计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2020,29(5):189−195 [doi: 10.15888/j.cnki.csa.007374]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 SVM 与 Inception-v3 的手势识别①

吴斌方[1], 陈　涵[1], 肖书浩[2]

[1](湖北工业大学 机械工程学院, 武汉 430068)
[2](武昌首义学院 机电工程研究所, 武汉 430064)
通讯作者: 陈　涵, E-mail: 231633388@qq.com

摘　要: 针对传统机器视觉的手势识别方法识别准确率低, 抗干扰能力差等问题, 提出了一种基于支持向量机(Support Vector Machine, SVM) 手势分割和迁移学习的静态手势识别方法. 本文使用 SVM 和迁移学习方法相结合构建新的手势识别模型, 利用 SVM 对样本进行手势分割, 将 Inception-v3 模型作为卷积神经网络模型基础, 对网络参数进行 fine-tuning, 将预先经过手势分割处理后的样本导入模型训练, 调整超参数得到新的最优手势识别模型, 并在一定干扰环境下测试, 得到测试结果. 测试结果表明该方法识别准确率和实时反馈效率均高于传统方法, 能高效识别手势, 满足实际应用需求.

关键词: 支持向量机; 手势分割; 迁移学习; 手势识别

引用格式: 吴斌方,陈涵,肖书浩.基于 SVM 与 Inception-v3 的手势识别.计算机系统应用,2020,29(5):189–195. http://www.c-s-a.org.cn/1003-3254/7374.html

Gesture Recognition Based on SVM and Inception-v3

WU Bin-Fang[1], CHEN Han[1], XIAO Shu-Hao[2]

[1](School of Mechanical Engineering, Hubei University of Technology, Wuhan 430068, China)
[2](Institute of Mechanical and Electrical Engineering, Wuchang Shouyi University, Wuhan 430064, China)

Abstract: Aiming at the problems of low recognition accuracy and poor anti-interference ability of traditional machine vision gesture recognition methods, a static gesture recognition method based on Support Vector Machine (SVM) gesture segmentation and transfer learning is proposed. This study uses SVM and transfer learning method to build a new gesture recognition model, uses SVM to segment the sample gesture, uses the Inception-v3 model as the basis of Convolutional Neural Network (CNN) model, carries out fine tuning on the network parameters, imports the sample processed by gesture segmentation into the model training, adjusts the super parameters using fine-tuning to get the new optimal gesture recognition model. The test results, obtained in disturbed environment, show that the recognition accuracy and real-time feedback efficiency of this method are higher than those of traditional methods, which can effectively recognize gesture and meet the practical application requirements.

Key words: Support Vector Machine (SVM); hand segment; transfer learning; hand-gesture recognition

引言

近年来, 机器学习 (machine learning) 领域快速发展, 图像识别技术日益成熟, 人机交互方式也随之改变. 手势识别是人机交互中最简单、最直观的一种交互方式. 该方式摆脱键盘、鼠标、按键等硬件束缚, 具有简单易学、操作方便、动作直观等特点, 极大增加用户

① 基金项目: 湖北省自然科学基金 (2018CFC810)
Foundation item: Natural Science Foundation of Hubei Province, China (2018CFC810)
收稿时间: 2019-09-25; 修改时间: 2019-10-22; 采用时间: 2019-10-30; csa 在线出版时间: 2020-05-07

体验感和人机互动性.

手势识别技术可分为两种: 基于数据手套技术和基于机器视觉技术[1]. 国内外学者针对机器视觉手势识别技术都有相应的研究与发展, Mahmoud 等[2]利用 YCbCr 颜色空间和深度信息结合高斯混合概率模型 (GMM) 计算手部区域, 利用隐马尔可夫模型 (HMM) 进行手势识别. Saha 等[3]利用 Kinect 传感器采集数据和隐马尔可夫模型 (HMM) 进行手势识别; 隐马尔可夫模型 (HMM) 是手势识别领域常用的方法, 该方法需要大量参数, 对识别时间和效率有一定影响. Tusor 等[4]利用模糊神经网络 (FNN) 根据预处理后的手势数据建立手势特征模型, 用模糊推理进行手势识别; 该方法网络层数浅, 学习能力较弱, 训练过程容易出现过拟合, 识别效果不好. Marin 等[5]利用 Leap motion 传感器和深度相机提取指尖间角度、距离和空间坐标等参数作为手部特征, 将特征馈送到 SVM 和随机森林进行手势识别, 该方法对硬件的要求高, 样本预处理较为复杂. 任彧等[6]运用方向梯度直方图 (Histograms of Oriented Gradient, HOG) 提取手势特征, 利用 SVM 学习识别手势, 消除了光照和手部旋转对手势识别的影响, 但背景要求单一, 识别准确率不高. 朱越等[7]利用 HSV 和 RGB 颜色空间联合进行手势分割, 根据手势轮廓像素变化判断手势, 该模型对肤色的抗干扰能力较差, 适应面窄, 识别种类局限. 操小文等[8]利用 8 层的卷积神经网络对手势样本进行训练和识别, 该方法需花费长时间设计定义网络模型, 且样本背景需单一, 抗干扰能力差.

本文针对手势识别提出了一种 SVM 手势分割与迁移学习相结合的方法, 利用 SVM 对样本进行手势分割, 采用迁移学习方法将 Inception-v3 模型进行 fine-tuning, 通过实验对比获得最优性能的超参数, 得到新的手势识别网络模型. 本文使用 SVM 对样本手势分割增加了手势分割的鲁棒性和强适应性, 消除了肤色、光照、旋转和背景等因素的干扰, 运用迁移学习简化定义和设计 CNN 的工作, 节省大量网络设计和网络训练时间, 构建的模型在识别准确率和识别效率上均有一定提升.

1 卷积神经网络与迁移学习

卷积神经网络是一种深度前馈神经网络, 是深度学习领域的一个重要分支[9], 它在图像处理领域表现出优越的性能. 卷积神经网络将原始图像信息分块处理, 能适应图像特征的平移旋转, 且分块处理特征信息后参数明显减少, 对提高模型学习效率有显著影响.

传统机器学习方法学习训练的过程需要庞大的训练数据集, 且测试数据集的数据分布需与训练数据集相同. 在大数据时代背景下, 轻松获取所需领域且满足工作任务需求的庞大数据集仍存在一定的难度. 另一方面, 在监督学习完成学习任务时, 需要大量的人工将训练数据集进行逐一对应的标注, 耗费大量的人力物力, 对于一般的高校的机器学习研究或者小型公司的机器学习技术开发都有极大的障碍. 测试数据集的数据分布亦常难以与训练数据集的数据分布一致, 给传统的机器学习方法带来一定的难度.

迁移学习可以很好的解决上述问题, 迁移学习是运用已存有的知识对不同但相关的领域的问题求解的一种新的机器学习方法[10]. 迁移学习在源领域模型上仅需少量的训练数据集便可以建立一个针对目标领域的新模型, 对数据分布不同的目标领域进行预测和分析.

2 实验数据准备与预处理

2.1 样本采集

网络上的手势数据集较匮乏且不满足实际需求, 遂利用实验室设备采集手势数据集, 采集 10 类不同手势各 250 张, 共 2500 张手势样本, 取出每种手势样本中的 25 张作为验证数据集. 采集到的 2500 张样本进行尺寸归一化处理, 得到 2500 张 640×360 像素的手势样本, 提高手势识别准确率. 部分样本如图 1 所示.

2.2 样本增强处理

将 2500 张样本随机高斯模糊化处理, 将样本与二维的高斯分布的概率密度函数作卷积, 随机模糊样本.

$$G(u,v)=\frac{1}{2\pi\sigma^2}e^{-(u^2+v^2)/(2\sigma^2)} \tag{1}$$

其中, σ 为正态分布的标准偏差, (u^2+v^2) 为模糊半径 r 的平方.

对样本进行旋转偏移处理, 随机对样本添加少量噪声, 增强网络识别的鲁棒性, 防止网络产生过拟合现象对测试结果造成影响.

2.3 SVM 手势分割

手势识别的重要的一个步骤是对样本中的手势进行分割, 提取重要的感兴趣区域信息, 剔除多余背景和环境对识别准确率造成的干扰. 传统的通过肤色阈值对手势进行分割的方法一般有 2 种, 在 RGB 空间肤色

阈值分割、HSV 空间肤色检测.

RGB 空间肤色阈值分割中 R(0~255), G(0~255), B(0~255)3 种像素值同时满足式 (2), 式 (3) 则为肤色.

$$\begin{cases}(R>95)\cap(G>40)\cap(B>20)\\ \max(R,G,B)-\min(R,G,B)>15\\ |R-G|>15\cap(R>G)\cap(R>B)\end{cases} \tag{2}$$

$$\begin{cases}(R>220)\cap(G>210)\cap(B>170)\\ (R-G)\le 15\cap(R>G)\cap(R>B)\end{cases} \tag{3}$$

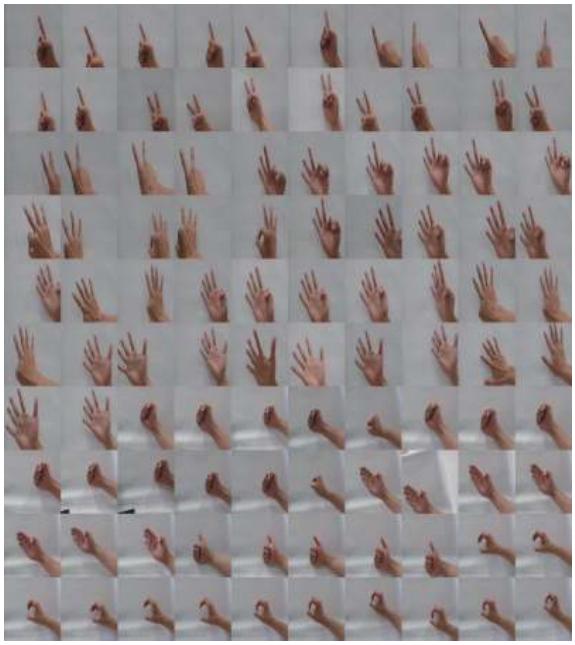


图 1 样本采集

当背景颜色与肤色相同或相似时, 会对肤色分割造成一定干扰, 对背景要求较高.

HSV 空间肤色建模要求 H(色调 Hue), S(饱和度 Saturation), V(亮度 Value) 满足式 (4) 则为肤色.

$$\begin{cases}(0^0\le H\le 25^0)\cup(335^0\le H\le 360^0)\\ (0.2\le S\le 0.6)\cap(0.4\le V)\end{cases} \tag{4}$$

H、S、V 三通道的值对应 HSV 空间中的某一点, 实际环境中光照的亮度会带来色调的改变, 对光照强度的强适应性给肤色检测带来一定噪声.

本文利用 SVM 将手部区域与背景区分开, 形成手势分割. SVM 在对图像的二分类处理问题有出色的表现, 泛化能力较强. 其基本的思想是将在低维空间非线性可分的两类映射到高维空间, 求解出一个超平面 (hyper lane) 在高维空间线性可分的两类数据完成分类. 本次实验采用线性 (liner) 核函数 (kernel): $K(x, y)=x^{\mathrm{T}}\cdot y$, 目标函数惩罚系数 C=1.0. 利用 Python 的 Tkinter 模块编写可视化界面对样本进行标记, 将手部区域与背景区域区分开, 如图 2 所示, 其中红色标记为手部区域, 绿色点标记为背景区域.

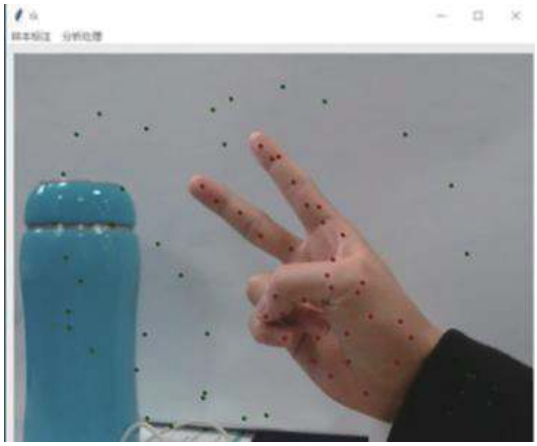


图 2 SVM 样本标记

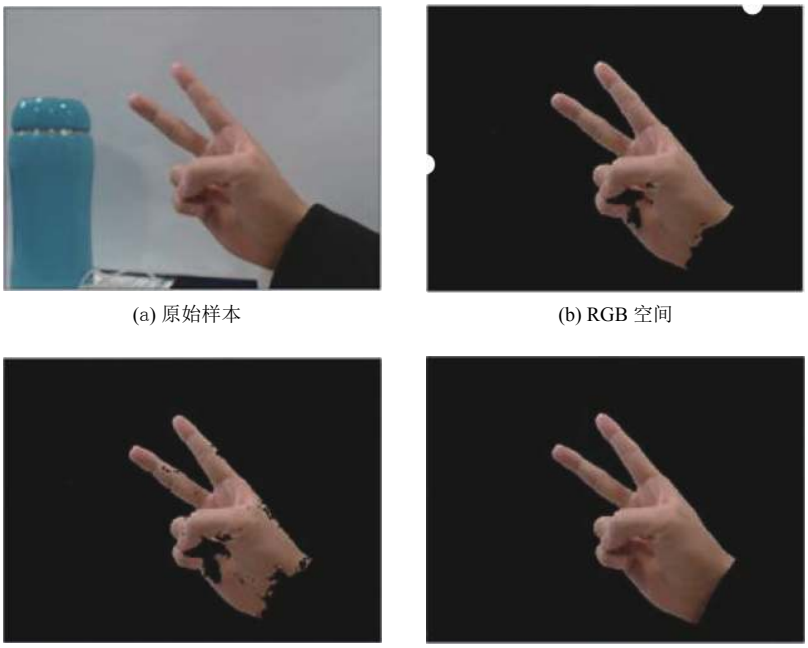
标记完成后利用 SVM 学习并显示结果, 各手势分割方法分割结果如图 3 所示. 通过对比实验效果, 上述前两种方法都有一定局限性, RGB 空间肤色阈值分割只有在背景单一和光照稳定的条件下肤色分割效果较好, 有一定的局限性, 如图 3(b) 所示; HSV 空间肤色建模对光线的鲁棒性较强, 但分割离散, 肤色区域不连续, 无法分割出完整手势, 如图 3(c) 所示. 利用 SVM 学习后对手势进行分割的效果明显优于基于 RGB 和 HSV 肤色分割的方法, 如图 3(d) 所示, 不仅在手部区域连续性较好, 对环境的要求也较低, 且该方法鲁棒性和灵活性将强, 对于肤色区别较大的实验者只需重新学习即可得到满足需求的手势样本.

利用训练好的 SVM 模型对 2500 张样本进行批量手势分割处理, 最后将手势分割后的样本选取适当的全局阈值, 经过全局二值化处理, 得到 2500 张手势二值样本, 如图 4 所示.

3 实验方法

本次实验采用迁移学习方法将 Inception-v3[11]模型结构作为训练模型的结构基础. Inception-v3 模型由谷歌提出, 其网络思想与其他深度网络主要有几点不同, 一方面网络使用更小的卷积核代替尺寸相对较大的卷积核, 例如将两个 3×3 的卷积核代替一个 5×5 的卷积核. 另一方面网络将 $n\times n$ 例如 3×3, 7×7 的二维卷

积拆分成两个 1×n, n×1 例如 3×1, 1×3 和 7×1, 1×7 的二维卷积, 这种方式让网络参数量大大减少, 在加快运算速度的同时也减少了过拟合的情况, 且这种对卷积结构的不对称拆分使特征空间保留完整, 网络非线性表达的能力也更强. 网络使用 Batch Nomalization (BN) 算法[12], 通过规范化方法将输入分配到均值为 0 方差为 1 的正态分布, 有效解决深层网络的梯度消失问题, 大幅增加训练效率和收敛后的样本分类的准确率.



(a) 原始样本 (b) RGB 空间

(c) HSV 空间 (d) SVM

图 3 各方法手势分割结果

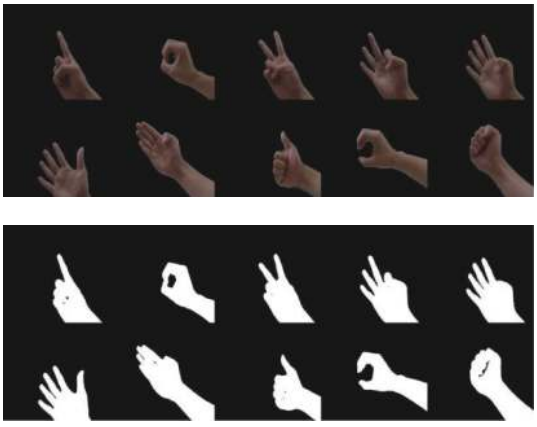


图 4 二值化处理样本

将 Inception-v3 中 Softmax 回归层一维输出大小从 1000 类定义为所需识别手势的 10 类, 保留除 Softmax 层外所有层的参数, 将网络的底层作为一个特征提取器, 只训练最后一层参数达到模型能够识别 10 类手势的目的. 模型通过标签平滑方式进行模型正则化, 首先对于输入的手势样本 x, 使用式 (5) 计算对应标签的概率.

$$p(k|x)=\frac{\exp(z_k)}{\sum_{i=1}^{K}\exp(z_i)} \tag{5}$$

其中, k 为手势标签类别, Z_i 为尚未归一化的对数概率. 手势样本在对应标签上在分布为 $q(k|x)$, 将样本损失定义为交叉熵损失函数:

$$l=-\sum_{k=1}^{K}\log(p(k))q(k) \tag{6}$$

最小化交叉熵等价于最大化标签对数似然期望, 其梯度为:

$$\frac{\partial l}{\partial z_k}=p(k)-q(k) \tag{7}$$

用 $q'(k|x)$ 代替标签分布 $q(k|x)=\delta_{k,y}$:

$$q'(k|x)=(1-\varepsilon)\delta_{k,y}+\frac{\varepsilon}{K} \tag{8}$$

其中, $\delta_{k,y}$ 为狄拉克 δ 函数, $(1-\varepsilon)$ 与 ε 分别为实际分布和固定分布的权重.

Fine-tuning 微调是训练深度卷积神经网络的技巧之一, 原理是采用模型原有参数作为网络的初始化参数, 冻结部分网络层, 降低学习效率, 以目标数据作为

输入在原有参数基础上训练参数. Fine-tuning 后的网络模型更易训练, 节省大量训练时间, 精度会相较直接随机初始化参数的网络有所提高.

完成定义网络模型后将手势分割后的二值化样本导入模型训练, 冻结网络 175 层参数, 调整网络超参数得到不同准确率和交叉熵损失函数曲线.

4 实验与结果

4.1 实验

本次实验环境为 Windows 10 操作系统, 采用 GTX1060 显卡在 Tensorflow 深度学习框架下完成实验.

为增强模型性能, 快速达到收敛, 让网络具有较好的识别效果和节省适当的训练、识别时间, 调整网络模型的一系列超参数作纵向对比实验. 训练集训练与测试集测试均在 GPU 加速环境下运行. 模型采用 RMSPro[13]梯度下降算法, 设置衰减值为 0.9, ε=1.0.

设定默认批次大小 (batch size) 为 64, 学习效率 (learning rate) 为 0.045, 迭代次数 (epoches) 为 2000, 得到迭代次数与准确率和损失函数的关系. 由图 5 看出, 模型在 1000 次迭代后基本趋于稳定, 准确率随迭代次数增长的波动较小, 损失函数也基本趋于稳定, 考虑迭代次数增加对网络模型产生过拟合情况的影响, 迭代次数恒定设置为 1000.

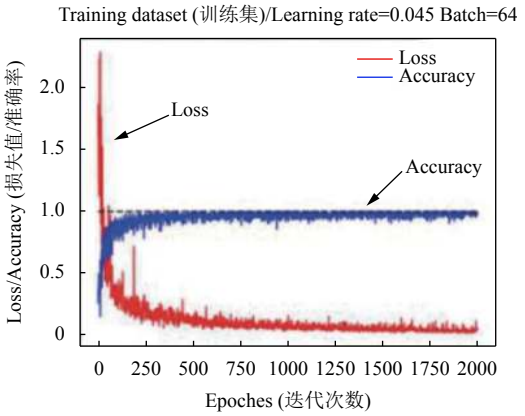


图 5 迭代次数对模型的影响

对比学习效率对模型识别准确率和损失函数的影响, 学习效率对模型训练起至关重要的作用, 较低的学习效率导致模型收敛速度较慢, 训练时间较长; 而较高的学习效率则可能会导致模型不收敛, 损失函数值波动较大. 因此设定 4 种分别为 0.001, 0.015, 0.045, 0.080 不同的学习效率对模型进行训练, 对比结果如图 6 所示.

从图 6 可以看出, 图 6(a) 中损失函数和准确率函数图像震动剧烈, 准确率较低, 损失函数值始终高于 0.5, 网络模型收敛情况较差; 图 6(b) 中图像震动有些许减少, 相较图 6(c)、6(d) 波动任较大. 图 6(c)、6(d) 在迭代初期损失函数值较大, 迭代后期收敛明显且图像震动幅度较小, 适宜训练模型, 对比图 6(c)、6(d) 对训练集的平均准确率和损失函数值, 选择图 6(c) 学习效率 0.045作为最终模型学习效率.

最后探究每次迭代样本批次大小 (batch size) 的选择, 不同的批次大小会对网络模型的准确率和训练效率产生影响, 选择过小的批次会导致准确率震荡较大, 模型无法收敛; 选择较大的批次会导致内存容量不足, 参数更新缓慢, 降低运行速度, 徒增训练时间. 实验选择批次大小为 32、64、128 和 256 等 4 种不同批次大小作对比, 得到结果如表 1.

从表 1 得出, 批次大小为 32 和 64 的模型在训练时间上与大批次模型相比有一定的优势, 准确率与批次大小为 128 的模型相比有略微差距, 综合考虑选择批次大小为 128 个样本导入网络模型完成训练.

4.2 实验结果

网络模型训练完成后对实时拍摄不同测试者手势样本进行预测, 选取 3 位测试者 10 种不同手势各 50 张共 1500 张样本, 收集手势识别预测结果, 如表 2 所示, 可知, 本文利用迁移学习训练卷积神经网络模型对实时获取的静态手势识别准确率较理想, 平均准确率达到 96.3%, 平均识别时间达到 39.2 ms, 在识别速度上基本满足实际应用需求.

4.3 方法对比与分析

将本文方法与传统的手势识别方法做对比, 验证本文提出算法模型性能, 对比结果如表 3 所示.

文献[2]中方法参数量大, 计算速度慢导致模型识别速度慢效率低. 文献[6]中方法对于手势区域在整体图像所占比例的干扰较大. 文献[7]识别速度快, 无法排除手指并拢的手势干扰, 识别准确率较低. 文献[8]中能排除一定噪声干扰, 网络结构较简单泛化性能较差. 本文方法在识别准确率突出, 识别反馈速度较快, 内因在于采用 SVM 进行手势分割后得到的二值化样本纯净, 特征利于网络学习, 迁移学习方式构建的网络模型层数深, 对特征的学习能力强, 善于分类, 在模型设计和参数训练的时间上相较传统卷积神经网络花费少, 泛化性能强.

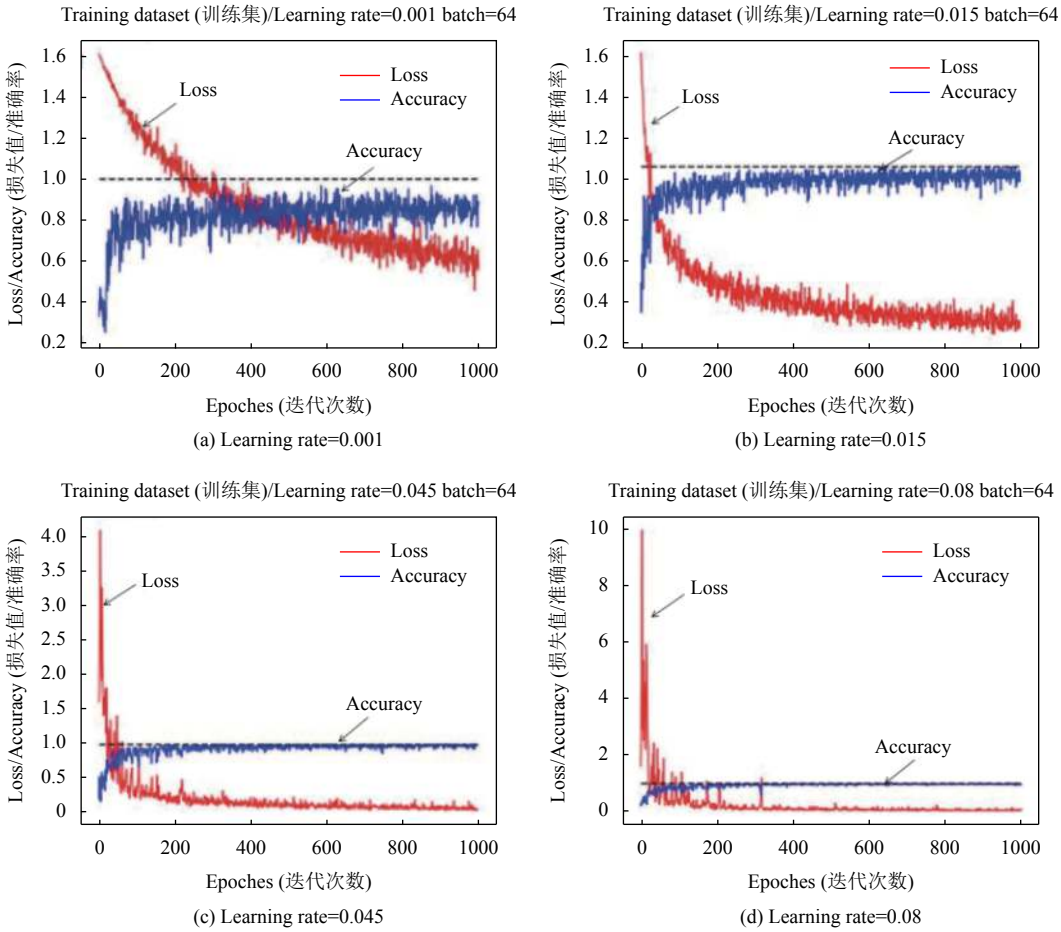


图 6 学习效率对模型的影响

表 1 不同批次大小模型训练结果

模型批次大小	训练时间 (ms)	准确率 (%)
32	53 430	98.3
64	102 713	98.6
128	202 724	99.3
256	406 234	97.8

表 2 测试结果

手势类别编号	识别准确率 (%)	平均准确率 (%)	平均识别时间 (ms)
1	98.0	96.3	39.2
2	97.3		
3	98.0		
4	93.3		
5	97.3		
6	96.7		
7	92.7		
8	98.0		
9	93.3		
10	98.6		

表 3 各种手势识别算法对比

实验方法	平均识别准确率 (%)
HMM[2]	94.7
HOG+SVM[6]	93.7
像素变化[7]	85.9
CNN[8]	95.2
本文方法	96.3

5 结论

本文方法将支持向量机和迁移学习相结合, 利用 SVM 进行手势分割取得的效果相比其他颜色空间手势分割方式的效果较好, 具有较好的鲁棒性和灵活调整能力; 利用迁移学习将已训练好的卷积神经网络作为基础, 训练全连接层参数, 需要的训练数据集较少, 大量缩减卷积神经网络构建和网络的训练时间, 取得 96.3% 的平均识别准确率和 39.2 ms 的平均识别反馈时间, 基本能满足实际应用需求.

参考文献

1 焦家祥. 手势识别技术前沿概述. 电子世界, 2018, (15): 29–30.

2 Elmezain M, Al-Hamadi A, Michaelis B. Real-time capable system for hand gesture recognition using hidden Markov models in stereo color image sequences. Journal of WSCG, 2008, 16(1): 65–72.

3 Saha S, Lahiri R, Konar A, *et al*. HMM-based gesture recognition system using kinect sensor for improvised human-computer interaction. Proceedings of 2017 International Joint Conference on Neural Networks. Anchorage, AK, USA. 2017. 2776–2783.

4 Tusor B, Varkonyi-Koczy AR. Circular fuzzy neural network based hand gesture and posture modeling. Proceedings of 2010 IEEE Instrumentation & Measurement Technology Conference. Austin, TX, USA. 2010. 815–820.

5 Marin G, Dominio F, Zanuttigh P. Hand gesture recognition with jointly calibrated Leap Motion and depth sensor. Multimedia Tools and Applications, 2016, 75(22): 14991–15015. [doi: 10.1007/s11042-015-2451-6]

6 任彧, 顾成成. 基于 HOG 特征和 SVM 的手势识别. 科技通报, 2011, 27(2): 211–214. [doi: 10.3969/j.issn.1001-7119.2011.02.012]

7 朱越, 李振伟, 杨晓利, 等. 基于视觉的静态手势识别系统. 计算机技术与发展, 2019, 29(2): 69–72.

8 操小文, 薄华. 基于卷积神经网络的手势识别研究. 微型机与应用, 2016, 35(9): 55–57, 61.

9 李彦冬, 郝宗波, 雷航. 卷积神经网络研究综述. 计算机应用, 2016, 36(9): 2508–2515, 2565.

10 庄福振, 罗平, 何清, 等. 迁移学习研究进展. 软件学报, 2015, 26(1): 26–39. [doi: 10.13328/j.cnki.jos.004631]

11 Szegedy C, Vanhoucke V, Ioffe S, *et al*. Rethinking the inception architecture for computer vision. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 2818–2826.

12 Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on International Conference on Machine Learning. Lille, France. 2015. 448–456.

13 Tieleman T, Hinton G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning, 2012, 4: 26–31.