部分。其中,训练集用于导入所设计的 CNN 卷积神经网络进行模型训练,测试集则是对所得到的训练模型进行准确率测试,当测试准确率达标,则输出保存训练模型,以此作为实时手势图像的手势判断模型。

作者简介: 黄彦铭(1996-),男,硕士研究生,主要研究方向:检测技术与自动化装置;宁媛(1968-),女,硕士,教授,主要研究方向:检测技术与自动化装置。

收稿日期: 2021-02-08



图1 总体设计流程

Fig. 1 Overall design

2 图像预处理

2.1 图像预处理概述

图像预处理在数字图像中占有很重要的地位,图像质量的好坏,直接影响分析,例如分类、识别、分割等,图像预处理部分就是基础,也是重中之重。图像预处理即是对手中的原始图像进行加工,处理。在本研究中针对手势图像,采用了数据增广,肤色提取和形态学处理3种方法做图像预处理。

2.2 数据增广

作为深度学习预处理中的常用技巧之一,数据增广的作用是增加数据集的量,增加数据集的多元性以及多样性,使所得到的训练模型具有更强的泛化能力。通过数据增广提升数据集中的相关数据,能防止网络学习到不相关的特征,更多的学到与数据有关的性能,显著的提升整体的性能,目前数据增广主要包括:水平/垂直翻转、旋转、缩放、裁剪、剪切、平移、对比度、色彩抖动、噪声等^[4]。本研究中主要运用图像平移、翻转以及旋转3种常用数据增广方法。

假设图像的原始坐标为 (x_0, y_0) , 平移后的坐标为 (x, y) , 则平移前和平移后的坐标关系如式(1)所示:

$$\begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{1} \end{pmatrix} = \begin{pmatrix} \hat{x}_0 \\ \hat{y}_0 \\ \hat{1} \end{pmatrix} + H \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \quad (1)$$

平移是指所有的像素在 x 和 y 方向各平移相应单位,平移变换对应的数学矩阵为式(2):

$$H = \begin{pmatrix} \hat{x} & 0 & d_x \\ \hat{y} & 0 & d_y \\ \hat{1} & 0 & 1 \end{pmatrix}, \quad (2)$$

图像翻转包括水平翻转和垂直翻转,其中水平翻转对应的数学矩阵如式(3)所示,垂直翻转对应的数学矩阵如式(4)所示:

$$H = \begin{pmatrix} \hat{x} & -1 & 0 & 0 \\ \hat{y} & 0 & 1 & 0 \\ \hat{1} & 0 & 0 & 1 \end{pmatrix}, \quad (3)$$

$$H = \begin{pmatrix} \hat{x} & 1 & 0 & 0 \\ \hat{y} & 0 & -1 & 0 \\ \hat{1} & 0 & 0 & 1 \end{pmatrix}, \quad (4)$$

图像旋转则需要确定旋转角,旋转对应的数学矩阵如(5)所示:

$$H = \begin{pmatrix} \hat{x} \cos \theta & \sin \theta & 0 \\ \hat{y} - \sin \theta & \cos \theta & 0 \\ \hat{1} & 0 & 1 \end{pmatrix}. \quad (5)$$

本文对原始手势图像所做的数据增广示例,如图2所示。

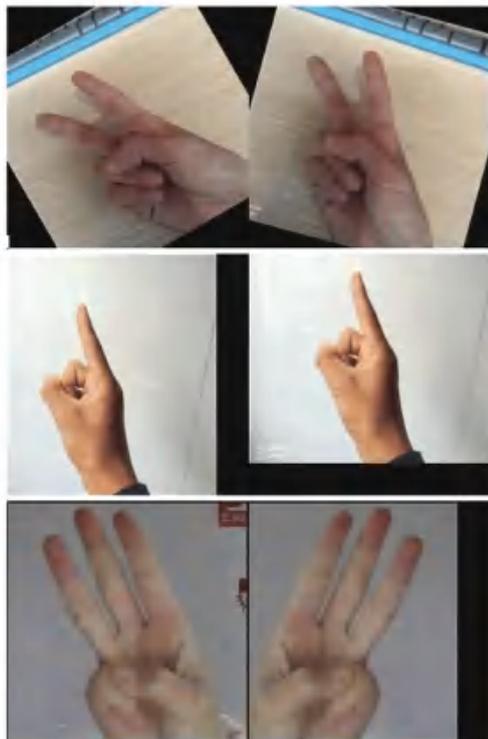


图2 数据增广示例图

Fig. 2 Data augmentation example diagram

2.3 肤色提取

常规下原始手势图像都是基于 RGB(三原色)空间下的图像,这样的图像更便于在硬件中理解与处理。而在机器视觉中,为了提取 RGB 原始手势图像中的手势部分,通常是将 RGB 转换为 HSV 颜色空间,其中 H 代表色度 hue, S 代表饱和度 saturation, V 代表亮度 value, 各个参数的范围见表 1^[5]。因此,在 HSV 颜色空间中能通过对 3 个参数的范围进行一个设定,在此范围内的即为肤色区域,也就是手势区域。而肤色区域的 H 、 S 、 V 3 个参数范围如式(6)所示。

表 1 HSV 参数范围

Tab. 1 HSV parameter range

	黑	灰	白	红	橙	黄	绿	青	蓝	紫	
h_{\min}	0	0	0	0	156	11	26	35	78	100	125
h_{\max}	180	180	180	10	180	25	34	77	99	124	155
s_{\min}	0	0	0	43	43	43	43	43	43	43	43
s_{\max}	255	43	30	255	255	255	255	255	255	255	255
v_{\min}	0	46	221	46	46	46	46	46	46	46	46
v_{\max}	46	220	255	255	255	255	255	255	255	255	255

$$\begin{aligned} 0 &\leq H \leq 20, \\ S &\geq 48, \\ V &\geq 80. \end{aligned} \quad (6)$$

基于 HSV 颜色空间的肤色提取预处理图像示例如图 3 所示。



图 3 基于 HSV 颜色空间的肤色提取示例图

Fig. 3 Skin color extraction based on HSV color space example diagram

2.4 形态学处理

数学形态学中有 2 个基本运算,即腐蚀和膨胀。腐蚀和膨胀是针对图像高亮部分而言,其操作就是将图像或图像的一部分 A ,与卷积核 B 进行卷积后的取值与赋值操作。

图像腐蚀操作就像“领域被蚕食”一样,将图像的高亮部分缩减化,其运行所得到的结果图与原图中的高亮区域相比面积更小,线条更细。其数学表达式(7)如下:

$$A \ominus B = \{x, y \mid (B)_{xy} \subseteq A\}. \quad (7)$$

式(7)表示图像 A 用卷积核 B 来进行腐蚀处理,通过卷积核 B 与图像 A 进行卷积计算,得出 B 覆盖区域的像素最小值,并用这个最小值来替代参考点的像素值,即是求局部最小值操作,这样原图中的高亮区域就会逐渐减小。

图像膨胀操作是腐蚀操作的逆操作,就像“领域的扩张”一样,将图像中的高亮区域扩张,其运行所得到的结果图与原图的高亮区域相比面积更大,线条更粗。其数学表达式(8)如下:

$$A \oplus B = \{x, y \mid (B)_{xy} \cap A \neq \emptyset\}. \quad (8)$$

式(8)表示图像 A 用卷积核 B 来进行膨胀操作,即是求局部最大值,达到增大原图高亮区域的目的。

腐蚀与膨胀两者结合形成了开运算和闭运算。开运算是先腐蚀再膨胀,主要目的为去除孤立的小点;闭运算则是先膨胀再腐蚀,主要目的为填平小孔,弥合小裂缝。本研究中运用了两次开运算,用以加强对图像的处理效果,处理示例如图 4 所示。



图 4 腐蚀膨胀示例图

Fig. 4 Corrosion and expansion example diagram

3 CNN 卷积神经网络模型设计与实验结果

3.1 模型建立

本文采用深度学习中的 CNN 卷积神经网络来对所得到的预处理手势图片进行模型训练,相较于简单的神经网络模型,CNN 卷积神经网络的优势在于不再单一地对每个图片像素做处理,而是对一小块区域的处理。这种做法加强了图像信息的连续性,使得神经网络看到的是一个图像,而非一个点,同时也加深了神经网络对图像的理解,卷积神经网络有一个批量过滤器,通过重复的收集图像的信息,每次收集的信息都是小块像素区域的信息,将信息整理,先得到边缘信息,再用边缘信息总结出更高层的信息结构,得到部分轮廓信息,然后得到完整的图像信息特征,最后将特征输入全连接层进行分类,得到分类结果。本文中卷积神经网络的设计流程如图 5 所示。

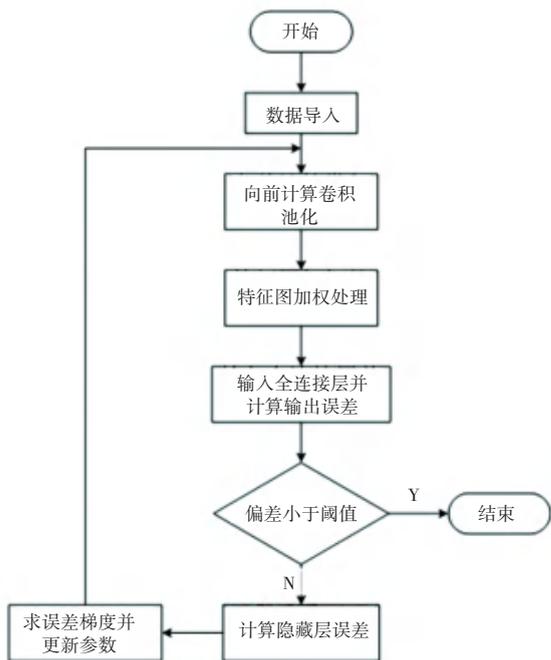


图5 CNN卷积神经网络模型设计流程

Fig. 5 CNN convolutional neural network model design process

3.2 模型训练

使用预处理数据集训练的模型的准确率和损失函数如图6和图7所示。可以看到,随着训练模型步骤的增多,其深度的加深,即采集到的训练集的特征的逐步增多,训练的准确率逐步增高,最后逐渐稳定在0.96~0.97之间,而损失函数也逐步的降低到0.6左右。

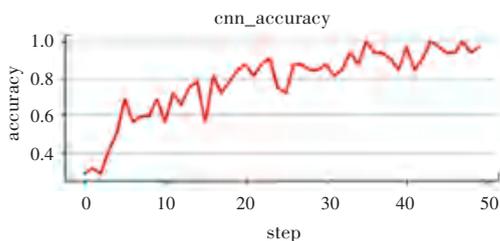


图6 训练集准确率

Fig. 6 Training set accuracy

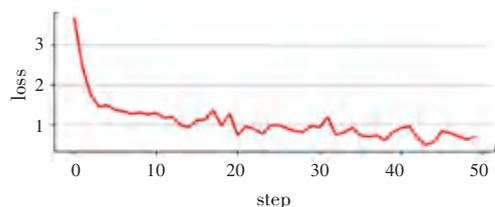


图7 损失函数

Fig. 7 Loss function

3.3 实验结果

为了更好的实现人机交互,本文利用 PyQt5 的 designer 工具生成了人机交互界面,调用电脑摄像头对实时拍摄的手势图像进行识别,测试结果如图8所示。



图8 测试结果图

Fig. 8 Test result chart

4 结束语

本文选用的工具是 Spyder,对一、二、三、四、4种手势图像进行图像增广、肤色提取和腐蚀膨胀等预处理操作,设计了 CNN 卷积神经网络对预处理图像进行分类、验证以及识别,并用 PyQt5 的 designer 工具实现了人机交互界面显示实验结果,得到了较为满意的实验结果,为图像预处理以及人机交互奠定了良好的基础。

参考文献

- [1] 李清水,方志刚,沈模卫,等. 手势识别技术及其在人机交互中的应用[J]. 人类工效学,2002(1):27-29,33.
- [2] 代雨锴. 基于可穿戴数据手套的手势识别方法研究[D]. 广州:华南理工大学,2019.
- [3] 郭志敏. 基于双目视觉的人手定位与手势识别系统研究[D]. 郑州:郑州大学,2011.
- [4] 任海兵,祝远新,徐光祐,等. 复杂背景下的手势分割与识别[J]. 自动化学报,2002(2):256-261.
- [5] 刘俊梅,阮秋琦. 一种复杂背景下的手势分割新方法[J]. 北京电子科技学院学报,2006(2):23-27.