基于 BFDS 改进的 YOLOv8 疲劳驾驶检测算法

郑凯东[1] 舒 心[1]
ZHENG Kaidong SHU xin

摘 要 近年来随着国内私家车保有量的逐年增多，我国的交通事故的发生率也是逐年升高，其中由于司机的疲劳驾驶导致的交通事故占比最多，针对司机在开车中的驾驶安全问题，提出了一种疲劳检测算法 BFDS-YOLO，融合了 DCNv2 与 BiFPN，并引入了视觉通道注意力机制，以提高疲劳检测的效率。引入了 Yolov8n-DCNv2、YOLOv8n-BF 等变种模型，通过详细实验比较它们的性能。提出的 BFDS-YOLO 模型对主干网络和颈部网络进行了改进，将 DCNv2 与 BiFPN 结合，并引入了视觉通道注意力机制。实验结果表明，BFDS-YOLO 相较于其他模型在多个性能指标上均有显著提升，提出的算法在疲劳检测任务中具有广泛的应用前景。

关键词 深度学习；疲劳检测；可变形卷积；视觉通道注意力机制；加权双向特征金字塔网络

doi：10.3969/j.issn.1672-9528.2024.02.048

0 引言

在近年来的研究中，疲劳驾驶检测系统取得了显著的进展。研究者们不断改进算法，结合深度学习的发展，提高了疲劳驾驶检测系统的准确性和实时性。在疲劳驾驶检测方法方面，主要分为两类方法：主动检测法、被动检测法。主动检测法包含但不限于：填写问卷、口头问答判断、主动检测模型。被动检测法包括：基于车辆行为[1]、基于司机行为[2]、基于生理信号[3]、基于面部特征[4]以及基于多特征融合的疲劳检测[5]。

目标检测在深度学习的推动下取得了革命性进展。基于深度学习的目标检测方法通常可以分为单阶段和两阶段两种类型以及卷积神经网络（CNN）方法。CNN 方法[6]在图像特征提取方面提取出色彩和纹理信息等更高级别的抽象特征，从而能够更好地捕捉目标，通常具有较高的准确性和鲁棒性，适用于各种目标检测任务。而 CNN 方法需要大量的标注数据来训练，对于小尺寸目标和遮挡目标，CNN 方法的性能可能会下降。

两阶段目标检测算法通常分为两个阶段：首先生成一些候选区域，然后对这些区域进行分类和精确定位。代表性的两阶段目标检测算法有 Faster R-CNN[7] 和 Mask R-CNN[8]。两阶段目标检测算法的优点为准确度高、多任务处理，一些两阶段算法如 Mask R-CNN 可以同时完成目标检测和语义分割等多个任务。两阶段目标检测算法的缺点为速度相对较慢，不太适用于实时应用，复杂度高、网络结构较为复杂、训练和部署相对困难。

单阶段目标检测算法通常将目标检测任务视为一个回归问题，直接从输入图像中预测目标的类别和位置。YOLO（you only look once）[9] 和 SSD（single shot multibox detector）[10] 是代表性的单阶段目标检测算法。YOLO 算法将目标检测问题转化为一个回归问题，通过将图像划分为网格单元并在每个单元中预测边界框的类别和位置。单阶段目标检测算法的优点为速度快，适用于实时应用，算法结构相对简单，易于理解和实现。单阶段目标检测算法的缺点为小目标检测效果差、定位精度较低、定位精度通常不如两阶段算法高。

鉴于疲劳驾驶检测的应用场景所需要的快速、实时运行的必须要求，本文采用 YOLOv8n 作为疲劳检测算法，并且为了进一步增加检测的准确度，本文融合了 DCNv2（deformable convolutional networks version 2）[11] 与 BiFPN（bi-directional feature pyramid network）[12] 等先进技术，提出了一个新型的 BFDS-YOLO 模型。

1 基于 BFDS 改进的 yolov8n 目标检测算法

本文针对最新的目标检测模型 YOLOv8 中的轻量化模型 YOLOv8n 融合先进的卷积模块、注意力机制以及加权双向特征拼接操作，相对于原始模型在轻量化的同时有着更高的精确度、更好的评价指标，命名为 BFDS-YOLO。图 1 展示了 BFDS-YOLO 的模型架构，其中的红色虚线所框出来的为改进部分。BFDS-YOLO 模型的改进主要在以下三个方面体现。首先，改变原始的 C2f 模块，融合 DCNv2 变为 C2f_DCN 模块，以增强骨干网络的特征提取能力，更好地提取到图片的复杂

1. 西安石油大学 陕西西安 710065

细节并处理一些形状变化的检测目标。其次，通过引入 BiFPN，增强了颈部网络的特征融合能力。最后，在颈部网络中的原始 C2f 之后添加视觉通道注意力机制 SEAttention[13]，使得模型更加关注学习到的重要特征，提升模型的鲁棒性。

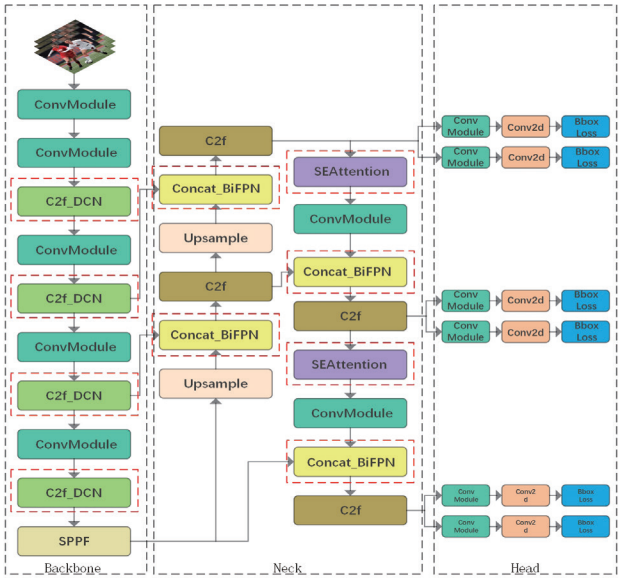

图 1 BFDS-YOLO 模型结构图

1.1 C2f_DCN 模块设计

疲劳驾驶检测中包含一些驾驶员的姿态的变化，例如东张西望、接听电话等动作。因此，在模型对目标进行特征的提取阶段，仅采用标准卷积（standard convolution）可能未能有效捕捉目标的细粒度细节。标准卷积无法适应目标的形变，无法自适应地调整卷积核的采样位置，导致目标的定位准确性下降，容易产生误检。

为有效地减少上述的几个问题，在特征提取中采用了 DCNv2 可变形卷积，着重解决的问题为常规的图像增强，仿射变换（线性变换加平移）中不能处理的，多种形式目标变化的几何变换问题。正常卷积计算和可变形卷积采样的方式不同，如图 2 所示。

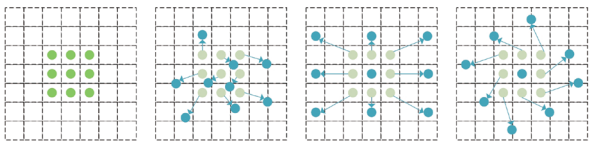

图 2 卷积核大小为 3*3 的正常卷积和可变形卷积的采样方式

以 3*3 大小的卷积核为例，每个卷积核的输出结果 $y(p_0)$ 都为 x 上以 $x(p_0)$ 为中心点的九宫格上所有参数的计算结果，(-1,-1) 代表 $x(p_0)$ 的左上角，(1,1) 代表 $x(p_0)$ 的右下角，其他类似，计算公式为：

$$R = \{(-1,-1),(-1,0),...,(0,1),(1,1)\} \quad (1)$$

$$y(p_0) = \sum_{p_n} w(p_n) \cdot x(p_0 + p_n) \quad (2)$$

可变形卷积（deformable conv）操作在卷积计算方法的功能范围上，增加了一种可学习的参数 Δp_n。虽然该计算方法对于每个输出的 $y(p_0)$ 仍然是从 x 上的中心位置 $x(p_0)$ 的四周不同的 9 个位置进行采样计算，但是增加了一个偏移量 Δp_n，则采样范围变成了不规则范围，计算公式为：

$$y(p_0) = \sum_{p_n} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \quad (3)$$

DCNv2 的核心创新在于引入了可变形卷积层，通过自适应地调整卷积核的形状和采样位置，能够灵活地适应复杂的目标形变，提高了模型对图像细节的感知能力。DCNv2 网络结构图如图 3 所示。

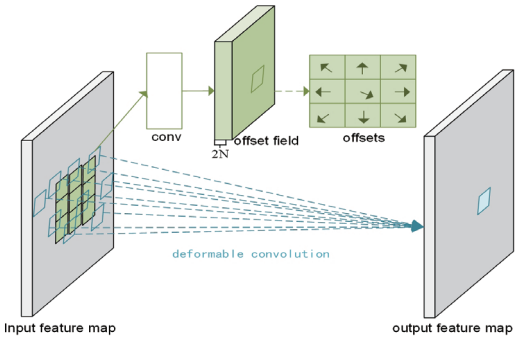

图 3 DCNv2 网络结构图

对加入 DCNv2 的模型，其计算时间会有明显增加，其原因为 DCNv2 中对可变形卷积的频繁使用。为了减少其计算量并且保持原始 YOLO 模型的轻量化，本文只对原始网络中 backbone 的 C2f 模块中的 Bottleneck 增加了 DCNv2 改变为 bottleneck_DCN 模块，如图 4 所示。

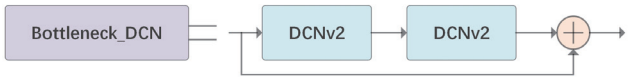

图 4 可变形卷积替换普通卷积后的 Bottleneck_DCN 模块

1.2 Concat_BiFPN 模块设计

BiFPN 是一种适用于目标检测的特征金字塔网络，基于特征金字塔网络，结合了双向路径和跨层特征聚合的思想，在提高特征表征能力的同时，实现了底层与高层之间特征信息的相互传递，有效地解决了特征金字塔网络中存在的信息丢失和冗余问题。

BiFPN 采用了一种跨层特征聚合的策略，将相邻层的特征进行加权融合，在牺牲小幅检测速度的前提下使模型能够对大小不同的特征进行融合，保留了不同尺度的特征信息，增强了网络的表征能力。BiFPN 网络结构如图 5 所示。本文借鉴了 BiFPN 的结构，将其融合进了原始模型的 neck 网络中的 Concat 变为 Concat_BiFPN 模块。

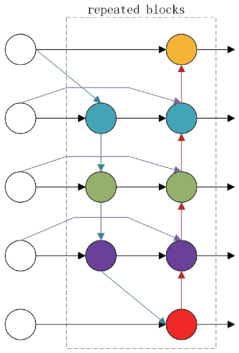

图 5 BiFPN 网络结构

1.3 融合 SEAttention 的 neck 设计

SEAttention 模块结合了空间激活和通道激活的双重注意机制，以增强模型对输入特征的表征能力。将全局信息与局部信息相结合，通过学习通道之间的关系动态调整通道权重，从而选择性地强化有用的特征信息。SEAttention 引入了两个子模块：Squeeze 操作和 Excitation 操作，前者用于压缩通道维度，后者则用于自适应地赋予通道不同的权重。SE-Block 模型如图 6 所示。

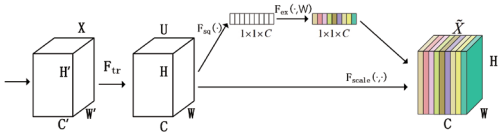

图 6 SE-Block 模型结构

本文在借鉴 SEAttention 后将 SEAttention 模块融合进原始的 YOLO 模型的 neck 网络中的 C2f 之后，该改进版 neck 模块能够自适应地学习特征图的权重，从而更好地捕捉到不同空间位置和通道之间的信息关系。

2 实验结果分析

2.1 所用数据集

本文实验所用数据集为 WiderFace 数据集和自制公交车司机驾驶疲劳状态数据集，如图 7 所示。

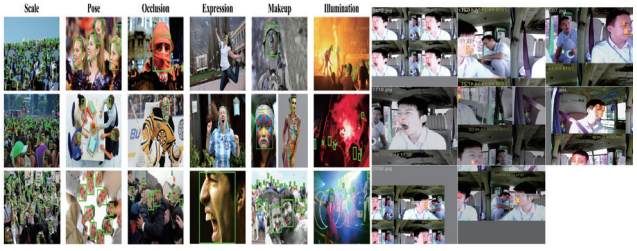

图 7 WiderFace 与公交司机数据集展示

2.2 实验过程

在 WiderFace 数据集上对面部检测进行训练，得到稳定准确的面部检测模型，然后将模型放在公交司机疲劳状态数据集上进行验证对比。

原始 YOLOv8n 和 BFDS-YOLO 的训练过程使用相同的数据集和参数，如图 8 所示，经过 30 轮迭代，几种模型的损失值下降速度开始趋于平缓，80 轮迭代后几种模型损失值误差趋于稳定。实验结果表明，BFDS-YOLO 和原模型相比，收敛速度更快，损失值更小，网络收敛能力明显提高。

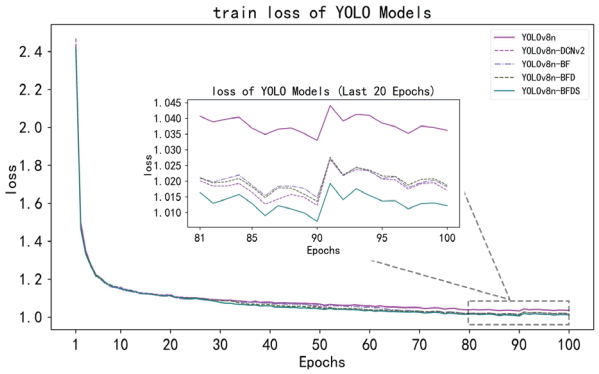

图 8 模型训练损失曲线

2.3 评价标准

在计算机视觉领域，模型的性能通常使用多个评估指标来全面衡量。其中，mAP50（mean average precision at IoU=0.5），mAP50-95，*precision*（精确率）和 *recall*（召回率）是几种常用的指标，这些评估指标能够帮助研究人员全面了解模型的性能，从而更好地选择适用于特定任务的模型，并进行进一步的性能优化，本文即采用上述四个指标作为模型的评价指标。*AP*、*mAP*、*precision* 和 *recall* 的定义式为：

$$AP=\int_0^1 P(r)\mathrm{d}r \tag{4}$$

$$mAP=\frac{\sum_{i=1}^{k} AP_i}{k} \tag{5}$$

$$precision=\frac{True\ Positive}{True\ Positive+False\ Positive} \tag{6}$$

$$recall=\frac{True\ Positive}{True\ Positive+False\ Negative} \tag{7}$$

式中：*AP* 为平均精度，使用积分的方式来计算 PR 曲线与坐标轴围成的面积。

2.4 实验结果和分析

消融实验的目的是对每个模型的改进版本进行测评其效果的变化。第一个模型为 YOLOv8n 本身的模型，不加改动在相同的参数与数据集上训练 100 轮后的最佳数据。实验结果如表 1 所示，最终所有指标对于原始的 YOLOv8n 模型在不同程度上都有所提升。mAP50、mAP50-95、recall、precision 增长了 1.23%、0.82%、0.9%、0.88%。

表 1 消融实验中每个模型的评估指标

模型	BiFPN	DCNv2	SE	mAP50 /%	mAP50-95 /%	Recall /%	Precision /%
YOLOv8n	×	×	×	64.72	35.26	56.92	83.67
YOLOv8n-BF	√	×	×	65.10	35.51	56.90	84.43
YOLOv8n-DCNv2	×	√	×	65.61	35.97	57.77	84.35
YOLOv8n-BFD	√	√	×	65.93	35.97	57.99	84.16
BFDS-YOLO	√	√	√	65.95	36.08	57.82	84.55

表 2 为 BFDS-YOLO 模型与原始 YOLOv8n 模型在自制公交车司机数据集上进行 50 轮实验的对比分析。

表 2 公交司机数据集上模型的评估指标

模型	mAP50/%	mAP50-95/%	Recall/%	Precision/%
YOLOv8n	98.98	57.50	97.49	97.26
BFDS-YOLO	98.98	59.44	97.52	97.35

从消融实验的结果来看，每个模型的改进方法都或多或少相对原始模型的精度有所提升，由此可见模型之间并不存在太大的冲突，以至最终的模型相比原始模型有着明显的提升。

3 结语

本文提出了一种创新的疲劳检测算法，名为 BFDS-YOLO，该算法融合了 DCNv2 与 BiFPN 并引入了视觉通道注意力机制 SEAttention，以提高检测的性能。

通过以上的改进和融合，BFDS-YOLO 在目标检测方面比原始模型有明显的提高。在 WiderFace 数据集上 mAP50 上升了 1.23%，mAP50-95 上升了 0.82%，*recall* 上升了 0.9%，*precision* 上升了 0.88%。当然，本模型主要为轻量化模型结构，在车内驾驶场景中场景较为简单，快速实时运行则成为首要考虑的因素，故不可避免地放弃了高精度的复杂模型，转而选择了检测更快速的轻量级模型。未来的研究可以选择模型与嵌入式设备相融合的方向，在可移植设备上实施进一步的研究。

参考文献：

[1] CHEN L W,CHEN H M.Driver behavior monitoring and warning with dangerous driving detection based on the internet of vehicles[J].IEEE transactions on intelligent transportation systems, 2020,22(11):7232-7241.

[2] LI R,CHEN Y V,ZHANG L. A method for fatigue detection based on driver's steering wheel grip[J].International journal of industrial ergonomics, 2021,82:83-103.

[3] FAN Y,GU F,WANG J,et al.SafeDriving：an effective abnormal driving behavior detection system based on EMG signals[J]. IEEE internet of things journal, 2021,9(14): 12338-12350.

[4] LIU Z, PENG Y, HU W. Driver fatigue detection based on deeply-learned facial expression representation[J].Journal of visual communication and image representation, 2020,71:102-123.

[5] VARDHAN V V S, REDDY N R K, SURYA K J, et al. Driver's drowsiness detection based on facial multifeature fusion[J].Journal of physics：conference series, 2021, 1998(1): 12-34.

[6] KRIZHEVSKY A , SUTSKEVER I , HINTON G .Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2017, 60(6): 84-90.

[7] REN S, HE K, GIRSHICK R, et al.Faster R-CNN: towards real-time object detection with region proposal networks[J]. In IEEE transactions on pattern analysis and machine intelligence, 2017,39(6):1137-1149.

[8] HE K, GKIOXARI G,DOLLÁR P,et al. Mask R-CNN[C]// 2017 IEEE international conference on computer vision (ICCV). Piscataway: IEEE, 2017: 2980-2988.

[9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Computer Vision & Pattern Recognition. Piscataway:IEEE,2016:779-788.

[10] WEI L , DRAGOMIR A , DUMITRU E ,et al.SSD:single shot multibox detector[EB/OL].(2015-12-08)[2023-10-05]. https://arxiv.org/abs/1512.02325.

[11] ZHU X,HU H.,LIN S,et al. Deformable ConvNets V2: more deformable, better results[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019:9300-9308.

[12] TAN M,PANG R,LE Q V. Efficientdet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).Piscataway: IEEE, 2020:10778-10787.

[13] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 7132-7141.

【作者简介】

郑凯东（1964—），男，广东汕头人，副教授，硕士生导师，研究方向：图形学与虚拟现实、深度学习与计算机视觉、程序设计。

舒心（1999—），男，江苏徐州人，硕士研究生，研究方向：计算机视觉、深度学习。

（收稿日期：2023-12-04）