

# 基于 ResNet 的安全监控目标检测

孙毅, 吴斯曼, 方伟, 吴双卿\*, 胡超

(浙大宁波理工学院, 信息科学与工程分院 宁波 315100)

**摘要:** 随着电动自行车骑行人员必须佩戴头盔的法规出台, 相关智能视觉检测技术的需求也应运而生, 本文以交通监控视频图像中骑行人员佩戴头盔情况为研究对象, 以 YOLO 目标检测框架为基础, 首先采用了分支吸收模块改善残差骨干网络, 然后通过结构通道重组提升卷积层特征融合, 最后应用设计的结构融合剪枝进一步压缩模型超参数。实验结果表明, 该算法具有较优的精度和实时性, 小目标检测效果也较好, 其多分类平均精度达到 88.8%, 检测速度可达 29.5fps, 基本满足交通视频监控的需求。

**关键词:** 目标检测; 特征融合; 残差骨干网络; 通道重组; 结构融合剪枝

中图分类号: TP 249

文献标志码: A

doi: 10.12146/j.issn.2095-3135.20231108001

## Object Detection of Security Monitoring Based on ResNet

Sun Yi, Wu Siman, Fang Wei, Wu Shuangqing\*, HU Chao

(School of Information Science and Engineering, NingboTech University, Ningbo 315100, China)

**Abstract:** With the regulations of wearing helmets while driving the electric bicycle, it is urgent to develop a detection algorithm that can accurately detect whether the drivers are wearing helmets. This paper introduces a novel method to detect the helmets based on the YOLO framework. The branch absorption module is proposed to improve the residual backbone network, then the feature fusion is improved through the channel recombination. Finally, the designed structural fusion pruning is applied to further compress the hyper-parameters of the model. The experimental results showed that, the proposed algorithm has higher accuracy and faster speed. Performance of small targets detection also can be improved, with the average accuracy of multiple classification up to 88.8% and detection speed of 29.5fps, which can meet the demand of video surveillance in real applications.

**Keywords:** Object detection; Feature fusion; Residual backbone network; Channel recombination; Structure fusion pruning

### 1 引言

头盔的佩戴能起到关键的安全保护作用, 很多城市出台了电动自行车骑行人员必须佩戴头盔的相关法规, 并安排交警在城市道路旁检测电动自行车骑行人员头盔佩戴情况。但这种人工检测的方式效率较低、成本较高、覆盖范围小, 难以达到理想的效果。基于交通监控视频图像的电动自行车安全头盔自动检测方法, 需要考虑以下问题: (1) 实际场景环境复杂, 目标易受到环境光线和天气状况的影响; (2) 检测目标尺度变化大, 目标大小随运动而变化, 会存在小目标与大目标同时出现等情况; (3) 目标移动速度较快, 检测系统不仅要满足精度, 也要满足应用于视频监控的实时性需求。

对于头盔检测的研究, 目前主要集中在施工场所中工作人员安全帽的佩戴情况, 电动自行车(或摩托车)驾驶员的头盔佩戴情况的监测也逐步得到了关注。Saumya 等人<sup>[1]</sup>首先使用 YOLOv3 算法, 把电动自行车和人作为一个整体进行检测, 计算电动自行车和人之间的边界框的重叠面积, 来定位电动自行车上的人的位置。王江涛等人<sup>[2]</sup>对主流的一步检测网络 SSD-Net 引入类似视觉机制的模块, 对网络的特征图在通道和空间上进行了权重的重新选择, 并采用余弦衰减学习率来优化网络。Wu 等<sup>[3]</sup>利用 Densenet 模型参数, 替

收稿日期: 2023-11-08

修回日期: 2023-12-27

基金项目: 浙江省自然科学基金(LQ17F030002); 浙江省大学生科技创新活动计划(2022R438A007)

代 YOLOv3 的骨干网络进行特征提取，从而形成一种基于 Densenet 的深度学习。周等<sup>[4]</sup>则寻求多种权重的检测模型应用在工地场景的安全帽检测。深度学习在头盔佩戴的自动检测方面已取得一些进展，但对于电动自行车头盔佩戴检测的应用研究，还有待公开数据集的建立和完善，检测算法的泛化能力、检测精度和效率等方面还可以进一步开展工作。

本文从交通安全的需求出发，进行驾驶员头盔佩戴情况的自动检测，建立了相应的数据集，以 YOLO 算法框架构建检测模型，采用了分支吸收模块改善其残差骨干网络，通过结构通道重组提升卷积层特征融合，最后应用设计的结构融合剪枝进一步压缩模型超参数。该算法具有较优的精度和实时性，小目标检测效果也较好，能够基本满足视频监控需求。

## 2 ResNet 目标检测

### 2.1 残差网络

ResNet (Residual Network)<sup>[5]</sup>的关键特点是引入了残差连接 (Residual Connection) 或者跳跃连接 (Skip Connection)，ResNet 的设计者们采用了残差学习的策略，利用神经网络不容易拟合一个恒等映射的特点，通过在学习过程中加入原始的特征向量，从而优化了特征映射的学习机制。残差的定义为  $\mathcal{H}(x) = \mathcal{F}(x) + x$ 。其中  $x$  代表输入图像，而  $\mathcal{F}(x)$  则是网络需要学习的目标映射。这样即使残差为 0，网络也只会学习检测映射，残差网络由多层跳跃连接的神经元子模块组成：

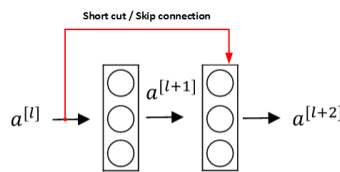


图 1 残差结构

Fig.1 Residual structure

其中 Short cut/Skip connection(捷径分支)直接建立起和的隔层关系，与直连参数进行高低阶级的特征混合，其向前传播的计算步骤为：

$$\begin{cases} z^{[l+1]} = W^{[l+1]}a^{[l]} + b^{[l+1]} \\ a^{[l+1]} = g(z^{[l+1]}) \\ z^{[l+2]} = W^{[l+2]}a^{[l+1]} + b^{[l+2]} \\ a^{[l+2]} = g(z^{[l+2]} + a^{[l]}) \end{cases} \quad (1)$$

### 2.2 剪枝

神经网络 (DNN, Deep Neural Network)<sup>[6]</sup>通常需要耗费大量的算力资源。剪枝作为一种模型压缩的方法，它基于一个假设，即 DNN 的过参数化 (Over-parameterization)<sup>[7]</sup>，而过参数化是指训练阶段我们需要大量的参数来捕捉数据中的微小信息。但有一些浮点型的神经元连接结构权重较小，对整个模型的影响较小，因此可以采取删减一些节点，来获得更小的模型和更快的推理速度。一般说来剪枝会带来精度上的下降，采用有效的裁剪模型并维持精度损失最小化是论文研究的重点。

## 3 本文方法

对于电动自行车安全头盔识别任务，本文算法的构建过程以骨干网络为基础，设计残差网络的优化框架，在特征融合网络层中构建通道重组，优化模型的特征交互，采用结构融合剪枝方法，进一步减少模型参数，提升推理速度。

### 3.1 分支吸收骨干网络

在 Darknet53 骨干网络<sup>[8]</sup>中，使用全卷积网络和残差结构的方式，其降维连接方式借鉴了 GoogLeNet<sup>[9]</sup>相关思想，通过  $1 \times 1$  与  $3 \times 3$  的小型卷积核堆叠，有效降低了深层网络训练的复杂性。然而，该结构在每次降维采样过程中损失了过多信息，由于维度的大幅减少。为了解决这一问题，本研究引入了分支吸收的策略，改进网络结构，采用多模态耦合的降维方式。具体来说，在步长  $s$  确定的情况下，网络被分为两个分支：在  $s=2$  的场景下，过渡模块由三层卷积块、两层卷积块以及池化层构成，经过卷积与池化操作后输出尺寸减半的融合特征图；在  $s=1$  的场景下，特征通道被分为两个分支，各自含有  $c-c'$  和  $c'$  数量的通道，通过带有跳跃连接的三层卷积块残差模块进行处理，最终将两个分支通过拼接操作合并，以保持通道总数的不变。如图 2 所示，(a) Darknet53 骨干网络中的特征提取模块；(b) 为本文步长为 1 时构建的通道分割残差结构；(c) 为本文构建的步长为 2 时的三层特征提取结构。

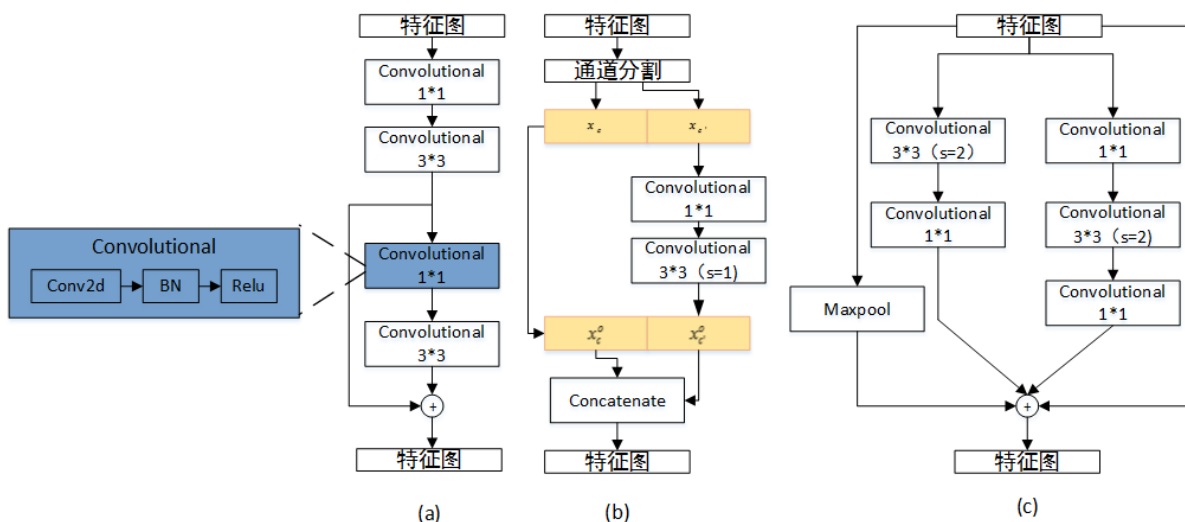


图 2 骨干网络中的特征提取模块

Fig.2 Feature Extraction Module in the Backbone Network

### 3.2 通道重组

本文使用多模式耦合策略，进行多特征通道划分构建分支吸收骨干网络，运用通道重组相关思想，具体应用为每层卷积后部署相关模块。算法步骤如下：

输入特征图  $A$

输出重组图  $P$

Step1. 将特征图展开为  $g * n * w * h$  四维度矩阵；步骤一为特征图  $A$  的展开操作，其中特征图被重构为四维矩阵形式。

Step2. 对该四维矩阵执行维度置换操作，具体是沿着给定尺寸的矩阵的第  $g$  轴与第  $n$  轴进行转置

Step3. 进行转置矩阵的扁平化处理，形成折叠成一维的数组。

Step4. 对得到的特征图实施组内的  $1 \times 1$  卷积运算，此过程在特征学习中促进了信息的有效整合，并有助于进一步的特征抽象化。

### 3.3 结构融合剪枝

文献<sup>[10]</sup>何易辉等考虑在权重修剪中加入正则性，即过滤器修剪，通道修剪<sup>[11]</sup>等，从而生成规则的和更小的权重矩阵，以便在 CPU / GPU 上更快地执行。如图 3 所示，过滤器剪枝对应的是去掉权重矩阵中的一行，而通道剪枝对应的是减少多个连续列。结构剪接方法在压缩率较高时，由于剪接滤波器信道的信息丢失，会造成明显的精度损失。

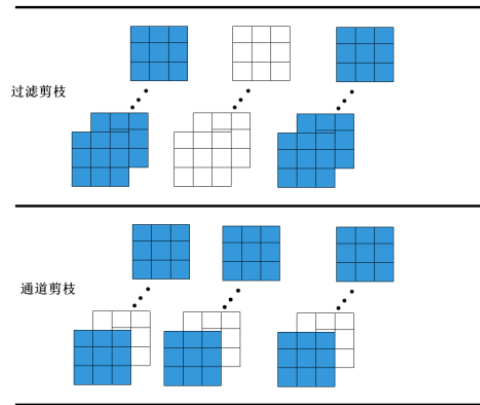


图 3 剪枝方法示意

Fig.3 Pruning Method Illustration

本文所采用的结构融合剪枝，引入新的剪枝视角，如图所示，每一层的权值矩阵是由多个大小相同的权值块  $m \times n$  组成。对每个块应用独立的行和列修剪，在每个块中可能有不同的修剪率（被修剪的行/列的数量），以确保高度的灵活性，每个块中剩余的权值仍然构成一个完整且更小的矩阵。对  $N$  层的深度神经网络设  $w$  表示第  $i$  层的权重，则  $w = \{w_i\}_{i=1}^N$ ，有  $w_{ij} \in \mathbb{Z}^{m \times n}$ ，每一个  $w_i$  会被均匀分成  $m \times n$  大小的  $k$  数量的矩阵，则  $w_i = [w_{i1}, w_{i2}, \dots, w_{ik}]$ ， $[w_{ij}]_p$  和  $[w_{ij}]_q$  分别表示  $w_{ij}$  的第  $p$  行和第  $q$  列。

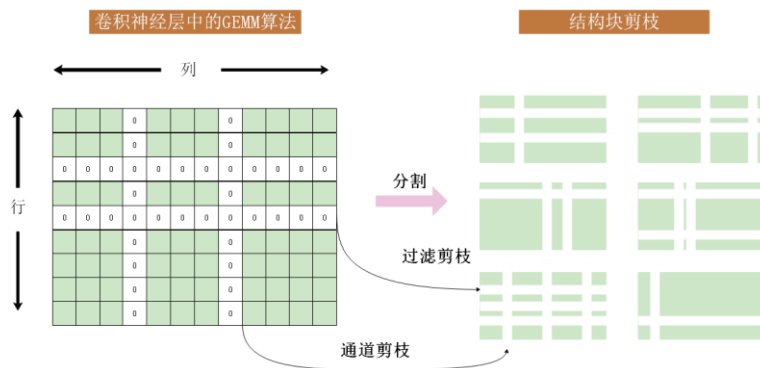


图 4 结构化融合剪枝过程

Fig.4 Structured Fusion Pruning Process

使用重加权方法来解决组 lasso 正则化，从而消除了对重要和不重要的权重应用相同惩罚的缺点，对

于块结构融合剪枝中的列剪枝模块，计算方式为：

$$f(w_{loss}) + \lambda \sum_{i=1}^N \sum_{j=1}^K \left( \omega_i^t \odot \left\| [w_{ij}]_p \right\|_{Fro} \right) \quad (2)$$

其中  $f(w_{loss})$  表示训练损失值， $\lambda$  是调整精度和稀疏度的相对重要性的惩罚参数， $\odot$  为 element-wise multiplication,  $\| \cdot \|_{Fro}$  表示 Frobenius norm 范数的计算， $\omega_i^t$  每一次迭代  $t$  更新一次罚权值集合，有助于增加组 lasso 正则化之外的稀疏度。在每一次迭代时， $w_i$  转化为  $w_i^t$ ，更新  $\omega_i^t$ ，通过设置：

$$\omega_i^{t+1} = \frac{1}{\left\| [w_{ij}]_p^t \right\|_{Fro}^2 + \Delta\beta} \quad (3)$$

$\Delta\beta$  为防止分母除零的小值参数。

结构融合行剪枝模块有：

$$f(w_{loss}) + \lambda \sum_{i=1}^N \sum_{j=1}^K \left( \omega_i^t \odot \left\| [w_{ij}]_q \right\|_{Fro} \right) \quad (4)$$

可以更新  $\omega_i^t$ ，通过：

$$\omega_i^{t+1} = \frac{1}{\left\| [w_{ij}]_q^t \right\|_{Fro}^2 + \Delta\beta} \quad (5)$$

本文采用的结构融合剪枝，首先使用预先训练的模型初始化，并为修剪后的模型预先定义块大小，在 DNN 训练中加入了重加权组 lasso 正则化，迭代更新惩罚参数，重新加权正则化项(s)，即在优化问题中有界，在重新加权步骤后，去除接近于零的权值(或一组权值)，并使用非零权值来细化 DNN。

算法：结构化融合剪枝的重加权正则化	
1	<b>初始化：</b> 预先训练的DNN模型初始化 $w$ ，设置迭代次数为T;定义块的大小为 $m \times n$
2	每层K: $w_i / m \times n$
3	<b>while</b> $t \leq T$ <b>do</b>
4	对 (1) 或者 (2) 使用梯度更新法则；
5	使用 $w_i^t$ 的处理结果更新 $\omega_i^{t+1}$ ；
6	<b>end</b>
7	移除接近于零的权重组，并重新训练其余的非零权重，以提高准确性
8	<b>结果：</b> 融合剪枝的DNN模型

图5 结构融合剪枝实现方法

Fig.5 Implementation Method of Structured Fusion Pruning

### 3.4 评价指标

为分类任务选择正确的衡量指标也是重要的一环，在电动自行车安全头盔佩戴检测的场景中，是一个

不平衡的分类问题，需要鉴别的类别有两个：佩戴安全头盔的人和没有佩戴安全头盔的人，其中前者代表了大类的数据点。在这种情况下，准确率不是一种良好的评价指标，同时应该考虑召回率，它是精确率 (Precision) 和召回率 (Recall) 的调和平均数，调和平均可以惩罚极端情况。其中，F1-score 的值范围为 0 到 1，1 表示最好的性能，0 表示最差的性能，这个指标也能衡量假阳性和假阴性的问题：

$$F1 = 2 * \frac{precision * recall}{precision + recall} \quad (6)$$

## 4 算法测试

### 4.1 数据集建立

考虑到实际场景的需求，数据集样本采用多点位人工收集方式（主要位于宁波市北仑区恒山路与四明山路区域），由拍摄视频、截取视频帧的方式进行处理，样本中包含不同的环境光照条件、不同颜色种类的头盔。并采用网络数据爬虫，采取数据融合、视频帧截取、数据增广等常用数据清洗手段补充部分数据集。合计训练集 3108 张，验证集 1206 张。

基于原始数据之上，我们设计 MATLAB GUI 对图像任务流处理。通过引入多样性的噪声、以及对图像的亮度和颜色等进行调整，提高模型对不同光照条件下的图像数据的适应性，再采用灰度调整，并结合姿态变化等要素，使模型有效地学习目标在不同角度和方向上的特征。



图 6 使用 MATLAB GUI 进行数据归一化

Fig.6 Data normalisation using the MATLAB GUI

### 4.2 结果与分析

训练模型以及检测实验，使用本文优化过残差结构的 Darknet53 骨干网络<sup>[8]</sup>，使用 50 个副本，每个副本在 GPU 上运行，批次大小为 8，为期 300 个 epoch。使用衰减为 0.9 的动量，decay 为 0.9，学习率为 0.045，每两个纪元衰减一次，指数衰减率为 0.94。

将本文的方法 (Result B) 分别与 Yolov3 (Result A)<sup>[8]</sup>和 MobileNetv2 (Result C)<sup>[11]</sup>进行对比，评价指



标结果如图 6 所示。本文算法的 mAP 为 91.78%，F1-score 为 0.91，而 Yolov3 的 mAP 为 88.52%，F1-score 为 0.87。MobileNet 是可分离卷积模块的堆叠，其可分离卷积模块包含深度卷积和点卷积，与之对应的是较小的模型参数与极快的训练速度，在 50epoch 就能达到较好的训练效果，但漏检了部分目标，体现在查全率和 F1-score 等参数。

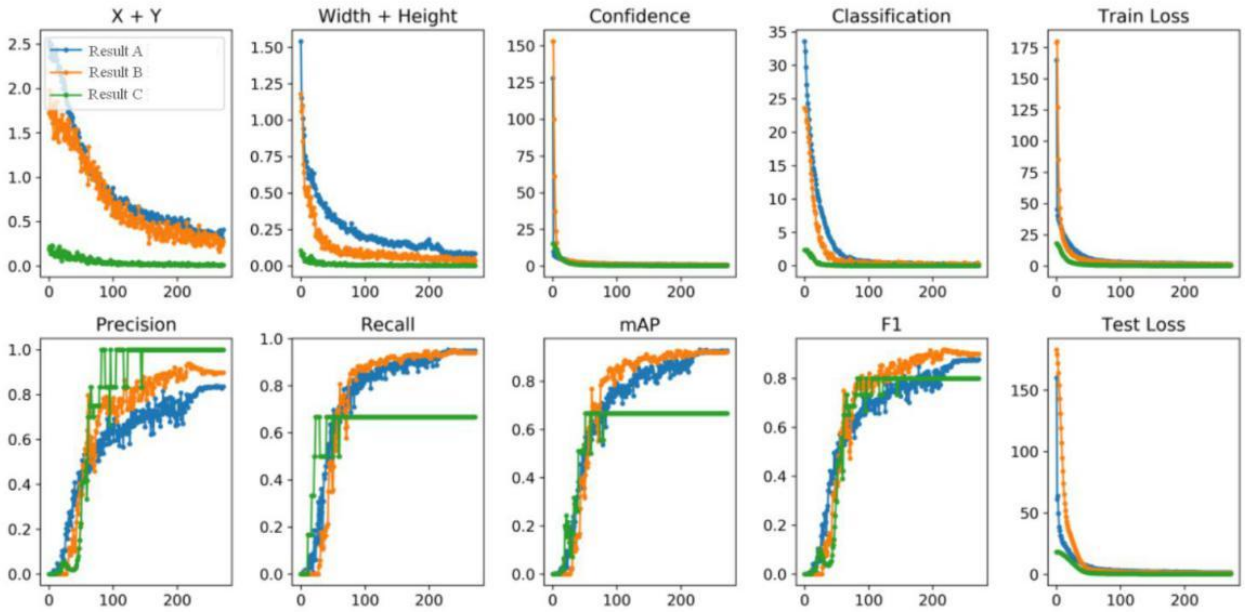


图 7 不同检测算法的评价指标结果比较

Fig.7 Comparison of Evaluation Metric Results for Different Detection Algorithms



图 8 本文算法与 YOLOV3 实况检验对比

如图 8 所示, 在目标重叠、特征类似、或者特征尺度极小的复杂情况下, 本文采用分支吸收骨干网络与通道重组的方法, 有更精准的回归锚框和更好的小目标检测效果, 在部分检测中能减少假正例的产生。

本文采用的结构化剪枝比过滤器剪枝和通道剪枝, 有更优的剪枝效果, 体现在较小的精度损失和更大的实时性收益。取 200 个 epoch 为例, 使用本文结构融合方法优化后的模型, 能够以 3% 的精度损失换取 12.5% 的模型检测加速效果。

表 1 测试结果与算法对比率

算法	<i>mAP</i>	<i>F1</i>	<i>FPS</i>
YOLOV3	88.52%	0.87	24.8
Mobilenet	69.87%	0.71	42.5
本文方法 (引入剪枝压缩前)	91.78%	0.91	23.6
本文方法	88.8%	0.9	29.5

## 5 总结与展望

本文围绕安全出行的应用需求, 进行驾驶员头盔佩戴情况的自动检测, 建立了相应的数据集, 以 YOLO 算法框架构建检测模型, 采用了分支吸收模块改善其残差骨干网络, 通过结构通道重组提升卷积层特征融合, 最后应用设计的结构融合剪枝进一步压缩模型超参数。实验结果表明, 该算法具有较优的精度和实时性, 小目标检测效果也较好, 其多分类平均精度达到 88.8%, 检测模型 FPS 为 29.5, 能够基本满足视频监控需求。接下来将跟踪研究最新的深度学习算法进行检测, 研究不同的主干网络结构, 优化损失函数算法, 并训练更多样化的模型进行验证和测试, 并通过模型蒸馏技术, 进一步消减模型冗余, 以适配于资源受限的嵌入式系统环境。

### 参考文献

- [1] Saumya A, Gayathri V, Venkateswaran K, et al. Machine Learning based Surveillance System for Detection of Bike Riders without Helmet and Triple Rides[C]. 2020 International Conference on Smart Electronics and Communication (ICOSEC). IEEE, 2020: 347-352.
- [2] 刘琛, 王江涛, 王明阳. 引入视觉机制的 SSD 网络在摩托车头盔佩戴检测中的应用[J]. 电子测量与仪器学报, 2021, v.35; No.243(03): 144-151.
- [3] Wu F, Jin G, Gao M, et al. Helmet detection based on improved YOLO V3 deep model[C]//2019 IEEE 16th International conference on networking, sensing and control (ICNSC). IEEE, 2019: 363-368.
- [4] Zhou F, Zhao H, Nie Z. Safety Helmet Detection Based on YOLOv5[C]//2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA). IEEE, 2021. DOI:10.1109/ICPECA51329.2021.9362711.
- [5] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [6] Sze V, Chen Y H, Yang T J, et al. Efficient processing of deep neural networks: A tutorial and survey[J]. Proceedings of the IEEE, 2017, 105(12): 2295-2329.
- [7] Allen-Zhu Z, Li Y, Song Z. A convergence theory for deep learning via over-parameterization[C]//International conference on machine learning. PMLR, 2019: 242-252.
- [8] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
- [9] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [10] He YH, Zhang XY, Sun J. Channel pruning for accelerating very deep neural networks[C]//ICCV. 2017: 1398-1406
- [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.



**作者简介:** 孙毅, 硕士, 研究方向为深度学习与人工智能; 吴斯曼, 本科, 研究方向为图像与信息处理; 方伟, 讲师, 研究方向为传感器与智能检测; 吴双卿(通讯作者), 副教授, 研究方向为机器视觉、光电检测与信息处理, E-mail: wsqing1999@163.com; 胡超, 教授, 研究方向为自动化与传感器技术、图像处理和机器视觉。

收稿日期: 2023-11-08 修回日期: 2023-12-27

基金项目: 浙江省自然科学基金(LQ17F030002); 浙江省大学生科技创新活动计划(2022R438A007)

[键入文字]