DOI:10.19652/j.cnki.femt.2305621

基于上下文语义联合 YOLOv7 的分心驾驶检测算法*

李 富[1] 徐 凯[2] 朱灵龙[1] 沈昊君[1] 王 泉[1]
(1.无锡学院物联网工程学院 无锡 214105;2.南京信息工程大学自动化学院 南京 210044)

摘 要:针对分心驾驶检测方法存在实时性差、精度低、可部署性差的问题,提出了一种基于上下文语义增强联合 YOLOv7 的分心驾驶检测算法。首先将模型 backbone 和 head 部分的 ELAN 模块替换成语义上下文增强模块(contextual transformer,CoT),提高上下文语义信息的捕获能力。其次,将语义关联增强机制(triplet attention)融入卷积块中,插入 backbone 和 head 的连接头之间以及融合 MP2 模块,强化目标间的关联关系以及提升目标特征提取能力。最后,将自注意力双向 Transformer 模块(Biformer)模块融合 SPPCSPC 模块,提升模型对分心驾驶中的复杂场景和遮挡目标的处理能力。改进的 YOLOv7 算法在分心驾驶数据集下平均精度均值(mean average precision,mAP)达到了 87.3%,比原算法提高了 4.3%,模型参数量减少了 4.7%,每秒传输帧数达到了 90 fps,具有较好的检测精度与速度。

关键词:YOLOv7;分心驾驶检测;CoT;Biformer;Triplet Attention
中图分类号:TP391 **文献标识码**:A **国家标准学科分类代码**:520.6

Distracted driving detection algorithm based on contextual semantics combined with YOLOv7

Li Fu[1] Xu Kai[2] Zhu Linglong[1] Shen Haojun[1] Wang Quan[1]
(1. School of Internet of Things Engineering, Wuxi University, Wuxi 214105, China;
2. School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: Aiming at the problems of poor real-time, low precision and poor deployable in distracted driving detection methods, a distracted driving detection algorithm based on contextual semantic enhancement combined with YOLOv7 was proposed. Firstly, ELAN modules in backbone and head of the model are replaced with contextual transformer (CoT) block to improve the ability to capture contextual semantic information. Secondly, the Triplet Attention mechanism is integrated into the convolutional block, inserted between the connectors of backbone and head, and the MP2 module is fused to strengthen the correlation between targets and improve the capability of target feature extraction. Finally, the self-attention bidirectional transformer (Biformer) module is integrated with the SPPCSPC module to improve the model's processing ability for complex scenes and occlusive targets in distracted driving. The mean average precision (mAP) of the improved YOLOv7 algorithm in the distracted driving data set reaches 87.3%, which is 4.3% higher than that of the original algorithm, and the number of model parameters is reduced by 4.7%. The number of frames per second reached 90 fps, with good detection accuracy and speed.

Keywords: YOLOv7; distracted driving detection; CoT; Biformer; Triplet Attention

0 引 言

随着现代交通系统的蓬勃发展,道路安全和驾驶行为的监控成为了日益重要的课题。分心驾驶行为指驾驶员无法集中注意力驾驶的行为,比如在车上打电话、吃喝、抽烟等。分心驾驶会导致驾驶员失去对路况的实时判断,增加交通事故发生的风险。根据美国国家公路安全管理局的数据,每 10 起交通事故就有 4 起是因为驾驶员驾驶时分心[1]。因此,其实时监测和及时干预具有重要的社会意义。传统的监测方法存在盲区和误判的问题,而基于深度

收稿日期:2023-10-08
* **基金项目**:国家自然科学基金青年科学基金(42305158)项目资助

学习的目标检测技术为解决这一难题提供了新的途径。

近年来，随着信息化时代发展，深度学习[2-4]目标检测算法进行了多次迭代更新，主流的目标检测算法分为单阶段和两阶段算法[5]，单阶段算法以 YOLO(you only look once)[6]、SSD(single shot multibox detector)[7]等为主，单阶段算法直接检测目标标签和位置，在训练检测方面要比两阶段算法更快。两阶段算法包含 Fast R-CNN[8]、Faster R-CNN[9]、SPP(spatial pyramid pooling)[10]等，由于先要提取候选框再分类，算法步骤更多，所以它的速度比单阶段算法要慢，但是准确率要更高。尽管两阶段算法准确率很高，但是其可部署性和实时性不强，很少用于驾驶员行为检测任务中，相对而言，单阶段算法凭借更快的速度而广泛应用。目前，国内外众多学者在提升驾驶员分心行为检测精度和检测速度做出了大量算法研究。尹智帅等[11]建立 2D/3D 姿态估计的驾驶员分心检测算法，选取随机旋转以及色彩抖动来增强数据，ResNet-56 作为分类网络。最终验证了模型的泛化性。但其只能对一个时刻的驾驶员分心行为进行检测，无法做到实时检测，并且检测行为精度不高。曹立波等[12]利用轻量化目标检测模型 NanoDet 进行驾驶员分心行为检测，虽然该方法检测速度较快，但是没有通过改进创新算法去提升检测验证精度。杜琥龙等[13]建立了轻量化神经网络的驾驶员分心行为检测，以 SSD 作为主要网络，将轻量级的 MobileNet 网络替换 VGG16 的浅层网络，在 MobileNet 网络引入 HDC 模块来优化误检漏检的问题。由于模型复杂度的降低，参数量的减少，虽然该模型提升了检测速度，但是其部分目标的检测精度有所下降。魏启康等[14]通过改进 YOLOv4-tiny 算法，增加一个小目标检测头并在特征提取主干添加 ECA 注意力机制，提高了抽烟的检测精度，但是没有对其他分心驾驶类别做到有效的精度提升。Mou 等[15]提出了一种基于卷积神经网络 (convolutional neural network，CNN)和 Transformer 的端到端的双通道算法用于分心驾驶检测，为了提高模型对时间序列的拟合，采用中点残差分别对 CNN 通道和 Transformer 通道进行建模，该模型的检测精度较高，但是其网络结构较复杂，检测速度较慢，不适合部署实时检测。

综上所述，基于计算机视觉的检测方法能有效识别分心驾驶行为。但是上述研究在提升分心驾驶检测精度的同时导致网络模型复杂度大幅提升，不利于实时检测，在降低网络模型复杂度的同时又造成了检测精度的损失，并且忽略了分心驾驶数据集中丰富的上下文语义。针对上述问题，本文提出了基于上下文语义增强联合 YOLOv7 的分心驾驶检测算法，旨在保证网络模型复杂度不增加的情况下，结合不同层次的上下文语义信息来大幅提升检测精度，实现检测精度和实时性的平衡。本文结合上下文语义的思想，在模型的 backbone 和 head 部分用多个语义上下文增强模块(contextual transformer，CoT)[16]来替换原模型的 ELAN 块，在减少参数量的同时增加上下文信息的捕获，增加视觉表现能力，加强对目标局部特征的提取；将语义关联增强机制(triplet attention[17])融合卷积层，形成一个新的 CBS_ATT1 模块，插入 backbone 和 head 的连接头之间，并将 CBS_ATT1 模块与 MP2 相融合，强化目标间的语义关联关系以及增加重要目标特征的提取能力；在模型的 head 部分将自注意力双向 Transformer 模块(Biformer[18])融合 SPPCSPC 模块，提升模型对分心驾驶中复杂场景和遮挡目标的处理能力，减少漏检和错检，使得模型更具鲁棒性。将改进后的 YOLOv7 模型在分心驾驶数据集上进行试验，本文研究的模型对比原本 YOLOv7 模型的平均精度均值(mAP)提升了 4.3%。

1 YOLOv7 网络模型

YOLOv7[19]算法主要由 3 部分组成，分别为输入端(input)、主干网络(backbone)、head 结构。YOLOv7 网络结构如图 1 所示。输入端由尺寸为 640×640 的三通道图像组成；backbone 部分主要有卷积，ELAN 模块和 MP 模块 3 部分组成，其中 ELAN 模块作为特征提取以及通道数控制，使得网络学习到更多的特征；head 结构由 SPPCSPC 模块、PAFP 结构和输出端组成，通过特征融合高层特征图与底层特征图，然后再进行特征提取，最后对生成的特征图进行预测输出。

2 改进 YOLOv7 算法

改进 YOLOv7 的网络结构如图 2 所示，主要包括如下 3 个部分。

1) 在 backbone 部分和 head 部分用 CoT 模块来替换 ELAN 模块，加入位置如图 2 所示。由于原 ELAN 模块容易忽略分心驾驶图像中上下文语义关系，导致分心特征提取不准确，CoT 通过结合不同层次的上下文信息来从多个尺度上捕获和利用上下文信息，加强了对局部特征的提取。

2) 将 Triplet Attention 融合卷积层，形成一个新的 CBS_ATT1 模块，将 CBS_ATT1 模块插入 ELAN 与 SPPCSPC 之间，并且与 MP2 模块相融合，引入位置如图 2 所示。这是一种接近无参的注意力机制，几乎在不增加模型参数量的同时提升检测性能。

3) 在原本的 SPPCSPC 模块末端处插入 Biformer 模块，加入位置如图 2 所示。该模块作用是使模型更加细致地感知目标周围的环境和语境，有助于减少误检和漏检，提高检测的精确。

2.1 自注意力双向 Transformer 模块

虽然原 YOLOv7 的 SPPCSPC 模块能够增大感受野，更好的区分大目标和小目标，但是其在复杂驾驶场景中对于驾驶员的手遮挡目标物时容易出现漏检、误检的问题，所以本文在 SPPCSPC 模块的末端插入了自注意力双向 Transformer 模块。自注意力双向 Transformer 模块是一种基于自注意力机制的双向 Transformer 模型，能够在输

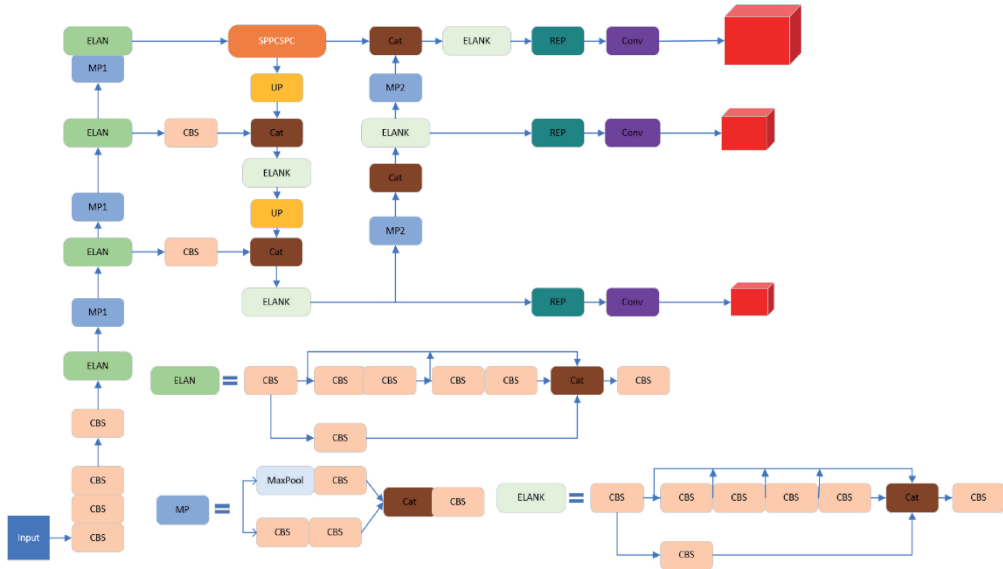


图 1　YOLOv7 网络结构

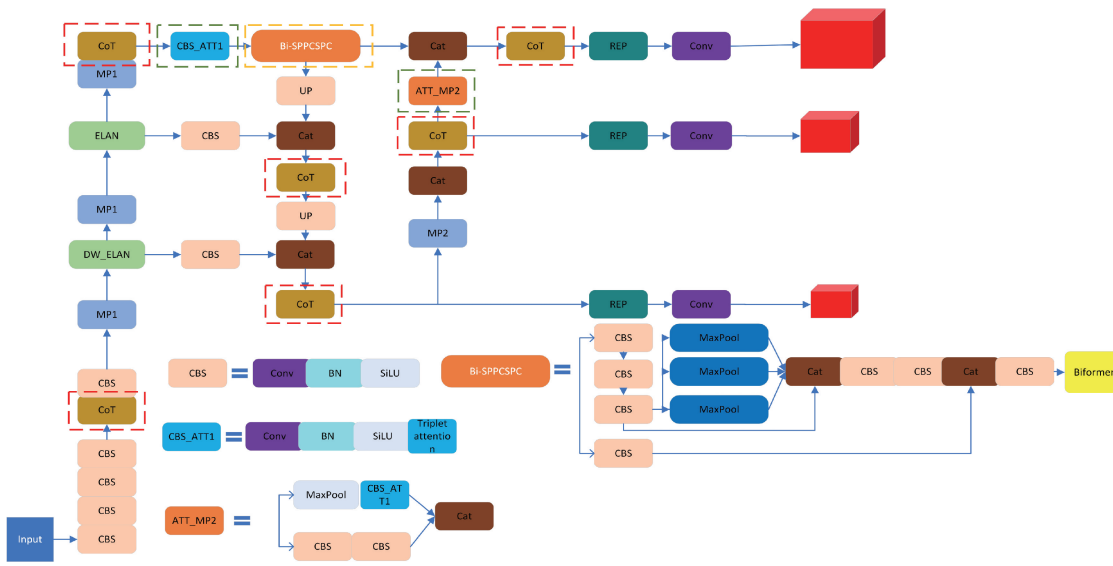


图 2　改进 YOLOv7 网络结构

入序列中捕捉长距离的依赖关系。在 YOLOv7 中引入自注意力双向 Transformer 模块可以帮助模型更好地理解物体之间的关联性和上下文信息，从而提升目标检测的准确性。在 YOLOv7 中加入自注意力双向 Transformer 模块可以使模型更加细致地感知目标周围的环境和语境，有助于减少误检和漏检，提高检测的精确性。自注意力双向 Transformer 模块利用了 Transformer 的自注意力机制，能够学习全局上下文信息，增强了模型的学习能力。将自注意力双向 Transformer 模块（Biformer）引入 YOLOv7 可以提升模型对分心驾驶中复杂场景和遮挡目标的处理能力，使得模型更具鲁棒性。

Biformer 块的细节结构如图 3 所示。先使用 3×3 卷积隐式编码相对位置信息。然后，依次使用 BRA 模块和一个 2 层的膨胀率为 e 的多层感知机（multi-layer perceptron，MLP）模块进行跨位置关系建模和逐位置嵌入。具体来说，自注意力双向 Transformer 在第 1 阶段使用重叠块嵌入，在第 2 到第 4 阶段使用块合并模块来降低输入空间分辨率，同时增加通道数，然后是采用连续的自注意力双向 Transformer 块做特征变换。需要注意的是，在每个块的开始均是使用 3×3 的深度卷积来隐式编码相对位置信息。随后依次应用 BRA 模块和扩展率为 e 的 2 层的 MLP 模块，分别用于交叉位置关系建模和每个位置嵌入。首先，有输入向量 $\boldsymbol{X}$，其中 $\boldsymbol{X}$ 是一个形状为（N，D）的矩阵，其中 N 是输入序列的长度，D 是每个序列元素的维

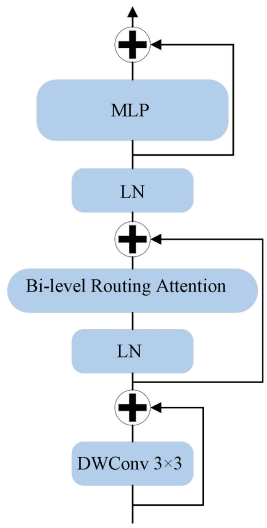


图 3　Biformer 结构

度。接下来，定义查询(Q)、键(K)和值(V)的投影权重，这些权重分别表示为 $\boldsymbol{W}^q$、$\boldsymbol{W}^k$ 和 $\boldsymbol{W}^v$，形状为(C,C)的矩阵，其中 C 是投影后的向量维度。然后，将输入序列 X 进行投影，得到投影后的 $\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$：

$$\boldsymbol{Q} = X^r \cdot \boldsymbol{W}^q \tag{1}$$

$$\boldsymbol{K} = X^r \cdot \boldsymbol{W}^k \tag{2}$$

$$\boldsymbol{V} = X^r \cdot \boldsymbol{W}^v \tag{3}$$

式中：$\boldsymbol{W}^q$、$\boldsymbol{W}^k$ 和 $\boldsymbol{W}^v$ 分别是 query、key、value 的投影权重。接下来，计算注意力权重 $\boldsymbol{A}$。注意力权重表示了查询与键之间的相似度，它可以通过将查询与键进行点积，并进行归一化来计算。可以使用 softmax 函数来实现归一化，如下式所示：

$$\boldsymbol{A} = softmax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{D_k}}\right)\boldsymbol{V} \tag{4}$$

其中，D_k 是键的维度，softmax 函数将注意力权重限制在 0～1 之间，并确保所有权重的总和等于 1。最后，可以使用注意力权重 $\boldsymbol{A}$ 来加权求和值 $\boldsymbol{V}$，从而得到 Biformer 的输出 $\boldsymbol{O}$。可以使用矩阵乘法来实现加权求和，如下式所示：

$$\boldsymbol{O} = \boldsymbol{A}\boldsymbol{V} \tag{5}$$

式(5)的乘法是矩阵乘法，将每个注意力权重与相应的值进行加权求和。

2.2　语义关联增强机制

目前的注意力机制几乎都需要一定大小的参数量才能建立通道之间的相互依赖，例如 CBAM[20] 和 SENet[21]，这些注意力虽然能提升模型检测精度，但是也会大幅增加模型参数量。所以本文引入了一种几乎不需要参数的 Attention。Triplet Attention 是一种基于注意力机制的模块，可以强化目标之间的语义关系。通过在 YOLOv7 中引入 Triplet Attention，模型可以对目标进行更准确的分组和聚合，从而提高目标检测的准确性和目标之间的关联性。Triplet Attention 可以自适应地调整特征的权重，增加重要目标特征的提取能力。在 YOLOv7 中应用 Triplet Attention，可以提升模型对目标特征的表达能力，使得模型更准确的区分目标和背景，提高检测的准确度。在 YOLOv7 中插入 Triplet Attention 可以在几乎不参加参数量的同时提升模型性能。图 4 所示为 Triplet Attention 网络结构，其中 Z-pool 数学公式如下所示：

$$Z\text{-}pool(\chi) = [MaxPool_{0d}(\chi), AvgPool_{0d}(\chi)] \tag{6}$$

式中：0d 是发生最大池化操作和平均池化操作的第 0 维度。从左往右第 1 个分支为空间注意力计算分支，输入特征经过 Z-Pool，再接着 7×7 卷积，最后 Sigmoid 激活函数生成空间注意力权重。第 2 个分支是通道 C 和空间 W 维度交互捕获分支，输入特征先经过 permute，变为 $H\times C\times W$ 维度特征，接着在 H 维度上进行 Z-Pool，后面操作类似。最后需要经过 permuter 变为 $C\times H\times W$ 维度特征，方便进行 element-wise 相加。第 3 个分支是通道 C 和空间 H 维度交互捕获分支，输入特征先经过 permute，变为 $W\times H\times C$ 维度特征，接着在 W 维度上进行 Z-Pool，后面操作类似。最后需要经过 permuter 变为 $C\times H\times W$ 维度特征，方便进行 element-wise 相加。最后对 3 个分支输出特征进行相加求 Avg 并得到最后的输出。

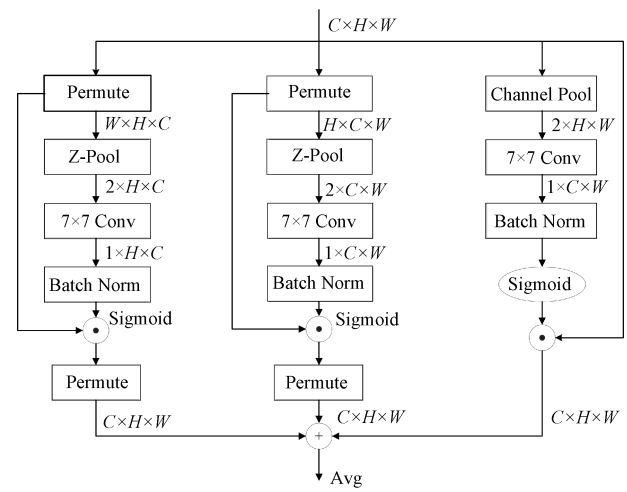


图 4　Triplet Attention 网络结构

2.3　语义上下文增强模块

现有的驾驶员分心识别方法通常忽略了图像中的上下文语义关系，导致识别结果不准确或不稳定。为了解决这个问题，提出了一种新的神经网络结构，为了提高视觉识别任务的性能，通过结合不同层次的上下文信息，称为 CoT，它可以在多个尺度上捕获和利用上下文信息。CoT 使用 Transformer 模块来编码和解码图像中的局部特征，并使用注意力机制来聚合不同层次的上下文信息。

本文设计了一个新颖的 Transformer 风格模块，即语义上下文增强块，用于视觉识别。该模块结构如图 5 所示。这种设计充分利用了输入键之间的上下文信息来指

导动态注意力矩阵的学习，从而增强了视觉表示的能力。从技术上讲，语义上下文增强块首先通过 3×3 卷积对输入键进行上下文编码，使输入的静态上下文表示。进一步将编码密钥与输入查询连接起来，通过两个连续的 1×1 卷积来学习动态多头注意力矩阵。学习的注意力矩阵乘以输入值，以实现输入的动态上下文表示。最终将静态和动态上下文表示的融合作为输出。

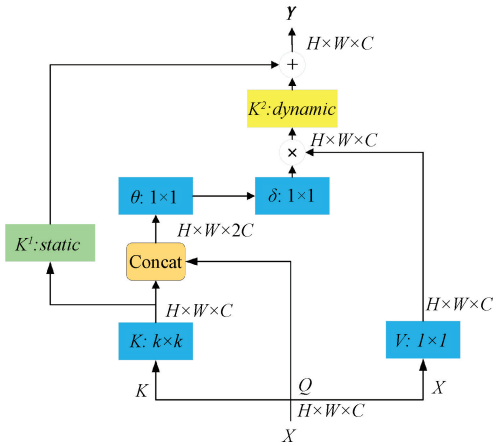


图 5　CoT 网络结构

3　实验与结果分析

3.1　实验环境

本文实验环境如表 1 所示。

表 1　实验环境配置

设备名称	设备信息
操作系统	Windows11
CPU	i5-12500H
GPU	RTX3060(6.0 GB)
Python 版本	3.9
CUDA 版本	12.0
Pytorch 版本	1.11.0

设置批量大小为 4，训练 100 个 epoch，并将权重衰减设置为 0.000 5，以防止过拟合。训练网络的学习率为 1×10^{-5}。此外，采用了多学习率策略，其中当前学习率会在训练的每个迭代中乘以一个系数，以实现更稳定的收敛。

3.2　实验数据集

本文实验所用的数据集为互联网自收集数据集，主要用来对驾驶员分心检测来进行训练。此数据集总共包含 3 643 张图片，总共包含 4 种类别，分别是 smoke、face、drink、phone。按照 8.2∶1∶0.8 划分训练集、测试集、验证集，其中训练集 3 003 张，测试集 356 张，验证集 284 张。数据集示例图如图 6 所示。

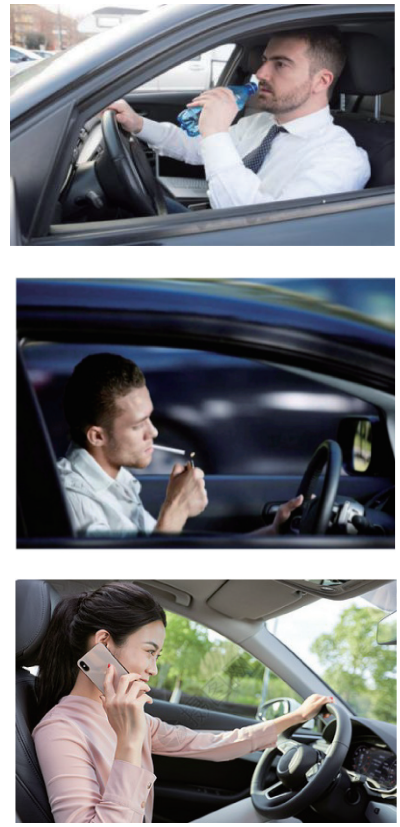


图 6　数据集示例图

3.3　实验评价指标

为了证明此模型的优势，使用了如下判断指标：mAP、参数量(Params)、每秒传输帧数。其中 mAP 公式如下：

$$Precision(P)=\frac{TP}{TP+FP} \tag{7}$$

$$Recall(R)=\frac{TP}{TP+FN} \tag{8}$$

$$AP_i=\int_0^1 P_i(R_i)\mathrm{d}R_i \tag{9}$$

$$mAP=\frac{1}{n}\sum_{i=1}^{n}AP_i \tag{10}$$

式中：TP 为真阳性样本；FP 为假阳性样本；FN 为假阴性样本。另外，P 表示整个预测结果中真正预测的个数，R 表示所有基础真理中真正预测的个数，N 是总类别。AP 通过考虑 P 和 R 指标来评估每个类别的模型性能。mAP 表示用来衡量目标检测算法的整体检测精度。

3.4　实验结果分析

当进行分心驾驶检测任务时，本文引入了基于上下文语义增强联合 YOLOv7 的分心驾驶检测算法，以提高检测性能。为了深入了解这一模型的贡献和优势，本文进行了一系列消融实验，逐步评估其各个组成部分的影响。如表 2 所示，本文将提出的各个模块进行了组合实验，以全面验证本文所提方法的有效性能。本文在 baseline 算法的基础上得到了原始的 mAP 指标的值为 83.0%。在此基础上，本文分别引入了自注意力双向 Transformer 模

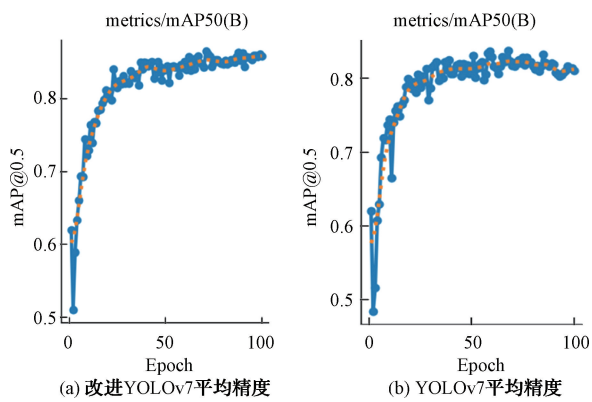
块，语义上下文增强模块和语义关联增强机制，得到的mAP值在原模型基础上分别提升了1.1%、0.9%、1.2%。进一步本文引入了自注意力双向Transformer模块和语义关联注意力机制得到的mAP指标的值为86.4%，在baseline的基础上提升了3.4%。更进一步的，本文引入语义关联注意力机制和语义上下文增强模块，这在YOLOv7原模型上提升了2.8%。当同时添加表2中3个模块时，可以看出，对原网络有显著提升，整体提升了4.3%。实验结果表明，本文方法的各个模块都具有明显的提升，证明了对分心行为检测的有效性。

表2 消融实验

Algorithms	Biformer	CoT	Triplet	mAP@0.5/%
YOLOv7				83.0
YOLOv7-B	√			84.1
YOLOv7-C		√		83.9
YOLOv7-T			√	84.2
YOLOv7-BT	√		√	86.4
YOLOv7-CT		√	√	85.8
本文算法	√	√	√	87.3

3.5 实验对比

为了验证网络改进效果，对原YOLOv7网络和改进后的mAP@0.5的数据进行对比，结果如图7所示，图7(a)为改进YOLOv7平均精度，图7(b)为YOLOv7平均精度。从图7可知，改进后的YOLOv7比原YOLOv7模型平均精度更高，并且模型训练曲线更加平滑，在分心驾驶检测中更具优势。

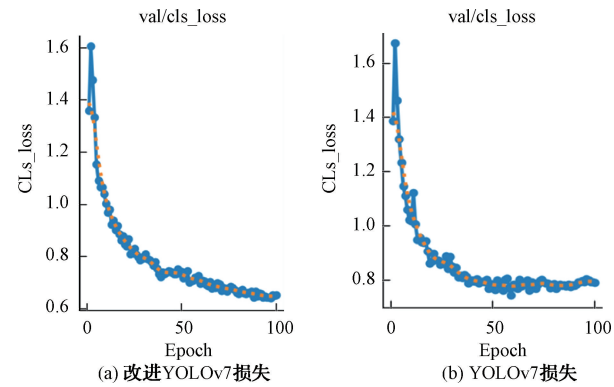


(a) 改进YOLOv7平均精度　(b) YOLOv7平均精度

图7 平均精度对比

损失对比结果如图8所示，图8(a)为改进YOLOv7损失，图8(b)为原YOLOv7损失。从图8可知，改进后的YOLOv7网络验证损失值始终低于原网络，并且曲线更加平滑，说明改进后的网络模型性能高于原网络模

为了进一步验证算法的有效性，将本文改进后的算法与DAMO-YOLO-M、DETR、RT-DETR、YOLOv3、YOLOv4、YOLOv5m、YOLOv7、YOLOv8m等主流目标



(a) 改进YOLOv7损失　(b) YOLOv7损失

图8 验证损失对比

检测算法在分心行为数据集进行Params、mAP和帧率指标进行对比，实验结果如表3所示。通过对比不同算法的实验结果，本文提出的改进YOLOv7模型的mAP相较于其他YOLO系列算法大幅提升，达到了87.3%。本文算法相较于基于Transformer的DETR提升了2.6%，并且检测速度大幅提升了64%。相较于改进后的RT-DETR，本文算法的检测精度虽然下降了1.2%，但是参数量也下降了55%，并且检测速度也提升了16.7%。本文所提出的方法相较于原本YOLOv7模型牺牲了小幅的检测速度，但是其在降低参数量的同时，大幅提升了检测精度，达到了检测精度和速度的平衡。虽然本文改进的算法模型参数量相较于最新改进的DAMO-YOLO-M提高了18.9%，但是本文算法的检测精度比DAMO-YOLO-M提升了0.5%，同时检测速度也要优于DAMO-YOLO-M。实验证明，本文的方法在与最先进的各个方法的对比中，具有显著性的优势，这足以证明本文方法的有效性能。

表3 对比试验

methods	Params /($\times10^6$)	mAP@0.5/%	帧率 /fps
YOLOv3	61.5	82.1	30
YOLOv4	27.6	81.8	58
YOLOv5m	21.2	82.6	54
YOLOv7	36.5	83.0	95
YOLOv8m	25.9	82.3	65
DETR	159.0	84.7	32
RT-DETR[22]	76.5	88.5	75
DAMO-YOLO-M[23]	28.2	86.8	86
本文算法	34.8	87.3	90

为了更直观体现实验效果，分别使用YOLOv7和本文改进的方法对分析驾驶图像进行可视化，如图9所示，图9(a)为改进YOLOv7检测图像，图9(b)为YOLOv7检测图像。从图9可以看出，改进后的YOLOv7相较于原

(a) (b)

图 9 检测结果对比

模型检测的置信度更高。这证明改进后的 YOLOv7 模型有更强的特征提取能力和分析能力，检测的精度更高。

4 结 论

本文以驾驶员分心行为检测应用为背景，针对目标检测中分心驾驶检测精度低、漏检错检率高等尚不完善的问题，对基于上下文语义增强联合 YOLOv7 方法进行实验研究，通过替换掉原有的 ELAN 模块改为语义上下文增强模块、融合自注意力双向 Transformer 模块和 SPPCSPC 模块、语义关联增强机制融合卷积层来优化原先基础网络 YOLOv7。实验结果表明，改进的 YOLOv7 相较于原始网络，在分心行为检测数据集上检测精度达到了 87.3%，提升了 4.3%，在精度提升的同时减少了参数量，检测速度达到了 90 fps，满足实时检测需求。随着研究领域的发展，未来的工作可以进一步关注改进方法的性能、可扩展性和实时性能，并将对本文算法在嵌入式设备上进行部署，进一步提升模型的实用性，以促进道路环境的安全性。

参 考 文 献

[1] PRAVEENA R, BABU T R G, SUDHA G, et al. Distracted driver detection using deep learning classifier of image net models[C]. 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS). IEEE, 2022: 1-5.

[2] 史浩琛，金致远，唐文婧，等. 基于深度学习的高精度晶圆缺陷检测方法研究[J]. 电子测量与仪器学报，2022，36(11)：79-90.

[3] 冯笑，代少升，黄炼. 基于可解释深度学习的单通道脑电跨被试疲劳驾驶检测[J]. 仪器仪表学报，2023，44(5)：140-149.

[4] 张彦生，王成龙，刘远红. 基于深度学习的绝缘子故障检测研究[J]. 电子测量技术，2023，46(8)：105-111.

[5] 陈娟，李燕，阚希，等. 基于 EfficientNet 的轻量化行人检测算法[J]. 国外电子测量技术，2023，42(6)：1-9.

[6] REDMON J, DIVVALA S K, 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.

[7] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multi box detector [C]. European Confe-

rence on Computer Vision. Cham: Springer,2016:21-37.

[8] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2015:1440-1448.

[9] REN S Q, HE K M, GIRSHICK R, et al. Faster RCNN: Towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems,2015,28:91-99.

[10] HE K, ZHANG X Y, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2015, 37(9):1904-1916.

[11] 尹智帅,钟恕,聂琳真,等. 基于人体姿态估计的分心驾驶行为检测[J]. 中国公路学报,2022,35(6):312-323.

[12] 曹立波, 杨洒, 艾昌硕, 等. 基于深度学习的分心驾驶行为检测方法[J]. 汽车技术, 2023(6):49-54.

[13] 杜虓龙,余华平. 基于改进 Mobile Net-SSD 网络的驾驶员分心行为检测[J]. 公路交通科技,2022,39(3):160-166.

[14] 魏启康, 朱文忠, 江嘉文, 等. 基于改进 YOLOv4-tiny 的分心驾驶行为检测[J]. 四川轻化工大学学报(自然科学版), 2023, 36(2):67-76.

[15] MOU L, CHANG J, ZHOU C, et al. Multimodal driver distraction detection using dual-channel network of CNN and Transformer[J]. Expert Systems with Applications, 2023, 234: 121066.

[16] LI Y, YAO T, PAN Y, et al. Contextual Transformer networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2022, 45(2): 1489-1500.

[17] MISRA D, NALAMADA T, ARASANIPALAI A U, et al. Rotate to attend: Convolutional triplet attention module[C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 3139-3148.

[18] ZHU L, WANG X, KE Z, et al. BiFormer: Vision transformer with bi-level routing attention[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2023: 10323-10333.

[19] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2023: 7464-7475.

[20] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.

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

[22] LV W, XU S, ZHAO Y, et al. Detrs beat YOLOs on real-time object detection[J]. Computer Science, 2023, DOI:10.48550/arXiv.2304.08069.

[23] XU X, JIANG Y, CHEN W, et al. Damo-YOLO: A report on real-time object detection design[J]. Computer Science, 2022, DOI: 10.48550/arXiv.2211.15444.

作者简介

李富,硕士生导师,高级工程师,主要研究方向为嵌入式系统设计,图像处理。

E-mail:lifu@cwxu.edu.cn

徐凯,硕士研究生,主要研究方向为计算机视觉,车联网。

E-mail:202212490529@nuist.edu.cn

王泉(通信作者),博士,正高级工程师,主要研究方向为通信工程,智能制造。

E-mail:seepwq@163.com