

DOI:10.19651/j.cnki.emt.2313834

# 复杂背景下多特征结合的深度学习手势识别<sup>\*</sup>

赵鸿图 李 豪 梁梦华

(河南理工大学物理与电子信息学院 焦作 454003)

**摘 要:**科学技术的快速发展使得基于深度学习的人机交互已经得到广泛的应用。手势识别作为人机交互领域的重要组成部分,同样具有重要的研究意义和应用价值。通过对传统的手势识别方法进行研究,发现主要是利用肤色检测算法实现手势识别和分类,但是传统方法在针对具有复杂背景的手势图像时会出现识别效果差等问题。为解决这一问题,提出一种基于卷积神经网络的肤色特征和边缘特征结合的手势识别方法。首先,在 YCrCb 颜色空间采用椭圆肤色模型和 Otsu 阈值肤色识别算法获取手势肤色特征,经算法判断后,对手势肤色图像采用改进 Canny 边缘检测算法获得手势边缘特征。其次,提出一种边缘填充方法对手势边缘图像处理,得到手势轮廓完整的手势边缘图像。最终,采用逻辑运算和形态学运算得到手势分割图像,并输入卷积神经网络进行训练和识别。实验结果表明,该方法在复杂背景下具有较好的手势识别效果,在 NUS-II 数据集上的平均识别率为 98.83%。

**关键词:**复杂背景;肤色检测;边缘检测;手势识别;卷积神经网络

**中图分类号:** TP391.41 **文献标识码:** A **国家标准学科分类代码:** 520.2060

## Deep learning gesture recognition based on multi-feature combination in complex background

Zhao Hongtu Li Hao Liang Menghua

(Physics &amp; Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China)

**Abstract:** With the rapid development of science and technology, human-computer interaction based on deep learning has found widespread applications. As an important component of human-computer interaction, gesture recognition holds significant research and application value. However, traditional gesture recognition methods utilizing skin color detection algorithms have limited effectiveness in recognizing gestures against complex backgrounds. To address this problem, a novel gesture recognition method based on convolutional neural network that combines skin color and edge features is proposed. Initially, the ellipse skin color model and Otsu threshold skin color recognition algorithm are used to obtain gesture skin color features in the YCrCb color space. Subsequently, the improved Canny edge detection algorithm is used to obtain the edge features of the gesture skin color images. Following this, the edge filling method is used to process the edge images. Finally, the gesture segmentation images are obtained by logical operation and morphological operation, which are as input to the convolutional neural network for training and recognition. Experimental results demonstrate the effectiveness of approach, with an average recognition rate of 98.83% on the NUS hand posture dataset II. The proposed method shows a significant improvement over traditional gesture recognition methods and can effectively recognize gestures against complex backgrounds.

**Keywords:** complex background; skin color detection; edge detection; gesture recognition; convolutional neural network

## 0 引 言

手势是人际交往中除口语语言外最为常用的交流语言,具有自然性、灵活性、丰富性和通用性等特点,可以直观

地表达人们的情感和内容<sup>[1]</sup>。手势识别作为人机交互领域的重要组成部分,能以更符合人类行为的方式实现更为直观和自然的人与计算机之间的信息交流<sup>[2]</sup>。从国内外对手势识别技术的研究历程来看,研究者在手势识别方法上的

收稿日期:2023-06-06

<sup>\*</sup> 基金项目:河南省科技厅科技攻关和软科学项目(192102310446)、河南省高校基本科研业务费专项资金项目(NSFRF210406)资助

探索一直处于稳步提升的状态,主要研究方法可以分为基于接触式外设穿戴设备的手势识别和基于计算机视觉的手势识别<sup>[3]</sup>。

近几年,深度学习技术的快速发展为手势识别领域提供了新的研究思路,并取得了许多显著的研究成果,基于深度学习的手势识别方法已经成为主流。王龙等<sup>[4]</sup>采用肤色高斯模型来提取图像中手的肤色特征,根据手的肤色和背景颜色的不同实现手势分割,并建立神经网络模型实现手势识别和分类,算法有效避免了人工选取特征的缺点,但是特征的提取较为单一。黄杰<sup>[5]</sup>将手势识别技术应用在手语识别领域,采用基于注意力机制的三维卷积神经网络来提取视频中手势的时空特征,有效提高了大规模手语孤立词的识别精度。Jiang 等<sup>[6]</sup>为了解决复杂背景对于手势识别精度的影响,在采用肤色检测算法分割手势区域的基础上,利用骨架优化算法提取手部关节的空间信息,实现复杂背景下的手势识别,在 ASL 数据集上的识别率为 96.01%。Wu<sup>[7]</sup>提出了一种双通道卷积神经网络模型用于手势识别,该算法是对预处理后的手势图像采用 Canny 边缘检测算法提取手势边缘特征,并将手势图像和边缘图像分别作为神经网络的两个输入通道进行学习和训练,提高了模型的泛化能力,但是仅能处理具有简单背景的手势图像。柏丽银等<sup>[8]</sup>提出一种改进 Canny 边缘检测结合椭圆肤色模型的手势识别算法,可以通过融合边缘特征和肤色特征实现手势分割,针对复杂背景可以较好地分割出手势边缘轮廓,但图像仍会存在多余的背景信息,对识别精度产生影响。杨镛<sup>[9]</sup>利用手部运动产生的肌电信号实现手势识别,其设计的卷积-循环混合神经网络模型的原理是借助卷积网络和循环网络分别提取肌电信号的空间和时间特征,手势识别率得到有效的提升。张晓俊等<sup>[10]</sup>提出的肤色融合算法虽然解决了复杂背景下手势分割效果和识别精度差的问题,但是该算法针对无肤色特征的手势识别具有局限性,这是因为算法是借助肤色在 YCrCb 和 HSV 颜色空间的聚类性才得以实现。徐玥等<sup>[11]</sup>采用基于 YCrCb 颜色空间的 Otsu 肤色识别算法实现手势分割,并通过椭圆傅里叶描述子来拟合手势边缘以提取手势特征,可以基本满足简单背景下的手势识别的性能要求。

上述手势识别方法主要是基于深度学习的手势识别方法,对于手势识别效果比较理想,但各自算法仍存在不足之处。本文的研究目的是解决复杂背景下的手势识别精度低和效果差等问题。研究方法为基于手势肤色特征和手势边缘特征相结合来实现手势分割,并通过网络模型训练,达到较高的手势识别准确率,使其对于复杂背景下的手势识别场景具有一定的社会实用价值。算法的主要创新包括如下 3 个方面:1) 通过分析现有肤色识别算法在获取肤色特征时容易受到光照强度变化和肤色背景等因素影响,因此提出一种基于 YCrCb 颜色空间的肤色融合算法,通过采用椭圆肤色模型和 Otsu 肤色识别算法,获取效果更佳的手势肤

色特征;2) Canny 边缘检测算法是一种常用的提取边缘特征的方法,针对其存在的不足之处,对其改进以实现“保边去噪”,此外提出一种边缘更新算法,得到轮廓封闭的手势边缘图像;3) 利用形态学运算和逻辑运算处理得到的肤色一边缘手势图像,去除图像噪点,可以有效地提高神经网络模型的性能和运行效率。最终实验结果表明,本文算法对复杂背景的手势识别具有较好的识别效果,对复杂背景下的 10 类静态手势的识别效果较为理想。

## 1 复杂背景的手势分割

肤色手势识别方法中,复杂背景往往是影响识别效果的主要原因,如肤色背景和光照变化等<sup>[12]</sup>。因此,复杂背景下的手势识别是本文的重点研究内容,图 1 所示为 NUS-II 数据集的 10 类手势图像。

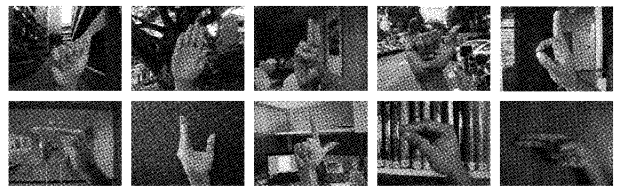


图 1 复杂背景下的手势图像

手势分割算法的工作流程如图 2 所示。首先,利用 Meidiapipe 应用框架中的 Hands 模型<sup>[13]</sup>获取图像中手势的 21 个关键点的坐标值,在得到的坐标值选取最大和最小的横纵坐标值,为获得完整的手势区域,对最大值加 10 和最小值减 10 得到裁切区间,并对图像进行裁切,而针对未能识别出手势的图像和裁切后图像宽或高小于等于 30 的图像,则按照人工设定的值进行裁切并保存,经验证横纵坐标设定的裁切区间分别为 $[15:120]$ 和 $[20:140]$ ;其次,在 YCrCb 颜色空间采用椭圆肤色模型和 Otsu 阈值肤色识别算法对裁切图像提取肤色特征,得到肤色图像;然后,采用改进 Canny 边缘检测算法对肤色图像处理,得到手势边缘图像;随后,采用提出的边缘填充方法对边缘图像更新,得到手势填充图像;最终,采用形态学运算、去除多余轮廓和与运算等方法得到手势分割图像。

### 1.1 基于 YCrCb 颜色空间的肤色融合算法

肤色作为手势的重要特征之一,是手势分割的重要手段。本文在对手势分割时,采用肤色融合算法获取手势区域,但是由于图像多为 RGB 类型,肤色受亮度影响较大导致难以准确识别肤色点,因此本文利用 YCrCb 颜色空间可以将亮度与色度分离开和肤色具有聚类性的特点<sup>[14]</sup>,实现肤色特征的提取。

基于 YCrCb 颜色空间的肤色融合算法是结合椭圆肤色模型和 Otsu 阈值肤色识别算法共同实现。椭圆肤色模型是指肤色信息映射到 YCrCb 颜色空间时,肤色点的红色色度 Cr 和蓝色色度 Cb 分量经过 K-L 变换(Karhunen-Loeve Transform)后在 Cr-Cb 二维空间分布近似为一个椭

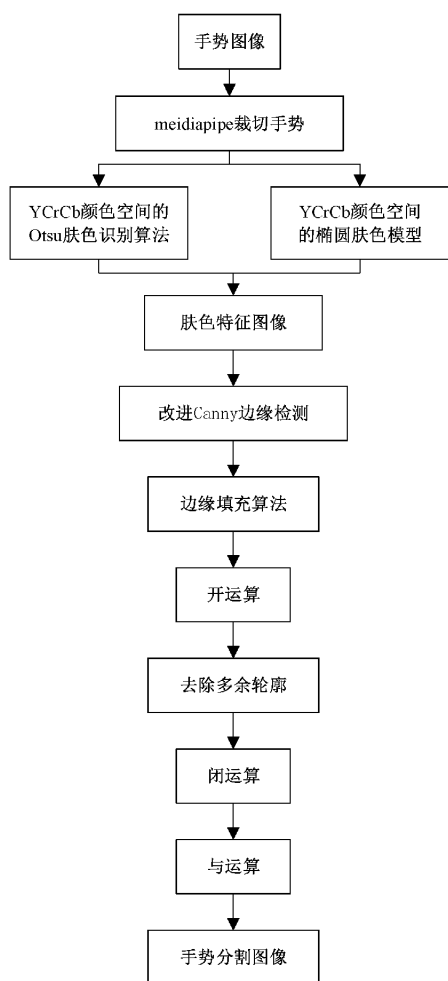


图 2 手势分割算法流程

圆,其计算公式如式(1)所示, $x-y$  坐标空间和  $Cr-Cb$  坐标空间的映射关系如式(2)所示。

$$\frac{(x - ec_x)^2}{a^2} + \frac{(y - ec_y)^2}{b^2} = 1 \quad (1)$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} Cb' - c_x \\ Cr' - c_y \end{bmatrix} \quad (2)$$

式中: $(ec_x, ec_y)$ 和 $(c_x, c_y)$ 分别为椭圆模型在  $x-y$  和  $Cr-Cb$  坐标空间的中心点的坐标, $a$  和  $b$  分别为椭圆的长轴和短轴, $\theta$  表示椭圆与水平方向的夹角, $Cr'$ 和  $Cb'$ 为  $Cr$  和  $Cb$  经过 K-L 变换得到的值。本文中的椭圆模型通过 OpenCV 库的 `ellipse` 函数构建,参数设置如下,中心点坐标设为  $(113, 155)$ ,椭圆的长轴和短轴分别设为 23 和 15,椭圆以中心点顺时针旋转  $43^\circ$ 。Otsu 阈值肤色识别算法同样需要先将图像的颜色空间由 RGB 转换到 YCrCb,然后,对提取的  $Cr$  分量做二值化阈值分割算法处理,得到肤色区域。

肤色融合算法的具体实现方式为分别利用椭圆肤色模型和 Otsu 肤色识别算法得到两类肤色特征图像,对两类图像一一对应采用与运算和或运算实现肤色特征融合,并根据与运算图像和或运算图像的肤色点占比来判断所要保存

的图像,具体判断规则如图 3 所示,其中  $width$  和  $height$  为图像的宽高, $a$  和  $b$  分别表示与运算和或运算图像的肤色点占比。经多次实验验证, $width$  和  $height$  的阈值为 100,这是因为当图像宽高都大于 100 时,说明手势区域在图像中的占比小,背景占比大,此时与运算图像的肤色点占比过小,因此不再作为判断依据。 $a$  的阈值为 0.45,当小于阈值时,说明与运算操作对肤色特征的提取效果差,需要根据或运算图像的肤色点占比进一步判断。 $b$  的阈值在不同条件下分别为 0.5、0.3 和 0.9,当阈值为 0.5 时,用于判断宽高都大于 100 的图像,若值小于 0.5,此时或运算操作可以较好地去除背景;当肤色点占比在 0.3 和 0.9 之间时,说明或运算图像效果更好,否则保存原图像或与运算图像。

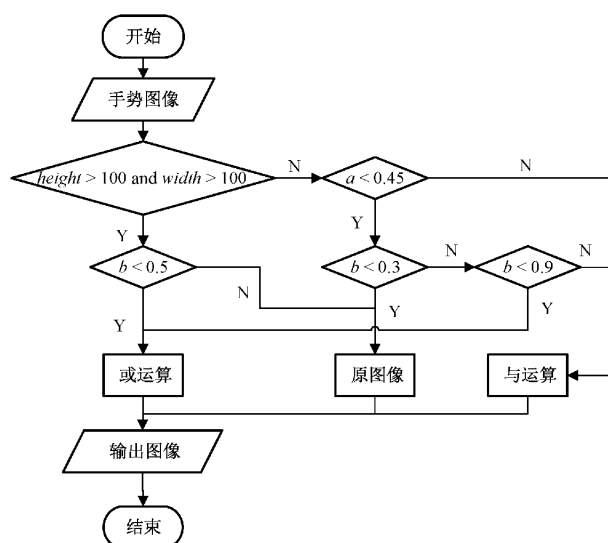


图 3 算法判断流程

肤色融合算法识别效果如图 4 所示,示例图像从左至右依次为原图像、与运算图像和或运算图像,因为原图像经过两种肤色算法运算后得到的图像为二值图,因而两张二值图像进行与运算和或运算操作后仍会得到二值图,不方便观察肤色特征的提取效果,因此,分别将与运算图像(二值图)和或运算图像(二值图)分别和原图像进行与运算操作得到新的与运算图像和或运算图像,此外,所选示例图像的宽高都不满足同时大于 100,此处无特殊用意。图 4(a)所示,与运算图像的肤色点占比小于 0.45,但或运算图像的肤色点占比大于 0.3 且小于 0.9,因此保存或运算图像,即  $a < 0.45$  且  $0.3 < b < 0.9$ 。图 4(b)所示,与运算图像的肤色点占比大于 0.45,因此保存与运算图像,即  $a > 0.45$ 。图 4(c)所示,与运算和或运算图像的肤色提取效果不佳,因此保存原图像,此时与运算和或运算图像的肤色点占比分别小于 0.45 和 0.3,即  $a < 0.45$  且  $b < 0.3$ 。由图 4 示例可知,本文肤色融合算法可以较好的保留手势区域,但对于具有肤色背景和光照变换的图像仍需进一步处理,以提高手势分割效果。



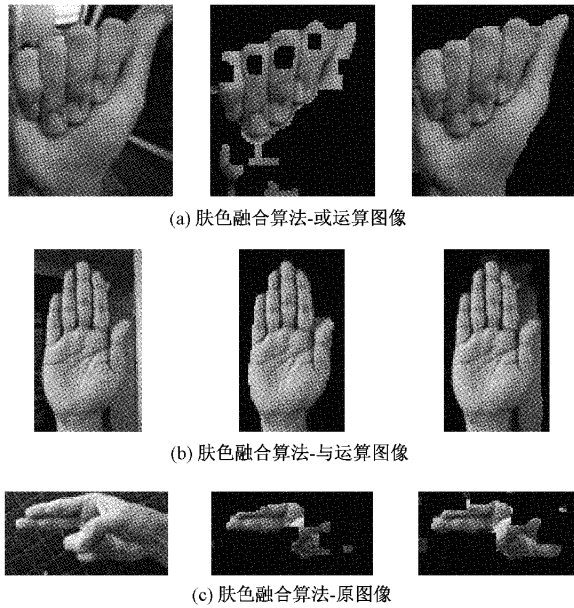


图 4 肤色融合算法的识别效果

### 1.2 改进的 Canny 边缘检测算法

Canny 算子是找到最优边缘的一种边缘检测算法,具有最优检测、最优定位和最优响应的特点,是实际应用中常见的边缘检测算法。Canny 算子的缺点是高斯滤波对图像的平滑处理会影响手势边缘的提取,此外,人为设定的双阈值也会导致边缘信息的丢失<sup>[15]</sup>。

针对 Canny 算子存在的不足之处,对其改进以获得效果更佳的手势边缘图像。首先,采用自适应中值滤波和双边滤波替换高斯滤波,可以在去除噪声的同时保留更多的边缘信息<sup>[16]</sup>。其次,梯度幅值和梯度方向的计算通过 Sobel 算子在 0°、45°、90° 和 135° 4 个方向上的一阶梯度分量计算得到,其中一阶梯度分量则由 4 个一阶梯度方向模板与滤波后的图像分别进行卷积运算得到,具体可参考文献[15],梯度幅值和梯度方向的计算公式为:

$$grad = \sqrt{grad\_x^2 + grad\_xy^2 + grad\_y^2 + grad\_yx^2} \quad (3)$$

$$dire = \arctan(grad\_y/grad\_x) \quad (4)$$

式中:grad\_x、grad\_xy、grad\_y 和 grad\_yx 分别表示四个方向的一阶梯度分量。然后,应用非极大值抑制方法去除虚假边缘,使边缘更清晰,并根据梯度方向将(-π/2, π/2)等分为 4 个区域,利用权重计算梯度两端的值,通过比较大小的判断是否为边缘点。最后,采用 Otsu 算法自适应求得阈值,并将轮廓周长小于 20 的边缘去除,达到强化边缘和减少噪点的效果。

改进的边缘检测算法的实验效果如图 5 所示,中间图像为传统 Canny 边缘检测算法,右侧图像为改进的 Canny 边缘检测算法。由实验结果图 5(a)和(b)可知,本文算法相较于传统算法在保留更多手势边缘的前提下,可以去除更多的背景边缘,但是仍然存在较多的背景边缘,对手势的

识别造成影响。

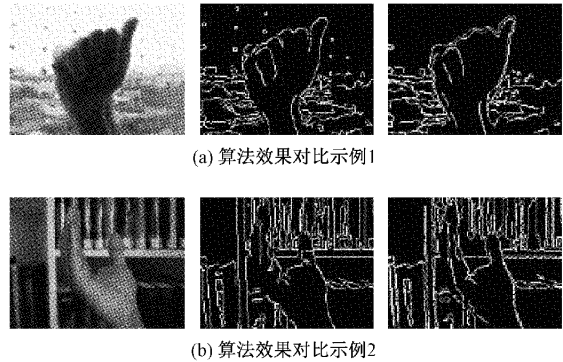


图 5 改进边缘检测算法对比

### 1.3 结合肤色和边缘特征的手势边缘填充

基于 YCrCb 颜色空间的肤色融合算法在有肤色背景和光照变化等情况下会出现识别背景和无法检测手势等情况。边缘检测算法可以较好地识别手势边缘,但同时也会检测到背景边缘。因此,根据上述肤色检测和边缘检测的优缺点,本文在肤色检测的基础上结合改进边缘检测算法提取手势边缘信息,并采用本文提出的边缘填充算法得到手势分割图像。

结合肤色和边缘特征的边缘填充算法具体分为 3 个步骤:1)对手势图像采用改进边缘检测算法,得到手势边缘特征。2)对上一步骤得到边缘图像采用边缘填充方法,具体为保留每行和每列的第一个和最后一个白点(肤色点),其余像素点的值设为 0,遍历图像,在以肤色点为中心的 3×3 领域内更新边缘点。更新规则如表 1 所示,“第 1 个”和“最后 1 个”分别表示边缘点在行(自左向右)或列(自上向下)中是第 1 个和最后 1 个白点,“中间”表示边缘点在行或列中位于第 1 个和最后 1 个白点之间,“第 1 个和最后 1 个”表示在行或列中仅有 1 个白点。3)对更新后的边缘图像填充,分别将每行和每列的第 1 个和最后 1 个白点之间的像素点的值设为 255,得到的两幅图像先进行与运算操作,再对图像做形态学运算处理。

表 1 边缘点更新规则

| 编号 | 每行白点的位置      | 每列白点的位置                   | 更新边缘点 |
|----|--------------|---------------------------|-------|
| 1  | 第 1 个        | 第 1 个、中间                  | 下和右   |
| 2  | 第 1 个        | 第 1 个和最后 1 个              | 右     |
| 3  | 中间           | 中间                        | 下和右   |
| 4  | 中间           | 第 1 个、最后 1 个、第 1 个和最后 1 个 | 右     |
| 5  | 最后 1 个       | 第 1 个或中间                  | 下     |
| 6  | 最后 1 个       | 最后 1 个                    | 左和下   |
| 7  | 第 1 个和最后 1 个 | 第 1 个、中间                  | 下     |

边缘点更新方法示例如图 6 所示,图像中数字“1、4、5、6”表示表 1 中的编号,即对应的边缘点更新规则,图像中数

字“255”表示边缘点,数字“0”表示新的边缘点。例如左上方的编号 1 所示,在  $3 \times 3$  领域内的中心点为边缘点,并且该点是本行的第一个白点,是本列中间位置的白点,因此根据表 1 应将边缘点下侧和右侧的像素点更新为新的边缘点。实验测试结果如图 7 所示,图中与运算图像的肤色点占比大于 0.45,根据图 3 判断输出与运算图像。由图 7 可知,基于 YCrCb 颜色空间的肤色融合算法可以较好的获取手势区域,但肤色背景同样会被保留,但是由于肤色背景特征单一,因此可以利用边缘检测算法得到完整的手势边缘,最终采用边缘填充算法得到效果最佳的手势轮廓。综上所述,肤色融合算法可以最大程度得到准确的手势区域,结合肤色和边缘特征的手势边缘填充算法则可以得到完整的手势轮廓,对于手势的分类和识别具有较好的识别效果。

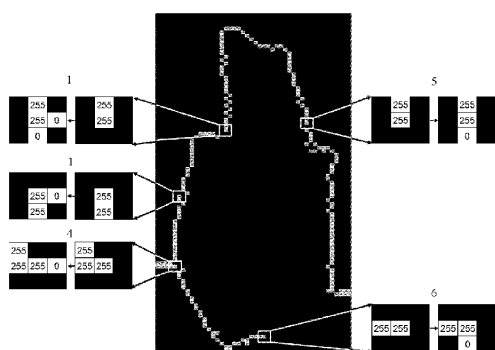


图 6 边缘点更新方法示例

## 2 基于卷积神经网络模型的手势识别

卷积神经网络(convolutional neural networks, CNN)是一种对数据快速响应的深度学习算法,本质是通过对大量数据的学习,得到训练网络从输入到输出的映射关系,可以通过改变其模型的深度和宽度来控制运算容量,即模型中“层”的数量和神经元的个数,因而常用于图像识别和图像分类<sup>[17]</sup>。CNN 的基本结构主要由输入层、卷积层、池化层、全连接层和输出层 5 个“层”组合而成<sup>[18]</sup>。卷积层是网络模型的关键部分,也是提取特征的重要手段,可以通过连续移动较小的卷积核实现局部特征到总体特征的提取。池化层的作用是降维,这是因为数据经过卷积层的计算后,其输出的结果可能存在信息冗余问题,所以要通过降维减少特征量,根据降采样方式的不同,可以分为平均池化和最大池化两种方式<sup>[19]</sup>。全连接层的作用主要是更改输出的维数,一般位于输出层的前面,便于分类或者回归模型发挥作用。

随着神经网络的层数不断增加,模型的训练也变得困难,因此需要对神经网络进行优化以降低模型的训练难度,常用的网络优化方法主要包括批量归一化层和 Dropout 层。批量归一化(batch normalization, BN)的本质是将较为分散的数据进行“中心化”以转换为紧密的形式。此外,由于神经网络的权重在训练过程中会不断更新,使得激活

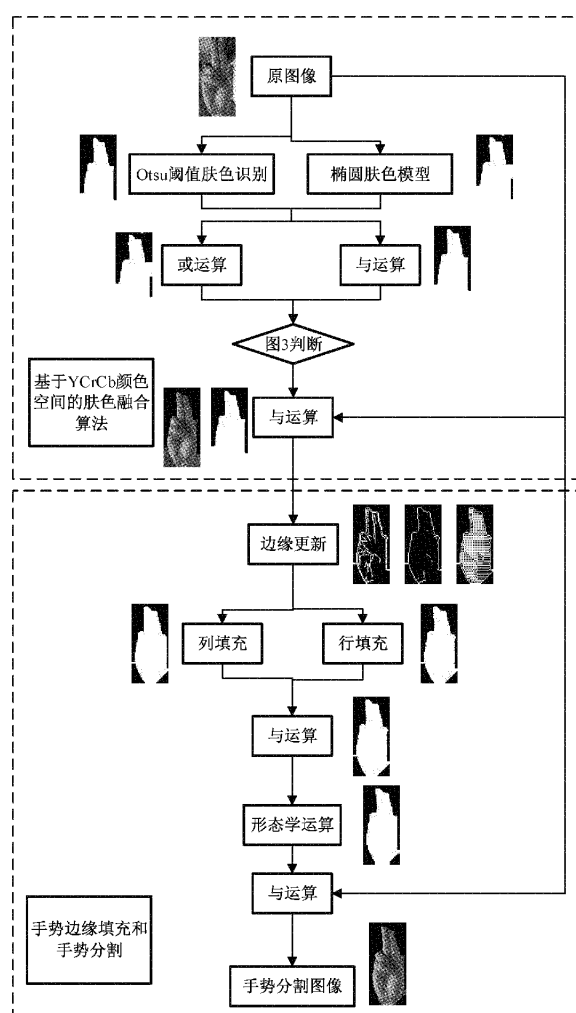


图 7 肤色和边缘特征结合的手势分割流程

函数的输出分布产生协方差偏移,最终导致梯度消失和梯度爆炸,而批量归一化层可以较好地解决这一问题。在卷积神经网络中,Dropout 层的作用是正则化,减少网络模型的计算量,并提高运行速率,这是因为 Dropout 层会在每批数据训练时随机舍弃部分输出特征,可以有效解决过拟合问题。

图 8 所示为本文设计的卷积神经网络模型的具体结构,由 3 个卷积块和 1 个全连接分类器组成。卷积块由 2 个卷积层、1 个批量归一化层和 1 个池化层组成,激活函数采用 ReLU 函数以提高模型非线性能力。此外,为了防止网络出现过拟合问题,在卷积块 3 后添加 1 个 Dropout 层,参数为 0.25。全连接分类器由 1 个 Flatten 层、2 个 Dense 层和 1 个批量归一化层组成,其中,2 个 Dense 层分别采用 ReLU 和 Softmax 作为激活函数。

## 3 实验结果与分析

本文的实验操作系统为 64 bit 的 Windows 11, CPU 型号为 AMD Ryzen 7 5800U with Radeon Graphics 1.90 GHz

(8核),GPU型号为 NVIDIA GeForce RTX 3050 Laptop GPU。编程语言为 Python 3.7,代码编辑器为 Pycharm,深度学习框架为 TensorFlow-GPU 2.7.0 和 Keras 2.7.0,并通过 CUDA 11.3.1 和 CUDNN 11.3 对 GPU 进行加速。采用的手势数据集是 NUS-II 数据集,这是一个包含 10 类手势的数据集,共计 2 000 张图像。数据集的图像在新加坡国立大学拍摄,图像具有复杂的自然背景,并由 40 名 22~56 岁的不同种族的男性和女性参与拍摄,因此,图像中手势的形状和大小都存在差异。

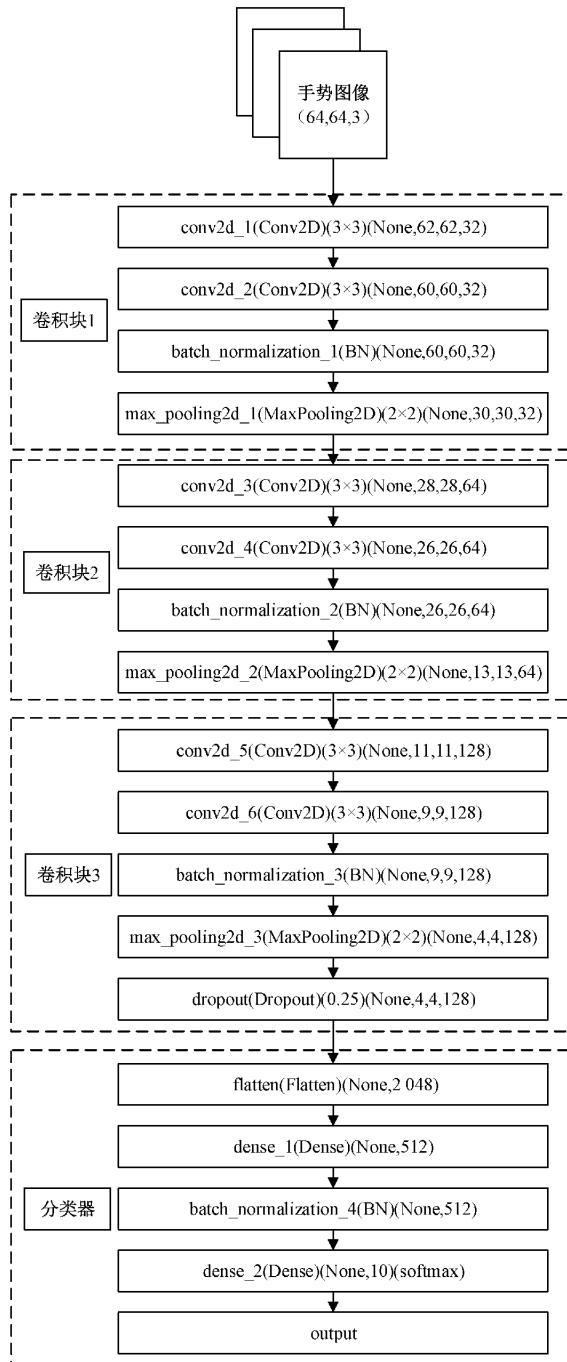
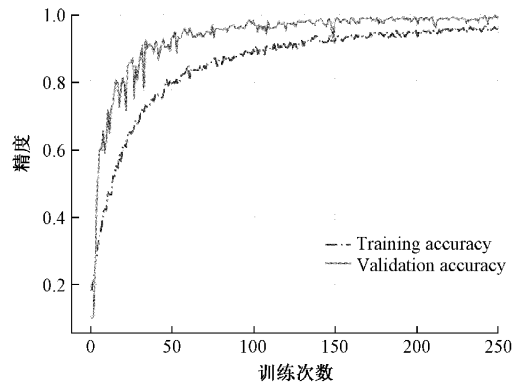
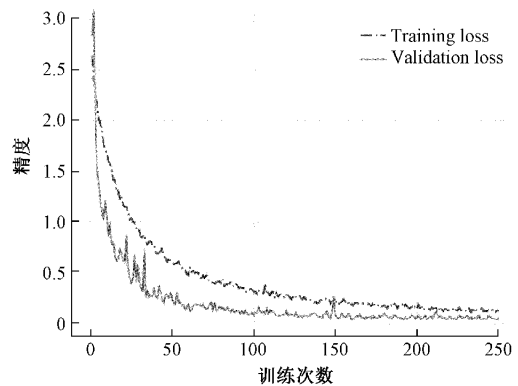


图 8 卷积神经网络结构

在网络模型训练时,训练集和验证集的比例为 4:1,即训练集 1 600 张(每类手势 160 张),验证集 400 张(每类手势 40 张),图像经过数据增强后大小为 (64,64,3),模型采用 Adam 优化器,初始学习率为 0.000 1,batch\_size 设为 10。因为数据集只有 2 000 张手势图像,所以采用 Python 生成器实现数据增强,生成器每次产生 (10,64,64,3) 的图像和 (10,) 的二进制标签所组成的批量,训练生成器和验证生成器分别生成 160 个和 40 个批量,即为完成一个轮次,共计训练 250 轮。网络模型的精度和损失如图 9(a)和图 9(b)所示,训练集上模型在 0~150 个轮次之间快速收敛,并在 225 个轮次后达到稳定,此时的精度和损失约为 96.16%和 0.119 4;验证集上模型在前 50 个轮次内快速拟合,并在 125 个轮次后趋于稳定,精度和损失约为 98.56%和 0.056 8。



(a) 训练集和验证集的精度



(b) 训练集和验证集的损失

图 9 网络模型的精度和损失

为了测试模型对于 10 类手势的识别率,采用验证集评估模型的性能,并以精确度 (Precision)、召回率 (Recall) 和 F1-score 作为评价指标对模型进一步分析,实验结果如图 10 和表 2 所示,“a~j”分别表示 10 类手势。图 10 为评估模型 10 类手势的混淆矩阵,横轴和纵轴分别表示模型预测类别和真实类别,由图可知模型仅在 3 类手势中预测错误,且错误识别数很小。表 2 所示为 10 类手势的精确度、召回率和 F1-score,3 种评价指标均能达到 0.95 以上,平均值分别 0.99、0.989 和 0.99。从图 10 和表 2 的数据来看,

模型针对复杂背景下的手势图像具有较好的识别效果,模型的性能优越。

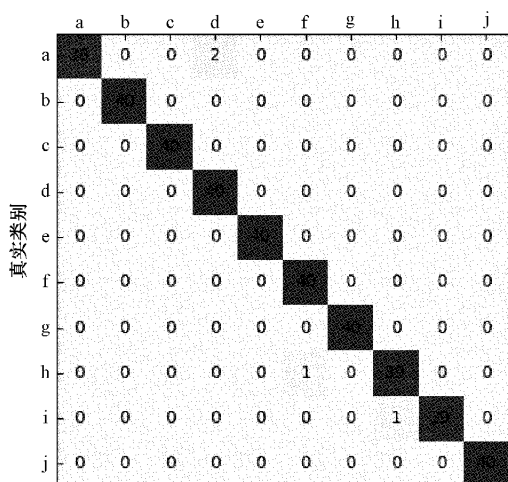


图 10 评估模型的混淆矩阵

表 3 所示为验证本文算法性能的消融实验,实验数据如图 11 所示。如图 11(a)所示,当消融对象为 Otsu 肤色识别算法时,由于椭圆肤色模型构建的椭圆大小固定,因此存在肤色点位于椭圆外而导致误判的情况,手势图像的肤色特征提取效果较差,手势识别的平均精度最低,为

表 3 消融实验

| 消融对象        | 训练集平均损失 | 验证集平均损失 | 训练集平均精度/% | 验证集平均精度/% |
|-------------|---------|---------|-----------|-----------|
| Otsu 肤色识别算法 | 0.157 4 | 0.086 8 | 94.79     | 98.13     |
| 椭圆肤色模型      | 0.141 1 | 0.776 7 | 95.27     | 98.26     |
| 改进边缘检测算法    | 0.120 2 | 0.047 9 | 95.87     | 98.55     |
| 无(本文算法)     | 0.109 2 | 0.034 2 | 96.25     | 98.83     |

表 2 10 类手势的验证结果

| 手势类别 | 样本数 | 精准度  | 召回率  | F1-score |
|------|-----|------|------|----------|
| a    | 40  | 1.00 | 0.95 | 0.97     |
| b    | 40  | 1.00 | 1.00 | 1.00     |
| c    | 40  | 1.00 | 1.00 | 1.00     |
| d    | 40  | 0.95 | 1.00 | 0.98     |
| e    | 40  | 1.00 | 1.00 | 1.00     |
| f    | 40  | 0.98 | 1.00 | 0.99     |
| g    | 40  | 1.00 | 1.00 | 1.00     |
| h    | 40  | 0.97 | 0.97 | 0.97     |
| i    | 40  | 1.00 | 0.97 | 0.99     |
| j    | 40  | 1.00 | 1.00 | 1.00     |

98.13%;如图 11(b)所示,当仅采用 Otsu 肤色识别算法提取肤色特征时,即消融对象为椭圆肤色模型,通过对 Cr 分量进行二值化处理可以避免光照变化的影响,有效地分割肤色点,但是此时也会提取类肤色背景特征,因而平均精度提升有限为 98.26%;如图 11(c)所示,当采用传统边缘检测算法消融改进算法以提取边缘特征时,最小和最大阈值分别设为 100 和 200,此时对手势识别效果影响相对较小,平均精度为 98.55%,说明肤色融合算法可以在提取手势肤色特征同时降低类肤色背景对算法性能的影响。

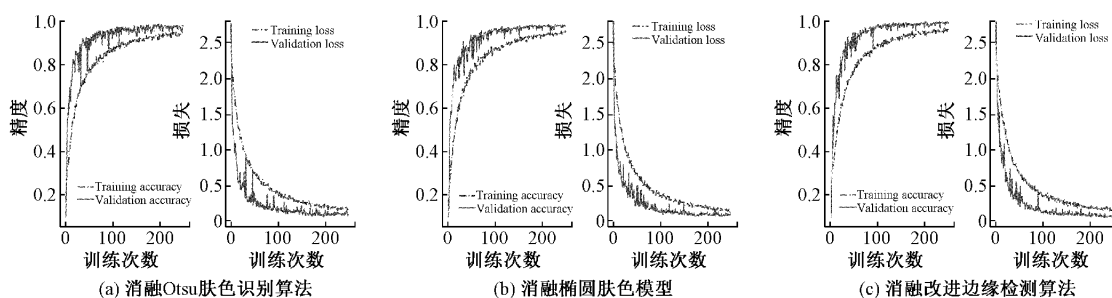


图 11 消融实验的精度和损失

表 4 所示为本文算法和其他文献算法的性能比较。本文算法相较于文献[8],使用了相同的 NUS-II 数据集,且识别率相近,但是文献算法在边缘特征提取时没有进行图像预处理,会增加背景边缘,影响算法效果。此外在使用文献中的卷积神经网络模型训练本文处理后的数据集时,卷积神经网络模型在运行效率方面相对较慢,并且手势识别的准确率也有所降低。文献[10]的手势识别算法是基于 HSV 颜色空间的肤色识别和椭圆肤色模型的融合

算法,结合 Otsu 法实现手势分割,文献算法相较于本文算法复杂度相对简单,并且基于 HSV 颜色空间的肤色识别效果比基于 YCrCb 颜色空间的肤色识别效果较差。文献[11]采用基于 YCrCb 颜色空间的 Otsu 阈值肤色识别算法提取肤色特征,并在边缘检测算法基础上采用椭圆傅里叶描述子拟合边缘,实现肤色特征和边缘特征的结合,文献可以实现 13 类手势的识别和分类,但是算法在简单背景下的手势识别率较低。综上所述,本文算法的手势识别



性能最好,在复杂背景下提取更多手势特征的基础上,可以提高网络模型的运行效率。

表 4 本文算法与其他算法对比

| 算法     | 手势种类 | 平均识别率/% |
|--------|------|---------|
| 文献[8]  | 10   | 98.81   |
| 文献[10] | 10   | 98.80   |
| 文献[11] | 13   | 89.96   |
| 本文算法   | 10   | 98.83   |

## 4 结 论

本文提出了一种基于卷积神经网络的肤色特征和边缘特征结合的手势识别方法,考虑了肤色背景和光照变化对于肤色检测算法提取手势肤色特征的影响,采用肤色一边缘检测算法和边缘检测算法的双通道手势识别方法,实现肤色和边缘特征相结合的手势分割方法,结果表明本文方法可以有效分割手势区域。此外,构建的卷积神经网络具有较好的鲁棒性,可以准确地提取和学习手势特征,提高手势识别的精度,平均识别率为 98.83%。

本文算法仍然存在一些不足之处和问题。一方面,尽管本文算法已经可以降低肤色背景和光照变化所带来的影响,但是部分图像的手势分割效果仍旧较差,存在进一步的改进空间。另一方面,本文算法仍然属于传统的图像处理问题,卷积神经网络仅仅用于图像预处理后的分类,因此未来的研究方向是利用全卷积神经网络或 U-Net 等网络模型实现手势分割,再结合 CNN 实现手势分类。

## 参考文献

- [1] 谢铮桂. 基于改进的卷积神经网络的手势识别的研究[J]. 计算机应用与软件, 2019, 36(3): 192-195,281.
- [2] XU C, WU X, WANG M M, et al. Improving dynamic gesture recognition in untrimmed videos by an online lightweight framework and a new gesture dataset ZJUGesture [J]. Neurocomputing, 2023, 523(2023): 58-68.
- [3] CHEOK M J, OMAR Z, JAWARD M H. A review of hand gesture and sign language recognition techniques [J]. International Journal of Machine Learning and Cybernetics, 2019, 10(1): 131-153.
- [4] 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.
- [5] 黄杰. 基于深度学习的手语识别技术研究[D]. 合肥: 中国科学技术大学, 2018.
- [6] JIANG D, LI G F, SUN Y, et al. Gesture recognition based on skeletonization algorithm and CNN with ASL

database [J]. Multimedia Tools and Applications, 2019, 78(21): 29953-29970.

- [7] WU X Y. A hand gesture recognition algorithm based on DC-CNN[J]. Multimedia Tools and Applications, 2019, 79(13): 9193-9205.
- [8] 柏丽银, 彭亚雄, 陆安江, 等. 复杂背景下基于卷积神经网络的手势识别[J]. 计算机工程与设计, 2020, 41(11): 3199-3203.
- [9] 杨镛. 基于混合神经网络的肌电手势识别算法研究[D]. 南京: 南京邮电大学, 2021.
- [10] 张晓俊, 李长勇. 基于深度学习多特征融合的手势分割识别算法[J]. 济南大学学报(自然科学版), 2022, 36(3): 286-291.
- [11] 徐玥, 周辉. 简单背景下基于 OpenCV 的静态手势识别[J]. 计算机科学, 2022, 49(S2): 393-398.
- [12] 王新, 王赛. 基于改进 YOLOv5 算法的交警手势识别[J]. 电子测量技术, 2022, 45(2): 129-134.
- [13] 孟杰, 杨鹏程, 杨朝, 等. 基于 Mediapipe 的幻影成像装置自然手势交互系统设计[J]. 国外电子测量技术, 2023, 42(3): 116-122.
- [14] SAHOO J P, SAHOO S P, ARI S, et al. Hand gesture recognition using densely connected deep residual network and channel attention module for mobile robot control [J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72(2023): 1-11.
- [15] 许宏科, 秦严严, 陈会茹. 一种基于改进 Canny 的边缘检测算法[J]. 红外技术, 2014, 36(3): 210-214.
- [16] 赵芳, 周旺辉, 陈岳涛, 等. 改进的 Canny 算子在裂缝检测中的应用[J]. 电子测量技术, 2018, 41(20): 107-111.
- [17] 常志磊. 基于 RGB-D 的双手手势识别方法研究及系统设计[D]. 呼和浩特: 内蒙古大学, 2022.
- [18] 孙进, 张道周, 张洋, 等. 基于双通道空洞卷积神经网络的手势识别[J]. 传感器与微系统, 2022, 41(3): 126-128.
- [19] 牛雅睿, 武一, 孙昆, 等. 基于轻量级卷积神经网络的手势识别检测[J]. 电子测量技术, 2022, 45(4): 91-98.

## 作者简介

**赵鸿图**(通信作者), 博士, 副教授, 主要研究方向为智能信息处理、数字信号处理、嵌入式系统等。

E-mail: hongtuzhao@hpu.edu.cn

**李豪**, 硕士研究生, 主要研究方向为计算机视觉、图像处理等。

E-mail: 1623716992@qq.com

**梁梦华**, 硕士研究生, 主要研究方向为智能信息处理、图像处理等。

E-mail: 1768394588@qq.com