基于多特征信息融合的疲劳驾驶检测算法

张兴旺[1,2],王凤随[1,2],杨海燕[1,2]
1. 安徽工程大学 电气工程学院,安徽 芜湖 241000
2. 高端装备先进感知与智能控制教育部重点实验室,安徽 芜湖 241000

摘 要:目的 针对目前驾驶员疲劳驾驶检测不能兼顾检测速度与检测准确率的问题,提出一种基于 YOLOv7 改进模型的疲劳驾驶检测算法。**方法** 首先,为提高模型的收敛速度,增强模型的检测性能,算法将传统的卷积层替换为深度过度参数化卷积层,通过增加可学习的参数加快拟合过程;其次,针对存在遮挡目标的场景,传统下采样过程容易导致特征丢失严重,为提高对遮挡目标检测的准确性,算法引入了基于 SE 注意力改进的 DS-Conv 模块;再次,为了提高模型对小目标的检测能力,算法在特征提取层中添加了多尺度特征提取 MSS 注意力模块,能够在不同尺度上捕捉目标的细节和上下文信息;最后,在检测出的人脸上根据 PERCLOS 准则进行疲劳驾驶判定。**结果** 实验结果表明:改进算法在 WIDER FACE 数据集的 Easy、Medium、Hard 子集上分别达到了 96.0%、94.6%、88.1%。**结论** 改进算法结构简单,参数量较少,满足人脸目标实时检测的要求,适合部署在车载系统等资源有限的环境中,有效保障驾驶员的驾驶安全。

关键词:疲劳驾驶检测;多特征信息融合;通道注意力;多尺度特征提取;PERCLOS 准则
中图分类号:U495;TP391.41 **文献标识码:**A doi:10.16055/j.issn.1672-058X.2025.0004.008

Fatigue Driving Detection Algorithm Based on Multi-feature Information Fusion

ZHANG Xingwang[1,2], WANG Fengsui[1,2], YANG Haiyan[1,2]
1. School of Electrical Engineering, Anhui Polytechnic University, Anhui Wuhu 241000, China
2. Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment, Ministry of Education, Anhui Wuhu 241000, China

Abstract: Objective Due to the current problem that driver fatigue detection cannot simultaneously balance detection speed and accuracy, this paper proposed a fatigue driving detection algorithm based on an improved YOLOv7 model. **Methods** Firstly, to improve the model's convergence speed and enhance its detection performance, the traditional convolutional layer was replaced with a depthwise over-parameterized convolutional layer, which accelerated the fitting process by adding learnable parameters. Secondly, in scenes with occluded targets, traditional downsampling processes can lead to significant feature loss. To improve the accuracy of occluded target detection, the algorithm introduced a Depthwise Separable Convolution (DS-Conv) module based on improved Squeeze-and-Excitation Attention. Thirdly, to enhance the model's ability to detect small targets, an MSS attention module with multi-scale feature extraction was added to the feature extraction layer, which can capture the details and contextual information of targets at different scales.

收稿日期:2024-01-04 **修回日期:**2024-03-28 **文章编号:**1672-058X(2025)04-0062-10
基金项目:安徽省自然科学基金(2108085MF197);安徽高校省级自然科学研究重点项目(KJ2019A0162);安徽工程大学国家自然科学基金预研项目(XJKY2022040).
作者简介:张兴旺(1998—),男,安徽六安人,硕士研究生,从事图像处理研究.
通信作者:王凤随(1981—),男,安徽宿州人,博士,教授,硕士生导师,从事图像与视频信息处理和计算机视觉等研究. Email:fswang@ahpu.edu.cn.
引用格式:张兴旺,王凤随,杨海燕. 基于多特征信息融合的疲劳驾驶检测算法[J]. 重庆工商大学学报(自然科学版),2025,42(4):62-71.
ZHANG Xingwang, WANG Fengsui, YANG Haiyan. Fatigue driving detection algorithm based on multi-feature information fusion [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(4): 62-71.

Finally, fatigue driving determination was performed on the detected faces according to the PERCLOS criterion. **Results** The experimental results showed that the improved algorithm achieved accuracies of 96.0%, 94.6%, and 88.1% on the Easy, Medium, and Hard subsets of the WIDER FACE dataset, respectively. **Conclusion** The improved algorithm, with its simple structure and small number of parameters, is conducive to real-time face target detection and is suitable for deployment in resource-limited environments such as in-vehicle systems, effectively ensuring driver safety.
Keywords: fatigue driving detection; multi-feature information fusion; channel attention; multi-scale feature extraction; PERCLOS criterion

1　引　言

疲劳驾驶占总交通事故发生原因的 20%,在大型卡车和公路交通事故中约占 37%,在重大交通事故中约占 43%,因此为了保障人们的生命财产安全,实施驾驶员疲劳驾驶检测具有十分重要的意义。

Zhang 等[1]提出基于面部行为分析的驾驶员疲劳检测方法,通过检测眼睛和嘴巴的状态,判断驾驶员是否处于疲劳状态。现阶段卷积神经网络(Convolutional Neural Networks,CNN)已广泛应用于图像分类、目标检测、人脸识别以及自动驾驶等任务中[2-3],并且在各任务中都卓有成效。目前卷积神经网络目标检测算法分为单阶段检测和两阶段检测。两阶段检测算法基本思想是先生成一些候选框,再对这些候选框进行分类和回归,以得到最终的检测结果。常见的两阶段目标检测算法有 MaskR-CNN[4]、R-CNN[5]、FastR-CNN[6]、FasterR-CNN[7]等。两阶段目标检测算法虽然准确率较高,但计算量较大,检测速度较慢,不能满足驾驶员疲劳驾驶检测实时性的要求。单阶段检测算法则可以直接预测边界框和类别,不需要通过两个阶段(即候选区域生成和分类)来完成。由于没有额外的候选区域生成和筛选过程,单阶段检测算法能够在减少计算资源需求的同时,实现较快的检测速度,常见的单阶段目标检测算法有 YOLO 系列[8-10]、SSD[11]、EfficientDet[12]、RetinaNet[13]等。但相较于两阶段检测算法,单阶段检测算法在准确率上可能略有不足。

臧露奇[14]在 YOLOv7 中进行了一些改进,他增加了感受野增强模块,通过捕获多尺度信息和不同距离的依赖关系,以提升模型的检测性能。此外,他还引入了排斥损失和注意力模块,以增强在复杂场景下的目标检测能力。然而,这些改进导致了模型的检测速度下降。Xu 等[15]提出一种不同的思路来解决准确率问题,他引入浅层特征和深层特征的融合机制,通过使用两种类型的金字塔结构,将上下文信息和背景信息引入金字塔结构中。这种融合机制避免了特征融合过程中的不适应现象,并提高了对于小尺度和遮挡人脸的检测能力。另外,Jia 等[16]采用 GhostNet 为 backbone 来提取特征,并通过特征融合进一步增强模型的性能,同时设计了一个新的损失函数,以平衡模型的检测速度和准确率,然而最终表现的准确率并不优异。

为了解决现有算法在疲劳驾驶检测任务上准确率和检测速度无法兼顾的问题,提出了一种基于改进 YOLOv7 的疲劳驾驶检测算法。首先,通过将传统卷积层替换为深度过度参数化卷积层[17](Depthwise Over-parameterized Convolutional,DO-Conv),增加可学习的参数,使模型能够学习到更复杂和精细的特征信息,提高目标检测的准确性,还可以帮助模型更好地训练数据,降低欠拟合的风险;其次,在特征提取层中添加多尺度特征提取 MSS 注意力模块[18],该模块能够在不同尺度上捕捉目标的细节和上下文信息,从而提高对小目标的检测能力;最后,引入基于 SE[19]注意力改进的 DS-Conv 模块,用于替换原有模型的下采样卷积层。DS-Conv 模块能够帮助模型关注与目标相关的通道信息,减少对于无关信息的依赖,提高目标检测的准确性,它还可以学习到对于遮挡目标更具区分度的通道信息,从而提高对于遮挡目标的检测能力。通过以上改进,提出的基于改进 YOLOv7 的疲劳驾驶检测算法在提高检测准确性的同时,也保持了较高的检测速度,从而有效解决了现有算法在疲劳驾驶检测任务中的问题。

2　YOLOv7 网络模型及其改进

2.1　YOLOv7 网络模型

Yolov7 是目标检测算法 YOLO 系列的一种改进版本,它以快速且准确的目标检测能力受到广泛关注。

Yolov7 主干网络采用了 Darknet,不仅具有较少参数,还拥有高效的计算性能,同时采用了更深的网络层次结构,使得模型能够学习更丰富的特征。Yolov7 通过引入特征金字塔,可以在不同层次的特征图上进行目标检测,可以更好地处理不同尺度的目标,并使用 PANet 来进行特征融合。PANet 通过自上而下和自下而上的路径,将来自不同尺度特征图的信息进行聚合,以提高目标检测的准确性。Yolov7 在训练和推理过程中采用了一些优化策略,如使用 Mish 激活函数、使用 Group Normalization(组归一化)等,以提高模型的效率

和准确性。Mish 激活函数是一种非线性激活函数,可以提供更广泛的激活范围和更好的梯度性质,进而改善模型的准确性。组归一化用来规范化特征图,相比于传统的批量归一化,组归一化对小批量数据更稳定,并且在训练过程中不依赖批量大小,有助于提高模型的泛化性能。Yolov7 在训练过程中还应用了多种数据增强技术,如随机缩放、随机裁剪、图像翻转等,以增加数据样本的多样性,提高模型的鲁棒性。

将 Yolov7 用于人脸检测任务中,主要有以下一些优点:Yolov7 相对轻量而且具有高效的性能,可以在实时或近实时的条件下进行人脸检测,这对于需要快速响应和处理的实时人脸识别应用非常关键。虽然 Yolov7 主要用于目标检测,但它可以提取人脸图像中的丰富特征。这些特征可以用于进一步的人脸识别任务,例如提取人脸特征向量并进行比对。Yolov7 的网络结构可以根据需要进行修改和扩展,以适应特定的人脸识别任务,同时可以在 Yolov7 的基础上添加额外的网络层或模块,以提高人脸识别的准确性。

2.2 改进的 YOLOv7 网络模型

2.2.1 下采样卷积层 DS-Conv

下采样操作通常会减小特征图的尺寸,从而丢失一些重要的特征信息,通过添加通道注意力,可以对每个通道进行自适应特征加权,使得模型能够更有针对性地保留重要的特征,将更多的注意力集中在对当前任务有用的通道上,从而增强模型的表达能力和性能。通道注意力还可以通过对通道进行加权乘法运算,减少对冗余信息的关注,这有助于降低模型的计算和存储成本,提高模型的计算效率。并且通过引入通道注意力,模型可以更好地适应不同的输入情况和场景变化,自适应地调整通道权重,从而提高模型的鲁棒性,使其在各种输入条件下都能有更好的性能表现。

SE(Squeeze-and-Excitation)注意力模块结构如图 1 所示。它通过学习通道间的关系,动态地调整每个通道的权重,以更好地捕捉重要的特征,增强模型的表达能力。SE 注意力机制由 3 个模块组成,分别为压缩(Squeeze)模块、激励(Excitation)模块和缩放(Scale)模块。压缩模块将卷积神经网络的输出特征图通过全局平均池化操作进行压缩,将通道维度压缩为一个特征向量。这个特征向量捕获了整个特征图的全局统计信息,反映了每个通道的重要性。激励模块通过使用两个全连接层(或卷积层)对压缩后的特征向量进行处理,以产生一个激励向量。第一个全连接层用于降低维度,第二个全连接层用于增强特征。最后使用激活函数将激励向量的值限制在 0~1 之间。缩放模块将激励向量乘以输入特征图,对每个通道进行缩放。这个缩放操作根据通道的重要性调整特征图中每个通道的权重,从而使更重要的特征得到增强,不重要的特征得到抑制。

SE 注意力机制的作用是通过自适应地学习通道权重,增强模型对重要特征的关注,抑制对于任务无关或不重要的特征。它可以帮助模型更好地适应不同样本和场景,提高模型的表达能力和准确性。同时 SE 注意力机制可以嵌入到现有的卷积神经网络中,而不需要额外的复杂结构。

DS-Conv 模块结构如图 2 所示。除了添加 SE 通道注意力以外,DS-Conv 还增加了几个卷积层。首先通过添加第一个普通卷积层将输入的通道数增加至原来的 4 倍,增加网络的表达能力和特征提取能力;然后添加一个深度可分离卷积层,减少参数量和降低计算量并获取各个通道的信息,目的是通过深度可分离卷积层来提取更加丰富和有表达力的特征;之后将特征输入到 SE 模块中,SE 模块对每个通道的特征进行降维,从而减少计算量,并在学习各个通道的权重后,融合各个通道的特征信息;最后增加一个普通卷积层恢复增加的通道数,为后续的操作提供适当的输入。DS-Conv 最终通过整合和融合各个通道的特征信息,提高了网络的性能和表达能力,从而更好地处理输入数据并提取有用的特征。

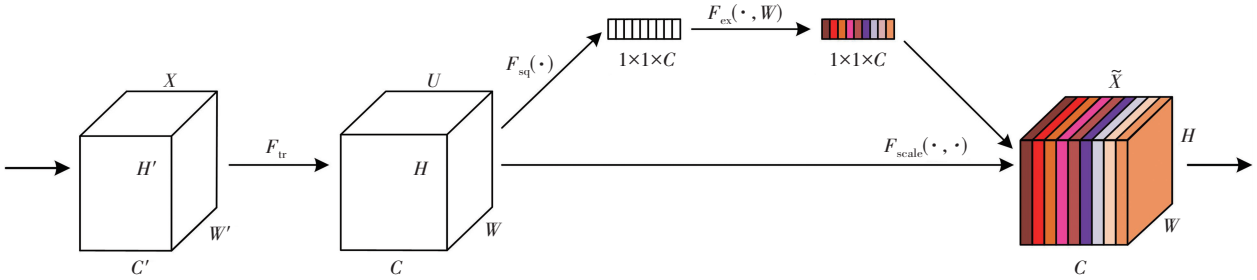


图 1 SE 注意力模块结构图

Fig. 1 Structure of SE attention module

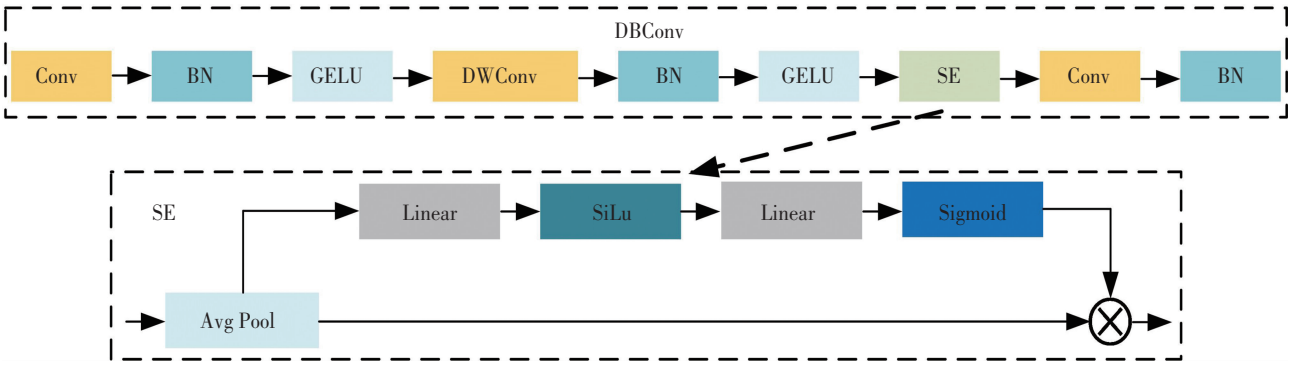


图 2　DS-Conv 模块结构图

Fig. 2　Structure of DS-Conv module

2.2.2　MSS 注意力

MSS 注意力主要由多头混合卷积(Multi-Head Mixed Convolution,MHMC)、多尺度感知聚合(Scale-Aware Aggregation,SAA)和尺度感知调制器(Scale-Aware Modulation,SAM)3 个模块组成,结构如图 3 所示。

MHMC 设置了 N 个卷积核,并将卷积核的大小初始化为 3×3,以 2 为步幅逐头递增,对每个检测头应用独立的深度可分离卷积操作。通过增加检测头的数量,可以使用更大的卷积核来扩大感受野,从而增强模型建模长距离依赖关系的能力。由于逐渐增加卷积核的大小,从而可以捕捉到不同尺度上的空间特征,较小的卷积核关注局部细节,较大的卷积核则能够捕捉更大范围的上下文信息。通过多个检测头的并行操作,模型能够同时处理多个尺度上的特征,从而提高目标检测的性能。因此 MHMC 具有通过调整头的数量来调节感受野的范围和多粒度信息的能力,MHMC 结构图如图 3(b)所示,过程可表述为

$$O_{\mathrm{MHMC}(\boldsymbol{X})}=\mathrm{Concat}\left(DW_{k_1\times k_1}(x_1),\cdots,DW_{k_n\times k_n}(x_n)\right) \tag{1}$$

式(1)中,$O_{\mathrm{MHMC}(X)}$ 表示 MHMC 的输出,$DW_{k_i\times k_i}(\boldsymbol{x})=[DW_{k_1\times k_1}(x_1),\cdots,DW_{k_n\times k_n}(x_n)]$表示对每个独立检测头应用深度可分离卷积的输出,$\boldsymbol{x}=[x_1,x_2,\cdots,x_n]$表示分割成的 N 个检测头,$k_i\in\{3,5,\cdots,K\}$ 表示卷积核大小单调增加 2。

由于 MHMC 检测头之间存在信息交互不足的问题,因此引入了 SAA 模块,设计如图 3(c)所示。SAA 模块从每个检测头中选择一个通道来共同构建一个组,这样每个组包含了来自不同检测头的特征通道。通过这种方式,SAA 模块对 MHMC 生成的不同粒度的特征进行重新组合和分组,以增强多尺度特征的多样性。然后 SAA 模块使用一个 1×1 卷积操作来进行组内和组间的信息融合,这种跨组信息融合的方式可以实现轻量化和高效的聚合效果。通过将不同组内的特征进行融合,SAA 模块促进了多尺度特征之间的交互,从而提高了目标检测模型的性能。当给定输入 $X\in\mathbf{R}^{H\times W\times C}$,组的数量 $N_{\mathrm{Groups}}=C/N_{\mathrm{Heads}}$,每个组中包含了 N 个特征通道,其中 C 为通道数,N 为检测头的数量,SAA 的过程可表述为

$$\begin{aligned}M&=\boldsymbol{W}_{\mathrm{inter}}([G_1,G_2,\cdots,G_M])\\G_i&=\boldsymbol{W}_{\mathrm{intra}}([H_1^i,H_2^i,\cdots,H_N^i])\\H_j^i&=DW\mathrm{Conv}_{k_j\times k_j}(x_j^i)\in\mathbf{R}^{H\times W\times 1}\end{aligned} \tag{2}$$

其中,$\boldsymbol{W}_{\mathrm{inter}}$ 和 $\boldsymbol{W}_{\mathrm{intra}}$ 是点向卷积的权矩阵,$j\in\{1,2,\cdots,N\}$,$i\in\{1,2,\cdots,M\}$。$H_j\in\mathbf{R}^{H\times W\times M}$ 表示深度可分离卷积第 j 个检测头的输出,H_j^i 表示第 j 个检测头第 i 个通道的输出。

图 3(a)中,在使用 MHMC 捕捉多尺度空间特征并通过 SAA 进行信息融合后,得到一个输出特征图,然后通过 SAM 进行调制。对于输入 $X\in\mathbf{R}^{H\times W\times C}$,计算输出 Z 如下:

$$\begin{aligned}Z&=M\odot V\\V&=\boldsymbol{W}_vX\\M&=\mathrm{SAA}(\mathrm{MHMC}(\boldsymbol{W}_SX))\end{aligned} \tag{3}$$

其中,$\boldsymbol{W}_v$ 和 $\boldsymbol{W}_s$ 是线性层的权矩阵。SAM 保留了信道维度,可以增强模型对不同通道特征的建模能力和灵活性。同时还可以在不同的输入下动态变化,实现自适应自调制,使模型能够独立调节每个通道的特征权重,实现特征选择和多样性的特征融合,从而提高目标检测模型的性能。

通过 MSS 多尺度特征的提取,即使目标在图像中可能以不同的尺度出现,也可以帮助模型对不同尺度的目标进行感知和定位,同时帮助模型适应这种尺度变化,捕

捉目标在不同尺度下的视觉特征。多尺度特征提取还可以提高目标定位的准确性,较小的尺度能够更好地捕捉目标的局部细节,而较大的尺度能够提供目标的全局上下文信息,通过综合利用不同尺度的特征,模型可以更准确地定位目标并减少定位误差。最后,MSS 通过融合不同尺度的特征来提高目标检测的性能,提供更全面和丰富的信息,有助于提高模型的准确性、鲁棒性和泛化能力,使其在各种场景中更加有效和可靠。

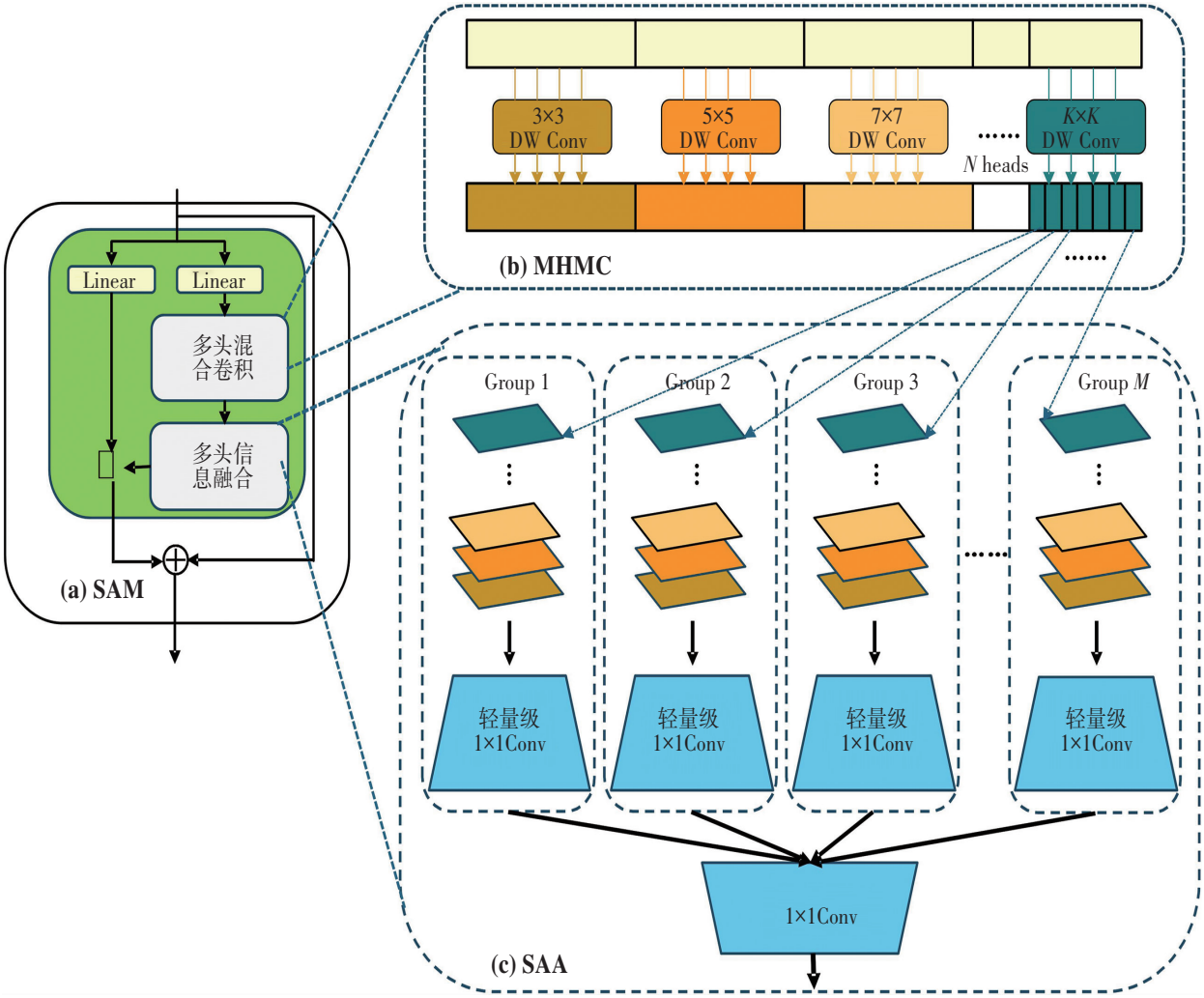


图 3 SAM 注意力结构图

Fig. 3 Structure of SAM attention

2.2.3 深度过度参数化卷积层 DO-Conv

深度过度参数化卷积层(DO-Conv)是由深度卷积核 $\boldsymbol{D}$ 和传统卷积核 $\boldsymbol{W}$ 两个可训练核共同组成。给定一个输入特征 $\boldsymbol{P}$,DO-Conv 的输出为 $\boldsymbol{O}=(\boldsymbol{D},\boldsymbol{W})\otimes\boldsymbol{P}$。如图 4 所示,DO-Conv 可以用图中数学等效的方式应用,即

$$\boldsymbol{O}=(\boldsymbol{D},\boldsymbol{W})\otimes\boldsymbol{P}=(\boldsymbol{D}^{\mathrm{T}}\circ\boldsymbol{W})*\boldsymbol{P} \tag{4}$$

其中,$\boldsymbol{W}\in\mathbf{R}^{C_{\mathrm{out}}\times D_{\mathrm{mul}}\times C_{\mathrm{in}}}$,$\boldsymbol{D}\in\mathbf{R}^{(M\times N)\times D_{\mathrm{mul}}\times C_{\mathrm{in}}}$,$\boldsymbol{P}\in\mathbf{R}^{(M\times N)\times C_{\mathrm{in}}}$,$\boldsymbol{D}^{\mathrm{T}}\in\mathbf{R}^{D_{\mathrm{mul}}\times(M\times N)\times C_{\mathrm{in}}}$ 是 $\boldsymbol{D}\in\mathbf{R}^{(M\times N)\times D_{\mathrm{mul}}\times C_{\mathrm{in}}}$ 在第一维度和第二维度上的转置。

如图 4 所示,在使用卷积核重参化时,$\circ$ 和 $*$ 的组合是通过一个复合核 $\boldsymbol{W}'$实现的,即深度过度参数化卷积算子首先通过可训练核 $\boldsymbol{D}^{\mathrm{T}}$ 来转换 $\boldsymbol{W}$,得到 $\boldsymbol{W}'=\boldsymbol{D}^{\mathrm{T}}\circ\boldsymbol{W}$,然后在 $\boldsymbol{W}'$和 $\boldsymbol{P}$ 之间应用传统卷积算子,得到 $\boldsymbol{O}=\boldsymbol{W}'*\boldsymbol{P}$。DO-Conv 是把传统卷积层做了一个过度参数化的处理。传统卷积层的参数 A 和 DO-Conv 的参数 A'分别可表示为

$$\begin{aligned}A&:C_{\mathrm{out}}\times(M\times N)\times C_{\mathrm{in}}\\A'&:C_{\mathrm{out}}\times D_{\mathrm{mul}}\times C_{\mathrm{in}}+(M\times N)\times D_{\mathrm{mul}}\times C_{\mathrm{in}}\end{aligned} \tag{5}$$

其中,$D_{\mathrm{mul}}\geqslant(M\times N)$,即使在 $D_{\mathrm{mul}}=(M\times N)$ 的情况下,参数的数量也增加了 $(M\times N)\times D_{\mathrm{mul}}\times C_{\mathrm{in}}$。在推理阶段,计算得到的权重 $\boldsymbol{W}'$用于推理,由于 $\boldsymbol{W}'$的形状与卷积层的核相同,所以 DO-Conv 在推理阶段的计算与传统卷积

层的计算完全相同。

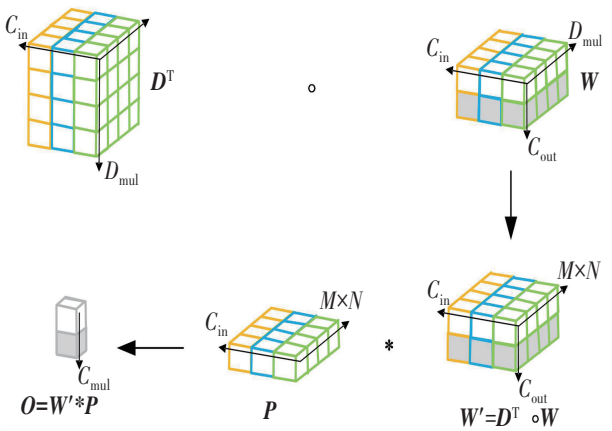


图 4　卷积核重参化

Fig. 4　Reparameterization of convolutional kernel

通过增加可学习参数,模型可以更好地拟合训练数据,提高模型的表达能力,从而更好地适应复杂的数据分布和任务要求。同时也增加了模型的灵活性,降低了欠拟合的风险,提高了模型的拟合能力。在一些复杂的任务和大规模数据集上通常需要更强大的模型来提取更复杂的特征信息,通过增加可学习参数,模型可以更好地处理这些挑战,从而提高任务的准确性和性能。

3　仿真实验与结果分析

3.1　实验数据集及环境配置

数据集采用的是 WIDER FACE[20] 数据集。WIDER FACE 数据集是一个大规模的人脸检测数据集,由伯克利大学的研究人员于 2016 年发布,共包含 32 203 张训练图像和 393 703 张个人脸部标注,覆盖大量的人脸大小、姿势、表情和遮挡情况。同时数据集基于 61 个事件类别进行组织,对于每个事件类别,随机选择 40%、10%、50%数据作为训练集、验证集和测试集,使得数据集适合和评估人脸检测算法。除了人脸边界框之外,WIDER FACE 还提供了 5 个面部关键点(左右眼、鼻子、左右嘴角)的标注,这有助于开展面部关键点检测和姿势估计等研究。WIDER FACE 数据集已成为人脸检测领域的一个重要基准,许多研究人员和工程师利用它来开发和评估新的人脸检测方法。

WIDER FACE 数据集的训练集包含了 12 876 张图片,验证集包含了 3 226 张图片。实验均在 Linux 操作系统的环境下对数据集进行训练及验证,具体运行环境配置如表 1 所示。实验训练 100 个 epoch,具体的网络模型参数设置如表 2 所示。

表 1　实验运行环境

Table 1　Experimental operating environment

类　型	参　数
操作系统	Linux
CPU	12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50 GHz
GPU	GeForce RTX 3090
PyTorch	1.90
CUDA	11.4
python	3.8

表 2　网络模型参数

Table 2　Network model parameters

参　数	数　值
Batchsize	16
epoch	100
device	0
workers	8

3.2　驾驶疲劳判定与评价指标

通过检测驾驶员眼睛是否处于闭合状态以及嘴巴是否打哈欠的方法,来判定驾驶员是否处于疲劳状态。具体步骤如下:

首先通过改进的 YOLOv7 网络模型检测输入图像,从图像中检测到脸部区域后,通过基于 Dlib 库的人脸关键点检测器获取关键点,并在图像中绘制出关键点位置。如图 5(a)为人脸 68 个关键点描绘图,图 5(b)为右眼及嘴巴关键点坐标。通过提取眼部特征并计算眼部纵横比(E_{AR})判断眼睛是否闭合。以左眼为例,E_{AR} 计算公式为

$$E_{AR}=\frac{\|P_2-P_6\|+\|P_3-P_5\|}{2\|P_1-P_4\|} \tag{6}$$

同理,计算出右眼的 E_{AR},然后将计算得出的两个 E_{AR} 进行平均后获得最终 E_{AR}。得到的 E_{AR} 结果结合 PERCLOS 准则,判断驾驶员是否处于疲劳状态。PERCLOS 是一种用于疲劳检测的指标,全称为“闭眼时间百分比”。目前,PERCLOS 有 EM、P70 和 P80 3 种准则判断被检测者是否处于疲劳状态。研究表明:当

处于驾驶状态时,P80 准则与驾驶员疲劳状态的相关性最高,即瞳孔被眼睑覆盖超 80%的面积时判定眼睛处于闭合状态,因此实验中的 PERCLOS 采用 P80 准则。一般来说,闭眼时间百分比在 0~15%之间的驾驶员为正常状态,闭眼时间百分比在 15%~30%之间的驾驶员可能处于疲劳状态,闭眼时间百分比超过 30%的驾驶员则极有可能处于严重的疲劳状态。闭眼时间百分比 P_{P80} 计算方式如下:

$$P_{P80}=\frac{F_{close}}{F_a} \tag{7}$$

式(7)中,F_a 为检测周期时间内的总帧数,F_{close} 为检测周期内处于闭眼状态时的总帧数。

通过提取嘴部特征并计算嘴部纵横比(M_{AR})判断嘴巴是否打哈欠,M_{AR} 计算公式为

$$M_{AR}=\frac{\| K_2-K_8 \| + \| K_3-K_7 \| + \| K_4-K_6 \|}{3 \| K_1-K_5 \|} \tag{8}$$

驾驶员在说话时嘴巴也处于张开状态,但当打哈欠时,M_{AR} 值会持续上升,并会持续一定的时间。因此将得到的 M_{AR} 结果结合 PERCLOS 准则,判断驾驶员是否处于疲劳状态。相比于使用神经网络分类,计算眼部状态和嘴部状态来判断驾驶员是否疲劳的方法更加简便,同时可以提高判断的准确性。

实验通过精确率 P(Precision)、召回率 R(Recall)、平均精度(mean Average Precision,mAP)值作为评价网络模型优劣的指标。精确率 P 表示分类器预测为正样本且实际为正样本的目标在所有被分类器预测为正样本的目标中所占的比例。精确率 P 计算公式如下:

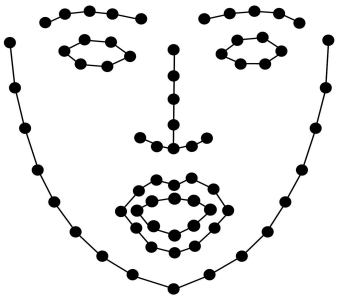
$$P=\frac{T_P}{T_P+F_P} \tag{9}$$

式(9)中,T_P 表示实际为正样本且预测结果为正样本的目标,即正确分类的正样本;F_P 表示实际为负样本但预测结果为正样本的目标,即错误分类的负样本。

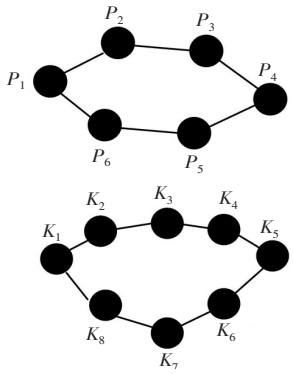
召回率 R 表示分类器预测为正样本且实际为正样本的目标在所有实际为正样本的目标中所占的比例。召回率计算公式如下:

$$R=\frac{T_P}{T_P+F_N} \tag{10}$$

式(10)中,T_P 与精确率计算中的含义相同;F_N 表示实际为正样本但预测结果为负样本的目标,即错误分类的正样本。以 epoch 为横坐标,精确率 P 为纵坐标,绘制 P-R 曲线,平均精度 mAP 的值可以表示为该曲线下的面积大小。



(a) Dlib 人脸关键点坐标



(b) 右眼及嘴巴关键点坐标

图 5 Dlib 人脸关键点检测器的 68 个关键点坐标和右眼及嘴巴关键点坐标

Fig. 5 Dlib face keypoint detector with 68 keypoint coordinates and right eye and mouth keypoint coordinates

3.3 实验结果

模型以平均精度 mAP 作为评价指标,mAP 是所有类别 AP 的均值。在 WIDER FACE 验证集的 3 个子集上的 mAP 曲线如图 6 所示,提出的算法最终在 WIDER FACE 验证集的 Easy、Medium、Hard 子集上分别达到 96.0%、94.6%、88.1%。为了进一步证明使用模块的有效性,将所提出的算法与其他算法进行对比,结果如表 3 所示。

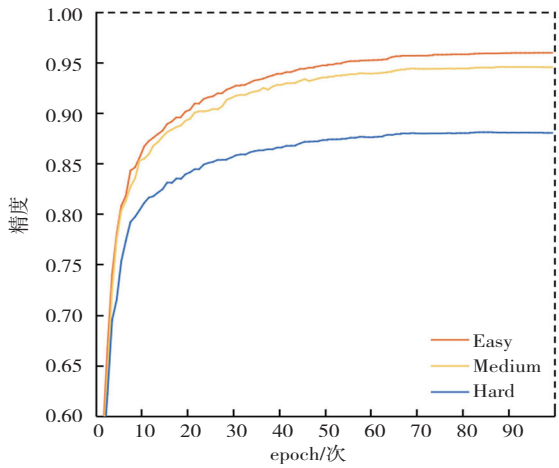


图 6 算法在 WIDER FACE 验证集 Easy、Medium、Hard 子集上的平均精度

Fig. 6 Algorithms on the WIDER FACE validation set Easy、Medium、Hard subset of the mAP

表 3　不同算法在 WIDER FACE 验证集上的结果(mAP)对比

Table 3　Comparison of results(mAP) of different algorithms on the WIDER FACE validation set

模　型	Easy	Medium	Hard
DSFD[21]	96.6	95.7	90.4
SFDet[22]	95.4	94.5	88.8
FANet[23]	95.2	94.0	90.0
FA-RPN[24]	94.9	94.1	89.4
文献[25]	94.9	93.0	83.6
SFD[26]	93.7	92.5	85.9
HR[27]	93.7	92.1	83.1
SSH[28]	93.1	92.1	84.5
CMS-RCNN[29]	89.9	87.4	62.4
Ours	**96.0**	**94.6**	**88.1**

通过表 3 可以看出:提出的算法在平均精度上仍有提升的空间,不过参数量较少,仅有 49.5 M,并且采用单阶段检测,适合在车载系统等资源有限的环境下部署。相比之下,DSFD 算法的参数量为 459 M,需要更多的计算资源。因此提出的算法在性能和计算资源之间取得了一定的平衡。此外与一些多阶段的人脸检测算法,如 CMS-RCNN 相比,提出的算法在准确率上取得了显著的提升。

3.4　消融实验

通过对 YOLOv7 模型进行一系列的改进,包括将传统卷积层替换为 DO-Conv,把多尺度特征提取 MSS 注意力模块添加在特征提取层中,引入基于通道注意力 SE 改进的 DS-Conv 模块,用于替换原有模型的下采样卷积层。实验结果如表 4 所示。

表 4　消融实验检测结果(mAP)对比

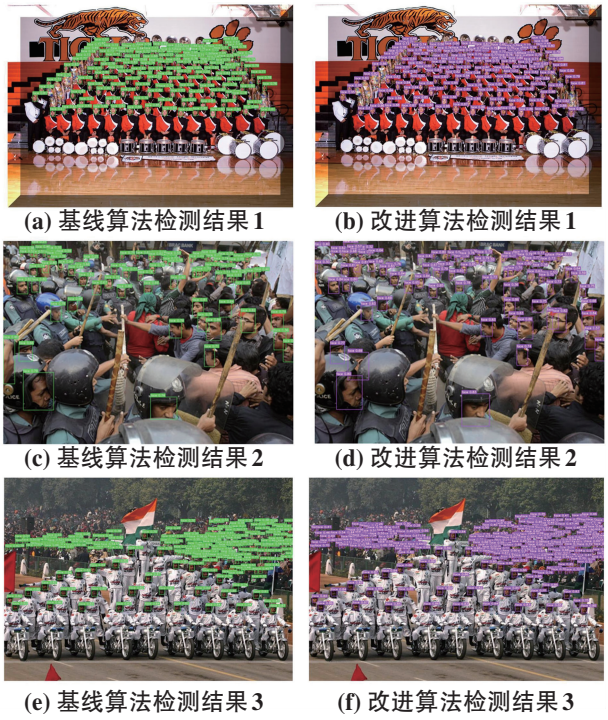
Table 4　Comparison of ablation test results(mAP)

模　型	DO-Conv	DS-Conv	MSS	参　数	Easy	Medium	Hard
基线算法	×	×	×	37 196 556	94.7	93.2	86.1
实验 1	✓	×	×	38 518 962	95.0	93.6	86.3
实验 2	✓	✓	×	41 562 290	95.3	94.0	87.0
实验 3	✓	✓	✓	49 545 394	96.0	94.6	88.1

根据表 4 中的数据可以看出:在增加多尺度特征提取 MSS 注意力模块后,模型可以更好地捕捉目标在不同大小和比例下的特征信息,Easy、Medium 和 Hard 子集上的精度均有较高提升,分别提升了 0.7%、0.6%和 1.1%。另外在添加 DS-Conv 模块后,Hard 子集提升较多,精度提高了 0.7%,这是因为在 Hard 子集中小脸所占的像素太少,导致下采样过程中特征丢失严重,通过引入通道注意力,可以更好地关注细节信息,从而提升在 Hard 子集上的准确率。最后引入 DO-Conv 增加了模型可学习的参数,Easy、Medium 和 Hard 子集上的准确率也有所提升,分别提高了 0.3%、0.4%和 0.2%。综上所述,通过增加 MSS 注意力模块、DS-Conv 模块和 DO-Conv 模块,可以改善模型在不同环境下的特征提取能力,提升检测准确性。

3.5　可视化结果

图 7 比较了使用改进算法与基线算法进行人脸检测任务的结果。图 7(a)、图 7(c)、图 7(e)展示了基线算法的检测结果,图 7(b)、图 7(d)、图 7(f)展示了改进算法的检测结果。在图 7(a)、图 7(b)中可以看到经过训练的 YOLOv7 模型在人脸识别任务中取得了出色的表现,但与改进算法结果相比,基线算法出现了一例误检,证明了改进算法在检测任务中的稳定性。在图 7(c)、图 7(d)中,当人脸存在遮挡情况时,基线算法的漏检率和误检率较高。相比之下,改进算法基本完成了检测任务,极大降低了漏检率和误检率,这表明改进算法具备较高的准确性,能够应对遮挡情况下的人脸检测挑战。在图 7(e)、图 7(f)中,在人脸较为密集且包含大量小脸的场景下,基线算法对于图片中远处小脸的检测效果较差,然而改进算法检测出的小脸数量大幅提升,对人脸的定位也更加精准。基线算法和改进算法漏检率和误检率如表 5 所示,在 Hard 子集中,改进算法漏检率下降了 2.7%,对复杂环境下的目标检测检出率大幅度提高。在误检率上,改进算法较基线算法在 Easy、Medium 和 Hard 上分别下降 0.6%、0.7%和 0.9%,表明改进算法在不同难度级别的场景中能够减少错误检测的情况。实验结果表明改进算法在人脸检测任务中相较于基线算法展现了更好的稳定性、准确性,以及在遮挡和小脸等复杂场景情况下也表现了更出色的检测能力。



(a) 基线算法检测结果 1 (b) 改进算法检测结果 1

(c) 基线算法检测结果 2 (d) 改进算法检测结果 2

(e) 基线算法检测结果 3 (f) 改进算法检测结果 3

图 7 图像识别对比结果

Fig. 7 Comparison of image recognition results

表 5 漏检率和误检率检测结果对比

Table 5 Comparison of detection results leakage rate and false detection rate /%

模 型	数据集	漏检率	误检率
基线算法	Easy	12.5	7.6
	Medium	14.2	8.1
	Hard	21.1	10.2
改进算法	Easy	11.2	7
	Medium	12.7	7.4
	Hard	18.4	9.3

4 结 论

提出一种基于 YOLOv7 的改进疲劳驾驶检测算法,在显著提高准确率的同时,保持了较高的检测速度。首先将深度过度参数化卷积层 DO-Conv 替换为传统卷积层,在不增加推理计算复杂度的情况下加快拟合过程,增强模型的检测性能;其次引入基于通道注意力 SE 改进的 DS-Conv 模块,用于替换原有模型的下采样卷积层,极大改善了遮挡目标场景下下采样过程特征丢失严重的问题,提高了目标检测的准确性;再次在特征提取层中添加多尺度特征提取 MSS 注意力模块,能够从不同尺度上捕捉目标的细节和上下文信息,增强对小目标的检测能力。实验结果表明:提出的算法在 WIDER FACE 数据集的 Easy、Medium、Hard 子集上的检测精度分别达到了 96.0%、94.6%、88.1%。综上所述,提出的算法结构简单、参数量较少、准确率较高,适用于资源受限和复杂条件下的环境,因此具有实际应用的潜力,这使得提出的算法成为一种值得考虑的疲劳驾驶检测模型解决方案。

参考文献(References):

[1] ZHANG F, SU J, GENG L, et al. Driver fatigue detection based on eye state recognition[C]//Proceedings of the International Conference on Machine Vision and Information Technology. IEEE, 2017: 105-110.

[2] FLOREZ R, PALOMINO-QUISPE F, COAQUIRA-CASTILLO R J, et al. A CNN-based approach for driver drowsiness detection by real-time eye state identification[J]. Applied Sciences, 2023, 13(13): 7849-7855.

[3] KIM M, JAIN A K, LIU X. AdaFace: Quality adaptive margin for face recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2022: 18750-18759.

[4] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2017: 2961-2969.

[5] 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. IEEE, 2014: 580-587.

[6] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2015: 1440-1448.

[7] 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 and Machine Intelligence, 2017, 39(6): 1137-1149.

[8] 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.

[9] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 7263

-7271.

[10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: Optimal speed and accuracy of object detection[C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 321-333.

[11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.

[12] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2020: 10781-10790.

[13] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2017: 2980-2988.

[14] 臧露奇. 基于 YOLOv7 改进的人脸检测算法[D]. 南昌: 南昌大学, 2023.
ZANG Lu-qi. Improved face detection algorithm based on YOLOv7[D]. Nanchang: Nanchang University, 2023.

[15] XU Z, BAI H, XIAO J, et al. Occluded and tiny face detection with deep and shallow features fusion and compensation[M]//Advances in Intelligent Information Hiding and Multimedia Signal Processing. Singapore: Springer Nature Singapore, 2022: 181-190.

[16] JIA H, XIAO Z, JI P. Real-time fatigue driving detection system based on multi-module fusion[J]. Computers & Graphics, 2022, 108(2): 22-33.

[17] CAO J, LI Y, SUN M, et al. Do-conv: Depthwise over-parameterized convolutional layer[J]. IEEE Transactions on Image Processing, 2022, 31(10): 3726-3736.

[18] LIN W, WU Z, CHEN J, et al. Scale-aware modulation meet transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2023: 6015-6026.

[19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 7132-7141.

[20] YANG S, LUO P, LOY CC, et al. Wider face: A face detection benchmark[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016: 5525-5533.

[21] LI J, WANG Y, WANG C, et al. DSFD: Dual shot face detector[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2019: 5060-5069.

[22] ZHANG S, WEN L, SHI H, et al. Single-shot scale-aware network for real-time face detection[J]. International Journal of Computer Vision, 2019, 127(6): 537-559.

[23] ZHANG J, WU X, HOI S C H, et al. Feature agglomeration networks for single stage face detection[J]. Neurocomputing, 2020, 380: 180-189.

[24] NAJIBI M, SINGH B, DAVIS L S. FA-RPN: Floating region proposals for face detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2019: 7723-7732.

[25] 董子平, 陈世国, 廖国清. 基于 YOLOv5s 的密集多人脸检测算法[J]. 计算机工程与科学, 2023, 45(10): 1838-1846.
DONG Zi-ping, CHEN Shi-guo, LIAO Guo-qing. A dense multi-face detection algorithm based on YOLOv5s[J]. Computer Engineering & Science, 2023, 45(10): 1838-1846.

[26] ZHANG S, ZHU X, LEI Z, et al. S^3FD: Single shot scale-invariant face detector[C]//Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2017: 192-201.

[27] HU P, RAMANAN D. Finding tiny faces[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 951-959.

[28] NAJIBI M, SAMANGOUEI P, CHELLAPPA R, et al. SSh: Single stage headless face detector[C]//Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2017: 4875-4884.

[29] ZHU C, ZHENG Y, LUU K, et al. CMS-RCNN: Contextual multi-scale region-based CNN for unconstrained face detection[C]//Deep Learning for Biometrics. Cham: Springer International Publishing, 2017: 57-79.

责任编辑:李翠薇