

# 用于自动视力检测的手势识别方法研究\*

何启莉, 何家峰, 郭娟

(广东工业大学 信息工程学院, 广东 广州 510006)

**摘要:** 对于自动视力检测系统, 手势识别是关键问题, 但是采用传统卷积神经网络模型识别手势存在过拟合、计算量大等问题。提出了一种 GR-AlexNet 模型, 对 AlexNet 网络模型进行了适应性修改和优化: 为了加快计算速度, 用  $7 \times 7$ 、 $5 \times 5$ 、 $1 \times 1$  的三个小卷积核替代原来的  $11 \times 11$  的大卷积核, 并删除 LRN 层和一个全连接层; 为了减轻过拟合效应, 在每次卷积后都加上一个 Dropout 优化。对同一数据集分别使用 LeNet 模型、AlexNet 模型、VGG16 模型与 GR-AlexNet 模型进行对比实验。实验表明 GR-AlexNet 模型在识别准确率上较传统的模型有一定的提高, 能抑制过拟合现象, 并且具有更快的训练速度。

**关键词:** 自动视力检测; OpenCV; 手势识别; GR-AlexNet

中图分类号: TP391.41

文献标识码: A

DOI: 10.19358/j.issn.2096-5133.2021.03.006

引用格式: 何启莉, 何家峰, 郭娟. 用于自动视力检测的手势识别方法研究[J]. 信息技术与网络安全, 2021, 40(3): 32-37, 47.

## Research on gesture recognition method for automatic vision detection

He Qili, He Jiafeng, Guo Juan

(School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China)

**Abstract:** For automatic vision detection systems, gesture recognition is a key issue, but the traditional convolutional neural network model to recognize gestures has problems such as over-fitting and large amount of calculation. This paper proposes a GR-Alexnet model, which adaptively modifies and optimizes the Alexnet network model. In order to speed up the calculation, three small convolution kernels of  $7 \times 7$ ,  $5 \times 5$ , and  $1 \times 1$  are used to replace the original  $11 \times 11$  large convolution kernel, and delete the LRN layer and a fully connected layer; in order to reduce the over-fitting effect, a dropout optimization is added after each convolution. The LeNet model, the Alexnet model, the VGG16 model and the GR-Alexnet model were used for comparative experiments on the same data set. Experiments show that the GR-Alexnet model has a certain improvement in recognition accuracy compared with the traditional model, can suppress the over-fitting phenomenon, and has a faster training speed.

**Key words:** automatic vision detection; OpenCV; gesture recognition; Gesture Recognition AlexNet (GR-AlexNet)

### 0 引言

随着人工智能技术的进步, 智能化设备逐渐融入到人们生活的方方面面。传统的医疗检测仪器逐渐被智能电子仪器所替代, 如心率测量仪、血压检测仪等, 然而视力检测这一基本的体检项目仍然沿用传统的人工检测方法, 检测效率低, 消耗人力且极不方便。随着计算机视觉技术迅速发展, 手势识别也逐渐成为智能人机交互的重要研究领域<sup>[1-4]</sup>。本

文通过对视力检测进行手势识别, 达到自动化视力检测的目的。

目前对手势识别技术的研究主要包括基于外部设备的手势识别技术和基于计算机视觉的手势识别技术两方面。基于外部设备的手势识别技术在硬件造价上过于昂贵, 比如利用数据手套<sup>[5-6]</sup>或者固定在手臂上的动作传感器来获取手势数据<sup>[7]</sup>, 并且需要全套的配套设施, 限制较多。基于计算机视觉的手势识别技术核心在于提取手势的特征。传统的识别算法多是由人工设计提取特征, 然后再用分

\* 基金项目: 国家自然科学基金(61571140)

类器分类识别,比如刘小建提出一种使用深度信息进行多特征提取的手势识别算法,再应用一个多分类的 SVM 分类器对手势进行分类<sup>[8]</sup>。人工提取特征繁琐、费时,可能还会因为提取特征产生的误差而导致识别准确率低。随着卷积神经网络在计算机视觉领域的迅速发展,对于将卷积神经网络用在手势识别上有了突破性的进展<sup>[9]</sup>。朱雯文在对 LeNet-5 卷积神经网络进行分析的基础上,提出了一种适合加速信号的 LeNet-A 网络,其识别准确率达到 90.37%<sup>[10]</sup>。龙辉基于深度学习理论将 MobileNet 和 SSD 网络相结合得到一种新的网络模型,该模型的识别准确率达到 93.7%<sup>[11]</sup>。为了提高识别的准确率,神经网络层次越来越复杂,大幅度地提高了网络的计算量的同时,容易出现过拟合现象<sup>[12]</sup>。为此,本文提出的方法是:手势识别前先对其进行肤色阈值分割处理,并将传统的 AlexNet 网络模型中的大卷积核替换成小卷积,删除全连接层,不再使用 LRN 层,在每次卷积之后都加上 Dropout 层<sup>[13]</sup>,以此来解决计算量大、过拟合问题,增强了模型的非线性,提高识别率。

## 1 自动视力检测系统

设计一个视力检测系统,该系统通过手势识别技术实现自动化视力检测<sup>[14]</sup>。系统包括手势采集模块、手势识别模块、输出控制模块以及显示器、扬声器等输出模块,其整体框架图如图 1 所示。

视力检测者看到显示器显示的“E”图标时,直接用手部动作给出“E”开口方向,通过摄像头采集该手势信息,对采集到的手势图片进行预处理后输

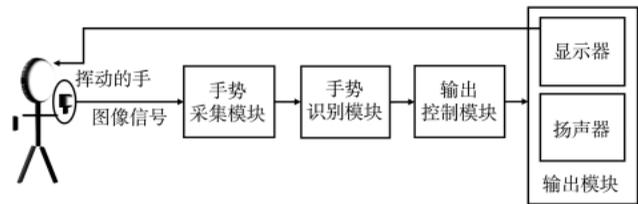


图 1 视力检测系统整体框架图

入手势识别模块进行识别,根据控制模块判断结果,控制显示器显示不同视力等级的“E”图标,再次重复进行上述步骤,最后得出结果。比对医疗检测标准对应的人体视力值,从而得到被测者的视力值并显示最终结果。

## 2 手势采集以及预处理

### 2.1 手势采集

视力检测者在进行视力检测时,看到“E”开口方向,直接用手势动作信息确定方向。本文数据集包括视力检测时的五种手势(上、下、左、右、看不清),利用 Python 中的 OpenCV 调用电脑摄像头,对录制手势视频按照一定的帧数保存图片,图片的命名将直接作为对应的标签。五种手势,每种手势在不同角度、不同背景、不同亮度下采集 1 300 张图片,整个数据集包含 6 600 张图片。

### 2.2 手势图像预处理

电脑摄像头采集的手势图片大小为 300 像素×300 像素,为了加快训练网络的收敛性,将采集到的手势图片进行预处理,并对图片进行归一化,得到大小为 227 像素×227 像素的图片。采集到的手势数据集中部分手势如图 2 所示。

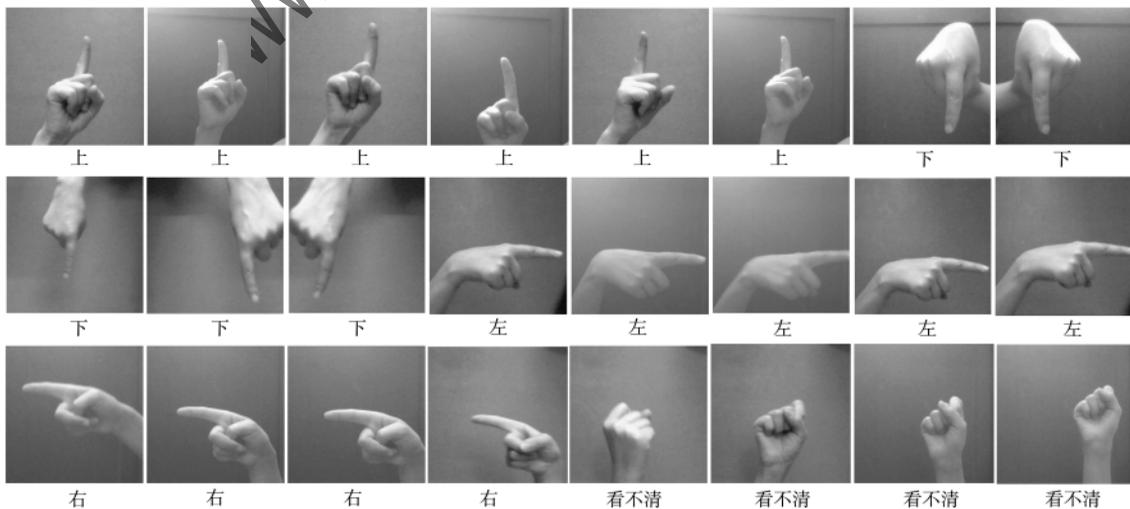


图 2 手势数据库中样例

本文利用基于 YCrCb 颜色空间的 Cr 分量和 Otsu 法阈值分割算法来对手势训练集进行降噪处理。

YCrCb 是彩色空间的一种，其中 Y 为亮度成分，Cb 和 Cr 分别为蓝色和红色的浓度偏移量成分。因为 YCrCb 空间受亮度影响很小，肤色会产生很好的聚类，这样就能把三维空间降为二维的 CrCb 空间，肤色点会形成一定的形状。根据经验，满足肤色的 CrCb 的值为： $133 \leq Cr \leq 173, 77 \leq Cb \leq 127$ ，其他的为非肤色点。

Otsu 算法是由日本学者大津于 1979 年提出的一种对图像进行二值化的高效算法，也称为最大类间算法，计算简单，不受图像亮度和对比度的影响。Otsu 算法是按图像的灰度特性将图像分为背景和前景两部分。用方差作为灰度分布均匀的一种度量，前景和背景间的类间方差越大，说明构成图像的两部分的差别越大。公式表示如下：

$$g = \omega_0 \omega_1 (\mu_0 - \mu_1)^2 \quad (1)$$

其中  $\omega_0$  为前景的像素点占整幅图像的比列，其平均灰度为  $\mu_0$ ， $\omega_1$  为背景的像素点占整幅图像的比列，其平均灰度为  $\mu_1$ ，类间方差记为  $g$ 。

对手势图像预处理过程分为以下三步：

- (1) 将 RGB 图像转换到 YCrCb 颜色空间，提取 Cr 分量图像；
- (2) 对 Cr 分量进行高斯滤波；
- (3) 对高斯滤波后的 Cr 分量图像用 Otsu 法做二值化阈值分割处理。处理后的图像效果如图 3 所示。

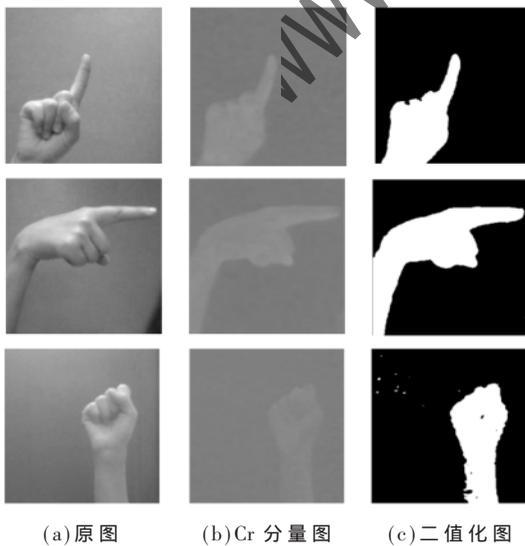


图 3 阈值分割处理

### 3 卷积神经网络及改进

#### 3.1 卷积神经网络

卷积神经网络的特点是局部链接、权值共享和空间或时间上的下采样(池化)。卷积神经网络具有表征学习的能力，能按照其阶层结构对所输入的信息进行平移不变分类，采用梯度下降法最小化损失函数，对网络中的权重参数逐层进行反向调节，通过高频的迭代训练来提高网络的精度。卷积是 CNN 的核心，用卷积核扫描图像，提取图像中的特征信息。不同的卷积核与上一层得到的特征图作卷积运算，然后加上偏置值得到输出，再通过一个激活函数推出卷积结果，得到特征图。卷积层的输出特征  $F_c$  表示为：

$$F_c = f \left( \sum_{i=1}^m \text{con}(M_i, C_i) + B_i \right) \quad (2)$$

其中有  $m$  个卷积层，第  $i$  个出入特征为  $M_i$ ，对应的卷积核为  $C_i$ ， $B_i$  为偏置值， $f$  表示激活函数， $\text{con}$  表示卷积函数。

对卷积输出特性进行池化操作，减小维度，再通过对全连接层进行权值变换和激活，可获得分类结果。全连接层的分类函数可表示为：

$$F_o = f(WT + B_2) \quad (3)$$

式中： $f$  为激活函数， $W$  是相应权重， $T$  表示完全连接层的输入特性， $B_2$  是偏置值。

#### 3.2 Dropout 优化

Dropout 是一种防止模型过拟合的技术，基本思想是在训练时随机地丢弃一些神经元，使网络更稀疏，模型不会过于依赖某些局部特征，从而让模型具有更好的鲁棒性，提高模型的泛化能力。Dropout 优化前后的对比如图 4 所示。

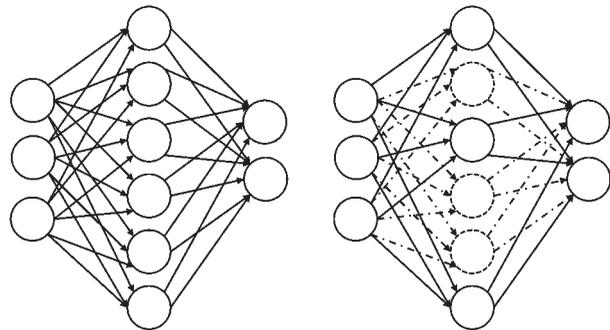


图 4 Dropout 优化前后对比图

#### 3.3 AlexNet 网络

AlexNet 网络结构<sup>[15]</sup>是在 2012 年的 ImageNet 竞

赛中取得冠军的一个网络模型。该模型共有 8 层, 包括 5 个卷积层和 3 个全连接层, 在每一个卷积层中包含了激励函数 ReLU 以及局部响应归一化 (LRN) 处理, 然后经过降采样 (pooling 处理), 最后通过一个 Softmax 分类器进行分类。Softmax 模型是 logistic 回归模型在多分类问题上的推广, 可以为每个输出分类的结果都赋予一个概率值, 表示属于每个类别的可能性, Softmax 函数表示如下:

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{c=1}^C e^{z_c}} \quad (4)$$

其中  $z_i$  为第  $i$  个节点的输出值,  $C$  为分类的类别个数。AlexNet 的基本架构如图 5 所示。

### 3.4 GR-AlexNet 模型

GR-AlexNet (Gesture Recognition AlexNet) 模型是本文针对视力检测所提出的模型。GR-AlexNet 模型如图 6 所示。

大尺寸的卷积核具有更大的感受野, 但是大尺寸的卷积核会带来更多的参数, 增加计算量。受文献[16]的启发, 用卷积核大小分别为  $5 \times 5$ 、 $7 \times 7$  的两个小卷积核来替代  $11 \times 11$  的大卷积核, 减少参数个数, 加快计算。再级联一个  $1 \times 1$  的卷积层, 增加网

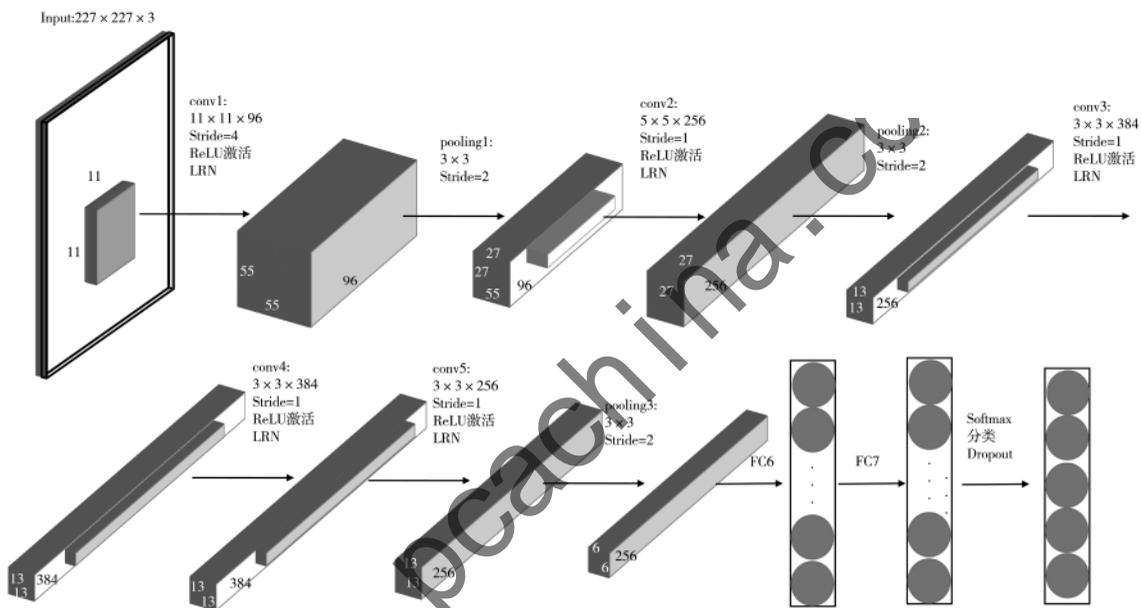


图 5 AlexNet 模型结构图

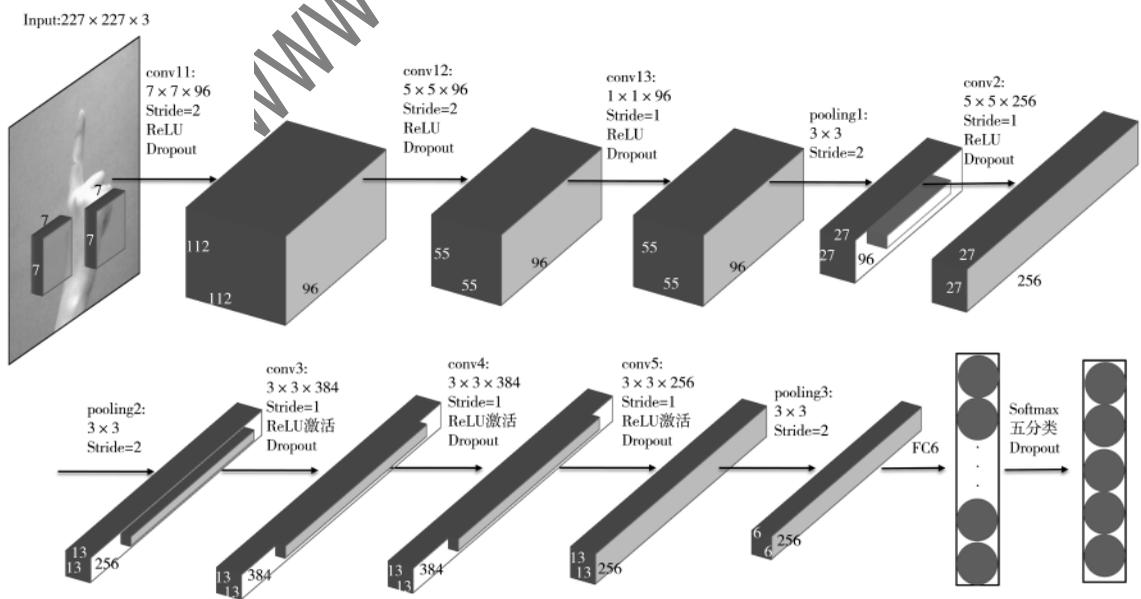


图 6 GR-AlexNet 模型结构图

络模型的深度及非线性<sup>[17]</sup>。同时,为了避免过拟合,每次卷积后增加一个 Dropout 操作。

(1)在 GR-AlexNet 网络中,用  $7 \times 7$ 、 $5 \times 5$ 、 $1 \times 1$  的三个小卷积核代替了原网络中  $11 \times 11$  的大卷积核,改进后的卷积层的具体参数如表 1 所示。去掉一层全连接层(FC7),具体的参数变化如表 2 所示。参数对比:改进后参数减少 1 458 144 个。

表 1 Conv1-1, Conv1-2, Conv1-3 具体参数

卷积层	Kernel-size	卷积核个数	stride	padding
Conv1-1	7	96	2	same
Conv1-2	5	96	2	same
Conv1-3	1	96	1	same

表 2 原网络与改进网络的参数变化对比

原网络中层次及参数	改进网络层次及参数
Conv1 : $(11 \times 11 \times 3 + 1) \times 96 = 34\ 944$	Conv1-1 : $(7 \times 7 \times 3 + 1) \times 96 = 14\ 208$
FC7 : $4\ 096 \times 4\ 096 = 1\ 677\ 216$	Conv1-2 : $(5 \times 5 \times 96 + 1) \times 96 = 230\ 496$
	Conv1-3 : $(1 \times 1 \times 96 + 1) \times 96 = 9\ 312$

(2)在每次卷积后都增加一个 Dropout 优化操作,并且删除激活函数与池化层之间的 LRN 层。

## 4 实验与结果分析

### 4.1 实验环境

实验环境为 64 位 Win10 系统,CPU 为 AMD Ryzen 7 4800H with Radeon Graphics,CPU 的主频为 2.9 GHz,内存为 8 GB。训练模型是在 Anaconda 下基于 Python 语言的 Pycharm 编译工具,采用 Tensor-Flow 深度学习框架。

### 4.2 网络训练

本文实验数据集是采集的关于视力检测时手势,包括上、下、左、右和看不清 5 种手势,每种手势有 1 300 张,共计 6 600 张图片。图片的大小统一

为 227 像素  $\times$  227 像素,随机地选取其中的 80% 作为训练集,20% 作为测试集在网络训练中,学习率设置为 0.001,迭代次数设置为 320,批次的大小设置为 32,使用随机梯度下降法来训练网络,更新参数,最小化损失函数,Dropout 的概率设置为 0.9,使用 ReLU 激活函数,ReLU 激活函数的表达式如下:

$$f(x) = \max(0, x) \tag{5}$$

### 4.3 对比实验

#### 4.3.1 识别准确率对比

分别训练 LeNet5 模型、AlexNet 模型以及 GR-AlexNet 模型,并且对其识别准确率进行对比。识别准确率与迭代次数的曲线如图 7 所示。

从图 7 中可以看出,识别准确率在迭代步数达到 320 时,趋于饱和。其中,LeNet5 识别准确率为 90.62%,AlexNet 识别准确率为 93.75%,VGG16 识别准确率为 87.5%,GR-AlexNet 模型识别准确率为 96.88%,由此可以看出,针对于本文中的数据集,GR-AlexNet 模型有更高的识别率。

#### 4.3.2 损失率对比

分别训练 LeNet5 模型、AlexNet 模型、VGG16 模型以及 GR-AlexNet 模型,将其损失率进行对比。损失率与迭代次数的曲线如图 8 所示。

针对本文的数据集,神经网络的损失率随着训练步数的增加而降低,并且均在 320 步时趋于最小值。从图 8 中可以看出,传统卷积神经网络和改进后的神经网络的损失率均随着训练步数的增加而降低,并且均在 320 步时趋于最小值。通过大量的实验,VGG16 虽然具有最低的损失率,但是存在过拟合现象,所以 VGG16 针对本文的数据集的识别准确率很低。

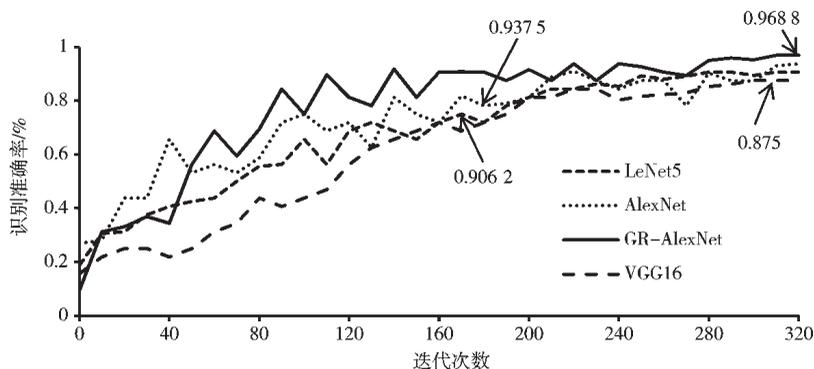


图 7 识别准确率与迭代次数关系曲线

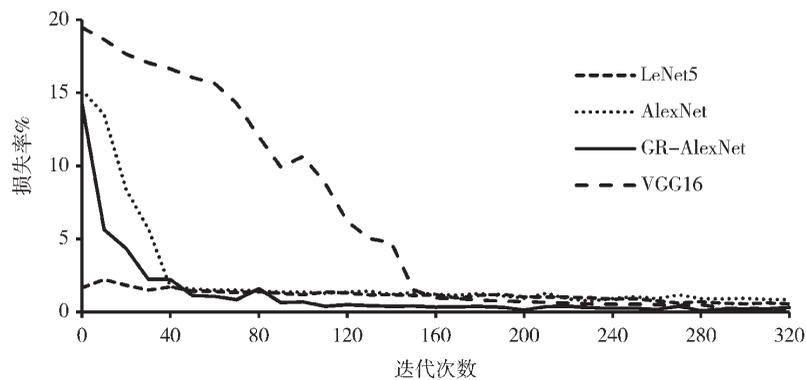


图8 损失率与迭代次数关系曲线

#### 4.3.3 训练时间对比

对于本文中的手势数据集,分别计算用 AlexNet 模型、VGG16 模型和 GR-AlexNet 模型训练的耗时,且两种模型的训练迭代次数均相同,其耗时时长如表 3 所示。

表3 AlexNet 和 GR-AlexNet 下的训练耗时

实验次数	AlexNet 耗时/min	VGG16 耗时/min	GR-AlexNet 耗时/min
1	23	42	19
2	24	39	20
3	27	41	17
平均耗时	24.7	40.7	18.7

从表 3 中可以看出, AlexNet 模型训练耗时在 25 min 左右, GR-AlexNet 模型训练耗时在 19 min 左右,而 VGG16 模型训练耗时在 40 min 左右。所以 GR-AlexNet 模型具有较快的训练速度。

## 5 结论

本文提出了一种改进的卷积神经网络模型: GR-AlexNet 模型,用于对视力检测涉及的 5 种手势进行识别。模型在足够的样本学习之后,可以自动获取样本特征,多次训练后得到高性能模型。与传统的模型相比, GR-AlexNet 模型网络参数减少,计算速度提高,模型的非线性增强,具有更好的泛化性能,损失率也更低。用测试样本图片对改进的模型进行测试,并且与传统的 LeNet5、AlexNet 模型以及 VGG16 模型作对比实验。GR-AlexNet 模型的识别准确率较 LeNet5 提高了 6.26%,较 AlexNet 提高了 3.13%,较 VGG16 提高了 9.38%。实验结果表明 GR-AlexNet 模型在加快了训练速度的基础上,也提高了识别精度。

## 参考文献

- [1] SHU Y, XIONG C, FAN S. Interactive design of intelligent machine vision based on human-computer interaction mode[J]. *Microprocessor and Microsystems*, 2020, 75(59): 12-17.
- [2] JEGHAM I, KHALIFA A B, ALOUANI I, et al. Vision-based human action recognition: an overview and real world challenges[J]. *Digital Investigation*, 2020, 32(1): 42-59.
- [3] FENG X, JIANG Y, YANG X. Computer vision algorithms and hardware implementations: a survey[J]. *Integration, the VLSI Journal*, 2019(69): 309-320.
- [4] ALAM M, SAMAD M D, VIDYARATNE L, et al. Survey on deep neural networks in speech and vision systems[J]. *Neurocomputing*, 2020(417): 302-321.
- [5] MAITRE J, CLEMÉNT RENDU A, KÉVIN BOUCHARD A, et al. Basic daily activity recognition with a data glove[J]. *Procedia Computer Science*, 2019(151): 108-115.
- [6] 刘雪花, 洪琴, 何玉麟, 等. 基于数据手套的手势识别系统[J]. *单片机与嵌入式系统应用*, 2020(6): 16-19.
- [7] KIM J H, HONG G S, KIM B G, et al. DeepGesture: deep learning-based gesture recognition scheme using motion sensors[J]. *Displays*, 2018(55): 38-45.
- [8] 刘小建, 张元. 基于多特征提取和支持向量机分类的手势识别[J]. *计算机工程设计*, 2017, 38(4): 953-958.
- [9] ZHANG S, MENG W, LI H, et al. Multimodal spatiotemporal networks for sign language recognition[J]. *IEEE Access*, 2019(10): 180270-180280.

(下转第 47 页)

- [11] Zhao Minghang, Tang Baoping, Deng Lei, et al. Multiple wavelet regularized deep residual networks for fault diagnosis[J]. Measurement, 2020, 152.
- [12] 熊剑, 邓松, 时大方. 基于改进残差网络的滚动轴承故障诊断[J]. 轴承, 2020(11): 50-55.
- [13] 梁志芳. 电子鼻系统中干扰抑制算法的研究[D]. 重庆: 重庆大学, 2017.
- [14] 刘伟玲, 吴龙焦, 张思祥, 等. 基于电子鼻技术的混合气体识别研究[J]. 现代电子技术, 2020, 43(6): 57-60.
- [15] 谢亮. 基于支持向量机的混合气体定量分析[J]. 电脑与信息技术, 2009, 17(2): 19-21, 24.  
(收稿日期: 2020-12-08)

---

作者简介:

- 詹灿坚(1995-), 男, 硕士研究生, 主要研究方向: 基于深度学习的混合气体浓度分析。
- 何家峰(1970-), 男, 博士, 副教授, 主要研究方向: 利用电子鼻进行混合气体定量分析。
- 骆德汉(1958-), 男, 博士, 教授, 主要研究方向: 基于机器嗅觉的气味复现。

---

(上接第 37 页)

- [10] 朱雯文, 叶西宁. 基于卷积神经网络的手势识别算法[J]. 华东理工大学学报(自然科学版), 2018, 44(2): 260-269.
- [11] 龙辉. 基于深度学习的手势识别算法设计[J]. 智能处理与应用(物联网技术), 2019(2): 57-59.
- [12] EL KORCHI A, GHANOU Y. DropWeak: a novel regularization method of neural networks[J]. Procedia Computer Science, 2018(127): 102-108.
- [13] PIERRE BALDI, PETER SADOWSKI. The dropout learning algorithm[J]. Artificial Intelligence, 2010(14): 78-122.
- [14] 张琴. 基于手势识别的视力检测系统的设计[J]. 电脑与电信, 2019(6): 5-8.
- [15] KRIZHEVSKY, ALEXSUTSKEVER, et al. ImageNet classification with deep convolutional neural networks[J].

IEEE Access, 2012, 32(5): 102-110.

- [16] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2016(308): 2818-2826.
- [17] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2015(6): 1-9.  
(收稿日期: 2020-12-16)

---

作者简介:

- 何启莉(1995-), 女, 硕士研究生, 主要研究方向: 机器学习、图像识别。
- 何家峰(1970-), 男, 副教授, 主要研究方向: 图像处理与模式识别、人工嗅觉信号处理。
- 郭娟(1997-), 女, 硕士研究生, 主要研究方向: 机器学习、仿生嗅觉。

---

(上接第 42 页)

- [10] WANG X, SHRIVASTAVA A, GUPTA A. A-Fast-RCNN: hard positive generation via adversary for object detection[C]. Honolulu: 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2017.
- [11] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Las Vegas: 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016.
- [12] 李大军, 何维龙, 郭丙轩, 等. 基于 Mask-RCNN 的建筑物目标检测算法[J]. 测绘科学, 2019, 44(10): 172-180.
- [13] OK A O. Automated detection of buildings from single VHR multispectral images using shadow information

and graph cuts[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2013(86): 21-40.

- [14] 赵春晖, 李思聪. 基于多特征融合的 SVM 高分辨率遥感影像建筑物提取方法[J]. 沈阳大学学报(自然科学版), 2017, 29(4): 314-319.
- [15] 刘文涛, 李世华, 覃驭楚. 基于全卷积神经网络的建筑物屋顶自动提取[J]. 地球信息科学学报, 2018, 20(11): 1562-1570.  
(收稿日期: 2020-12-29)

---

作者简介:

- 余慧明(1994-), 男, 硕士, 主要研究方向: 目标检测。
- 周志祥(1983-), 男, 本科, 主要研究方向: 通信技术。
- 彭杨(1993-), 男, 本科, 主要研究方向: 目标检测。