摩托车驾驶员和行人检测算法综述

陈俊豪　苏山杰

（重庆交通大学机电与车辆工程学院，重庆 400074）

【欢迎引用】陈俊豪，苏山杰. 摩托车驾驶员和行人检测算法综述[J]. 汽车文摘，2023(2): 34-43.

【Cite this paper】CHEN J H, SU S J. A Review on Detection Algorithms for Motorcycle Driver and Pedestrian [J]. Automotive Digest (Chinese), 2023(2): 34-43.

【摘要】行人与摩托车驾驶员的违规操作是造成事故发生的重要原因，加大摩托车驾驶员头盔佩戴与行人违规的监督力度对摩托车驾驶安全至关重要。论述了驾驶员头盔与行人检测的主要任务与特点，梳理了当前基于深度学习的检测算法优点、缺点，对近年来安全头盔与行人检测研究进展进行总结，结合检测难点对该领域进行分析与展望。

主题词：深度学习　目标检测　行人检测　头盔检测　摩托车检测

中图分类号：TP183　　**文献标识码：**A　　**DOI:** 10.19822/j.cnki.1671-6329.20220096

A Review on Detection Algorithms for Motorcycle Driver and Pedestrian

Chen Junhao, Su Shanjie

(School of Mechatronics and Vehicle Engineering, Chongqing Jiaotong University, Chongqing 400074)

【Abstract】 Illegal operations by pedestrians and motorcycle drivers are the primary cause of accidents. Increasing supervision of helmet wear and pedestrian violations by motorcycle drivers is critical to motorcycle driving safety. This paper discusses the main tasks and characteristics of driver helmet and pedestrian detection, introduces the principle and advantages and disadvantages of the current deep learning-based detection algorithms, summarizes the research status of safety helmets and pedestrians in recent years, and analyzes and prospects in this field in combination with detection difficulties.

Key words: Deep learning, Target detection, Pedestrian detection, Helmet detection detection, Motorcycle detection

缩略语

SIFT	Scale-Invariant Feature Transform
HOG	Histogram of oriented gradient
NMS	Non-Maximum Suppression
FPN	Feature Pyramid Network
YOLO	You Only Look Once
IOU	Intersection Over Union
GIOU	Generalized Intersection Over Union
CIOU	Complete Intersection Over Union
SSD	Single Shot Multibox Detector
ROI	Region Of Interest
CA	Coordinate Attention
mAP	mean Average Precision
SE	Squeeze-and-Excitation
SK	Selective Kernel
SS	Selective Search
CBAM	Convolutional Block Attention Module
MLP	Multilayer Perceptron
SVM	Support Vector Machine
RCNN	Region-CNN
RPN	Region Proposal Network
CNN	Convolutional Neural Network
NLP	Natural Language Processing
FC	Fully Connected
CV	Computer Vision
QKV	Query-Key-Value
TPU	Tensor Processing Unit
LBP	Local Binary Patterns
DIOU	Distance Intersection Over Union
FPGA	Field Programmable Gate Array
HLS	HTTP Live Streaming

RGB	Red-Green-Blue
BCNN	Binary Complex Neural Network
FLOPs	Floating Point Operations Per second
AP	Average Precision
KL	Kullback-Leibler Divergence
LBP	Local Binary Pattern

1 引言

2018年，世界卫生组织发布了《2018年全球道路安全状况》报告[1]，报告显示每年全球约有135万人死于道路交通事故，其中28%的交通事故由摩托车引发。特别是在一些不发达地区，由于城市基础设施结构和经济条件的限制，摩托车已经成为主要的交通工具，这些地区的道路交通死亡率大约是发达地区的3倍。比如印度、越南、印度尼西亚等国家，摩托车交通事故死亡人数分别占所有交通事故死亡人数的43%和36%。世界卫生组织指出，事故发生的主要原因是驾驶员以及行人违反交通规则。摩托车手的头部受伤是死亡的主要原因[2]，行人闯红灯也会增加事故发生率。因此，非常有必要对驾驶员及行人的违规行为进行管控，但要人为督促驾驶员与行人遵守交规，需要投入极大的警力和财力。通过监控抓拍违规车辆与行人，可以对事故多发地段进行精准管控，从而有效地降低成本和风险。

行人与摩托车是重要的交通参与者，国内外少有文献对摩托车、行人检测进行概述。本文通过综述国内外文献，总结常用的摩托车、行人检测方法与难点，前沿的改进措施，以及可行的改进方向。

2 检测特点与难点

由于摩托车头盔与行人检测不同于交通路况中的车辆检测，需要从摩托车头盔与行人的检测特点与任务出发才能更好的应对复杂环境。

2.1 检测特点

监控视频中的头盔类目标与行人目标尺度变化大，容易出现小目标。在小目标检测过程中，随着检测网络层数的增加与图片进行池化操作容易造成信息丢失。摩托车头盔与行人这类目标容易发生聚集现象，导致目标之间互相遮蔽，增大检测难度。复杂的背景，如光线、天气、噪音等不确定因素，可能将自行车、电瓶车检测为摩托车，从而出现误检。驾驶员头盔的颜色与行人的姿态多种多样，使得检测样本不均衡。

2.2 检测任务

精准捕捉到监控视频或图片信息中的驾驶员与行人的信息，并判断驾驶员和乘员是否佩戴头盔，检测任务主要分目标定位与分类：

（1）目标定位：边界框（Bounding box）在目标检测中用来找到检测物体并分类。用边界框框选出驾驶员头部与行人范围，并标记其中心位置。

（2）目标分类：定位到头部位置与行人后，需要判断出头部是否佩戴头盔，以及行人是否出现违反交通规则的行为。

3 目标检测算法

传统的机器学习算法需要手工提取特征，对操作人员的实际操作经验和理论知识都有较高要求，而且有泛化能力差的劣势。随着近年来软件、硬件和大数据的发展，深度学习算法得到了飞跃发展，深度学习算法精度、速度、泛化能力都明显优于传统的机器学习算法。因此，利用深度学习网络检测成为了目标检测发展新方向[3]。

3.1 二阶段算法

过去十几年中，目标识别算法主要是以尺度不变特征变换方法（Scale- Invariant Feature Transform, SIFT）[4]和方向梯度直方图方法（Histogram of Oriented Gradient, HOG）[5]为主流的传统机器学习算法。2012年Krizhevsky等提出的AlexNet[6]在ILSVRC目标识别大赛中获得第1名，Top5错误率达到15%。之后涌现出VGG[7]、GoogleNet[8]、ResNet[9]等网络，随后R-CNN将AlexNet在ImageNet目标识别的能力泛化到PASCAL VOC目标检测上来。表1为R-CNN[10]、Fast R-CNN[11]、Faster R-CNN[12]的网络组成。

表1 R-CNN、Fast R-CNN、Faster R-CNN网络组成

<table>
<tr><th colspan="2">R-CNN[10]</th><th>Fast R-CNN[11]</th><th>Faster R-CNN[12]</th></tr>
<tr><td>候选框</td><td>Crop/Warp</td><td>候选框</td><td rowspan="3">特征提取
ROI池化
Softmax+边框回归</td></tr>
<tr><td colspan="2">特征提取</td><td rowspan="2">特征提取
ROI池化
Softmax+边框回归</td></tr>
<tr><td>SVM分类</td><td>BBox回归</td></tr>
</table>

3.1.1 R-CNN[10]

R-CNN的流程图如图1所示，具体流程如下：

（1）图片使用选择性搜索（Selective Search，SS）生成1 000~2 000个候选区域；

（2）依次将候选区域送入深度网络提取特征；

（3）将特征送入支持向量机（Support Vector Machine，SVM）分类器，判断类别；

(4)使用回归器精修候选区域位置。

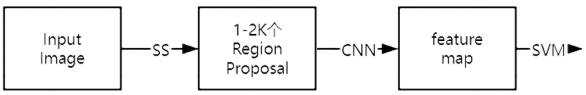

图1　R-CNN流程[10]

3.1.2　Fast R-CNN[11]

Fast R-CNN的流程图如图2所示，具体流程如下：

(1)图片通过SS生成1 000~2 000个候选区域；

(2)将图片输入深度网络获取相应的特征图，将第一步生成的候选框投影到特征图上获得特征矩阵；

(3)将每个特征矩阵通过兴趣池(ROI Pooling)缩放到统一大小的特征图(如7×7)；

(4)展平特征矩阵，通过一系列全连接层得到预测结果。

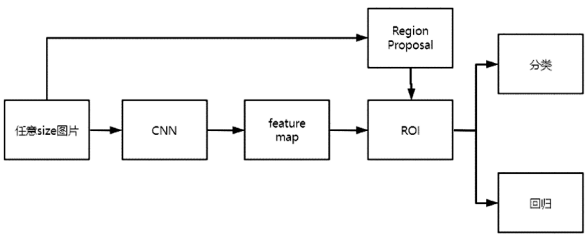

图2　Fast-RCNN流程[11]

3.1.3　Faster R-CNN[12]

Faster R-CNN的流程图如图3所示，具体流程如下：

(1)图片通过深度网络生成特征图；

(2)特征图使用区域生成网络(Region Proposal Network, RPN)生成候选框，再将候选框投影到特征图上获得特征矩阵；

(3)特征矩阵通过ROI Pooling缩放到统一大小的特征图(如7×7)；

(4)展平特征图并通过一系列全连接层获得预测结果。

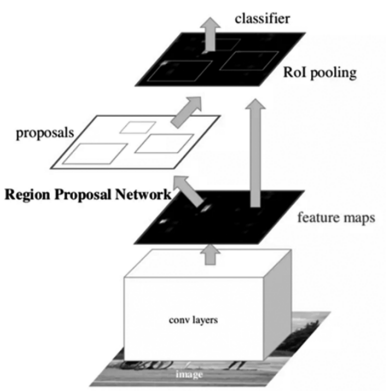

图3　Faster R-CNN流程图[12]

R-CNN先将SS网络中生成候选区域依次输入网络中，所以消耗的时间较长。R-CNN虽然将检测率从35.1%提升到53.7%，但训练速度慢。在Fast R-CNN中，图片分别通过RPN网络与特征提取网络，随后将RPN网络生成的候选区域一起映射到特征图中，速度得到了改善，但是候选区域的生成还是依靠SS，所以耗时仍然较长，每张图片耗时3 s。Faster R-CNN将原本的SS算法替代为RPN网络，速度得到了较大的提升，速度达到了10 ms但仍然达不到实时检测的要求。

3.2　一阶段算法

二阶段算法需要先通过CNN网络得到候选框，再进行分类与回归，所以检测速度较慢，Faster RCNN检测速度最快，但只能达到17帧/s，达不到实时检测的要求。一阶段算法直接对物体进行分类与预测，主要是单次多边框检测(Single Shot Multibox Detector，SSD)算法与YOLO系列算法，速度上能够达到实时检测的要求。

3.2.1　SSD算法[13]

SSD算法与YOLO算法都属于一阶段算法，SSD算法在算法推出时，其速度和精度都优于YOLO算法，相对于Faster R-CNN而言，SSD算法不需要先获取候选框再进行分类与回归，这是SSD算法在速度上优于二阶段算法的原因。SSD算法与YOLO算法主要分为3个方面：

(1)采用卷积直接做检测；

(2)SSD算法提取了不同尺度的特征图来做检测，大尺度特征图(较靠前的特征图)可以用来检测小物体，而小尺度特征图(较靠后的特征图)用来检测大物体；

(3)SSD算法采用了先验框，特征图的每个像素处都会生成不同尺寸、不同长宽比例的先验框，SSD 300中一共有5个不同尺寸的特征图，每个像素有4个先验框，经过计算一共有8 732个先验框。SSD算法算采用了密集采样方法，避免了正负样本不平衡的问题。

YOLO算法缺点是难以检测小目标，而且定位不准，SSD算法使用先验框、多尺度检测的方法在一定程度上克服这些缺点。

3.2.2　YOLO V1算法[14]

2016年，Joseph Redmon等提出了YOLO算法，YOLO算法的核心思想就是利用整张图作为网络的输入，直接在输出层回归边界框的位置及其所属的类别，将目标检测问题转换为回归问题。损失函数由坐标损失、置信度损失和类别预测损失构成。

YOLO V1算法将图片分为7×7的网格，每个网格负责预测中心点落到该网格内部的目标，这无需通过

RPN网络来获得感兴趣区域，所以YOLO算法可以获得快速的预测速度。

主要步骤如下：

（1）划分图片为$S\times S$个网格；

（2）每个网格负责B个边界框的预测，边界框由（***x, y, w, h, Confidence***）张量组成（图4）；

（3）每个网格预测一个类别信息（一共有C类）；

（4）网络输出$S\times S\times(5\times B+C)$的张量。

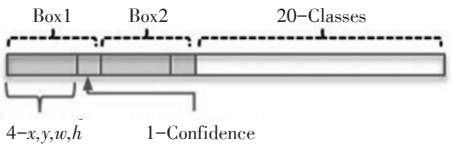

图4　YOLO V1预测张量[14]

YOLO V1算法的优点：

（1）速度快，标准版可达到45帧/s，极速版达到150帧/s，能以小于25 ms的延迟进行实时检测；

（2）在实时检测系统中，YOLO算法的实时检测精度是其它系统的2倍；

（3）迁移能力强，能够应用到其它领域。

YOLO V1算法的缺点：

（1）小目标和拥挤的目标检测（由于一个网格只能预测2个框，并且只能是同一类），当一个网格同时出现多个类时，无法检测所有类；

（2）定位误差大（损失函数影响）；

（3）对不常见的角度目标泛化性较差。

3.2.3　YOLO V2算法[15]

YOLO V2算法在速度与精度上都超过了YOLO V1算法，同时速度上也超过了SSD算法，在YOLO V1算法的基础之上进行了8项改进。

（1）批标准化（Batch Normalization）使得性能提高了2%；

（2）高分辨率分类器，图片从224×224分辨率增加到448×448的分辨率，使得mAP提高了4%；

（3）引入锚框（Anchor box）；

（4）使用聚类方法选择合适的锚框；

（5）特征融合；

（6）去掉了全连接（Fully Connected，FC）层；

（7）多尺度训练；

（8）使用了Darknet-19。

YOLO V2算法主要是提高了召回率和定位能力，借鉴Faster R-CNN的思想预测边界框的偏移，移除了全连接层，删掉了一个池化层（Pooling）使特征的分辨率更大。调整了网络的输入，使得位置坐标为奇数，这样就只有一个中心点。加上锚框能预测超过1 000个检测目标，将mAP提高了15.2%。通过表1可以看出YOLO V2算法在精度上已经超过了SSD算法，速度上已经达到了实时检测的要求。YOLO V2算法虽然在YOLO V1算法的基础上精度得到了改良，但仍存在正负样本不平衡的问题。

3.2.4　YOLO V3算法[16]

YOLO V3的主要创新点如下：

（1）骨干网络（Backbone）使用更深的网络，采用Darknet-53替换掉原来YOLO V2的Darknet-19；

（2）利用特征金字塔进行多尺度融合。选择3种不同形状的Anchor，每种Anchor有3种不同的尺度，一共使用9种不同的Anchor预测目标；

（3）分类方法上使用逻辑回归代替Softmax；使用多标签分类替换YOLO V2中的单标签分类。

3.2.5　其它YOLO算法

2020年Redom退出计算机视觉（Computer Vision，CV）领域，关于YOLO算法的研究大多数是在YOLO系列算法的基础上利用数据增强、Backbone、损失函数方法进行改进，精度和速度上都取得了进步。

Alexery Bochkovskiy在2020年推出了YOLO V4算法[17]，同年ULtralytics发布了YOLO V5算法，首次在YOLO算法中引入Focus结构。

2021年，旷视科技[18]推出了YOLOX，YOLOX-L版本，以68.9帧/s的速度在COCO数据集上实现了50.0%的mAP，比同版本的YOLO V5-L算法高出1.8%，YOLOX算法最大的改进如下：

（1）抛弃了YOLO算法以往的检测头，使用双检测头（Decoupled head），实现分离、分类与定位操作；

（2）增加无锚框（Anchor-free）机制，预测框的数量从3个降为1个，直接预测网格左上角的偏移量、预测框的宽与高。减缓了YOLO正负样本不匹配带来的影响，同时降低了参数量与GFLOPS，加快预测速度；

（3）搭配SimOTA标签分配策略。

3.3　转变器（Transformer）模型

Transformer模型[19]开始用于NLP领域，随着多模态的发展趋势，Transformer模型被用于分类、检测、分割等图像领域。Transformer模型的工作原理是利用QKV矩阵计算检测图片各像素的相关性，采用Encoder-Decoder结构，不断的实现解码与编码，结构如图5。

在CV领域中采用Transformer模型的结构主要有2种结构框架。一种是纯Transformer模型，另一种是CNN与Transformer结合的混合模型。

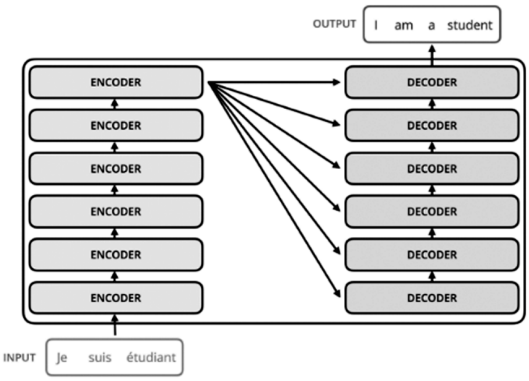

图5 Encoder-decoder结构[19]

纯Transformer模型中主要以Vit[20]模型为代表，2020年Vit推进Transformer模型在视觉领域的发展。金字塔视觉转换器（Pyramid Vision Transformer，PVT）和Swin Transformer都是建立在Vit的基础之上的。主要流程是将图片切块，通过FC将图像转换为向量，然后加上位置编码，最后输入Enconder中，通过MLP head完成分类任务。通过JFT300M的预训练，训练结果可以超过ResNet152和EfficienNet[21]，需要的TPU消耗更少，得到结果数据量越大，训练的结果越好，因为没有限制网络自身的表达能力，在数据量增大的情况下，卷积网络效果会出现瓶颈（图6）。

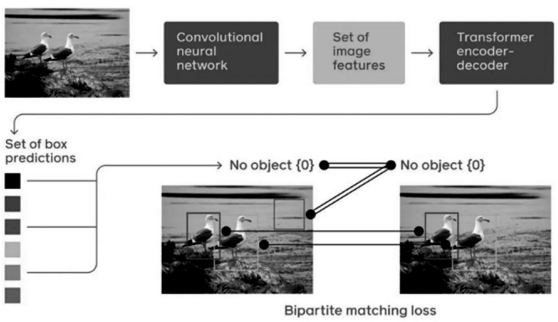

图6 Vit结构[20]

原始的Transformer模型需要计算每个像素间的相关性，计算规模与图片的大小成平方关系，此时在目标检测领域仍然是卷积神经网络处于主导地位。2021年3月，微软推出了Swin Transformer[22]，设计思想吸收了残差神经网络（ResNet），从全局到局部，极大减小了运算规模，并应用到目标检测领域，检测效果超过了Faster R-CNN、YOLO、SSD检测算法。

对Transformer模型的改进方面，主要是注意力窗口（Window attention）与滑动注意力窗口（Shift window attention）（图7）。Window attention是按照一定的尺寸将图像划分为不同的Windows，每次Transformer的Attention只在Window内部进行计算，只是使用Window attention会使得像素点的感受野（Receptive Field）受到限制；Shift window attention则是划分Window的方式，起到提高像素感受野的作用。

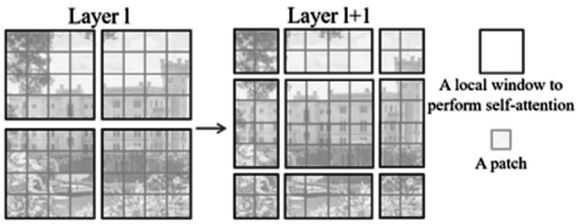

图7 Shift window attention结构[22]

表2 目标检测算法在VOC数据集的mAP与FPS

检测框	mAP/%	FPS/帧·s^{-1}
Fast R-CNN[11]	70.0	0.5
Faster R-CNN VGG16[12]	73.2	7.0
Faster R-CNN ResNet[12]	76.4	5.0
SSD 300[13]	74.3	46.0
SSD 500[13]	76.8	19.0
YOLO V1[14]	63.4	45.0
YOLO V2[15]	78.6	40.0
YOLO V3[16]	79.4	77.0

从表2中可以看出，YOLO系列算法的精度赶上了一阶段算法，速度上也是远超一阶段算法，达到了77帧/s，所以在智能驾驶行业，常用YOLO系列算法对道路安全进行实时监控。

4 基于深度学习的摩托车头盔检测研究进程与分析

4.1 摩托车头盔检测研究现状

Doungmala等[23]使用Haar特征提取器对全头盔进行检测，使用圆形霍夫变换（Hough transform）对半头盔进行检测。Dhwani等[24]采用阈值法对移动车辆进行检测。根据宽高比和面积判断是否为摩托车手，然后确定感兴趣区域用级联分类器检测相应区域。Li等[25]使用ViBe背景减法来检测移动的物体，使用梯度直方图和SVM分类器对头盔进行检测。Adam等[26]在安全帽的位置获得圆弧，利用质心位置计算圆弧参数，使用几何特征来验证该集合中是否存在安全帽。Silva等[27]根据几何形状和纹理特征的模型检测无头盔的车辆，然后使用Hough transform和SVM来检测摩托车手头部。

Shine等[28]得出CNN网络在检测速度与精度上更优于手工提取特征。Vishnu等[29]使用自适应背景减法来获得运动目标，利用CNN网络在运动对象中选择摩托车手，最后在图片局部应用CNN来进一步识别不戴头盔驾驶的摩托车手。Yogamccna等[30]利用高斯混合模型对前

景目标进行分割并标记，然后采用Faster R-CNN检测已标记的前景对象中的摩托车，利用字符序列编码的CNN模型和空间转换器识别无头盔摩托车手的编号。这类算法先使用传统机器学习算法从前景分割出摩托车，然后使用深度学习网络对头盔进行分类，与传统机器学习算法比，这类算法速度与精度更好。虽然头盔检测采用了深度学习的方法，但在摩托车检测阶段使用传统方法来获取前景目标，在拥挤的场景中效果会很差。

Kietikul等[31]使用目标滤波技术和2个CNN从监控摄像机中检测出无头盔的摩托车手。Khan[32]使用YOLO V3-tiny，在COCO数据集上预训练，然后在摩托车检测数据集进行微调，实现对摩托车驾驶员头盔佩戴情况的检测。Saumya等[33]使用YOLO V3模型来识别自行车驾驶者，二值图像的垂直投影用于计算超超载人数。Han等[34]在SSD低层特征中加入空间注意力机制，在高层特征采用通道注意力机制，最终设计了一种锚框的自适应方法。

Dailey M等[35]使用YOLO摩托车检测和头盔违章分类。Santhosh等[36]与Dasgupta等[37]分别使用SSD算法与YOLO V3算法来检测摩托车区域，最后提取图像的上部，使用分类算法来识别是否佩戴头盔。当摩托车上有多人时，分类算法容易失效。Jiawei等[38]使用改进的YOLO V5算法从视频监控中检测出摩托车，再次通过改进YOLO V5以判断是否佩戴头盔。其中YOLO V5算法设计了三通道（Tiplet）注意力机制，加强特征融合，使用缓和极大值（Soft—NMS）取代NMS，解决遮蔽问题。Boonsirisumpun等[39]使用SSD算法检测头盔。用一个CNN网络来检测摩托车和骑手的包围盒区域，同时对配戴头盔进行分类。Lin等[40-41]在泰国的7个城市收集了91 000帧摩托车安全帽佩戴视频数据集，使用RetinaNet设计新的位置编码，克服了正负样本不平衡的问题。蒋良卫等[42]使用YOLO V3算法，骨干网络（Backbone）使用ResNeXt50[43]作为新的特征提取网络并融合SPPNet，使用CIOU优化损失函数，结合DeepSort完成对安全头盔佩戴情况的实时跟踪。

4.2 行人检测研究现状

Mu等[44]基于行人特点对局部二值模式（Local Binary Pattern，LBP）的2种变化进行运用；Wang等[45]利用SVM同时融合HOG与LPB特点，解决图片遮蔽问题的缺陷；邓健锋等[46]使用FPGA采集图像信息，利用HLS对图像进行加速，然后使用CENTRIST+SVM完成定位。

Ren等[47]融合R-CNN与SSD来检测RGB图像中的人物，在光线不足的复杂背景下，mAP达到91.5%。

Wu等[48]提出了一种增强卷积神经网络（Binary Complex Neural Network，BCNN）系统，该算法提出了一种加权损失函数，该函数在训练CNN时突出了对可能目标的选定加权，增加了检测的真阳率，同时算法复杂度较低，检测时间对比同等级算法并不处于劣势。在Caltech数据集上对数平均缺失率为11.40%。

Mao等[49]针对行人检测中行人与背景区分度低的情况，改进Faster R-CNN算法，提出超参数学习（Hyper-Learner）算法。

陈一潇等[50]针对拥挤行人检测难点，使用Res2Block作为主干网络，提高网络的特征融合能力，使用坐标注意力机制（Coordinate Attention，CA）增加目标定位能力，最后在行人检测任务中，速度达到了51帧/s，精度提高了3.75%。

4.3 研究改进方向分析

对摩托车的头盔与行人检测需要实时性能，所以一阶段目标检测成为了首选，在检测上主要会遇到以下难点，即遮蔽、误判、小目标和轻量化网络所带来的精度减少的问题。一般从数据、主干网络、注意力机制、损失函数和轻量化出发解决以上的检测问题。

4.3.1 数据

数据增强可以防止训练过程中模型过拟合，以及降低标注成本与扩充数据集。为了防止CNN过于关注一个小的中间激活集或输入图像上的一个小区域，提出了随机特征去除正则化。在随机失活（Dropout）用于随机删除隐藏激活，用于删除输入上的随机区域。特征去除策略通过让模型不仅关注对象中最具辨识性的部分，而且关注整个对象区域，从而提高泛化和定位能力。

YOLO V4、YOLO V5、YOLOX算法没有在YOLO V3的结构上进行明显改进，而是使用各种数据增强方法在数据层面使得模型的精度与速度得到提升（如Focs，Cutmix，Cutout，Mosaic等）。

关于使用位置编码，Chen等[51]提出一种将骑车人数量、位置和头盔使用情况数据进行集成的新方法。利用一种新的编码方案，将摩托车头盔使用分类问题由多类分类问题转化为多类二值分类问题的组合（图8）。

4.3.2 主干网络

优化网络结构方面，多尺度特征融合通过融合低层位置信息和高层分类信息，可以明显提高检测精度。张永亮等[52]在YOLO V4算法中采用设计多尺寸特征，提取模块和增强特征融合模块，改进算法同时提升了定位和分类交通标志的能力。金雨芳等[53]在

YOLO V4的基础上，增加128×128特征图的输出，基于密集连接的思想，将特征提取网络输入的4个不同尺度的特征层进行重新连接，平均精度高达91.17%，相比原网络检测精度提高了2.96%，检测速度达53帧/s。

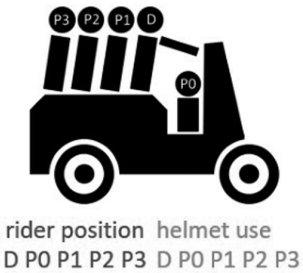

图8　位置编码[53]

Wang等[54]在YOLO V4中采用像素洗牌法代替插值采样法，解决了低分辨率特征图与高分辨率特征图融合时信息丢失的问题。引入Ghost Net和Squeeze Net减少了网络的时延和网络参数。头盔类和无头盔类的mAP增加了约5%，相应层的参数和浮点数（FLOPs）都减少了50%，模型尺寸减少了20%。

4.3.3　注意力机制

通道注意力机制（Squeeze-and-Excitation Networks, SE）使用2维全局池化计算通道注意力，以相当低的成本提升了性能，基于卷积核的注意力机制（Selective Kernel Networks, SK）则是在SE的基础上增加不同尺寸的卷积核，完成对不同尺寸目标的注意，SE与SK只考虑编码通道间信息而忽略位置信息的重要性。通道空间注意力机制（Channel Block Attention Model, CBAM）使用空间注意力完成了对位置信息的利用，而卷积只能获得局部的位置关系。坐标注意力机制（Coordinate Attention, CA）使用2种转化分别沿着2个空间方向聚合，一个保存沿着一个空间方向的长期依赖关系，并保存另一个方向的精确位置关系。

赵梦等[55]针对模糊情况、失真场景使用了SK结合YOLO V5算法，有效的解决了水下模糊图片的识别问题。

王玲敏等[56]在YOLO V5的主干网络中加入了CA，使网络可以在更大的感受野中进行注意，使用加权双向特征金字塔网络加权双向特征网络（BiFPN）网络替代原有的特征融合层，实现高效的双向尺度连接与特征融合，在自制安全帽数据集上进行了测试，平均精度提高了5%。

王静等[57]针对白细胞数据样本量上、类间差别小以及目标尺寸小的问题，在YOLO V5的主干网络中添加坐标注意力机制CA，同时使用四尺度特征检测来提高小目标的识别精度，mAP提高了3.8%。

Zhang等[58]在YOLO V5算法的骨干网络中加入深度方式坐标注意力机制（Depthwise Coordinate Attention, DWCA）和K-Means++，使网络可以独立学习各个通道的权重，增加区分前景、背景的能力，提高了先验锚框的匹配度，比YOLO V5头盔检测的平均精度提高了3%。

4.3.4　损失函数

对于一阶段检测模型的损失函数，定位信息与分类信息适合不同的卷积参数，如YOLOX的双检测头。

交并比（IOU）作为损失函数时，如果2个物体不重叠，不会反映2个形状之间的距离，此时梯度为0，无法进行梯度回传。广义交并比（GIOU）则是采用IOU减去2框交集与并集的比，解决了损失（Loss）为零时梯度不回传的问题，但出现边界框覆盖真实框（Ground truth）时，只要边界框与真实框的面积不变，不管2个框的位置在哪里，IOU Loss与GIOU Loss都是一个定值，因此可以忽略2个框的中心位置。距离交并比（DIOU）则是引入了2个框中心的距离作为惩罚项来限制中心点的位置。考虑到边界框回归3要素中的长宽比还没被考虑到计算中，进一步在DIOU的基础上提出了完整交并比完整交并比（CIOU）。

表3　IOU、GIOU、DIOU、CIOU在COCO数据集上的表现

Loss/Evaluation	mAP50/%	mAP75/%
IOU	37.9	40.8
GIOU	38.0	41.1
DIOU	38.1	41.1
CIOU	38.7	42.0

于娟等[59]针对无人机图像中的小目标与遮蔽目标，在YOLO V5的基础上使用相对熵（Kullback-Leibler, KL）散度损失函数替代原有的置信度损失的交叉熵函数，提高模型的泛化能力。研究人员将残差模块替换为沙漏（LSandGlass）模块，减少低分辨率的语义损失。

Wu等[60]采用多尺度骨干网络增强结构（Im-Res2Net），利用增强的EN-PAN进行特征融合，并使用焦点损失（Focal loss）缓解了正负本不平衡的状况，在IOU等于0.5的水平上获得92.8%的高准确率。徐守坤等[61]通过对Faster-RCNN运用多尺度训练和增加锚点数量，提高网络检测不同目标的鲁棒性，使用在线困难样本的挖掘策略，解决了正负样本的不均衡问题，提高了对安全帽检测精度，比原始的Faster R-CNN准确率提高了7%。Zhang等[62]在YOLO V3的基础上使用Anchor free和Focal loss方法避免正负样本的不平衡带来的问题，同时采用GIOU作为损失函数，改善了均方误差。

4.3.5 轻量化

实现网络轻量化使网络参数、模型结构变小，不可避免的牺牲检测精度。

经典目标检测依赖于CNN作为主干进行特征提取。一些优秀的基础网络，如VGGNet、ResNet等，虽然能很好地提取特征，但是这些网络的计算量很大，特别是用于嵌入式设备时，依靠这些网络设计的物体检测模型很难达到实时性的要求。因此要对模型进行加速，一方面对训练好的复杂模型进行压缩得到小模型，另一方面直接设计小模型并进行训练，如Mobilenet、Shufflenet、Mixnet等。

在摩托车头盔检测领域中，主要分为传统的检测算法、混合算法、深度学习算法。传统的检测算法通过头盔的外形轮廓进行目标定位与分类，忽略了更深的语义信息导致检测效果偏差。混合算法利用传统算法对目标进行摩托车与背景分类，再通过深度学习算法对检测出的摩托车进行头盔分类，混合算法在检测的第一步依赖传统算法，难以在复杂工况下检测出摩托车，导致混合算法的精度较低。深度学习算法直接通过检测网络从图片中检测出驾驶员与乘客的头部，可以成功地从复杂工况下检测出驾驶员是否佩戴头盔。近年来深度学习算法成为了摩托车头盔检测领域的主流算法。

在行人检测领域中，传统的检测算法大多数建立在二值法的基础上，检测中图片丢失了颜色信息，不利于模型分类，同时传统算法对遮蔽情况的处理也是不利的；使用深度学习算法，解决了传统算法常见的色彩丢失问题，精度上超过了传统的检测算法，所以深度学习算法是行人检测算法的发展趋势。

近年来深度学习算法成为摩托车头盔检测与行人检测的主流算法，出现了大量针对模型改进的方法。主要从数据、主干网络、注意力机制、损失函数轻量化5个角度出发。数据主要使用各种的数据增强方法，近年来出现的位置编码成功解决了样本的不平衡问题。骨干网络主要从位置与分类信息出发，融合低层位置信息与高层分类信息。注意力机制从通道、空间、位置之间的关系出发，增强模型对目标的敏感程度。损失函数从预测框的角度出发，文章研究趋势倾向于无锚框方式。轻量化趋向使用轻量化网络替换检测网络中的骨干网络。

5 总结与展望

本文首先对摩托车头盔与行人的检测特点与任务进行总结。其次针对常见的目标检测模型进行分析，二阶段算法精度高但是速度慢，满足不了实时检测的要求，一阶段算法在精度上已经满足了检测精度，速度也可以满足实时检测的要求。然后总结国内外具有代表性的摩托车与行人检测相关文献，深度学习算法在速度与精度上都远超传统算法，更适合作为检测模型。最后从数据、主干网络、注意力机制、损失函数、轻量化5个角度出发，总结最新的检测模型改进方向。

学者可以通过阅读本文对摩托车违规检测与行人检测进行了解，通过本文总结的改进方向，得到高精度、高速度、泛化能力强、小参数的检测模型，应用于交通的监管中，提高检测效率，降低事故率。

参 考 文 献

[1] LIU B C, IVERS R, NORTON R, et al. Helmets for preventing injury in motorcycle riders[J]. Cochrane Database of Systematic Reviews, 2008, 1(2): CD004333.

[2] 俞春俊, 王长君. 摩托车头盔与摩托车交通事故的相关研究[J].中国安全生产科学技术, 2009, 5(2): 76-80.

[3] 赵永强, 饶元, 董世鹏, 等. 深度学习目标检测方法综述[J].中国图象图形学报, 2020, 25(4): 629-654.

[4] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91-110.

[5] WATANABE T, ITO S, YOKOI K. Co-occurrence Histograms of Oriented Gradients for Human Detection[J]. IPSJ Transactions on Computer Vision and Applications, 2010(2): 39-47.

[6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E.Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[7] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C]// International Conference on Learning Representations (ICLR), 2015.

[8] 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.

[9] 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.

[10] 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.

[11] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE international conference on computer vision, 2015: 1440-1448.

[12] 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 & Machine Intelligence, 2017, 39(6):1137-1149.

[13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.

[14] 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.

[15] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 7263-7271.

[16] REDMON J, FARHADI A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[17] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[18] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.

[19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017(11): 6000-6010.

[20] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.

[21] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//International conference on machine learning, PMLR, 2019: 6105-6114.

[22] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.

[23] DOUNGMALA P, KLUBSUWAN K. Helmet Wearing Detection in Thailand Using Haar Like Feature and Circle Hough Transform on Image Processing[C]//2016 IEEE International Conference on Computer and Information Technology (CIT), Nadi, 2016: 611-614.

[24] CONTRACTORRD, PATHAK K, SHARMA S, et al.Cascade Classifier based Helmet Detection using OpenCV in Image Processing[C]//National Conference on Recent Trends in Computer and Communication Technology (RTCCT 2016), 2016: 195-200.

[25] JIE L, LIU H, WANG T, et al. Safety helmet wearing detection based on image processing and machine learning[C]//2017 Ninth International Conference on Advanced Computational Intelligence (ICACI), 2017: 201205.

[26] ADAM A, RIVLIN E, SHIMSHONI I, et al. Robust real-time unusual event detection using multiple fixed-location monitors[J]. IEEE transactions on pattern analysis and machine intelligence, 2008, 30(3): 555-560.

[27] WEN C Y, CHIU S H, LIAW J J, et al. The safety helmet detection for ATM's surveillance system via the modified Hough transform[C]//IEEE 37th Annual 2003 International Carnahan Conference on Security Technology, 2003. Proceedings, IEEE, 2003: 364-369.

[28] SHINE L, JIJI C V. Automated detection of helmet on motorcyclists from traffic surveillance videos: a comparative analysis using hand-crafted features and CNN[J]. Multimedia Tools and Applications, 2020(79): 14179-14199.

[29] VISHNU C, SINGH D, MOHAN C K, et al. Detection of motorcyclists without helmet in videos using convolutional neural network[C]//2017 International Joint Conference on Neural Networks (IJCNN), IEEE, 2017: 3036-3041.

[30] YOGAMEENA B, MENAKA K, PERUMAAL S S. Deep learning-based helmet wear analysis of a motorcycle rider for intelligent surveillance system[J]. IET Intelligent Transport Systems, 2019, 13(7): 1190-1198.

[31] JEARANAITANAKIJ K, IAMTHAMMARAK K, WANGC HAROEN N. Fast Classifying Non-helmeted Motorcyclists by Using Convolutional Neural Networks[J]. SNRU Journal of Science and Technology, 2021, 13(1): 1-10.

[32] KHAN F A, NAGORI N, NAIK A. Helmet and Number Plate detection of Motorcyclists using Deep Learning and Advanced Machine Vision Techniques[C]//2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020: 714-717.

[33] 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.

[34] HAN G, ZHU M, ZHAO X, et al. Method based on the cross-layer attention mechanism and multiscale perception for safety helmet-wearing detection[J]. Computers & Electrical Engineering, 2021, 95(10): 107458.

[35] CHAIRAT A, DAILEY M, LIMSOONTHRAKUL S, et al. Low cost, high performance automatic motorcycle helmet violation detection[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020: 3560-3568.

[36] SANTHOSH A, AUGUSTHY J, SUNIL L, et al. Real-Time Helmet Detection of Motorcyclists without Helmet using Convolutional Neural Network[J]. International Journal for Research in Applied Science and Engineering Technology IJRASET, 2020, 8(7): 1112-1116.

[37] DASGUPTA M, BANDYOPADHYAY O, CHATTERJI S. Automated helmet detection for multiple motorcycle riders using CNN[C]//2019 IEEE Conference on Information and Communication Technology, IEEE, 2019: 1-4.

[38] JIA W, XU S, LIANG Z, et al. Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector[J]. IET Image Processing, 2021, 15(14): 3623-3637.

[39] BOONSIRISUMPUN N, PUARUNGROJ W, WAIROTCHANAPHUTTHA P. Automatic detector for bikers with no helmet using deep learning[C]//2018 22nd International Computer Science and Engineering Conference (ICSEC), IEEE, 2018: 1-4.

[40] SIEBERT F W, LIN H. Detecting motorcycle helmet use with deep learning[J]. Accident Analysis & Prevention, 2020, 134(1): 105319.

[41] LIN H, DENG J D, ALBERS D, et al. Helmet Use Detection of Tracked Motorcycles Using CNN- Based Multi-Task Learning[J]. IEEE Access, 2020(8): 162073-162084.

[42] 黄林泉, 蒋良卫, 高晓峰. 改进YOLOv3的实时性视频安全帽佩戴检测算法[J]. 现代计算机, 2020(30): 32-38+43.

[43] XIE S, GIRSHICK R, DOLLÁR, PIOTR, et al. Aggregated Residual Transformations for Deep Neural Networks[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2016.

[44] MU Y, YAN S, LIU Y, et al. Discriminative local binary patterns for human detection in personal album[C]//2008 IEEE conference on computer vision and pattern recognition, IEEE, 2008: 1-8.

[45] 陈锐, 王敏, 陈肖. 基于PCA降维的HOG与LBP融合的行人检测[J]. 信息技术, 2015, 39(2): 101-105.

[46] 邓健峰, 谢云. 车载环境下实时行人检测系统研究[J]. 自动化仪表, 2018, 39(11): 78-82.

[47] 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 & Machine Intelligence, 2017, 39(6):1137-1149.

[48] BO W, NEVATIA R. Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors[C]//Tenth IEEE International Conference on Computer Vision, IEEE, 2005.

[49] MAO J, XIAO T, JIANG Y, et al. What Can Help Pedestrian Detection?[C]// Computer Vision & Pattern Recognition, IEEE, 2017.

[50] 陈一潇, 阿里甫·库尔班, 林文龙, 等. 面向拥挤行人检测的 CA- YOLOv5[J]. 计算机工程与应用, 2022, 58(9): 239-245.

[51] LIN H, CHEN G, SIEBERT F W. Positional Encoding: Improving Class-Imbalanced Motorcycle Helmet use Classification[C]//2021 IEEE International Conference on Image Processing (ICIP), 2021: 1194-1198.

[52] 张永亮, 陆阳, 朱芜强, 等. 基于多尺度特征提取和特征融合的交通标志检测[J]. 计算机工程, 2021(12): 1-11.

[53] 金雨芳, 吴祥, 董辉, 等. 基于改进YOLO v4的安全帽佩戴检测算法[J]. 计算机科学, 2021, 48(11): 268-275.

[54] WANG H, GE H, LI M. PFG-YOLO: A Safety Helmet Detection based on YOLOv4[C]//2021 IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), IEEE, 2021(5): 1242-1246.

[55] 赵梦, 于红, 李海清, 等. 融合SKNet与YOLOv5深度学习的养殖鱼群检测[J]. 大连海洋大学学报, 2022(2): 312-319.

[56] 王玲敏, 段军, 辛立伟. 引入注意力机制的YOLOv5安全帽佩戴检测方法[J]. 计算机工程与应用, 2022(9): 303-312.

[57] 王静, 孙紫雲, 郭苹, 等. 改进YOLOv5的白细胞检测算法[J].计算机工程与应用, 2022, 58(4): 134-142.

[58] JIN Z, QU P, SUN C, et al. DWCA-YOLOv5: An Improve Single Shot Detector for Safety Helmet Detection[J]. Journal of Sensors, 2021, 2021: 4746516.

[59] 于娟, 罗舜. 基于YOLOv5的违章建筑检测方法[J].计算机工程与应用, 2021, 57(20): 236-244.

[60] WU P, LI H, ZENG N, et al. FMD-Yolo: An efficient face mask detection method for COVID-19 prevention and control in public[J]. Image and Vision Computing, 2021,117: 104341.

[61] 徐守坤, 王雅如, 顾玉宛, 等. 基于改进Faster RCNN的安全帽佩戴检测研究[J]. 计算机应用研究, 2020, 37(3): 901-905.

[62] ZHANG L, LI Y, CHEN H, et al. Anchor-free YOLOv3 for mass detection in mammogram[J].Expert Systems with Applications, 2022, 191: 116273.

【作者简介】

陈俊豪(1998-),重庆交通大学硕士研究生,主要从事深度学习和图像识别研究。

E-mail:1471272619@qq.com