JOURNAL OF IMAGE AND GRAPHICS

主办：中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2021
04
VOL.26

ISSN1006-8961
CN11-3758/TB

细粒度图像分类 *P847*

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第26卷第4期（总第300期）
2021年4月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主　　编 吴一戎
编辑出版 《中国图象图形学报》编辑部
通信地址 北京市海淀区北四环西路19号
邮　　编 100190
电子信箱 jig@radi.ac.cn
电　　话 010-58887035
网　　址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订　　购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
（邮政信箱：北京399信箱　邮编：100048）
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial Office of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@radi.ac.cn
Telephone 010-58887035
Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

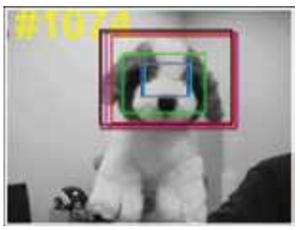
CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

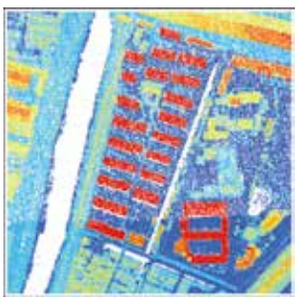
2021年4月 第26卷 第4期（总第300期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

目次

YOLOv3剪枝模型的多人姿态估计(第0837页)



融入时序和速度信息的自适应更新目标跟踪(第0883页)

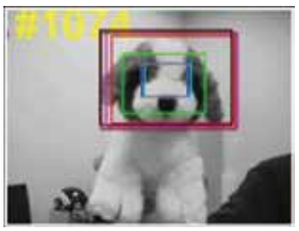


可变半径Alpha Shapes提取机载LiDAR点云建筑物轮廓(第0910页)

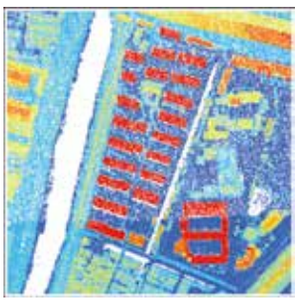
Volume 26, Number 4
Published April 16, 2021

CONTENTS
JOURNAL OF IMAGE AND GRAPHICS

Research on multiperson pose estimation combined with YOLOv3 pruning model(P0837)



Adaptive update object tracking algorithm incorporating timing and speed information(P0883)



Extraction of building contours from airborne LiDAR point cloud using variable radius Alpha Shapes method(P0910)

中图法分类号:TP391　　**文献标识码**:A　　**文章编号**:1006-8961(2021)04-0815-13

论文引用格式:Wang Y, Chen Y L and Sun Q L. 2021. Hand gesture recognition in complex background. Journal of Image and Graphics,26(04):0815-0827(王银,陈云龙,孙前来. 2021. 复杂背景下的手势识别. 中国图象图形学报,26(04):0815-0827)[DOI:10.11834/jig.200211]

复杂背景下的手势识别

王银,陈云龙*,孙前来

1. 太原科技大学电子信息工程学院,太原　030024;2. 先进控制与装备智能化山西省重点实验室,太原　030024

摘　要:**目的**　手势识别是人机交互领域的热点问题。针对传统手势识别方法在复杂背景下识别率低,以及现有基于深度学习的手势识别方法检测时间长等问题,提出了一种基于改进 TinyYOLOv3 算法的手势识别方法。**方法**　对 TinyYOLOv3 主干网络重新进行设计,增加网络层数,从而确保网络提取到更丰富的语义信息。使用深度可分离卷积代替传统卷积,并对不同网络层的特征进行融合,在保证识别准确率的同时,减小网络模型的大小。采用 CIoU(complete intersection over union)损失对原始的边界框坐标预测损失进行改进,将通道注意力模块融合到特征提取网络中,提高了定位精度和识别准确率。使用数据增强方法避免训练过拟合,并通过超参数优化和先验框聚类等方法加快网络收敛速度。**结果**　改进后的网络识别准确率达到 99.1%,网络模型大小为 27.6 MB,相比原网络(TinyYOLOv3)准确率提升了 8.5%,网络模型降低了 5.6 MB,相比于 YOLO(you only look once)v3 和 SSD(single shot multibox detector)300 算法,准确率略有降低,但网络模型分别减小到原来的 1/8 和 1/3 左右,相比于 YOLO-lite 和 MobileNet-SSD 等轻量级网络,准确率分别提升 61.12% 和 3.11%。同时在自制的复杂背景下的手势数据集对改进后的网络模型进行验证,准确率达到 97.3%,充分证明了本文算法的可行性。**结论**　本文提出的改进 TinyYOLOv3 手势识别方法,对于复杂背景下的手势具有较高的识别准确率,同时在检测速度和模型大小方面都优于其他算法,可以较好地满足在嵌入式设备中的使用要求。

关键词:手势识别;TinyYOLOv3;深度可分离卷积;CIoU 损失

Hand gesture recognition in complex background

Wang Yin, Chen Yunlong*, Sun Qianlai

1. College of Electronic Information Engineering, Taiyuan University of Science and Technology, Taiyuan　030024, China

2. Shanxi Key Laboratory of Advanced Control and Equipment Intelligence, Taiyuan　030024, China

Abstract: **Objective** The rapid development of artificial intelligence and target detection technology has accelerated the iteration of intelligent devices and also promoted the development of related technologies in the field of human-computer interaction. As an important body language and an important means to realize human-computer interaction, gesture recognition has attracted considerable attention. It has the characteristics of simplicity, high efficiency, directness, and rich content. This interaction mode is more in line with people's daily behavior and easier to understand. Gesture recognition has a wide application prospect in smart homes, virtual reality, sign language recognition, and other fields. Gesture recognition involves a wide range of disciplines such as image processing, ergonomics, machine vision, and deep learning. In addi-

收稿日期:2020-06-16;**修回日期**:2020-09-16;**预印本日期**:2020-09-23

*通信作者:陈云龙　s20180411@stu.tyust.edu.cn

基金项目:国家自然科学基金项目(61905172);山西省科技成果转化引导专项项目(201904D131023);山西省重点研发计划项目(201803D121025,201903D121130);山西省研究生教育创新项目(2020SY422)

Supported by: National Natural Science Foundation of China(61905172)

tion, due to the variety of gestures and the complexity of practical application environment, gesture recognition has become a very challenging research topic. **Method** The traditional vision-based gesture recognition method mainly uses the skin color or skeleton model of the human body to partially segment gestures and realizes gesture classification through manual design and extraction of effective features. However, the collected RGB images are greatly affected by light conditions, skin color, clothing, and background. In the case of backlight, dark light, and dark skin color, the effects of segmentation and gesture recognition are poor. Features such as texture and edge are extracted manually, feature omission and misjudgment are easily generated during the extraction process. The recognition rate is low, and the robustness is poor under complex background. In recent years, deep learning technology has attracted more and more attention due to its robustness and high accuracy. The convolutional neural network model based on deep learning has gradually replaced the traditional manual feature extraction method and gradually become the mainstream method of gesture recognition. Although existing mainstream deep learning methods such as you only look once (YOLO) and single shot multibox detector (SSD) have achieved high accuracy in gesture recognition under complex backgrounds, their models are generally large, which is a difficult requirement to meet. It is difficult to achieve real-time detection effect with embedded devices and detection time. Therefore, how to reduce the complexity of the model and algorithms while ensuring the detection accuracy and meeting the requirements of real-time detection in practical applications has become an urgent problem that needs to be solved. The TinyYOLOv3 algorithm has the advantages of fast detection speed and small model, but its recognition accuracy is far from meeting the requirements of practical application. Therefore, to solve the above problems, this study proposes a gesture recognition method based on the improved TinyYOLOv3 algorithm. In this study, the TinyYOLOv3 backbone network is redesigned. The convolution operation with stride of 2 is used to replace the original maximum pooling layer, and the number of network layers is increased to ensure that the network extracts richer semantic information. At the same time, the depthwise separable convolution is used to replace the traditional convolution, and the characteristics of different network layers are integrated to reduce the size of the network model, ensure the recognition accuracy, and avoid the loss of feature information due to the deepening of the network structure. In the improvement of the loss function, the CIoU (complete intersection over union) loss is used to replace the original bounding box coordinate to predict loss. The experimental results show that CIoU is helpful to speed up the convergence of the model, reduce the training time, and improve the accuracy to a certain extent. The channel attention module is integrated into the feature extraction network, and the information of different channels is recalibrated to reduce the increase of parameters and improve the recognition accuracy. The data enhancement method is used to avoid overfitting training, and super parameter optimization, dynamic learning rate setting, prior frame clustering, and other methods are used to accelerate network convergence. **Result** This study uses the NUS-II (National University of Singapore) gesture dataset for verification experiments. The experimental results show that the accuracy rate of the improved network recognition reaches 99.1%, which is 8.5% higher than that of the original network (TinyYOLOv3, 90.6%), and the size of the network model is reduced from 33.2 MB to 27.6 MB. Compared with YOLOv3, the recognition accuracy of the improved algorithm is slightly reduced; however, the detection speed is nearly doubled, the model size is about one-eighth that of YOLOv3, and the number of parameters is also reduced by nearly 10 times, verifying the feasibility of the algorithm. At the same time, ablation experiments were carried out on different improved modules, and the results showed that the improvement of each module helped to improve the accuracy of the algorithm. By comparing and analyzing the accuracy and loss rate changes of Tiny YOLOv3, improved Tiny YOLOv3 by CIoU and the algorithm in this paper, the advantages of the algorithm in this paper in training time and convergence speed are verified. The advantages of this algorithm in terms of training time and convergence speed are verified. This study also compared the improved algorithm with some classical traditional and deep learning gesture recognition algorithms. In terms of model size, detection time, and accuracy, the algorithm in this study achieved better results. **Conclusion** Gesture recognition in complex background is a key and difficult problem in the field of gesture recognition. To solve the problems of low gesture recognition rate of traditional gesture recognition methods in complex background and long detection time of existing gesture recognition methods based on deep learning, a gesture recognition method based on improved TinyYOLOv3 algorithm is proposed in this study. The network structure, loss function, feature channel optimization, and prior frame clustering are improved. The use of depthwise separable convolution makes it possible to deepen the network while reducing the number of parameters. The

deepening of the network structure and the optimization of the feature channel enable the network to extract more effective semantic information and improve the detection effect. The improved network not only ensures the accuracy, but also takes into account the balance between the network model size and detection time, which can meet the use requirements of embedded equipment.

Key words: gesture recognition; TinyYOLOv3; depthwise separable convolution; CIoU loss

0 引 言

机器视觉和人工智能技术的兴起有效推动了人机交互技术的发展。手势作为人与人之间自然的交流方式,蕴含着丰富的信息,在智能家居、虚拟现实和手语识别等众多领域都有着广泛的应用。目前,手势识别方法主要分为基于计算机视觉或者相关硬件设备的传统方法和基于卷积神经网络等的深度学习方法两种。

基于计算机视觉的传统手势识别方法依靠提取到的纹理形状等特征进行手势分类。Priyal 和 Bora (2013)通过人体肤色模型提取手臂和人手区域,用人体测量学方法分离手臂和手部区域,最后使用归一化矩特征对二值化的手部轮廓进行表征,并使用最小距离分类法进行手势识别,对于相似变换和透视失真具有较强的鲁棒性。Gupta 等人(2016)把 HOG(histogram of oriented gradient)和 SIFT(scale invariant feature transform)两种特征进行融合,通过计算测试图像和训练集的特征矩阵的相关性,并将其相关性输入到 K 近邻分类器(K-nearest neighbor, KNN)中实现对不同手势的分类和识别。刘淑萍等人(2015)通过人体肤色模型对人手区域进行提取,使用手指个数信息和 HOG 特征建立适当的分类器进行手势识别。以上几种手势识别方法都是通过传统手工提取特征的方式进行手势识别,但容易受到人手分割效果的影响,不同种族的人肤色模型差异较大,现实环境中也存在较多类肤色背景的影响,难以完整、有效地提取人手部位,使得相应的手势识别的准确率较低。

另一种传统手势识别方法是基于硬件设备的方法。Lyu 等人(2017)使用数据手套进行手势信息采集,对原始数据进行稀疏分解后应用支持向量机(support vector machine,SVM)分类器进行识别。由于数据手套操作的复杂性,这种方法会严重影响用户体验。Jiang 等人(2019)通过双目摄像机采集手势图片,采用 BM(bidirectional matching)立体匹配算法提取深度图像,并将深度信息重新映射到原始 RGB 图像上,实现 3 维重建和点云图像生成。根据点云图像信息,对手部信息进行提取。这种检测方法依赖于特定的设备,这些设备硬件成本较高,一般用户无法接受。

传统的目标检测方法需要人工设计并提取相应特征,而深度学习的目标检测方法通过逐层特征变换的方式对不同层次的特征进行分析和计算,进而可以自动学习到待检测目标的不同特征。深度学习的方法在手势识别领域也有着广泛的应用。虽然现在比较流行的基于深度学习的目标检测算法,如 YOLO(you only look once)(Redmon 等,2016;Redmon 和 Farhadi,2017,2018)、SSD(single shot multibox detector)(Liu 等,2016)、RCNN(region convolutional neural network)(Girshick 等,2014)和 Faster R-CNN(Ren 等,2015)等算法在目标检测和分类问题中取得了较高的准确率,但这些算法往往通过设计更深层次的网络结构来提取更多的深度特征,对硬件的计算能力和存储能力的要求很高,这些检测模型普遍存在模型较大和检测时间长等问题,难以在嵌入式设备中普及,也不能满足许多场合中对于实时性的要求。

为了有效解决以上问题,满足深度神经网络在嵌入式设备中的使用要求,许多轻量级目标检测网络结构被提出。YOLO-lite(Huang 等,2018)是基于 YOLOv2 提出的简化版本,在其网络结构的设计中,使用了更少的卷积层数,并且去掉了 BN(batch normalization)层,以此减小网络计算开销,提高检测速度。YOLO-lite 在无 GPU 设备中检测速度可达到 21 帧/s,但其在 COCO(common objects in context)数据集上的 mAP(mean average precision)值仅为 12.26%,难以达到实际应用中的精度要求。同理,YOLOv3 也提出了其对应的轻量化版本 TinyYOLOv3。另外,还有许多基于 MobileNet(Howard 等,2017)、MobileNetV2(Sandler 等,2018)、SqueezeNet

(Iandola 等,2016)和 ShuffleNet(Zhang 等,2018)等经典的轻量级目标分类网络构建的目标检测网络。SqueezeDet(Wu 等,2017)是基于改进的 SqueezeNet 提出的目标检测网络,SqueezeNet 大量使用 1×1 卷积核来替换 3×3 卷积核,同时通过减少 3×3 卷积核的数量来减小参数量,该网络在减小模型大小和运算量的同时,尽可能保证检测精度,但是这种网络设计方法使得网络深度有所增加,导致检测时间变长。MobileNet-SSD(Howard 等,2017)使用 MobileNet 代替 SSD 的骨干网络,使网络参数大幅减少,同时检测速度也得到有效提升,在 PASCAL VOC(pattern analysis statistical modeling and computational learning visual object classes)2007 数据集上的 mAP 值为 68%。MixNet(Tan 等,2019)在使用深度可分离卷积的同时,通过不同大小卷积核的组合来提升检测精度,但是过多的分支结构和较大的卷积核设计使得检测速度变慢。以上目标检测方法在网络模型轻量化方面做出了很大贡献,但相关网络结构的设计更加注重减小网络参数量并提升网络的检测速度,而忽略网络的特征表达能力,难以兼顾检测时间、准确率和模型大小。

为了有效解决以上手势识别方法存在的复杂背景下识别率低、实时性差等问题,以及满足在相关嵌入式设备中的使用要求,本文提出了一种改进的 TinyYOLOv3 的手势识别方法。主要贡献如下:针对传统卷积网络参数量和计算量大的问题,使用深度可分离卷积对 TinyYOLOv3 网络的主干结构进行改进,在不影响实时性的同时,有效减少了网络参数量;使用 CIoU(complete intersection over union)损失代替原损失函数中的边界框坐标预测损失项,同时使用先验框聚类和数据增强等方法,加快了网络收敛速度并使得手势的定位精度得到改善;在特征提取网络中加入通道注意力模块,改善网络的特征提取能力,提高了手势识别的准确率。

1 网络结构

1.1 TinyYOLOv3 网络

TinyYOLOv3 是 YOLOv3 的简化版本,并继续沿用了 YOLOv3 网络的回归思想,将目标检测问题视为回归问题,通过网络回归直接得到检测目标的位置以及分类结果。该方法有效地减少了网络计算开支,极大地提升了网络处理速度。TinyYOLOv3 的网络结构如图 1 所示(Gong 等,2020),在 YOLOv3 的基础上去掉了许多特征提取层,在多尺度检测上只保留了 13×13 和 26×26 两个预测分支,采用上采样的方式将 13×13 的特征层和 26×26 的特征层进行融合,最后在两个特征图上分别作出预测。其主

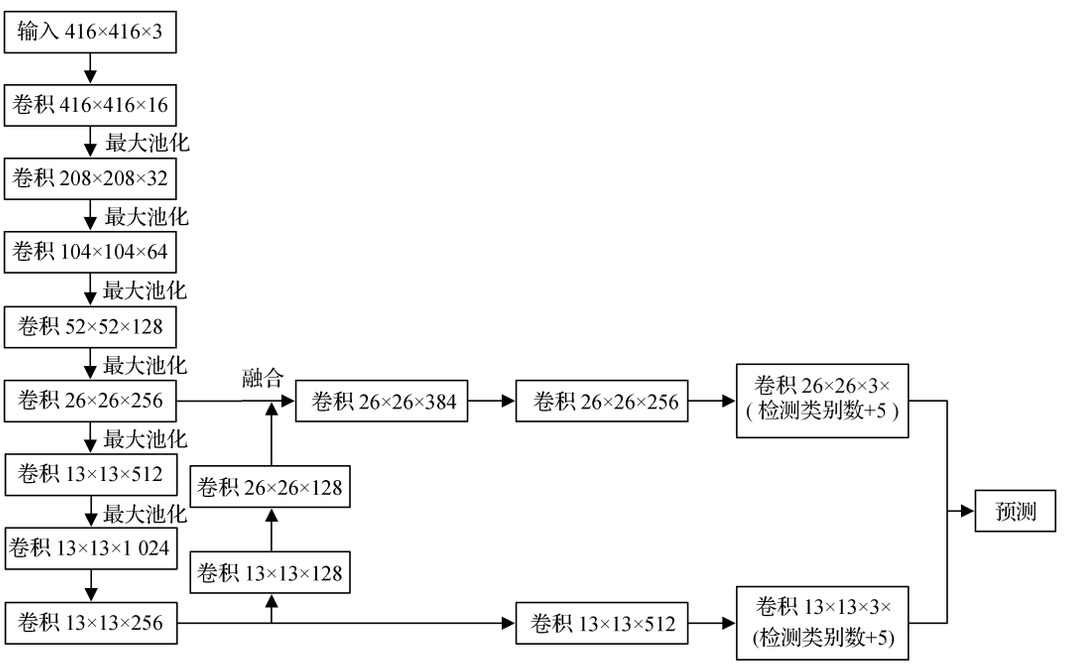


图 1 TinyYOLOv3 网络结构(Gong 等,2020)

Fig. 1 TinyYOLOv3 network structure(Gong et al., 2020)

干网络由卷积层和最大池化层拼接而成，并且去掉了 YOLOv3 中的残差结构，由于主干网络比较浅，很难提取到更深层次的语义特征。由于 TinyYOLOv3 的网络结构简单，计算量小，检测时间快，更适合在移动端和嵌入式设备中使用。但是这种简单的网络结构，使得在前向传递过程中很容易造成特征损失，因此其检测精度相比 YOLOv3 网络也大大降低。

1.2　深度可分离卷积模块

由于 TinyYOLOv3 的识别率较低，很难在实际应用中使用，而现有的经典分类网络如 YOLOv3 和 SSD 等，通过增加网络层数来提升识别准确率，这种设计方法使得网络模型和计算量不断增大，检测时间也随之增加，而一般的嵌入式设备根本无法满足此类大型网络对于存储和计算资源的要求，这将严重限制深度学习算法在嵌入式设备中的使用和发展。

考虑到上述问题，本文通过增加 TinyYOLOv3 网络的特征提取层数，提升网络对于复杂背景下的手势识别准确率，同时采用深度可分离卷积代替传统卷积层，在保证检测精度的前提下，降低网络模型的大小，使其满足在嵌入式设备中的使用要求。其中深度可分离卷积示意见图 2。

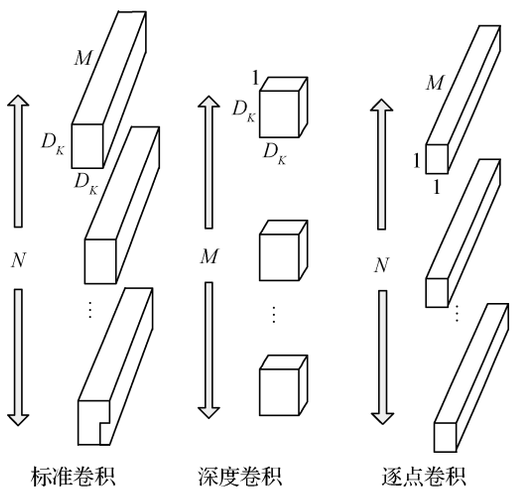


图 2　深度可分离卷积示意图

Fig. 2　Diagram of depth-wise separable convolution

在常规卷积中，输入特征图像的所有通道数都会被同时考虑，这种传统的卷积方式存在很大的计算冗余。深度可分离卷积提出了一种新的卷积思路（Howard 等，2017）：将传统卷积改为深度卷积和 1 × 1 卷积两层卷积操作，首先使用深度卷积对不同的通道分别进行卷积加操作，然后使用 1 × 1 卷积进行特征融合。对于一个输入大小为 $H \times W \times C$ 的特征图使用 stride = 1 的 K 个 3 × 3 的卷积核进行卷积操作后的输出大小为 $H \times W \times K$，若使用普通卷积，则计算量为 $H \times W \times C \times K \times 3 \times 3$，而使用深度可分离卷积进行卷积时，depthwise 计算量为 $H \times W \times C \times 9$，pointwise 的计算量为 $H \times W \times C \times K$，总计算量为 $H \times W \times C \times (9 + K)$，深度可分离卷积的计算量缩减普通卷积的$\frac{1}{9}$左右。

本文使用如图 3（a）（b）两种不同的深度可分离卷积模块代替 TinyYOLOv3 中的传统卷积，当 stride = 1 时，首先使用 1 × 1 卷积对输入特征进行升维，避免由于非线性激活造成的特征损失，然后进行标准的深度可分离卷积操作，参考 ResNet（residual network）的特征融合思想对浅层特征进行融合，同时使用 1 × 1 卷积减少融合的特征数，在保证网络提取到更多的手势信息的同时，有效减少网络的参数量。当 stride 为 2 时使用步长为 2 的卷积层代替原始 TinyYOLOv3 的池化层。改进后的特征提取主干网络如表1所示，其中 ***X***1 和 ***X***2 为主干网络不同尺

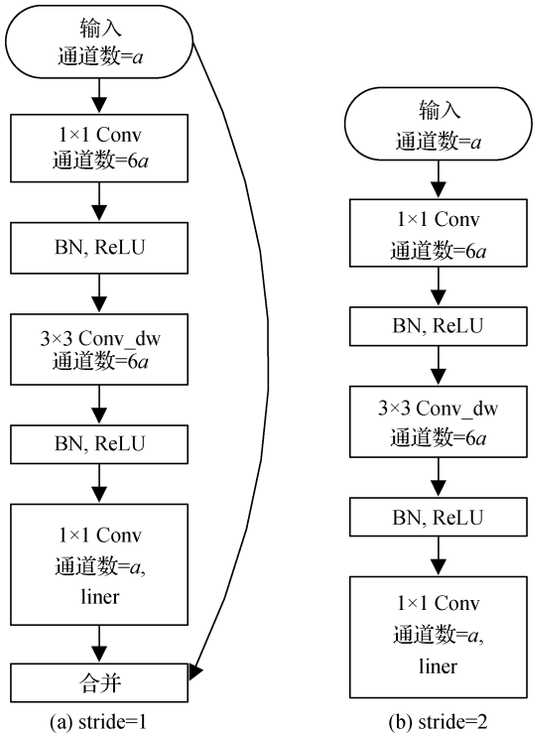


图 3　深度可分离卷积模块

Fig. 3　Depth-wise separable convolution

（（a）stride = 1；（b）stride = 2）

表 1　特征提取主干网络

Table 1　Backbone network of feature extraction

	卷积类型	卷积步长	通道数	特征图尺寸	主干输出
	Conv2d, 3×3	2	24	208×208	
	Conv_dw	1	12	208×208	
	Conv_dw	2	18	104×104	
	Conv_dw	1	18	104×104	
	Conv_dw	2	24	52×52	
2×	Conv_dw	1	24	52×52	
	Conv_dw	2	48	26×26	
	Conv_dw	1	48	26×26	
4×	Conv_dw	1	72	26×26	*X*2
	Conv_dw	2	120	13×13	
3×	Conv_dw	1	120	13×13	
	Conv_dw	1	240	13×13	
	Conv2d, 1×1	1	960	13×13	*X*1

度的特征输出,在主干网络中,除了第一层和最后一层使用普通卷积外,其他层次都使用深度可分离卷积,并在保证准确率的同时,通过压缩 channel 数量来减少网络参数量。

1.3　IoU 和 CIoU

目标定位是目标检测的一个重要任务,通常通过边界框的坐标位置来表示。YOLO 系列引入交并比(intersection over union, IoU)间接作为衡量预测边界框置信度的重要参数。是衡量目标定位精度的一个常用指标。IoU 具有尺度不变性的优点,任意两个方框 A 和 B 的相似性与它们的空间尺度大小无关。但是 IoU 也存在明显的缺点。一方面,对于两个不相交的目标,对应的 IoU 将变为 0,此时无论两个目标距离远近,IoU 都无法体现,当使用 IoU 作为损失函数时,有可能出现梯度为 0 的情况,导致无法优化;另一方面,相同的 IoU 并不能代表检测框具有相同的定位效果,即 IoU 无法区分两个对象之间不同的对齐方式。

GIoU(generalized intersection over union)损失函数(Rezatofighi 等,2019)引入了真实框和检测框的最小外接矩形,可以很好地解决真实框和检测框不相交的问题。但是当真实框和检测框出现包含现象时,GIoU 则退化为 IoU,与 IoU 效果相同,将无法区分真实框和检测框之间的相对位置。DIoU(distamce intersection over union)和 CIoU 损失函数(Zheng 等,2020)考虑了真实框和检测框的中心点之间的距离信息(如图 4 中的 d),解决了 GIoU 存在的不足,同时 CIoU 添加了检测框和真实框各自宽高比的相似性损失项,使得预测框更加贴近于真实框。因此本文使用 CIoU 损失代替原损失函数中的边界框坐标预测损失项。

真实框和检测框的交并比示意图如图 4 所示,其中 $\boldsymbol{B}$ 为检测框,$\boldsymbol{B}^{gt}$ 为真实框,d 为 $\boldsymbol{B}$ 和 $\boldsymbol{B}^{gt}$ 中心点之间的距离,c 为 $\boldsymbol{B}$ 和 $\boldsymbol{B}^{gt}$ 最小外接矩形的对角线的距离。

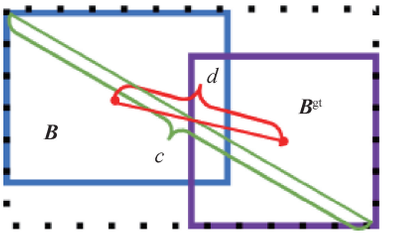


图 4　交并比(IoU)示意图(Zheng 等,2020)

Fig. 4　The schematic diagram of intersection over union (IoU) (Zheng et al., 2020)

CIoU 损失计算为

$$L_{\mathrm{CIoU}} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{1}$$

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{2}$$

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{3}$$

式中,b 和 b^{gt} 分别表示检测框和真实框的中心点,$\rho(\cdot)$ 表示欧氏距离,IoU 表示 B 和 B^{gt} 的交并比,w 和 h 表示检测框的宽和高,w^{gt} 和 h^{gt} 表示真实框的宽和高。

1.4　损失函数

本文使用 CIoU 损失对 TinyYOLOv3 算法中的边界框预测损失项进行替换,改进后的损失函数由 CIoU 误差、置信度误差和分类误差 3 部分构成,计算为

$$L = L_{\mathrm{C}} + L_{\mathrm{con}} + L_{\mathrm{class}} \tag{4}$$

具体各部分损失为

$$L_{\mathrm{C}} = \sum_{i=0}^{s^2}\sum_{j=0}^{B} I_{ij}^{\mathrm{obj}} L_{\mathrm{CIoU}} \tag{5}$$

$$L_{\mathrm{con}} = \sum_{i=0}^{s^2}\sum_{j=0}^{B} I_{ij}^{\mathrm{obj}}(C_i - \tilde{C}_i)^2 +$$

$$\lambda_{\text{noobj}} \sum_{i=0}^{s^2} \sum_{j=0}^{B} I_{ij}^{\text{noobj}} (C_i - \tilde{C}_i)^2 \tag{6}$$

$$L_{\text{class}} = \sum_{i=0}^{s^2} \sum_{j=0}^{B} I_{ij}^{\text{obj}} \sum_{c \in \boldsymbol{class}} (p_i(c) - \tilde{p}_i(c))^2 \tag{7}$$

式中，s^2 为输入图像中的网格数，B 为每个网格预测生成的边界框数，本文中设计的网络采用输出为 13×13 和 26×26 的两个尺度的特征融合，所以 s 取值为 13 和 26，B 取值为 2。$I_{ij}^{\text{obj}} = 1$ 表示目标物体落入到第 i 个网格的第 j 个边界框中，没有落入时 $I_{ij}^{\text{obj}} = 0$。在置信度误差中，λ_{noobj} 是置信度误差的权重，参考 TinyYOLOv3 网络的设置，取 $\lambda_{\text{noobj}} = 0.5$，$I_{ij}^{\text{noobj}} = 1$ 时表示第 i 个网格的第 j 个边界框中不包含目标物体，若包含目标则取值 $I_{ij}^{\text{noobj}} = 0$。$C_i$ 为真实置信度，$\tilde{C}_i$ 为目标边界框的预测置信度。***class*** 为不同手势的类别。在分类误差中，c 为检测到的目标的所属类别，$\tilde{p}_i(c)$ 指 c 类目标在网格 i 中的置信度，$p_i(c)$ 为预测值。

1.5　通道注意力模块

经过不同类型的卷积核进行卷积后的图片会产生不同的信息，不同通道的信息对于检测目标的特征表达能力是不同的。通道注意力模块（squeeze and excitation，SE）（Hu 等，2018）通过对不同特征通道之间的关系进行建模，可以有效地学习到不同通道对于关键信息的权重值，并通过抑制无关信息，加强相关信息的权重值，达到优化网络性能的目的。通道注意力模块结构如图 5 所示。

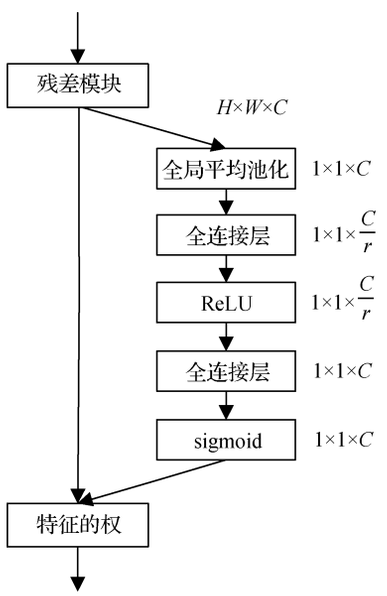


图 5　通道注意力模块（Hu 等，2018）

Fig. 5　Module of channel attention (Hu et al., 2018)

1）特征压缩。经过 C 个不同卷积核卷积后得到的是 $H \times W \times C$ 大小的特征图，而每个特征通道是 $H \times W$ 的 2 维矩阵。通过全局平均池化对特征图进行空间维度上的压缩，使其变为 1 个 1 维矢量的特征描述子。

2）特征激励。包含两个全连接层和 1 个 ReLU（rectified linear units）激活函数。通过第 1 个全连接层对输入特征进行降维，激活函数为 ReLU，然后再使用 1 个全连接层恢复特征维数为 C，最后经过 sigmoid 激活得到 0～1 之间归一化的权重。通过两个全连接层对特征压缩后的结果做非线性变化，同时可以减少参数量和运算量。

3）特征重标定。通过特征激励实现对特征图中不同通道的重要性的评估，并得到 1 个 1 维权重，使用该权重信息完成对原始特征的重标定。

为了提升网络性能，本文将通道注意力模块融合到 TinyYOLOv3 的特征提取网络中。过多地添加注意力模块会增加网络模型的复杂度，本文中将通道注意力添加到主干网络后的特征提取层上，在尽可能地减少参数量和计算量的同时，增加网络的特征提取能力，改进后的网络结构如图 6 所示。图中 ***X***1 和 ***X***2 为特征提取主干网络中不同尺度的输出，SE 为在 13×13 和 26×26 两个不同尺度的特征输出后添加的通道注意力模块，通过添加的通道注意力模块重新分配不同通道的权重信息，提高网络的检测精度。***y***1 和 ***y***2 为不同尺度的特征输出，对两个尺度的输出进行融合，最后通过非极大值抑制得出最终的检测结果。

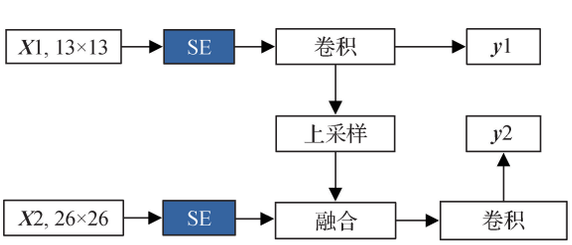


图 6　特征融合模块

Fig. 6　Module of feature fusion

2　实验结果及分析

2.1　数据集及实验环境

实验所使用的数据集是公开的 NUS-II（National

University of Singapore)手势数据集(Pisharady 等,2013),该数据集的部分数据图像如图7所示。该手势数据集由50位不同年龄不同种族的个体在室内及室外各种复杂的背景环境下采集制作的,共包含10种不同的手势,而且采集者的肤色也存在巨大的差异。该数据集包括两个子集,其中一个子集是在不同的自然背景下制作完成的,共包含2 000幅RGB图像;另一个子集包含人体肤色的干扰,背景包含人脸和行人等,共包含750幅RGB图像。为了验证本文算法对于类肤色等复杂背景的鲁棒性,实验中将两个子集中的所有手势图像合并为一个数据集,按照7∶2∶1的比例将数据集随机划分为训练集、测试集和验证集,并使用随机翻转、平移等实时的数据增强方式来避免过拟合并提高训练模型的泛化能力。

本实验的运行环境为:Windows10操作系统,8 GB内存,NVIDIA GTX1650显卡,Python3.6,使用keras深度学习框架进行模型训练和测试,并使用Adam优化器进行优化,训练初始学习率设置为0.005,最大轮次(epoch)为100,batchsize为4,并使用衰减学习率和提前终止训练策略,衰减参数设置为$\alpha=0.000\ 5$,训练过程中监控每个epoch中验证集的损失大小,当损失值在连续的10个epoch的训练过程中都没有减小,则结束模型训练。

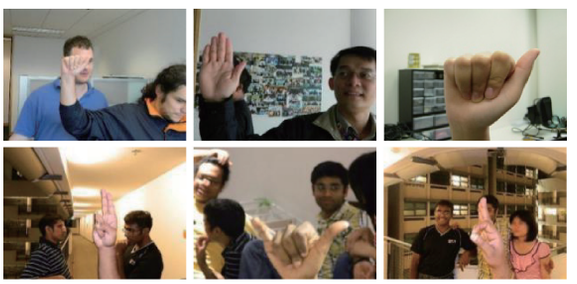


图7 NUS-II数据集

Fig. 7 NUS-II dataset

2.2 结果与分析

2.2.1 不同模块效果对比

为了加快网络模型的收敛速度并降低训练难度,针对本文数据集使用K-means聚类算法确定先验框尺寸。K-means算法通过计算每个样本点到初始聚类中心的欧氏距离,将每个样本点分配到距离最近的聚类中心,并更新聚类中心点,直至函数收敛。对于数据集的目标框进行聚类,得到6个候选框,宽高分别为(38,75),(47,92),(58,105),(59,63),(81,81),(103,152)。

为了验证改进TinyYOLOv3算法的各个模块的效果,在改进算法的基础上,对通道注意力模块(SE模块)、CIoU损失、深度可分离卷积的网络结构改进和目标框聚类等不同模块进行消融实验,其结果如表2所示。由表2可以得知,当对所有模块进行融合时,手势识别率有显著提升,相比TinyYOLOv3算法,准确率提升了8.5%。同时可以看出,对于每一个单独模块的改进,准确率都有1%~3%左右的提升。而在改进整体结构后,准确率的提升最为显著。通过结构改进,使用下采样操作代替TinyYOLOv3中的最大池化,并使用深度可分离卷积在尽可能减少网络参数量的同时增加网络的深度,确保提取到更多的有效特征,最终准确率提升了6.41%。

表2 不同模块对网络的影响

Table 2 Influence of different modules on the network

SE模块	CIoU	结构改进	聚类	准确率/%
				90.60
√				93.43
	√			93.83
		√		97.01
			√	91.31
√	√	√		98.53
√	√		√	94.61
√		√	√	97.63
	√	√	√	98.21
√	√	√	√	99.10

2.2.2 CIoU性能分析

为了更清晰直观地展现CIoU损失函数的优势,本文进行了一系列对比实验。图8为TinyYOLOv3算法、使用CIoU改进的TinyYOLOv3算法和本文算法在训练过程中的损失曲线对比图。由图中结果可知,使用CIoU改进损失函数后的模型比原始TinyYOLOv3的收敛速度更快,TinyYOLOv3、CIoU改进的TinyYOLOv3和本文算法的最终训练损失分别约为4.63、2.63和1.21。相比原算法,本文算法的最终损失降低了3.42;同时表明单独使用CIoU对原算法进行修改也有助于损失的降低,加快网络收敛

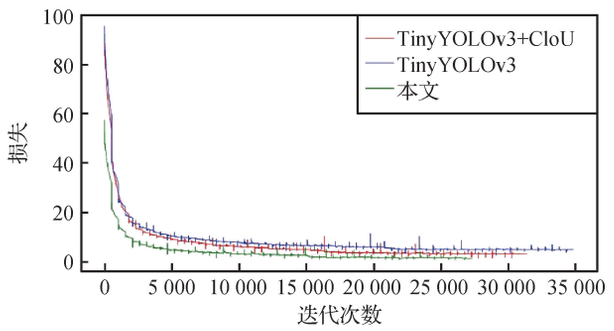


图 8　不同模型训练损失曲线

Fig. 8　Training loss curves of different models

速度。

表 3 为在本文算法中使用 GIoU、DIoU 和 CIoU 损失替换原有边界框预测损失函数的检测结果对比图。由表 3 对比结果可知，使用 CIoU 损失时，准确率达到最高，分别比 GIoU 和 DIoU 提升 0.08%、0.06%。图 9 为使用不同损失的模型在训练过程中验证集的准确率曲线对比图。由图 9 可以得知，本文算法比 TinyYOLOv3 算法的准确率更高，而且由于网络模型的加深和添加的通道注意力模块，使得训练初期能够提取到更多的有效特征，加快了模型收敛速度，减少训练时间。本文算法在训练 40 个 epochs 后基本收敛，而 TinyYOLOv3 则在 50 个 epochs 后才达到收敛状态。同时，在使用不同损失函数的准确率对比曲线上可以看出，在使用 CIoU 损失时，网络的收敛速度更快，并且波动更小。图 10 为使用不同损失的检测结果对比图，其中浅蓝色框为真实标注框，绿色框为检测框。对比图中结果可知，使用 CIoU 损失的检测框更加贴近真实标注框。

表 3　边界框预测损失函数的对比结果

Table 3　Comparison results of bounding box prediction loss function

损失函数	准确率/%
GIoU	99.02
DIoU	99.04
CIoU	99.10

2.2.3　本文算法与其他算法比较

为了验证本文算法的有效性，将本文算法与 YOLOv3（Redmon 和 Farhadi，2018）和 SSD300（Liu 等，2016）等经典深度学习目标检测算法以及 TinyYOLOv3、TinyYOLOv3 MobileNetV1、TinyYOLOv3 MobileNetV2、MobileNet-SSD（Howard 等，2017）和 YOLO-lite（Huang 等，2018）等轻量级目标检测算法进行比较，其中 TinyYOLOv3 MobileNetV1 和 TinyYOLOv3 MobileNetV2 算法是指分别用 MobileNetV1 和 MobileNetV2 替换 TinyYOLOv3 的主干网络而构成的算法。各算法在准确率、检测时间、模型大小和参数量等不同指标的对比结果如表 4 所示。

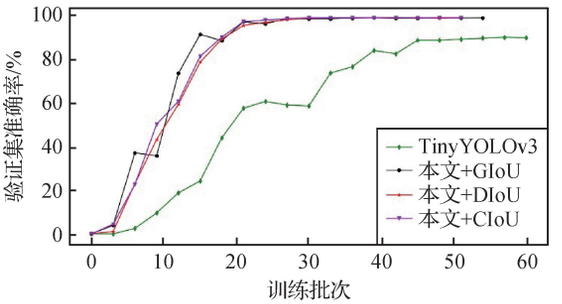
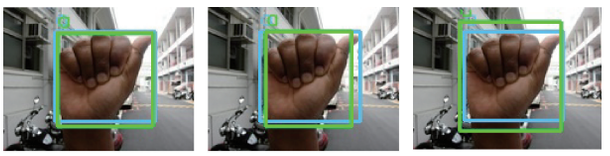


图 9　不同模型准确率变化曲线

Fig. 9　Accuracy curve for different models



(a) CIoU　(b) DIoU　(c) GIoU

图 10　不同损失函数检测结果对比

Fig. 10　Comparison of different loss function detection results((a) CIoU; (b) DIoU; (c) GIoU)

由表 4 中结果可知，在准确率的比较上，YOLOv3 算法比本文算法高 0.54%，但是在检测时间、模型大小和参数量等其他指标的比较上，本文算法都远优于 YOLOv3。由于 YOLOv3 网络存在较多的卷积层和残差连接，使其模型和参数量都比较大，而本文算法的模型大小仅有 27.6 MB，约为 YOLOv3 的 1/8，参数量约为 YOLOv3 的 1/9。SSD 使用 VGG16（visual geometry group 16-layer net）作为主干特征提取网络，并使用 6 个尺度融合的检测方式，但 VGG16 的网络结构比较简单，其卷积层和全连接层共 16 层，SSD 在网络结构设计上使用卷积层代替 VGG16 的全连接层，并且增加了额外的卷积层来提取特征，使得 SSD 的模型和参数量都比较大，性能远不如本文算法。在同轻量级网络的比较上，本文算法也取得了较优的结果。由于 TinyYOLOv3 的结构简单，具有良好的实时性，但是其准确率却比较

表 4　不同算法对比结果

Table 4　Comparison results of different algorithms

算法	准确率/%	检测时间/ms	模型大小/MB	参数数量
YOLOv3	99.64	58	236	6.1×10^7
SSD300	99.37	103	92.1	2.42×10^7
TinyYOLOv3	90.60	20	33.2	9.0×10^6
TinyYOLOv3 MobileNetV1	82.95	33	46.7	1.2×10^7
TinyYOLOv3 MobileNetV2	99.04	41	45.6	1.1×10^7
MobileNet-SSD	95.99	45	23.3	5.7×10^6
YOLO-lite	37.98	11	2.30	6.0×10^5
本文	99.10	28	27.6	7.13×10^6

低。本文算法的检测时间比 TinyYOLOv3 多 8 ms，但是准确率远高于 TinyYOLOv3，同时由于本文算法采用了深度可分离卷积，并适当减小了特征图的数量，因此模型大小和参数量也略优于 TinyYOLOv3。同理，由于 YOLO-lite 简单的结构设计，使其在所有算法的对比中，检测时间、模型大小和参数数量都达到了最优，但该算法准确率只有 37.98%，难以达到实际应用的要求。在不同指标的对比结果上，本文算法均优于 TinyYOLOv3 MobileNetV1 和 TinyYOLOv3 MobileNetV2。MobileNet-SSD 的模型大小虽然只有 23.3 MB，但其检测时间比本文算法多 17 ms，且准确率也比本文算法低 3.11%。由此可知，本文算法在不同指标的比较上都达到了较优的结果，而且较小的网络模型为在移动端和嵌入式设备上的移植提供可能，同时也保证了检测精度和实时性。

表 5 是本文算法与机器学习以及传统手势识别相关算法在准确率上的对比。由于 NUS-II 手势数据集的背景复杂，存在许多类肤色背景的干扰，导致使用 CNN 和 SVM 算法（吴晴，2018）提取的手势信息有限，识别率较低。基于贝叶斯注意力和 SVM 的算法（Pisharady 等，2013）将手势识别分为人手检测和手势分类两部分，首先使用注意力模型提取图像中的人手区域部分，然后使用视觉皮层提取手势特征并设计 SVM 分类器进行手势识别。由于视觉皮层对于手势特征的表达相对简单，尤其是当图片中存在肤色或者其他类肤色背景时，提取到的肤色特征可能是无效的，因此该算法在复杂背景下的手势识别率较低。同理，由于复杂背景下类肤色的干扰，导致高斯模型（Jia 等，2008）和弹性图匹配（Triesch 和 von der Malsburg，2001）等算法的手势识别率严重降低。

表 5　不同算法对比结果

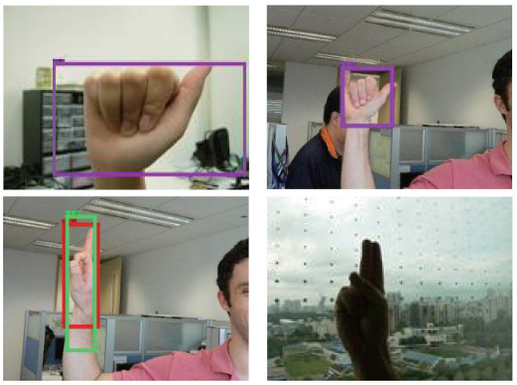
Table 5　Comparison results of different methods

算法	准确率/%
CNN + SVM	93.01
贝叶斯注意力 + SVM	93.71
高斯模型	89.12
弹性图匹配	65.96
本文	99.10

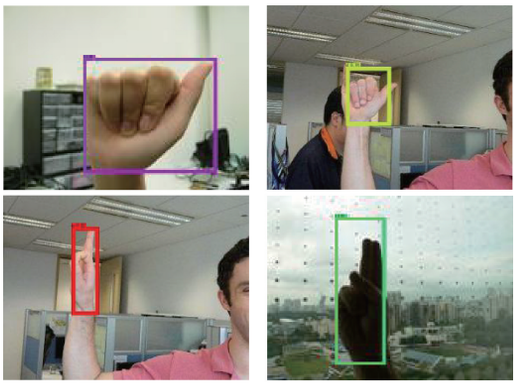
图 11 为 TinyYOLOv3 和本文算法在本文所划分的测试集中部分图片检测结果的对比图，其中图 11（a）为待检测图像，图 11（b）（c）分别为 TinyYOLOv3 和本文算法的检测结果。由图中的 TinyYOLOv3 检测结果可以看出，第 1 幅图中手势定位精度较低，第 2 幅和第 3 幅图中出现了错检，第 4 幅图像中出现了漏检的情况。而本文算法能够准确地识别手势，并取得了较好的定位精度。由于 TinyYOLOv3 网络结构简单，只能够提取到有限的手势特征，因此检测结果的定位精度较低，当手势区域较小或背景比较复杂时容易出现错检和误检的情况，而且当光线较暗、手势特征不明显时易出现漏检。相较TinyYOLOv3，本文算法使用深度可分离卷积对其网络结构进行加深，并添加通道注意力模块，增强了网络的特征提取能力，并使用 CIoU 损失改善定位精度，使用数据增强

和参数优化等多种方法来提高网络模型的泛化能力，使得改进后的模型在复杂背景下的手势识别取得了较好的效果。

(a) 待检测原图



(b) TinyYOLOv3检测结果



(c) 本文算法检测结果

图 11　本文算法和 TinyYOLOv3 检测结果对比

Fig. 11　The comparison between our algorithm and TinyYOLOv3 ((a) original images to be tested; (b) detection results of TinyYOLOv3; (c) detection results of ours)

为了充分验证本文算法的鲁棒性以及泛化能力，实验人员在室内以及室外等现实环境中采集了 4 种不同手势的图片进行验证实验，其中每种手势采集 100 幅图像，共 400 幅图像。实验结果表明，本文算法能较好地克服复杂背景的干扰，手势识别准确率达到 97.3%，部分检测结果如图 12 所示。此外，本文使用 Marcel 手势数据集(Marcel，1999)进行验证实验，该数据集共有 6 种不同手势，5 494 幅图像。与其他算法的实验结果对比如表 6 所示。由于 Marcel 数据集中的图像背景较为简单，且基本不存在类肤色背景的干扰，因此肤色 + CNN(王龙 等，2017)和高斯模型(Jia 等，2008)算法都取得了较好的检测结果，但是本文算法的准确率达到 99.21%，比肤色 + CNN 和高斯模型算法分别高 2.01% 和 5.21%。在以上实验中，本文算法均取得了较高的准确率，充分验证了本文算法的鲁棒性以及泛化能力。

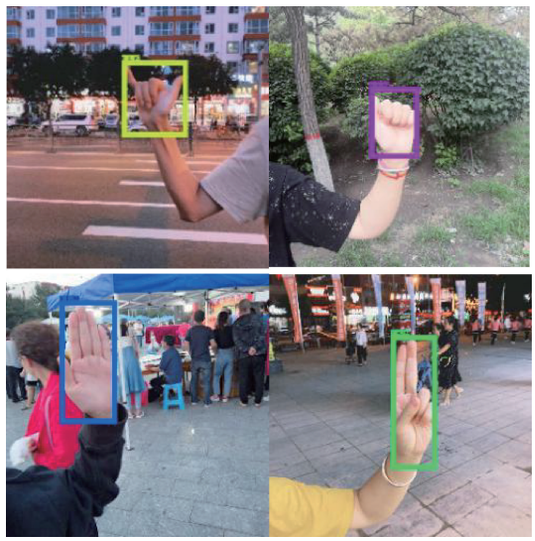


图 12　本文算法部分检测结果

Fig. 12　Partial detection results of proposed algorithm

表 6　Marel 数据集不同算法对比结果

Table 6　Comparison results of different algorithms on Marcel dataset

算法	准确率/%
肤色 + CNN	97.20
高斯模型	94.00
本文	99.21

3　结　论

本文提出一种改进的 TinyYOLOv3 手势识别算法。使用深度可分离卷积对 TinyYOLOv3 特征提取主干网络进行重新设计，通过减少特征通道数降低模型复杂度，使用 CIoU 对损失函数进行改进，在特

征融合输出通道上添加通道注意力模块,使用数据增强、超参数优化和动态学习率更新等方法对网络模型进行训练。实验结果表明,本文算法对复杂背景下的手势具有良好的识别能力,相比其他检测算法,本文算法兼顾了识别率、检测时间和模型大小之间的平衡。

后续工作将改进多尺度特征融合方法,提取更丰富的浅层特征,以改善对小尺度手势的识别效果;通过模型量化压缩和模型剪枝相关方法对模型进行进一步压缩,在保证准确率的同时,进一步提升检测速度。

参考文献(References)

Girshick R, Donahue J, Darrell T and Malik J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE: 580-587 [DOI: 10.1109/CVPR.2014.81]

Gong X T, Ma L and Ouyang H K. 2020. An improved method of Tiny YOLOV3. IOP Conference Series: Earth and Environmental Science, 440(5): #052025 [DOI: 10.1088/1755-1315/440/5/052025]

Gupta B, Shukla P and Mittal A. 2016. K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion//Proceedings of 2016 International Conference on Computer Communication and Informatics. Coimbatore, India: IEEE: 1-5 [DOI: 10.1109/ICCCI.2016.7479951]

Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T, Andreetto M and Adam H. 2017. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2020-06-02]. https://arxiv.org/pdf/1704.04861.pdf

Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]

Huang R, Pedoeem J and Chen C X. 2018. YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers//Proceedings of 2018 IEEE International Conference on Big Data (Big Data). Seattle, USA: IEEE: 2503-2510 [DOI: 10.1109/bigdata.2018.8621865]

Iandola F N, Han S, Moskewicz M W, Ashraf K, Dally W J and Keutzer K. 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size [EB/OL]. [2020-06-02]. https://arxiv.org/pdf/1602.07360.pdf

Jia J, Jiang J M and Wang D. 2008. Recognition of hand gesture based on Gaussian mixture model//Proceedings of 2008 International Workshop on Content-Based Multimedia Indexing. London, UK: IEEE: 353-356 [DOI: 10.1109/CBMI.2008.4564968]

Jiang D, Zheng Z J, Li G F, Sun Y, Kong J Y, Jiang G Z, Xiong H G, Tao B, Xu S, Yu H, Liu H H and Ju Z J. 2019. Gesture recognition based on binocular vision. Cluster Computing, 22(6): 13261-13271 [DOI: 10.1007/s10586-018-1844-5]

Liu S P, Liu Y, Yu J and Wang Z F. 2015. Hierarchical static hand gesture recognition by combining finger detection and HOG features. Journal of Image and Graphics, 20(6): 781-788 (刘淑萍, 刘羽, 於俊, 汪增福. 2015. 结合手指检测和HOG特征的分层静态手势识别. 中国图象图形学报, 20(6): 781-788) [DOI: 10.11834/jig.20150607]

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C. 2016. SSD: single shot MultiBox detector//Proceedings of the European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 21-37 [DOI: 10.1007/978-3-319-46448-0_2]

Lyu N, Yang X H, Jiang Y and Xu T. 2017. Sparse decomposition for data glove gesture recognition//Proceedings of the 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics. Shanghai, China: IEEE: 1-5 [DOI: 10.1109/CISP-BMEI.2017.8302114]

Marcel S. 1999. Hand posture recognition in a body-face centered space//Proceedings of the CHI'99 Extended Abstracts on Human Factors in Computing Systems. Pittsburgh, USA: ACM: 302-303 [DOI: 10.1145/632716.632901]

Pisharady P K, Vadakkepat P and Loh A P. 2013. Attention based detection and recognition of hand postures against complex backgrounds. International Journal of Computer Vision, 101(3): 403-419 [DOI: 10.1007/s11263-012-0560-5]

Priyal S P and Bora P K. 2013. A robust static hand gesture recognition system using geometry based normalizations and Krawtchouk moments. Pattern Recognition, 46(8): 2202-2219 [DOI: 10.1016/j.patcog.2013.01.033]

Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 779-788 [DOI: 10.1109/CVPR.2016.91]

Redmon J and Farhadi A. 2017. YOLO9000: better, faster, stronger//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 6517-6525 [DOI: 10.1109/CVPR.2017.690]

Redmon J and Farhadi A. 2018. YOLOv3: an incremental improvement [EB/OL]. [2020-06-02]. https://arxiv.org/pdf/1804.02767.pdf

Ren S Q, He K M, Girshick R and Sun J. 2015. Faster R-CNN: towards real-time object detection with region proposal networks//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: NIPS: 91-99

Rezatofighi H, Tsoi N, Gwak J Y, Sadeghian A, Reid I and Savarese S.

2019. Generalized intersection over union: a metric and a loss for bounding box regression//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 658-666 [DOI: 10.1109/CVPR.2019.00075]

Sandler M, Howard A, Zhu M L, Zhmoginov A and Chen L C. 2018. MobileNetV2: inverted residuals and linear bottlenecks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 4510-4520 [DOI: 10.1109/CVPR.2018.00474]

Tan MX, Brain G and Le Q V. 2019. MixConv: mixed depthwise convolutional kernels [EB/OL]. [2020-06-02]. https://arxiv.org/pdf/1907.09595.pdf

Triesch J and von der Malsburg C. 2001. A system for person-independent hand posture recognition against complex backgrounds. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(12): 1449-1453 [DOI: 10.1109/34.977568]

Wang L, Liu H, Wang B and Li P J. 2017. Gesture recognition method combining skin color models and convolution neural network. Computer Engineering and Applications, 53(6): 209-214 (王龙, 刘辉, 王彬, 李鹏举. 2017. 结合肤色模型和卷积神经网络的手势识别方法. 计算机工程与应用, 53(6): 209-214) [DOI: 10.3778/j.issn.1002-8331.1508-0251]

Wu B C, Wan A, Iandola F, Jin P H and Keutzer K. 2017. SqueezeDet: unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu, USA: IEEE: 129-137 [DOI: 10.1109/CVPRW.2017.60]

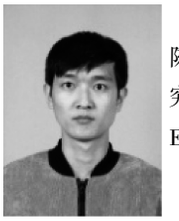
Wu Q. 2018. Research on Gesture Recognition Algorithm Based on Improved CNN and SVM. Nanchang: Jiangxi Agricultural University (吴晴. 2018. 基于改进的 CNN 和 SVM 手势识别算法研究. 南昌: 江西农业大学)

Zhang X Y, Zhou X Y, Lin M X and Sun J. 2018. ShuffleNet: an extremely efficient convolutional neural network for mobile devices//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 6848-6856 [DOI: 10.1109/CVPR.2018.00716]

Zheng Z H, Wang P, Liu W, Li J Z, Ye R G and Ren D W. 2020. Distance-IoU loss: faster and better learning for bounding box regression//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA: AAAI: 12993-13000 [DOI: 10.1609/aaai.v34i07.6999]

作者简介

王银，1982 年生，男，副教授，主要研究方向为计算机视觉、机器学习。
E-mail: yinwang@tyust.edu.cn



陈云龙，通信作者，男，硕士研究生，主要研究方向为图像处理与人工智能。
E-mail: s20180411@stu.tyust.edu.cn

孙前来，男，副教授，主要研究方向为图像处理、机器视觉。
E-mail: sql_sun@tyust.edu.cn