DOI:10.19651/j.cnki.emt.2107568

基于 YOLOv3 的轻量化口罩佩戴检测算法

薄景文 张春堂
(青岛科技大学 自动化与电子工程学院 青岛 266061)

摘 要: 当前疫情防控形势严峻,在人群密集场所进行实时快速的口罩佩戴检测可以有效降低病毒传播的风险。针对目前人工检测效率低的问题,提出一种基于 YOLOv3 的轻量化口罩佩戴检测算法。使用 ShuffleNetv2 替换原来的主干特征提取网络,降低网络参数量,减少计算功耗。提出将 SKNet 注意力机制引入到特征融合网络部分,增强不同尺度的特征提取能力;使用 CIoU 作为边界框回归损失函数,进一步提高检测精度。在构建的人脸口罩检测数据集上实验表明,与原 YOLOv3 相比,所提算法在保持较高检测精度的情况下,检测速度提高了 34 FPS,有效地实现了准确快速的口罩佩戴检测,与其他主流目标检测算法相比,该算法也具有更好的检测效果。

关键词: 口罩检测;轻量化;YOLOv3;注意力机制;损失函数
中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Lightweight mask wearing detection algorithm based on YOLOv3

Bo Jingwen Zhang Chuntang
(College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: At present, the situation of epidemic prevention and control is grim. Real time and rapid mask wearing detection in crowded places can effectively reduce the risk of virus transmission. Aiming at the low efficiency of manual detection, a lightweight mask wearing detection algorithm based on YOLOv3 is proposed. ShuffleNetv2 is used to replace the original backbone feature extraction network to reduce the amount of network parameters and computing power consumption. SKNet attention mechanism is introduced into the feature fusion network to enhance the ability of feature extraction at different scales. CIoU is used as the boundary box regression loss function to further improve the detection accuracy. Experiments on the constructed face mask detection data set show that, compared with the original YOLOv3, the proposed algorithm improves the detection speed by 34 FPS while maintaining high detection accuracy, and effectively realizes accurate and fast mask wearing detection. Compared with other mainstream target detection algorithms, the algorithm also has better detection effect.

Keywords: mask detection;lightweight;YOLOv3;attention mechanism;loss function

0 引 言

新型冠状病毒肺炎席卷全球,影响着世界各国人民的生命安全,中国秉承人类命运共同体的理念率先在疫情防控方面取得了阶段性的成功。目前国外疫情形式依然严峻,境外入境人员第一时间会出现在机场、火车站、港口码头等人员密集的场所,稍有不慎便会增加病毒传播风险。在人员密集场所佩戴口罩可有效降低病毒传播风险,然而人工检查口罩佩戴情况存在效率低,易漏检的问题,因此本文提出一种基于深度学习的轻量化目标检测算法,实现高效准确的口罩佩戴检测。

自从 2012 年 AlexNet[1] 提出,卷积神经网络再度兴起,Girshick 等[2] 在 2014 年提出 R-CNN(regions with CNN features)用于目标检测任务,基于卷积神经网络的目标检测领域开始了飞速发展。目前,目标检测主要分为两类[3]:1)基于候选区域的两阶段算法,主要以 R-CNN 系列算法为代表,比如 Fast R-CNN[4] 和 Faster R-CNN[5];2)基于回归思想的一阶段算法,主要以 SSD(single shot multibox detector)[6]、YOLO(you only look once)[7] 系列算法为代表。两种算法各有优缺点,两阶段算法的检测精度高,速度慢,难以达到实时性的要求,一阶段算法检测速度快,精度上相对而言较低。YOLOv3[8] 算法的综合性能较强,较好地平衡了检测速度与检测精度,应用领域广泛。例如文献[9]提出改进 YOLOv3 的商品包装检测,应用在无人销售领域,文献[10]提出一种基于 MSRCR 和 MobileNetV2[11]

收稿日期:2021-08-12

的改进 YOLOv3 算法，实现了对海洋动物的实时检测。然而 YOLOv3 的网络结构复杂，对硬件平台资源配置要求较高，难以在常用的低算力设备上达到检测的实时性，因此有必要根据具体目标场景需求的不同来对 YOLOv3 算法做出优化。本文基于 YOLOv3 算法进行改进，提出一种对平台计算能力需求小且检测效果稳定的轻量化 YOLOv3 算法，在保持检测精度基本不变的前提下，有效地提高了检测速度，兼顾了口罩佩戴检测的实时性与准确性。

本文首先将 YOLOv3 的主干特征提取网络 DarkNet53 替换为 ShuffleNetv2[12]，降低了网络参数量，提高了模型推理速度和训练速度；然后将自适应动态选择注意力机制 SKNet(selective kernel networks)[13]引入到 YOLOv3 的特征融合网络部分，避免了轻量化主干网络所导致的特征提取能力不足的问题，仅增加了少量参数同时进一步提高了检测精度；最后使用 CIoU(complete IoU)[14]作为边界框回归损失函数，将目标与先验框间的重叠度、中心距离和长宽比都充分考虑在损失计算当中，再次提高了检测精度。构建了口罩检测数据集，实验结果表明，所提方法与原 YOLOv3 和其他主流算法相比，更好地实现了精度与速度的平衡，可有效地检测人脸与口罩。

1　改进的 YOLOv3 算法

1.1　主干特征提取网络

旷视于 2018 年发布 ShuffleNetv2，指出浮点计算量并非是衡量模型推理速度的唯一标准，内存访问成本也会影响模型推理速度。因此 ShuffleNetv2 作者提出轻量高效的卷积神经网络需具备如下 4 条设计准则：输入输出通道数相同时内存访问成本最低；大量使用组卷积会增加内存访问成本；网络分支过多会降低模型并行度；元素级操作虽然浮点计算量少，但内存访问成本大。基于以上 4 条准则，在 ShuffleNetv1[15]的基础上提出了轻量级网络 ShufflNetv2 的两种核心模块，如图 1 所示。

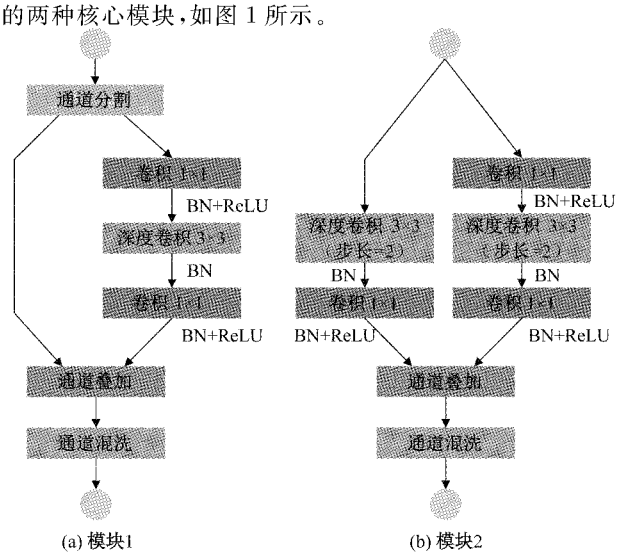

图 1　ShufflcNctv2 两种基本模块

当步长为 1 时选用模块 1，此时输入输出尺寸不变，通道数加深；当步长为 2 时选用模块 2，此时不仅通道数变为两倍，特征图尺寸也要减半。本文算法将 ShuffleNetv2 作为主干特征提取网络，具体结构如表 1 所示。

表 1　本文算法的主干特征提取网络结构

类型	步长	重复次数	输出尺寸
输入	—	—	416×416×3
卷积 3×3	2	1	208×208×24
最大池化 3×3	2	1	104×104×24
模块 2	2	1	52×52×116
模块 1	1	3	52×52×116
模块 2	2	1	26×26×232
模块 1	1	7	26×26×232
模块 2	2	1	13×13×464
模块 1	1	3	13×13×464
卷积 1×1	1	1	13×13×1 024

1.2　SKNet 注意力机制

人类视觉皮层根据不同大小目标刺激程度的不同会动态调整神经元的感受野，受此启发 SKNet 提出一种可自适应选择卷积核尺寸(selective kernel，SK)的注意力机制，它融合了分组卷积、空洞卷积和 SENet (squeeze-and-excitation networks)[16]通道注意力机制的思想，实现了对不同大小的卷积核添加注意力机制。SKNet 的核心结构为 SK 卷积模块，其结构如图 2 所示。

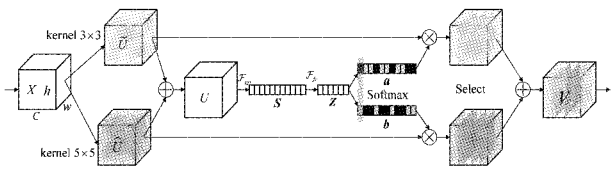

图 2　SK 卷积结构

SK 卷积主要由 Split、Fuse 和 Select 三个步骤组成。Split 将大小为 $h\times w\times C$ 的输入特征图 X 分成两个分支，分别使用 3×3 卷积和 5×5 卷积进行卷积操作得到 $\tilde{U}$ 和 $\hat{U}$，为了增强特征提取能力，实际用 3×3 空洞卷积代替 5×5 卷积。Fuse 操作与 SENet 的方法类似，先将 $\tilde{U}$ 和 $\hat{U}$ 通过元素相加进行融合得到 U，然后通过全局平均池化 F_{gp} 得到 $1\times1\times C$ 的特征向量 $\boldsymbol{S}$，再通过全连接 F_{fc} 降维得到 $1\times1\times d$ 的特征向量 $\boldsymbol{Z}$，最后经由两个 Softmax 函数输出得到包含不同分支通道权重信息的矩阵 $\boldsymbol{a}$ 和 $\boldsymbol{b}$，大小为 $1\times1\times C$。Select 将权重矩阵 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 分别与原始特征图 $\tilde{U}$ 和 $\hat{U}$ 进行加权操作后再相加最终得到输出特征图 V。

SK 卷积不再是局限于通道层面和空间层面的注意力机制，而是给不同尺寸的卷积核实施注意力机制，从而让网络自适应地调整自身结构。同时 SK 卷积也是一个轻量化

的即插即用模块,既不会给网络增加太多的计算量,又能带来精度上的提升。

1.3 边界框回归损失函数

原 YOLOv3 的边界框回归损失使用的是交并比损失函数(intersection over union,IoU)。IoU 表示预测框与真实框的重合程度,如式(1)所示。其中 P 和 G 分别表示预测框和真实框的面积,$IoU(P,G)$ 表示两框的交集与并集之比,可以看出 IoU 越大则模型预测的目标位置与真实位置越接近。

$$IoU(P,G)=\frac{P\cap G}{P\cup G} \tag{1}$$

但是 IoU 仍然存在一些缺点,当预测框和真实框不相交时,根据式(1)可知 IoU 为 0,此时边框回归损失函数为零,网络反向传播时梯度无法更新。针对此问题,本文提出使用 CIoU 损失函数代替原 IoU 函数,如式(2)所示。

$$L_{CIoU(P,G)}=1-IoU(P,G)+\frac{\rho^2(p,g)}{c^2}+\alpha\upsilon \tag{2}$$

式中:p 和 g 分别表示预测框和真实框的坐标中心,$\rho(p,g)$ 表示两个中心点间的欧氏距离,c 表示能同时包含预测框和真实框的最小矩形的对角线长度,$\alpha\upsilon$ 表示惩罚因子;其中 α 是平衡比例的参数;υ 是衡量预测框与真实框长宽比一致性的参数,计算公式分别如式(3)和(4)所示,w^{gt}、h^{gt} 表示预测框的宽度和高度,w、h 表示真实框的宽度和高度。

$$\alpha=\frac{\upsilon}{1-IoU(P,G)+\upsilon} \tag{3}$$

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

由于 CIoU 同时考虑了预测框和真实框的重合程度、中心距离和长宽比,因此模型训练的收敛速度更快,边框回归更精确,检测效果更好。

1.4 改进的 YOLOv3 算法结构

本文提出的改进 YOLOv3 算法的整体网络结构如图 3 所示。输入大小为 416×416×3 的图像,经过改进的主干网络提取特征后,分别输出大小为 52×52×116、26×26×232、13×13×1024 的特征图送入特征融合网络。特征融合网络是一种自下而上的金字塔结构,首先对 3 层不同尺度的特征图进行 SK 卷积,然后将 13×13×1 024 的特征图经过 5 次 DBL 卷积后,再通过上采样和 SK 卷积与上层特征图拼接融合,以此类推最终输出包含丰富语义信息的 3 层特征图,实现了对不同尺度的特征图进行信息融合。最后将这 3 层特征图分别进行 1 次 DBL 卷积和标准卷积后进行预测,通过含有 CIoU 的损失函数进行模型训练,完成网络反向传播和参数迭代更新。

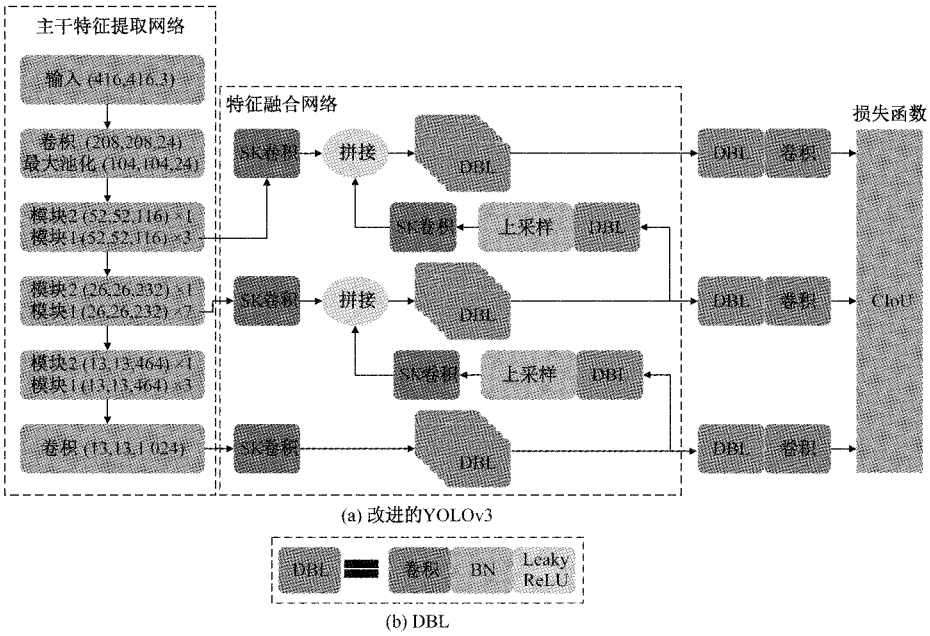

图 3　改进 YOLOv3 的算法结构

2 实　　验

2.1 数据集

本文从公开数据集 WIDER Face 和 MAFA(Masked Faces)中选取人脸图片共 7 959 张,其中包含戴口罩人脸图片 4 065 张和未佩戴口罩人脸图片 3 894 张,标记两种类别 face_mask(戴口罩人脸)和 face(不戴口罩人脸),划分训练验证集共 6 120 张,测试集 1 839 张。为了避免模型训练过拟合,在训练过程中采用随机翻转、随机裁剪和随机颜色变换等数据增强方法提高数据复杂度,增强模型的泛化能力。

2.2 实验环境

本文模型训练环境:Ubuntu16.04 64 位操作系统;

GPU 为 NVIDIA RTX2080Ti，显存 11 GB；CPU 为 Intel(R)Xeon(R)E5-2640 v4，板载内存 64 GB。模型测试环境：Windows10 64 位操作系统，CPU 为 Intel Core i7-9750H，板载内存 16 GB，GPU 为 NVIDIA RTX2060，显存 6 GB。基于 TensorFlow 深度学习框架，使用 Python3.6 编程。

2.3　模型训练及评价指标

总共训练 200 个轮次，batchsize 设为 8，选用常用的 Adam 优化器，学习率为 0.001，使用余弦学习率衰减法。模型训练损失曲线如图 4 所示。由图 4 可知，模型在训练 50 个轮次后损失缓慢下降并逐渐收敛。

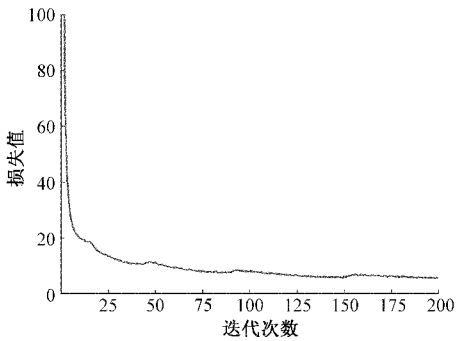

图 4　训练损失变化曲线

本文主要采用平均精度(average precision，AP)、平均精度均值(mean average precision，mAP)和每秒检测帧数(frame per second，FPS)作为评价指标来衡量算法性能。其中 AP 的计算涉及到精确率(Precision)和召回率(Recall)的概念，如式(5)、(6)所示。

$$Precision = \frac{tp}{tp + fp} \times 100\% \tag{5}$$

$$Recall = \frac{tp}{tp + fn} \times 100\% \tag{6}$$

式中：tp 表示模型预测为正类实际为正类的样本数，fp 表示模型预测为正类实际为负类的样本数，fn 表示模型预测为负类实际为正类的样本数。AP 和 mAP 的计算公式如式(7)、(8)所示。

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{7}$$

$$mAP = \frac{\sum_{i=1}^{C} AP_i}{C} \tag{8}$$

以不同召回率及其对应的最高精确率分别作为横、纵坐标，将得到一条 P-R 曲线，由式(7)可知，对 P-R 曲线求积分即为平均精度值 AP。C 为目标类别数，因此由式(8)可知 mAP 为所有类别的 AP 均值。

2.4　实验结果分析

为了进一步验证本文所提改进点的有效性，在测试集上进行消融实验，实验结果如表 2 所示。表中方法 1 表示使用 ShuffleNetv2 作为主干网络的 YOLOv3；方法 2 表示在方法 1 的基础上添加 SK 卷积模块；方法 3 即为本文算法，在方法 2 的基础上使用 CIoU 作为边框回归损失函数。

表 2　消融实验结果

方法	ShuffleNetv2	SK 卷积	CIoU	AP/%		mAP/%	FPS/(帧/s)
				face	face_mask		
YOLOv3	×	×	×	94.91	93.87	94.39	25.62
方法 1	√	×	×	90.95	92.12	91.54	67.66
方法 2	√	√	×	93.58	92.15	92.87	58.95
方法 3(本文算法)	√	√	√	93.66	93.10	93.38	59.09

通过表 2 可以看出，方法 1 的 mAP 值相比于原 YOLOv3 降低了 2.85%，说明引入轻量级主干网络 ShuffleNetv2 使得网络特征提取能力稍显减弱，但同时模型检测速度大幅提升，达到了 67.66 帧/s，满足检测实时性要求。对比方法 1 和 2，方法 2 的 mAP 值提高了 1.33%，说明注意力机制 SK 卷积模块有效地抑制了无用特征信息，实现了特征增强，进一步提高了特征融合网络的特征提取能力，同时带来了小部分计算量的增加，因此检测速度略微降低。方法 3 在不增加模型计算量的前提下，mAP 值相比于方法 2 提高了 0.51%，验证了 CIoU 的有效性。与原 YOLOv3 算法相比，本文算法的检测精度虽降低了 1.01%，但依然保持在高水平，且检测速度达到了 59.09 帧/s，是原算法的 2.3 倍，远远满足检测实时性。

图 5 为本文算法的部分测试结果，由图 5 可以看出，本文算法在不同的复杂环境下都能有效地检测出佩戴口罩人脸和未佩戴口罩人脸。对于存在多目标、小目标、目标被手遮挡和目标远近不同的情况下，本文算法均没有出现漏检和误检现象。

表 3 为本文算法和原 YOLOv3 算法的模型复杂度性能对比。由表 3 可以看出，本文算法的模型体积缩减了 63.1%，网络参数量仅为 22.4 M(Million)，浮点运算量为 17.112 BFLOPs(Billion floating point operations)，相比原算法降低了 18.2 BFLOPs。参数量和计算量的大幅减少也更有利于模型部署在算力低、功耗小的硬件平台，保证模型的检测实时性。

为进一步验证本文算法的有效性，在相同的实验环境下，将本文算法与原 YOLOv3、EfficientDet[17]、YOLOv4[18] 和 RetinaNet[19] 等主流算法进行对比，结果如表 4 所示。

图 5 本文算法检测效果

表 3 本文算法与 YOLOv3 的模型复杂度

方法	模型体积/MB	参数量/M	浮点运算量/BFLOPs
YOLOv3	235	61.6	65.312
本文算法	86.6	22.4	17.112

表 4 本文算法和其他主流算法的性能对比

方法	AP/%		mAP/%	FPS/(帧/s)
	face	face_mask		
YOLOv3	94.91	93.87	94.39	25.62
EfficientDet-D0	92.45	93.35	92.90	34.63
YOLOv4	94.66	94.05	94.36	20.76
RetinaNet	93.76	93.67	93.71	15.59
本文算法	93.66	93.10	93.38	59.09

由表 4 可以看出,本文改进的 YOLOv3 算法与其他主流算法相比,不仅在检测速度上占有显著优势,在检测精度上同样保持较高的水平,更好地平衡了精度与速度。因此本文算法的综合性能最佳,更具鲁棒性。

3 结 论

本文针对口罩佩戴检测任务提出一种轻量化的改进 YOLOv3 算法。为了减少网络复杂度,提高模型推理速度,提出用 ShuffleNetv2 作为主干特征提取网络。在保证模型轻量化的前提下,将注意力机制 SKNet 中的 SK 卷积模块融入到特征融合部分,增强特征复用性,提高了检测精度。采用 CIoU 作为边界框回归损失函数,充分考虑预测框和真实框的重合程度、中心距离和长宽比,进一步提高了模型检测性能。实验结果表明,本文算法减少了模型容量、参数量和计算量,牺牲了 1.01%检测精度,实现了检测速度的成倍提升,满足检测实时性。与其他主流算法相比,本文算法在精度与速度的平衡性上表现更优,可以更好地应用在口罩佩戴检测领域。下一步将考虑如何在保证网络模型轻量化的前提下,进一步提高检测准确性。

参考文献

[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.

[2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.

[3] 张培培,王昭,王菲.基于深度学习的图像目标检测算法研究[J].国外电子测量技术,2020,39(8):34-39.

[4] GIRSHICK R. Fast r-cnn [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.

[5] REN S, HE K, 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, 2016, 39(6): 1137-1149.

[6] LIU W, ANGUELOV D, ERHAN D, et al. SSD:

Single shot multibox detector [C]. European Conference on Computer Vision, Springer, Cham, 2016: 21-37.

[7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[8] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. ArXiv Preprint, 2018, ArXiv: 1804. 02767.

[9] 方仁渊,王敏. 基于改进型 YOLO 网络的商品包装类型检测[J]. 电子测量技术,2020,43(7):108-112.

[10] 贾振卿,刘雪峰. 基于 YOLO 和图像增强的海洋动物目标检测[J]. 电子测量技术,2020,43(14):84-88.

[11] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.

[12] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 116-131.

[13] LI X, WANG W, HU X, et al. Selective kernel networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 510-519.

[14] ZHENG Z, WANG P, REN D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. ArXiv Preprint, 2020, ArXiv: 2005. 03572.

[15] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6848-6856.

[16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.

[17] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.

[18] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal Speed and Accuracy of Object Detection[J]. ArXiv Preprint, 2020, ArXiv: 2004. 10934.

[19] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.

作者简介

薄景文,硕士研究生,主要研究方向为深度学习、目标检测、检测技术与智能装置。

E-mail: bjw287916174@163.com

张春堂(通信作者),副教授,主要研究方向为深度学习、图像处理、模式识别、检测技术与智能装置。

E-mail: zct1999@163.com