DOI:10.19651/j.cnki.emt.2313606

基于 YOLOv7 的驾驶人使用手机与抽烟行为识别方法*

娄 文[1] 郭杜杜[2] 张 杰[1] 赵 亮[1] 徐勤功[1]
(1.新疆大学智能制造现代产业学院 乌鲁木齐 830017;2.新疆大学交通运输工程学院 乌鲁木齐 830017)

摘 要:针对机动车驾驶人驾驶过程中使用手机与抽烟行为威胁交通安全的问题,本文提出了一种基于 YOLOv7 的改进网络模型。首先使用 MobileNetv3 主干网络代替原版 YOLOv7 的主干网络,减少模型参数量与计算量,提升模型的处理速度;利用深度可分离卷积、亚像素卷积搭建改进特征金字塔分支并与原版特征金字塔的输出特征层进行融合,丰富特征信息,增强特征提取效果;最后利用特征加强模块对融合特征层进行强化,提升特征层通道及区域两个方面的关注度。实验结果表明,改进网络模型的平均精度均值为 95.33%,检测速度为 75.31 fps,相比于原版 YOLOv7 网络的平均精度均值提高了 6.84%,检测速度增加了 17.25 fps。改进网络模型在满足实时检测的基础上具有较高的检测精度,能够实现对驾驶人使用手机与抽烟行为的实时、准确识别。

关键词:YOLOv7;目标检测;使用手机行为;抽烟行为
中图分类号:TP391 **文献标识码**:A **国家标准学科分类代码**:520.6040

Identification method of mobile phone use and smoking behavior of drivers based on YOLOv7

Lou Wen[1] Guo Dudu[2] Zhang Jie[1] Zhao Liang[1] Xu Qingong[1]
(1. School of Intelligent Manufacturing Modern Industry, Xinjiang University, Urumqi 830017, China; 2. School of Transportation Engineering, Xinjiang University, Urumqi 830017, China)

Abstract: To address the problem of motorists using cell phones and smoking behaviors during driving threatening traffic safety, this paper proposes an improved YOLOv7-based network model. Firstly, the MobileNetv3 backbone network is used instead of the original YOLOv7 backbone network to reduce the number of model parameters and computation and improve the processing speed of the model. The depth separable convolution and sub-pixel convolution are used to build an improved feature pyramid branch and fuse it with the output feature layer of the original feature pyramid to enrich the feature information and enhance the feature extraction effect. Finally, the feature enhancement module is finally used to enhance the fused feature layer to improve the attention of both the feature layer channels and regions. The experimental results show that the mean average precision of the improved network model is 95.33%, and the detection speed is 75.31 frames per second. Compared with the original YOLOv7 network, the mean average precision is increased by 6.84%, and the detection speed is increased by 17.25 frames per second. It has higher detection accuracy on the basis of satisfying real-time detection and can realize real-time and accurate detection of drivers' use of cell phones and smoking behavior.

Keywords: YOLOv7;object detection;cell phone use behavior;smoking behavior

0 引 言

驾驶人的不良行车习惯在很大程度上会影响道路交通安全,增加交通事故发生的可能性[1]。使用手机与抽烟行为是常见的驾驶人不良行车习惯。大约 12%的致命交通事故是驾驶人使用手机造成的[2]。一项驾驶行为调查报告说明,在受访者中超过 5 成的驾驶人在驾驶过程中会使用手机[3]。研究表明,吸烟者发生交通事故的概率是非吸烟

收稿日期:2023-05-11
* 基金项目:自治区重点研发计划项目(2022B01015-3)、重点实验室开放课题(2023ZDSYSKFKT06)项目资助

者的 1.5 倍[4]。因此，如何准确、高效地识别驾驶人使用手机与抽烟行为显得十分重要。

目前主要利用目标检测算法对驾驶人使用手机与抽烟行为进行识别，包括传统目标检测方法和基于深度学习的目标检测方法。传统的目标检测方法一般通过手动设计特征[5]，对提取的特征通过 SVM(support vector machine)或 Adaboost 等分类器进行分类，实现目标检测。王肖[6]与赵雄等[7]利用 OpenPose 和 Alpha pose 算法提取人体骨架关键点，虽然可以较好地识别驾乘人员打电话的行为，但会受限于姿态和角度因素。渠皓然[8]通过在 YCbCr 空间检测类肤色像素以实现驾驶人打电话行为检测，具有较高的正确率；但在驾驶人出现揉眼等相似动作时会出现误判的情况。Chen 等[9]使用 Vibe 算法与 HSV(hue saturation value)模型提取特征，分析形状变化及运动特性检测吸烟行为，可以达到实时检测的要求，但检测精度不佳。传统的目标检测算法需要人工设计特征，在特征设计及泛化能力方面仍然存在不足。

深度学习目标检测算法通过训练数据自动学习特征，实现目标的检测和分类，突破了传统目标检测算法的瓶颈。李俏[10]以 YOLOv3 方法及 RetinaNet 为基础进行改进，在人体动作目标检测方面获得了一定的提升，接打手机行为的识别精度达到了 87%，检测效果优于传统目标检测算法，但仍存在提升空间。针对空间特征识别方法不足的问题，关文喆[11]设计基于 SSD(single shot multibox detector)的脸－手位置网络，将空间特征与时域特征进行了结合，进一步提升了检测效果，对公交车驾驶员接电话行为的识别精度达到了 93%，满足了实际检测的精度需求。针对以往手机检测方法容易出现物体遮挡[12]、图像旋转、光照变化、难以提取图像深层特征等问题，Xiong 等[13]通过人脸检测跟踪和手机目标检测两个步骤有效地检测了驾驶人打电话的行为，算法的准确率达到 96%，处理速度达到 25 fbs，虽然具有较好的检测效果，但难以满足高帧率视频实时检测的要求。针对传统 HSV 模型吸烟检测精度不足的问题，Huang 等[14]设计了一种基于深度学习的抽烟检测模型，提取烟雾多特征信息，检测准确率达到 90%以上，但其检测速度仍然无法满足实时要求。为了提升目标检测速度，郭佳伟[15]通过剪枝和量化方法压缩模型大小，以改进 MobileNet-SSD 为基础设计行为决策算法实现驾驶人抽烟行为识别，识别速度提升了约 34.3%，为优化算法模型提供了新的思路。Zhao 等[16]设计了改进的 YOLOv4 轻量级吸烟检测模型，检测速度比原网络提升了 57%，但是检测精度有所下降。

目前大多数研究只针对驾驶人使用手机或抽烟的单一行为进行识别；且由于目标检测网络检测精度与检测速度的矛盾性，大多数研究侧重于单独提升识别精度或识别速度，综合性能存在不足。因此为了实现对驾驶人使用手机与抽烟行为的实时、准确识别，本文提出了一种基于 YOLOv7 的改进网络模型(mobilenet-YOLOv7，M－YOLOv7)，首先利用 MobileNetv3 网络替换输入及主干网络部分，减少模型参数量和计算量，提升网络运行速度；其次利用深度可分离卷积(depthwise separable convolution，DSC)、亚像素卷积(pixel shuffle，PS)搭建改进特征金字塔分支(improved-feature pyramid network，I-FPN)并与原版特征金字塔(feature pyramid network，FPN)的输出特征层进行融合，丰富特征信息，提升目标检测的效果；然后利用特征加强模块(convolutional block attention module，CBAM)对融合特征层进行强化，提升特征层通道及区域两个方面的关注度，进一步提高网络的识别准确率及泛化能力。实验结果显示，本文提出的改进网络在检测精度和速度上都较改进前有较大提升。

1　YOLOv7 网络

YOLOv7[17]使用了创新的多分支堆叠结构进行特征提取，使用了创新的下采样结构，并且将 YOLOv5 的跨网格搜索以及 YOLOX[18]的匹配策略结合起来，还提出了辅助头的一个训练方法，通过增加训练成本，提升精度，同时不影响推理预测的时间，在准确率和速度上超越了以往的 YOLO(you only look once)系列。

YOLOv7 网络的整体结构如图 1 所示，网络主要包括输入及主干网络、颈部、头部 3 个部分。输入端为图像输入；主干网络学习输入图像的特征信息；颈部强化特征层信息；头部利用输出特征层进行目标检测。

2　M-YOLOv7 网络

本文以 YOLOv7 网络为基础搭建了 M-YOLOv7 网络，其整体结构如图 2 所示。其中输入及主干网络部分替换为 MobileNetv3 主干网络；颈部利用深度可分离卷积、亚像素卷积搭建 I-FPN 并与原版 FPN 的输出特征层进行融合；然后利用特征加强模块对融合特征层进行强化，最后将特征层传入头部获得预测结果。

2.1　输入及主干网络

MobileNetv3[19-20]具有更轻量化的网络架构，其模型的参数量和计算量较少，可以达到较高准确率的情况下减少计算资源和模型大小并且更加适用于移动设备。MobileNetv3 网络的特点是使用了深度可分离卷积以及具有线性瓶颈的逆残差结构(inverted resblock，IR)，而且在逆残差结构中加入了轻量型的注意力模块(squeeze and excitation，SE)。利用 MobileNetv3 替换 YOLOv7 的主干网络，可以减少网络总体参数量，提升网络的运行速度。

1)深度可分离卷积

深度可分离卷积由逐通道卷积(depthwise convolution，DC)和逐点卷积(pointwise convolution，PC)组成。深度可分离卷积的结构如图 3 所示。

以输入通道为 3，卷积核大小为 3×3，数量为 4 个示

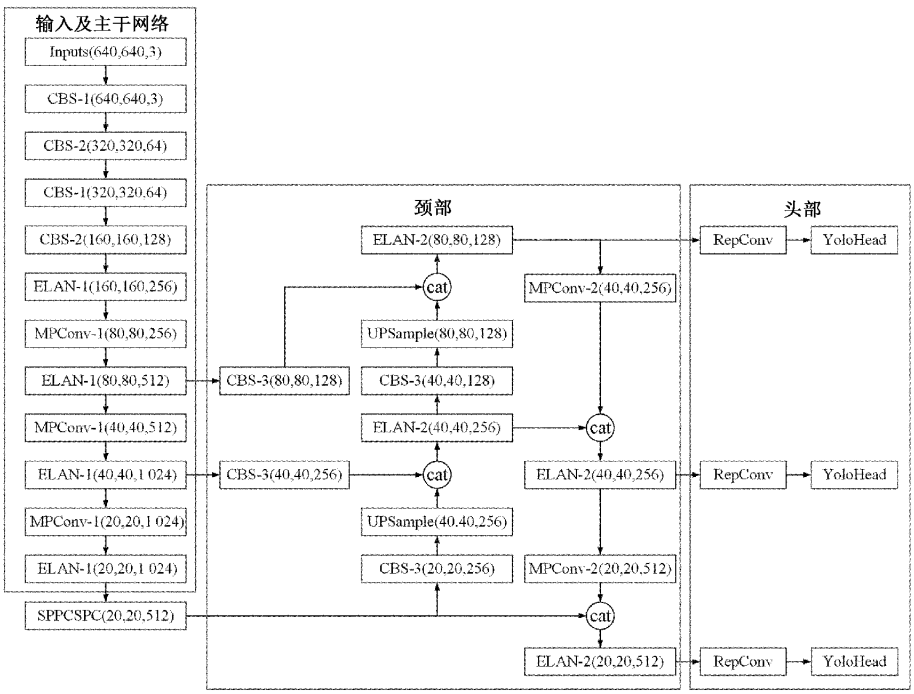


图 1　YOLOv7 网络结构

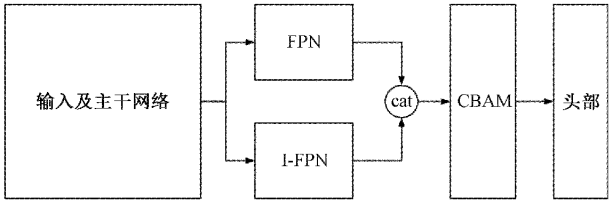


图 2　M-YOLOv7 网络结构图

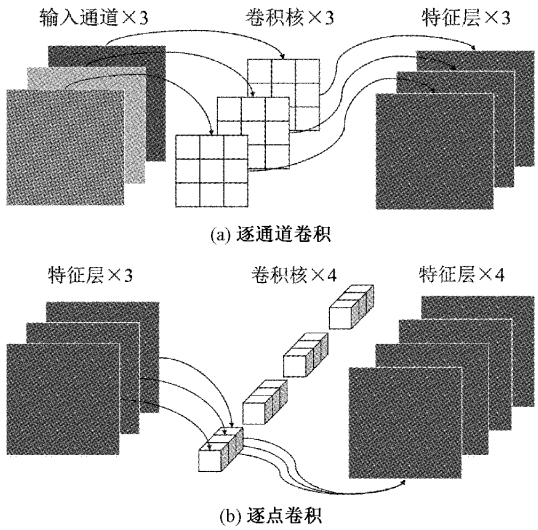


图 3　深度可分离卷积

例,普通卷积的参数个数为 108,而深度可分离卷积的参数个数为 39,深度可分离卷积的参数大约是常规卷积的 1/3。深度可分离卷积比普通卷积具有更少的参数量,在卷积后特征图尺度相同的情况下,深度可分离卷积具有更少的计算量。

2)逆残差结构

MobileNetv3 网络使用了线性瓶颈的逆残差结构,并在其中加入轻量型的注意力模块。逆残差结构如图 4 所示。其中 NL 指激活函数,不同的输入对应不同的激活函数类型;Pool 指池化操作;FC 指全连接层(fully connected,FC)。

逆残差结构首先利用 1×1 的卷积对输入的通道数进行扩张,然后利用 3×3 的逐通道卷积进行学习,通过池化得到 1×1 的列向量,利用全连接层及 ReLU 激活函数处理使其通道数变为 1/4,再利用全连接层及 H-Sigmoid 函数恢复为原通道数得到通道权重,将通道权重与池化前的特征层相乘,得到赋权的特征层,最后利用 1×1 的卷积调整赋权特征层的通道数。当逐通道卷积步长为 1、输入与输出尺寸相同时,最下方的跳跃连接启用,将原输入与输出特征层对应元素相加;当逐通道卷积步长为 2,输入与输出尺寸不相同时,跳跃连接不会启用。

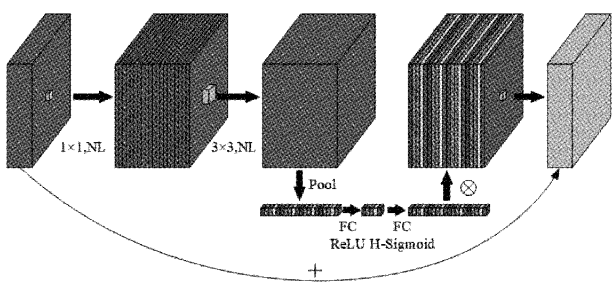


图 4　逆残差结构图

3)MobileNetv3 主干网络

MobileNetv3 主干网络结构如图 5 所示。

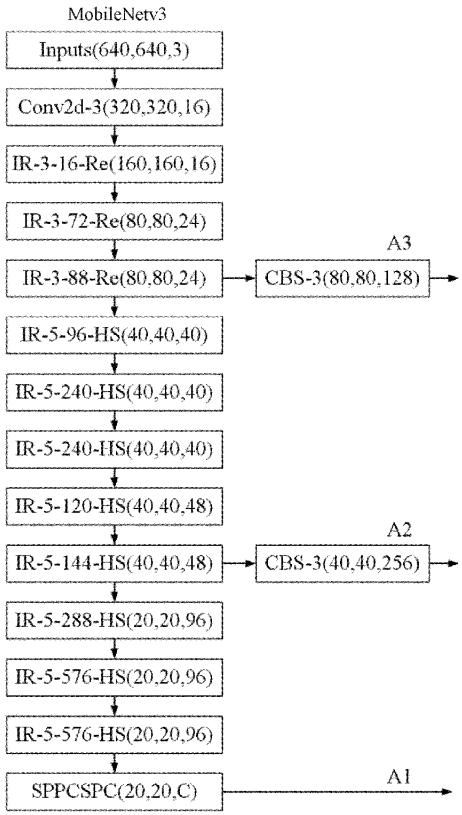


图 5　MobileNetv3 主干网络结构图

Conv2d-3 指卷积核大小为 3×3 的普通卷积；IR-3-16-Re 指逆残差结构中逐通道卷积的卷积核大小为 3×3，扩张的通道数为 16，使用 ReLU 激活函数，其他部分的含义与此类似；HS 指使用 H-Swish 激活函数；C 指特征层调整后的通道数，对于不同的颈部分支有不同的数值。当 MobileNetv3 网络与 FPN 结构相连时，C 取值为 512，当 MobileNetv3 网络与 I-FPN 结构相连时，C 取值为 1 024。

利用 MobileNetv3 网络替换 YOLOv7 的主干网络，最后输出 3 个特征层，分别利用 CBS-3 和 SPPCSPC 处理得到特征层 A3、A2、A1，输入到后续结构中进行特征强化。

2.2　颈部

颈部利用深度可分离卷积对堆叠特征进行特征学习及整合，利用亚像素卷积进行上采样，搭建 I-FPN 结构，并与原版 FPN 的输出特征层进行融合，丰富特征信息，提升目标检测的效果。

1)亚像素卷积

亚像素卷积[21]的结构如图 6 所示。对于尺寸为 2×2，4 个通道的特征层，通过亚像素卷积将不同通道的同一位置的数据进行重组，得到一个新的特征层，其尺寸是原来的两倍，通道数变为原来的 1/4。

亚像素卷积的主要功能是通过卷积对特征层多通道间的信息进行重组，提高特征层的分辨率。亚像素卷积可以更好的利用通道之间的信息，具有更好的上采样效果。

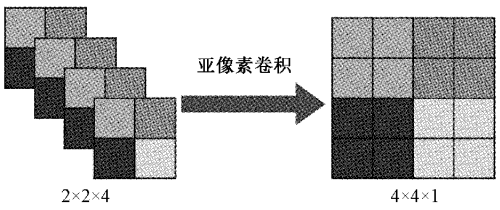


图 6　亚像素卷积

因此在特征金字塔分支中采用亚像素卷积可以更好地利用特征层信息。

2)I-FPN

I-FPN 结构如图 7 所示，图中 DBL 指深度可分离卷积加 BN(batch normalization)加 Leaky ReLU 激活函数的操作，PSSample 指亚像素卷积。以主干网络提取的 A1(20×20×1 024)特征层为基础，利用亚像素卷积进行上采样，并与 A2(40×40×256)拼接后利用 DBL 进行整合得到 A20(40×40×512)。A20 利用亚像素卷积进行上采样，与 A3(80×80×128)拼接后利用 DBL 进行整合得到 A30(80×80×256)。A30 调整通道后获得第 1 个输出特征层 B3。B3 利用 DBL 调整尺寸后与 A20 拼接并利用 DBL 进行整合得到第 2 个输出特征层 B2。B2 利用 DBL 调整尺寸后与 A1 拼接并利用 DBL 进行整合得到第 3 个输出特征层 B1。

3)原版 FPN

将 MobileNetv3 主干网络得到的特征层(A3、A2、A1)输入到原版 YOLOv7 的颈部结构中，通过两次上采样及拼接操作完成自下而上的特征金字塔分支，通过两次过渡下采样及拼接操作完成自上而下的特征金字塔分支，最终输出三个特征层 F3(80×80×128)、F2(40×40×256)和 F1(20×20×512)。

将两个分支输出的特征层 F3、F2、F1 分别与 B3、B2、B1 进行拼接，利用卷积调整其通道数，得到 3 个输出特征层 Out3、Out2、Out1，进行后续处理。

2.3　CBAM

CBAM[22]结构如图 8 所示，其主要由通道注意力部分(channel attention，CA)和空间注意力部分(spatial attention，SA)组成。该结构通过将通道注意力机制与空间注意力机制组合起来，可以有效提高特征层信息的表达能力。

通道注意力部分的结构如图 9 所示。通道注意力部分首先使用最大池化和平均池化聚合特征层的空间信息，使其成为两个列向量，然后通过共享感知器对两个列向量的参数进行处理。在共享感知器中，给定 1 个压缩率 r，先将通道数压缩为原来的 1/r，再扩张到原通道数，经过 ReLU 激活函数得到两个激活后的结果。将这两个激活后的结果逐元素相加，再通过一个 Sigmoid 激活函数得到通道权重。将该通道权重与原输入特征层相乘，即可得到通道赋权的特征层。

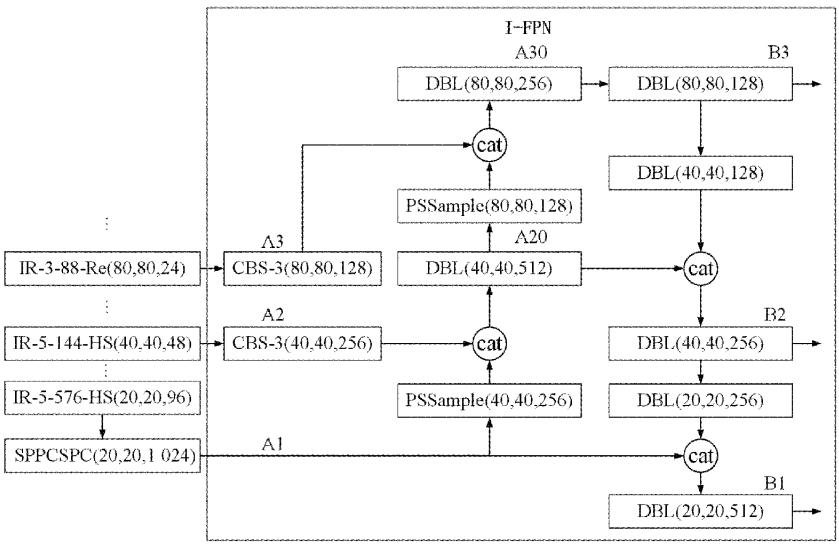


图 7　I-FPN 结构图

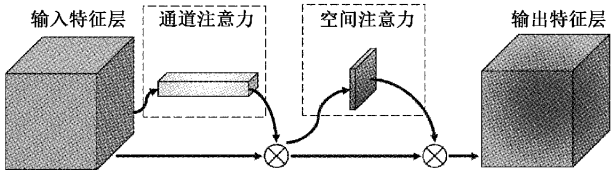


图 8　CBAM 模块

通道注意力可以提升特征层通道之间的信息连接,给定不同通道的权重系数,通过权重系数的大小表示不同通道的重要程度。在网络结构中加入通道注意力机制可以使网络对不同通道的信息加以重视,关注更加重要的通道特征,提高神经网络对图像的表示和识别效果,进而提高网络的识别准确率及泛化能力。

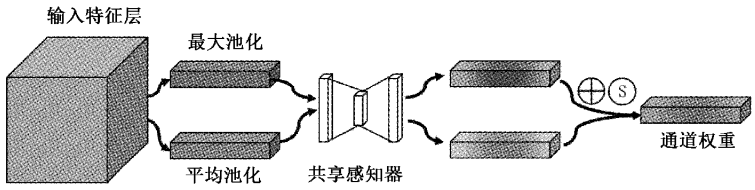


图 9　通道注意力

空间注意力部分的结构如图 10 所示。空间注意力将通道赋权的特征层沿通道轴方向进行平均池化和最大池化,得到两个尺寸不变、通道为 1 的特征层。将这两个特征层进行拼接,然后利用尺寸为 7×7 的卷积进行融合获得 1 个单通道的特征层,经 Sigmoid 函数处理后得到空间权重,如式(1):

$$M_s(u_c)=\delta(w_1(\mathrm{Max}+\mathrm{Avg})) \tag{1}$$

其中,$M_s(u_c)$为空间权重,u_c 为通道赋权的特征层,对图像进行 Max 和 Avg 两种池化并融合,w_1 为图像 7×7 卷积,δ 为 Sigmoid 激活函数。

将该空间权重与通道赋权的特征层相乘,即可得到强化后的特征层,如式(2):

$$M_F=M_s(u_{ci})\times u_{ci} \tag{2}$$

其中,M_F 为强化后的特征层,u_{ci} 为输出特征层 Out_i,i 的范围为 1～3。

空间注意力可以通过给定不同区域的权重系数来表示特征层不同区域的重要程度。在网络结构中加入空间注意力机制可以使网络在处理图像时更加关注图像的重要区域,进而提高网络的识别准确率及泛化能力。

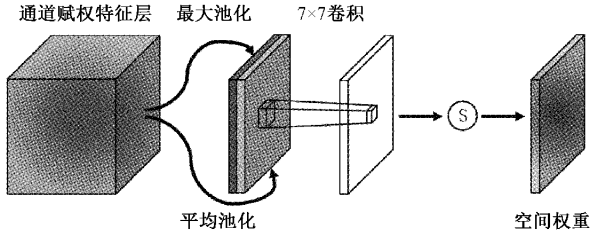


图 10　空间注意力

3　实验环境与数据集

3.1　实验环境

本研究实验采用的系统为 Windows10,内存为 32 G,CPU 型号为 AMD Ryzen 9 5950X,GPU 型号为 NVIDIA GeForce RTX3090Ti,采用的深度学习框架为 PyTorch 1.10.0。

3.2　数据集

本文建立了使用手机与抽烟检测数据集，共包含 10 427 张图片，其中使用手机数据 5 192 张，抽烟数据 5 235 张。数据来源可分为 3 部分：第 1 部分为网络查找的数据；第 2 部分为自采数据，包括模拟驾驶环境与实车驾驶环境下拍摄的数据；第 3 部分为 StateFarm 公共数据集中选取的包含手机目标的部分数据。数据集示例如图 11 所示。将数据集按照数量比例 8∶1∶1划分为训练集、验证集和测试集。

图 11　数据集示例

4　实验与分析

4.1　网络训练

M-YOLOv7 网络训练参数设置如表 1 所示。在训练过程中使用了在 ImageNet 上训练好的 MobileNetv3 网络权重，首先在前 50 个迭代中冻结网络模型的特征提取部分，此时特征提取网络不发生改变，仅对网络进行微调，系统显存占用较小，将批次处理数设置为 24；50 个迭代训练完成之后，解冻模型所有层，再次进行 250 个迭代的训练，此时网络所有的参数都会发生改变，系统显存占用较大，将批次处理数设置为 12。设置最大迭代次数为 300，当完成 300 个迭代训练后，停止训练过程。

表 1　训练参数配置

参数	值
Input	640×640×3
Freeze_Epoch	50
Batch_Size1	24
UnFreeze_Epoch	250
Batch_Size2	12
Max_Epoch	300
Learning Rate	0.01
Optimizer	SGD

M-YOLOv7 网络在训练过程中的 Loss 下降曲线如图 12 所示。在前 50 个迭代训练过程中，Loss 值下降较快；在后 250 个迭代训练过程中，Loss 值下降逐渐缓慢，在第 250 个迭代之后，训练集的损失已接近平稳，此时训练基本拟合。

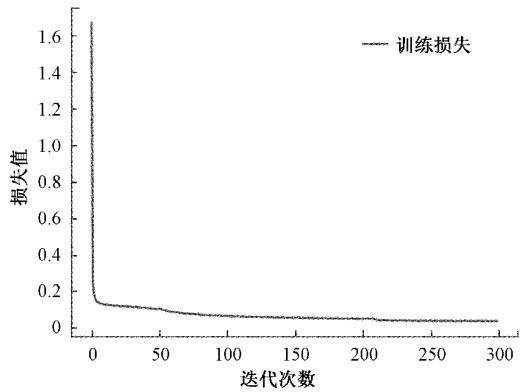


图 12　训练损失下降曲线

4.2　评价指标

精度评价指标主要有精确率(precision，P)、召回率(recall，R)、平均精度均值(mean average precision，mAP)，速度评价指标主要为每秒处理帧数(frames per pecond，FPS)。精确率指的是被分类为正样本的集合中实际为正样本的比例，如式(3)：

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

其中，TP 为正样本集合中分类为正样本的样本数量；FP 为负样本集合中分类为正样本的样本数量。

召回率指的是实际为正样本的集合中被分类为正样本的比例，如式(4)：

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

其中，FN 为正样本集合中分类为负样本的样本数量。

以精确率为横轴坐标，召回率为纵轴坐标，可绘制 P-R 曲线，该曲线与坐标轴围成的面积为单一类别目标检测的平均精度(average precision，AP)。mAP 定义为多类目标 AP 的均值，用来衡量数据集中多类目标的平均精度，如式(5)：

$$mAP = \frac{\sum_{i=0}^{N} AP}{N} \tag{5}$$

其中，AP 为平均精度，N 为目标类别。

FPS 可以代表算法模型对图像帧数的处理速度，单位为帧/秒。FPS 越高，表示模型的检测速度越快，可以更快地对输入数据进行预测和处理。

4.3　对比实验

选择多个对比网络，利用数据集进行训练测试。在训练过程中，设置各对比网络的迭代次数为 50(冻结训练)和 250(解冻训练)，设置批次处理数为 24(冻结训练)和 12(解冻训练)，选择相同的训练环境。对比实验的结果如表 2 所示。

由对比实验的结果可知，M-YOLOv7 网络的 mAP 为 95.33%，比其他对比网络的值都要高，说明 M-YOLOv7 网络的检测精度最好。相比于原版 YOLOv7 算法，改进

表 2　对比实验结果

网络	手机 AP/%	抽烟 AP/%	mAP/%	FPS
SSD	83.10	77.16	80.13	150.21
CenterNet	86.42	83.83	85.13	107.13
YOLOv5_L	91.04	88.92	89.98	50.83
YOLOv5_S	85.23	83.25	84.24	76.17
YOLOX_L	91.32	89.54	90.43	48.21
YOLOX_S	79.80	78.93	79.37	75.41
YOLOv7	89.10	87.87	88.49	58.06
M-YOLOv7	**95.85**	**94.80**	**95.33**	**75.31**

YOLOv7 算法的 mAP 和 FPS 值都有所提升。在所有对比网络中,SSD 算法的 FPS 最高,但其 mAP 值较低,检测效果不好;YOLOX_L 算法的 mAP 值最高,但其 FPS 值仅有 48.21 帧/秒。综合考虑网络模型的识别精度与识别速度,本文提出的 M-YOLOv7 网络识别效果最佳。

4.4　消融实验

为了验证各个改进模块的优化效果,利用数据集对 YOLOv7 原版算法进行训练,在此基础上通过替换主干网络、添加 I-FPN 分支以及加入 CBAM 模块进行消融实验,实验结果如表 3 所示。A 模型为 YOLOv7 原版算法,B 模型表示在 A 模型基础上替换 MobileNetv3 主干网络,C 模型表示在 B 模型基础上添加 I-FPN 分支,D 模型表示在 C 模型基础上加入 CBAM 模块。

由表 3 可知,当引入改进 1 后,精度降低了 0.46%,速度提升了 30.62 fps,这表明替换 MobileNetv3 主干网络后降低了参数量和计算量,网络的运行速度得到了提升,但精度有小幅度下降。当引入改进 2 后,精度提高了 3.82%,速度降低了 4.81 fps,这表明添加了 I-FPN 分支后,增加少量参数及计算量,网络的运行速度有所下降,但将 I-FPN 分支与原版 FPN 的输出融合,加强了特征表达,提高模型的性能,精度得到了提高。当引入改进 3 后,精度提高了 3.48%,速度降低了 8.56 fps,这表明 CBAM 模块通过通道注意力和空间注意力机制自适应地学习了特征层不同通道的重要性权重,使得网络更加关注图像中的重要信息通道,同时自适应地学习到特征层不同空间位置的重要性权重,并将这些权重应用于特征图的每个通道上,从而使得网络更加关注图像中的重要区域,实现了通道和空间的交互调整,提高模型的表达能力、泛化性能及稳定性,但是由于增加了网络的计算量,因此使得网络的运行速度有所下降,但仍满足实时检测(60 fps)的要求。综合来看,通过引入 3 个改进点,模型的检测精度提高了 6.84%,模型的检测速度提升了 17.25 fps,证实了 M-YOLOv7 网络模型的优越性能。

表 3　消融实验结果

模型	改进 1	改进 2	改进 3	mAP/%	FPS
A	×	×	×	88.49	58.06
B	√	×	×	88.03	88.68
C	√	√	×	91.85	83.87
D	√	√	√	95.33	75.31

4.5　测试结果

为了更好的感受 M-YOLOv7 网络与其他对比网络的检测效果,从测试集中选取部分图片进行检测;综合考虑到检测精度与检测速度的性能,选择 M-YOLOv7、YOLOv5_S、YOLOX_S 3 种网络进行测试对比,结果如表 4 所示。

表 4　网络测试结果

目标	网络	实际帧数	目标帧数	检测正确帧数	精确率/%	召回率/%	FPS
手机	YOLOv5_S	3 445	3 275	3 057	93.34	88.74	74.17
	YOLOX_S	3 445	3 427	3 216	93.84	93.35	71.54
	M-YOLOv7	3 445	3 468	3 350	96.60	97.24	71.23
抽烟	YOLOv5_S	3 217	3 145	2 759	87.73	85.76	74.36
	YOLOX_S	3 217	3 158	2 855	90.41	88.75	71.44
	M-YOLOv7	3 217	3 164	3 015	95.29	93.72	71.10

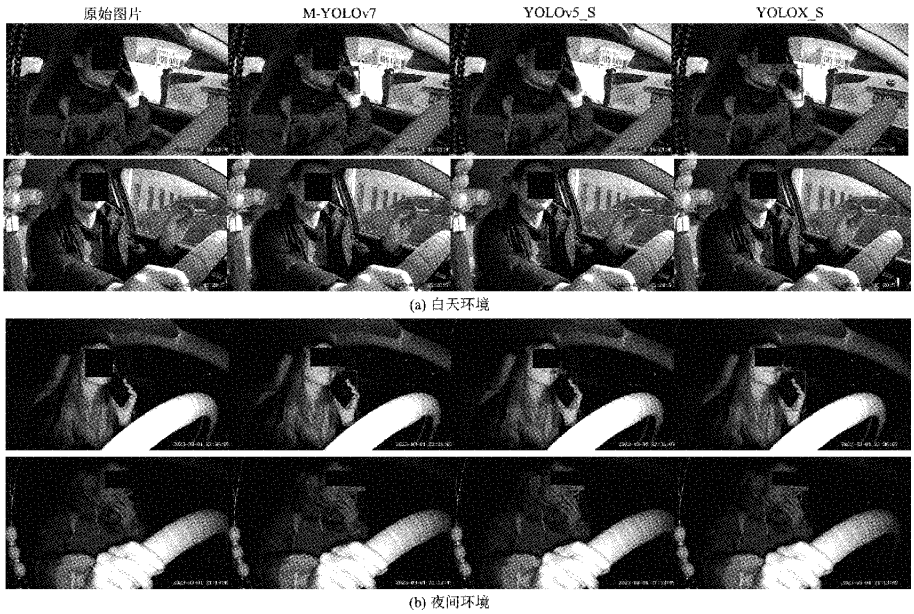
表 4 中,实际帧数指实际包含手机或抽烟目标的图片数量,目标帧数为网络检测到包含手机或抽烟目标的图片数量,检测正确帧数指实际包含手机或抽烟目标的图片中被正确检测到的图片数量。

由此可知,在手机和抽烟目标的检测过程中,M-YOLOv7 网络的精确率与召回率普遍高于 YOLOv5_S 和 YOLOX_S 网络,证明 M-YOLOv7 网络可以实时、准确地识别使用手机和抽烟行为。

此外,分别选取了白天环境和夜间环境下的检测结果进行对比,结果如图 13 所示。

图 13 中第 1 行和第 3 行的检测框代表手机目标,第 2 行和第 4 行的检测框代表香烟目标。3 个网络目标检测的置信度信息如表 5 所示。

对于白天环境检测效果,YOLOv5_S 在检测手机目标时出现了漏检情况,YOLOX_S 与 M-YOLOv7 均检测出手机与抽烟目标,但 M-YOLOv7 的置信度普遍较高。对



(a) 白天环境

(b) 夜间环境

图 13　检测效果对比图

表 5　目标检测置信度

环境	目标	M-YOLOv7	YOLOv5_S	YOLOX_S
白天	手机	0.85	—	0.74
	抽烟	0.77	0.69	0.74
夜间	手机	0.85	0.69	0.83
	抽烟	0.85	0.71	0.76

于夜间环境检测效果，3 种网络都正确检测出了手机与抽烟目标，但 M-YOLOv7 的置信度普遍较高，证明 M-YOLOv7 的检测效果具有更高的可信度和准确性。综上分析，M-YOLOv7 具有较好的目标检测效果，综合性能优越。

5　结　　论

针对机动车驾驶人驾驶过程中使用手机与抽烟行为威胁交通安全的问题，本文提出了基于 YOLOv7 网络改进的 M-YOLOv7 网络模型。首先利用 MobileNetv3 主干网络替换了原版 YOLOv7 的主干网络，减少模型参数量与计算量，提升模型的处理速度。其次搭建了 I-FPN 分支，并与原版 FPN 的输出融合，加强了特征表达，提高模型的性能。最后利用 CBAM 对融合特征层进行强化，提升特征层通道及区域两个方面的关注度，实现了通道和空间的交互调整，提高模型的表达能力、泛化性能及稳定性。通过对比实验、消融实验及测试结果对比可知，M-YOLOv7 的 mAP 值为 95.33%，FPS 值为 75.31 fps，相比于原版 YOLOv7 网络的 mAP 值提高了 6.84%，FPS 值增加了 17.25 fps；在满足实时检测的基础上具有更高的检测精度，综合性能超越了其他对比网络，能够实现对驾驶人使用手机与抽烟行为的实时、准确识别。在后续工作中，尝试优化算法结构，考虑将算法移植到车载移动端设备或嵌入式设备，使算法具有更高的可应用性。

参考文献

[1] 张智腾. 基于卷积神经网络的驾驶员疲劳检测[D]. 长沙：湖南大学，2018.

[2] SHAABAN K，GAWEESH S，AHMED M M. Investigating in-vehicle distracting activities and crash risks for young drivers using structural equation modeling[J]. PLOS ONE，2020，15(7)：e0235325.

[3] ATWOOD J，GUO F，FITCH G，et al. The driver-level crash risk associated with daily cellphone use and cellphone use while driving[J]. Accident Analysis & Prevention，2018，119：149-154.

[4] CHIEN T C，LIN C C，FAN C P. Deep learning based driver smoking behavior detection for driving safety[J]. Journal of Image and Graphics，2020，8(1)：15-20.

[5] 刘星，蔡乐才，陈波杰，等. 基于 YOLOv5 的轻量化端到端手机检测方法[J]. 电子测量技术，2023，46(1)：188-196.

[6] 王肖. 典型异常驾驶行为识别与预警方法研究[D]. 重庆：重庆邮电大学，2020.

[7] 赵雄，陈平，潘晋孝. 基于骨架关键点的车内异常行为识别方法[J]. 机械与电子，2021，39(3)：10-15.

[8] 渠皓然. 基于机器视觉的驾驶员异常行为检测[D]. 天津：天津工业大学，2020.

[9] CHEN S，JIA K，LIU P，et al. Taxi drivers' smoking behavior detection in traffic monitoring video[C]. 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference（APSIPA

ASC). IEEE, 2019: 968-973.
[10] 李俏. 基于深度学习的机车驾驶员异常行为识别方法研究[D]. 石家庄:石家庄铁道大学, 2020.
[11] 关文喆. 基于机器视觉的司机异常驾驶行为研究[D]. 北京:北京邮电大学, 2020.
[12] 李智伟, 杨亚莉, 钟卫军, 等. 基于改进的 SSD 模型手机违规使用目标检测[J]. 电子测量与仪器学报, 2021, 35(1): 120-127.
[13] XIONG Q, LIN J, YUE W, et al. A deep learning approach to driver distraction detection of using mobile phone[C]. 2019 IEEE Vehicle Power and Propulsion Conference(VPPC). IEEE, 2019: 1-5.
[14] HUANG J, LI R. Smoking driving behavior detection based on deep learning [J]. Academic Journal of Science and Technology, 2023, 5(2): 59-62.
[15] 郭佳伟. 基于计算机视觉的驾驶员异常行为识别与预警[D]. 大连:大连海事大学, 2019.
[16] ZHAO K. Real time detection of drivers' smoking behavior using the improved YOLO-V4 model[C]. 2022 2nd International Conference on Computer Technology and Media Convergence Design (CTMCD 2022). Atlantis Press, 2022: 126-134.
[17] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [J]. ArXiv Preprint, 2022, ArXiv: 2207.02696.
[18] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding yolo series in 2021 [J]. ArXiv Preprint, 2021, ArXiv: 2107.08430.
[19] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3 [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1314-1324.
[20] 伍锡如, 邱涛涛, 王耀南. 改进 Mask R-CNN 的交通场景多目标快速检测与分割[J]. 仪器仪表学报, 2021, 42(7): 242-249.
[21] SHI W, CABALLERO J, HUSZAR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1874-1883.
[22] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 3-19.

作者简介

娄文,硕士研究生,主要研究方向为图像处理与目标检测、智能交通管控技术与装备。
E-mail:2495835634@qq.com

郭杜杜(通信作者),副教授,主要研究方向为智能交通。
E-mail:guodd@xju.edu.cn

张杰,硕士研究生,主要研究方向为智能交通管控技术与装备。
E-mail:1345047991@qq.com

赵亮,硕士研究生,主要研究方向为智能交通管控技术与装备。
E-mail:2105203262@qq.com

徐勤功,硕士研究生,主要研究方向为智能交通管控技术与装备
E-mail:1446158165@qq.com