



JIANGXI NORMAL UNIVERSITY

江西师范大学

# 本科生毕业设计(论文)

中文题目 YOLO-DFMnet:动态聚焦调制与注意力引导的车辆检测方法

外文题目 YOLO-DFMnet: Vehicle Detection with Dynamic Focal Modulation and Attention Guidance

学 号 202026701170

姓 名 刘骏华

学 院 软件学院

专 业 软件工程

指导教师 王文乐 副教授

完成时间 2024.4

江西师范大学教务处制

# 声 明

本人郑重声明：

所呈交的毕业设计（论文）是本人在指导教师指导下进行的研究工作及取得的研究成果。其中除加以标注和致谢的地方，以及法律规定允许的之外，不包含其他人已经发表或撰写完成并以某种方式公开过的研究成果，也不包含为获得其他教育机构的学位或证书而作的材料。其他同志对本研究所做的任何贡献均已在文中作了明确的说明并表示谢意。

本毕业设计（论文）成果是本人在江西师范大学读书期间在指导教师指导下取得的，成果归江西师范大学所有。

特此声明。

声明人（毕业设计（论文）作者）学号：

声明人（毕业设计（论文）作者）签名：

签名日期：            年        月        日

## 摘要

在自动驾驶和智慧交通领域中，现有的车辆目标检测方法在处理类别分布不均和稀有目标识别问题时常常遇到挑战，这些方法倾向于多数类而忽略少数类（如罕见目标），严重限制了模型的识别与定位能力。

为突破此瓶颈，本文提出了 YOLO-DFMnet 检测框架，引入动态聚焦调制平滑 IoU 损失函数（DFMS-IoULoss），并结合改进的激活函数及卷积块注意力模块（CBAM），旨在提高车辆检测的精度和实时性，同时增强模型对少数类的识别能力。

在 BDD100K 数据集上的广泛测试表明，与原始模型相比，YOLO-DFMnet 模型在平均精度(mAP)提升了 8.4%，显著提高了对所有类别的检测准确度，并满足了实时性要求。本研究为自动驾驶和智能交通系统的实际应用提供了有力支持。

**关键词：**车辆检测；动态聚焦调制平滑损失函数；卷积块注意力模块

# Abstract

In the fields of autonomous driving and intelligent transportation, existing vehicle target detection methods often struggle with imbalanced class distribution and the identification of rare targets, tending to overlook minority classes (such as rare targets), which severely limits the model's recognition and localization capabilities.

To break through this bottleneck, this paper introduces the YOLO-DFMnet detection framework, which incorporates a Dynamic Focus Modulation Smooth IoU Loss function (DFMS-IoULoss) along with improved activation functions and the Convolutional Block Attention Module (CBAM). This is aimed at enhancing the accuracy and real-time performance of vehicle detection, while also boosting the model's ability to recognize minority classes.

Extensive testing on the BDD100K dataset shows that compared to the original model, the YOLO-DFMnet model has improved its mean Average Precision (mAP) by 8.4%, significantly enhancing detection accuracy across all categories, and meeting real-time performance requirements. This research provides strong support for the practical applications of autonomous driving and intelligent traffic systems.

**Key Words:** Vehicle Detection; DFMS-IoU; Convolutional Block Attention Module - CBAM

# 目 录

第 1 章 引言.....	1
1.1 项目背景及意义.....	1
1.2 国内外研究现状.....	1
1.3 本文的研究内容.....	2
1.4 论文的组织结构.....	3
第 2 章 相关理论介绍 .....	4
2.1 卷积神经网络.....	4
2.2 YOLO 系列模型介绍.....	4
第 3 章 基于 YOLO-DFMnet 的车辆检测方法 .....	8
3.1 YOLO-DFMnet 的网络结构.....	8
3.2 DFMS-IOU 损失函数 .....	9
3.3 激活函数改进.....	11
3.4 添加 CBAM 模块.....	12
第 4 章 实验结果与分析 .....	15
4.1 实验环境及参数设置.....	15
4.2 数据集及评价指标.....	15
4.3 实验结果与分析.....	17
第 5 章 总结和展望 .....	23
5.1 总结.....	23
5.2 展望.....	24
参考文献.....	25
致谢.....	27

# 第1章 引言

## 1.1 项目背景及意义

随着全球交通需求的激增，道路车辆目标检测技术日益成为道路交通安全和流量管理的关键支撑技术。传统的车辆检测方法主要依赖基于手工特征提取，不仅耗时费力，且难以应对复杂环境下的目标多样性。深度学习技术的引入，特别是 YOLO 系列算法的发展，极大地提升了检测速度和准确性。然而，这些高级模型在处理大规模数据集时仍面临类别不平衡问题，尤其在识别稀有目标时、重叠目标及以及复杂遮挡时表现不佳，这直接影响了系统的可靠性和实用性<sup>[1]</sup>。

因此，本文提出 YOLO-DFMnet 框架，旨在解决图像类别不平衡、尺度变化等问题。进而提升智慧交通系统的整体效能、确保道路行车安全，并对抗城市交通拥堵和减少交通事故，为构建更智能、安全的交通环境提供坚实的技术基础。

## 1.2 国内外研究现状

早期的车辆检测技术主要依赖人工设计的特征，该方法包括三个连续环节：先是界定潜在关注区域，并采取滑动窗口机制来截取可能含有目标的候选区域；随后对这些候选片段执行细致的手工特征抽取操作；最终借助分类模型来完成对象的区分与确认。这种方法虽然在早期有一定的效果，但是面对复杂多变的交通环境时，其局限性明显<sup>[2]</sup>：跨领域应用难度大，针对变化多端的目标形态，必须定制特征提取策略，这极度依赖研究者经验积累；在手动特征提取环节，信息抽取不充分会导致关键信息在训练阶段无法复原，直接削弱了检测成效；此外经典的滑窗扫描策略，通过分割图像为多样尺寸的小单元并逐一识别，优选高置信度区域而舍弃低概率片段，此过程不仅计算密集、充斥冗余步骤大幅拖慢了处理速度，还在实际工程部署上提出了较高门槛。

传统的车辆目标检测算法主要分为 3 类：帧间差分法<sup>[3]</sup>，背景差分法<sup>[4]</sup>，光流法<sup>[5]</sup>。与经典方法相比较，深度学习技术在训练模型时展现出优越的泛化能力，摆脱了对人工特征规则的依赖，转而直接从数据中自动发掘内在特征，这些特征更能深刻揭示数据的本质。得益于此，模型不仅能够在大规模数据集上得到有效训练，还能够较为简便地迁移到新的任务场景或不同数据集之中，促进模型的通用性和适应范围。深度学习驱动的目标检测技术开始于 LeCun 等人的开创性工作——1989 年问世的 LeNet，这一里程碑事件首次将卷积神经网络(Convolutional

Neural Networks, CNN) 应用于图像识别的舞台。及至 1998 年 LeNet5 的登场, 真正意义上揭开了 CNN 的序幕, 但当时受制于 GPU 算力的局限大规模数据运算难以实现, 导致 CNN 未能立即引起广泛关注。随着硬件技术的进步, 基于 CNN 的目标检测算法逐步分化为两大阵营。其中两阶段 (Two Stage) 方法作为一种基于区域的策略, 涵盖了 R-CNN、SPP-Net、Fast R-CNN 和 Faster R-CNN 等里程碑式的模型。R-CNN 作为深度学习在目标检测领域的先驱, 与先前的传统检测技术相比, 实现了显著的性能飞跃。它通过 Selective Search 策略生成候选区域, 随后利用神经网络抽取图像特征, 并将这些特征输入至分类器与回归器中完成检测。尽管这类两阶段方法在图像特征提取上表现出色, 但检测速度缓慢是其局限所在。YOLO 系列算法<sup>[6]</sup>作为单阶段(one stage)方法的典范, 以其实时处理速度而著称, 只在精度上略显不足。尤其 YOLO-v5 相较于其先驱版本, 在确保检测精确度大幅度增进的同时, 也显著加快了检测速度。这种优化使得 YOLO 系列在诸如自动驾驶、尖端监控系统等对时效性有着严苛要求的应用场景中占据了核心位置展现了其高效与准确性。

### 1.3 本文的研究内容

尽管最近 YOLO 系列算法取得了显著进展, 但在实际应用中仍面临着一些问题: 如何有效地处理各类样本间的均衡问题, 特别是对于稀有目标的精准识别与定位。这一难题源于大规模数据集中固有的类别分布不均衡现象, 其中多数类样本数量远超少数类。现有主流检测方法在面对这种数据分布失衡时, 往往呈现出对多数类学习的偏好性, 即模型在训练过程中容易被大量常见类别主导, 从而导致其对少数类 (稀有目标) 的学习效果大打折扣。具体表现为模型在预测阶段对少数类目标的召回率偏低、误检率偏高, 或者定位精度不足, 严重影响了整体检测系统的稳健性和实用性。

针对上述的问题, 本文提出了一种新的车辆目标检测框架 YOLO-DFMnet。本文主要工作如下:

(1) 动态聚焦调制平滑 IoU 损失函数(DFMS-IoULoss): 设计了一种新的损失函数, 动态地根据样本的检测难度调整损失权重, 对难检测的目标施以更高的关注, 从而优化了样本间不均衡的问题。通过结合平滑 IoU 策略, 改进了梯度计算, 使得模型在处理车辆尺寸变化、复杂形状等挑战时能更精准地回归边界框。DFMS-IoULoss 特别提高了对稀疏样本的关注度, 提升了模型泛化性能。

(2) Hard-swish 激活函数的引入: 采用非线性表达能力强的 Hard-swish 激活函数增强模型捕捉车辆特征的能力, 通过分段处理逻辑, 减少了内存访问频率,

加速了计算过程，同时保持了良好的数值稳定性，避免了梯度消失或爆炸问题。

(3)CBAM 注意力机制的集成：通过整合通道注意力（CAM）和空间注意力（SAM），CBAM 机制使得模型能够自适应地加权显著特征，同时抑制背景噪声。这个过程提升了特征图的质量，加强了对重要区域的关注，从而提高了特征输入的利用效率和模型在多层特征上的表现，促进了不同尺度和空间信息的有效整合。

## 1.4 论文的组织结构

第1章：介绍了车辆目标检测的背景及意义，阐述了国内外相关研究情况，最后说明了本文研究内容及各章节结构。

第2章：对CNN的定义和架构进行了介绍，随后介绍了YOLO模型的发展历程及其网络架构。

第3章：针对优化的YOLO系列模型进行了讲述。

第4章：描述了本文的实验环境、参数设置、数据集和评价指标，然后分析指标并对各模块做数据对比，最后对以上数据进行分析，得出研究结果。

第5章：总结并概括本文工作内容，并且展望未来研究的方向。



## 第2章 相关理论介绍

### 2.1 卷积神经网络

卷积神经网络 (CNN)<sup>[5]</sup> 是深度学习领域中一种基于卷积运算的神经网络结构, 适用于处理具有空间层级结构的数据。相较于传统的全连接神经网络, CNN 通过卷积操作能够有效地提取和表征输入数据中的局部特征, 这些特征在空间层级结构上是共享的, 减少了模型参数量并且提高了计算效率。CNN 的架构通常遵循前馈网络的设计原则, 其中卷积层负责捕获输入数据的空间结构信息, 通过卷积核的滑动与原始数据进行点积操作, 从而生成特征图; 而池化层则通过对特征图进行下采样操作, 实现了特征降维, 降低了模型的复杂度, 同时也增强了模型对空间位置的鲁棒性。全连接层位于 CNN 的顶层, 用于整合卷积层和池化层提取的特征, 将其映射到样本标记空间, 从而完成分类或回归任务。

在目标检测领域, 基于 CNN 的算法通常遵循以下步骤: 先通过区域提取算法 (如选择性搜索、边界框回归等) 从输入图像中生成一系列候选区域; 然后利用预训练的 CNN 模型对这些候选区域进行特征提取以获得丰富的区域描述; 最后将提取到的特征送入分类器中进行目标类别判断, 并通过回归器确定目标在图像中的精确位置。

YOLO (You Only Look Once)<sup>[6]</sup> 算法是近年来优秀的基于单一网络的目标检测算法。它摒弃了传统的区域提取和特征提取分离的设计思路, 通过前向传递网络直接预测出目标的边界框和置信度。YOLO 算法在保持较高检测精度的同时, 极大地提高了检测速度, 使其在自动驾驶、智能监控等实时性要求较高的场景中得到了广泛应用。通过不断的优化和改进, YOLO 系列算法已成为当前目标检测领域的主流方法之一。

### 2.2 YOLO 系列模型介绍

YOLO 系列始于 2015 年 Joseph Redmon 提出的 YOLO-v1, 采用卷积网络融合全连接层进行目标检测。历经迭代, 至 YOLO-v2 通过优化边界框及借鉴其他模型提升速度与精度。YOLO-v3 引入 darknet53 作为新骨干网络, 调整 anchor 并采用 Gaussian loss 及 GIOU、DIOU 提升性能。YOLO-v4 在 v3 基础上微调, 加入 Mosaic 数据增强、自对抗训练等策略, 采用 CSPDarknet53、Mish 激活函数及 SPP、FPN+PAN 结构, 进一步优化特征提取与损失函数, 推动目标检测技术的演

进。

YOLO-v5 模型包含四种不同型号。模型大致构造相似，但它们的深度和宽度不同。本文的研究工作主要基于 YOLO-v5s 模型，它是这个系列中深度和特征图宽度最小的，具有最快的检测速度和相对较高的识别精确率，非常适用于嵌入式设备中使用。YOLO-v5 结构如图 2.1 所示，下面对其详细介绍：

Input 输入端对图像数据进行自适应填充和锚框计算，训练时采用 Mosaic 数据增强技术，替代了传统的 CutMix 方法。Mosaic 通过随机处理四张图片（翻转、平移、缩放、裁剪和色域调整）并拼接成新图片，增加了训练样本的多样性和丰富度，提供了更多背景信息和目标交互，使模型能更好的学习目标特征。Mosaic 数据增强效果如图 2.2 所示。

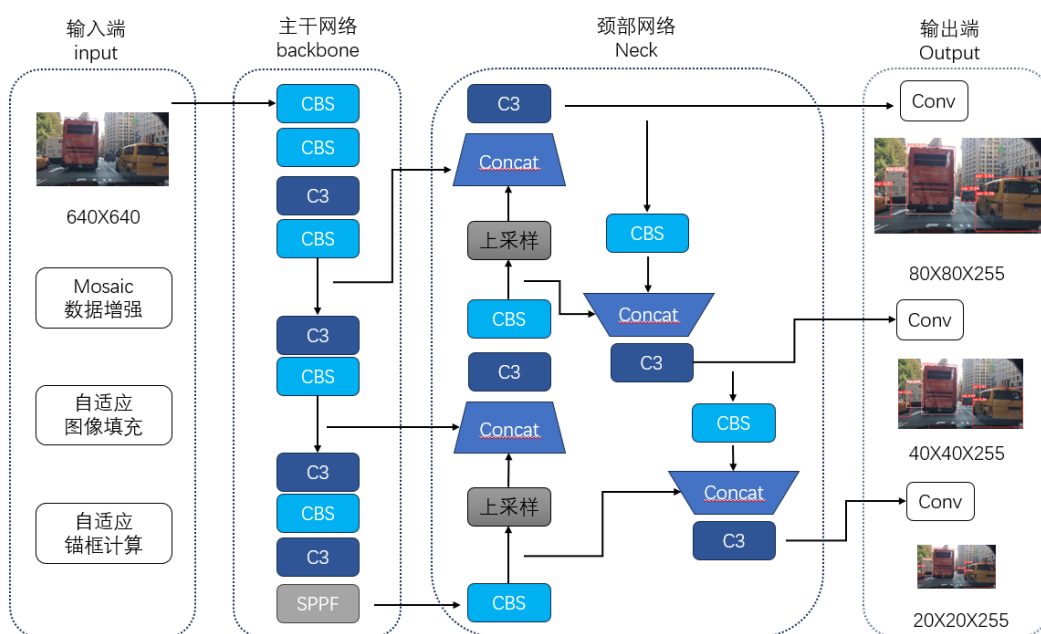


图 2.1 YOLO-v5 结构图

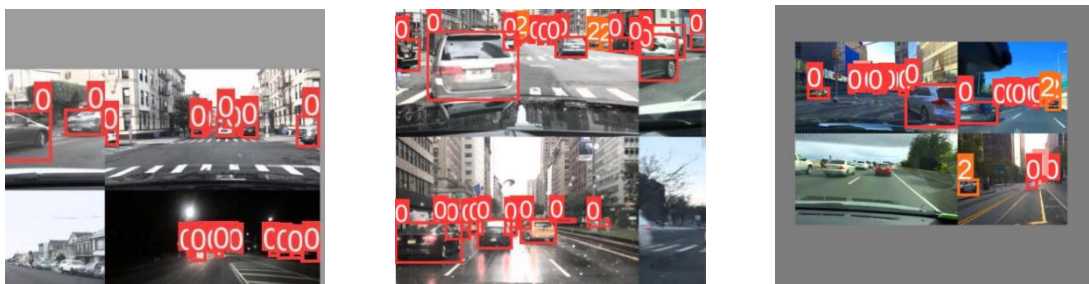


图 2.2 部分 Mosaic 数据增强效果

主干网络 (Backbone) 是特征提取的核心，它由 Focus、CBL、CSP 和 SPP 等模块构成。其中，Focus 结构通过 2 倍下采样将图像转换为 4 张特征图，有效保留了图像的宽度、高度信息并集中到通道维度，确保特征不丢失。这使网络能

在多个尺度上感知目标，增强了网络对不同尺寸目标的处理能力，从而提升目标检测网络的性能。Focus 模块切片操作示意图如图 2.3 所示。

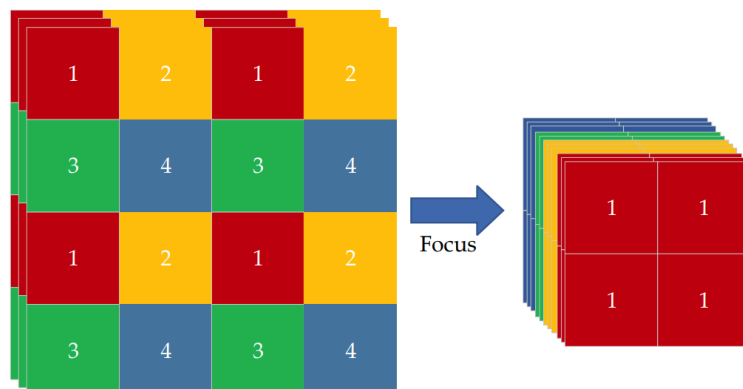


图 2.3 Focus 切片操作示意图

颈部网络（Neck）主要负责融合底层和高层特征，通过 FPN 向上采样并结合高层的丰富语义特征与底层的空间位置细节，以此促进不同尺度信息的有效整合。同时采纳自下而上的路径聚合网络 PAN 架构设计理念，该设计促进了定位特征从浅层到深层的传播，进而强化了模型在辨识和构建多样尺寸目标表示上的效能。FPN-PAN 结构使得网络在目标检测任务中能够更准确地理解和定位目标，提升整体性能。FPN-PAN 结构图如图 2.4 所示。

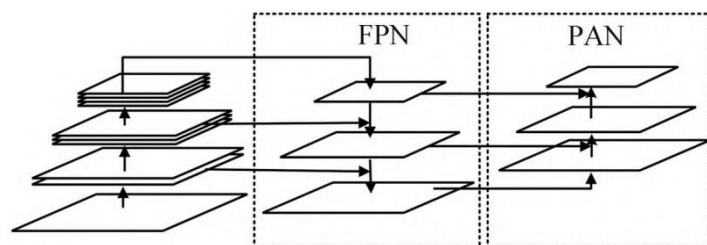


图 2.4 FPN-PAN 结构示意图

Output 输出端生成检测结果，包括 Bounding box 损失函数和 NMS。损失函数衡量预测与真实框的差距，优化网络参数以提高预测准确性。YOLOv5 使用 C IOU 来优化 Bounding box。非极大值抑制（Non-Maximum Suppression）是一种后处理技术，用于消除冗余的检测结果，保留最具代表性的目标框。在 YOLOv5 中，NMS 根据预测框的置信度排序，并去除与最高置信度框重叠度高的其他框<sup>[7]</sup>，以减少冗余并提高目标检测的精度和效率。

尽管近年来 YOLO 系列算法取得了显著进展，但在实际应用中仍面临一些问题：如何有效地处理各类样本间的均衡问题，特别是对于稀有目标的精准识别与定位。这一问题源于大规模数据集中固有的类别分布不均衡现象，其中多数类

样本数量远超少数类。现有主流检测方法在面对这种数据分布失衡时，往往呈现出对多数类学习的偏好性，即模型在训练过程中容易被大量常见类别主导，从而导致其对少数类（稀有目标）的学习效果大打折扣。具体表现为模型在预测阶段对少数类目标的召回率偏低、误检率偏高，或者定位精度不足，严重影响了整体检测系统的稳健性和实用性。

针对上述的问题，本文提出了一种新型的车辆目标检测框架 YOLO-DFMnet。该框架集成了动态聚焦调制平滑 IoU 损失函数（DFMS-IoULoss），巧妙融合动态聚焦调制与平滑 IoU 策略，前者根据样本难度动态调整损失权重，侧重优化困难目标，有效应对样本分布不均；后者通过改进梯度计算提升边界框回归精度，适用于车辆尺度变化、复杂形状等情形。为增强 DFMS-IoULoss 的性能，我们采用非线性表达能力的激活函数 Hard-swish 以捕捉车辆细微特征，并嵌入 CBAM 注意力机制以自适应加权显著特征、抑制背景干扰，从而提升特征输入质量与多层特征的有效利用。

## 第3章 基于 YOLO-DFMnet 的车辆检测方法

### 3.1 YOLO-DFMnet 的网络结构

本文针对 YOLO 模型，提出网络级动态聚焦机制与注意力引导特征精细化的车辆检测方法 YOLO-DFMnet，该模型的运行流程如图 3.1 所示：

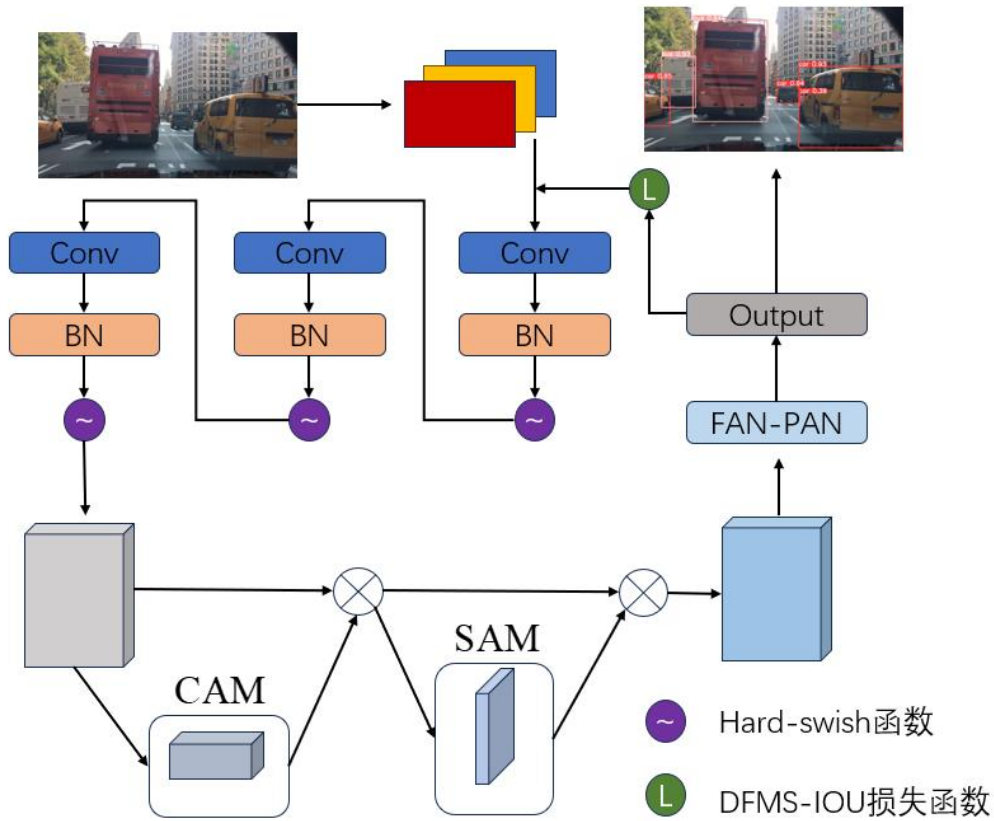


图 3.1 YOLO-DFMnet 流程图

先将目标图像输入模型，通过一系列的处理环节——包括卷积层、批量标准化层和 Hard-Swish 激活机制来开展特征提取工作，利用卷积核在图像上的滑动操作，挖掘出深层次的视觉特征。随后借助批量归一化层，不仅均衡了卷积输出的特征分布，还显著提升了模型的稳定性和对数据的适应能力。Hard-Swish 激活函数的融入，则为这一流程增添了非线性的维度，进一步放大了特征表达的丰富性和深度。特征经过卷积层和批量归一化层的处理后，本文引入了 CBAM 模块，以提取特征图的空间和通道信息。CBAM 模块整合了通道注意力和空间注意力

机制，分别学习特征图中各通道的重要性权重，并增强对局部信息的关注，从而使模型更准确地关注到物体。本文处理了一系列卷积层的特征后，采用改进后的损失函数 DFMS-IOU 来进行目标检测任务的训练和优化。DFMS-IOU 新定义了惩罚措施，聚焦的解决了难例样本和类别不平衡和高低质量样本的梯度差异问题，在训练过程中优化了边界框的预测。最终获得了模型的输出结果，既涵盖预测的物体边界框信息，也明确指出了框定对象的分类标识及置信度。

接下来分别详细介绍 YOLO-DFMnet 中 DFMS-IOU、Hard-swish 和 CBAM 的工作原理。

### 3.2 DFMS-IOU 损失函数

YOLO 系列模型采用了 CIOU 损失函数。然而，在面对更为复杂的检测场景，特别是当涉及大规模数据集和类别不平衡问题时，标准的 CIOU 损失函数可能仍不足以全面优化模型表现，尤其是对于罕见目标的识别与精确定位。因此，本文提出了一个新的损失函数——动态聚焦调制的平滑 IoU 损失函数（DFMS-IOU 损失函数）。

(1) 姿态估计综合损失（Pose Estimation Consolidated Loss）描述了中心点与 x-y 轴之间的最小角度。它可以引导锚框找到与目标框最接近的轴，从而减少 BBR（Bounding Box Regression）的总自由度。预测框  $B$  和真实框  $B^{gt}$  的中心点坐标、宽度和高度分别为  $(b_{cx}, b_{cy}, w, h)$  和  $(b_{cx}^{gt}, b_{cy}^{gt}, w, h)$ ， $\sigma$  是预测框和真实框中心点之间的距离， $c_w$  和  $c_h$  如图是外接矩形的宽和高。如图 3.2 所示。

$$C_h = \max(b_{cy}^{gt}, b_{cy}) - \min(b_{cy}^{gt}, b_{cy}), \sigma = \sqrt{(b_{cx}^{gt} - b_{cx})^2 + (b_{cy}^{gt} - b_{cy})^2} \quad (3.1)$$

$$x = \frac{C_h}{\sigma}, \Lambda = \sin(2 \sin^{-1} x), \Delta = \sum_{t=x,y} (1 - e^{-\gamma \rho_t}) \quad (3.2)$$

$$\rho_x = \frac{b_{c_x}^{gt} - b_{c_x}}{c_w}, \rho_y = \frac{b_{c_y}^{gt} - b_{c_y}}{c_h}, \gamma = 2 - \Lambda \quad (3.3)$$

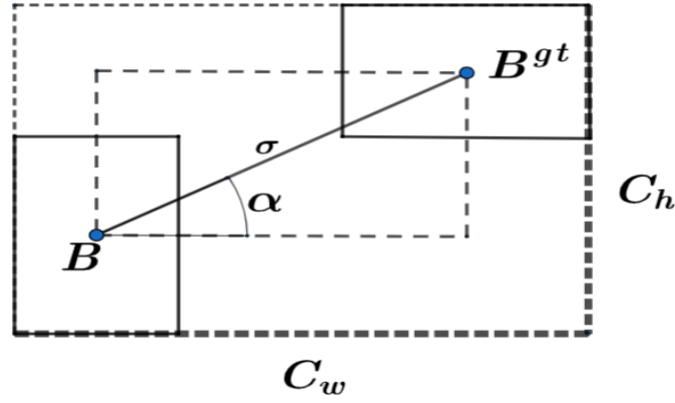


图 3.2 姿态损失

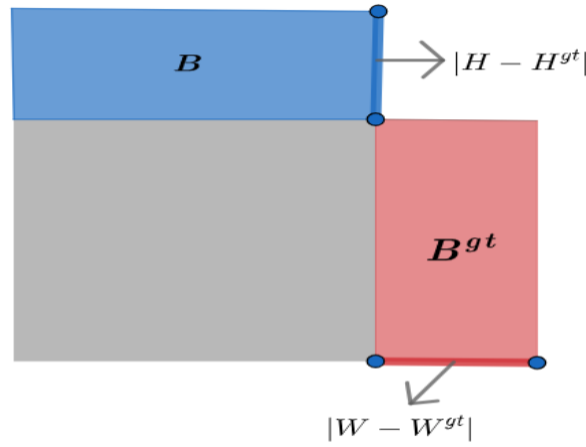


图 3.3 几何损失

(2) 几何一致性损失 (Geometric Consistency Loss) 考虑两个边界框之间的长宽比。它通过计算两个边界框的宽度差异以及它们之间的最大宽度比例来定义。在这个方程中  $\theta$  的值设为 4，它控制着形状的代价应该被关注的程度。其损失函数如(3.6)式。

$$\Omega = \sum_{i=x,y} (1 - e^{-\omega_i})^\theta \quad (3.4)$$

$$\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}, \omega_h = \frac{|H - H^{gt}|}{\max(H, H^{gt})} \quad (3.5)$$

$$L_{IoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (3.6)$$

(3) 动态聚焦调制的平滑 IoU 损失 (Dynamically Focal-Modulated Smooth IoU Loss) 本文新调整了姿态和几何方面的损失计算方式，确保随着对象形状复杂度的提升，对错误匹配的惩罚力度也随之增强<sup>[10]</sup>。这种方法促使模型训练加速并降低了预测边界的误差。但是训练中遭遇了样本不平衡的困境，特别是那些精确匹



配的目标实例相比不准确的匹配显得更为稀缺。现有研究表明, 这些不寻常的样本可能导致训练中梯度过大, 对模型学习产生负面影响。因此, 优化策略着重于确保高精度样本在训练中发挥更大的作用, 贡献更多的梯度信息<sup>[8]</sup>。

为了解决这个问题, 本文引入了一种聚焦机制, 以使损失函数更专注于高质量的样本, 分析如下。

(1) 当回归误差较小时, 梯度幅值应该较大, 而在回归误差较大的区域, 梯度幅值应该较小。

(2) 可以设置一些超参数, 以便能够灵活地控制难以检测的目标的关注程度。

(3) 随着超参数取值的变化, 梯度函数具有一个归一化的尺度, 在 $(0, 1]$ 范围内, 这有助于在常见和稀疏样本之间保持平衡。

基于上述考虑, 我们定义了 DFMS-IOU 损失如(3.8)式:

$$L_{\text{DFMS-IOU}} = R_{\text{DFMS}} L_{\text{IOU}} \quad (3.7)$$

$$R_{\text{DFMS}} = -(1 - \text{IoU})^\gamma \ln(\text{IoU}) \quad (3.8)$$

其中,  $\gamma$  是控制异常值抑制程度的参数。通过引入 $(1 - \text{IoU})^\gamma$ 项, 可以提高低 IOU 样本的权重, 有助于模型对困难样本的关注和学习。同时,  $-\ln(\text{IoU})$ 项用于减小高 IOU 值样本的梯度, 以降低它们在训练过程中的影响。具体的 $\gamma$ 值可以根据问题和数据集的特点进行调整。经过实验证明, 使用 $\gamma = 0.5$ 可以取得出色的性能, 既平衡了异常值的抑制程度, 又提高了对不同 IOU 值样本的敏感性<sup>[9]</sup>。

目前的研究普遍忽视了有效挖掘 BBR (Bounding Box Regression) 实例的重要性, 这导致低质量样本对梯度的贡献过大, 从而限制了 BBR 的性能。基于这一观察, 我们提出了 DFMS-IOU 损失, 旨在解决现有损失函数的缺陷, 并平衡高质量和低质量样本所导致的梯度差异。该方法减少了易于检测目标上的预测误差对损失函数的贡献, 更加关注那些难以检测的目标, 从而显著提升了模型的泛化能力。

采用 DFMS-IOU 损失后, 模型能够有效地学习并挖掘 BBR 实例的重要特征, 提高了模型在目标检测任务中的性能和泛化能力。这一改进方法对于解决低质量样本导致的梯度贡献过大、平衡样本差异以及提升模型的鲁棒性具有重要意义, 并为相关领域的研究和应用提供了一种有效的解决方案。

### 3.3 激活函数改进

激活函数对于提升模型的表达力和运算速率非常重要。本研究采用 Hard-swish 函数来替换 ReLU 函数, 基于其展现出的额外优点。Hard-swish 通过实施分段处理逻辑, 有效减少了内存访问频次, 这不仅显著缩短了等待周期, 还为模



型运算提速提供了强大助力，促进了整体计算效率的飞跃。**Hard-swish** 函数具有很好的数值稳定性，它能够有效地避免梯度爆炸或梯度消失使得模型在训练过程中更加稳定。这样可以避免一些数值上的不稳定性带来的影响，使得模型的训练更加可靠和高效。此外，采用 **Hard-swish** 函数还能够使得模型具有更丰富的表达能力。**Hard-swish** 函数可以更好地保留和传递特征信息，将对输入进行非线性变换的能力最大化。这样网络可以更好地学习到输入数据的复杂关系，提升模型的特征提取能力和分类准确度。

**Hard-swish** 表达式如式(3.9)，图 3.4 为其函数图像：

$$\text{Hardswish}(x) = \begin{cases} 0, & \text{if } x \leq -3 \\ x, & \text{if } x \geq +3 \\ \frac{x(x+3)}{6}, & \text{otherwise} \end{cases} \quad (3.9)$$

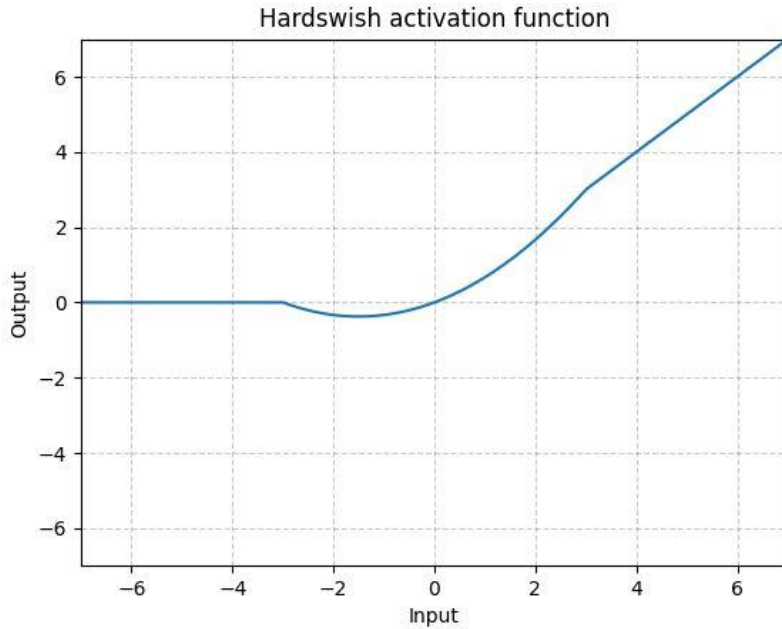


图 3.4 **Hard-swish** 函数图像

### 3.4 添加 CBAM 模块

如图 3.5 的 **CBAM** 架构中，特征信息先通过考量通道间的相互依赖生成通道注意力映射（**CAM**），该过程强调了图像中富含信息的部分。随后再经过通道注意力增强的特征经历一个空间注意力（**SAM**）分析步骤，进一步细化并补充了 **CAM**，确保空间维度上的关键特征得到凸显。**CBAM** 处理流程的最终产物是一个在通道和空间方向上均被赋予差异化注意力权重的新特征图，这样的特征表示加深了通道内的特征关联，同时促进了空间特征的协同作用，共同提升特征的有

效性和模型的识别能力。

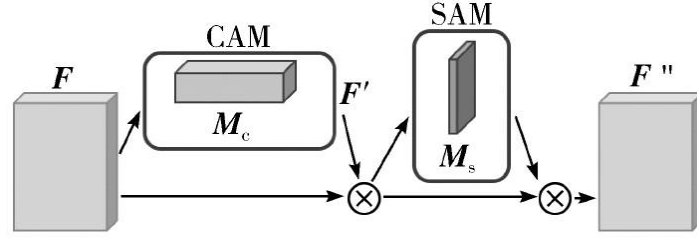


图 3.5 CBAM 结构图

如图 3.6 所示，CAM 结构通过对输入特征图  $F$  执行全局最大池化（Global Max Pooling, GMP）<sup>[11]</sup>和全局平均池化（Global Average Pooling, GAP）<sup>[12]</sup>，然后经过多层感知器（Multi Layer Perception, MLP）<sup>[13]</sup>，将 MLP 的输出进行逐元素相乘和相加，并通过激活操作生成通道注意力特征图  $M_c$ 。公式为：

$$M_c = \sigma \left( MLP(GAP(F)) + MLP(GMP(F)) \right) \quad (3.10)$$

其中， $\sigma$  代表激活操作。

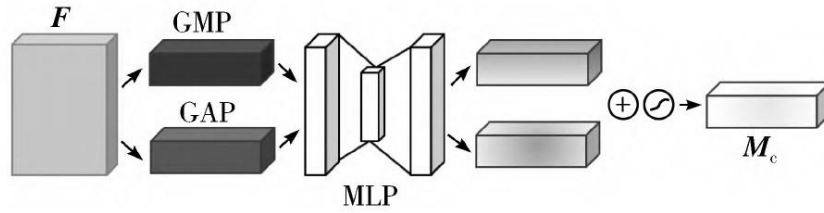


图 3.6 CAM 结构图

根据图 3.7，通过对输入  $F'$  的各通道进行均值池化和最大值池化操作，可以在维度上对通道信息进行整合。对获得的两个  $H \times W \times 1$  大小的通道特征进行融合时，需先将其输入到一个  $7 \times 7$  的卷积层中。最终通过使用 Sigmoid 激活函数对权重系数和输入特征  $F'$  进行相乘，得到空间注意力图  $M_s$ 。具体公式如下：

$$M_s = \sigma(f^{7 \times 7}([GAP(F); GMP(F)])) \quad (3.11)$$

其中， $f^{7 \times 7}$  代表  $7 \times 7$  卷积。

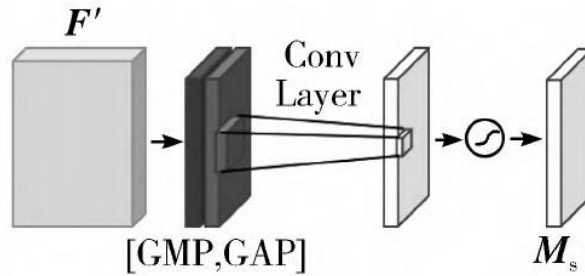


图 3.7 SAM 结构图

在自动驾驶应用的背景下<sup>[14]</sup>,检测目标间的外观形态展现出显著多样性。受摄像头成像原理的制约,接近摄像头的物体倾向于在图像画面中占据较大比例,相反,远处物体所占视觉空间则相对狭小。基于这一现象,我们可依据目标的外观尺寸,将待识别的对象科学划分为微形、中型与大型三类,以此作为分类处理的初步步骤。

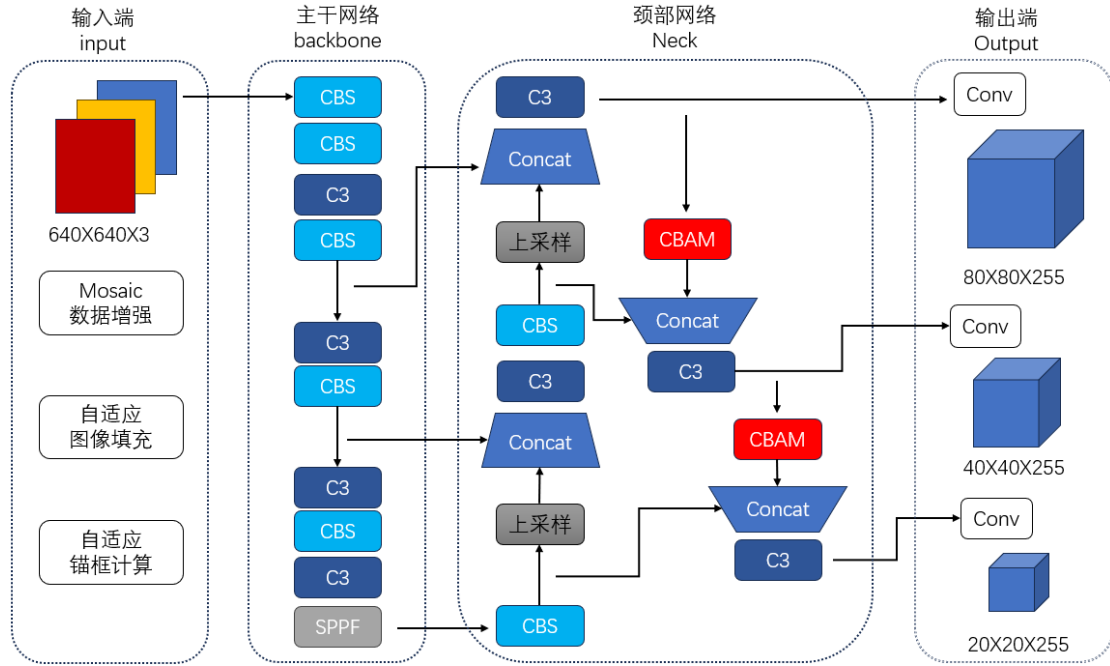


图 3.8 优化后的网络结构

本研究对 YOLO 模型的中间网络部分实施了创新性调整,集成两个 CBAM 组件,其优化结构如图 3.8。在采用 CBAM 模块处理特征图的流程中,我们先进行通道注意力 (CAM) 分析,随后执行空间注意力 (SAM) 计算,以此顺序获得经过双重注意力机制优化的输出。CBAM 模块的核心优势在于同步捕获 SAM 和 CAM 两方面的特征重视度信息,使得通过注意力引导的特征处理流程,模型能够聚焦于图像内最关键的信息区域。这种机制促进了特征图中关键元素的有效突出,增强了模型在特征提取层面的敏锐度,继而提升了目标检测的精确性。

## 第4章 实验结果与分析

### 4.1 实验环境及参数设置

本文实验环境为 Linux 操作系统，内存为 12 GB 的 NVIDIA GTX3060 GPU 上执行，并在 PyTorch 1.12.0、Torchvision0.13.0 和 CUDA10.2 环境下进行数据训练和目标检测。设置网络超参数具体如表 4.1 所示。

表 4.1 网络超参数表

训练参数名称	参数值
初始学习率	0.01
学习率下降参数	0.0001
权重衰减系数	0.0005
动量(momentum)	0.937
批处理大小(batch)	16
训练批次(epoch)	100
图像输入尺寸	640*640

### 4.2 数据集及评价指标

#### 4.2.1 数据集

本实验数据集以 Berkley 大学发布的 BDD100K<sup>[15]</sup>自动驾驶数据集为基础进行选择 and 调整，按照 7:2:1 的比例分配为训练集、测试集和验证集。该数据集有大量的数据信息，如环境、天气、地理因素。在原来的数据集中，特征类型包括公共汽车、交通灯、交通标志、人、自行车、卡车、摩托车、汽车、火车和乘车人等，本文剔除了冗余和不相关的特征，将需要识别的特征类型减少至三种，即 car、bus 和 truck，这样可以提高检测速度，减少不必要的等待时间，提高模型的性能和可解释性，使实验结果更加准确。

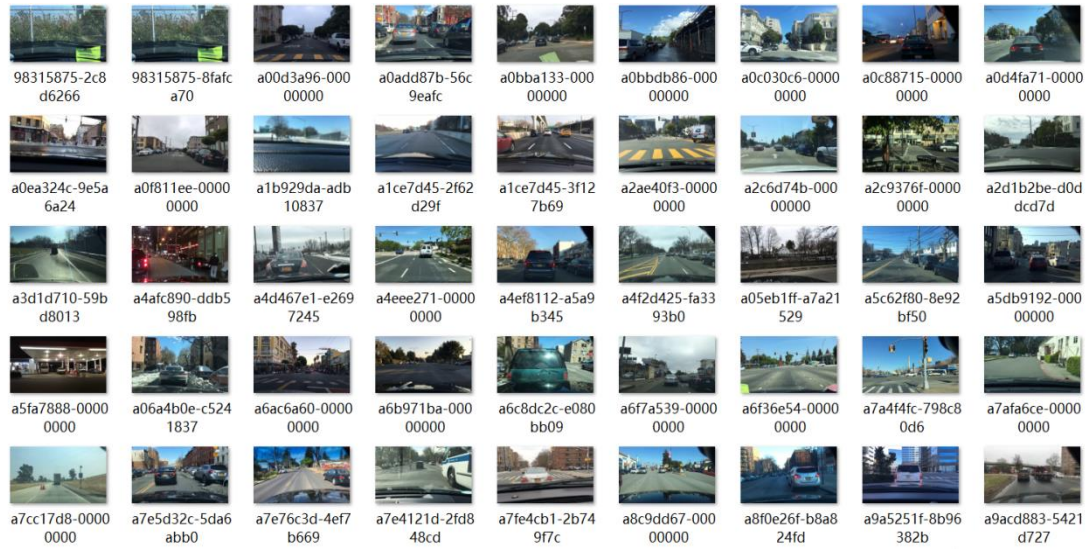


图 4.1 部分数据集展示

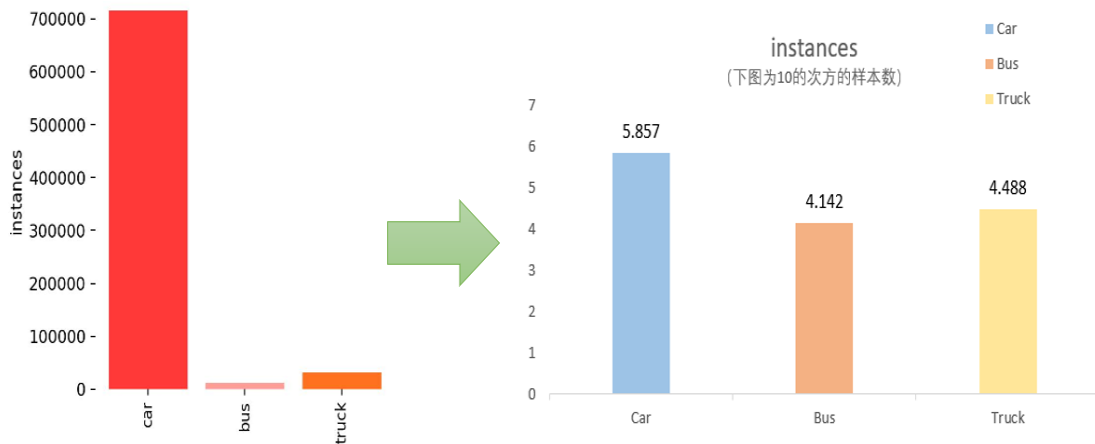


图 4.2 标记的对象

实验将原 BDD100K 文件中的 JSON 文件转为 YOLO 格式的 TXT 文件，将冗余和不相关的特征类型剔除后，保留下来的 **car**、**bus** 和 **truck** 的标记数量如图 4.2 所示。我们将数量柱状图，转化为 10 的指数的柱状图，便于观察和比较不同类别之间的数量差异。转化后的柱状图如下：**car** 类别数量最多；**bus** 类别数量最少；**truck** 类别数量居中。通过这样的转化，我们可以更直观地看出各类别之间的数量差距，从而更好地理解数据集的特点和分布情况。图 4.3 说明了车辆目标检测的建模流程，从数据准备开始到模型训练<sup>[16]</sup>。从 JSON 格式的原始图像标注到 YOLO 格式标签的转换过程是至关重要的初始步骤，保证了与 YOLO 训练框架的兼容性。

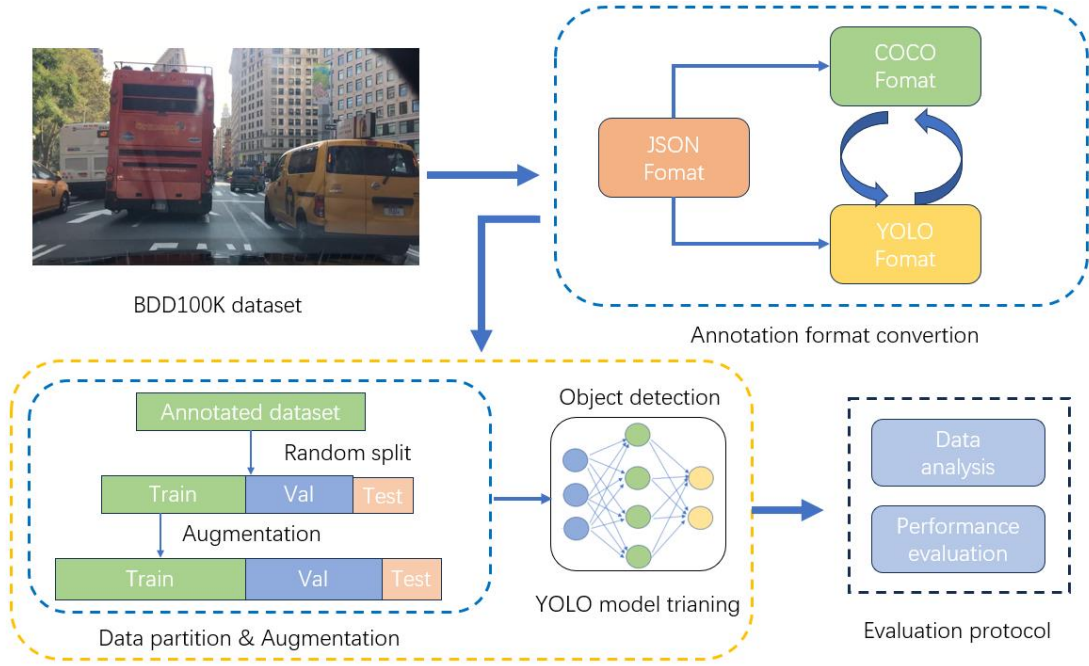


图 4.3 数据准备和目标检测流程

### 4.2.2 评价指标

针对车辆目标检测问题，本文主要使用精确率<sup>[17]</sup>（Precision, P）、召回率（Recall, R）、平均精度（Average Precision, AP）、类别的平均检测精度（mean Average Precision, mAP）对检测结果进行评价<sup>[18]</sup>。计算公式如式(4.1)-式(4.4)所示：

$$P = \frac{TP}{TP + FP} \quad (4.1)$$

$$R = \frac{TP}{TP + FN} \quad (4.2)$$

$$AP = \int_0^1 \text{Precision}(\text{Recall})dR \quad (4.3)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (4.4)$$

## 4.3 实验结果与分析

### 4.3.1 各性能指标对比

如表 4.1 所示，是 baseline 和 YOLO-DFMnet 的 PR 值以及平均精度的对比，改进后的 YOLO-DFMnet 的整体性能比 YOLO 高，精确度达到 0.80234，召回率达到 0.60398，平均精度值为 0.73987，推理时间达到了 0.009s。

表 4.1 结果分析

模型	Precision	Recall	mAP@0.5	推理时间
YOLO	0.73642	0.57304	0.65154	0.014s
YOLO-DFMnet	<b>0.81126</b>	<b>0.64079</b>	<b>0.73558</b>	<b>0.009s</b>

从上表中可以得知，平均精度值提升了 8.4%，精确率(P)和召回率(R)也分别提升了 7.5%和 6.6%，推理时间下降了 0.005s，图 4.4 是模型 B 的训练结果。

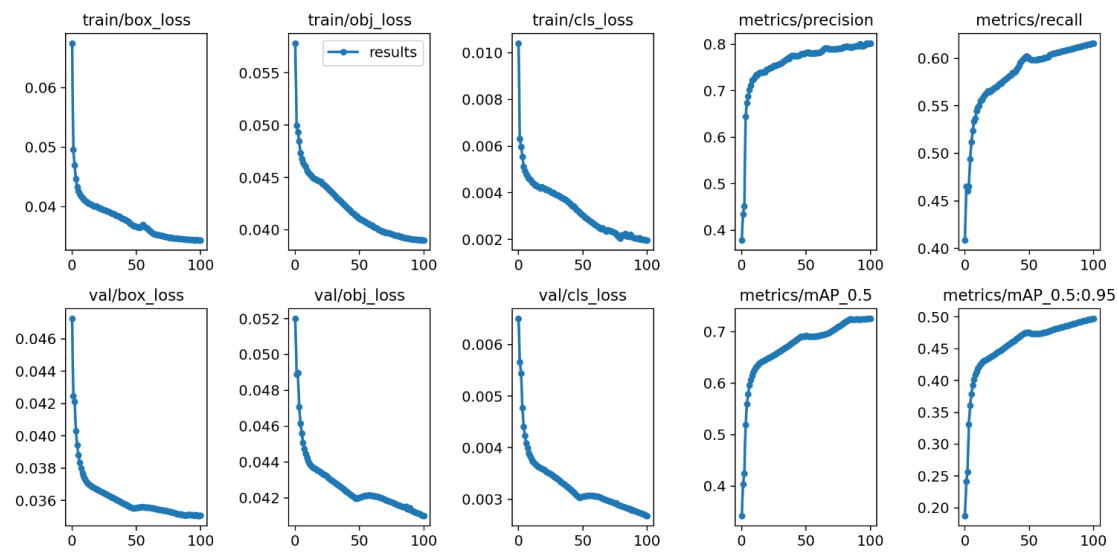


图 4.4 训练结果

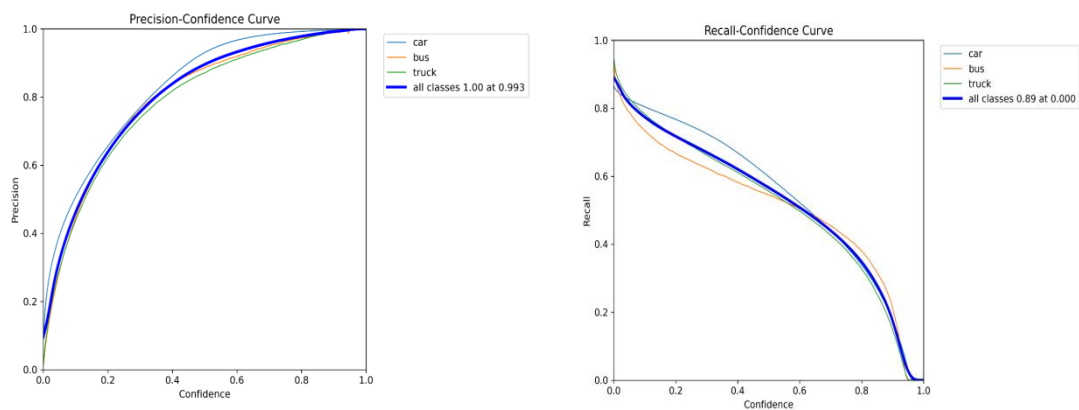


图 4.5 YOLO-DFMnet 的精确率-置信度和召回率-置信度的性能曲线



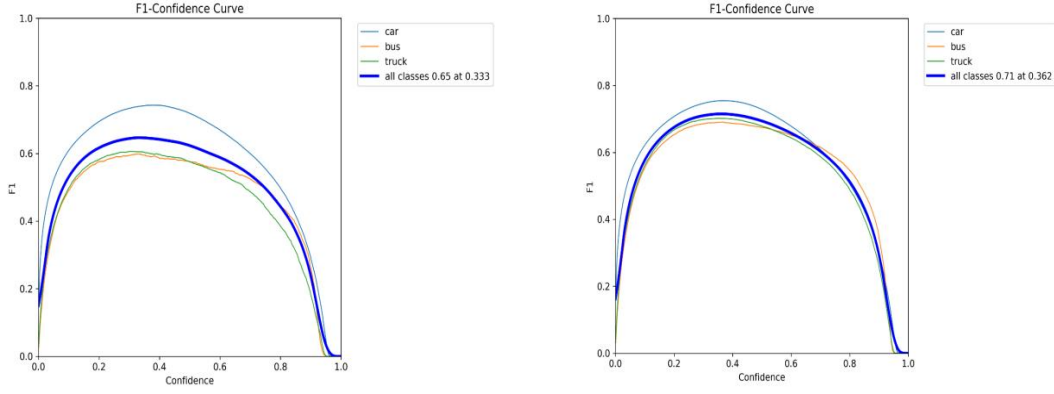


图 4.6 F1 置信度曲线对比图

基于先前计算的评估指标，本文进一步引入了一系列补充性评价指标，旨在为所提模型的性能提供一个直观且全面的剖析。这些额外的度量标准涵盖 F1-confidence 曲线、Precision-confidence 曲线以及 Recall-confidence 曲线，具体展示于图 4.5 与图 4.6 中。通过这些扩展的评估工具，我们能够获得关于模型在识别及分类图像中车辆任务上的性能更为细腻且深入的理解。精确率-召回率曲线与 F1 分数曲线共同描绘了在变化的决策阈值背景下，精确率与召回率之间的微妙平衡，揭示了模型在不同信任度要求下的表现权衡。Precision-confidence 曲线的分析强调了高置信度预测与高精度结果之间的正向关联，意味着模型在高度确信时的预测准确性更高。F1-confidence 曲线的深入探讨指出，本研究构建的模型在设定的置信阈值为 0.362 时，达到了最佳的 F1 分数 0.71，这一发现凸显了模型在特定置信度水平下的最优综合性能。

通过对比 F1-confidence 曲线，明显观察到 YOLO-DFMnet 在处理具有挑战性的样本类别（即车辆中的巴士和卡车）时，其识别精度实现了显著提升，并且这种提升促使这两类对象的识别性能更加接近“汽车”类别的识别曲线，显示出模型在提高各类别识别均衡性方面的有效改进。

#### 4.3.2 消融实验

如表 4.2 所示，是各部分的消融实验。表 4.3 是目标平均准确率（AP）对比，图 4.4 是目标类别的 PR 曲线，实验结果明确地揭示了 YOLO-DFMnet 在应对样本分布不均的挑战时，对该网络针对稀疏检测对象的聚焦能力有显著增强，极大地推动了对稀疏目标检测精度的跃升，并促成了全面检测效能的优化。集成 Hard-Swish 激活函数与注意力机制 CBAM 的策略，有效提升了特征图的表达质量和促进了跨层级特征的有效融合，进一步辅助 DFMS-IoULoss 损失函数有效解决了复杂样本的识别难题，彰显了该方法在提高检测系统鲁棒性和准确性方面的潜力。

综合分析表 4.3、图 4.7 的 PR 曲线以及图 4.8 呈现的混淆矩阵，可以明确观



察到，在本文的模型中，car 这一大样本类别的识别精度经历了轻微但稳定的增长。尤为值得注意的是，针对样本数量极为有限的 Bus 与 Truck 类别，识别准确率实现了显著的提升，增幅逾越 10% 阈值，这一进步凸显了模型在处理罕见类别时的优越性能。进一步探究 PR 曲线，可以发现，随着小样本类别识别性能的增强，三类对象的 PR 曲线趋向更加紧密的拟合状态，这表明各类别间的检测均衡性得到了实质性的改善，进而对总体样本集合的检测精度产生了积极的推动作用，实现了全面且显著的性能提升。

表 4.2 各部分消融实验结果

YOLO	DFMS-IOU	Hard-Swish	CBAM	P	R	mAP@0.5
✓	×	×	×	0.73642	0.57304	0.65154
✓	×	✓	×	0.75292	0.57036	0.65456
✓	×	✓	✓	0.77629	0.60504	0.69247
✓	✓	×	×	0.76165	0.59095	0.71570
✓	✓	✓	✓	<b>0.81126</b>	<b>0.64079</b>	<b>0.73558</b>

表 4.3 目标平均准确率 (AP) 对比

目标类别	YOLO	YOLO-DFMnet
Car	0.762	<b>0.777</b>
Bus	0.596	<b>0.701</b>
Truck	0.606	<b>0.729</b>
all classes	0.655	<b>0.736</b>

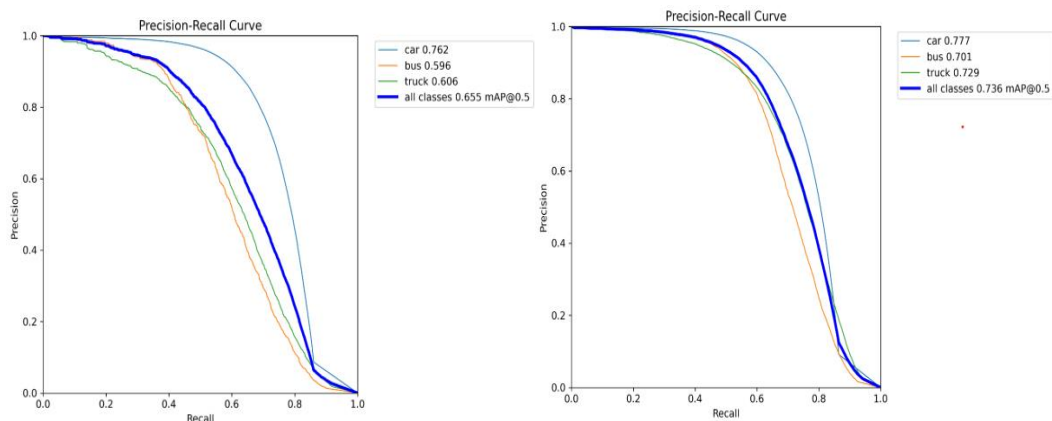


图 4.7 PR-curve

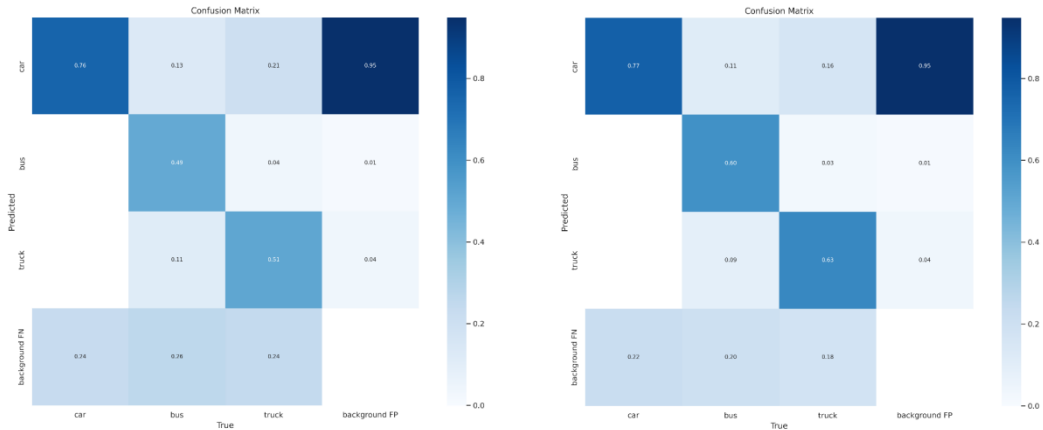


图 4.8 confusion matrices

### 4.3.3 不同模型性能比较

表 4.4 各个网络结果分析

Network	mAP@0.5(%)	Speed(fps)
MultiNet	60.2	8.6
DLT-Net	68.4	9.3
Faster R-CNN	64.9	8.8
YOLOv5s	65.1	71.4
YOLOP	76.5	41
YOLO-DFMnet	73.6	111

如表 4.4, YOLO-DFMnet 在精确度上取得了 73.6% 的成绩, 相比于 MultiNet、DLT-Net 和 Faster R-CNN 有明显的提升。虽然 YOLOP 在 mAP@0.5 上表现出色, 达到了 76.5%, 原因可能是 epoch 次数较多(本实验数据均为 100 个 epoch 的结果), 但其速度只有 41fps<sup>[19]</sup>, 远低于 YOLO-DFMnet 的 111fps。YOLO-DFMnet 在保证一定检测精度的前提下, 拥有更快的速度, 更适合实时应用。

### 4.3.4 检测结果比较

在图 4.9 中, 呈现了原始模型与 YOLO-DFMnet 在多种复杂场景下的物体检测效果对比。从左至右, 依次是原始图像、原模型的检测结果以及我们的模型的检测结果。通过对比分析, 可以明确地看出 YOLO-DFMnet 在处理复杂路况、低光照环境以及图像质量欠佳等情况时, 具备更高的检测精度和车辆类别识别率。无论是晴天、雪天还是夜晚, YOLO-DFMnet 都能准确地识别出道路上的各种车辆类型, 包括汽车、公共汽车和卡车等, 而原模型则在某些情况下出现了漏检或误检的问题, 进一步证实了 YOLO-DFMnet 在应对真实世界复杂驾驶环境时的优

越性能。



图 4.9 模型检测对比图

实验表明 YOLO-DFMnet 在自动驾驶与智慧交通相关数据集上的车辆目标检测任务中表现出显著优势,关键指标如 P-R 曲线、mAP 值、FPS 等均优于 YOLOv5 及其他主流方法。消融实验验证了各组件的独立贡献与协同效应,可视化分析揭示了模型在复杂场景下精准检测车辆目标及注意力机制引导关键特征识别的能力。YOLO-DFMnet 通过集成 DFMS-IoULoss、改进激活函数与 CBAM 注意力机制。有效地处理了在大规模数据集中固有的类别分布不均衡情况下,现有模型对多数类学习的偏好性的问题,实现了高精度、实时、鲁棒的检测,为提升自动驾驶安全性和智能交通管理效率提供有力的技术支撑。本研究不仅丰富了目标检测理论与方法论,也为自动驾驶与智慧交通技术的未来发展提供了高质量的实践指导与理论借鉴。

## 第 5 章 总结和展望

### 5.1 总结

本文针对自动驾驶和智能交通系统中车辆目标检测的挑战，尤其是类别分布不均和稀有目标识别难题，提出了 YOLO-DFMnet 的新型车辆检测框架。该框架引入了动态聚焦调制平滑 IoU 损失函数（DFMS-IoULoss），该损失函数动态调整样本难度相关的损失权重，关注难以检测的目标，有效解决了样本不均衡问题。同时，结合 Hard-swish 激活函数和卷积块注意力模块（CBAM）进一步增强模型的特征表示能力，提高了对车辆细微特征的捕捉和背景干扰的抑制，促进了模型的精度和鲁棒性。

经过详尽的实验验证，本文提出的 YOLO-DFMnet 模型在 BDD100K 这一自动驾驶与智慧交通数据集上，不仅将平均精度（mAP）提升了显著的 8.4%，还在各个类别上实现了全面的精度提升，尤其在 bus 和 truck 这两个类别上，检测性能的提升幅度超过了 10%，这一进步突显了模型在解决类别分布不均衡问题上的有效性。与诸如 MultiNet、DLT-Net、Faster R-CNN 等其他经典模型相比，YOLO-DFMnet 在 mAP 指标上实现了显著超越，达到了 73.6%。特别值得注意的是，虽然 YOLOP 在 mAP@0.5 上达到了 76.5%，与 YOLO-DFMnet 相近，但 YOLO-DFMnet 凭借其 111fps 的高速推理时间，在实时性上远超 YOLOP 的 41fps，这对于需要即时响应的自动驾驶场景来说至关重要。这表明 YOLO-DFMnet 在维持较高检测精度的同时，优化了计算效率，使其在资源有限的嵌入式设备上也能高效运行，满足了实时应用的严格需求。

在处理复杂场景，如低光照、遮挡或是目标密集情况时，YOLO-DFMnet 依旧展现出卓越的检测性能，其强大的鲁棒性和准确性在与原始模型的直接对比中更为突出。通过模型的检测结果比较，我们可以清晰地看到，YOLO-DFMnet 能更准确地识别出各种条件下的车辆，即便是图像质量不佳或者目标信息模糊的情况也不例外，进一步证明了其在实际应用中的可靠性。YOLO-DFMnet 在车辆检测任务上不仅实现了精度与速度的双重飞跃，还克服了复杂场景下的挑战，完全符合自动驾驶和智慧交通系统对于快速、精确车辆检测的高标准要求，为未来的智能交通基础设施建设提供了坚实的基石。

## 5.2 展望

尽管 YOLO-DFMnet 在车辆检测任务中展现出了显著的性能提升,尤其在处理类别分布不均、尺度变异及复杂场景方面,未来研究需深化几个关键点:智能化调节动态聚焦策略以更好适应各场景特征,优化 CBAM 模块提高注意力机制效率同时减轻过拟合风险,尤其是在大规模数据应用中,且需探索轻量化设计,确保高速推理满足实时应用需求。后续研发应侧重于模块效率与效果的提升、模型泛化能力和极端条件下的性能增强,推动车辆检测领域更全面深入的发展。

## 参考文献

- [1] 李明熹,林正奎,曲毅. 计算机视觉下的车辆目标检测算法综述[J].计算机工程与应用,2019,55(24):20-28.
- [2] 陈凯. 基于深度学习的遥感影像目标检测与显示技术研究[D].西安电子科技大学,2020.
- [3] Tsai DuMing, Lai ShiaChih. Independent Component Analysis-Based Background Subtraction for Indoor Surveillance.[J]. IEEE transactions on image processing,2009,18(1):158-167.
- [4] Dar-Shyang Lee. Effective Gaussian mixture learning for video background subtraction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(5): 827-832.
- [5] Horn Berthold K, Schunck Brian G. Determining Optical Flow[J]. Massachusetts Institute of Technology (United States),1981,17(1-3):185-203.
- [6] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[M]. arXiv, 2016[2023-08-02]. <http://arxiv.org/abs/1506.02640>.
- [7] Zhang Y F, Ren W, Zhang Z, et al. Focal and Efficient IOU Loss for Accurate BoundingBoxRegression[M].arXiv,2022[2023-09-08].
- [8] Gevorgyan, Zhora. "SIoU Loss: More Powerful Learning for Bounding Box Regression." ArXiv abs/2205.12740 (2022): n. pag.
- [9] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection n[M]. arXiv, 2018[2023-09-25]. <http://arxiv.org/abs/1708.02002>.
- [10] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[M]. arXiv, 2020[2023-09-22]. <http://arxiv.org/abs/2004.10934>.
- [11] WOO S Y, PARK J C, LEE J Y, et al. CBAM:Convolutional block attention module[C]//Proceedings of the 2018 European Conference on Computer Vision(ECCV). 2018:3-19.
- [12] Min Lin, Qiang Chen 0007, Shuicheng Yan. Network In Network.[J]. CoRR,2013,abs/1312.4400.
- [13] Zhou B, Khosla A, Lapedriza A, et al. Learning Deep Features for Discriminative Localization[C]//2016 IEEE Conference on Computer Vision and Pattern

- n Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 2921-2929.
- [14]Meng Zhe, Zhao Feng, Liang Miaomiao. SS-MLP: A Novel Spectral-Spatial MLP Architecture for Hyperspectral Image Classification[J]. Remote Sensing, 2021, 13(20):4060.
- [15]Yu F, Chen H, Wang X, et al. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 2633-2642.
- [16]Zhen J, Xie B. Fused attention mechanism-based ore sorting network[M]. arXiv, 2024[2024-05-20]. <http://arxiv.org/abs/2405.02785>.
- [17]Xiong C, Zayed T, Abdelkader E M. A novel YOLOv8-GAM-Wise-IoU model for automated detection of bridge surface cracks[J]. Construction and Building Materials, 2024, 414: 135025.
- [18]Zhang J, Zhang R, Yan X, et al. BG-YOLO: A Bidirectional-Guided Method for Underwater Object Detection[M]. arXiv, 2024[2024-05-20]. <http://arxiv.org/abs/2404.08979>.
- [19]Han C, Zhao Q, Zhang S, et al. YOLOPv2: Better, Faster, Stronger for Panoptic Driving Perception[M]. arXiv, 2022[2024-05-10]. <http://arxiv.org/abs/2208.11434>.

## 致谢

在此论文即将完成之际，我心中满怀感激之情，特借此篇章向那些在我求学路上留下深刻印记的人们表达最诚挚的感谢。

我衷心感谢江西师范大学这片知识的沃土，以及风景如画的青山湖校区，四年的时光里，这里不仅是我学术探索的起点，更是我成长道路上的一片温馨港湾。校园的每一处景致，每一次晨读夜习，都见证了我从青涩到成熟的蜕变。

特别感谢我的两位辅导员——赵老师与廖老师。赵老师温暖人心的关怀和廖老师严谨治学态度，如同灯塔般照亮了我前行的道路。在我迷茫时给予方向，在挑战面前鼓励我勇往直前，是他们用无私的付出，为我的大学生活增添了无限色彩。

在此，我更要向我的论文指导老师——王文乐老师致以最深的敬意和感激。王老师不仅在专业领域内以其渊博的知识引领我深入研究，更在生活中给予了我家人般的关怀和支持。您的每一次悉心指导，每一个耐心解答，都是我论文顺利完成不可或缺的力量源泉。您的教诲如同春雨般润物无声，滋养了我的学术理想，也教会了我如何以更加坚韧和乐观的态度面对人生的风雨。

同时我也要感谢同窗好友们的陪伴与互助，我们在图书馆的并肩作战，在宿舍的深夜畅谈，都是这段青春记忆中最宝贵的片段。你们的存在，让这段旅程充满了欢笑与温暖。

最后我要向我的家人表达最深切的感激。是家人的理解与支持，构成了我不断前进的强大后盾。无论是在精神上的鼓励还是生活上的照顾，都是我能够专心学业、勇敢追梦的基石。

这篇论文的每一页都凝聚了众多师长与亲朋的关爱与期待。虽然文字有限，但感恩之情无尽。未来，我将铭记这份恩情，带着在这里收获的一切，继续在人生的旅途上探索前行，努力成为一个对社会有贡献的人。

再次感谢所有在我成长道路上留下足迹的人，愿我们都能在各异的领域绽放光芒，共赴美好未来。