

# 基于局部和全局知识蒸馏的危险驾驶行为检测

李坤彪,杨晓晖\*,张风,徐涛,郭庆北

(济南大学信息科学与工程学院,山东 济南 250022)

**摘要:**为提高道路安全,预防交通事故,提出一种基于局部与全局知识蒸馏(local and global knowledge distillation, LGD)的危险驾驶行为检测算法。在知识蒸馏框架基础上,引入融合局部特征与全局特征的蒸馏损失函数,引导学生网络高效学习教师模型判别能力。这一方法有效促进学生网络的学习,使其能够在保持参数减少时,达到与教师网络相当的检测准确率。试验结果显示,本方法在仅包含31.85 M参数大小的模型下,实现91.79%准确率,表明该方法在处理分心驾驶检测问题上的有效性。

**关键词:**知识蒸馏;学生网络;分心驾驶检测

**中图分类号:**TP391

**文献标志码:**A

**引用格式:**李坤彪,杨晓晖,张风,等. 基于局部和全局知识蒸馏的危险驾驶行为检测[J]. 山东大学学报(工学版), 2025, 55(6): 13-20.

LI Kunbiao, YANG Xiaohui, ZHANG Feng, et al. Risky driving behavior detection based on local and global knowledge distillation[J]. Journal of Shandong University (Engineering Science), 2025, 55(6): 13-20.

## Risky driving behavior detection based on local and global knowledge distillation

LI Kunbiao, YANG Xiaohui\*, ZHANG Feng, XU Tao, GUO Qingbei

(School of Information Science and Engineering, University of Jinan, Jinan 250022, Shandong, China)

**Abstract:** To enhance road safety and prevent traffic accidents, a distracted driving behavior detection algorithm based on local and global knowledge distillation (LGD) was proposed. Built upon a knowledge distillation framework, the method introduced a distillation loss function that integrated both local and global features, guiding the student network to effectively learn the discriminative capabilities of the teacher model. While maintaining a lightweight network structure, the approach significantly improved the recognition accuracy of distracted driving behaviors. By effectively facilitating the learning process of the student network, the method enabled it to achieve detection accuracy comparable to that of the teacher model, despite having fewer parameters. Experimental results demonstrated that the proposed method achieved an accuracy of 91.79% with only 31.85 M parameters, highlighting its effectiveness in addressing the problem of distracted driving detection.

**Keywords:** knowledge distillation; student network; distracted driving detection

## 0 引言

据世界卫生组织报告,全球每年约有135万人死于道路交通事故,2000万~5000万人受伤,其中多人致残。分心驾驶是道路交通事故的主要原因之一<sup>[1]</sup>。分心驾驶会将驾驶员的注意力从驾驶任务转移到其他活动上,如发短信、接打电话等。这

种行为会显著降低驾驶人的反应速度和判断能力,增加交通事故发生风险<sup>[2-3]</sup>。据美国国家公路交通安全管理局数据,2022年,美国有3308人因分心驾驶丧生,289310人因分心驾驶受伤<sup>[4]</sup>。

认知分心、行为分心和视觉分心是驾驶员分心的3种典型类型<sup>[5-6]</sup>。驾驶员是车辆的主要决策者和控制者<sup>[7]</sup>,行为分心会极大影响正常驾驶,易导致交通事故。近年来,车载信息系统发展使车辆信

收稿日期:2024-07-17

基金项目:山东省自然科学基金资助项目(ZR2023LZH013);济南市市校融合发展战略工程资助项目(JNSX2023025, JNSX2023015)

第一作者简介:李坤彪(2000—),男,河南商丘人,硕士研究生,主要研究方向为驾驶员驾驶分心检测。E-mail: ise\_libk@stu.ujn.edu.cn

\*通信作者简介:杨晓晖(1985—),男,山东菏泽人,教授,博士生导师,博士,主要研究方向为图像处理和视频处理。

E-mail: ise\_xhyang@ujn.edu.cn

息娱乐和通信功能更加丰富流畅,易分散驾驶员注意力<sup>[8]</sup>。同时,手机通信已成为人们日常生活重要需求,一些司机甚至在开车时频繁接打电话、发信息<sup>[9]</sup>。

驾驶员分心检测方法大致分为基于驾驶人生理信号方法、基于可穿戴设备方法和基于计算机视觉方法。基于驾驶人生理信号的方法<sup>[10]</sup>通过生理传感器检测驾驶员的心率、脑电等指标判断驾驶员状态。该方法准确性高,但传感器具有侵入性,可能影响驾驶体验。基于可穿戴设备的方法,例如数据手套、头戴式眼动仪等,虽然能够捕捉驾驶员特定动作,但受限于设备的舒适性和可及性,难以广泛应用。基于计算机视觉的方法,通过深度学习算法分析驾驶员面部、眼球运动及肢体动作的视觉数据实现分心检测。因其非侵入性、成本低且易于部署,近年来受到研究人员广泛关注。

尽管计算机视觉方法展现出诸多优势,但在实际应用中仍面临严峻挑战。当前深度学习模型往往依赖高复杂度网络结构以实现高精度检测,导致推理计算量大,对设备性能要求较高,从而在边缘设备或实时场景下部署仍存在一定困难。现有检测方法通常依赖显著特征,忽视局部关键区域精细化处理,导致检测细微或局部异常行为时准确性不足,难以适应复杂多变的驾驶场景。如何在保证检测精度前提下实现网络轻量化,同时增强其对局部重要区域的关注能力,成为分心驾驶检测技术落地关键问题。

为解决上述问题,本研究提出基于局部和全局知识蒸馏的单阶段全卷积目标检测网络(local and global knowledge distillation fully convolutional one-stage object detection, LGD-FCOS)。该网络依托深度学习,从驾驶员图像中提取行为特征,通过教师-学生网络架构结合 LGD 方法在降低计算成本的同时保持高检测精度。在知识蒸馏过程中,引入注意力机制蒸馏,使学生网络能够自适应地关注图像中与分心行为最相关的区域,增强特征表征能力。

## 1 相关工作

### 1.1 目标检测

目标检测是计算机视觉领域中的一项任务,旨在识别图像或视频中的特定物体,确定其在图像中的位置<sup>[11]</sup>。目标检测方法主要分为二阶段目标检测方法和单阶段目标检测法2种。二阶段目标检测

方法将算法生成候选区域作为样本,通过卷积神经网络对样本进行分类。经典算法如更快基于区域卷积神经网络(faster region-based convolutional neural network, Faster R-CNN)<sup>[12]</sup>,通过引入区域选取网络(region proposal network, RPN)生成候选目标区域。相较 R-CNN 算法,Faster R-CNN 将候选区域的生成与目标分类和位置回归过程进行解耦,提高检测速度。级联区域卷积神经网络(cascade region-based convolutional neural network, Cascade R-CNN)<sup>[13]</sup>是 Faster R-CNN 的改进算法,通过级联结构引入多个级联分类器和级联回归器,提高检测器召回率和准确性。单阶段目标检测算法无需生成候选区域,直接从图像中预测目标位置和类别。文献[14]提出 YOLO(you only look once)算法,使用整个图作为卷积神经网络(convolutional neural networks, CNN)的输入,直接回归输出层中边界框的位置和类别。文献[15]提出 YOLOv3 算法,该算法使用 Darknet-53 作为主干网络,引入多尺度预测提高检测精度。单阶段全卷积目标检测网络(fully convolutional one-stage object detection, FCOS)<sup>[16]</sup>是一种先进的目标检测算法,它采用全卷积网络结构实现端到端的目标检测。FCOS 将每个像素点视为潜在目标中心,通过预测目标中心、类别及边界框大小进行目标检测,这使 FCOS 具有天然密集预测能力,无需使用锚框或候选区域生成网络。

### 1.2 知识蒸馏

知识蒸馏(knowledge distillation, KD)是一种模型压缩和知识迁移技术,将大型教师模型中知识传递给小型学生模型,减少计算成本,同时保持较高性能。知识蒸馏最早由文献[17]提出,通过将教师模型软标签作为监督信号来训练学生模型,通常用于提高模型泛化能力。在 FitNet<sup>[18]</sup>中,教师模型和学生模型通常都是深度卷积神经网络。与传统知识蒸馏不同,FitNet 专注于通过引导学生网络直接学习教师网络深层表示。FitNet 利用深度网络中不同层之间的信息流动,使得学生网络可以学会从教师网络的深层表示中进行有效的知识提取。注意力迁移(attention transfer, AT)蒸馏<sup>[19]</sup>是一种专注于将教师模型的注意力机制传递给学生模型的知识蒸馏方法。通过教会学生模型复制教师模型的注意力行为来提高性能。

一些研究将知识蒸馏用于目标检测领域。文献[20]通过降低分级头中背景蒸馏损失权重处理

蒸馏损失正负态之间的极端不平衡问题,同时保留对主干中完整特征图学习。文献[21]为二阶段检测器设计一个蒸馏框架,将 L2 蒸馏损失应用于学生模型 RPN 采样特征。该模型由随机采样的负样本和正样本组成,并按一定比例由真实标签区分。文献[22]提出一种新颖的本地化蒸馏方法,大幅提高定位蒸馏在目标检测中的蒸馏效率。文献[23]提出焦点和全局蒸馏。焦点蒸馏将前景和背景分离,迫使学生专注于老师的关键像素和通道。全局蒸馏重建不同像素之间关系,将其从教师传递给学生,补偿焦点提取中缺失的全局信息。

现有大多数工作主要侧重于整体特征蒸馏,忽略局部和全局特征在复杂任务中的重要性。在处理驾驶分心检测等需要同时关注局部行为细节与整体驾驶环境的复杂感知任务时,往往无法兼顾细节信息捕捉和全局场景理解,导致模型在实际应用中表现不够理想。

### 1.3 驾驶分心检测

文献[24]提出能够检测驾驶员在车内执行正常驾驶、说话等行为的多模态分析平台 DarNet。文献[25]通过 Faster R-CNN 检测手机使用情况。该算法优于以往方法,准确率为 94.20%。文献[26]结合车载摄像头与 YOLOv3-Tiny 算法识别驾驶员使用手机的行为,利用 MQ-3 传感器进行酒精检测,引入发动机锁定机制预防酒驾。文献[9]提出一种使用驾驶员手部图像进行分心检测的方案,使用基于阈值的重建误差准则训练一个无监督低秩非负字典。该方法在 Youtube 自我驾驶、国营农场驾驶、CVPR-3D、VIVA、AUC 分心驾驶员数据集<sup>[27]</sup>,准确率较高。文献[28]提出通过驾驶员姿势分类方法检测驾驶员使用手机和吃东西。基于两台红外相机采集新数据集,训练并评估不同深度学习模型,

以覆盖更广泛头部角度。文献[29]针对吸烟、打电话、回头和打哈欠四类常见注意力不集中行为,提出一种检测系统。该系统采用深度卷积神经网络 Inception v3,并结合两种数据增强策略与少数类过采样方法,提升分类模型泛化能力。文献[30]提出一种基于多源信息融合危险驾驶检测模型,该模型融合面部疲劳特征和分心驾驶特征,提高危险驾驶检测准确性和系统响应速度。文献[31]利用驾驶模拟器,结合眼动和瞳孔变化,分析认知分心对驾驶表现的影响。该研究为认知分心检测和干预措施提供了基础参考。文献[32]通过 EEG 数据结合最小生成树拓扑方法,提出一种有效识别驾驶分心状态模型为智能驾驶分心预警系统提供技术支持。文献[33]利用姿态检测提取驾驶员姿态特征,使用 CatBoost 等 4 种分类器检测驾驶分心,结果显示 CatBoost 模型在 F1 得分和预测时间上表现优异。文献[34]通过 Mediapipe 提取眼部位置关键点,结合长短期记忆网络和门控循环单元深度学习模型检测驾驶员视觉分心活动。

## 2 研究方法

### 2.1 总体框架

本研究提出一种基于知识蒸馏框架的目标检测网络检测驾驶分心。教师模型为 FCOS-ResNet101 网络,学生模型为 FCOS-ResNet50 网络。研究提出的结合局部和全局知识蒸馏的方法,通过全面捕捉细节和整体信息,提升模型在复杂驾驶场景中的精度。双重蒸馏策略提高模型的泛化能力和检测细微行为的能力。进行知识蒸馏操作后,该方法可以使得学生模型在不增长参数量的情况下准确率接近教师模型。方法的整体框架图如图 1 所示。

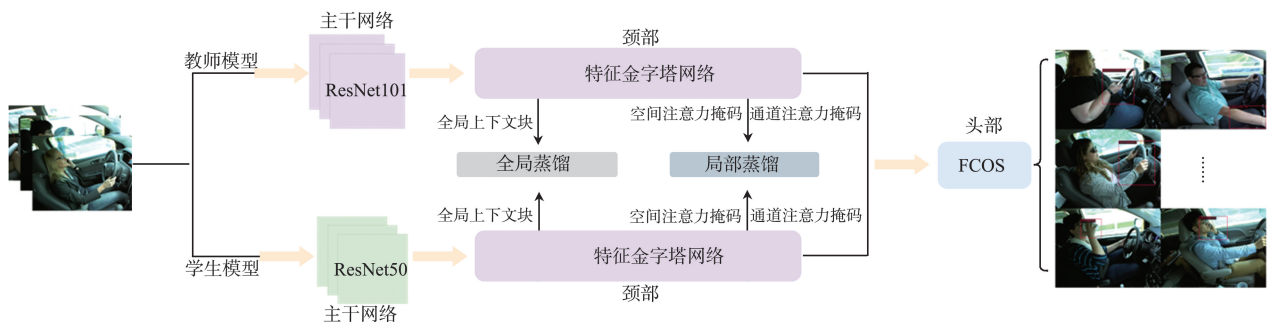


图 1 LGD-FCOS 网络整体框架图

Fig.1 Overall framework diagram of LGD-FCOS network

## 2.2 FCOS 目标检测算法

FCOS 算法主要特点如下。

(1)无锚框:FCOS 不依赖于预定义锚框,而是直接从网络输出中预测目标位置。

(2)边界框回归:FCOS 通过预测每个像素点到目标边界框四个角距离确定边界框位置。

(3)类别感知的边界框回归:FCOS 在边界框回归时,会同时考虑类别信息,实现类别感知的边界框预测。

(4)全卷积网络:FCOS 完全基于全卷积网络,不需要区域建议网络或类似组件。

(5)单阶段检测:FCOS 是单阶段检测器,直接从输入图像到最终检测结果。

(6)端到端训练:FCOS 网络结构是端到端的,可以直接从输入图像到最终检测结果进行训练。

FCOS 目标检测算法的网络结构图如图 2 所示,网络由主干网络、特征金字塔和检测头组成。

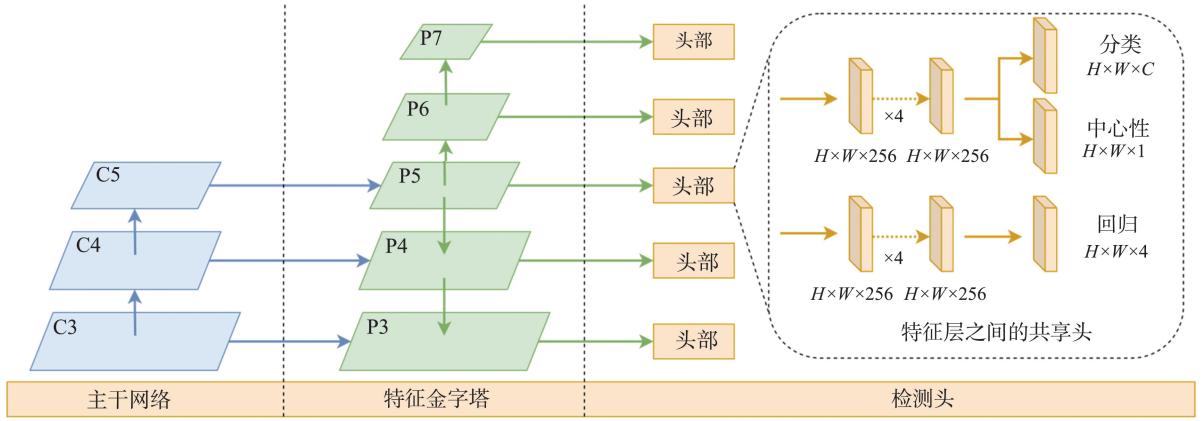


图 2 FCOS 目标检测网络结构图

Fig.2 FCOS object detection network structure diagram

特征金字塔是用于提取多尺度特征的关键结构,它通常由多个卷积层组成。这些卷积层能够从输入图像中提取不同层次特征。在特征金字塔中,较低层特征具有较高空间分辨率,适合检测小尺寸目标;较高层特征空间分辨率较低,但能捕捉到更抽象语义信息,适合检测大尺寸目标。这种多尺度特征提取方式使 FCOS 能有效检测不同尺寸目标。检测头是用于进行最终目标检测的关键部分。它通常连接在特征金字塔的每个层级上,以实现针对不同尺度特征的进一步处理。检测头负责以下几个关键任务。(1)分类:预测每个像素点属于各个类别的概率。(2)中心性:预测每个像素点作为目标中心点的可信度,这有助于提高检测的精度。(3)回归:预测每个像素点到目标边界框四个角距离,从而确定边界框位置。

FCOS 算法通过特征金字塔和检测头的结合,实现一种高效的目标检测框架。特征金字塔负责提取多尺度的特征,检测头则通过对这些特征进行分类、中心性评估和边界框回归,从而对图像中目标精确检测。该框架能够同时处理不同尺寸的目标,且通过端到端的训练方式提高检测的速度和准确性。

## 2.3 知识蒸馏方法

### 2.3.1 局部蒸馏

文献[18]使用取绝对平均值的方法来获得空间注意力掩码,提高蒸馏性能。这里用类似方法选择关键像素和通道,得到相应注意力掩码。不同像素和不同通道上的绝对平均值计算公式为:

$$G^P(F) = \frac{1}{C} \cdot \sum_{k=1}^C |F_k|, \quad (1)$$

$$G^C(F) = \frac{1}{HW} \cdot \sum_{i=1}^H \sum_{j=1}^W |F_{i,j}|. \quad (2)$$

注意力掩码的计算公式为:

$$A^Q(F) = H \cdot W \cdot \text{Softmax}(G^P(F)/T), \quad (3)$$

$$A^C(F) = C \cdot \text{Softmax}(G^C(F)/T). \quad (4)$$

特征的蒸馏损失计算公式为:

$$L_{\text{fea}} = \frac{1}{CHW} \sum_{k=1}^C \sum_{i=1}^H \sum_{j=1}^W A_{i,j}^Q A_k^C (F_{k,i,j}^T - f(F_{k,i,j}^S))^2, \quad (5)$$

式中, $F$ 为特征, $G^P$ 为不同像素的绝对平均值, $G^C$ 为不同通道的绝对平均值, $H$ 、 $W$ 为特征的高度和宽度, $C$ 为通道数, $F^T$ 和 $F^S$ 分别为教师和学生的特征, $A^Q$ 和 $A^C$ 分为教师检测器的空间和通道注意力掩码, $f(F_{k,i,j}^S)$ 是将 $F^S$ 重塑为与 $F^T$ 相同维度

适应层的操作,  $T$  为用来控制温度的超参数。

使用注意力损失  $L_{at}$  来使学生检测器模仿教师检测器的空间和通道注意力掩码, 即

$$L_{at} = \gamma \cdot (I(A_T^O, A_S^O) + I(A_T^C, A_S^C)), \quad (6)$$

式中,  $A_T^O$  和  $A_S^O$  分别为教师网络和学生网络的空间注意力,  $A_T^C, A_S^C$  分别为教师网络和学生网络的通道注意力,  $I(A_T^O, A_S^O)$  和  $I(A_T^C, A_S^C)$  为正则化损失计算,  $\gamma$  为用于平衡损失的超参数。

局部蒸馏损失  $L_{local}$  是特征损失  $L_{fea}$  和注意力损失  $L_{at}$  的总和, 即

$$L_{local} = L_{fea} + L_{at} \circ \quad (7)$$

### 2.3.2 全局蒸馏

全局上下文块 (global context block, GcBlock)<sup>[35]</sup> 是一种增强特征表示的模块, 旨在通过全局上下文信息的引入, 提升卷积神经网络的性能。GcBlock 通过全局池化和上下文建模捕捉整个图像全局信息, 改善特征表达能力, 其结构如图 3 所示。

用通过 GcBlock 模块计算后的结果,  $W_k$ 、 $W_{v1}$  和  $W_{v2}$  为卷积操作, LN 为层归一化,  $N_p$  为特征中的像素数, ReLU 为激活函数,  $\lambda$  为平衡损失的超参数。

总损失  $L$  的计算为

$$L = L_{original} + L_{local} + L_{global}, \quad (10)$$

式中  $L_{original}$  是检测器的原始损失。

蒸馏损失仅在特征图上计算, 该特征图可以从检测器颈部获得。局部蒸馏主要通过学习教师模型局部特征, 学生模型能够更精准地捕捉到图像中的重要细节和局部特征, 如危险驾驶行为具体姿态。这部分蒸馏主要通过公式 (7) 中特征损失和注意力损失实现, 确保学生模型在关键特征区域的注意力与教师模型一致。全局蒸馏通过对比教师模型和学生模型的全局特征关系, 促使学生模型更好学习全局上下文信息。公式 (8) 中定义全局蒸馏损失则通过比较教师模型和学生模型全局特征, 促使学生模型在较大空间范围内捕捉有用信息。由于局部特征往往容易受到细微噪声影响, 单独依赖局部蒸馏可能导致模型过拟合于特定细节。通过引入全局蒸馏, 模型能够更好地平衡对局部和全局信息学习, 避免因过分关注局部细节而导致过拟合。

## 3 试验验证

### 3.1 数据集

试验中使用的数据集是将 Kaggle 竞赛中 State Farm 分类数据集重新划分并且进行目标检测格式标注。该数据集包含约 22 000 张图像, 其中每个驾驶员都有约 2 200 张包含整个驾驶员的身体和车内环境的图像, 图像分辨率为 640 像素×480 像素。本研究将原数据集中的左、右手打电话合并为打电话, 左、右手发信息合并为玩手机。本研究图像分类及各类中图片数量如表 1 所示。

表 1 每个类别中的图像数量

Table 1 The number of images in each category

类别	数量/张
正常驾驶	2 467
喝水	2 325
打电话	4 643
玩手机	4 609
操控中控台	2 311

数据集按照 8 : 1 : 1 的比例被划分为训练集、验证集和测试集。

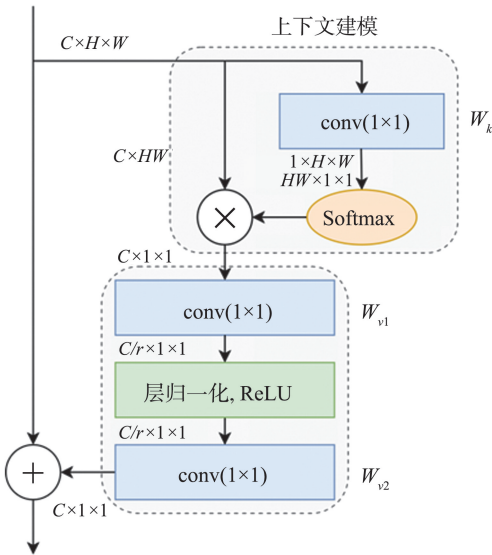


图 3 GcBlock 全局蒸馏

Fig.3 GcBlock global distillation

全局蒸馏损失使用 GcBlock 模块分别对教师网络、学生网络全局上下文信息进行计算, 得出全局蒸馏损失。全局蒸馏损失  $L_{global}$  如公式 (8) 和公式 (9) 所示:

$$L_{global} = \lambda \cdot \sum (R(F^T) - R(F^S))^2, \quad (8)$$

$$R(F) = F + W_{v2} \left( \text{ReLU} \left( \text{LN} \left( W_{v1} \left( \sum_{j=1}^{N_p} \frac{e^{W_k F_j}}{\sum_{m=1}^{N_p} e^{W_k F_m}} F_j \right) \right) \right) \right), \quad (9)$$

式中,  $R(F^T)$  和  $R(F^S)$  为教师网络和学生网络使

### 3.2 试验策略

试验所用硬件设备是 24 GB 显存的 RTX3090 显卡,Ubuntu 20.04.1 系统。模型训练 100 个轮次,优化器采用随机梯度下降,初始学习率为 0.01,动量为 0.9,权重衰减为 0.000 1。偏置项学习率倍数为 2.0,偏置项权重衰减倍数为 0。设置梯度裁剪,使用 L2 范数,梯度最大范数为 35。学习率策略选择阶梯型,即在特定的迭代步骤时调整学习率。warmup 参数为学习率预热的方式,这里选择的是常数方式。预

热的迭代步数为 500,预热时的学习率比例为 0.333 33。梯度列表包含具体迭代步骤,在第 30、60 和 90 个迭代步骤时会进行学习率调整。

训练过程采用 2.3 节所述知识蒸馏方法。在提供的数据集上独立训练教师网络,以确保其达到较高性能水平。利用其学习到的权重指导学生网络训练,并通过 2.3 节所提蒸馏损失函数,使学生网络学习教师模型提取的深度特征,从而提升自身性能。本研究所提知识蒸馏过程如图 4 所示。

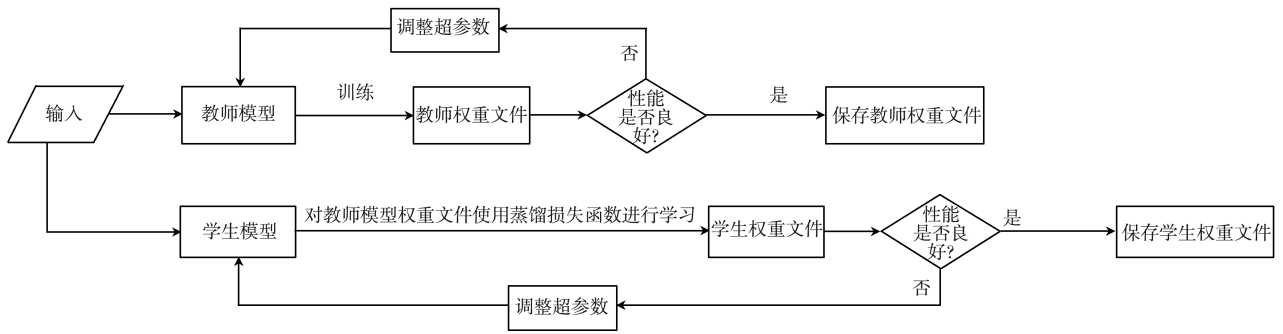


图 4 局部和全局知识蒸馏流程具体示意图

Fig.4 Schematic diagram of the specific process of local and global knowledge distillation

### 3.3 对比试验

在相同试验环境下对经典目标检测网络 GFL<sup>[36]</sup>、Reppoints<sup>[37]</sup>、RetinaNet<sup>[38]</sup>、Cascade\_RCNN<sup>[13]</sup>、YOLOv3<sup>[15]</sup> 在数据集上进行训练。通过平均准确率  $m_{AP}$ 、参数量  $P_{arams}$ 、每秒浮点运算次数  $F_{LOPs}$  3 个评价指标对算法性能评估,结果如表 2 所示。

表 2 各种算法的比较试验结果

Table 2 Comparative experimental results of various algorithms

网络模型	$m_{AP}/\%$	$P_{arams}/M$	$F_{LOPs}/G$
GFL	91.15	51.03	280.78
Reppoints	80.60	57.81	271.59
RetinaNet	91.70	55.18	282.20
Cascade_RCNN	91.30	87.93	310.55
YOLOv3	91.71	61.95	195.55
LGD-FCOS	91.79	31.85	196.86

由表 2 可以看出,本研究所提知识蒸馏方法精确度高于其他算法,且参数量和运算量较低。局部和全局知识蒸馏结合,有效提升模型对细节特征捕捉能力。这种多层次特征学习策略使得模型能够更准确识别不同类型驾驶分心行为,因此准确率优于其他方法。同时知识蒸馏方法目的在于降低模型参数量和计算量,研究所提知识蒸馏方法使学生模型在原有参数量和计算量下达到教师

模型精度。

### 3.4 消融试验

为验证本研究所提方法有效性,进行消融试验。试验结果如表 3 所示。由表 3 可以看出,学生网络的  $m_{AP}$  在未进行蒸馏操作时为 91.63%,仅使用局部蒸馏(local distillation, LD)时为 91.71%,仅使用全局蒸馏(global distillation, GD)时为 91.69%,将两者结合后的蒸馏结果为 91.79%。证明本研究所提方法有效,并且在以学生网络参数量的前提下取得接近教师网络精度。

表 3 消融试验结果

Table 3 The results of the ablation experiment

网络模型	方法	$m_{AP}/\%$	$P_{arams}/M$	$F_{LOPs}/G$
FCOS+R101(T)		91.83	50.79	272.92
FCOS+R50(S)		91.63	31.85	196.86
FCOS+R50(S)+LD	LD	91.71		
FCOS+R50(S)+GD	GD	91.69		
FCOS+R50(S)+LGD	LGD	91.79	31.85	196.86

注:其中 T 表示教师网络,S 表示学生网络。空格表示无法获得。

测试集上对算法进行测试的部分样例如图 5 所示,从图 5 中可以看出算法可以很好识别出目标类别,并且使用锚框准确检测到目标位置。图 6 为 FCOS\_LGD 算法在验证集上 100 个轮次的  $m_{AP}$  图,该图反映算法在不同目标类别上平均检测性能,有助于评估算法准确性和精确性。



(a) 正常驾驶

(b) 操控中控台

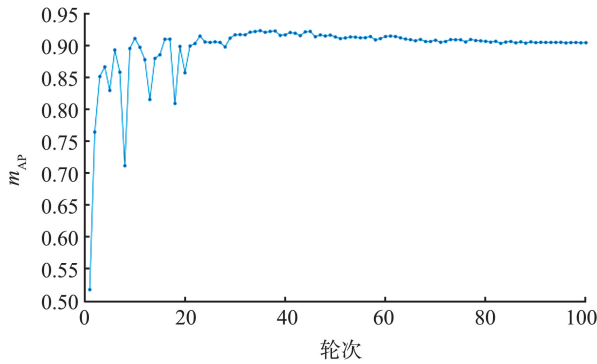
(c) 喝水

(d) 玩手机

(e) 打电话

图5 FCOS\_LGD 算法检测测试结果样例

Fig.5 Sample test results of FCOS\_LGD algorithm detection

图6 FCOS\_LGD 在验证集上 100 个轮次的  $m_{AP}$  图像Fig.6  $m_{AP}$  images of FCOS\_LGD over 100 epochs on the validation set

## 4 结论

本研究提出一种基于知识蒸馏的驾驶分心检测算法。训练一个网络层较深的教师模型用于引导学生模型进行训练,教师模型可以更好地学习深层特征然后通过提出的知识蒸馏方法将提取的深层特征传递给学生模型,使学生模型以较少参数量达到接近教师模型精度。为检验模型效果,对经典分类数据集 State Farm 数据集重新划分并进行目标检测格式数据标注。最终学生模型在该数据集上的检测精度达 91.79%,参数量为 31.85 M。在未来工作中,将对现有模型结构进行改进,尝试提出更轻量化学生模型,使学生模型可以以更少参数量达到更高精度,或尝试新知识蒸馏方法,比如对目标检测任务中锚框定位信息进行蒸馏。

### 参考文献:

- [1] World Health Organization. Road traffic injuries [EB/OL]. (2023-12-13) [2024-07-17]. <https://www.who.int/newsroom/factsheets/detail/road-traffic-injuries>
- [2] NASR AZADANI M, BOUKERCHE A. Driving behavior analysis guidelines for intelligent transportation systems [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(7): 6027-6045.
- [3] WANG J Y, CHAI W H, VENKATACHALAPATHY A, et al. A survey on driver behavior analysis from in-vehicle cameras [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(8): 10186-10209.
- [4] NHTSA. Put the phone away or pay[EB/OL]. (2024-04-01) [2024-07-17]. <https://www.nhtsa.gov/press-releases/2022-traffic-deaths-2023-early-estimates>
- [5] U. S. Center for Disease Control and Prevention. Distracted driving[EB/OL]. (2024-05-16) [2024-07-17]. [https://www.cdc.gov/distracted\\_driving/about/](https://www.cdc.gov/distracted_driving/about/)
- [6] ZANGI N, SROUR-ZREIK R, RIDEL D, et al. Driver distraction and its effects on partially automated driving performance: a driving simulator study among young-experienced drivers[J]. Accident Analysis & Prevention, 2022, 166: 106565.
- [7] HANG P, LV C, XING Y, et al. Human-like decision making for autonomous driving: a noncooperative game theoretic approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(4): 2076-2087.
- [8] JHA S, BUSSO C. Estimation of driver's gaze region from head position and orientation using probabilistic confidence regions[J]. IEEE Transactions on Intelligent Vehicles, 2022, 8(1): 59-72.
- [9] ROY K. Unsupervised sparse, nonnegative, low rank dictionary learning for detection of driver cell phone usage [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 18200-18209.
- [10] ZUO X, ZHANG C, CONG F Y, et al. Mobile phone use driver distraction detection based on MSaE of multi-modality physiological signals[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(11): 17650-17665.
- [11] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(1): 142-158.
- [12] 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, 2016, 39(6): 1137-1149.
- [13] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 6154-6162.
- [14] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 779-788.
- [15] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2024-07-17]. <https://arxiv.org/abs/1804.02767>
- [16] TIAN Z, SHEN C H, CHEN H, et al. FCOS: fully convolutional one-stage object detection [C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 9626-9635.
- [17] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[J]. Computer Science, 2015, 14(7):38-39.
- [18] ROMERO A, BALLAS N, KAHOU S E, et al. Fitnets: hints for thin deep nets[EB/OL]. (2014-12-19) [2024-07-17]. <https://arxiv.org/abs/1412.6550>
- [19] ZAGORUYKO S, KOMODAKIS N. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer[EB/OL]. (2016-12-12) [2024-07-17]. <https://arxiv.org/abs/1612.03928>
- [20] CHEN G B, CHOI W G, YU X, et al. Learning efficient object detection models with knowledge distillation[C]//Proceedings of the Advances in neural information processing systems. Long Beach, USA: NIPS, 2017:742-751.
- [21] LI Q Q, JIN S Y, YAN J J. Mimicking very efficient network for object detection [C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017: 7341-7349.
- [22] ZHENG Z H, YE R G, HOU Q B, et al. Localization distillation for object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(8): 10070-10083.
- [23] YANG Z D, LI Z, JIANG X H, et al. Focal and global knowledge distillation for detectors[C]//Proceedings of 2022 the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE, 2022: 4633-4642.
- [24] STREIFFER C, RAGHAVENDRA R, BENSON T, et al. Darnet: a deep learning solution for distracted driving detection [C]//Proceedings of the 18th ACM/IFIP/USENIX Middleware Conference Industrial Track. Las Vegas, USA: ACM, 2017: 22-28.
- [25] LE T H N, ZHENG Y T, ZHU C C, et al. Multiple scale faster-RCNN approach to driver's cell-phone usage and hands on steering wheel detection[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Las Vegas, USA: IEEE, 2016: 46-53.
- [26] JAIN H, ZOPE H, JAIN H, et al. Mobile phone detection and alcohol detection with engine locking for the prevention of car accidents[C]//Proceedings of the 2023 International Conference on Digital Applications, Transformation & Economy. Miri, Malaysia: IEEE, 2023: 1-6.
- [27] ABOUENLAGA Y, ERAQI H M, MOUSTAFA M N. Real-time distracted driver posture classification [EB/OL]. (2017-06-28) [2024-07-17]. <https://arxiv.org/abs/1706.0948>
- [28] WAGNER B, TAFFNER F, KARACA S, et al. Vision based detection of driver cell phone usage and food consumption [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(5): 4257-4266.
- [29] JIANG L D, XIE W, ZHANG D, et al. Smart diagnosis: deep learning boosted driver inattention detection and abnormal driving prediction [J]. IEEE Internet of Things Journal, 2021, 9(6): 4076-4089.
- [30] CAI J F, BAI J J, ZHOU T Q, et al. Dangerous driving behavior detection based on multi-source information fusion[C]//Proceedings of the 2022 8th International Conference on Big Data and Information Analytics (BigDIA), Guiyang, China: IEEE, 2022: 366-370.
- [31] NARAYANA P, ATTAR N. Analyzing the impact of distractions on driver attention: insights from eye movement behaviors in a driving simulator [C]// Proceedings of the 2023 Seventh IEEE International Conference on Robotic Computing (IRC). Laguna Hills, USA: IEEE, 2023: 356-359.
- [32] LIU R, QI G Q, GUAN W, et al. Driver distraction state recognition based on minimum spanning tree features using EEG data[C]//Proceedings of the 2023 3rd International Conference on Digital Society and Intelligent Systems (DSInS). Chengdu, China: IEEE, 2023: 437-440.