

基于改进YOLOv4-tiny算法的手势识别

卢迪* 马文强

(哈尔滨理工大学 哈尔滨 150080)

摘要: 随着人机交互的发展, 手势识别越来越重要。同时, 移动端应用发展迅速, 将人机交互技术在移动端实现是一个发展趋势。该文提出一种改进YOLOv4-tiny的手势识别算法。首先, 在YOLOv4-tiny网络基础上, 添加空间金字塔池化(SPP)模块, 融合了图像的局部和全局特征, 增强网络的准确定位能力。其次, 在YOLOv4-tiny原网络的3个最大池化层和新增SPP模块后各添加一个 1×1 的卷积模块, 减少了网络的参数, 提高网络的预测速度。在此基础上, 利用K-means++算法生成适合检测手势的先验框, 加快网络检测手势。在手势数据集NUS-II上, 与YOLOv3-tiny算法和YOLOv4-tiny算法进行对比, 改进算法平均精度均值(mAP)为100%, 每秒传输帧数(fps)为377, 可以快速准确地检测识别手势。将该文改进算法部署在安卓(Android)移动端, 实现了移动端实时的手势检测与识别, 对人机交互的发展有很大的研究意义。

关键词: 手势识别; 人机交互; YOLOv4-tiny; 安卓

中图分类号: TN911.73

文献标识码: A

文章编号: 1009-5896(2021)11-3257-09

DOI: [10.11999/JEIT201047](https://doi.org/10.11999/JEIT201047)

Gesture Recognition Based on Improved YOLOv4-tiny Algorithm

LU Di MA Wenqiang

(Harbin University of Science and Technology, Harbin 150080, China)

Abstract: With the development of human-computer interaction, gesture recognition is becoming more and more important. At the same time, mobile terminal applications are developing rapidly, it is a development trend to implement human-computer interaction technology on the mobile terminal. An improved YOLOv4-tiny gesture recognition algorithm is proposed. Firstly, on the basis of YOLOv4-tiny network, the Spatial Pyramid Pooling(SPP) module is added to integrate the local and global features of the image to enhance the accurate positioning ability of the network. Secondly, a 1×1 convolution is added after the 3 maximum pooling layers of the original YOLOv4-tiny network and the newly added SPP module, which reduces the network parameters and improves the prediction speed of the network. On this basis, the K-means++ algorithm is used to generate an anchor box suitable for detecting gestures to speed up the network detection of gestures. In the gesture dataset NUS-II, compared with the YOLOv3-tiny algorithm and the YOLOv4-tiny algorithm, the improved algorithm mean Average Precision(mAP) is 100%, frames per second (fps) is 377, which can detect and recognize gestures quickly and accurately. The improved algorithm of this paper is deployed on the Android mobile terminal to realize the real-time gesture detection and recognition on the mobile terminal, which has great research significance for the development of human-computer interaction.

Key words: Gesture recognition; Human computer interaction; YOLOv4-tiny; Android

1 引言

在人机交互领域, 手势有很多种应用场景, 如游戏控制、智能家居、手语实时翻译等^[1]。因此, 对于手势的快速准确检测和识别十分重要。自20世纪末, 国外就开始了基于手势识别交互的研究, 主要分为基于外部设备的手势识别和基于计算机视觉的手势识别。相较于基于外部设备的手势识别方

式, 基于计算机视觉的手势识别方式更为方便灵活、降低了设备成本。传统的基于计算机视觉的手势识别方法分为手势的检测和分割、手势特征提取与手势识别3个阶段。常常使用肤色分割模型分割出手势区域, 再提取尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)、加速稳健特征(Speeded Up Robust Features, SURF)、方向梯度直方图(Histogram of Oriented Gradient, HOG)、Hu矩等特征, 采用支持向量机(Support Vector Machine, SVM)、Adaboost等机器学习方法进行分

收稿日期: 2020-12-14; 改回日期: 2021-04-15; 网络出版: 2021-04-30

*通信作者: 卢迪 ludizeng@hrbust.edu.cn

类。该类方法对数据集的依赖性大,算法的鲁棒性不强,且人工提取特征的方式过程烦琐复杂,依赖工程人员的自身经验,算法的应用局限性较大。

随着深度学习的研究与发展,基于深度学习的方法识别速度快,准确率高,成为主流的手势识别算法。Oyedotun等人^[2]将图像二值化,利用阈值分割来获取手势区域,使用卷积神经网络(Convolutional Neural Networks, CNN)和编码器进行手势识别。王龙等人^[3]利用肤色模型进行手势检测,并将得到的手势区域重建为灰度图像,最终将手势区域输入CNN进行特征提取和识别。Mohanty等人^[4]不进行手势检测处理,将原始图像直接输入简单的2层CNN进行特征提取和识别。上述手势识别分类算法中,当图像中存在人体肤色等复杂背景干扰时,CNN识别率显著下降。随着目标检测算法的发展,许多学者将手势识别的分类问题转换为目标检测问题,主要分为两种:基于回归的1阶段目标检测方法和基于区域推荐的2阶段目标检测方法。1阶段目标检测算法主要有YOLO (You Only Look Once)^[5-8], SSD (Single Shot multibox Detector)^[9], RetinaNet^[10]和CornerNet^[11]等,2阶段目标检测算法主要有R-CNN (Region-based CNN)^[12], Fast R-CNN(Fast Region-based CNN)^[13], Faster R-CNN (Faster Region-based CNN)^[14]和R-FCN (Region-based Fully Convolutional Networks)^[15]等。Soe等人^[16]用Faster R-CNN算法进行手势识别的研究,采用ZFNet网络作为特征提取的主干网络,网络虽然简单,但不能充分提取到手势的特征,识别率不高,在NUS-II数据集^[17]所有手势的精确率只有90.08%,手势c的精确率更是只有65.82%。常建红^[18]提出了一种改进Faster R-CNN的手势识别算法,在NUS-II手势数据集上精确率达到99.89%,但是由于该算法先经过RPN推荐候选框,再进行识别,属于2阶段目标检测算法,速度较慢。1阶段的目标检测方法比2阶段法检测速度快,更适用于一些实时性要求高的情况。张勋等人^[19]基于SSD算法,使用轻量化的ASSD模型, fps为30,算法可以实时对手势进行检测与识别,平均精度均值为93.3,识别效果不是很好。彭玉青等人^[20]提出一种基于YOLO定位和CNN分类识别的HGDR-Net算法,在复杂背景下取得不错的效果,在处理时间为0.065 s的情况下F1平均值达到98.43%,但算法要先通过YOLO算法定位,再进行识别,识别过程复杂。王粉花等人^[21]用改进YOLOv3-tiny算法进行手势识别, fps达到220,平均精度均值为92.24%,YOLOv3-tiny-T算法识别

速度虽然很快,但是平均精度均值相对较低。Jiang等人^[22]提出一种改进YOLOv4-tiny算法,该算法具有更快的目标检测速度,适合在移动端和嵌入式等运算能力不强的设备开发。

经过上述分析,以上手势检测识别算法都存在一定缺陷,不能在准确率和速度之间取得最好的均衡。基于此问题,本文选取适合在移动端部署的YOLOv4-tiny算法来进行手势的检测与识别,并在此基础上改进算法实现准确率和速度的最佳平衡。

2 相关网络与模块

2.1 YOLOv4-tiny算法

YOLOv4-tiny算法是在YOLOv4方法的基础上设计的,其具有更快的目标检测速度,预测过程与YOLOv4相同。该算法首先调整输入图像的大小,使所有输入图像具有相同的固定大小 $A \times A$ (本文为 416×416)。将 $A \times A$ 大小的图片分为两个尺度的网格进行预测,这两个尺度分别为 $B_1 = A/32$ 和 $B_2 = A/16$ 。如果真实标注框中某个目标的中心坐标落在哪个网格中,那么就由该网格来预测该目标。网络预测得到的输出有两个维度是提取到的特征图的维度,还有一个维度(深度)大小为 $n \times (5+N)$,其中 n 表示每个网格预测的边界框的数量,YOLOv4-tiny中 n 为3; N 表示检测的类别数,5表示4个坐标信息和1个置信度。模型会输出 $(B_1 \times B_1 + B_2 \times B_2) \times n$ 个边界框,最后通过预测置信度和非极大值抑制算法来排除多余的边界框,得到模型的最终检测框。

YOLOv4-tiny方法使用CSPDarknet53-tiny网络作为主干网络来代替YOLOv4方法中使用的CSPDarknet53网络。在特征融合部分,YOLOv4-tiny方法使用特征金字塔网络提取不同尺度的特征并进行融合,提高了目标检测精度。YOLOv4-tiny使用 13×13 和 26×26 两种不同比例的特征图来预测检测结果,YOLOv4-tiny网络结构如图1所示。

2.2 SPP空间金字塔池化

一般卷积神经网络结构中,在卷积层后面通常连接全连接层。而全连接层的特征数是固定的,所以在网络输入的时候,会固定输入图像的大小。在实际中,输入的图像尺寸不能满足要求,需要裁剪和拉伸,但会使图像失真。空间金字塔池化(Spatial Pyramid Pooling, SPP)是由He等人^[23]提出的,用来解决计算机视觉领域中输入图像尺度不同而造成图片失真的问题。如图2(a)所示,SPP空间金字塔池化使用多级大小空间窗口,在同一个特征图的不同尺度上提取特征并做池化操作,能够对任意大小输入图像产生相同尺度的输出。

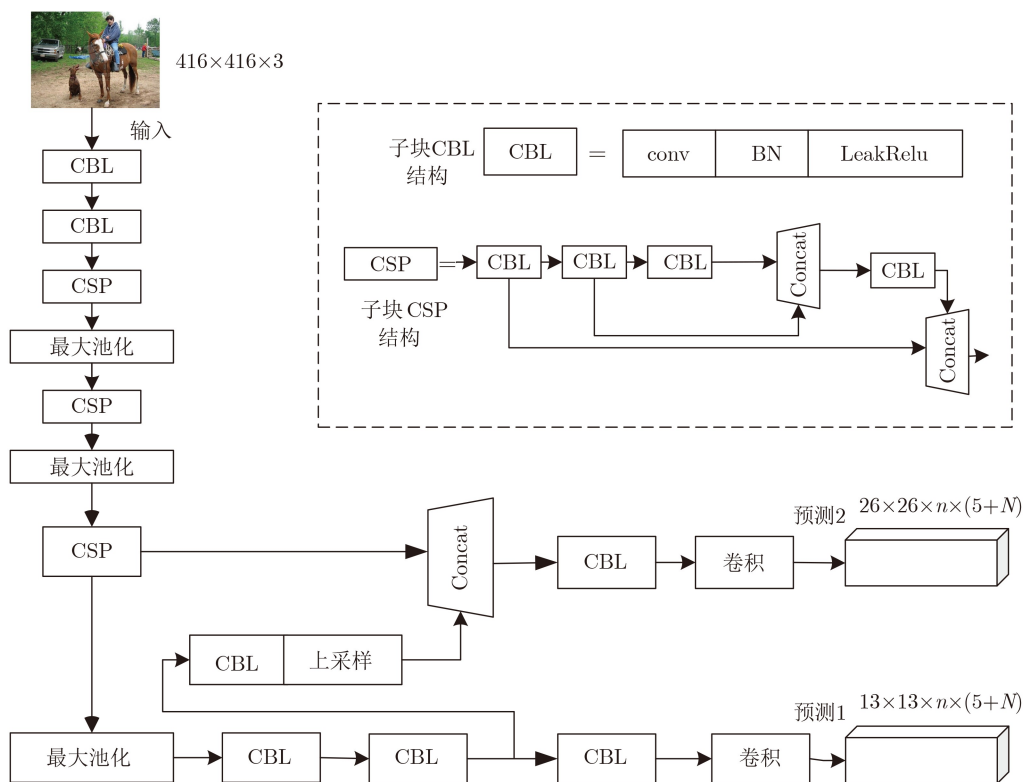


图1 YOLOv4-tiny网络结构图

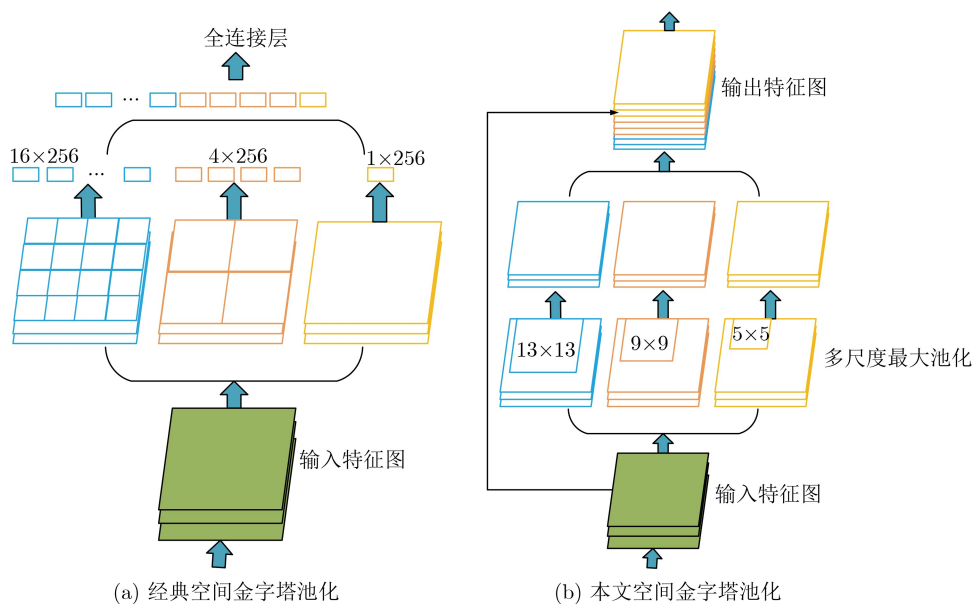


图2 空间金字塔池化

借鉴SPP空间金字塔池化思想,本文使用的是如图2(b)所示的SPP空间金字塔池化模块,对特征图的局部区域进行最大池化操作。这里的多尺度空间金字塔池化模块由3个最大池化层组成,池化窗口的大小分别为 13×13 , 9×9 和 5×5 ,步长均为1。对输入特征图填充0,从而确保池化后输出的特征图与输入特征图大小相同。该空间金字塔池化模块融合了多重感受野,同时利用局部特征和全局特征来提高目标检测的精度。

3 手势识别网络构建及改进

3.1 基于K-means++算法的先验框选取

原始YOLOv4-tiny算法中的先验框尺寸是对COCO数据集聚类得到的,有80种检测目标,这些目标物体尺寸不一。本文手势检测识别网络的检测目标只有手势,手势目标数据集聚类生成的多数先验框大小相近、宽高比相似,为了使得改进YOLOv4-tiny算法更加快速准确地预测出目标手势

的位置, 本文利用K-means++聚类算法对手势数据集NUS-II进行重新聚类, 得到更适合检测手势的先验框。K-means++算法是K-means算法的改进版, K-means++算法逐个选取 k 个簇中心, 离其他簇中心越远的样本点越有可能成为下一个簇中心, 然后计算目标与簇中心的距离, 并把它们分配给最近的簇中心成为 k 个簇中心, 重复计算更新簇中心, 直到簇中心不再改变, 得到最终的聚类结果。K-means++算法通常以欧氏距离作为计算的度量距离, 但在该文算法中, 利用簇类中心先验框与其他先验框的面积交并比(Intersection Over Union, IOU)作为度量距离。对NUS-II手势数据集重新进行聚类, 生成6个先验框为(92, 165), (176, 165), (118, 256), (138, 335), (258, 284), (202, 303)。

3.2 手势识别网络改进

本文在原始YOLOv4-tiny的基础上增加SPP空间金字塔池化模块, 通过SPP模块实现了局部手势特征和全局手势特征的融合, 特征图经过局部手势特征与全局手势特征相融合后, 丰富了特征图的表达能力, 所以在检测精度上有了较好的提升。同时, 本文在YOLOv4-tiny每个最大池化层和SPP空间金字塔池化模块后添加 1×1 卷积模块, 减少参数数量和计算量。NUS-II手势数据集上分类个数为10, 改进YOLOv4-tiny算法的手势识别网络结构如图3所示。

4 手势识别实验与结果分析

4.1 实验环境配置和训练参数设置

本实验操作系统为Ubuntu16.04, 设备为Intel(R) Core(TM) i5-9400F和NVIDIA GeForce RTX 2080Ti, 采用深度学习框架为Darknet, 调用CUDA(Computer Unified Device Architecture)和英伟达神经网络库(NVIDIA CuDNN, cuDNN)为显卡加速, CUDA和cuDNN版本分别为10.1和7.6, 同时安装计算机视觉库OpenCV3.4.11显示检测的结果。采用迁移学习思想, 将YOLO官网在COCO数据集上预训练的yolov4-tiny.conv.29权重作为初始权重进行训练, 加快网络收敛。网络输入图像为 416×416 大小的RGB图像, 批量大小设置为64, subvision设置为8, 动量参数设置为0.9, 权重衰减为0.0005, 训练最大迭代数设置为40000, 初始学习率设置为0.001, 分别在总迭代数的80%和90%即32000次和36000次时将学习率调为0.0001和0.00001。

4.2 实验数据集

实验采用的是公开手势数据集NUS-II, 这是一个10类手势数据集, 手势背景比较复杂, 是各种自然场景。这些形状不同、尺寸各异的手势由40名不同种族人士分别在不同背景下拍摄, 分别展示字母“a”到“j”总共10种手势, 每个字母手势拍摄5次。同时, 数据集中还包括带有人类干扰的750张

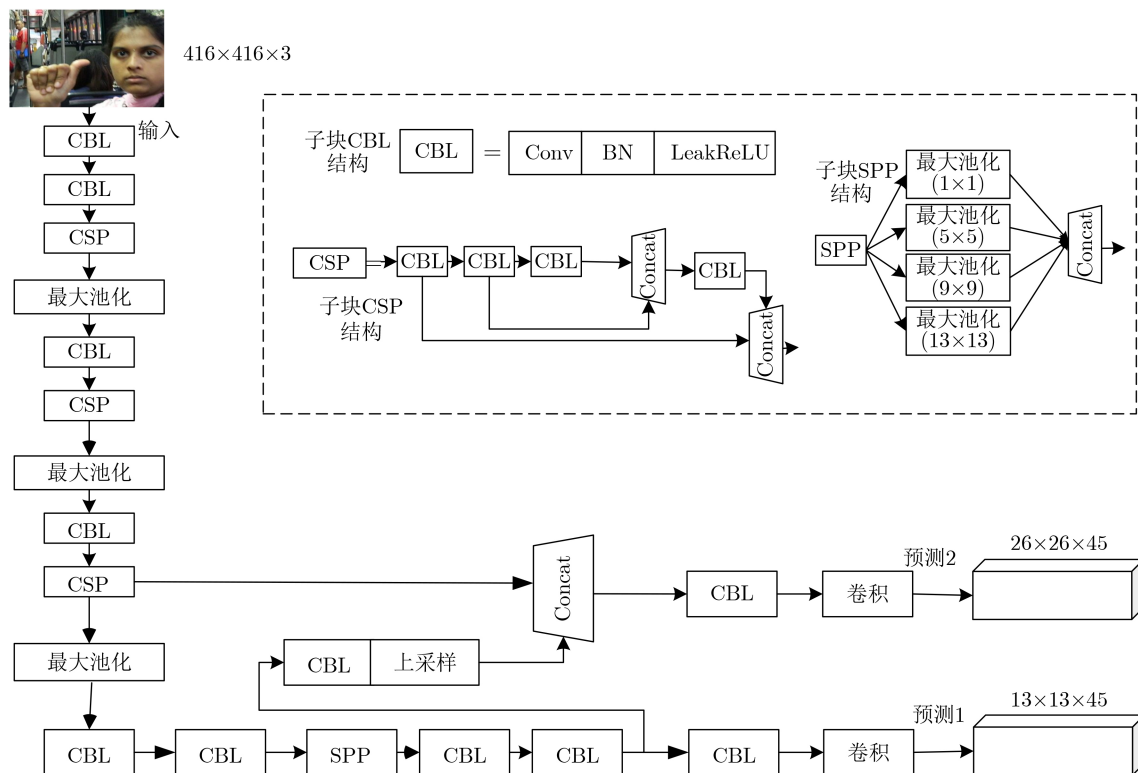


图3 改进YOLOv4-tiny算法手势识别结构图

图片，数据集合计2750张。数据集贴合实际生活，可以研究不同复杂背景、不同光照、不同手势下的识别，有很大的研究价值。将2750张数据集分为2305张训练集和445张测试集，数据集部分图片如图4所示。

4.3 模型结果与分析

训练过程中，每4个epoch在测试集上计算1次mAP，图5显示了手势检测识别模型在训练期间的损失曲线和mAP曲线变化。从图5看出，在模型训练的初级阶段，模型学习效率较高，训练曲线收敛速度快。随着训练的深入，训练曲线的斜率逐渐减小，最后，当训练迭代数到达大约8000，模型学习效率逐渐达到饱和，损失在0.006左右。mAP在迭

代数2000时基本达到100%并稳定维持。训练的YOLOv4-tiny算法对手势的检测识别效果如图6所示，可以看出该算法能够准确检测并识别出NUS-II测试集上复杂背景、带有肤色和人脸干扰的“a”——“j”10种手势。

实验采用精确率、平均精度均值和每秒传输帧数作为评价手势检测识别方法的性能评价标准。精确率和召回率计算方法如式(1)和式(2)所示

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

其中，TP表示正样本被正确识别为正样本的个

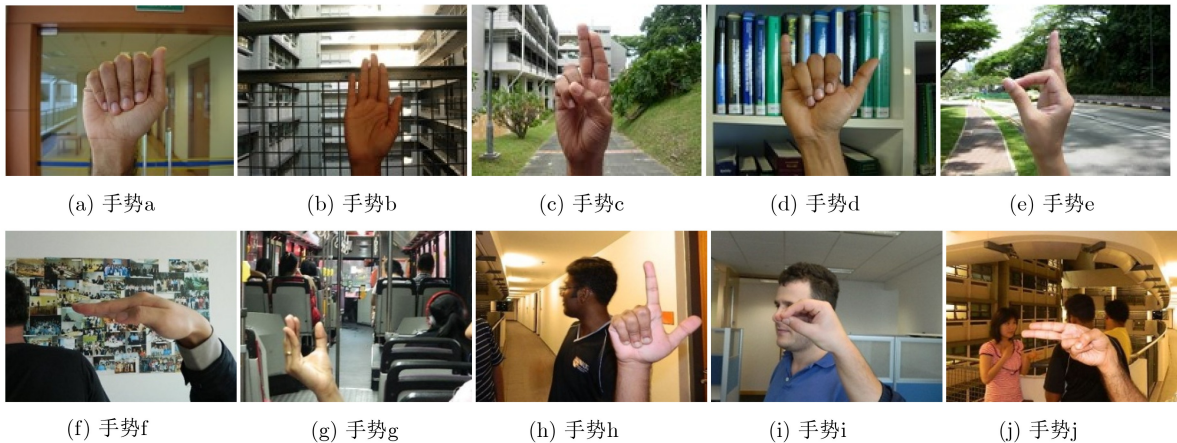


图 4 NUS-II手势数据集

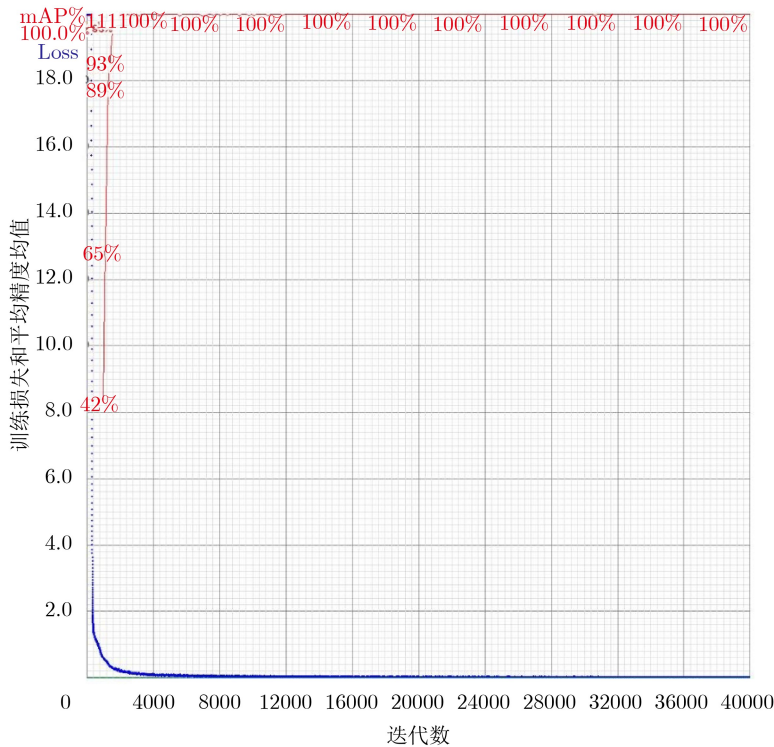


图 5 手势检测模型的mAP和损失曲线

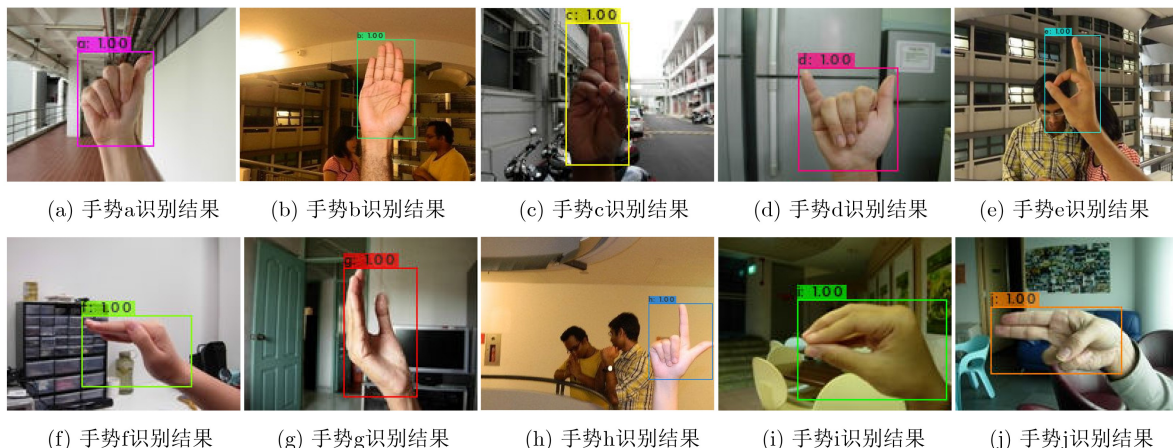


图6 手势检测识别结果

数, FP表示负样本被错误识别为正样本的个数, FN表示正样本被错误识别为负样本的个数。通过取不同的精确率和召回率值可以画出PR曲线, PR曲线下的面积就定义为AP, 取所有检测类别AP的均值即为mAP。性能评价指标AP和mAP计算方法如式(3)和式(4)所示

$$AP = \int_0^1 p(r)dr \quad (3)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (4)$$

与文献[16,18]、YOLOv3-tiny和YOLOv4-tiny进行比较, 以测试该改进算法的性能, 其中, YOLOv4-tiny1表示用K-means++聚类算法对手势数据集生成的先验框进行的实验; YOLOv4-tiny2表示在K-means++聚类的基础上, 添加 1×1 卷积模块进行的实验; YOLOv4-tiny3表示在K-means++聚类的基础上, 添加SPP空间金字塔池化模块进行的实验。

结果如表1所示, 文献[16]精确率只有90.08%, 文献[18]精确率99.89%最高, 但是由于文献[18]使用的是2阶段目标检测算法Faster R-CNN, 因此速度很慢, 不能实时检测手势。由于NUS-II数据集相对简单, YOLOv3-tiny, YOLOv4-tiny和本文改进YOLOv4-tiny算法都能很容易学习到数据集手势特征, 在IOU取0.5的时候, mAP@0.5都很高, 分别为99.97%, 100%, 100%。本文也选取对检测要求更高的即IOU为0.9和IOU为0.5:0.05:0.9的均值作为衡量算法性能的标准。可以看出虽然YOLOv3-tiny算法速度最快, FPS为420, 但是IOU为0.9的时候, mAP@0.9仅为22.13%, 说明YOLOv3-tiny对手势定位不准, 不能很好地预测出手势的准确位置。同理, IOU为0.9的时候, YOLOv4-tiny算法的

mAP@0.9为61.87%, 而本文改进YOLOv4-tiny算法的mAP@0.9为71.36%, 提高了9.49个百分点。计算在IOU取0.5到0.95时候的10个mAP均值, YOLOv3-tiny, YOLOv4-tiny和本文改进YOLOv4-tiny算法的mAP@0.5:0.95分别为77.05%, 86.10%和88.01%, 本文改进算法提高了1.91%。综合考虑算法的精度和速度, 本文改进算法有着最好的检测效果。

为了更直观地看出算法的性能, 用3种算法分别对测试集中的图片预测, 结果如图7至图9所示, 可以看出, 对于手势a, 3种算法都能很好地检测识别出手势a; 对于手势d, YOLOv3-tiny算法定位手势稍有偏差, YOLOv4-tiny和本文改进算法都能准确定位并识别; 对于手势c, YOLOv3-tiny和YOLOv4-tiny算法都出现了一定的定位偏差, 本文改进算法依然可以准确定位到手势并识别出来。本文改进算法是以上算法中速度和准确率最佳权衡的算法。

5 Android手势识别系统

5.1 Android介绍

安卓(Android)是Google开发的基于Linux平台

表1 实验结果对比

算法	精确率 (%)	mAP@0.5 (%)	mAP@0.9 (%)	mAP@0.5:0.95 (%)	fps
文献[16]	90.08	—	—	—	—
文献[18]	99.89	—	—	—	—
YOLOv3-tiny	98.87	99.97	22.13	77.05	420
YOLOv4-tiny	99.09	100.00	61.87	86.10	382
YOLOv4-tiny1	99.10	100.00	69.39	87.10	384
YOLOv4-tiny2	99.33	100.00	66.66	86.96	387
YOLOv4-tiny3	99.10	100.00	73.99	88.20	353
本文算法	99.77	100.00	71.36	88.01	377

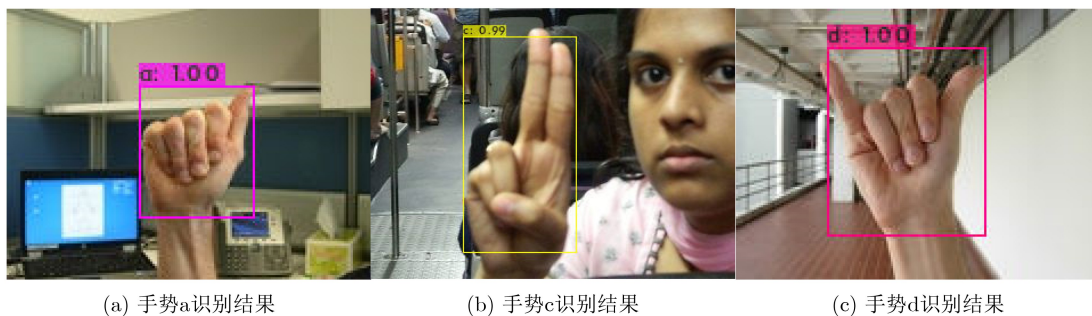


图7 YOLOv3-tiny算法手势检测识别结果

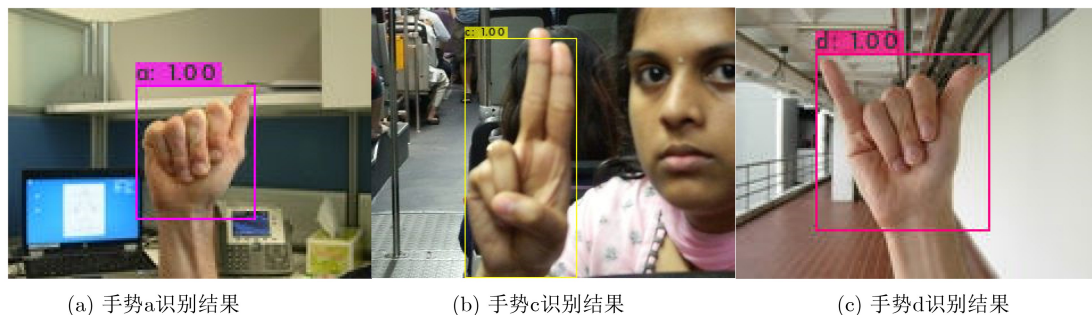


图8 YOLOv4-tiny算法手势检测识别结果



图9 改进YOLOv4-tiny算法手势检测识别结果

的开源手机操作系统，主要用于智能手机与平板电脑，现在已经拓展到互联网电视、智能家居等。它包括操作系统、用户界面和应用程序。应用非常广泛，开发环境不会受到各种条条框框的限制，开发者任意修改开放的源代码来实现与开发各种实用的手机APP软件，具有高级图形显示、界面友好等特点，对人机交互有着广泛的应用前景。

5.2 手势识别移动端部署

为了方便在移动端部署深度学习模型，出现了很多移动端深度学习工具，本文使用ncnn^[24]框架进行模型部署。腾讯优图实验室的ncnn是一个为手机端极致优化的高性能神经网络计算框架。设计初心就考虑手机端的部署和使用，手机端CPU的处理速度快于目前已知的移动端深度学习框架。开发者能够将深度学习算法轻松移植到手机端并高效执行，开发出人工智能手机APP，将人工智能算法在手机端运行。

首先将在darknet环境中训练的改进YOLOv4-

tiny手势识别权重文件gesture.weights和网络配置文件gesture.cfg文件转成ncnn格式的gesture.param和gesture.bin文件。然后搭建Android手势识别项目工程，在主程序中调用手势识别算法进行检测识别。手势识别算法在小米10手机进行测试，如图10所示，对拍摄手势图进行检测识别，手机APP可以正确识别出“a”——“j”10种手势，一张手势图处理时间仅需4.4 ms，fps为22，可以达到移动端实时检测识别的效果。

6 结论

手势识别算法的发展趋势要求识别准确率高，识别速度快，同时识别算法要小，可以部署在移动端。本文提出一种改进YOLOv4-tiny算法，通过增加SPP空间金字塔池化模块来融合图像全局和局部特征，增强算法的手势定位准确度，丰富了特征图的表达能力；通过增加1×1卷积模块减少网络参数，同时提升网络处理速度；通过K-means++聚

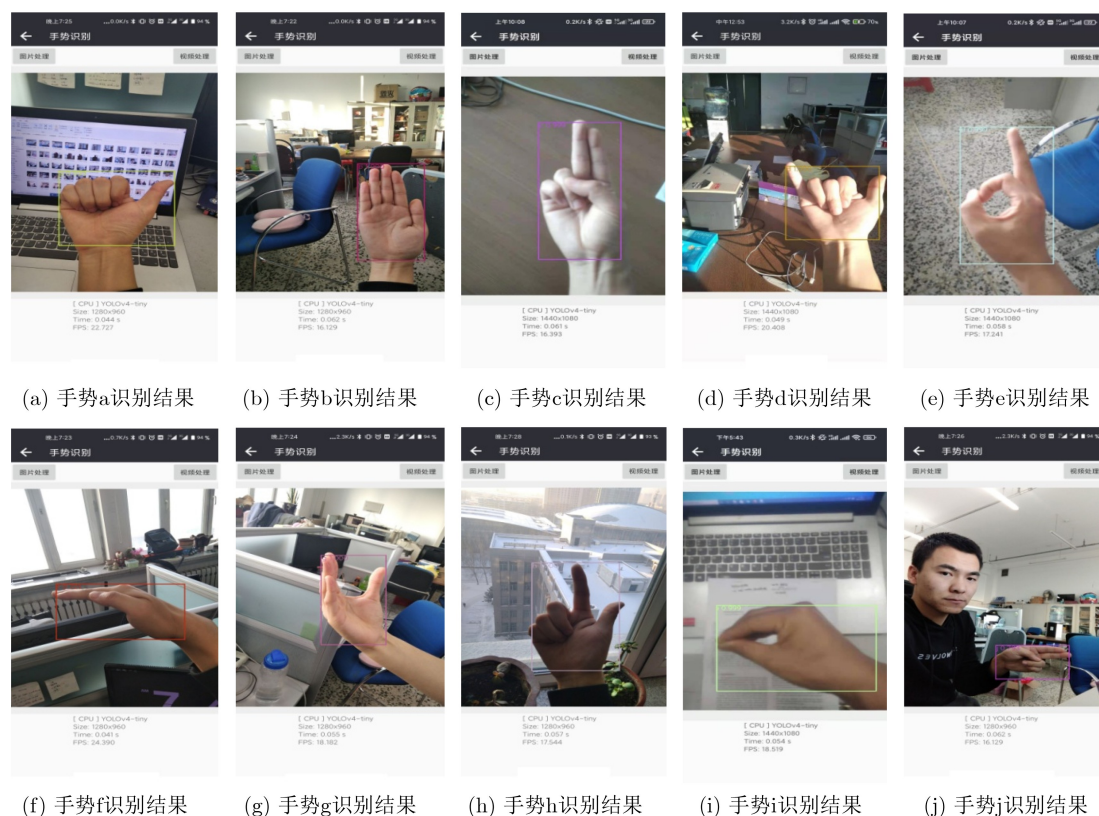


图10 移动端手势识别

类算法对手势数据集重新聚类, 得到更适合检测手势的先验框。实验表明, 本文改进算法在几乎不影响处理速度的情况下提升检测性能, $mAP@0.9$ 为 71.36%, 相比改进前提高了 9.49 个百分点, $mAP@0.5:0.95$ 为 88.01%, 相比改进前提高了 1.91%, 检测效果得到明显的提升。将本文算法移植到移动端, 仅需要 4.4 ms 即可准确检测并识别出手势, 可以在移动端进行实时的手势检测与识别, 对人机交互领域有很大的研究意义。

参考文献

- [1] 夏朝阳, 周成龙, 介钧誉, 等. 基于多通道调频连续波毫米波雷达的微动手势识别[J]. 电子与信息学报, 2020, 42(1): 164–172. doi: [10.11999/JEIT190797](https://doi.org/10.11999/JEIT190797).
- [2] XIA Zhaoyang, ZHOU Chenglong, JIE Junyu, *et al.* Micro-motion gesture recognition based on multi-channel frequency modulated continuous wave millimeter wave radar[J]. *Journal of Electronics & Information Technology*, 2020, 42(1): 164–172. doi: [10.11999/JEIT190797](https://doi.org/10.11999/JEIT190797).
- [3] OYEDOTUN O K and KHASHMAN A. Deep learning in vision-based static hand gesture recognition[J]. *Neural Computing and Applications*, 2017, 28(12): 3941–3951. doi: [10.1007/s00521-016-2294-8](https://doi.org/10.1007/s00521-016-2294-8).
- [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. doi: [10.3778/j.issn.1002-8331.1508-0251](https://doi.org/10.3778/j.issn.1002-8331.1508-0251).
- [5] WANG Long, LIU Hui, WANG Bin, *et al.* Gesture recognition method combining skin color models and convolution neural network[J]. *Computer Engineering and Applications*, 2017, 53(6): 209–214. doi: [10.3778/j.issn.1002-8331.1508-0251](https://doi.org/10.3778/j.issn.1002-8331.1508-0251).
- [6] MOHANTY A, RAMBHATLA S S, and SAHAY R R. Deep gesture: Static hand gesture recognition using CNN[C]. *International Conference on Computer Vision and Image Processing*, Singapore, 2017: 449–461. doi: [10.1007/978-981-10-2107-7_41](https://doi.org/10.1007/978-981-10-2107-7_41).
- [7] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: Unified, real-time object detection[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, 2016: 779–788. doi: [10.1109/CVPR.2016.91](https://doi.org/10.1109/CVPR.2016.91).
- [8] REDMON J and FARHADI A. YOLO9000: Better, faster, stronger[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, 2017: 6517–6525. doi: [10.1109/CVPR.2017.690](https://doi.org/10.1109/CVPR.2017.690).
- [9] REDMON J and FARHADI A. YOLOv3: An incremental improvement[EB/OL]. <http://arxiv.org/abs/1804.02767>, 2018.
- [10] BOCHKOVSKIY A, WANG C Y, and LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detector[EB/OL]. <https://arxiv.org/abs/2004.10934v1>, 2020.
- [11] LIU Wei, ANGUELOV D, ERHAN D, *et al.* SSD: Single

- shot MultiBox detector[C]. 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 21–37. doi: [10.1007/978-3-319-46448-0_2](https://doi.org/10.1007/978-3-319-46448-0_2).
- [10] LIN T Y, GOYAL P, GIRSHICK R, *et al*. Focal loss for dense object detection[C]. The IEEE Transactions on Pattern Analysis and Machine Intelligence, Venice, Italy, 2017: 2999–3007. doi: [10.1109/TPAMI.2018.2858826](https://doi.org/10.1109/TPAMI.2018.2858826).
- [11] LAW H and DENG Jia. CornerNet: Detecting objects as paired keypoints[C]. The 15th European Conference on Computer Vision (ECCV), Munich, Germany, 2018: 765–781. doi: [10.1007/978-3-030-01264-9_45](https://doi.org/10.1007/978-3-030-01264-9_45).
- [12] GIRSHICK R, DONAHUE J, DARRELL T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014: 580–587. doi: [10.1109/CVPR.2014.81](https://doi.org/10.1109/CVPR.2014.81).
- [13] GIRSHICK R. Fast R-CNN[C]. IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1440–1448. doi: [10.1109/ICCV.2015.169](https://doi.org/10.1109/ICCV.2015.169).
- [14] REN Shaoqing, HE Kaiming, GIRSHICK R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [15] DAI Jifeng, LI Yi, HE Kaiming, *et al*. R-FCN: Object detection via region-based fully convolutional networks[C]. The 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 2016: 379–387. doi: [10.5555/3157096.3157139](https://doi.org/10.5555/3157096.3157139).
- [16] SOE H M and NAING T M. Real-time hand pose recognition using faster region-based convolutional neural network[C]. The First International Conference on Big Data Analysis and Deep Learning, Singapore, 2019: 104–112. doi: [10.1007/978-981-13-0869-7_12](https://doi.org/10.1007/978-981-13-0869-7_12).
- [17] PISHARADY P K, VADAKKEPAT P, and LOH A P. Attention based detection and recognition of hand postures against complex backgrounds[J]. *International Journal of Computer Vision*, 2013, 101(3): 403–419. doi: [10.1007/s11263-012-0560-5](https://doi.org/10.1007/s11263-012-0560-5).
- [18] 常建红. 基于改进Faster RCNN算法的手势识别研究[D]. [硕士学位论文], 河北大学, 2020. doi: [10.27103/d.cnki.ghebu.2020.001315](https://doi.org/10.27103/d.cnki.ghebu.2020.001315).
- CHANG Jianhong. The gesture recognition research based on the improved faster RCNN algorithm[D]. [Master dissertation], Hebei University, 2020. doi: [10.27103/d.cnki.ghebu.2020.001315](https://doi.org/10.27103/d.cnki.ghebu.2020.001315).
- [19] 张勋, 陈亮, 胡诚, 等. 一种基于深度学习的静态手势实时识别方法[J]. 现代计算机, 2017(34): 6–11. doi: [10.3969/j.issn.1007-1423.2017.34.002](https://doi.org/10.3969/j.issn.1007-1423.2017.34.002).
- ZHANG Xun, CHEN Liang, HU Cheng, *et al*. A real-time recognition method of static gesture based on depth learning[J]. *Modern Computer*, 2017(34): 6–11. doi: [10.3969/j.issn.1007-1423.2017.34.002](https://doi.org/10.3969/j.issn.1007-1423.2017.34.002).
- [20] 彭玉青, 赵晓松, 陶慧芳, 等. 复杂背景下基于深度学习的手势识别[J]. 机器人, 2019, 41(4): 534–542. doi: [10.13973/j.cnki.robot.180568](https://doi.org/10.13973/j.cnki.robot.180568).
- PENG Yuqing, ZHAO Xiaosong, TAO Huifang, *et al*. Hand gesture recognition against complex background based on deep learning[J]. *Robot*, 2019, 41(4): 534–542. doi: [10.13973/j.cnki.robot.180568](https://doi.org/10.13973/j.cnki.robot.180568).
- [21] 王粉花, 黄超, 赵波, 等. 基于YOLO算法的手势识别[J]. 北京理工大学学报, 2020, 40(8): 873–879. doi: [10.15918/j.tbit1001-0645.2019.030](https://doi.org/10.15918/j.tbit1001-0645.2019.030).
- WANG Fenhua, HUANG Chao, ZHAO Bo, *et al*. Gesture recognition based on YOLO algorithm[J]. *Transactions of Beijing Institute of Technology*, 2020, 40(8): 873–879. doi: [10.15918/j.tbit1001-0645.2019.030](https://doi.org/10.15918/j.tbit1001-0645.2019.030).
- [22] JIANG Zicong, ZHAO Liquan, LI Shuaiyang, *et al*. Real-time object detection method based on improved YOLOv4-tiny[EB/OL]. <https://arxiv.org/abs/2011.04244>, 2020.
- [23] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al*. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904–1916. doi: [10.1109/TPAMI.2015.2389824](https://doi.org/10.1109/TPAMI.2015.2389824).
- [24] GitHub, Inc. NIHUI. ncnn[EB/OL]. <https://github.com/Tencent/ncnn>, 2021.
- 卢迪: 女, 1971年生, 教授, 博士, 研究方向为数据融合、图像处理。
- 马文强: 男, 1992年生, 硕士生, 研究方向为图像处理、手势识别。

责任编辑: 余蓉