第 13 卷 第 6 期
2024 年 11 月
集 成 技 术
JOURNAL OF INTEGRATION TECHNOLOGY
Vol. 13 No. 6
Nov. 2024

引文格式：
孙毅, 吴斯曼, 方伟, 等. 基于 ResNet 的安全监控目标检测 [J]. 集成技术, 2024, 13(6): 44-52.
Sun Y, Wu SM, Fang W, et al. Object detection of security monitoring based on ResNet [J]. Journal of Integration Technology, 2024, 13(6): 44-52.

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学院 宁波 315100）

摘 要 《中华人民共和国道路交通安全法》要求摩托车驾驶人及乘坐人员应按规定戴安全头盔，因此，头盔佩戴智能视觉检测技术的需求应运而生。本文算法模型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，该算法的精度和实时性较优，小目标检测效果也较好，多分类平均精度为 88.8%，检测速度可达 29.5 帧/s，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 TP249 **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*)

**Corresponding Author*: *wsqing*1999@163.*com*

Abstract The “Road Traffic Safety Law of the People’s Republic of China” requires that motorcycle riders and passengers must wear safety helmets as stipulated by law. Consequently, the demand for intelligent visual detection technology for helmet wearing has emerged. This paper focuses on the study of helmet wearing by riders in traffic surveillance video images, based on the YOLO object detection framework. Initially, a branch absorption module is employed to improve the residual backbone network. Subsequently, the convolutional layer feature fusion is enhanced through structural channel recombination. Finally, a designed structural fusion pruning technique is applied to further compress the model’s hyperparameters. Experimental results indicate that the algorithm boasts superior accuracy and real-time performance, with effective detection of small targets.

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项目（LQ17F030002）；浙江省大学生科技创新活动计划项目（2022R438a007）
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，学士，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。

The average precision for multi-classification reaches 88.8%, and the detection speed can achieve up to 29.5 frames per second, which essentially meets the requirements of traffic video surveillance.

Keywords　object detection; feature fusion; residual backbone network; channel recombination; structure fusion pruning

Funding　This work is supported by Natural Science Foundation of Zhejiang Province (LQ17F030002), and Zhejiang Student's Platform for Innovation and Entrepreneurship Training Program (2022R438a007)

1　引　言

佩戴头盔能起到关键的安全保护作用，《中华人民共和国道路交通安全法》第五十一条中规定摩托车驾驶人及乘坐人员应当按规定戴安全头盔，交通部门在部分城市道路中安排了交警，以监督电动自行车骑行人员的头盔佩戴情况。但这种人工检测的效率较低、成本较高、覆盖范围较小，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需要考虑以下问题：(1)实际场景环境复杂，目标易受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，会存在小目标和大目标同时出现的情况；(3)目标移动速度较快，检测系统不仅要满足精度，还要满足应用于视频监控的实时性需求。

Saumya 等[1]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，通过计算电动自行车和人之间的边界框的重叠面积，定位电动自行车上人的位置。刘琛等[2]对主流的一步检测网络 SSD-Net，引入了一种类似视觉机制的模块，该模块在网络的特征图上，分别在通道和空间维度进行了权重的重新分配，并利用余弦衰减学习率优化网络。Wu 等[3]利用 DenseNet 模型参数替代 YOLOV3 的骨干网络进行特征提取，从而形成一种基于 DenseNet 的深度神经网络。Zhou 等[4]研究了在工地场景中应用多样化权重检测模型进行安全帽检测的方法。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需要进一步研究。

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，建立相应的数据集，以 YOLO 算法框架构建检测模型，采用分支吸收模块改善 YOLO 算法的残差骨干网络，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的需求。

2　残差网络目标检测

2.1　残差网络

残差网络(residual network，ResNet)[5]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，优化特征映射的学习机制。残差的定义为 $H(x)=F(x)+x$。其中，x 为输入图像，$F(x)$为残差网络需要学习的目标映射。所获得的残差网络构建的架构，即使残差为 0，残差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。初始层 $a^{[l]}$ 经过双层的残差结构运算，得到 $a^{[l+2]}$，如图 1 所示。

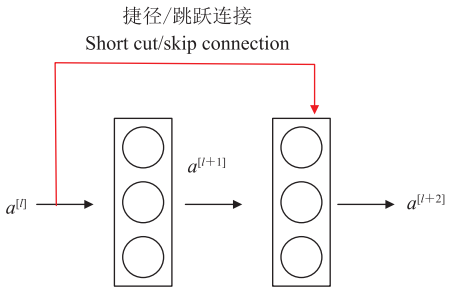


图 1 残差结构

Fig. 1 Residual structure

其中，short cut/skip connection（捷径/跳跃连接）建立起隔层相加的方式，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

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

计算过程从 $a^{[l]}$ 开始，首先进行线性激活，根据式（1），通过 $a^{[l]}$ 算出 $z^{[l+1]}$，即 $a^{[l]}$ 乘以权重矩阵 $\boldsymbol{W}^{[l+1]}$ 后加上偏差因子 $b^{[l+1]}$；其次，通过非线性函数 ReLU 激活得到 $a^{[l+1]}$；再次，进行线性激活，根据式（1）可以得出 $z^{[l+2]}$；最后，根据式（1）进行 ReLU 非线性激活，得到 $a^{[l+2]}$。其中，g 为 ReLU 非线性函数。

2.2 剪枝

深度神经网络（deep neural network，DNN）[6] 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，基于一个假设，即 DNN 的过参数化（over-parameterization）[7]。过参数化的训练阶段需要大量的参数来捕捉数据中的微小信息，然而由于一些浮点型的神经元连接结构的权重较小，对整个模型的影响较小，因此可以删减一些节点，以获得更小的模型和更快的推理速度。一般来说，剪枝会造成精度下降，因此，采用有效的裁剪模型并维持精度损失最小化是本研究的重点。

3 本文方法

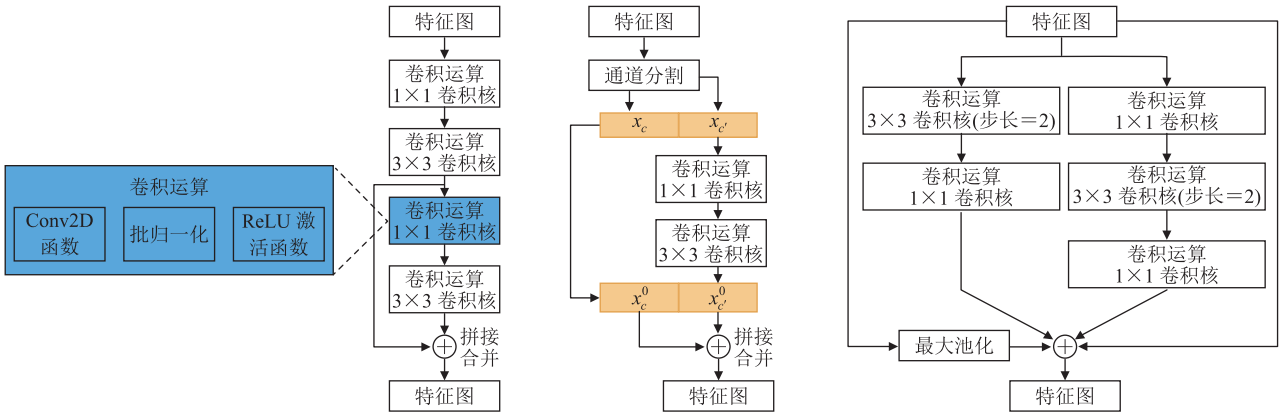
针对电动自行车安全头盔识别任务，本文算法的构建过程以骨干网络为基础，设计残差网络的优化框架，在特征融合网络层中构建通道重组，优化模型的特征交互，采用结构融合剪枝方法，进一步减少模型参数，提升推理速度。

3.1 分支吸收骨干网络

DarkNet53 骨干网络[8]使用全卷积网络和残差结构的方式有助于局部感知野和参数共享，并改善梯度消失的问题。所使用的降维连接方式借鉴了 GoogLeNet[9]相关思想，通过 1×1 与 3×3 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，上述方式在降维采样过程中损失了过多信息，原因是维度的大幅减少。为解决这一问题，本研究通过引入分支吸收的策略改进网络结构，并采用多模态耦合的降维方式，以求增加模型检测精度。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 3 层卷积块、两层卷积块及池化层构成，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，其中，c 为特征通道的总量，c' 为第二条特征通道的总量。经过带有跳跃连接的 3 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。

3.2 通道重组

本文采纳了多模式耦合策略，对多特征通道进行划分，并据此构建了分支吸收的骨干网络。本文在 DarkNet53 的框架下，融入了通道重组的



(a) DarkNet53 骨干网络中的特征提取模块　(b) 本文步长为 1 时构建的通道分割残差结构　(c) 本文步长为 2 时构建的 3 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

概念，具体体现在每一层卷积操作嵌入设计成型的相关算法模块。设计好的算法模块具体步骤如下。

步骤零：定义输入特征图 $\boldsymbol{A}$ 和输出重组图 $\boldsymbol{P}$。

步骤一：将特征图展开为 $c\times n\times w\times h$ 四维度矩阵，其中，c 是通道维度，n 是批次维度，w 是特征图宽，h 是特征图高。本步骤为特征图 $\boldsymbol{A}$ 的展开操作，其中，特征图被重构为四维矩阵形式。

步骤二：对上述四维矩阵执行维度置换操作，即沿着给定尺寸的矩阵的第 c 轴与第 n 轴进行转置。

步骤三：进行转置矩阵的扁平化处理，即折叠成一维的数组。

步骤四：对得到的特征图实施组内的 1×1 卷积运算，此过程在特征学习中促进了信息的有效整合，有助于进一步的特征抽象化。

3.3　结构融合剪枝

He 等[10]考虑在权重修剪中加入正则性，即过滤修剪、通道修剪[11]等，从而生成规则的和更小的权重矩阵，以便在 CPU/GPU 上更快地执行运算。如图 3 所示，过滤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。在高压缩率条件下，结构剪接技术因剪接滤波器信道中的信息丢失而引致显著的精度降低。

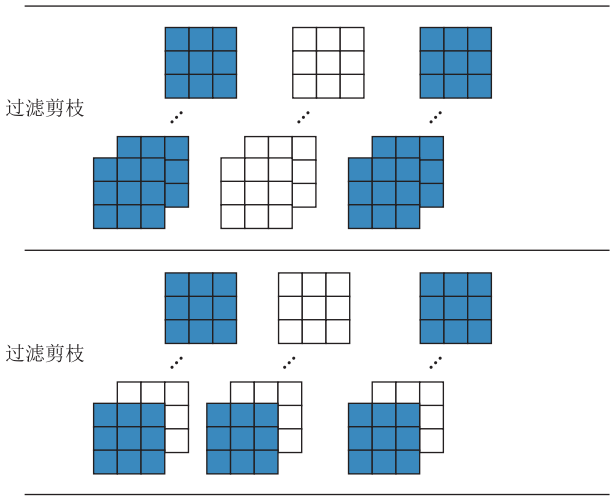


图 3　剪枝方法示意

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 4 所示，每一层的权值矩阵由多个大小不同的 $m\times n$ 的权值块组成。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，并确保每个块中剩余的权值仍然能够构成一

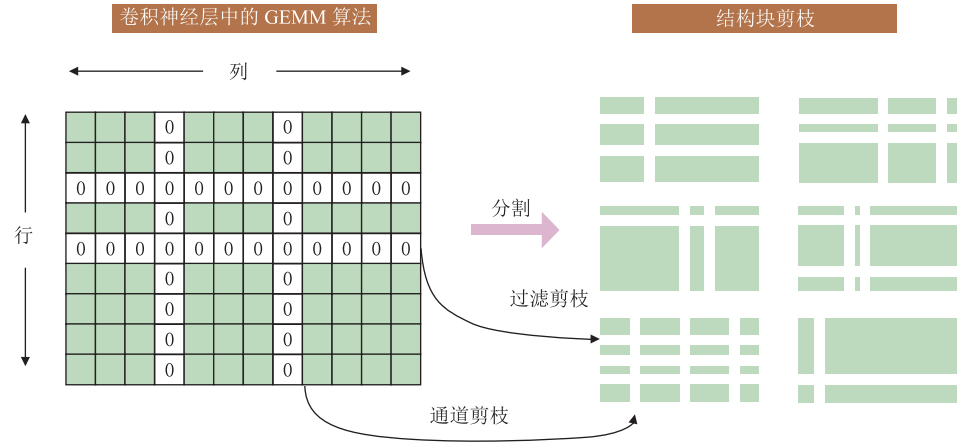


图 4 结构化融合剪枝过程

Fig. 4 Structured fusion pruning process

个完整且更小的矩阵。对于 N 层的深度神经网络来说，假设 w 为第 i 层的权重，则 $w=\{w_i\}_{i=1}^{N}$，假设有 $w_{ij}\in Z^{m\times n}$，其中 $Z^{m\times n}$ 为所有权块值之和，每个 w_i 会被均匀分成 j 个大小为 $m\times n$ 的矩阵，则 $w_i=\left[w_{i1},w_{i2},\cdots,w_{ij}\right]$。其中，$\left[w_{ij}\right]_p$ 为 w_{ij} 的第 p 行；$\left[w_{ij}\right]_q$ 为 w_{ij} 的第 q 列。

利用重加权方法解决组 lasso 正则化，从而消除对重要和不重要的权重应用相同惩罚的缺点。块结构融合剪枝中的列剪枝模块的计算方式为

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

其中，N 为对应的层数；$\boldsymbol{K}$ 为每个矩阵总和；$f\left(w_{\mathrm{loss}}\right)$ 为训练损失值；λ 为调整精度与稀疏度相对重要性的惩罚参数；$\odot$ 为 element-wise multiplication（两个矩阵对应位置元素进行乘积）；$\|\ \|_{\mathrm{Fro}}$ 为 Frobenius robe 范数的计算；ω_i^t 每一次迭代，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 ω_i^t：

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

式（3）在计算时，需要消除分母为 0 的干扰，而 $\Delta\beta$ 为防止分母为 0 的小值参数。

结构融合行剪枝模块有

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

可以通过式（5）更新 ω_i^t。

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

如图 5，本文采用的结构融合剪枝首先使用预先训练的模型进行初始化，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

	算法：结构化融合剪枝的重加权正则化
1	**初始化**：预先训练的 DNN 模型初始化权重 W 设置迭代次数为 T；定义块的大小为 $m\times n$
2	**层数** $i=\dfrac{w_i}{m\times n}$
3	**while** $t\leqslant T$ do
4	对权重或者层数使用梯度更新法则；
5	使用梯度更新的结果，来更新 ω_i^{t+1}
6	end
7	移除接近于 0 的权重组，并重新训练其余的非零权重，以提高准确性
8	结果：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

正则化项，即在优化问题中有界，在重新加权步骤后，去除接近于 0 的权值(或一组权值)，并利用非零权值细化 DNN。

3.4　评价指标

为分类任务选择正确的衡量指标也是评估算法的重要一环，在电动自行车安全头盔佩戴检测的场景中，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，前者体现了大多数的数据类，后者相对于前者只占小部分，所以是一个特征提取不平衡的分类问题。在这种情况下，准确率不是一种良好的评价指标，因此应同时考虑召回率和F1-score。其中，F1-score 是精确率(Precision)和召回率(Recall)的调和平均数，可以惩罚极端情况。F1-score(F1)的范围为 0～1，1 表示最好的性能，0 表示最差的性能，能衡量假阳性和假阴性的问题：

$$F1=2\times\frac{precision\times recall}{precision+recall} \tag{6}$$

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，地点为宁波市北仑区恒山路与四明山路区域，以拍摄视频、截取视频帧的方式进行处理，样本中包含不同的光照条件和不同颜色的头盔。此外，数据集样本还采用网络数据爬虫、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。本研究合计训练集 3 108 张，验证集 1 206 张。

基于原始数据，本文设计 MATLAB GUI 对图像任务进行处理(图 6)。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

本文采用经过优化的残差结构 DarkNet53[8]进行训练模型和检测实验。实验过程使用 50 个副本，每个副本均在 GPU 上独立运行，设定批次大小为 8，训练周期为 300 个 epoch。同时，设定动量为 0.9，初始学习率为 0.045，并在每个 epoch 周期内进行一次学习率衰减，其中指数衰减率为 0.94。将本文的方法(Result B)分别与 YOLOV3(Result A)[8]和 MobileNetV2(Result C)[11]进行对比，评价指标结果如图 7 所示。本文算法的 mAP 平均精度为 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 50 epoch 就能达到较好的训练效果，但漏检了部分目标，体现在 Recall Rate 查全率和 F1-score 等参数较差。

本文构建了多维度的评估检测标准体系，并通过统一收集和归一化的数据，以坐标系的模式进行展示。以下为一些评估指标的补充说明：(1)混淆矩阵是对分类问题预测结果的总结，以行表示预测的类别(y 轴)，以列表示真实的类别(x 轴)；(2) *Width* 和 *Height* 分别表示宽和高，其值对应的是定位锚框，在一定程度上，其锚框大小越小，其定位越准确；(3)置信度；(4)分类指标；(5)训练误差，指模型在训练集上的误差，反映模型的学习能力；(6)精确率(Precision)，表示模型正确预测为正样本的数量占所有预测为正样本的数量比例；(7)召回率(Recall)，表示模型正确预测为正样本的数量占所有实际正样本数量的比例；(8)多分类精度的平均值；(9)F1 为精确率和召回率的调和平均数；(10)测试误差，指模型在测试集上的误差，反映模型对未知数据的预测能力。

如图 8 所示，在目标重叠、特征类似、特征尺度极小的复杂情况下，本文采用分支吸收骨干

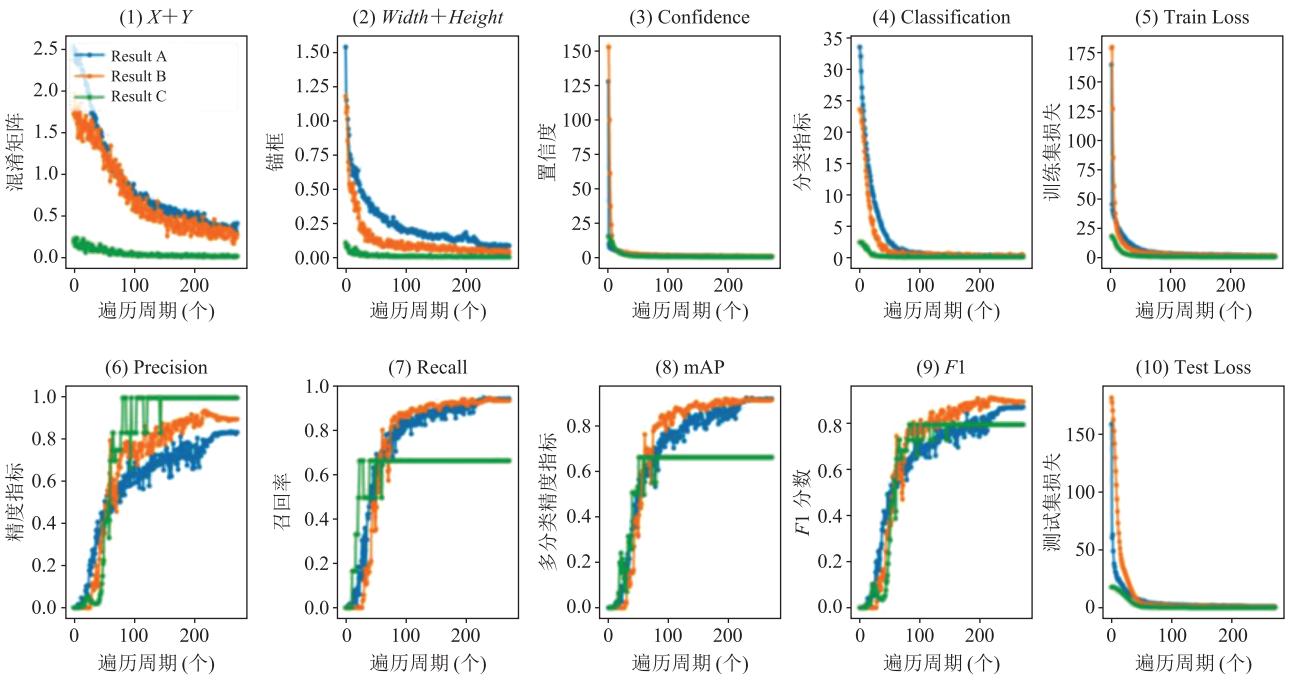


图 7 不同检测算法的评价指标结果比较

Fig. 7 Comparison of evaluation results for different detection algorithms

图 8　本文算法与 YOLOV3 实况检验对比

Fig. 8　Comparison of the algorithm with YOLOV3 in practical verification

网络与通道重组的方法，得到了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。

与过滤器剪枝和通道剪枝相比，本文采用的结构化剪枝具有更优的剪枝效果，体现在较小的精度损失和更大的实时性收益上。以 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% 的模型检测速度的增加，如表 1 所示。

表 1　测试结果与算法对比率

Table 1　Comparison in test results and algorithms

算法	*mAP*	*F*1	检测速度（帧/s）
YOLOV3	88.52%	0.87	24.8
MobileNet	69.87%	0.71	42.5
本文算法（引入剪枝压缩前）	91.78%	0.91	23.6
本文算法（引入剪枝压缩后）	88.80%	0.90	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，利用分支吸收模块改善 YOLO 算法的残差骨干网络，通过结构通道重组提升卷积层特征融合，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，本文算法的精度和实时性较优，小目标检测效果较好，多分类平均精度为 88.8%，检测速度为 29.5 帧/s，能够基本满足视频监控需求。未来，本研究团队将跟踪研究最新的深度学习算法，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。此外，本研究团队将通过模型蒸馏技术，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] // Proceedings of the 2020 International Conference on Smart Electronics and Communication (ICOSEC), 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, 35(3): 144-151.
Liu C, Wang JT, Wang YM. Application of SSD network with visual mechanism in motorcycle helmet wearing detection [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(3): 144-151.

[3] Wu F, Jin GQ, Gao MY, et al. Helmet detection based on improved YOLO V3 deep model [C] // Proceedings of the 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), 2019: 363-368.

[4] Zhou FB, Zhao HL, Nie Z. Safety helmet detection based on YOLOv5 [C] // Proceedings of the 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA), 2021: 6-11.

[5] He KM, Zhang XY, Ren SQ, 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 YH, Yang TJ, 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 YZ, Song Z. A convergence theory for deep learning via over-parameterization [C] // Proceedings of the 36th International Conference on Machine Learning, 2019: 242-252.

[8] Redmon J, Farhadi A. YOLOv3: an incremental improvement [Z/OL]. arXiv Preprint, arXiv: 1804.02767, 2018.

[9] Szegedy C, Liu W, Jia YQ, 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] // Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017: 1389-1397.

[11] Sandler M, Howard A, Zhu ML, et al. MobileNetV2: inverted residuals and linear bottlenecks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.