

徐红梅, 李亚林, 李中鑫, 等. 基于改进YOLOv7的复杂环境下拖拉机驾驶员面部检测[J]. 华中农业大学学报, 2025, 44(4): 288-301.  
DOI: 10.13300/j.cnki.hnlkxb.2025.04.028

## 基于改进YOLOv7的复杂环境下拖拉机 驾驶员面部检测

徐红梅, 李亚林, 李中鑫, 蒙焱仕, 阳康鑫, 李旭荣

华中农业大学工学院/农业农村部长江中下游农业装备重点实验室, 武汉 430070

**摘要** 为提高复杂场景下拖拉机驾驶员面部小目标检测精度, 基于YOLOv7算法提出了一种精度高、泛化性强的驾驶员面部小目标检测方法。首先, 构建改进空间金字塔池化模块, 有效聚合低频全局信息与高频局部信息; 其次, 采用跨级部分网络模块, 提高算法的计算效益; 最后, 调整检测层结构, 构建全新检测头 SC\_C\_detect, 提高小目标特征提取能力。消融实验及对比试验结果显示, 改进后的算法单张图片检测时间为 7.8 ms, mAP@0.5 为 97.29%, mAP0.5:0.95 为 69.45%, 优于 Faster-RCNN、YOLOv5l、YOLOv8l 等目标检测算法。在拖拉机不同振动水平下开展泛化性试验发现, 改进后面部小目标检测模型的背景误差与定位误差均有所降低。结果表明, 该算法兼具实时性与准确性, 且在不同振动水平下, 拥有良好的泛化性能。

**关键词** 拖拉机; 驾驶员; 面部检测; 小目标检测; YOLOv7

**中图分类号** TP391.4

**文献标识码** A

**文章编号** 1000-2421(2025)04-0288-14

在中国农业现代化进程中, 农机安全监理系统的发展对保障生产安全、优化生产策略具有重要意义。近年来, 拖拉机已成为农业生产的重要基础设施。随着产业规模不断扩大, 拖拉机危险驾驶、违规操作所引起的安全生产问题却日益突出。据公安部门统计, 近 20 年来, 中国拖拉机交通事故死亡人数高达 48 327 人, 直接财产损失 43 203 万元, 其中 53% 的灾难性事故与驾驶员的危险驾驶有关<sup>[1-2]</sup>。为此, 必须加强驾驶员状态监测方法与安全预警技术的研究, 以达到防范驾驶员危险驾驶、保障拖拉机安全生产的目的。

驾驶员状态检测是农机安全监理系统的关键环节, 而面部检测是驾驶员状态监测的重要基础。传统的面部检测方法主要通过人脸特征模板匹配或分类器来手工设计特征, 在变化复杂的背景下, 其鲁棒性与稳定性较差<sup>[3-5]</sup>。卷积神经网络是一种具有强表征能力和泛化性的深度学习网络, 可以提取图像层次化的特征, 大幅度地提升检测性能。近年来, 基于卷积神经网络的目标检测算法已取得较大进展, 并成功应用于生物医学、智能安防等重要领域。但

在绝对尺寸小于 32 像素×32 像素的面部小目标检测领域<sup>[6]</sup>, 由于检测时存在可利用像素较少、可视化特征不明显、易受复杂场景干扰等问题, 经常出现漏检和误检现象, 且目标检测精度也亟待提高。

近年来, 面部小目标检测已成为目标检测领域的研究热点与难点。为了提高复杂场景下面部小目标的检测精度, 研究人员从网络结构、训练策略等方面展开了大量的研究工作。Zhang 等<sup>[7]</sup>提出一种细化神经网络 RefineFace, 采用残差网络 ResNet<sup>[8]</sup>作为主干网络, 并引入特征监督模块与感受野增强模块, 从而更好地捕捉小目标人脸及某些极端姿态下的人脸。Li 等<sup>[9]</sup>提出的 ASFD (automatic and scalable face detector) 网络, 通过可微神经网络架构搜索 1 个独特的特征金字塔网络 (feature pyramid network, FPN), 并融合骨干网络输出 6 个不同尺度的特征图, 以灵活处理不同尺度的目标。Lin 等<sup>[10]</sup>提出的 FPN 可以融合不同层次的特征图以提高检测准确率, 但该方法计算成本较高, 且训练过程相对复杂。刘委坤<sup>[11]</sup>运用复色聚类性与检测速度优异的 YCbCr 颜色空间进行人脸肤色分割, 并采用积分图计算人脸

收稿日期: 2024-11-22

基金项目: 国家自然科学基金项目 (52175232)

徐红梅, E-mail: xhm790912@163.com

通信作者: 李旭荣, E-mail: 438409371@qq.com

的Haar特征<sup>[12]</sup>,检测速度明显提升,但鲁棒性较差且误检率较高。Liu等<sup>[13]</sup>通过在YOLOv7中引入坐标注意力机制(coordinate attention, CA),并增加一个小目标检测头,在提高动态追踪准确性的同时提高了对眼部、嘴部的定位精度。

综上分析,目前面部小目标的检测仍处于起步探索阶段,而且相关研究主要聚焦于汽车工程领域,针对拖拉机驾驶人员面部小目标检测的研究则相对较少。不同于汽车驾驶员面部检测,由于拖拉机作业场景特殊,其剧烈的颠簸及振动易造成拖拉机驾驶员面部模糊,致使驾驶员眼部、嘴部等关键特征失真。同时,随着拍摄端与拖拉机的距离逐渐增加,驾驶员面部可利用像素逐渐减少,特征提取愈加困难。此外,农业生产过程中,田间灰尘散射易造成面部对比度下降且水汽折射易引发面部特征模糊。针对上述问题,现有小目标检测算法普遍存在检测精度低、误检漏检严重等问题,因此难以直接应用于拖拉机驾驶员面部小目标的检测任务。

为此,本研究提出了一种基于改进YOLOv7的拖拉机驾驶员面部小目标检测算法YOLO-SOD,该方法基于融合注意力机制、跨级部分网络模块及融合空间和通道重建卷积SCConv的小目标检测头SC\_C\_detect,旨在提高复杂场景下拖拉机驾驶员面部小目标检测精度,为农机作业人员健康状态监测与主动安全预警提供重要参考。

## 1 材料与方法

### 1.1 数据采集及数据集制作

数据采集地点位于华中农业大学试验田,试验对象选取小、中、大3种体型的驾驶员共7位,试验拖拉机型号为东方红LX804,采集设备为尼康Z5微单相机。为充分考虑实际检测时的复杂场景,分别于各个正常耕作时间段、不同自然条件下,采集多种角度、不同距离以及具有多种遮挡情况的1515幅驾驶员面部图像作为样本数据集。数据集以640像素×640像素分辨率保存为JPG格式,并以8:2的比例划分训练集和验证集。

为验证模型检测的泛化性能,本研究在拖拉机不同振动水平下开展模型泛化试验。泛化试验数据集采用尼康Z5微单相机以视频格式采集,采集方式与训练数据集相同。试验拖拉机型号为约翰迪尔1054。试验人员驾驶拖拉机分别以5、15、25 km/h的平均速度行驶,各采集10段视频序列,每段视频平均长度为320帧,帧速率为30帧/s,分辨率为1920像

素×1080像素。随机抽取每段视频中60帧连续图片,共计1800幅驾驶员图像。

采用轻量级的图形标注软件Labelimg标注驾驶员面部。每个标签中包含一个边界框坐标,其由长度为4的数组( $x$ 、 $y$ 、width、height)表示。其中, $x$ 与 $y$ 表示边界框左上角的坐标值,width与height分别表示边界框的宽与高。

### 1.2 YOLOv7面部小目标检测模型的改进与优化

YOLO<sup>[14]</sup>(you only look once)架构基于全卷积神经网络(fully convolutional networks, FCN),是一种单阶段目标检测算法,通过将目标检测任务转化为回归问题以实现端到端的训练和预测。YOLOv7<sup>[15]</sup>源自YOLOv4、Scaled及YOLO-R,并引入多种改进模块以提高推理速度及准确性。YOLOv7主要包含Backbone、Neck、Head 3个组件。该网络通过引入扩展高效层注意力网络E-ELAN、空间金字塔池化层SPPCSPC以增强不同特征映射和参数利用率;在检测头方面,基于深度监督技术引入大、中、小3种目标尺寸的IDetect检测头框架,提高了模型的学习效果。

YOLOv7虽具有较高准确率与检测帧率,但未能有效解决小目标漏检、误检及检测精度低等问题,无法满足驾驶员面部小目标检测需求。

针对驾驶员面部小目标分辨率低、特征模糊等问题,本研究以YOLOv7网络为基线模型,提出了一种融合注意力机制卷积运算符、可切换空洞卷积、跨级部分网络模块、空间和通道重建卷积的驾驶员面部小目标检测算法YOLO-SOD。具体改进策略如下:(1)借鉴SPPF(spatial pyramid pooling-fast)的思想,在SPPCSPC模块中嵌入融合注意力机制的卷积运算符AttnConv以及可切换空洞卷积SAConv,构建AS\_SPPFCSPC模块,有效捕捉低频全局信息的同时整合高频局部信息,提高了网络在处理不同尺度和复杂度特征时的适应性和准确性。(2)在颈部网络中,采用跨级部分网络模块VoVGSDCSP替换原有的E-ELAN结构,通过深度可分离卷积(depth-wise separable convolution)及通道混洗操作,在保留更多语义信息的同时,实现更高的计算效益。(3)删除大目标检测层,新增小目标检测层,并基于空间和通道重建卷积构建了全新的检测头SC\_C\_detect,通过充分融合浅层空间信息和深层语义信息,实现小目标感受野的加权,提高模型的表征能力。改进后的网络结构如图1所示。





前,已有学者将SPPF与CSPC结合,提出了SPPFC-SPC结构<sup>[17]</sup>,在保证感受野不变的同时获得了速度的提升。

然而,SPPCSPC与SPPFCSPC利用多尺度空间金字塔池化提高模型的全局感受野,虽在低频全局信息的提取方面表现优异,但对于高频局部信息

的处理仍存在不足。为了解决这一问题,本研究在原有的SPPCSPC中引入融合注意力机制的卷积运算符AttnConv<sup>[18]</sup>,并结合可切换空洞卷积SAConv<sup>[19]</sup>以帮助网络更灵活地适应不同尺度的特征,缓解引入AttnConv后造成的计算量提升,改进后的AS\_SPPFCSPC结构如图2所示。

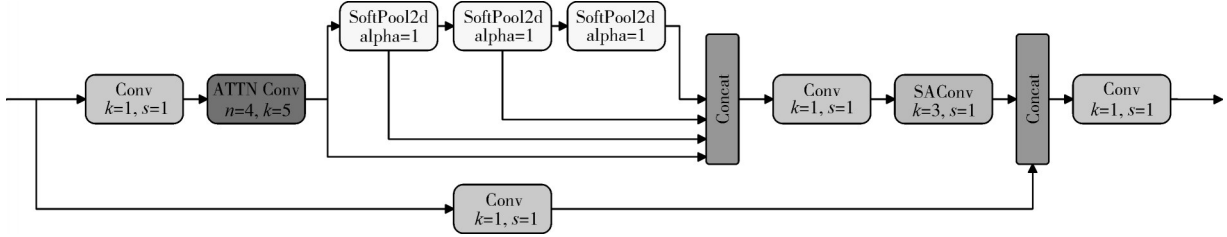


图2 AS\_SPPFCSPC模块示意图

Fig. 2 Schematic diagram of AS\_SPPFCSPC module structure

AttnConv通过线性变换得到查询向量 $Q$ 、键向量 $K$ 、数值向量 $V$ 。使用深度卷积DWConv<sup>[20]</sup>聚合局部信息,在完成局部信息与共享权值的集成之后,将 $Q$ 与 $K$ 结合后生成上下文感知权重。具体操作如下:首先使用2个DWConv分别聚合 $Q$ 与 $K$ 的局部信息,随后计算二者的Hadamard积,对结果进行一系列变换后获得上下文感知权重,最后利用生成的权值对局部特征进行增强。计算过程如式(1)所示:

$$\left\{ \begin{array}{l} Q, K, V = FC(X_{in}) \\ V_s = DWConv(V) \\ Q_t = DWConv(Q) \\ K_t = DWConv(K) \\ A_{tm} = \tanh\left(\frac{FC(\text{Swish}(FC(Q_t \odot K_t)))}{\sqrt{d}}\right) \\ X_{local} = A_{tm} \odot V \end{array} \right. \quad (1)$$

式(1)中, $X_{in}$ 表示输入特征矩阵,FC表示全连接层, $V_s$ 、 $Q_t$ 、 $K_t$ 分别表示聚合局部信息后的向量 $V$ 、 $Q$ 、 $K$ ,Swish及Tanh为非线性激活函数, $d$ 为token通道数, $A_{tm}$ 表示生成的上下文感知权重, $\odot$ 表示Hadamard积, $X_{local}$ 表示局部分支的输出特征矩阵。相比普通的注意力机制,AttnConv生成了更高质量的上下文感知权值。

局部分支与全局分支的输出融合过程如式(2)所示:

$$\left\{ \begin{array}{l} X_t = \text{Concat}(X_{local}, X_{global}) \\ X_{out} = FC(X_t) \end{array} \right. \quad (2)$$

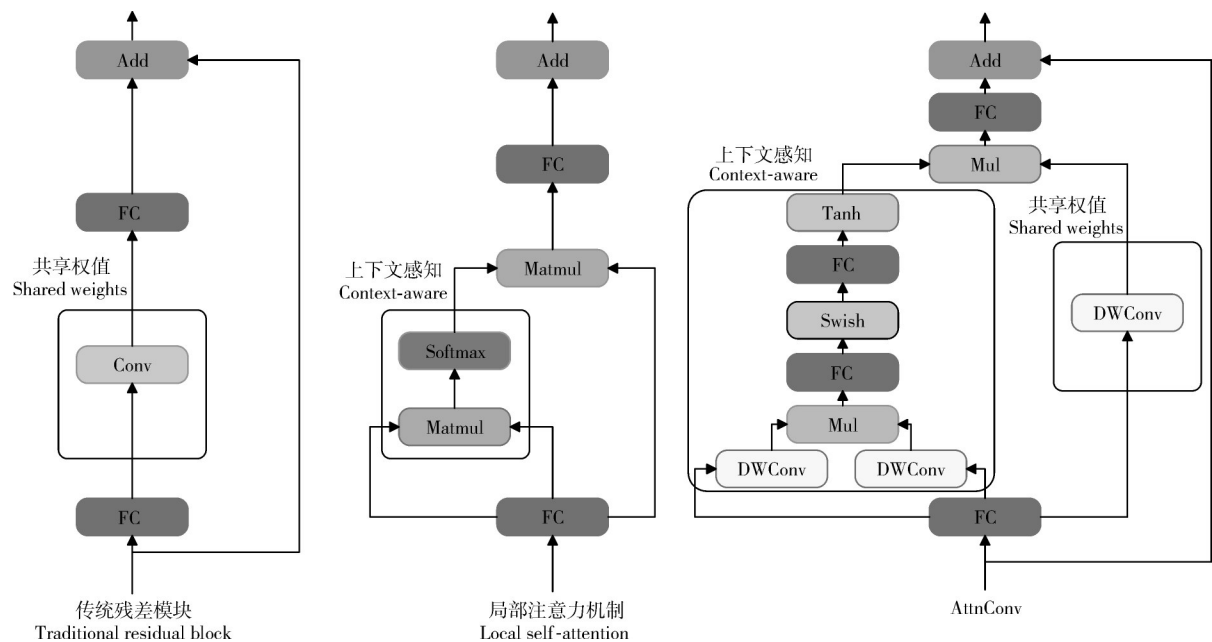
式(2)中, $X_{global}$ 表示全局分支输出, $X_t$ 表示局部与全局分支输出融合所得特征矩阵, $X_{out}$ 为输出特征矩阵。

如图3所示,传统的残差模块仅依靠卷积算子来获取高频局部信息,卷积算子通过全局共享的内核权重对相邻token执行加权和。局部自注意力机制与其不同,基于上下文感知权重,每个token都能够从相邻的token中提取高频局部信息。AttnConv结合传统卷积与局部自注意力卷积的优点,引入共享权重以更好地处理高频信息,并且在生成上下文感知权重的过程中引入了Swish与Tanh非线性激活函数,提高模型性能。

AttnConv的引入在提高模型检测精度方面表现优异,但融合了共享权值及上下文感知权重后计算量明显上升,且对于小目标中不同尺度的特征分割和检测能力有所不足。为此,本研究引入了可切换空洞卷积(switchable atrous convolution, SAConv),该模块通过对相同的输入特征应用不同的空洞率进行卷积,随后使用空间依赖的开关函数组合卷积结果,最后利用不同位置的特征图对应不同开关函数控制其输出,以扩大感受野。该模块的引入提高了网络在处理不同尺度和复杂度特征时的适应性和准确性,并且在一定程度上减少了模型参数及计算量。

2)嵌入跨级部分网络模块VoVGSDCSP。YOLOv7在颈部网络中采用ELAN-W模块,通过对通道维度及空间维度进行注意力加权以提高特征提取效率。ELAN模块考虑最短梯度路径使模型能够堆叠更多模块,丰富梯度来源。二者虽能采集丰富梯度信息,但计算量的提升不利于模型实时检测及部署。

针对上述问题,本研究采用基于轻量级卷积GSConv<sup>[21]</sup>及深度卷积DWConv的跨级部分网络模

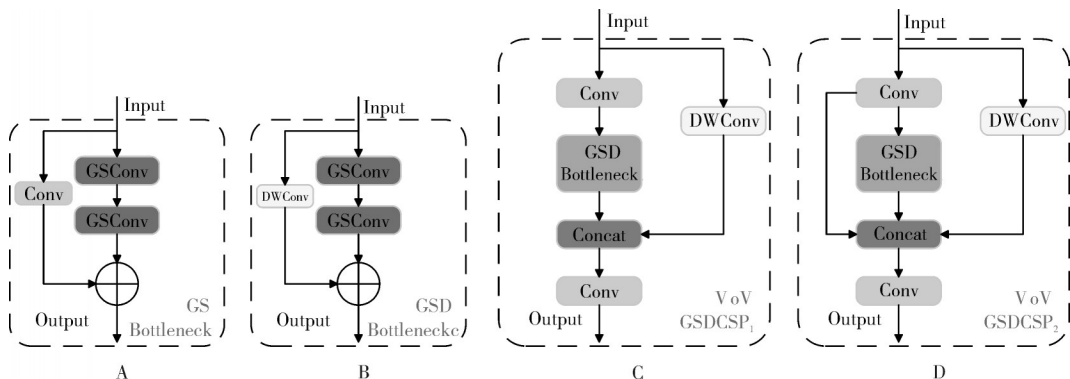


Conv 表示由 Conv2d、BN 及 ReLu 激活函数组成的卷积模块;FC 表示全连接层;Add 为特征矩阵相加;Mul 与 Matmul 表示特征矩阵相乘;Tanh、Swish 表示非线性激活函数;DWConv 表示深度卷积。Conv represents a convolutional module composed of Conv2d; BN, and Re-Lu activation function; FC represents a fully connected layer; Add denotes feature matrix addition; Mul and Matmul represent feature matrix multiplication; Tanh and Swish represent non-linear activation functions; DWConv represents depthwise convolution.

图 3 不同卷积方法比较

Fig. 3 Comparison between different convolutional methods

块 VoVGSDCSP 替代颈部网络中的 ELAN-W 模块,其结构如图 4 所示。在空间信息逐级向通道传递的过程中,特征图的空间压缩和通道扩展都会造成部分语义信息的丢失,GSCnv 通过增加深度可分离卷积(depth-wise separable convolution, DSC)和通道混洗操作以增强非线性表达能力,以较低的计算复杂度尽可能地保留重要信息。本研究引入基于 GSCnv 的细颈结构,并将原有结构第二条分支中的普通卷积替换为 DWConv,以更好聚合局部信息,新结构命名为 GSD Bottleneck。然后,采用单次聚合的方法引入跨级部分网络模块 VoVGSDCSP,并提出 2 种设计方案。其中,方案 C 选取 2 个输出相加,模型结构简洁;方案 D 选取 3 个输出相加,保留了更多的特征信息,但计算复杂度有所上升。



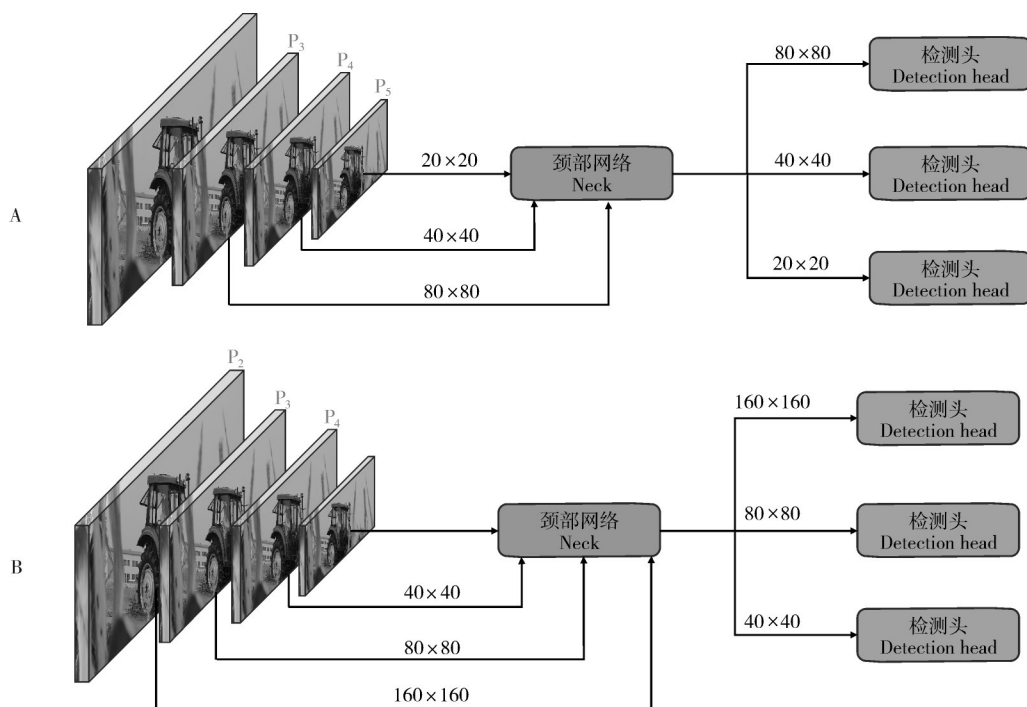
A:GS 细颈模块的结构;B:引入 DW 卷积后的 GSD 细颈模块结构;C—D:VoVGSDCSP 的 2 种方案,其中 C 方案结构较简单,推理速度更快,而 D 方案具有更高的特征重用率。A: Structure of the GS bottleneck module; B: Structure of the GSD Bottleneck module after the introduction of DW convolution; C—D: Two schemes for VoVGSDCSP, respectively, in which C scheme has a simpler structure and faster inference speed, and D has a higher reuse rate for the features.

图 4 跨级部分网络模块 VoVGSDCSP

Fig. 4 Cross stage partial network module VoVGSDCSP

3)检测层结构调整。YOLOv7网络检测层如图5所示, $P_3$ 、 $P_4$ 、 $P_5$ 检测层输出特征图依次对应小( $80 \times 80$ 像素)、中( $40 \times 40$ 像素)、大( $20 \times 20$ 像素)3种尺度的目标,并于特征融合阶段跨尺度聚合 $P_3$ 、 $P_4$ 、 $P_5$ 检测层的特征信息。然而,仅增加160像

素 $\times 160$ 像素的小目标检测层在有限提升检测精度的同时会造成网络参数量及计算开销的增大。基于此,本研究在新增160像素 $\times 160$ 像素小目标检测层 $P_2$ 的同时,剔除了20像素 $\times 20$ 像素大目标检测层 $P_5$ ,以提高模型对小目标图像的特征提取效率。



A:原始检测层 Raw detection layer;B:改进YOLO-SOD检测层 Improved YOLO-SOD detection layer.

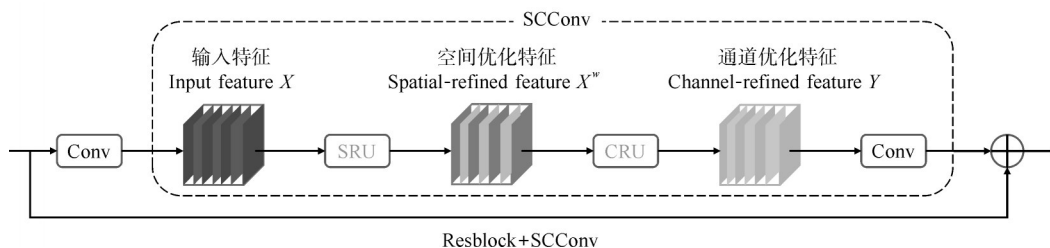
图5 2种算法检测层架构对比

Fig. 5 Architecture comparison of detection layer between two algorithms

4)嵌入小目标检测头SC\_C\_detect。YOLOv7采用基于深度监督技术的IDetect检测头以辅助模型训练,模型虽在较大目标检测任务中表现良好,但计算速度相对较慢。基于此,本研究引入空间和通道重建卷积SCConv<sup>[22]</sup>,构建全新的检测头SC\_C\_detect替换原有的检测头IDetect,以降低计算复杂性及

成本。

空间与通道重建卷积(spatial and channel reconstruction convolution,SCConv)由空间重构单元(spatial reconstruction unit,SRU)与信道重构单元(channel reconstruction unit,CRU)组成,二者通过顺序连接,其结构如图6所示。



输入特征映射 $X \in R^{N \times C \times H \times W}$ ,其中 $N$ 为Batch轴, $C$ 为Channel轴, $H$ 与 $W$ 分别为空间高度与宽度轴。通过SRU计算得空间细化特征 $X^w$ ,随后通过CRU计算获得通道细化特征 $Y$ 。Input feature mapping  $X \in R^{N \times C \times H \times W}$ , where  $N$  represents the batch axis,  $C$  represents the channel axis, and  $H$  and  $W$  represent the spatial height and width axes respectively. Spatial-refined feature  $X^w$  is obtained through SRU calculation, followed by the calculation of channel-refined feature  $Y$  through CRU.

图6 空间与通道重建卷积SCConv结构

Fig. 6 The architecture of spatial and channel reconstruction convolution

其中,空间重构单元SRU采用分离重构的方法抑制空间冗余。首先对输入特征进行标准化操作,如式(3)所示:

$$X_{\text{out}} = \text{GN}(X) = \gamma \frac{X - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \quad (3)$$

式(3)中, $\mu$ 与 $\sigma$ 为输入特征 $X$ 的均值与标准差, $\epsilon$ 为非零因子且 $\epsilon > 0$ , $\gamma$ 与 $\beta$ 为可训练的仿射变换。利用组归一化(group normalization, GN)<sup>[23]</sup>层中的比例因子评估各特征图的信息,归一化相关权重 $W_\gamma \in R^C$ 由式(4)所得:

$$W_\gamma = \frac{\gamma_i}{\sum_{j=1}^C \gamma_j}, i, j = 1, 2, \dots, C \quad (4)$$

然后,该权值通过Sigmoid函数映射至 $(0, 1)$ ,通过阈值进行门控(gate),并将大于阈值的权重设置为1,得到信息权重 $W_1$ ,小于阈值的权重设置为0,得到非信息权重 $W_2$ ,整个 $W$ 的计算方法如式(5)所示:

$$W = \text{Gate}(\text{Sigmoid}(W_\gamma(\text{GN}(X)))) \quad (5)$$

最后,将输入特征 $X$ 分别与 $W_1$ 、 $W_2$ 相乘得到空间内容矩阵 $X_1^w$ 与冗余内容矩阵 $X_2^w$ 。同时,为减少空间冗余,使用重构操作将二者特征交叉重构拼接,生成空间精细特征映射 $X^w$ ,重构过程如式(6)所示:

$$\begin{cases} X_1^w = W_1 \otimes X \\ X_2^w = W_2 \otimes X \\ X_{11}^w \oplus X_{22}^w = X^{w1} \\ X_{21}^w \oplus X_{12}^w = X^{w2} \\ X^{w1} \cup X^{w2} = X^w \end{cases} \quad (6)$$

式(6)中, $\otimes$ 为矩阵乘法, $\oplus$ 为矩阵加法, $\cup$ 为拼接操作。

通道重构单元CRU主要利用分割-转换-融合策略,减少通道冗余。对于给定的空间精细特征 $X^w \in R^{C \times H \times W}$ ,首先将其分割为 $\alpha C$ 与 $(1 - \alpha)C$ 两部分,并利用 $1 \times 1$ 卷积进一步压缩特征映射的通道以提高计算效率。随后,将空间精细特征 $X^w$ 划分为 $X_{\text{up}}$ 与 $X_{\text{low}}$ 两部分。通过组卷积(groupwise convolution, GWC)<sup>[24]</sup>、逐点卷积(pointwise convolution, PWC)<sup>[25]</sup>替换标准 $k \times k$ 卷积以提取丰富的代表性特征,特征提取过程如式(7)所示:

$$Y_1 = M^G X_{\text{up}} + M^{P_1} X_{\text{up}} \quad (7)$$

式(7)中, $M^G \in R^{\frac{\alpha C}{gr} \times k \times k \times C}$ , $M^{P_1} \in R^{\frac{\alpha C}{r} \times 1 \times 1 \times C}$ 为GWC与PWC的可学习权重矩阵,挤压比 $r$ 通常设置为2, $0 \leq \alpha \leq 1$ 为分割比,组大小设置为 $g=2$ , $X_{\text{up}} \in R^{\frac{\alpha C}{r} \times H \times W}$ 与 $Y_1 \in R^{C \times H \times W}$ 分别为输入与输出

的特征映射。

同时,通过低成本的卷积运算PWC及特征重用方法处理冗余特征,得到 $X_{\text{low}}$ 的输出 $Y_2$ 如式(8)所示:

$$Y_2 = M^{P_2} X_{\text{low}} \cup X_{\text{low}} \quad (8)$$

式(8)中, $M^{P_2} \in R^{\frac{(1-\alpha)C}{r} \times 1 \times 1 \times (1-\frac{1-\alpha}{r})C}$ 为PWC的可学习权重矩阵, $\cup$ 为拼接操作, $X_{\text{low}} \in R^{\frac{(1-\alpha)C}{r} \times H \times W}$ 与 $Y_2 \in R^{C \times H \times W}$ 分别为输入与输出特征映射。

转换完成后,首先采用全局平均池化的方法,通过通道统计 $S_m \in R^{C \times 1 \times 1}$ 收集全局空间信息,其计算公式为:

$$S_m = \text{Pooling}(Y_m) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W Y_c(i, j), m = 1, 2 \quad (9)$$

然后,将上下部全局通道描述符 $S_1$ 、 $S_2$ 叠加后使用通道注意力运算操作生成特征向量 $\beta_1$ 、 $\beta_2 \in R^c$ ,计算过程如式(10)所示:

$$\begin{cases} \beta_1 = \frac{e^{S_1}}{e^{S_1} + e^{S_2}} \\ \beta_2 = \frac{e^{S_2}}{e^{S_1} + e^{S_2}} \\ \beta_1 + \beta_2 = 1 \end{cases} \quad (10)$$

最后,将上部特征 $Y_1$ 与下部特征 $Y_2$ 按通道方向合并,得到细化特征 $Y$ ,其计算过程如式(11)所示:

$$Y = \beta_1 Y_1 + \beta_2 Y_2 \quad (11)$$

SCConv通过组归一化中的缩放因子评估不同特征所含信息,通过交叉重建操作将加权后的不同信息特征合并,最后通过特征的转换与融合将全局空间信息与通道统计信息合并,在提升小目标检测能力的同时,保留了较大目标的精细特征。此外,SRU与CRU两个单元的参数主要集中在转换阶段,因此理论内存占用较小。通过将CRU与SRU顺序组合为高效的SCConv结构,在减少计算量的同时提高模型性能。本研究采用SCConv替换原有IDetect中的标准卷积,搭建了全新的检测头结构SC\_C\_detect。

### 1.3 模型训练与测试

模型训练采用随机梯度下降法(stochastic gradient descent, SGD)<sup>[26]</sup>作为优化器对网络进行优化,初始学习率设为0.01,最终学习率设为0.002,采用余弦退火衰减调整学习率,权重衰减系数设为0.000 5,动量因子设为0.937;采用warm up策略,在前3轮内采用0.000 1的学习率进行预热训练,待模型趋于稳定后恢复到初始学习率,批次大小为8,共迭代100轮。



整个模型训练与测试过程采用PyTorch12.1.2框架, 基于PC端AMD Ryzen 5 5600 6-Core Processor、3.50 GHz处理器及NVIDIA GeForce 3060 12 GB实现。

#### 1.4 评价指标

试验采用准确率(precision,  $P$ )、召回率(recall,  $R$ )、 $mAP@0.5$ 以及 $mAP@0.5:0.95$ 作为评价指标, 其中 $mAP@0.5$ 表示交并比(intersection over union, IoU)<sup>[27]</sup>阈值为0.5时的检测精度, $mAP@0.5:0.95$ 表示IoU在0.50至0.95范围内的平均检测精度。采用参数量作为模型大小的评价指标,模型推理速度根据单张图片检测时间进行评价。其中,IoU表示为:

$$IoU = \frac{A \cap B}{A \cup B} \quad (12)$$

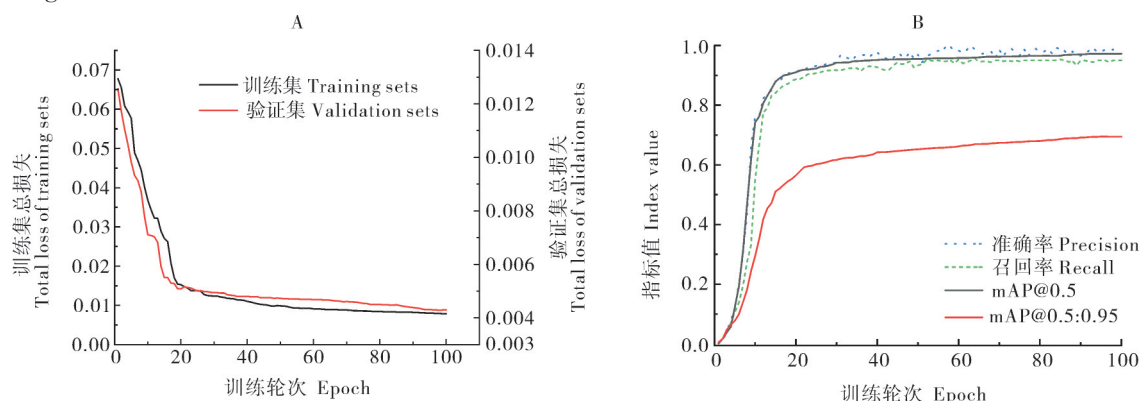
式(12)中, $A \cap B$ 表示预测框(bounding box)与真实框(ground truth)交集的面积, $A \cup B$ 表示预测框

与真实框并集的面积。

同时,引入TIDE<sup>[28]</sup>中的定位错误Loc、背景检测错误Bkg及漏检错误Miss作为小目标检测的额外评价指标。其中,Miss表示所有未被定位或分类误差覆盖的假阴性样本(false negatives),Loc与Bkg计算方法如下所示:

$$\begin{cases} \text{Loc: } t_b \leq IoU_{\max} \leq t_t \\ \text{Bkg: } IoU_{\max} \leq t_b \end{cases} \quad (13)$$

式(13)中, $IoU_{\max}$ 表示被模型预测为正类的负样本(false positives)与真实边框(ground truth)的最大IoU重叠, $t_b$ 表示前景IoU阈值,设置为0.5, $t_t$ 表示背景IoU阈值,设置为1。训练集和验证集总损失曲线以及训练集各项指标精度曲线如图7A所示,当模型迭代100次时,训练集与验证集损失分别为0.007 9和0.004 3,此时各项损失趋于平缓且精度指标达到最优。



A: 训练集和验证集总损失曲线 Total loss curves for the training set and validation set; B: 训练集各项指标精度曲线 Accuracy curves for each metric in the training set.

图7 YOLO-SOD模型训练过程

Fig. 7 Training process of the YOLO-SOD model

## 2 结果与分析

### 2.1 消融实验

本研究设计消融实验以验证各改进策略对模型整体性能的影响。由表1可知,当将颈部网络中的空间金字塔池化模块SPPCSPC替换为融合注意力机制及可切换空洞卷积的全新空间金字塔池化模块AS\_SPPFCSPC时, $mAP@0.5$ 和 $mAP@0.5:0.95$ 较原始模型分别提升1.3和5.3百分点。这是由于AttnConv融合了注意力机制,充分利用共享权重与上下文感知权重,提高了模型处理高频信息及适应局部感知信息输入的能力,有助于复杂背景情况下驾驶员面部的准确定位,但参数量及检测时间也大

幅度增加,原因在于AttnConv使用共享权重聚合全局信息的方式使得模型层数加深,且该模块生成上下文感知权重的方法较局部自注意力机制更为复杂,预测过程前向传播时间显著增加。

当将颈部网络中E-ELAN模块替换为跨级部分网络模块VoVGSDCSP时,模型召回率及平均精度明显上升,但准确率下降0.8百分点。原因在于GSConv在特征图的空间压缩和通道扩展过程中最大限度地保留了语义信息,并通过DWConv更有效率地聚合局部信息,同时,GSConv使用线性的通道混洗操作,降低模型计算量的同时,提高了检测速度,但随着网络层数的加深,深层特征会加剧数据流



的抵抗,导致检测准确率轻微下降。当调整检测层结构时,模型准确率、召回率、mAP@0.5 均有所提升,但模型参数量略微增大且检测速度下降,分析原因可知,删除原有  $P_5$  检测层并添加  $P_2$  检测层后,引入了更多的权重参数及偏置项,显著提高了小目标检测准确率的同时,也增加了部分参数量,但更换计算复杂性及成本较低的检测头  $SC\_C\_detect$  后,模型参数量下降了  $8.589\times 10^6$ ,检测时间减少了 0.7 ms。

表 1 不同改进策略消融实验结果

策略 Strategy			准确率/%	召回率/%	mAP@0.5/%	mAP@0.5:0.95/%	参数量	检测时间/ms
1	2	3	Precision	Recall			Parameter	Detection time
			96.4	89.4	94.80	62.62	$34.847\times 10^6$	8.0
✓			97.7	94.7	96.10	66.60	$37.694\times 10^6$	8.4
	✓		95.6	93.7	95.50	65.31	$26.258\times 10^6$	6.7
		✓	96.5	92.7	96.01	61.90	$27.791\times 10^6$	7.3
✓	✓		98.1	95.4	96.70	68.80	$30.866\times 10^6$	7.9
✓		✓	97.5	95.0	96.69	67.00	$29.786\times 10^6$	7.6
	✓	✓	96.9	94.6	96.40	66.04	$26.208\times 10^6$	7.1
✓	✓	✓	98.6	95.7	97.29	69.45	$31.096\times 10^6$	7.8

注:策略 1 表示引入 AS\_SPPFCSPC 模块,策略 2 表示引入 VoVGSDCSP 模块,策略 3 表示调整检测层结构并引入  $SC\_C\_detect$  模块。  
Note: Strategy 1 represents the introduction of the AS\_SPPFCSPC module. Strategy 2 represents the introduction of the VoVGSDCSP module. Strategy 3 represents the introduction of the  $SC\_C\_detect$  module after the structure of the detection layer is adjusted.

选取不同角度、距离、遮挡下的数据集样本,对改进后模型有效改善了漏检、误检等问题,检测精度检测结果进行可视化,如图 8 所示。检测结果表明,明显提升。



A:原图 Original; B:YOLOv7 检测结果 Detection results of the original YOLOv7; C: YOLO-SOD 检测结果 Detection results of the YOLO-SOD.

图 8 改进 YOLO-SOD 模型检测结果

Fig. 8 The detection results of the improved YOLO-SOD model

2.2 不同空间金字塔池化模块对模型的影响

空间金字塔池化模块通过对输入的特征图进行多尺度的空间金字塔池化,以捕捉不同大小的目标和场景信息。为探究空间金字塔池化模块的不同结构对模型的影响,分别在原始YOLOv7网络中引入SP-PCSPC、SPPFCSPC、AS\_SPPCSPC、AS\_SPPFC-SPC 4个模块进行对比试验,结果如表2所示,(1)SPPCSPC与SPPFCSPC模块大小相同,但SP-PFCSPC的平均准确率及检测时间略有提升。(2)引

入AS\_SPPCSPC与AS\_SPPFCSPC后,模型的mAP@0.5和mAP@0.5:0.95均明显提升,其中,mAP@0.5分别提高0.9和1.3百分点,mAP@0.5:0.95分别提高2.68和3.98百分点。(3)改进后的AS\_SP-PCSPC与AS\_SPPFCSPC的定位错误Loc与背景检测错误Bkg均有所下降,但参数量及浮点运算量均略有提升,在模型大小和检测时间方面,AS\_SPPFC-SPC表现更优。综合各项指标,采用AS\_SPPFC-SPC替换SPPCSPC后模型性能提升更为明显。

表2 不同空间金字塔池化模块对比试验结果

Table 2 Comparative experimental results of different spatial pyramid pooling modules

模块 Module	mAP@0.5/%	mAP@0.5:0.95/%	Loc	Bkg	参数量 Parameter	浮点运算量 Floating point operation	检测时间/ms Detection time
SPPCSPC	94.80	62.62	1.85	0.59	$34.847 \times 10^6$	$26.0 \times 10^9$	8.0
SPPFCSPC	95.13	62.85	1.31	0.84	$34.847 \times 10^6$	$26.0 \times 10^9$	8.1
AS_SPPCSPC	95.70	65.30	1.33	0.85	$42.398 \times 10^6$	$30.6 \times 10^9$	8.6
AS_SPPFCSPC	96.10	66.60	1.27	0.37	$37.694 \times 10^6$	$28.4 \times 10^9$	8.4

2.3 不同跨级部分网络模块对模型的影响

为了验证不同跨级部分网络模块结构对模型的影响,分别引入改进前的VoVGSCSP<sub>1</sub>、VoVGSCSP<sub>2</sub>模块与改进后的VoVGSDCSP<sub>1</sub>、VoVGSDCSP<sub>2</sub>模块替换颈部网络中原有E-ELAN模块。由表3可知:(1)VoVGSCSP<sub>2</sub>较VoVGSCSP<sub>1</sub>提升效果不明显,模型参数量随特征利用率提高而增加,且检测速度明显下降。(2)由于DWConv通过深度卷积与逐点卷积组合,对输入特征图进行降维和整合,改进后

的VoVGSDCSP<sub>1</sub>和VoVGSDCSP<sub>2</sub>参数量及浮点运算量均有所减小,mAP@0.5:0.95分别提升0.11和0.15百分点。(3)改进后的VoVGSDCSP<sub>1</sub>针对定位错误及漏检问题分别降低1.05和1.08,模型在复杂环境下对小目标的定位能力明显提升。此外,对于VoVGSDCSP<sub>2</sub>,引入DWConv改进后,模型mAP@0.5有所降低。综合各项指标,引入DWConv后的VoVGSDCSP<sub>1</sub>模块综合性能更优。

表3 不同跨级部分网络模块对比试验结果

Table 3 Comparative experimental results of different cross-stage network modules

模块 Module	mAP@0.5/%	mAP@0.5:0.95/%	Loc	Miss	参数量 Parameter	浮点运算量 Floating point operation	检测时间/ms Detection time
VoVGSCSP1	95.34	65.20	2.08	2.07	$27.862 \times 10^6$	$22.3 \times 10^9$	6.7
VoVGSCSP2	95.39	65.58	1.93	1.98	$28.057 \times 10^6$	$22.6 \times 10^9$	7.5
VoVGSDCSP1	95.50	65.31	1.03	0.99	$26.258 \times 10^6$	$21.7 \times 10^9$	6.7
VoVGSDCSP2	95.10	65.73	1.51	1.24	$27.671 \times 10^6$	$22.4 \times 10^9$	7.5

2.4 不同检测模型对比试验

本研究选取YOLOv5l、Faster R-CNN<sup>[29]</sup>、SSD<sup>[30]</sup>以及YOLOv8l<sup>[31]</sup>等主流目标检测算法验证改进后的模型对拖拉机驾驶员面部小目标的检测效果。由表4可知,改进后的YOLO-SOD模型mAP@0.5达到最优,较YOLOv5l、Faster R-CNN、SSD、YOLOv8l分别提高3.99、6.79、7.69、0.59百分点,模型大小较YOLOv5l、Faster R-CNN、YOLOv8l分别减少 $15.956 \times 10^6$ 、 $106.003 \times 10^6$ 、 $11.346 \times 10^6$ 。YO-

LOv8l网络的mAP@0.5:0.95最高,较改进后的YO-LO-SOD网络提高了0.85百分点,分析原因可知,YOLOv8l模型在骨干网络中引入了梯度流更丰富的C2f结构,并将检测头替换为解耦头结构,具有较高的泛化性,但其在参数量方面高于改进后的YOLO-SOD模型,且C2f模块对于硬件要求较高,不利于模型整体部署。综合考虑模型性能及参数量,改进后的YOLO-SOD驾驶员面部小目标检测算法表现更优。

表 4 不同检测模型性能对比

Table 4 Performance comparison between different detection models

模型 Models	mAP@0.5/%	mAP@0.5:0.95/%	参数量 Parameter
YOLOv5l	93.3	61.0	$47.052\times10^6$
Faster R-CNN	90.5	59.3	$137.099\times10^6$
SSD	89.6	59.1	$26.285\times10^6$
YOLOv8l	96.7	70.3	$42.442\times10^6$
YOLO-SOD	97.29	69.45	$31.096\times10^6$

本研究选取不同角度、距离、场景下驾驶员操纵图像,使用类激活热力图对输出层进行可视化,更直观地反映检测结果。由图 9 可知,在复杂环境及驾驶员面部部分受到遮挡的条件下,驾驶员面部存在漏

检、背景分割误差较大、过小目标、较多背景遮挡导致的目标背景检测误差较大等问题,改进后的 YOLO-SOD 模型在上述各场景下的检测效果均明显提升。



A:原图 Original picture;B:原始 YOLOv7 检测结果 Detection results of the original YOLOv7;C:改进后 YOLO-SOD 检测结果 Detection results of the improved YOLO-SOD model.

图 9 改进前后模型检测效果对比

Fig. 9 Comparison of model detection results before and after improvement

2.5 泛化试验

拖拉机在行驶过程中振动的频率及振幅通常与机械的转速、负载、结构等因素有关。为验证改进模型在不同振动水平下的检测泛化性,选取自建模型

泛化性验证数据集进行验证。试验环境与参数配置与本文“1.3”保持一致。由表 5 可知,YOLO-SOD 在 3 种行驶速度(5、15、25 km/h)下,mAP@0.5 分别较 YOLOv7 提升 2.6、2.3、4.9 百分点,mAP@0.5:0.95



分别提升 3.5、4.0、5.6 百分点。同时, YOLO-SOD 检测时间均有效降低, 有着更好的面部小目标检测在 3 种行驶速度下, 模型检测的 Bkg、Loc 以及检测能力。

表5 YOLOv7 与 YOLO-SOD 在自建泛化性验证数据集上的试验结果

Table 5 Experimental results of YOLOv7 and YOLO-SOD on the self-constructed generalization validation dataset							
模型 Models	行驶速度/(km/h) Speed	Bkg	Loc	mAP@0.5/%	mAP@0.5:0.95/%	参数量 Parameter	检测时间/ms Detection time
YOLOv7	5	0.94	3.06	94.1	61.4	34.847×10 <sup>6</sup>	8.7
	15	1.88	5.15	91.6	59.3		9.0
	25	1.53	7.63	89.2	56.5		10.8
YOLO-SOD	5	0.39	1.41	95.7	64.9	31.096×10 <sup>6</sup>	8.1
	15	1.39	3.75	93.9	63.3		8.6
	25	1.17	2.69	94.1	62.1		10.5

3 讨 论

本研究提出了一种基于改进 YOLOv7 模型的驾驶员面部小目标检测算法 YOLO-SOD。首先, 在颈部网络中构建 AS\_SPPFCSPC 模块以提高网络处理不同尺度、复杂特征时的适应性与准确性; 其次, 在颈部网络中引入改进后的跨级部分网络模块 VoVGSDCSP, 保证模型检测精度的同时实现更高的计算效益; 最后, 调整检测层结构并构建了全新检测头 SC\_C\_detect 以加强模型提取小目标特征的能力。本研究在自建拖拉机驾驶员面部小目标数据集上进行训练和测试。结果表明, 改进后的模型相比原模型准确率  $P$  与召回率  $R$  分别提升 2.2 和 6.3 百分点,  $mAP@0.5$  和  $mAP@0.5:0.95$  分别提升 2.49 和 6.83 百分点, 且模型整体参数量明显下降。

复杂环境下拖拉机驾驶员面部小目标检测易受外界因素干扰, 存在检测精度较低、背景检测错误等问题。针对上述问题, 本研究通过构建 AS\_SPPFC-SPC 模块与 VoVGSDCSP 模块, 增强了模型针对小目标的定位能力, 有效降低了模型的误检和漏检。此外, 通过构建小目标检测层并引入结构更为简洁的 SC\_C\_detect 检测头, 在提高检测准确率的同时平衡了模型计算成本。最后, 本研究在拖拉机不同振动水平下开展模型泛化试验。结果表明, 在 3 种不同振动状态下, 改进后的模型定位错误及背景检测错误均明显降低, 能够更有效地分割检测目标与背景。

本研究提出的方法虽有效提高了拖拉机驾驶员面部小目标检测精度及泛化性, 但仍存在以下不足: 拖拉机行驶过程中, 驾驶员自遮挡或拖拉机驾驶室遮挡导致部分视频帧存在漏检、误检等问题; 模型参数量虽有所下降, 但针对嵌入式系统的部署仍需进

一步降低功耗和计算成本。后续工作拟针对驾驶员面部遮挡的修复问题展开研究, 并基于嵌入式系统开发板进行部署与应用。

参考文献 References

[1] 张已民. 浅析农机安全监管中存在的问题及应对措施[J]. 农业技术与装备, 2023(10): 106-108. ZHANG Y M. Analysis of the existing problems in the current agricultural machinery safety supervision and countermeasures [J]. Agricultural technology & equipment, 2023 (10) : 106-108 (in Chinese with English abstract).

[2] 黎泽文. 新时期加强农机安全监管的措施和方法[J]. 农机质量与监督, 2023(2) : 26. LI Z W. Measures and methods to strengthen agricultural machinery safety supervision in the new period [J]. Agricultural machinery quality & supervision, 2023 (2): 26 (in Chinese).

[3] 黄帅凤, 汤丽娟. 人脸识别技术综述[J]. 中国高科技, 2022(8) : 10-11. HUANG S F, TANG L J. Overview of face recognition technology [J]. China high-tech, 2022(8) : 10-11 (in Chinese with English abstract).

[4] 肖阳. 人脸检测算法综述[J]. 电子技术与软件工程, 2014(4) : 113-116. XIAO Y. A survey of face detection algorithms [J]. Electronic technology & software engineering, 2014 (4) : 113-116 (in Chinese).

[5] 简讯. 人脸识别技术综述[J]. 电脑知识与技术, 2019, 15(2) : 171-172. JIAN X. Summary of face recognition technology [J]. Computer knowledge and technology, 2019, 15 (2) : 171-172 (in Chinese).

[6] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context [C]//Computer Vision-ECCV 2014. September 6-12, 2014, Zurich, Switzerland. Cham: Springer International Publishing, 2014: 740-755.

[7] ZHANG S F, CHI C, LEI Z, et al. RefineFace: refinement neural network for high performance face detection [J]. IEEE



- transactions on pattern analysis and machine intelligence, 2021, 43(11):4008-4020.
- [8] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 27-30, 2016, Las Vegas, NV, USA. Las Vegas: IEEE, 2016: 770-778.
- [9] LI J, ZHANG B, WANG Y B, et al. ASFD: automatic and scalable face detector [C]//Proceedings of the 29th ACM International Conference on Multimedia. Virtual Event China. New York: ACM, 2021: 2139-2147.
- [10] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. July 21-26, 2017, Venice, Italy. Venice: IEEE, 2017: 2117-2125.
- [11] 刘委坤. 基于机器视觉的拖拉机驾驶员疲劳检测方法研究 [D]. 长春: 吉林农业大学, 2019. LIU W K. Research on tractor driver fatigue detection method based on machine vision [D]. Changchun: Jilin Agricultural University, 2019 (in Chinese with English abstract).
- [12] PAPAGEORGIOU C P, OREN M, POGGIO T. A general framework for object detection [C]//Sixth International Conference on Computer Vision. January 7, 1998, Bombay, India. Bombay: IEEE, 1998: 555-562.
- [13] LIU S G, WANG Y J, YU Q G, et al. A driver fatigue detection algorithm based on dynamic tracking of small facial targets using YOLOv7 [J]. IEICE transactions on information and systems, 2023, E106.D(11): 1881-1890.
- [14] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). June 27-30, 2016, Las Vegas, NV, USA. Las Vegas: IEEE, 2016: 779-788.
- [15] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 17-24, 2023, Vancouver, BC, Canada. Vancouver: IEEE, 2023: 7464-7475.
- [16] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [C]// Computer Vision-ECCV 2014. Cham: Springer International Publishing, 2014: 346-361.
- [17] 齐向明, 柴蕊, 高一萌. 重构SPCSPC与优化下采样的小目标检测算法 [J]. 计算机工程与应用, 2023, 59(20): 158-166. QI X M, CHAI R, GAO Y M. Algorithm of reconstructed SPCSPC and optimized downsampling for small object detection [J]. Computer engineering and applications, 2023, 59(20): 158-166 (in Chinese with English abstract).
- [18] FAN Q H, HUANG H B, GUAN J Y, et al. Rethinking local perception in lightweight vision transformer [DB/OL]. arXiv, 2023: 2303.17803 [2024-11-22]. <https://arxiv.org/abs/2303.17803v5>.
- [19] QIAO S Y, CHEN L C, YUILLE A. Detectors: detecting objects with recursive feature pyramid and switchable atrous convolution [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 20-25, 2021. Nashville, TN, USA. Nashville: IEEE, 2021: 10208-10219.
- [20] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 21-26, 2017, Honolulu, HI, USA. Honolulu: IEEE, 2017: 1800-1807.
- [21] LI H L, LI J, WEI H B, et al. Slim-neck by GSConv: a better design paradigm of detector architectures for autonomous vehicles [DB/OL]. arXiv, 2022: 2206.02424v2 [2024-11-22]. <https://export.arxiv.org/abs/2206.02424v2>.
- [22] LI J F, WEN Y, HE L H. SCConv: spatial and channel reconstruction convolution for feature redundancy [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 17-24, 2023, Vancouver, BC, Canada. Vancouver: IEEE, 2023: 6153-6162.
- [23] WU Y X, HE K M. Group normalization [C]//Computer Vision-ECCV 2018, September 8-14, 2018, Munich, Germany. Cham: Springer, 2018: 3-19.
- [24] HAN K, WANG Y H, TIAN Q, et al. GhostNet: more features from cheap operations [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 13-19, 2020, Seattle, WA, USA. Seattle: IEEE, 2020: 1577-1586.
- [25] HUA B S, TRAN M K, YEUNG S K. Pointwise convolutional neural networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018, Salt Lake City, UT, USA. Salt Lake City: IEEE, 2018: 984-993.
- [26] RUDER S. An overview of gradient descent optimization algorithms [DB/OL]. arXiv, 2022: 1609.04747 [2024-11-22]. <https://arxiv.org/abs/1609.04747v2>.
- [27] YU J H, JIANG Y N, WANG Z Y, et al. UnitBox: an advanced object detection network [C]//Proceedings of the 24th ACM International Conference on Multimedia. Amsterdam: ACM, 2016: 516-520.
- [28] BOLYA D, FOLEY S, HAYS J, et al. TIDE: a general toolbox for identifying object detection errors [C]// Computer Vision-ECCV 2020. Cham: Springer International Publishing, 2020: 558-573.
- [29] REN S Q, HE K M, 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.

- [30] 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.
- [31] REIS D, KUPEC J, HONG J, et al. Real-time flying object detection with YOLOv8 [DB/OL]. arXiv, 2023: 2305.09972 [2024-11-22]. <https://arxiv.org/abs/2305.09972v2>.

## Improved YOLOv7 based facial detection of tractor drivers in complex environments

XU Hongmei, LI Yalin, LI Zhongxin, MENG Junshi, YANG Kangxin, LI Xurong

*College of Engineering/Ministry of Agriculture and Rural Affairs Key Laboratory of Agricultural Equipment in the Middle and Lower Reaches of the Yangtze River, Huazhong Agricultural University, Wuhan 430070, China*

**Abstract** A high-precision and highly generalized method of detecting facial small object of driver based on YOLOv7 algorithm was proposed to address the issues of falsely detecting facial small target and the low accuracy of detection caused by vibration and background occlusion for tractor drivers in complex environments of agriculture. An improved spatial pyramid pooling module AS\_SPPFCSPC was used to replace SPPCSPC to effectively aggregate low-frequency global information and high-frequency local information to enhance the accuracy of facial localization for drivers. The cross-level partial network module VoVGSDCSP was used to replace the E-ELAN module in the neck network to achieve higher computational efficiency of the algorithm. The structure of detection layer was adjusted and a new detection head SC\_C\_detect was introduced to improve the ability to extract small target features. The results of ablation and comparative experiments showed that the improved algorithm had a single-image detection time of 7.8 ms, with mAP@0.5 at 97.29% and mAP@0.5:0.95 at 69.45%, superior to object detection algorithms including Faster RCNN, YOLOv5l, and YOLOv8l. The results of generalization experiments conducted on tractors at levels of different vibration showed that the background error and localization error of the facial small target detection model after improvement were effectively reduced. It is indicated that the algorithm proposed combines real-time and accuracy, with good generalization performance at different levels of vibration.

**Keywords** tractor; driver; facial detection; small target detection; YOLOv7

(责任编辑:葛晓霞)