

文章编号: 2095-2163(2020)07-0064-03

中图分类号: TP391

文献标志码: A

# 关于残差网络的手势识别算法实现

郝禹哲<sup>1</sup>, 袁天夫<sup>1</sup>, 田海越<sup>2</sup>

(1 上海工程技术大学 电子电气工程学院, 上海 201620; 2 上海工程技术大学 机械与汽车工程学院, 上海 201620)

**摘要:** 残差网络作为卷积神经网络中的经典模型, 受到了研究者的广泛关注, 因此产生了多种衍生模型。同时, 手势识别也是当前的热点研究领域, 在利用残差网络实现手势识别方面已有大量研究成果。本文利用了多种残差网络模型的衍生模型, 对 ASL 手势数据集进行训练, 得到了不同模型下的实验结果。其中, 训练结果最好的模型是 ResNet18\_v1, 它的识别正确率最高可达到 93.3%。研究结果表明: 在残差网络的衍生模型中, 所堆叠的卷积层数越多, 对准确率的提升效果不一定越强, 需要根据任务要求, 灵活选择模型并应用。

**关键词:** 手势识别; 卷积神经网络; 残差网络

## Implementation of gesture recognition algorithm on residual network

HAO Yuzhe<sup>1</sup>, YUAN Tianfu<sup>1</sup>, TIAN Haiyue<sup>2</sup>

(1 School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;

2 School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

**[Abstract]** Residual network, as a classic model in convolutional neural networks, has received extensive attention from researchers, and a variety of derivative models have been produced. At the same time, gesture recognition is also a current hot research field, and there are a lot of research results in the use of residual network to realize gesture recognition. In this paper, a variety of derivative models of residual network models are used to train the ASL gesture data set, and experimental results under different models are obtained. The model with the best training result is ResNet18\_v1, and its recognition accuracy can reach 93.3%. The research results show that in the derivative model of the residual network, more convolutional layers stacked do not equal better accuracy. Therefore, the model needs to be flexibly selected and applied according to the task requirements.

**[Key words]** Gesture recognition; Convolutional neural network; Residual Network

## 0 引言

深度学习以其优异的性能越来越多地被应用在手势识别领域。王龙等人提出了一种结合肤色模型和卷积神经网络的手势识别方法<sup>[1]</sup>; 吴杰对比了 5 种不同的卷积神经网络的效果, 并提出了一种 CNN 和 RBM 联合网络<sup>[2]</sup>; 易生为克服传统 2D 卷积神经网络识别准确度较低的缺点, 提出了一种基于多列深度 3D 卷积神经网络的手势识别方法<sup>[3]</sup>。由于卷积神经网络的层数越来越深, 为了缓解了卷积神经网络中的网络退化问题, 何恺明等人在残差网络基础上, 讨论了各个残差单元对网络性能的影响, 做出了改进, 提出了 ResNet\_v2 模型<sup>[4]</sup>; 梁智杰等人提出了一种结合宽残差和长短时记忆网络的手势识别方法, 并取得了良好的效果<sup>[5]</sup>。熊才华等人为解决不同环境下光照强度的变化对手势识别准确率影响, 提出了一种基于残差网络的改进 Faster R-CNN 手势识别算法<sup>[6]</sup>。

## 1 残差网络 (Residual Network, ResNet)

由于对深度学习的认知加深, 越来越多的研究者认识到, 在一定条件下, 增加网络深度有助于增强网络的特征表达能力, 从而提高识别的准确率。但是同时, 网络深度的过度拓展不可避免的会带来网络的退化, 参数的成倍提升, 梯度的消失以及梯度的爆炸等问题。残差网络便应运而生。

(1) 残差块 (ResBlock)。残差块是 ResNet 的基本单元, 图 1 中每一个这样的单元称作 BasicBlock, 而在残差网络中的基本单元是 ResBlock, 每一个 ResBlock 是由若干个 BasicBlock 组成。假设输入为  $x$ , 输出为  $H(x)$ , 经过常规卷积运算输出为  $F(x)$ , 这就是网络要学习的内容, 它要比  $H(x)$  更容易学习。图 1 中, ResNet\_v1 与 ResNet\_v2 的计算顺序上有很大不同, 调整了激活函数与卷积操作的位置, 有效地防止了梯度消失的问题, 便于正则化。

**基金项目:** 国家大学生创新项目 (201910856009)。

**作者简介:** 郝禹哲 (1998-), 男, 本科生, 主要研究方向: 图像处理与模式识别; 袁天夫 (1963-), 男, 硕士, 副教授, 主要研究方向: 通信信号处理、嵌入式系统。

**通讯作者:** 袁天夫 Email: yuan.tf@163.com

收稿日期: 2020-04-08

(2) BN (Batch Normalization)。神经网络的训练过程本质是学习输入数据的分布。在训练之前, 一般会将图像标准化到 [0, 1]。为了解决输入数据分布不统一的问题, 文采用 BN 对每一批次的数据进行批量标准化, 其具体公式如 (1) ~ 公式 (4):

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i, \quad (1)$$

$$\sigma_B = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2, \quad (2)$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}, \quad (3)$$

$$y_i = \gamma \hat{x}_i + \beta. \quad (4)$$

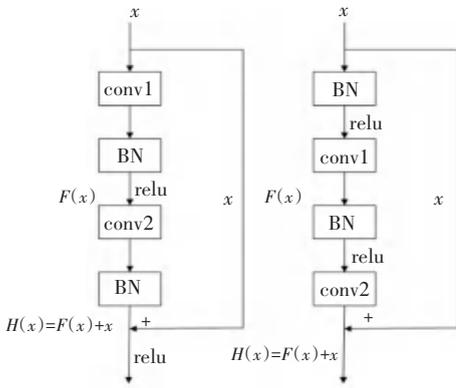


图 1 BasicBlock 结构图

Fig. 1 BasicBlock structure diagram

(3) GAP 层。GAP (Global Average Pooling) 可以代替全连接层, 减了模型的参数, 并不受输入大小的限制。它对输入的特征图在每个通道上计算像素的均值, 再将这些均值组成一个特征向量进行分类。假设 GAP 层的输入为  $28 \times 28 \times 256$ , 得到的就是  $1 \times 256$  的特征向量, 如图 2 所示。

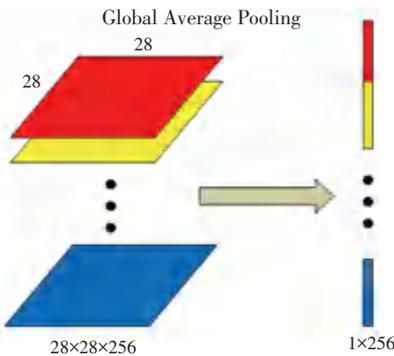


图 2 Global Average Pooling 示意图

Fig. 2 GAP schematic diagram

## 2 实验结果与分析

### 2.1 实验环境

本文实验基于 Windows10 操作系统, 利用

Tensorflow2.0 实验环境下完成。如图 3 所示, 采用的是 ASL (American Sign Language) 数据集, 包含 29 种手势图像, 每一类 3000 张。



图 3 手势训练集

Fig. 3 Gesture training dataset

### 2.2 残差网络参数

本文选取的数据集的输入为  $200 \times 200 \times 3$ , 按照 7:2:1 的比例将原始数据集划分为训练集、验证集、测试集。本文采用 6 种方式对数据集进行训练, 并比较其中各种方法的优劣, 见表 1, 每一种模型用 ResNet\_v1 与 ResNet\_v2 两种方式进行训练。

表 1 网络参数配置

Tab. 1 Network parameter configuration

模型	BasicBlock1 数量	BasicBlock2 数量	BasicBlock3 数量	BasicBlock4 数量
ResNet18	2	2	2	2
ResNet34	3	4	6	3
ResNet50	3	6	12	3

### 2.3 实验结果

为了对比 ResNet 结构和层数对于训练结果准确率的影响, 经过训练, 将 6 种模型训练过程的准确度绘制, 如图 4 所示, 其中 ResNet18\_v1, ResNet18\_v2, ResNet50\_v2 的训练准确度最高, 它们的准确率可以稳定在一个较大的准确率, 说明训练好的网络模型具有较好的稳定性。

为了评估不同模型对测试集图片的识别性能优劣, 将测试集作为输入, 利用训练好的网络模型对测试集手势图像进行检测, 不同模型的测试准确率如图 5 所示。从图 5 可以看出 ResNet18\_v1 的准确率最高, 可以达到 93.3%, ResNet18\_v2, ResNet50\_v2 的准确度次之, 均可以达到 90% 左右, 说明这 3 种模型均能有效地对手势进行识别。

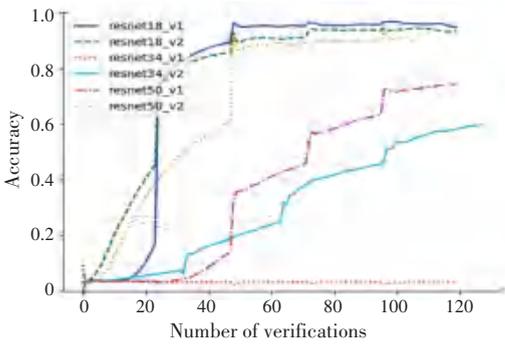


图4 训练准确度变化曲线

Fig. 4 Training accuracy curve

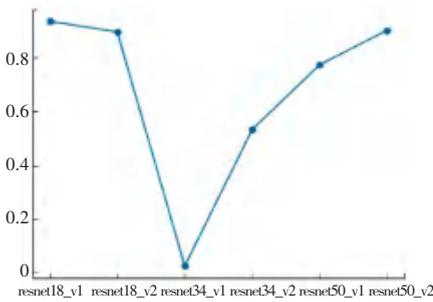


图5 测试准确度变化曲线

Fig. 5 Testing accuracy curve

(上接第63页)

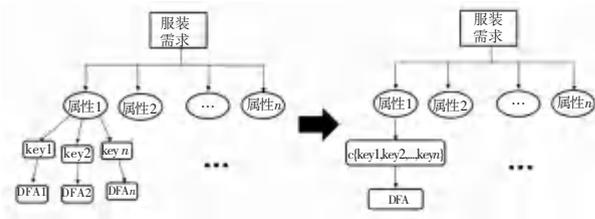


图3 DFA与服装需求对应关系

Fig. 3 Correspondence between DFA and clothing demand

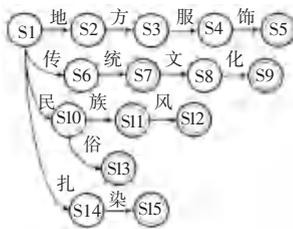


图4 构建出的有限状态自动机

Fig. 4 Finite state automata constructed

### 4 结束语

服装个性化定制的需求是一个抽象的描述,并没有准确的分类方法,行业内专家可以根据服装面料,生产工艺等属性和个人喜好对服装需求进行评判,而普通的消费者在提出定制需求时会根据以往服装购买经验和个人喜好进行抉择。本文通过对多

### 3 结束语

本文比较了多种残差网络模型的优劣。实验结果表明,残差网络的深度不一定与识别准确率呈正相关,即使是残差网络,也会随模型层数的增多,出现不同程度的退化。所以需要根据不同的任务,灵活地选用模型。本文中 ResNet18\_v1 的识别正确率可达到 93.3%,具有良好的性能。在未来工作中,将进一步考虑多背景和复杂背景下的手势识别问题。

### 参考文献

- [1] 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.
- [2] 吴杰. 基于深度学习的手势识别研究[D]. 电子科技大学,2015.
- [3] 易生,梁华刚,茹锋. 基于多列深度3D卷积神经网络的手势识别[J]. 计算机工程,2017,43(8):243-248.
- [4] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks[C]//European conference on computer vision. Springer, Cham, 2016: 630-645.
- [5] 梁智杰,廖盛斌. 融合宽残差和长短时记忆网络的动态手势识别研究[J]. 计算机应用研究, 2019, 36(12):3846-3852.
- [6] 熊才华, 巩言丽, 廉华, 等. 基于 ResNet-50 改进的 Faster R-CNN 手势识别算法[J]. 计算机时代, 2019, 327:7-10.

模式匹配算法、词性标注以及有限状态自动机等理论的研究,提出了一种能够抽取标准化服装定制需求的方法。随着行业的发展和研究的深入,将产生更多真实的服装定制需求描述样本,本方法结合其他模型进行优化,会为服装个性化需求提取服务提供更加高效的解决方案。

### 参考文献

- [1] 丁纯,李君扬. 德国“工业4.0”:内容、动因与前景及其启示[J]. 德国研究,2014,29(4):49-66,126.
- [2] 倪玉婷. 工业化服装定制,应该这样做——标准起草人解读服装定制通用技术规范[J]. 中国纤检,2019(3):80-81.
- [3] 刘静. 基于互联网+时代下服装个性化定制的设计与实现[J]. 西部皮革,2018,40(24):21.
- [4] 李炜. 基于情感语义的服装推荐系统研究与实现[D]. 东华大学,2013.
- [5] 潘璐. 基于层次分析法的服装推荐专家系统[D]. 东华大学,2016.
- [6] 余思雅,周莉英. 探析服装风格的流行与传播[J]. 现代装饰(理论),2013(9):126.
- [7] 张海花. 服装风格与传播研究[D]. 北京服装学院,2010.
- [8] Erhard Rahm, Philip A. Bernstein. A survey of approaches to automatic schema matching[J]. The VLDB Journal,2001,10(4).
- [9] 李伟男,鄂跃鹏,葛敬国,等. 多模式匹配算法及硬件实现[J]. 软件学报,2006(12):2403-2415.