Software Engineering and Applications 软件工程与应用, 2025, 14(5), 1035-1044
Published Online October 2025 in Hans. https://www.hanspub.org/journal/sea
https://doi.org/10.12677/sea.2025.145092

Hans汉斯

改进YOLO11n的疲劳驾驶目标检测方法

王业童，张丽艳*

大连交通大学轨道智能工程学院电子通信系，辽宁 大连

收稿日期：2025年7月29日；录用日期：2025年9月19日；发布日期：2025年9月28日

摘　要

近年来，随着汽车数量不断增多，因疲劳驾驶导致的交通事故频发。针对当前目标检测算法准确率不足、鲁棒性差、尺寸大等问题，本文基于YOLO11n提出一种用于疲劳驾驶分析的高效面部状态检测模型。本文首先针对因眼部、嘴部等小目标容易受到分辨率、角度偏移、遮挡等因素的影响，在C3K2特征提取单元中引入非对称填充的风车形卷积PSConv，使模块在提取特征时能够捕捉更广域的上下文信息；其次，针对原始模型C3PSA中自注意力机制消耗较高的计算资源问题，引入基于统计学的新型注意力算子TSSA，通过对token特征的统计分析来有效捕捉特征、精准聚焦目标区域，同时引入Dynamic Tanh作为注意力机制中的归一化层，在无需多余计算资源的基础上为模型添加非线性归一化操作，提升检测精度、减小模型尺寸、提高模型鲁棒性；最后，为进一步降低模型参数量，引入轻量级共享卷积头，并在此基础上集成细节增强卷积，补偿模型检测精度。本文在公开疲劳驾驶数据集上进行有效性验证实验，相较于基线模型，改进模型在检测准确率方面，mAP50提升1个百分点、mAP50-95提升4.1个百分点；模型参数量降低近23%；帧率提升7帧；在模块改进层面完成了轻量化和检测精度水平的优化任务，可以为进一步疲劳驾驶研判提供高精度的特征信息。

关键词

疲劳驾驶，YOLO11n，自注意力机制，细节增强卷积

Improved YOLO11n Fatigue Driving Target Detection Method

Yetong Wang, Liyan Zhang*

Department of Electronic Communications, School of Railway Intelligent Engineering, Dalian Jiaotong University, Dalian Liaoning

Received: July 29, 2025; accepted: September 19, 2025; published: September 28, 2025

*通讯作者。

文章引用: 王业童, 张丽艳. 改进 YOLO11n 的疲劳驾驶目标检测方法[J]. 软件工程与应用, 2025, 14(5): 1035-1044.
DOI: 10.12677/sea.2025.145092

Abstract

In recent years, with the increasing number of cars, traffic accidents caused by fatigue driving have become frequent. In response to the problems of insufficient accuracy, poor robustness, and large size of current object detection algorithms, this paper proposes an efficient facial state detection model for fatigue driving analysis based on YOLO11n. This article first addresses the impact of factors such as resolution, angle offset, and occlusion on small targets such as the eyes and mouth. In the C3K2 feature extraction unit, an asymmetric filled windmill shaped convolution PSConv is introduced to enable the module to capture a wider range of contextual information when extracting features; Secondly, in response to the high computational resource consumption of the self attention mechanism in the original model C3PSA, a new attention operator TSSA based on statistics is introduced to effectively capture features and accurately focus on the target area through statistical analysis of token features. At the same time, Dynamic Tanh is introduced as the normalization layer in the attention mechanism, which adds nonlinear normalization operations to the model without unnecessary computational resources, improves detection accuracy, reduces model size, and enhances model robustness; Finally, to further reduce the number of model parameters, a lightweight shared convolution head is introduced, and on this basis, detail enhanced convolution is integrated to compensate for model detection accuracy. This article conducted effectiveness verification experiments on a publicly available fatigue driving dataset. Compared to the baseline model, the improved model improved detection accuracy by 1 percentage point for mAP50 and 4.1 percentage points for mAP50-95; The number of model parameters decreased by nearly 23%; Frame rate increased by 7 frames; The optimization tasks of lightweighting and detection accuracy have been completed at the module improvement level, which can provide high-precision feature information for further fatigue driving analysis.

Keywords

Fatigue Driving, YOLO11n, Self Attention Mechanism, Detail Enhanced Convolution

1. 引言

根据国家统计局数据显示，截至 2023 年全国民用汽车拥有量已超 3.29 亿辆，汽车驾驶员人数超 4.64 亿人。伴随着汽车总量的不断增加，公路交通安全问题越来越凸显，交通事故也随之频发。其中，与驾驶员密切相关的交通事故占比近九成。与驾驶员主观因素行为相关的疲劳驾驶极易引发交通事故，进而造成人员伤亡和经济损失。对驾驶员的疲劳状态进行实时的监测，并及时进行预警，在源头上预防交通事故的发生诱因，可以一定程度上提高公路交通安全。

随着人工智能技术的不断发展，基于计算机视觉的检测方法直接通过车辆内部摄像头实时采集驾驶员包括其头部、面部等部位的图片，利用深度学习算法检测驾驶员的驾驶行为，具有更强的实用性和更高的灵活性。2023 年，张育榕等人[1]基于 RetinaFace 进行人脸检测，在头部、眼部、嘴部等多区域使用 CNN 提取局部特征，融合之后通过神经网络判定疲劳，眼部检测准确率为 97.1%、嘴部为 97.5%、头部为 88.1%，性能与容错并存；2024 年，Abderrahim 等人[2]研究使用 CNN 提取面部深层特征(眼、眉、嘴、表情)，计算五类结构特征；结合全局 CNN 特征通过 AlexNet＋LSTM 实现帧级到视频级疲劳分类，提升

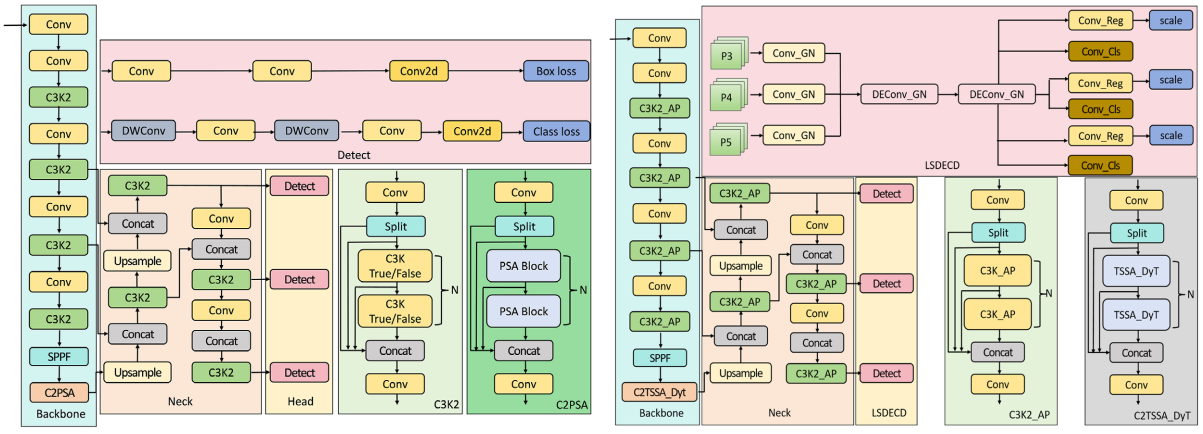
识别稳健性；2025 年，杜威等人[3]提出一种改进 YOLOv5s 的疲劳驾驶目标检测算法，使用轻量的 EfficientNet 骨干网络作为 YOLOv5s 的主干网络来进行特征提取，同时选用 SIoU 作为模型的损失函数，参数量和准确率实现小幅优化；Ding [4]提出“VA-YOLO”的目标检测算法，采用了轻量级的主干网络 Vanilla Net 来替代传统的复杂骨干网络，引入了 SE 注意力机制和 SIoU 损失函数，在精度上提升了 5.3% 同时参数量下降了 1.01 M。尽管当前基于 YOLO 的目标检测算法对于眼部、嘴部小目标区域检测的检测精度已经处于较高水平，但当图像分辨率低、光照不足及驾驶员面部存在遮挡等特殊情况出现时，模型检测精度会受到影响。另外，为保证模型在终端设备流畅运行，还需进一步缩小模型尺寸。

2. 本文算法原理

2.1. 改进的 YOLO11n 面部疲劳特征目标检测模型

YOLO11 [5]用于目标检测任务的网络结构由三部分组成，分别是主干网络、颈部网络以及检测头网络。主干网络用于特征提取，颈部网络用于特征融合，检测头网络用于实现最后的目标检测任务。YOLO11 主干网络和颈部网络中，有三大核心模块，分别是 C3K2、SPPF 以及 C2PSA：(1) C3K2 模块作为核心特征提取模块在 C2f 模块基础之上引入多尺度卷积核 C3K，可调卷积核的设计能够保持丰富的特征表达、捕捉更广泛的上下文信息，在复杂场景检测中表现出色。(2) 在主干网络末端，YOLO11 的 SPPF 模块(快速空间金字塔池化模块)，引入多尺度特征图池化，扩大网络感受野，捕捉不同尺度特征。(3) 在 SPPF 后引入 C2PSA 模块。C2PSA 模块包含金字塔压缩注意力 PSA，并配合前馈网络 FFN 对特征进行重新加权，增强特征捕捉能力。检测头网络是一种解耦的双塔结构，由分类分支和回归分支组成，能够在保证高精度的前提下，降低计算成本，提升检测实时性。

当面部存在遮挡及光照条件不足等特殊情况出现时，YOLO11n 在检测精准率上存在不足，为此本文提出一种改进 YOLO11n 的面部疲劳特征检测模型，有效的提高了模型的检测精度，同时，进一步对模型进行轻量化处理，改进网络结构如图 1 所示。本文的改进包括以下四点：(1) 将 YOLO11n 模型中的 C3K2 模块中引入非对称填充卷积成为 C3K2_AP 模块，实现卷积过程感受野的扩大，降低因分辨率低、角度偏移、遮挡等因素对面部疲劳特征提取效果造成的影响，完成高效的特征提取；(2) 在 C2PSA 模块中引入一种基于令牌统计的自注意力机制成为 C2TSSA_Dyt 模块，有效捕捉特征、精准聚焦目标区域，同时加入归一化操作，增强鲁棒性；(3) 在轻量级共享卷积检测头基础上引入细节增强卷积优化模型尺寸。



(a) (b)

Figure 1. Structure diagram of the YOLO11n network and its improved model. (a) YOLO11n; (b) Improved model
图 1. YOLO11n 网络及其改进模型结构图。(a) YOLO11n 模型；(b) 改进的模型

2.2. 基于非对称填充卷积的 C3K2_AP 模块

为进一步增强原模型的 C3K2 模块对眼部、嘴部小目标区域的特征提取能力，本文引入基于非对称填充的风车形卷积(Pinwheel-shaped Convolution, PSConv) [6]替换普通卷积，如图 2 所示。

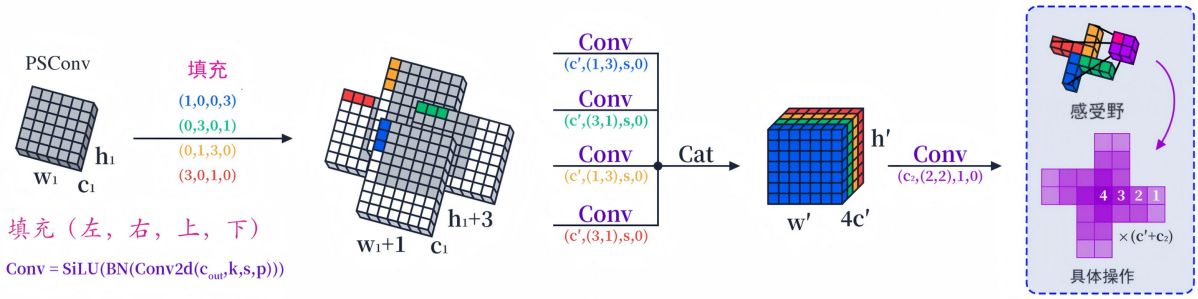


Figure 2. Structure diagram of PSConv
图 2. PSConv 结构图

PSConv 是一种四向不对称卷积结构可以大致分为四分支平行非对称卷积、整合、降维三部分。设输入特征图为 $X^{(h_1,w_1,c_1)}$，其中 h_1、w_1、c_1 表示高度、宽度、通道数。首先，对特征图 $X^{(h_1,w_1,c_1)}$ 同时用四个不对称并行卷积分支处理，每个分支中都使用不同形状的非对称填充和一维卷积核。填充参数为 $P_{(l,r,t,b)}$，其中的 l、r、t、b 分别表示左、右、上、下方向上填充的像素数；一维卷积核 K_i 在水平、垂直方向分别采用 1 × 3 和 3 × 1 两种形式，前者旨在水平方向上聚焦左右信息，后者旨在垂直方向上聚焦上下信息。四分支卷积构造过程如公式(1)，其中，BN 表示批归一化，$SiLU$ 表示激活函数，W_i 表示卷积核，当 $i = 1$ 或 3 时，卷积核尺寸为 1 × 3；当 $i = 2$ 或 4 时，尺寸为 3 × 1。非对称卷积的输出尺寸变化如式(2)所示，其中，s 是卷积步长，c_2 是输出特征图的通道数；其次，对四个不对称并行卷积分支的输出结果进行通道层面的拼接，如式(3)所示。最后，将拼接后的输出张量通过尺寸为 $K^{(2,2,c_2)}$ 的卷积核归一化处理，且不进行填充，输出为 $Y^{(h_2,w_2,c_2)}$ 如式(4)，尺寸表示如式(5)所示。

$$X_i^{(h',w',c')} = SiLU\left(BN\left(X_{P_{i(l,r,t,b)}}^{(h_1,w_1,c_1)} \otimes W_i\right)\right), i = 1,2,3,4 \tag{1}$$

$$h' = \frac{h_1}{s} + 1, w' = \frac{w_1}{s} + 1, c' = \frac{c_2}{4} \tag{2}$$

$$X'^{(h',w',c')} = Concat\left(X_{1,\cdots,4}^{(h',w',c')}\right) \tag{3}$$

$$Y^{(h_2,w_2,c_2)} = SiLU\left(BN\left(X'^{(h',w',c')} \otimes K^{(2,2,c_2)}\right)\right) \tag{4}$$

$$h_2 = h' - 1 = \frac{h_1}{s}, w_2 = w' - 1 = \frac{w_1}{s} \tag{5}$$

此外，本文在 PSConv 基础上引入 APBottleneck 瓶颈结构的设计思想，替换 C3K 模块形成 C3K2_AP 模块，在对四个分支进行非对称填充卷积基础上结合分组卷积，将四组通道独立进行卷积操作，各组的输出做通道拼接得到最终输出。相比于原始 PSConv 中每个分支的卷积操作处理完整输入通道的过程，分组卷积能够在一定程度上减少参数量，提高计算效率。在原始 YOLO11n 的核心特征提取模块 C3K2 中引入 PSConv，在水平和垂直两个主方向上突出中心像素与其周围像素的梯度变化；同时，实现了感受野的有效扩大，在保持模型轻量的前提下，兼顾上下文信息，有助于提升眼部、嘴部小目标的检测准确率。

2.3. 基于令牌统计自注意力的动态调整并行卷积块

本文在基于统计学的新型注意力算子(Token Statistics Self-Attention, TSSA) [7]模块基础上引入 DyT，形成 TSSA_DyT 模块用于增强改进网络的特征提取过程，如图 3 所示。用 TSSA_DyT 替换原始 C2PSA 模块中的 PSA 模块，组成基于令牌统计自注意力的动态调整并行卷积块 C2TSSA_DyT。一方面，TSSA 将原始自注意力中每个 token 两两配对，替换为在若干低秩子空间上做统计并回投，实现了对全局空间的二阶统计，自适应聚焦关键 token，极大地增强了模型对驾驶员面部信息的捕捉能力；另一方面，DyT 简化了归一化过程，保证了极值抑制与梯度稳定，使得整个检测网络在复杂背景以及轻微遮挡环境下表现出更强的鲁棒性。两者协同作用，使得改进模型相较于 YOLO11n 在检测精度和模型尺寸控制上都有提升。

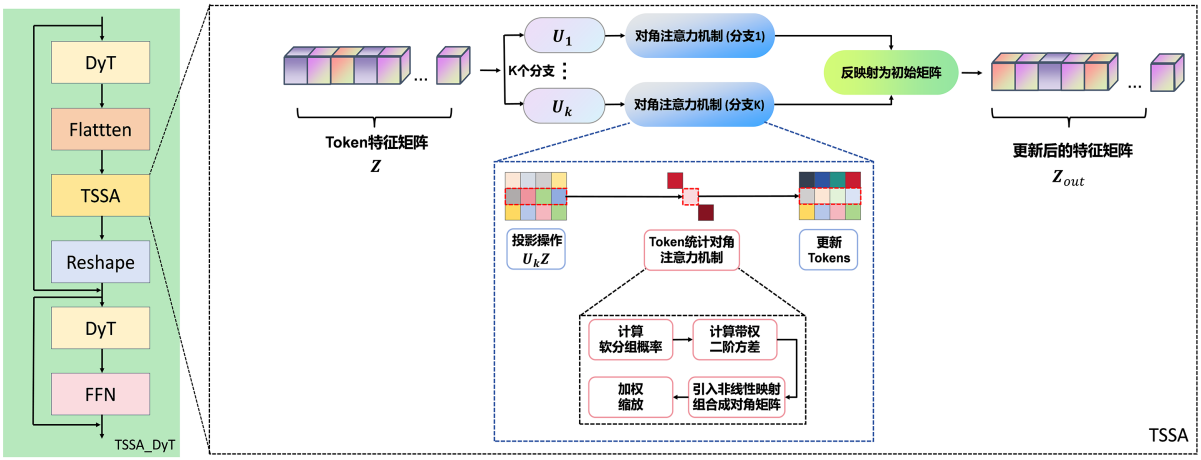


Figure 3. TSSA_DyT module
图 3. TSSA_DyT 模块

如图 3 右半部分所示，TSSA 将原始特征投影到若干子空间，以便于把眼、嘴相关方向进行聚合，用加权二阶统计识别每个子空间内真正有信息的通道，并通过对角权重放大重要方向、抑制噪声，把加权结果回投并按软分组概率融合到原特征，从而放大如眨眼、打哈欠等微小信号，同时显著降低注意力的时间与内存开销。具体来说，TSSA 模块核心在于无需构造 $n\times n$ 的点积相似度矩阵，而是将 d 维特征投影至若干低秩子空间，采用加权方差来判断各通道的重要性，最后通过软分组概率融合多头结果。设输入 token 特征矩阵为 $Z=[z_1,\cdots,z_n]\in\mathbb{R}^{d\times n}$，其中 d 表示通道维度、n 为 token 数量。TSSA 预定义 K 个可学习的低秩投影矩阵 $U_k\in\mathbb{R}^{d\times p}$，$K=1,2,\cdots,K$，其中 p 表示每个子空间投影后的维度，并且 $p\ll d$。所有 token 的投影表示为式(6)。在每个投影子空间中，TSSA 首先需要计算软分组概率 $\Pi\in\mathbb{R}^{n\times K}$ 如式(7)所示，使得每个 token 可以根据其所处子空间的能量进行动态分配。其中，j 表示第 j 个 token，ℓ 表示第 ℓ 个子空间，$\eta>0$ 是可学习的温度参数，$\left\|\mathrm{U}_k^{\top}z_j\right\|_2^2$ 衡量第 j 个 token 在子空间 U_k 上的投影能量，$\Pi_{j,k}$ 表示第 j 个 token 属于组 k 的概率。若某个 token 在第 k 个子空间上的能量较大，说明其对该子空间特征较相关，对应的软分组概率 $\Pi_{j,k}$ 较高。获得软分组概率后，在子空间 k 中进一步计算带权二阶方差如式(8)和式(9)所示。其中，$\tilde{Z}_k^{\odot 2}\in\mathbb{R}^{p\times n}$ 表示对投影矩阵 $\widetilde{Z_k}$ 中的每个元素逐元素平方，$\Pi_{:,k}\in\mathbb{R}^n$ 表示第 k 列软分组向量，$\mathrm{x}_{k,i}$ 表示第 i 个投影通道在子空间 k 上的加权方差。

$$\widetilde{Z_k}=U_kZ\in\mathbb{R}^{p\times n} \tag{6}$$

$$\Pi_{j,k}=\frac{\exp\left(\frac{1}{2\eta}\left\|\mathrm{U}_k^{\top}z_j\right\|_2^2\right)}{\sum_{\ell=1}^{K}\exp\left(\left\|\mathrm{U}_\ell^{\top}z_j\right\|_2^2\right)},\sum_{k=1}^{K}\Pi_{j,k}=1, \tag{7}$$

$$\mathrm{x}_k=\left(\mathrm{x}_{k,1},\mathrm{x}_{k,2},\cdots,\mathrm{x}_{k,p}\right)^{\top}=\tilde{Z}_k^{\odot 2}\Pi_{:,k}\in\mathbb{R}^p \tag{8}$$

$$\mathrm{x}_{k,i}=\sum_{j=1}^{n}\left(\widetilde{Z_k}[\mathrm{i},\mathrm{j}]\right)^2\Pi_{j,k} \tag{9}$$

为突出高方差、信息量大的通道并抑制小方差、存在噪声的通道，引入非线性映射$\nabla f\left(x_{k,i}\right)$，并将其组合成为一个对角矩阵$D_k$如式(10)和式(11)所示。由此可知，若某一投影通道加权方差很大，则$\nabla f\left(x_{k,i}\right)$较小，大方差方向得以保留；若加权方差很小，则$\nabla f\left(x_{k,i}\right)$近似为 1，对该通道仅做弱衰减。通过上述设计可以保留子空间内最具有判别力的方向，同时抑制不必要的噪声，实现对每个投影维度的加权收缩。接下来，将得到的对角矩阵D_k对投影特征$\widetilde{Z_k}$根据通道做加权缩放如式(12)所示，并将加权后的投影通过U_k反映射至原始的d维空间如式(13)。此时，即可得到子空间内的重要信息对原始特征的增强。

$$\nabla f\left(x_{k,i}\right)=\frac{1}{1+x_{k,i}},i=1,\cdots,p \tag{10}$$

$$D_k=diag\left(\nabla f\left(x_{k,1}\right),\cdots,f\left(x_{k,p}\right)\right)\in\mathbb{R}^{p\times p} \tag{11}$$

$$D_k\widetilde{Z_k}=\left[D_k\widetilde{Z_{k,1}},\cdots,D_k\widetilde{Z_{k,n}}\right]\in\mathbb{R}^{p\times n} \tag{12}$$

$$O_k=U_kD_k\left(U_k^{\top}Z\right)\in\mathbb{R}^{d\times n} \tag{13}$$

各个子空间并行不悖，共产生K个子空间的更新输出$O_k\in\mathbb{R}^{d\times n}$，在单个 token 维度上，对第$j$个 token 最终的融合输出设为$\widetilde{\mathrm{z}_j}$如式(14)，并将所有 token 以矩阵形式表示为$\hat{Z}$如式(15)。同时，为了进一步增强表达能力，使网络进一步微调融合权重，在融合操作前后引入一个可学习的线性映射$\mathrm{W}\in\mathbb{R}^{d\times d}$，最终引入残差结果定义 TSSA 的输出$Z_{out}\in\mathbb{R}^{d\times n}$如式(16)所示。本文采用上述的 TSSA 对 PSA 自注意力机制进行改进得到 C2TSSA 模块，使模型聚焦在眼部、嘴部这类微小变化上，同时可以减少因遮挡、光照等因素带来的误检问题，并降低注意力模块的计算开销。

$$\widehat{z_j}=\sum_{k=1}^{K}\Pi_{j,k}\left[O_k\right]_{:,j}=\sum_{k=1}^{K}\Pi_{j,k}U_kD_k\left(U_k^{\top}z_j\right) \tag{14}$$

$$\hat{Z}=\left[\widehat{z_1},\cdots,\widehat{z_n}\right]=\sum_{k=1}^{K}\left(U_kD_k\left(U_k^{\top}Z\right)\right)Diag\left(\Pi_{:,k}\right)\in\mathbb{R}^{d\times n} \tag{15}$$

$$Z_{out}=Z+W\hat{Z} \tag{16}$$

本文在注意力模块和前馈网络前加入归一化操作，引入 Dynamic Tanh (DyT)方法[8]对特征做归一化和非线性压缩操作，以此来抑制极端噪声、放大细微信号，进一步提升检测鲁棒性。DyT 定义如式(17)

$$DyT\left(x\right)=\gamma*\tanh\left(\alpha x\right)+\beta \tag{17}$$

其中，x是输入张量；α是可学习的标量参数，用于根据输入范围对输入进行缩放来适配不同的x尺度，使大多数元素集中在 tanh 的近似区间内，初始设置为 0.5；γ和β分别是与通道数相同的可学习向量，对应 LN 中的仿射缩放与偏移操作，允许网络在 tanh 压缩之后再恢复到适当的通道尺度与中心位置。DyT 的设计用一个标量参数α取代了层归一化中针对每个 token 计算均值、方差的过程，再通过γ和β完成

仿射变换，保持与层归一化相当甚至更优的收敛速度和泛化性能，在训练和推理阶段都可以节省计算资源。

2.4. 基于细节增强的轻量级共享卷积检测头

由于疲劳检测场景下对于模型检测的实时性和准确率要求极高，本文引入轻量级共享卷积检测头(Lightweight Shared Convolutional Detection Head, LSCD) [9]替代原模型的 head 检测部分，并将引入细节增强卷积(Detail-Enhanced Convolution, DEConv) [10]替代 LSCD 的两个共享卷积 Conv_GN 模块，重新设计了一个基于细节增强的轻量级共享卷积检测头(Lightweight Shared Detail Enhanced Convolutional Detection Head, LSDECD)如图 1 改进模型中的 LSDECD 模块所示。其中 DEConv 模块如图 4 所示。本文在 LSDECD 的两个 3 × 3 的共享卷积 Conv_GN 中引入 DEConv 组成 DEConv_GN，将 DEConv 应用在两次共享卷积阶段，相当于先对浅层特征进行高频升维，再通过第二次细节增强来融合全局与局部上下文，使 LSCD 并行学习边缘、纹理等高频细节，极大增强眼部小目标的可分辨性。其中模块 DEConv 将若干差分卷积与一条常规卷积并行，将其合并成一个等效卷积，以兼顾低频和高频信息，增强微小特征的识别能力，同时保持与普通卷积一致的计算开销。DEConv 由标准卷积(VC)、中心差分卷积(CDC)、角差分卷积(ADC)、水平差分卷积(HDC)以及垂直差分卷积(VDC)五个卷积分支组合成并行结构。CDC 和 ADC 分别关注图像的中心区域和角区域；HDC 和 VDC 分别关注图像水平和垂直方向上的特征。DEConv 的实现可由式(18)表示。

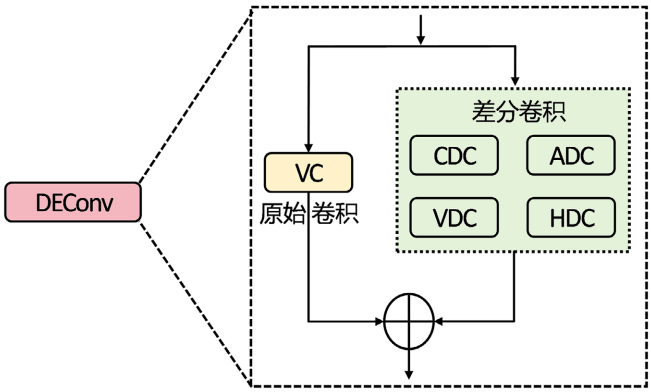


Figure 4. DEConv structure diagram
图 4. DEConv 结构图

$$F_{out} = VC(F_{in}) + CDC(F_{in}) + ADC(F_{in}) + HDC(F_{in}) + VDC(F_{in}) \tag{18}$$

其中，F_{in} 表示输入特征图，F_{out} 表示输出特征图。HDC 与 VDC 通过计算像素对之间的水平或垂直差分，有助于捕捉图像中的高频边缘信息；CDC 与 ADC 则分别聚焦于中心差分与极坐标方向的高频成分，补充普通卷积主要关注的低频内容。通过捕捉像素间的方向性差异，并将这些差异特征与卷积核权重融合，进而生成具有多层次视觉信息的特征图，增强了模型对图像特征的判别能力与泛化性能。同时，DEConv 采用了一种重参数化技术，在无额外推理开销的前提下，能够有效增强了模型对高频细节的表达能力。

3. 算法仿真

3.1. 数据集

本文选用 Yawning Detection Dataset (YawDD) [11]公开疲劳驾驶数据集进行模型改进效果的验证工

作。本文在数据集中随机选取 20 名男性、20 名女性共 40 段视频数据，驾驶员不同肤色、引入佩戴眼睛和太阳镜拓宽检测应用场景，驾驶员动作包括打哈欠和保持沉默两种。每间隔 6 帧抽取一张图片，去除失真、模糊等无效图片，利用 labelimg 工具进行面部、两个眼部、嘴部共五类特征的标注；最后，利用图像增强算法对闭眼类别进行数据扩充，同时便于进行模型鲁棒性验证。最终得到 7503 张图像数据，为确保类别分布平衡，按照 7:2:1 比例划分训练集、验证集和测试集。类别分为：人脸、睁眼、闭眼、张嘴、闭嘴四类。

3.2. 消融实验

为验证基于 YOLO11n 改进方案的有效性，本文分别在验证集、测试集上对改进点进行消融实验，以此评估 C3K2_AP、C2TSSA_DyT 以及 LSDECD 三个模块的改进效果以及模型的综合性能。消融实验结果如表 1 所示。从表实验结果可知，首先，引入 C3K2_AP 模块后，mAP50 及 mAP50-95 数值在验证集和测试集上均有提升，参数量降低约 7.7%，说明 C3K2_AP 可以通过提升卷积过程的感受野，兼顾上下文信息，降低参数量的同时增强小目标特征捕捉能力。其次，继续继承 C2TSSA_DyT 模块后，mAP50-95 有了明显提升，在验证集上增加 0.6 个百分点，验证集上则增加了 1 个百分点，这一结果说明，引入令牌统计自注意力机制，模型更专注于聚焦在眼部、嘴部这类微小变化上，减少了因遮挡、光照等因素带来的干扰，提高了检测准确率并确保模型参数量不会增大。最后，集成 LSDECD 高效检测头，可以发现参数量有所下降，FPS 明显提升实现模型的轻量化。

3.3. 对比实验

为更好评估改进的 YOLO11n 模型，本文选择若干现有的目标检测算法进行对比实验，实验结果如表 2 所示。实验结果表明相较于经典目标检测算法 SSD，本文所提出的改进模型在各方面均表现出色；同时，相比于 YOLO 系列的其他最轻量级尺寸模型，尽管改进模型在 FPS 上没有显现出明显改进优势，但在保持高帧率的同时，模型检测准确率以及参数量均有较大程度的优化；甚至，本文基于 YOLO11n 改进的模型，相较于 YOLO11s、YOLO12s 这些较大尺度模型在各方面也均有提升，进一步印证了本文改进方案的合理性。

Table 1. Results of ablation experiment
表 1. 消融实验结果

C3K2_AP	C2TSSA_DyT	LSDECD	mAP50/%		mAP50-95/%		R/%		Params/M		FPS/帧	
			Val	Test	Val	Test	Val	Test	Val	Test	Val	Test
-	-	-	98.1	98.1	73.5	73.2	96.9	97.1	2.6	2.6	390	411
√	-	-	98.9	99.1	76.7	76.8	97.7	98.3	2.4	2.4	370	379
-	√	-	99.0	99.1	76.5	76.3	98.1	**98.4**	2.5	2.5	395	419
-	-	√	**99.1**	**99.1**	76.8	76.9	98.2	97.8	2.3	2.3	405	420
√	√	-	99.0	99.1	76.7	76.9	98.1	98.0	2.4	2.4	367	383
√	-	√	99.0	99.2	76.8	77.4	98.1	98.1	2.1	2.1	379	391
-	√	√	99.0	99.2	76.8	77.4	98.1	98.0	2.2	2.2	**402**	**420**
√	√	√	99.0	99.0	**76.8**	**77.5**	**98.3**	98.0	**2.0**	**2.0**	379	394

注：其中 √ 表示添加模块，-表示未添加。

Table 2. Comparative test results
表 2. 对比试验结果

模型	mAP50/%	mAP50-95/%	参数量/M	FPS/帧
SSD [12]	91.6	68.5	144	187
YOLO5n [13]	97.4	73.2	2.5	394
YOLO8n [14]	98.3	73.3	3.0	409
YOLO9t [15]	98.3	73.0	2.0	365
YOLO10n [16]	98.2	73.1	2.3	484
YOLO11n [5]	98.1	73.5	2.6	390
YOLO11s [5]	98.5	75.2	9.4	290
YOLO12n [17]	98.4	72.3	2.5	388
YOLO12s [17]	98.5	74.2	9.1	255
改进模型	99.1	77.6	2.0	379

此外，本文在验证集中，将改进模型与基线模型 YOLO11n 进行了面部疲劳特征检测的结果进行对比。选择 IOU 在 0.5 到 0.95 范围内各类别平均检测精度均值 mAP50-95 这一评价指标进行对比，结果如图 5 所示。与原始模型对比，改进后的 YOLO11n 在面部、眼部、嘴部的检测精度均有显著提升。其中，面部类别检测精度提升 2.8 个百分点，提升面部识别的精确度能够为后续人脸关键点的应用奠定坚实的基础；闭眼作为疲劳特征直接影响疲劳行为的判断，该类别精度提升了 8.4%，证明了前文在提升小目标检测性能所做工作的有效性；张嘴类别也提升了 4.3 个百分点，能够进一步提升嘴部预警的准确性。

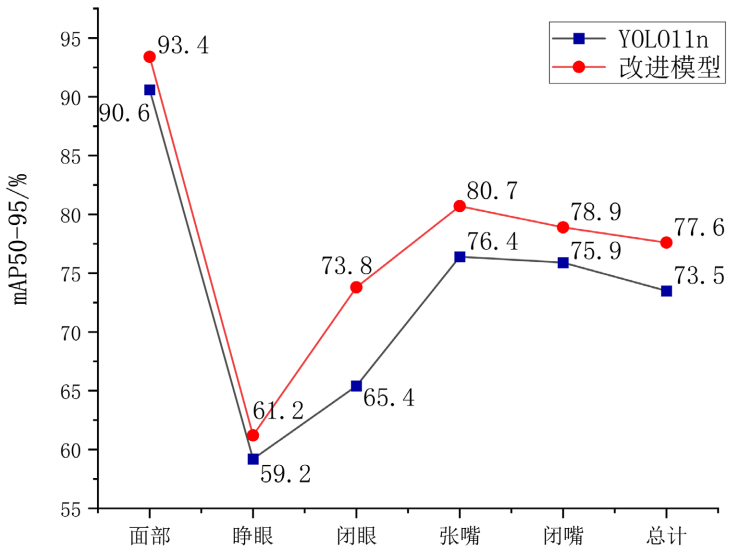


Figure 5. Comparison chart of accuracy rates for specific categories
图 5. 具体类别准确率对比图

4. 总结与展望

本文基于 YOLO11n 目标检测模型，在基本框架上做出优化，针对驾驶员面部、眼部及嘴部状态进行目标检测，属于疲劳预警系统的上游特征提取任务。主要改进工作有：首先，在 C3K2 模块中引入非对

称填充卷积实现卷积过程感受野的扩大，降低因分辨率低、角度偏移、遮挡等因素对面部疲劳特征提取效果造成的影响；其次，在 C2PSA 模块中引入一种基于令牌统计的自注意力机制，有效捕捉特征、精准聚焦目标区域，同时加入归一化层减小模型尺寸，增强鲁棒性；最后，在轻量级共享卷积检测头基础上引入细节增强卷积。本文在公开数据集 YawDD 上进行了网络模块消融实验、损失函数验证实验以及模型间对比实验，验证了改进方法的有效性。下一步，本文会将提取到的驾驶员面部特征信息与面部特征点检测模型以及头部姿态估计算法相结合，得到眼部、嘴部以及头部三部分疲劳特征做并行判断，以此实现更加可靠的疲劳状态判定。

参考文献

[1] 张育榕, 谷昆, 张轩雄. 基于神经网络的疲劳驾驶检测方法研究[J]. 理论数学, 2023, 13(5): 1298-1314.

[2] Benmohamed, A. and Zarzour, H. (2024) A Deep Learning-Based System for Driver Fatigue Detection. *Ingénierie des systèmes d'information*, **29**, 1779-1788. https://doi.org/10.18280/isi.290511

[3] 杜威, 宁武, 孟丽囡, 等. 基于改进 YOLO 的矿卡驾驶员疲劳检测算法[J]. 现代电子技术, 2025, 48(7): 126-131.

[4] Yin, L.F. and Ding, Z.Y. (2024) Lightweight Research on Fatigue Driving Face Detection Based on YOLOv8. Recent *Advances in Computer Science and Communications*, **19**.

[5] Khanam, R. and Hussain, M. (2024) Yolov11: An Overview of the Key Architectural Enhancements.

[6] Yang, J., Liu, S., Wu, J., Su, X., Hai, N. and Huang, X. (2025) Pinwheel-Shaped Convolution and Scale-Based Dynamic Loss for Infrared Small Target Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, **39**, 9202-9210. https://doi.org/10.1609/aaai.v39i9.32996

[7] Wu, Z., Ding, T., Lu, Y., *et al*. (2024) Token Statistics Transformer: Linear-Time Attention via Variational Rate Reduction.

[8] Zhu, J., Chen, X., He, K., LeCun, Y. and Liu, Z. (2025) Transformers without Normalization. 2025 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 10-17 June 2025, 14901-14911. https://doi.org/10.1109/cvpr52734.2025.01388

[9] 李军, 周科宇, 邹军, 等. 基于改进 YOLOv8n 的施工场景下防护装备佩戴检测算法[J]. 郑州大学学报(工学版), 2025, 46(3): 19-25+104.

[10] Chen, Z., He, Z. and Lu, Z. (2024) Dea-Net: Single Image Dehazing Based on Detail-Enhanced Convolution and Content-Guided Attention. *IEEE Transactions on Image Processing*, **33**, 1002-1015. https://doi.org/10.1109/tip.2024.3354108

[11] Omidyeganeh, M., Shirmohammadi, S., Abtahi, S., Khurshid, A., Farhan, M., Scharcanski, J., *et al*. (2016) Yawning Detection Using Embedded Smart Cameras. *IEEE Transactions on Instrumentation and Measurement*, **65**, 570-582. https://doi.org/10.1109/tim.2015.2507378

[12] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., *et al*. (2016) SSD: Single Shot Multibox Detector. In: *Lecture Notes in Computer Science*, Springer, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[13] Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., *et al*. (2022) Ultralytics/YOLOv5: v6. 2-Yolov5 Classification Models, Apple M1, Reproducibility, ClearML and Deci.ai Integrations.

[14] Yaseen, M. (2024) What Is YOLOv9: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector.

[15] Wang, C.Y., Yeh, I.H. and Mark Liao, H.Y. (2024) YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. In: *Lecture Notes in Computer Science*, Springer, 1-21. https://doi.org/10.1007/978-3-031-72751-1_1

[16] Wang, A., Chen, H., Liu, L., *et al*. (2024) YOLOV10: Real-Time End-to-End Object Detection. *Advances in Neural Information Processing Systems*, **37**, 107984-108011.

[17] Tian, Y., Ye, Q. and Doermann, D. (2025) YOLOV12: Attention-Centric Real-Time Object Detectors.