文章编号：2095-2163(2022)08-0102-05 中图分类号：TP391 文献标志码：A

改进目标检测网络的仰卧起坐测试计数

包梓群
（浙江理工大学 信息学院，杭州 310018）

摘 要：针对深度学习技术在仰卧起坐测试领域的实时性较差问题，提出了一种改进目标检测网络的仰卧起坐测试计数算法。该算法首先对被测试人员进行目标检测，然后对被测试人员进行关键点提取，最后对被测试人员的仰卧起坐动作进行分析。为了达到实时检测的效果，改进了 RetinaNet 骨干网络中的传统卷积层，以减小计算量，加快识别速度；提出了一种改进的边框损失函数，以达到实时检测效果的同时，保证检测的精度。经对其算法进行仿真实验，验证了其识别速度和检测精度，达到了预期效果。

关键词：仰卧起坐实时检测；目标检测；RetinaNet 网络；边框损失函数

Sit-ups test counting based on improved target detection network

BAO Ziqun
（School of informatics Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China）

【Abstract】 Aiming at the poor real-time performance of sit-ups test, a sit-ups test counting algorithm based on improved target detection network is proposed in this paper. The algorithm first detects the target of the tested person, then extracts the key points of the tested person, furtherly analyzes the sit-ups of the tested person. In order to achieve the effect of real-time detection, the traditional convolutional layer in the backbone network of RetinaNet is improved to reduce the amount of calculation and speed up the recognition speed. At the same time, an improved frame loss function is proposed to ensure real-time detection and detection accuracy. Finally, the simulation experiment of the algorithm is carried out to verify its recognition speed and detection accuracy, and the expected effect is achieved.

【Key words】 real time detection of sit-ups; target detection; RetinaNet; frame loss function

0 引 言

仰卧起坐是国内各个阶段学生体育测试中的一项重要运动。在日常测试过程中，需要人工对其动作是否规范进行评判并计数。随着深度学习中目标检测技术被广泛应用于生产生活中，同时也为机器检测的实现提供了技术基础。目标检测为当今计算机视觉的热门研究方向。其主要工作就是预测目标在视频或者图片中的具体位置，现已在安防、自动驾驶、行为分析等应用领域起着至关重要的作用[1]。

目标检测[2]，旨在从数字图像中检测出特定类别的实例，这是计算机视觉中一项基本且具有挑战性的任务[3]。但近年来，随着卷积神经网络（Convolutional Neural Network, CNN）[4]不断发展与演进，使得目标检测算法越来越成熟。采用 CNN 的目标检测算法因其在特征提取上具有良好的泛化性，逐渐取代了基于人工特征的目标检测算法。基于 CNN 的目标检测算法在不同的场景中产生多种类型：

（1）基于区域候选目标检测算法，如 Faster-RCNN[5-6]。

（2）基于端到端回归的算法模型，如 YOLO[2,6]、RetinaNet[7]。

2 种模型的特点较为明显：基于区域候选的模型可以得到较好的检测准确率，但检测速度较慢；基于回归的模型目标检测速度快，但准确率较低[8]。为了使模型能在实时检测的同时又不有损精度，本文提出了一种改进的 RetinaNet 网络目标检测算法。

1 网络模型

RetinaNet 网络模型主要由主干网络、颈部网络、分类子网络和回归自网络组成。其中，主干网络即为卷积池化层的堆叠网络，一般为 ResNet[9] 网络和 VGG[10] 网络；颈部网络则用于特征的堆叠和融

基金项目：国家级大学生创新创业训练计划项目（202010338024）；浙江省教育厅一般科研项目（Y202147659）；浙江省重点研发计划项目（2020C03094）；国家自然科学基金（6207050141）。

作者简介：包梓群（2001-），男，本科生，主要研究方向：计算机视觉、智能图像处理。

收稿日期：2022-01-27

合，一般使用特征金字塔网络（Feature Pyramid Network，FPN）[11]。FPN 将多尺度特征加以融合，使得最后的预测结果包含各帧图片内各个尺度的信息，模型的性能也得以提升；分类子网络利用全卷积层对颈部网络的输出进行处理，再对图像中的目标对象去做类别预测；回归子网络利用全卷积层对颈部网络的输出进行处理，并对图像中的目标对象实现定位。RetinaNet 网络模型结构如图 1 所示。

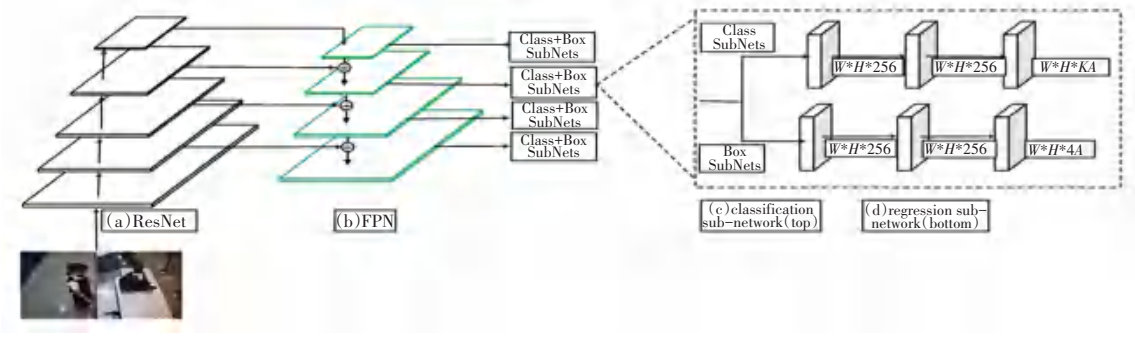


图 1　一般 RetinaNet 网络模型图

Fig. 1　General RetinaNet network model diagram

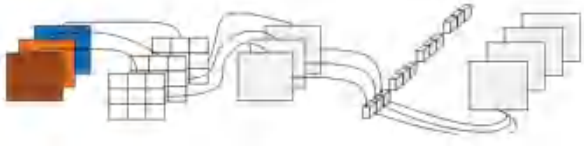
2　改进的 RetinaNet 网络目标检测算法

针对原始网络对于视频中目标检测精度不高的问题，采用 2 个 RetinaNet 网络模型级联，即将第一个 RetinaNet 网络的输出作为第二个 RetinaNet 网络的输入，用来对待检测图像进行目标检测，增加模型的泛化能力。但是网络模型的级联会增加资源消耗和参数数量。为了解决此问题，引入深度可分离卷积（Depth Separable Convolution，DSC）[12] 取代原始的卷积模块，以降低网络级联带来的资源消耗和计算量。使用深度卷积模块，虽然简化了模型的骨干网络，但会弱化模型的特征提取能力，导致模型精度下降。为此，提出了一种新的 *LIoU* 函数，用来计算定位框的损失，弥补丢失的精度。

2.1　深度可分离卷积

深度可分离卷积（Depth Separable Convolution，DSC）是把常规卷积分为深度卷积（Depthwise Convolution，DW）和点卷积（Pointwise Convolution，PW）两个阶段[12]。其中，DW 阶段实质上起到一个滤波的作用，通过使用和输入图像通道数相同的卷积核，提取每一个单独通道的特征信息。PW 阶段可以看作是对 DW 阶段的输出进行组合的过程，使用一个 $1 \times 1 \times C_{in} \times C_{out}$（这里，$C_{in}$ 为输入通道数，C_{out} 为输出通道数）的卷积核对 DW 的输出进行整合，其结构如图 2 所示。

深度可分离卷积的计算量和参数量都比一般卷积要小，可以极大地增加模型检测的速度，满足仰卧起坐实时检测的需求。但因其简化了特征提取模块，因此就需要改进预测回归的损失函数，来弥补准确率的丢失。



（a）深度卷积过程　　（b）点卷积过程

图 2　深度可分离卷积

Fig. 2　Depth Separable Convolution

2.2　损失函数的改进

在目标检测中，常常利用预测框（Prediction Box，P）与真实框（Ground Truth，G）之间的交并比（Intersection over Union，IoU）[13] 作为衡量两者之间关系的重要度量，IoU 的计算公式如下：

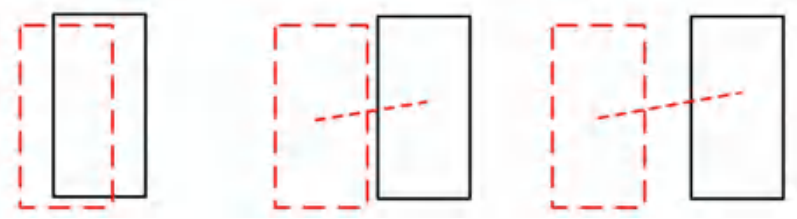
$$IoU = \frac{|P \cap G|}{|P \cup G|} \tag{1}$$

相对于其它预测回归的损失函数，IoU 具有更好的效果，但也存在一定的局限性。研究给出了几种预测框与真实框间的关系如图 3 所示。由图 3（b）、图 3（c）可知，当预测框与真实框没有重叠时，2 种情况的 IoU 损失值相同，但图 3（b）的效果略好于图 3（c）。另一方面，当损失函数的值为 0 时，在反向传播中其梯度为 0，无法对网络进行优化。当初始值选择不佳时，会使训练出来的模型拟合效果极差。

为了解决上述问题，对 IoU 回归损失函数进一步优化。对此可表示为：

$$F_{IoU} = 1 - IoU \tag{2}$$

由式（2）可知，当 IoU 的值为 0 时，F_{IoU} 恒为 1，可见单凭借 F_{IoU} 依然无法很好地进行优化。为解决此问题，在 F_{IoU} 中加入正则项 $Smooth_{L_1}$。$Smooth_{L_1}$ 函数可以计算预测框与标准框之间的位置偏差，此



(a) 预测框和真实框重合 (b) 预测框和真实框不重合 (c) 预测框和真实框不重合

图 3 几种预测框与真实框之间的关系

Fig. 3 Relationship between several prediction frames and real frames

时用到的数学公式可写为:

$$Smooth_{L1}(x)=\begin{cases}0.5x^2 & \text{if } |x|<1\\ |x|-0.5 & \text{otherwise}\end{cases} \tag{3}$$

其中, x 表示真实框与预测框之间的偏差。

将其求导得到:

$$\frac{\mathrm{d}Smooth_{L1}}{\mathrm{d}x}=\begin{cases}x & \text{if } |x|<1\\ \pm 1 & \text{otherwise}\end{cases} \tag{4}$$

从式(4)中可以看出, $Smooth_{L1}$ 在 x 较小时,对 x 的梯度也较小;当 x 变大时,也只能在 1 之内,限制了梯度跌涨自由,不会破坏网络参数,解决了当 x 值比较大时导致训练损失值出现明显变化而引起的训练不稳定问题。

比较 F_{IoU} 和 IoU 函数,引入正则项,改变了 I_{oU} 的正负号,使得其与正则项有一样的梯度朝向。构建预测回归的损失函数 L_{IoU} 公式如下:

$$L_{IoU}=F_{IoU}+\lambda\cdot Smooth_{L1}(y,\hat{y}) \tag{5}$$

其中, y 表示预测框; $\hat{y}$ 表示真实框; λ 为平衡因子。

由式(5)可知,若出现被测试人员半卧起、卧起姿态的特殊情况时,虽然 2 种情况具有相同的 IoU 值,即与公式(5)中的 F_{IoU} 值相同。但得益于 $Smooth_{L1}$ 正则项,边框损失函数 L_{IoU} 的梯度仍然可以得到反向传播。实验测得 $\lambda=3$ 时,在仰卧起坐测试上有着较好的效果。

3 人体姿态估计

人体姿态估计、即关键点检测, 目的是检测人体身上 K 个关键点的位置(头部、手肘、膝盖等),抽象出人体的当前行为。目前,最先进的方法是把该问题转变为估计 K 热图。需要一提的是,每个热图 Hk 的值,表示第 k 个关键点的位置置信度。

在网络设计方面,当前大多数方法都是将高分辨率到低分辨率的子网络串联起来,且每个子网络形成一个阶段,相邻子网络之间存在一个下采样层,将分辨率缩小一半。本文采用 HRNet 并行地连接高到低的子网,保持了高分辨率的表示,生成了整个过程的空间精确热图估计。通过重复融合高到低子网产生的高分辨率,生成可靠的高分辨率表示。

本文将 HRNet 引入到模型中,测试时被测试人员各个姿态的关键点效果如图 4 所示。

具体地,图 4(a)表示被测试人员平躺姿态的骨架图; 图 4(b)表示被测试人员半卧起姿态的骨架图;图 4(c)表示被测试人员卧起姿态的骨架图。

(a) 被测试人员平躺姿态 (b) 被测试人员半卧起姿态 (c) 被测试人员卧起姿态

图 4 被测试人员各个姿态的关键点提取

Fig. 4 Key points extraction of each pose of the tested target

4 实验结果分析

实验所用的计算机系统配置:CPU 为 Intel(R) Xeon(R) Silver 4110 CPU @ 2.10 GHz;GPU 为 24 G RTX3090 显卡;主频为 4.00 GHZ;系统为 CentOS 7.7。采用 Python3.6 语言编写实验代码,深度学习框架选取 Pytorch1.4.0。

4.1　性能评价指标

由于将网络用于仰卧起坐的实时检测具有一定的特殊性，无法使用召回率、F_1 等常用指标来进行评价。因此，本文设计了一些合理的评价指标，用于实验检测，对此拟做阐释分述如下。

采集 100 次 1 min 仰卧起坐测试实时计数数据集。其中，设定 $\hat{I}$ 为测试得到仰卧起坐数量的集合，内有 100(Num) 个元素；I 为实际测得仰卧起坐数量的集合，内有 100(Num) 个元素。

(1)平均测得仰卧起坐数量可由如下公式计算求出：

$$mTestNum = \frac{\Sigma \hat{I}}{Num} \tag{6}$$

(2)平均实际仰卧起坐数量。可由如下公式计算求出：

$$mTrueNum = \frac{\Sigma I}{Num} \tag{7}$$

(3)平均反应时间。可由如下公式计算求出：

$$\frac{1\ \text{min}}{mTestNum} - \frac{1\ \text{min}}{mTrueNum} \tag{8}$$

(4)平均测试准确率。可由如下公式计算求出：

$$1 - \frac{|mTrueNum - mTestNum|}{mTrueNum} \tag{9}$$

4.2　实验结果以及分析

为了验证本文提出的改进目标检测网络相对于原始效果有所提升，使用 4.1 节中采集的数据集展开对比试验，并使用上述指标进行评价。实验结果见表 1。为了得出式(5)中最好的超参数 λ，在[0，10]的区间内，设置步长为 1 进行调参。实验结果如图 5 所示。

表 1　改进前后网络各个指标具体数值表

Tab. 1　Specific values of network indicators before and after improvement

模型	平均测得仰卧起坐数量	平均实际仰卧起坐数量	平均模型反应时间/ms	平均测试准确率/%	FPS/(F·S^{-1})
改进后的网络(λ = 3)	42.32	43.89	50	96.4	25
改进前的网络	20.44	43.89	1 570	46.6	25

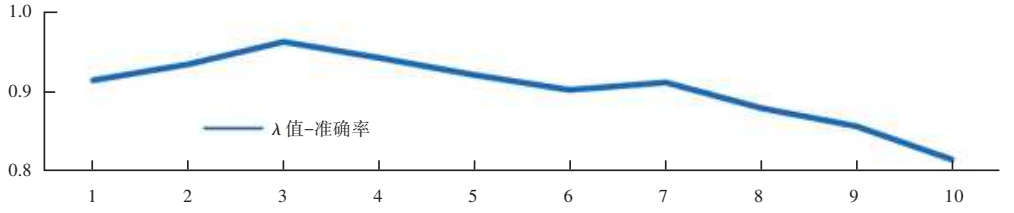


图 5　获取最佳超参数 λ 的实验结果图

Fig. 5　Experimental results of getting the best super parameters λ

由图 5、表 1 可以看出，改进后的网络相对于原网络在速度和精度上有了质的飞跃。模型反应时间由原来的 1 570 ms 降低到了 50 ms，达到了实时检测的效果。得益于速度的增长和损失函数的改进，模型的准确率提高了 0.498。以上结论验证了改进网络的有效性。

5　结束语

本文提出了一种改进的 RetinaNet 网络目标检测算法。为了提高检测效果，将 2 个 RetinaNet 网络级联，采用深度可分离卷积代替了原网络中的骨干模块，以减小级联网络带来的额外计算量；而后对边框损失函数加以改进，引入了 $Smooth_{L1}$ 正则项，在 IoU 给出了重合度信息的基础上，$Smooth_{L1}$ 又提供了预测框与真实框的位置信息，使得网络效果得到提升，并且训练也更加稳定。由实验结果可知，改进后的网络针对仰卧起坐测试计数具有良好的效果，满足正确检测的实时要求。

参考文献

[1] 倪金卉. 基于深度学习的目标检测研究[J]. 无线互联科技，2021，18(19)：115-116.

[2] ZOU Zhengxia，SHI Zhenwei，GUO Yuhong，et al. Object detection in 20 years：a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2020，84(1)：154-191.

[3] 赵立新，邢润哲，白银光，等. 深度学习在目标检测的研究综述[J]. 科学技术与工程，2021，21(30)：12787-12795.

[4] 李玲. 基于卷积神经网络的图像融合算法综述[J]. 信息与电脑(理论版)，2021，33(21)：55-57.

(下转第 109 页)