西安交通大学

博士学位论文

多光谱融合智能光电处理算法与系统设计

学位申请人：陈炜煌

指导教师：孙宏滨教授

学科名称：控制科学与工程

2025 年 12 月

Intelligent Electro-Optical Processing Algorithm and Systems Design based on Multispectral Fusion

A dissertation submitted to

Xi'an Jiaotong University

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

By

Weihuang Chen

Supervisor: Prof. Hongbin Sun

Control Science and Technology

December 2025

博士学位论文答辩委员会

多光谱融合智能光电处理算法与系统设计

答辩人：陈炜煌

答辩委员会委员：

西安交通大学教授：辛景民（注：主席）

西北工业大学教授：王鹏

西安电子科技大学教授：董伟生

西安交通大学教授：魏平

西安交通大学教授：杜少毅

答辩时间：2023 年 05 月 14 日

答辩地点：西安交通大学科学馆 324

摘 要

自动驾驶在节约驾驶成本、提高交通效率、减少环境污染等方面拥有巨大优势，成为了学术界和工业界的热门研究课题。为了实现安全可靠、稳定高效的驾驶行为，自动驾驶车辆需要精准地预测出周围环境中交通参与者的未来行为轨迹，并规划出自身无碰撞且运动学可行的短时运动轨迹。传统轨迹预测方法无法保证长期预测的精度，严重依赖启发式设计的传统运动规划方法也无法保证其泛化性能。近年来，基于数据驱动的深度学习方法得到了快速发展，为完成预测规划任务带来了新思路。从数据输入输出的角度考虑，预测和规划都是对交通参与者的历史特征进行建模后输出未来轨迹。因此，这两项具有共性的任务均可以采用具有强大特征拟合能力的深度学习方法来完成。然而，此类方法仍然存在交通参与者异质性处理能力差，缺少概率性预测结果以及无法保证轨迹平滑性等问题，使得自动驾驶的安全性受到威胁，阻碍了自动驾驶技术进一步的发展。

本文聚焦于利用 Transformer 网络解决上述核心难点问题：1）如何构造更精准、更快速的实用化轨迹预测网络模型；2）如何在保证完成自动驾驶任务的前提下促使运动规划方法尽可能地减少交通违规行为。主要研究工作如下。

1. 提出了一种基于时空 Transformer 网络的单模态轨迹预测网络模型，弥补了之前方法只能有效预测同质交通参与者的缺陷，提高了密集交通环境下时空交互建模能力。针对之前方法对时间序列数据进行串行处理造成的记忆能力弱以及空间邻域范围设置不合理等问题，该方法采用 Transformer 网络并构建了全感知域的时空图模型。整个网络包括时空 Transformer 编码器、时间 Transformer 编码器和时间 Transformer 解码器三个部分。时空 Transformer 编码器能够对时空图特征按照不同维度交替提取，从而充分融合时空信息。经过时间 Transformer 编码器对于时间信息的进一步处理后，时间 Transformer 解码器生成了关于异质交通参与者的单模态轨迹。在自动驾驶轨迹预测公开数据集上的实验结果表明，该方法比当时最好的方法在主要性能指标上提高了至少 7.2%。

2. 提出了一种基于概率性候选轨迹网络的多模态轨迹预测网络模型，在加快模型推理速度的同时，提高了多模态轨迹预测的精度。针对当前多模态轨迹预测方法无法提供概率性预测结果的问题，该方法设计了一种既能生成目标点引导信息，又能提供概率性结果的三阶段轨迹预测过程。首先，该方法利用无监督学习自动获取交通参与者的潜在意图集合，并应用分类网络筛选出符合当前交通参与者运动趋势的概率性目标点集合。然后，通过 Transformer 网络生成中间位置锚点。最后，使用连续曲线光滑连接当前位置、锚点和目标点，形成表达能力更强的概率性候选轨迹集。多个公开轨迹预测数据集的实验结果验证了该方法在提供高性能、高效率的概率性预测结果的同时，能够确保概率较高的预测结果更符合交通参与者的下一步行为。

3. 提出了一种基于安全轨迹树网络的运动规划网络模型，减少了之前基于学习的运动规划方法在完成自动驾驶任务时出现的大量交通违规行为。针对之前方法因不能满足相关运动约束而造成的违规问题，该方法提出了一种具有曲率连续性和运动学可行性的轨迹树。该轨迹树既能够用于运动规划主任务，也能够作用于共性的轨迹预测辅助任务，从而帮助模型通过学习预测规划间的交互提升性能。针对高维栅格化特征输入可解释性差、计算效率低的问题，该方法采用包含交通参与者和局部任务路线的离散化输入表达方式，增加了模型的可解释性。该方法还利用 Transformer 主干网络精准提取不同输入之间的空间交互信息。针对自动驾驶汽车在复杂场景中保持长期静止不动的问题，该方法在训练过程中引入了焦点损失函数，鼓励自动驾驶车辆安全高效地完成导航任务。多个自动驾驶闭环测试基准的实验结果表明，该方法不仅在自动驾驶任务完成度和违规得分方面比之前最好的方法分别提高了 39.2% 和 10.6%，而且推理速度加快了 1.5 倍。

综上所述，本文所提出的单模态轨迹预测、多模态轨迹预测和运动规划方法获得了高性能的表现，具有精度高、速度快和违规驾驶行为少的优势，为保证自动驾驶安全性发挥了重要作用。

关 键 词： 自动驾驶；轨迹预测；运动规划；自注意力模型

论文类型： 应用研究

ABSTRACT

Autonomous driving is an innovative and advanced research field in academia and industry, with potential to reduce road fatalities, improve traffic efficiency, and decrease environmental pollution. To achieve safe, reliable, stable, and efficient driving behavior, autonomous vehicles need to accurately predict the future trajectories of surrounding traffic participants and plan collision-free, kinematically feasible short-term motion trajectories. Traditional trajectory prediction methods often lack accuracy in long-term predictions, while motion planning methods based on heuristic design may lack generalization performance. In recent years, rapid advancements in data-driven deep learning methods have revolutionized prediction and planning tasks. Deep learning methods offer powerful feature fitting capabilities that enable accurate modeling of historical traffic patterns and output of future trajectories. Consequently, both prediction and planning tasks can be achieved using deep learning techniques. Despite these remarkable benefits, these methods still face various challenges, such as poor ability to deal with traffic participant heterogeneity, lack of probabilistic prediction results, and inability to guarantee trajectory smoothness. These issues pose significant safety concerns for autonomous driving and impede the further advancement of this technology.

This dissertation proposes to leverage Transformer network to address these core difficulties. The objectives of this study are twofold: 1) improving the accuracy and inference speed of practical trajectory prediction models, 2) enhancing motion planning methods to minimize traffic violations while guaranteeing the completion of autonomous driving tasks. The main contributions of this research are as follows.

1. This dissertation proposes a new Spatio-Temporal Transformer Network for unimodal trajectory prediction, which addresses the limitations of previous homogeneous prediction methods and improves spatio-temporal interactive modeling capabilities. To address the problems of weak memory ability and unreasonable setting of spatial neighborhood range caused by the serial processing of time series data in the previous method, we adopt Transformer network and constructs a spatio-temporal graph of the whole perceptual domain. The network consists of three parts, i.e. spatio-temporal Transformer encoder, temporal Transformer encoder, and temporal Transformer decoder. The first Transformer encoder extracts spatio-temporal features by alternating between different dimensions to fully integrate spatio-temporal information. The second Transformer encoder further processes temporal information, and the temporal Transformer decoder generates unimodal trajectories for heterogeneous traffic participants. Experimental results demonstrate that the proposed method enhances key performance metrics by at least 7.2% over state-of-the-art methods.

2. This dissertation proposes a new Probabilistic Proposal Network for multimodal trajectory prediction which not only enhances the prediction accuracy of multimodal trajectory prediction, but also accelerates the inference speed. To address the problem that previous multimodal trajectory prediction methods cannot provide probabilistic prediction results, we devise a three-stage trajectory prediction process that generates target point guidance information and provides probabilistic outcomes. Firstly, the proposed method employs unsupervised learning to automatically obtain the potential intention set of traffic participants and applies a classification network to filter out a set of probabilistic target points that comply with the current movement trend of traffic participants. Next, Transformer network generates intermediate position anchors. Finally, a continuous curve is used to smoothly link the current position, anchors, and target point, producing a more expressive set of probabilistic trajectory candidates. Experimental results demonstrate that the proposed method yields high-performance and high-efficiency probabilistic prediction results while ensuring that the prediction results with higher probability align more closely with the next behavior of traffic participants.

3. This dissertation proposes a new safe Trajectory Tree Network for motion planning, which can effectively reduce traffic violations while completing autonomous driving tasks. The key component of TTNet is a predefined trajectory tree that conforms to vehicle dynamics constraints and explicitly reflects different intentions. This tree is used for both the main planning task and an auxiliary trajectory prediction task. To enhance interpretability, we introduce input expressions typically used in traditional planning algorithms into our integrated framework. Additionally, to promote safe and efficient navigation, we incorporate a focal loss during training and employ a Transformer-based backbone network to accurately capture spatial interactions not only among the ego vehicle and its surroundings, but also among dynamic agents and the reference line. Experimental results demonstrate that the proposed method significantly improves task completion and violation scores by 39.2% and 10.6%, respectively, compared to SOTA methods while accelerating the inference speed by 1.5 times.

In summary, our proposed methods achieve outstanding performance for unimodal trajectory prediction, multimodal trajectory prediction and motion planning, with the advantages of high precision, high speed and less driving violations, thus playing crucial roles in ensuring the safety of autonomous driving.

KEY WORDS: Autonomous Driving; Trajectory Prediction; Motion Planning; Transformer

TYPE OF DISSERTATION: Application Research

目 录

CONTENTS

1 结论与展望

1.1 结论

本文围绕“多光谱融合智能光电处理算法与系统设计”这一主题，针对无人机在低空复杂环境中对实时、鲁棒感知与跟踪的迫切需求，展开了一系列从理论方法、关键算法到工程系统的深入研究。现有目标检测跟踪算法在无人机视角下仍存在技术瓶颈，对于像素占比极小的目标，其检测精度仍有较大提升空间，对于长时被遮挡目标，现有目标跟踪算法缺乏可靠的丢失判定与重检测机制。本文通过将先进的人工智能算法与严格的嵌入式工程约束相结合，致力于解决机载平台在有限算力、内存和功耗下实现高精度环境感知的挑战，构建了一套从核心处理算法到完整软硬件系统的完整解决方案。本文主要研究成果包括以下四个方面。

(1) 提出了面向可见光航拍图像的高性能小目标检测网络 BAP-DETR，在无人机航拍图像的目标检测任务中，算法需应对极端尺度变化、密集小目标不均匀分布及复杂背景干扰三重挑战。通用目标检测器在此类场景下存在显著性能差距，而直接提高输入图像分辨率又会引发计算量激增，难以满足机载平台的实时性要求。现有小目标检测方法的网络架构往往在特征传递过程中丢失对小目标至关重要的细粒度信息，导致精度与速度难以兼顾。针对这些问题，本文设计了双重注意力处理模块，通过通道分离策略实现卷积与自注意力全局建模能力的并行优化与交互，使模型能从复杂场景中提取更具判别力的细微特征。其次，设计了配备频域感知融合模块的双融合编码器，该设计能有效保留并融合包含丰富细节的低层特征与承载语义信息的高层特征，显著增强了对小目标的特征保留能力与多尺度感知性能。最后，在损失函数中，结合倒数归一化 Wasserstein 距离与 CIoU 损失，在不增加推理开销的前提下，进一步提升了对小目标定位的精确度与鲁棒性。在 VisDrone、UAVDT 和 AI-TOD 三个公开航拍数据集上的实验表明，BAP-DETR 在保持推理效率的同时，平均检测精度（AP）较基线模型提升 6.9%，并实现了 17.5% 的计算负载降低。这项工作为无人机可见光视角下的实时高精度小目标检测提供了一个有效的解决方案，有效平衡了机载场景下对精度与速度的要求。

(2) 设计了面向无人机红外图像的轻量化小目标检测网络 MFF-DCNet，无人机红外成像在夜间、烟尘等恶劣环境下具有不可替代的优势，但其图像固有的低分辨率、低对比度、高噪声及纹理缺失特性，使得目标检测，尤其是对远距离像素占比极小的微弱目标的检测，成为极具挑战性的难题。现有通用检测模型直接迁移至红外图像时性能显著下降，而计算密集的先进网络又难以在机载边缘设备的严格算力约束下实现实时推理。针对这些问题，本文设计了深度可分离跨阶段 Transformer 模块用于增强主干网络的特征提取能力，该结构有效建模了远

距离弱小目标与复杂背景之间的长距离上下文依赖关系，增强了特征的判别性。其次，提出了一个新颖的多特征聚焦颈部结构，通过自适应的跨尺度特征加权与融合策略，提升了网络对小目标特征的提取能力。在 HIT-UAV 和 DroneVehicle 红外航拍数据集上的实验表明，MFF-DCNet 不仅检测精度（AP）显著超越专用无人机图像检测器及 YOLO、DETR 系列等基线模型，更在处理效率上实现了提升。同时，该网络在 NVIDIA Jetson Orin NX 嵌入式边缘计算平台上达到了 39.6 FPS 的实时处理能力，验证了其满足实际机载任务对低功耗、高实时性的要求，为无人机全天候智能感知提供了可靠的红外视觉解决方案。

(3) 提出了一种面向边缘计算设备的抗遮挡长时目标跟踪框架 SKF-Tracker，在无人机执行持续监视、目标跟踪等任务时，目标被环境中的建筑、植被等障碍物完全遮挡是导致跟踪失败的主要原因。现有主流跟踪算法及公开数据集多聚焦于短时、部分遮挡的场景，且缺乏无人机视角下的图像，导致算法在实际复杂城市场景中鲁棒性不足。针对这一问题，本文设计了一种具备目标丢失判断与重捕获能力的抗遮挡长时跟踪框架，使用结构相似性指数作为跟踪置信度判断依据，结合轨迹预测与动态搜索的重捕获机制，使系统能够在识别到目标被遮挡的情况后，及时暂停目标模板更新，在目标重现时快速锁定，同时，自适应的模板更新策略确保了模型外观记忆的可靠性。为了针对性地验证算法抗遮挡能力，本文构建了一个多模态（可见光与红外）无人机视角目标跟踪数据集 MMUOT-1050，包含 353 段可见光视频序列与 697 段红外视频序列，每段视频均包含目标被完全遮挡的场景。SKF-Tracker 在该数据集上实现了 89.37% 的可见光视频成功率和 91.93% 的红外视频成功率，与基线方法相比提升了 14% 和 11.73%。在实时性方面，SKF-Tracker 在保持最高精度的同时，仍能保持 31.25 的帧率，其速度远超基于深度神经网络的 SiamRPN，并且不占用 NPU 资源。该工作为解决面向边缘计算设备的抗遮挡目标跟踪提供了一套兼具理论创新与工程落地的可行路径。

(4) 设计并实现了一套能够满足低功耗、高实时性要求的机载智能光电原型系统，将算法创新融入到完整的工程实践中。该系统采用高性能边缘计算模组，集成了模块化的多光谱数据采集、处理与通信软件框架，并通过大疆 M350 RTK 无人机平台进行了实地飞行验证。在硬件层面，系统以高性能异构边缘计算模组（RK3588）为核心，集成了可见光与红外传感器，实现了多光谱数据的采集与处理。利用芯片厂商提供的专用工具链（RKNN），对网络模型进行了针对性的优化，解决了网络模型在边缘侧部署的核心瓶颈。在软件层面，本文构建了一套模块化、高内聚、低耦合的嵌入式应用软件框架。数据接入层以工厂模式管理多源异构传感器，数据处理层以策略模式封装并动态调度检测、跟踪等智能算法，通信总线层以事件驱动与消息队列实现模块间异步通信，通过无锁环形缓冲区和内存池保障了数据流的高效与确定性。同时，系统提供了支持多协议的上位机交互接口并支持算法参数高度可配置，大幅提升系统的可扩展性与工程

可维护性。

本文围绕低空无人机在复杂环境下的智能感知需求，从理论方法、核心算法、系统实现到实验验证，实现了从算法设计到工程实现与性能验证的完整闭环。所提出的BAP-DETR、MFF-DCNet 及 SKF-Tracker 算法，针对无人机视角下的小目标检测与严重遮挡跟踪等难题提供了有效解决方案，基于软硬件协同设计的智能光电原型系统，为算法的工程化落地与性能验证提供了坚实基础。本工作不仅在提升无人机自主感知的精度、鲁棒性与实时性方面取得了进展，同时为机载智能光电系统原型设计提供了经验与参考。

1.2 展望

本文所提出的算法与系统，显著提升了机载平台在复杂环境下对特定目标的感知精度与跟踪鲁棒性，为执行明确的侦察、监视等任务提供了有效的技术支撑。然而，当前的系统能力仍集中在对环境的“感知”与“反应”层面，在结合场景上下文、任务意图进行高层语义理解与自主决策方面，与具备认知能力的智能体（Agent）仍存在本质差距。这一差距的根源，一方面在于现有嵌入式平台的硬件条件，难以承载需要大规模算力的先进模型，如开放世界目标检测、视觉-语言模型（Vision-Language Models, VLM）等，另一方面，在于现有算法多聚焦于单一任务的优化，缺乏对多任务、多模态信息的综合处理与理解能力。后续的研究将在现有基础上，面向资源受限平台与开放环境下智能感知的需求，融合边云协同与模型压缩蒸馏技术以引入具备开放世界识别与跨模态理解能力的模型，强化检测、跟踪、重识别与语义解析的协同工作，形成从低级感知到高层意图理解与策略生成的闭环，具体围绕以下几个方面展开：

（1） 开放世界感知与跨模态信息融合

本文中机载光电系统部署的目标检测模型是固定类别的闭集检测模型，仅能识别在训练集中预先定义好的有限类别目标，但对未见类别与长尾目标的适配能力有限。视觉-语言模型如 CLIP、GLIP 及其变体，具备开放词汇感知与零样本识别能力，这类模型通过在海量图像-文本对上进行对比学习，将视觉特征与语义概念在统一的高维空间中对齐，从而能够仅根据自然语言描述来检测和识别未见过的类别，而无需针对这些类别进行任何模型参数的更新或重新训练，使系统能够识别未知类别与长尾目标，实现对未知环境的感知。后续工作将探索如何在机载环境下融入视觉-语言模型的感知能力，可从硬件升级与架构优化两个方面进行探索，NVIDIA Jetson AGX Orin/Thor 等新一代边缘计算模组可提供数十至上百 TOPS 的 AI 算力，使在端侧部署轻量化视觉-语言模型成为可能，Jetson Orin Nano 上可部署约 3B 参数的蒸馏 VLM，在 AGX Orin 上可部署 70B 参数的模型，在 AGX Thor 上能够运行参数超过 100B 的模型。升级硬件是直接有效的方案，但是需要考虑机载环境下的物理约束，系统功耗，散热设计，成本与可维护性等因素，同时也要考虑整个软件框架迁移的成本。另一方面，在低算力平台上，

可通过边云协同计算架构，将复杂的视觉-语言模型部署在云端服务器上，机载平台仅运行轻量级的前端感知与数据处理模块，将关键帧或特征上传至云端进行推理，云端再将结果反馈至机载平台，实现低延迟的开放世界感知能力。该方案能够利用云端强大的计算资源，同时减轻机载平台的算力负担，但需要解决网络连接的稳定性与带宽限制问题，后续研究将结合模型压缩、蒸馏与边云协同技术，探索在机载环境下实现开放世界感知的可行路径。

（2） 意图理解与自主策略生成

本文实现了机载智能光电系统对目标的精准感知与稳定跟踪，但是主要聚焦于单一任务的优化，缺乏对场景高级语义与任务意图的理解能力。后续研究将探索面向动态任务场景的意图理解与自主策略生成，在目标与场景意图理解层面，发展基于多模态时序数据的目标行为建模与意图推断方法。构建一个融合目标运动状态、外观语境与场景语义的智能推理模型，通过一个通用的多模态编码器，将目标的外观、运动和历史轨迹提取为时空对齐、富含语义的统一特征表示，采用类似 TAMFormer 的多模态 Transformer 解码器，对特征进行时序建模与上下文关联，生成语义化意图标签。在任务驱动的自主策略生成层面，研究将高层意图理解转化为具体的感知与行动策略，使系统能根据任务目标、当前态势理解以及平台自身约束自主制定并动态调整传感器调度策略、平台机动建议以及信息收集优先级。同时，需要设计简单易用的人机交互接口，使自主生成的策略能够以清晰的方式呈现给操作人员，支持“人在回路”的核准、修改与控制，形成混合增强智能的闭环，确保最终决策的可靠性。

（3） 高性能边缘计算平台选型与工程化验证

本文中使用的边缘计算模组只能满足于初级模型的算力需求，主要面向经过优化的卷积神经网络，在面对 DETR 类基于注意力机制、计算复杂度更高的先进模型时，现有边缘算力难以满足实时处理的要求。至于参数规模庞大、需要跨模态对齐的多模态视觉-语言模型，其研究与验证更是高度依赖服务器集群与特定数据集，导致在实际工程应用，尤其是在资源受限的机载平台上，其性能与均缺乏充分验证。要实现一个具备高级认知与自主决策能力的智能体，必须首先构建强大的边缘算力基础，并围绕此基础设计合理的软硬件架构。NVIDIA Jetson AGX Orin/Thor 系列提供了可行的解决方案，其专用的 TensorRT 和 vLLM 服务框架，能够实现在边缘端实时运行包括视觉语言模型、视觉语言动作模型在内的大型生成式模型，平台原生支持 FP8、W4A16 等先进量化格式以及预测性解码等技术，能有效在模型精度、速度和内存占用之间取得平衡，为资源受限的机载平台运行复杂模型提供了可能。瑞芯微 RK3588 的 RKLLM 工具链也为在边缘设备上部署大模型提供解决方案，能够支持 TinyLlama 1.1B、Qwen2.5-1.5B 等参数量在十亿级别的模型，但是目前还处于原型验证阶段，局限于开发板环境，围绕固定模型和示例场景，距离工程化应用还有很大距离。此外，后续研究还应该探索构建实际工程应用场景与系统性测试验证体系，将先进算法从实验室的固定样例转变为能

够在真实、动态的机载环境中稳定工作的工程产品，为机载智能光电系统提供高级认知能力。

致　谢

漫漫求学路，最让人回味的，莫属于读博这几年。回首初入交大时情不自禁的喜悦，经历了硕博八年洗礼后，依旧幸福感满存。谨以此文聊表感激之心。

衷心感谢我的导师孙宏滨教授在博士期间对我的悉心指导与关怀。在刚进组时，孙老师就将新发现号的搭建工作交与我全权负责，帮助我从系统的角度对无人驾驶整体研究有了深刻认识；在博二时，孙老师就让我担任了发现号车队队长并认真细致地指导我们准备每年的未来挑战赛，极大地提高了我的组织管理能力；在科研工作中，孙老师指点迷津，引领我做好科研探索。孙老师严谨务实的科研态度，一丝不苟的治学精神，高屋建瓴的学术见地，勤奋谦虚的个人品质都深深感染着我，激励着我，使我受益终生。

感谢我们敬爱的郑南宁院士。郑老师对于无人驾驶车队的关心和指导使我们整个车队的技术水平得到不断提高。感谢我博士前两年的合作导师辛景民教授的关怀，感谢魏平教授在智能车未来挑战赛备赛和比赛过程中的悉心教导，感谢王乐教授在轨迹预测方面的支持，感谢薛建儒教授、兰旭光教授、任鹏举教授、杜少毅教授、徐林海高级工程师、陈仕韬助理教授、王芳芳工程师以及其他所有人工智能学院老师在我读博期间给予的帮助和支持。

感谢课题组张旭翀师兄和汪航师兄对我科研工作一直以来的帮助，两位师兄扎实的理论功底和极强的解决问题能力都给我留下了很深印象。感谢沈源、张婧、刘丹为我们的学习生活提供的便利。

感谢王潇、史菊旺、李庚欣、陶中幸、张璞等师兄师姐在科研上的关照。感谢冯洋、杨帅、吴金强、冯超、向钊宏、陈达、张志浩、王玉学、韩伟光、权柄章、钱成龙、葛冲、陈科、李诚、罗鑫凯、陈煜炜、王申奥、李天航等师弟在发现号无人驾驶平台开发和无人车比赛中的付出。感谢戴赫、孙长峰、郑方、段景海、石刘帅等师弟在小论文上的帮助。感谢同届张剑、杨少飞、李宝婷的帮助。感谢唐浩雯师妹在科研生活中的交流与帮助。感谢好友冯立琛、丁兆伦、雷洁、马晨、荣韧闲暇时度过的快乐时光。感谢和我一起在创新港并肩战斗的赵博然，在科研和为人处世方面都对我产生了很大影响。

最后，感谢我的父母和家人多年来对我学习和生活上的关心和支持，是你们的坚强后盾让我能够全身心地投入到科研探索中。感恩一路有你们相伴，你们永远是我内心最温暖的港湾。

参考文献

[1] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection[J/OL]. ArXiv, 2020, abs/2004.10934. https://api.semanticscholar.org/CorpusID:216080778.

[2] Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond Empirical Risk Minimization[C/OL]. 2018. https://openreview.net/forum?id=r1Ddp1-Rb.

[3] Yun S, Han D, Oh S J, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2019: 6023-6032.

[4] Kisantal M, Wojna Z, Murawski J, et al. Augmentation for small object detection[J]. arXiv preprint arXiv:1902.07296, 2019.

[5] Chen C, Zhang Y, Lv Q, et al. RRNet: A Hybrid Detector for Object Detection in Drone-Captured Images[C/OL]. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). 2019: 100-108. DOI: 10.1109/ICCVW.2019.00018.

[6] Xiao J, Guo H, Zhou J, et al. Tiny object detection with context enhancement and feature purification [J]. Expert Systems with Applications, 2023, 211: 118665.

[7] Ünel F Ö, Özkalayci B O, Çiğla C. The Power of Tiling for Small Object Detection[C/OL]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2019: 582-591. DOI: 10.1109/CVPRW.2019.00084.

[8] Yu X, Gong Y, Jiang N, et al. Scale Match for Tiny Person Detection[C/OL]. 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). 2020: 1246-1254. DOI: 10.1109/WACV45572.2020.9093394.

[9] Lin J, Jing W, Song H. SAN: Scale-aware network for semantic segmentation of high-resolution aerial images[J]. arXiv preprint arXiv:1907.03089, 2019.

[10] Zoph B, Cubuk E D, Ghiasi G, et al. Learning data augmentation strategies for object detection[C]. European conference on computer vision. 2020: 566-583.

[11] Yang F, Choi W, Lin Y. Exploit All the Layers: Fast and Accurate CNN Object Detector with Scale Dependent Pooling and Cascaded Rejection Classifiers[C/OL]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 2129-2137. DOI: 10.1109/CVPR.2016.234.

[12] Cai Z, Fan Q, Feris R S, et al. A Unified Multi-scale Deep Convolutional Neural Network for Fast Object Detection[C]. Leibe B, Matas J, Sebe N, et al. Computer Vision – ECCV 2016. Cham: Springer International Publishing, 2016: 354-370.

[13] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[EB/OL]. 2018. https://arxiv.org/abs/1804.02767. arXiv: 1804.02767 [cs.CV].

[14] Lin T Y, Dollár P, Girshick R, et al. Feature Pyramid Networks for Object Detection[C/OL]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 936-944. DOI: 10.1109/CVPR.2017.106.

[15] Ghiasi G, Lin T Y, Le Q V. NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection[C/OL]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 7029-7038. DOI: 10.1109/CVPR.2019.00720.

[16] Qiao S, Chen L C, Yuille A. DetectoRS: Detecting Objects with Recursive Feature Pyramid and Switchable Atrous Convolution[C/OL]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 10208-10219. DOI: 10.1109/CVPR46437.2021.01008.

[17] Li J, Liang X, Shen S, et al. Scale-Aware Fast R-CNN for Pedestrian Detection[J/OL]. IEEE Transactions on Multimedia, 2018, 20(4): 985-996. DOI: 10.1109/TMM.2017.2759508.

[18] Yang C, Huang Z, Wang N. QueryDet: Cascaded Sparse Query for Accelerating High-Resolution Small Object Detection[C/OL]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022: 13658-13667. DOI: 10.1109/CVPR52688.2022.01330.

[19] Singh B, Davis L S. An Analysis of Scale Invariance in Object Detection - SNIP[J/OL]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017: 3578-3587. https://api.semanticscholar.org/CorpusID:4615054.

[20] Singh B, Najibi M, Davis L S. SNIPER: Efficient Multi-Scale Training[J]. NeurIPS, 2018.

[21] Najibi M, Singh B, Davis L S. AutoFocus: Efficient Multi-Scale Inference[J]. ICCV, 2019.

[22] Chen Y, Zhang P, Li Z, et al. Dynamic Scale Training for Object Detection[EB/OL]. 2021. https://arxiv.org/abs/2004.12432. arXiv: 2004.12432 [cs.CV].

[23] Li Y, Chen Y, Wang N, et al. Scale-Aware Trident Networks for Object Detection[J]. ICCV 2019, 2019.

[24] Liu S, Qi L, Qin H, et al. Path Aggregation Network for Instance Segmentation[C/OL]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 8759-8768. DOI: 10.1109/CVPR.2018.00913.

[25] Tan M, Pang R, Le Q V. EfficientDet: Scalable and Efficient Object Detection[C/OL]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 10778-10787. DOI: 10.1109/CVPR42600.2020.01079.

[26] Zhang H, Wang K, Tian Y, et al. MFR-CNN: Incorporating Multi-Scale Features and Global Information for Traffic Object Detection[J/OL]. IEEE Transactions on Vehicular Technology, 2018, 67(9): 8019-8030. DOI: 10.1109/TVT.2018.2843394.

[27] Woo S, Hwang S, Kweon I S. StairNet: Top-Down Semantic Aggregation for Accurate One Shot Detection[J/OL]. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 2017: 1093-1102. https://api.semanticscholar.org/CorpusID:13681687.

[28] Zhao Q, Sheng T, Wang Y, et al. M2Det: A Single-Shot Object Detector based on Multi-Level Feature Pyramid Network[C]. The Thirty-Third AAAI Conference on Artificial Intelligence,AAAI. 2019.

[29] Liu Z, Gao G, Sun L, et al. IPG-Net: Image Pyramid Guidance Network for Small Object Detection[C/OL]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2020: 4422-4430. DOI: 10.1109/CVPRW50498.2020.00521.

[30] Gong Y, Yu X, Ding Y, et al. Effective Fusion Factor in FPN for Tiny Object Detection[C/OL]. 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). 2021: 1159-1167. DOI: 10.1109/WACV48630.2021.00120.

[31] Hong M, Li S, Yang Y, et al. SSPNet: Scale Selection Pyramid Network for Tiny Person Detection From UAV Images[J/OL]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5. DOI: 10.1109/LGRS.2021.3103069.

[32] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]. NIPS'14: Pro-

ceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2. Montreal, Canada: MIT Press, 2014: 2672-2680.

[33] Li J, Liang X, Wei Y, et al. Perceptual Generative Adversarial Networks for Small Object Detection [J/OL]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 1951-1959. https://api.semanticscholar.org/CorpusID:6704804.

[34] Bai Y, Zhang Y, Ding M, et al. SOD-MTGAN: Small Object Detection via Multi-Task Generative Adversarial Network[C]. Computer Vision – ECCV 2018. Cham: Springer International Publishing, 2018: 210-226.

[35] Pang Y, Cao J, Wang J, et al. JCS-Net: Joint Classification and Super-Resolution Network for Small-Scale Pedestrian Detection in Surveillance Images[J/OL]. IEEE Transactions on Information Forensics and Security, 2019, 14(12): 3322-3331. DOI: 10.1109/TIFS.2019.2916592.

[36] Kim J, Lee J K, Lee K M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks[C/OL]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 1646-1654. DOI: 10.1109/CVPR.2016.182.

[37] Cao J, Pang Y, Li X. Learning Multilayer Channel Features for Pedestrian Detection[J/OL]. IEEE Transactions on Image Processing, 2017, 26(7): 3210-3220. DOI: 10.1109/TIP.2017.2694224.

[38] Fu C Y, Liu W, Ranga A, et al. DSSD : Deconvolutional Single Shot Detector[EB/OL]. 2017. https://arxiv.org/abs/1701.06659. arXiv: 1701.06659 [cs.CV].

[39] Corsel C W, van Lier M, Kampmeijer L, et al. Exploiting Temporal Context for Tiny Object Detection[C/OL]. 2023 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW). 2023: 1-11. DOI: 10.1109/WACVW58289.2023.00013.

[40] Zhang S, Wen L, Bian X, et al. Single-Shot Refinement Neural Network for Object Detection [C/OL]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 4203-4212. DOI: 10.1109/CVPR.2018.00442.

[41] Yi K, Jian Z, Chen S, et al. Feature Selective Small Object Detection via Knowledge-based Recurrent Attentive Neural Network[EB/OL]. 2019. https://arxiv.org/abs/1803.05263. arXiv: 1803.05263 [cs.CV].

[42] Yang X, Yang J, Yan J, et al. SCRDet: Towards More Robust Detection for Small, Cluttered and Rotated Objects[C/OL]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 8231-8240. DOI: 10.1109/ICCV.2019.00832.

[43] Fu J, Sun X, Wang Z, et al. An Anchor-Free Method Based on Feature Balancing and Refinement Network for Multiscale Ship Detection in SAR Images[J/OL]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(2): 1331-1344. DOI: 10.1109/TGRS.2020.3005151.

[44] Lu X, Ji J, Xing Z, et al. Attention and Feature Fusion SSD for Remote Sensing Object Detection [J/OL]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-9. DOI: 10.1109/TIM.2021.3052575.

[45] Ran Q, Wang Q, Zhao B, et al. Lightweight Oriented Object Detection Using Multiscale Context and Enhanced Channel Attention in Remote Sensing Images[J/OL]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 5786-5795. DOI: 10.1109/JSTARS.2021.3079968.

[46] Li Y, Huang Q, Pei X, et al. Cross-Layer Attention Network for Small Object Detection in Remote Sensing Imagery[J/OL]. IEEE Journal of Selected Topics in Applied Earth Observations and

Remote Sensing, 2021, 14: 2148-2161. DOI: 10.1109/JSTARS.2020.3046482.

[47] Tian Z, Shen C, Chen H, et al. FCOS: A Simple and Strong Anchor-Free Object Detector[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(4): 1922-1933. DOI: 10.1109/TPAMI.2020.3032166.

[48] Yang F, Fan H, Chu P, et al. Clustered Object Detection in Aerial Images[C/OL]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 8310-8319. DOI: 10.1109/ICCV.2019.00840.

[49] Duan C, Wei Z, Zhang C, et al. Coarse-grained Density Map Guided Object Detection in Aerial Images[C/OL]. 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). 2021: 2789-2798. DOI: 10.1109/ICCVW54120.2021.00313.

[50] Li C, Yang T, Zhu S, et al. Density Map Guided Object Detection in Aerial Images[C/OL]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2020: 737-746. DOI: 10.1109/CVPRW50498.2020.00103.

[51] Wang Y, Yang Y, Zhao X. Object Detection Using Clustering Algorithm Adaptive Searching Regions in Aerial Images[C]. Computer Vision – ECCV 2020 Workshops. Springer International Publishing, 2020: 651-664.

[52] Deng S, Li S, Xie K, et al. A Global-Local Self-Adaptive Network for Drone-View Object Detection [J/OL]. IEEE Transactions on Image Processing, 2021, 30: 1556-1569. DOI: 10.1109/TIP.2020.3045636.

[53] Xu J, Li Y, Wang S. AdaZoom: Adaptive Zoom Network for Multi-Scale Object Detection in Large Scenes[EB/OL]. 2021. https://arxiv.org/abs/2106.10409. arXiv: 2106.10409 [cs.CV].

[54] Leng J, Mo M, Zhou Y, et al. Pareto Refocusing for Drone-View Object Detection[J/OL]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(3): 1320-1334. DOI: 10.1109/TCSVT.2022.3210207.

[55] Koyun O C, Keser R K, Akkaya İ B, et al. Focus-and-Detect: A small object detection framework for aerial images[J/OL]. Signal Processing: Image Communication, 2022, 104: 116675. DOI: https://doi.org/10.1016/j.image.2022.116675.

[56] Cui L, Lv P, Jiang X, et al. Context-Aware Block Net for Small Object Detection[J/OL]. IEEE Transactions on Cybernetics, 2022, 52(4): 2300-2313. DOI: 10.1109/TCYB.2020.3004636.

[57] Sun J, Gao H, Wang X, et al. Scale Enhancement Pyramid Network for Small Object Detection from UAV Images[J/OL]. Entropy, 2022, 24(11). https://www.mdpi.com/1099-4300/24/11/1699. DOI: 10.3390/e24111699.

[58] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters [C/OL]. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010: 2544-2550. DOI: 10.1109/CVPR.2010.5539960.

[59] Henriques J F, Caseiro R, Martins P, et al. High-Speed Tracking with Kernelized Correlation Filters [J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3): 583-596. DOI: 10.1109/TPAMI.2014.2345390.

[60] Danelljan M, Häger G, Khan F S, et al. Learning Spatially Regularized Correlation Filters for Visual Tracking[C/OL]. 2015 IEEE International Conference on Computer Vision (ICCV). 2015: 4310-4318. DOI: 10.1109/ICCV.2015.490.

[61] Valmadre J, Bertinetto L, Henriques J, et al. End-to-End Representation Learning for Correlation

Filter Based Tracking[C/OL]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 5000-5008. DOI: 10.1109/CVPR.2017.531.

[62] Bhat G, Danelljan M, Van Gool L, et al. Learning Discriminative Model Prediction for Tracking [C/OL]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 6181-6190. DOI: 10.1109/ICCV.2019.00628.

[63] Danelljan M, Van Gool L, Timofte R. Probabilistic Regression for Visual Tracking[C/OL]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 7181-7190. DOI: 10.1109/CVPR42600.2020.00721.

[64] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-Convolutional Siamese Networks for Object Tracking[C]. Computer Vision – ECCV 2016 Workshops. Cham: Springer International Publishing, 2016: 850-865.

[65] He A, Luo C, Tian X, et al. A Twofold Siamese Network for Real-Time Object Tracking[C/OL]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 4834-4843. DOI: 10.1109/CVPR.2018.00508.

[66] Li B, Yan J, Wu W, et al. High Performance Visual Tracking with Siamese Region Proposal Network [J/OL]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 8971-8980. https://api.semanticscholar.org/CorpusID:52255840.

[67] Zhu Z, Wang Q, Bo L, et al. Distractor-aware Siamese Networks for Visual Object Tracking[C]. European Conference on Computer Vision. 2018.

[68] Li B, Wu W, Wang Q, et al. SiamRPN++: Evolution of Siamese Visual Tracking With Very Deep Networks[C/OL]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 4277-4286. DOI: 10.1109/CVPR.2019.00441.

[69] Xu Y, Wang Z, Li Z, et al. SiamFC++: Towards Robust and Accurate Visual Tracking with Target Estimation Guidelines[J/OL]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(07): 12549-12556. https://ojs.aaai.org/index.php/AAAI/article/view/6944. DOI: 10.1609/aaai.v34i07.6944.

[70] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C/OL]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 770-778. DOI: 10.1109/CVPR.2016.90.

[71] Voigtlaender P, Luiten J, Torr P H, et al. Siam R-CNN: Visual Tracking by Re-Detection[C/OL]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 6577-6587. DOI: 10.1109/CVPR42600.2020.00661.

[72] Yu Y, Xiong Y, Huang W, et al. Deformable Siamese Attention Networks for Visual Object Tracking [C/OL]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 6727-6736. DOI: 10.1109/CVPR42600.2020.00676.

[73] Liu J, Wang H, Ma C, et al. SiamDMU: Siamese Dual Mask Update Network for Visual Object Tracking[J/OL]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2024, 8(2): 1656-1669. DOI: 10.1109/TETCI.2024.3353674.

[74] Chen X, Yan B, Zhu J, et al. Transformer Tracking[C]. CVPR. 2021.

[75] Wang N, Zhou W, Wang J, et al. Transformer Meets Tracker: Exploiting Temporal Context for Robust Visual Tracking[C/OL]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 1571-1580. DOI: 10.1109/CVPR46437.2021.00162.

[76] Yan B, Peng H, Fu J, et al. Learning Spatio-Temporal Transformer for Visual Tracking[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021: 10448-10457.

[77] Song Z, Yu J, Chen Y P P, et al. Transformer Tracking with Cyclic Shifting Window Attention [C/OL]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022: 8781-8790. DOI: 10.1109/CVPR52688.2022.00859.

[78] Liu Z, Lin Y, Cao Y, et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows[J/OL]. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 9992-10002. https://api.semanticscholar.org/CorpusID:232352874.

[79] Cui Y, Jiang C, Wang L, et al. MixFormer: End-to-End Tracking with Iterative Mixed Attention [C/OL]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022: 13598-13608. DOI: 10.1109/CVPR52688.2022.01324.

[80] Wu H, Xiao B, Codella N, et al. CvT: Introducing Convolutions to Vision Transformers[C/OL]. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 2021: 22-31. DOI: 10.1109/ICCV48922.2021.00009.

[81] Lin L, Fan H, Zhang Z, et al. SwinTrack: A Simple and Strong Baseline for Transformer Tracking [C/OL]. Advances in Neural Information Processing Systems: vol. 35. 2022: 16743-16754. https://proceedings.neurips.cc/paper_files/paper/2022/file/6a5c23219f401f3efd322579002dbb80-Paper-Conference.pdf.

[82] Ye B, Chang H, Ma B, et al. Joint Feature Learning and Relation Modeling for Tracking: A One-Stream Framework[C]. ECCV. 2022.

[83] Chen X, Peng H, Wang D, et al. SeqTrack: Sequence to Sequence Learning for Visual Object Tracking[C/OL]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023: 14572-14581. DOI: 10.1109/CVPR52729.2023.01400.

[84] Hong L, Yan S, Zhang R, et al. OneTracker: Unifying Visual Object Tracking with Foundation Models and Efficient Tuning[C/OL]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024: 19079-19091. DOI: 10.1109/CVPR52733.2024.01805.

[85] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]. Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part v 13. 2014: 740-755.

[86] Jocher G, Qiu J, Chaurasia A. Ultralytics YOLO[CP/OL]. 8.0.0. 2023. https://github.com/ultralytics/ultralytics.

[87] Cao Y, He Z, Wang L, et al. VisDrone-DET2021: The vision meets drone object detection challenge results[C]. Proceedings of the IEEE/CVF International conference on computer vision. 2021: 2847-2854.

[88] Lv W, Zhao Y, Chang Q, et al. Rt-detrv2: Improved baseline with bag-of-freebies for real-time detection transformer[J]. arXiv preprint arXiv:2407.17140, 2024.

[89] Li Y, Wu P, Zhang M. Rethinking the sparse mask learning mechanism in sparse convolution for object detection on drone images[J]. Computer Vision and Image Understanding, 2025: 104432.

[90] Leng J, Ye Y, Mo M, et al. Recent Advances for Aerial Object Detection: A Survey[J]. ACM Computing Surveys, 2024, 56(12): 1-36.

[91] Tan L, Liu Z, Liu H, et al. A Real-Time Unmanned Aerial Vehicle (UAV) Aerial Image Object Detection Model[C]. 2024 International Joint Conference on Neural Networks (IJCNN). 2024: 1-7.

[92] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]. European conference on computer vision. 2020: 213-229.

[93] Huang Y X, Liu H I, Shuai H H, et al. Dq-detr: Detr with dynamic query for tiny object detection [C]. European Conference on Computer Vision. 2024: 290-305.

[94] Du D, Qi Y, Yu H, et al. The unmanned aerial vehicle benchmark: Object detection and tracking [C]. Proceedings of the European conference on computer vision (ECCV). 2018: 370-386.

[95] Wang J, Yang W, Guo H, et al. Tiny Object Detection in Aerial Images[C/OL]. 2020 25th International Conference on Pattern Recognition (ICPR). 2021: 3791-3798. DOI: 10.1109/ICPR48806.2021.9413340.

[96] Xu X, Mao Z, Wang X, et al. Dynamic Anchor: Density Map Guided Small Object Detector for Tiny Persons[J]. Computer Vision and Image Understanding, 2025, 255: 104325.

[97] Li C, Yang T, Zhu S, et al. Density map guided object detection in aerial images[C]. proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 190-191.

[98] Du B, Huang Y, Chen J, et al. Adaptive sparse convolutional networks with global context enhancement for faster object detection on drone images[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 13435-13444.

[99] Akyon F C, Altinuc S O, Temizel A. Slicing aided hyper inference and fine-tuning for small object detection[C]. 2022 IEEE international conference on image processing (ICIP). 2022: 966-970.

[100] Zhu X, Su W, Lu L, et al. Deformable detr: Deformable transformers for end-to-end object detection [J]. arXiv preprint arXiv:2010.04159, 2020.

[101] Li F, Zhang H, Liu S, et al. Dn-detr: Accelerate detr training by introducing query denoising[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 13619-13627.

[102] Yao Z, Ai J, Li B, et al. Efficient detr: improving end-to-end object detector with dense prior[J]. arXiv preprint arXiv:2104.01318, 2021.

[103] Roh B, Shin J, Shin W, et al. Sparse detr: Efficient end-to-end object detection with learnable sparsity[J]. arXiv preprint arXiv:2111.14330, 2021.

[104] Zhao Y, Lv W, Xu S, et al. Detrs beat yolos on real-time object detection[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024: 16965-16974.

[105] Zhang H, Liu K, Gan Z, et al. UAV-DETR: Efficient End-to-End Object Detection for Unmanned Aerial Vehicle Imagery[J]. arXiv preprint arXiv:2501.01855, 2025.

[106] Xue H, Tang Z, Xia Y, et al. HCTD: A CNN-transformer hybrid for precise object detection in UAV aerial imagery[J]. Computer Vision and Image Understanding, 2025: 104409.

[107] Chen L, Fu Y, Gu L, et al. Frequency-aware feature fusion for dense image prediction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.

[108] Wang J, Chen K, Xu R, et al. CARAFE: Content-Aware ReAssembly of FEatures[C/OL]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 3007-3016. DOI: 10.1109/ICCV.2019.00310.

[109] Wang J, Xu C, Yang W, et al. A normalized Gaussian Wasserstein distance for tiny object detection [J]. arXiv preprint arXiv:2110.13389, 2021.

[110] Wang C Y, Yeh I H, Mark Liao H Y. Yolov9: Learning what you want to learn using programmable gradient information[C]. European conference on computer vision. 2024: 1-21.

[111] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.

[112] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]. Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[113] Zhu C, He Y, Savvides M. Feature selective anchor-free module for single-shot object detection [C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 840-849.

[114] Liu Z, Gao G, Sun L, et al. HRDNet: High-resolution detection network for small objects[C]. 2021 IEEE international conference on multimedia and expo (ICME). 2021: 1-6.

[115] Xu C, Wang J, Yang W, et al. Detecting tiny objects in aerial images: A normalized Wasserstein distance and a new benchmark[J/OL]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 190: 79-93. DOI: https://doi.org/10.1016/j.isprsjprs.2022.06.002.

[116] Guo G, Chen P, Yu X, et al. Save the Tiny, Save the All: Hierarchical Activation Network for Tiny Object Detection[J/OL]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34: 221-234. DOI: 10.1109/TCSVT.2023.3284161.

[117] Zhao M, Li W, Li L, et al. Single-frame infrared small-target detection: A survey[J]. IEEE Geoscience and Remote Sensing Magazine, 2022, 10(2): 87-119.

[118] Tong K, Wu Y. Deep learning-based detection from the perspective of small or tiny objects: A survey[J]. Image and Vision Computing, 2022, 123: 104471.

[119] Kou R, Wang C, Peng Z, et al. Infrared small target segmentation networks: A survey[J]. Pattern Recognition, 2023, 143: 109788.

[120] Liu C, Gao G, Huang Z, et al. YOLC: You Only Look Clusters for Tiny Object Detection in Aerial Images[J]. IEEE Transactions on Intelligent Transportation Systems, 2024.

[121] Tong K, Wu Y. Deep learning-based detection from the perspective of small or tiny objects: A survey[J]. Image and Vision Computing, 2022, 123: 104471.

[122] Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.

[123] Xu X, Feng Z, Cao C, et al. An Improved Swin Transformer-Based Model for Remote Sensing Object Detection and Instance Segmentation[J/OL]. Remote Sensing, 2021, 13(23). DOI: 10.3390/rs13234779.

[124] Xue J, He D, Liu M, et al. Dual Network Structure With Interweaved Global-Local Feature Hierarchy for Transformer-Based Object Detection in Remote Sensing Image[J/OL]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 6856-6866. DOI: 10.1109/JSTARS.2022.3198577.

[125] Suo J, Wang T, Zhang X, et al. HIT-UAV: A high-altitude infrared thermal dataset for Unmanned Aerial Vehicle-based object detection[J]. Scientific Data, 2023, 10(1): 227.

[126] Sun Y, Cao B, Zhu P, et al. Drone-Based RGB-Infrared Cross-Modality Vehicle Detection Via Uncertainty-Aware Learning[J/OL]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(10): 6700-6713. DOI: 10.1109/TCSVT.2022.3168279.

[127] Zhang G, Xu G, Chen S, et al. It’s Not the Target, It’s the Background: Rethinking Infrared Small-Target Detection via Deep Patch-Free Low-Rank Representations[J/OL]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 1-13. DOI: 10.1109/TGRS.2025.3608239.

[128] Dai Y, Wu Y, Zhou F, et al. Attentional local contrast networks for infrared small target detection [J]. IEEE transactions on geoscience and remote sensing, 2021, 59(11): 9813-9824.

[129] Min X, Zhou W, Hu R, et al. LWUAVDet: A Lightweight UAV Object Detection Network on Edge Devices[J]. IEEE Internet of Things Journal, 2024, 11(13): 24013-24023.

[130] Howard A G, Zhu M, Chen B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. ArXiv, 2017, abs/1704.04861.

[131] Zhang X, Zhou X, Lin M, et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices[C/OL]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 6848-6856. DOI: 10.1109/CVPR.2018.00716.

[132] Han K, Wang Y, Tian Q, et al. GhostNet: More Features From Cheap Operations[C/OL]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 1577-1586. DOI: 10.1109/CVPR42600.2020.00165.

[133] He K, Zhang X, 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.

[134] Zhang J, Lei J, Xie W, et al. SuperYOLO: Super resolution assisted object detection in multimodal remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-15.

[135] Xiong X, He M, Li T, et al. Adaptive Feature Fusion and Improved Attention Mechanism-Based Small Object Detection for UAV Target Tracking[J]. IEEE Internet of Things Journal, 2024, 11(12): 21239-21249.

[136] Guo C, Fan B, Zhang Q, et al. Augfpn: Improving multi-scale feature learning for object detection [C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 12595-12604.

[137] Fan H, Xiong B, Mangalam K, et al. Multiscale vision transformers[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2021: 6824-6835.

[138] Wang W, Xie E, Li X, et al. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2021: 568-578.

[139] Chen C F R, Fan Q, Panda R. Crossvit: Cross-attention multi-scale vision transformer for image classification[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2021: 357-366.

[140] Yu W, Si C, Zhou P, et al. Metaformer baselines for vision[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 46(2): 896-912.

[141] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1251-1258.

[142] Du Z, Hu Z, Zhao G, et al. Cross-Layer Feature Pyramid Transformer for Small Object Detection in Aerial Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 1-14.

[143] Yuan X, Cheng G, Yan K, et al. Small object detection via coarse-to-fine proposal generation and imitation learning[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2023: 6317-6327.

[144] Huang S, Lu Z, Cun X, et al. DEIM: DETR with Improved Matching for Fast Convergence[J/OL]. 2025: 15162-15171. DOI: 10.1109/CVPR52734.2025.01412.

[145] Huang S, Hou Y, Liu L, et al. Real-Time Object Detection Meets DINOv3[J]. arXiv, 2025.

[146] Peng Y, Li H, Wu P, et al. D-FINE: Redefine Regression Task in DETRs as Fine-grained Distribution Refinement[C]. The Thirteenth International Conference on Learning Representations. 2025.

[147] Kang M, Ting C M, Ting F F, et al. ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation[J]. Image and Vision Computing, 2024, 147: 105057.

[148] Yang G, Lei J, Tian H, et al. Asymptotic Feature Pyramid Network for Labeling Pixels and Regions [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(9): 7820-7829.

[149] Azad R, Niggemeier L, Hüttemann M, et al. Beyond self-attention: Deformable large kernel attention for medical image segmentation[C]. Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2024: 1287-1297.

[150] Rahman M M, Munir M, Marculescu R. Emcad: Efficient multi-scale convolutional attention decoding for medical image segmentation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 11769-11779.

[151] Chen L, Gu L, Zheng D, et al. Frequency-adaptive dilated convolution for semantic segmentation [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 3414-3425.

攻读学位期间取得的研究成果

I. 学术论文

[1] **Weihuang Chen**, Zhigang Yang, Lingyang Xue, Jinghai Duan, Hongbin Sun, Nanning Zheng. Multimodal pedestrian trajectory prediction using probabilistic proposal network[J]. IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), 2022. (SCI 1 区，IF：5.859，DOI：10.1109/TCSVT.2022.3229694)

[2] **Weihuang Chen**, Fang Zheng, Liushuai Shi, Yongdong Zhu, Hongbin Sun, Nanning Zheng. Multiple goals network for pedestrian trajectory prediction in autonomous driving[C]. IEEE International Conference on Intelligent Transportation Systems (ITSC), 2022:717–722.

[3] **Weihuang Chen**, Fangfang Wang, Hongbin Sun. S2tnet: Spatio-temporal transformer networks for trajectory prediction in autonomous driving[C]. Asian Conference on Machine Learning (ACML), 2021:454–469. (引用：10）

[4] **Weihuang Chen**, Yuwei Chen, Shen'ao Wang, Tianhang Li, Xuchong Zhang, Hongbin Sun. Motion planning using trajectory tree network for autonomous driving[J]. IEEE Transactions on Vehicular Technology (TVT), 2023, Under review. （投稿号：VT-2023-00733）

[5] Cheng Li, **Weihuang Chen**, Xinkai Luo, Fangfang Wang, Jingmin Zhang, Yanlong Yang, Hongbin Sun. Optimal preview distance control using model prediction for autonomous vehicle[C]. CAA International Conference on Vehicular Control and Intelligence (CVCI). 2021:1–8.

II. 专利

[6] 孙宏滨、**陈炜煌**、王玉学、章浩飞、李煊、吴彝丹，一种面向多场景的自动驾驶规划方法及系统 [P]，专利授权号：ZL202110276175.5

III. 科研获奖

[7] 第一届全国研究生智能挑战赛，三等奖，2019 年。（队长）

[8] 第六届中国研究生智慧城市技术与创意设计大赛，二等奖，2019 年。

[9] 第十二届中国智能车未来挑战赛，全国第 5 名，发现号自动驾驶平台，2020 年。（队长）

IV. 参与项目

[10] 国家重点研发计划项目（2018.05-2023.04）：“下一代深度学习理论、方法与关键技术”（项目编号：2017YFA0700800）

[11] 国家自然科学基金重大项目（2018.01-2022.12）：“极限工况下的人机协同机理及切换控制”（项目编号：61790563）

[12] 横向项目（2021.03-2021.09）：“基于深度学习的传感器数据融合”（项目编号：202103136）

答辩委员会会议决议

轨迹预测与规划是自动驾驶领域的重要研究问题。论文开展了基于深度神经网络的轨迹预测和运动规划方法研究，选题具有重要的研究与应用价值。主要创新点如下：

1. 提出了一种基于时空 Transformer 网络的单模态轨迹预测模型，提升了密集交通环境下不同类别交通参与者的轨迹预测能力。

2. 提出了一种基于概率性候选轨迹网络的多模态轨迹预测模型，提高了交通参与者的多模态轨迹预测速度和精度。

3. 提出了一种基于安全轨迹树网络的运动规划模型，提高了自动驾驶车辆的运动规划性能。

论文写作认真，结构清晰，论述清楚，工作量饱满，表明作者已掌握本学科宽广坚实的基础理论和系统深入的专业知识，独立从事科研工作的能力强，是一篇高质量的博士学位论文。

答辩中讲述清晰，回答问题正确，经答辩委员会讨论和无记名投票表决，一致同意通过学位论文答辩，并一致建议授予陈炜煌同学工学博士学位。

常规评阅人名单

本学位论文共接受 3 位专家评阅，其中常规评阅人 2 名，名单如下：

魏平　教授　　西安交通大学

邓成　教授　西安电子科技大学

学位论文独创性声明（1）

本人声明：所呈交的学位论文系在导师指导下本人独立完成的研究成果。文中依法引用他人的成果，均已做出明确标注或得到许可。论文内容未包含法律意义上已属于他人的任何形式的研究成果，也不包含本人已用于其他学位申请的论文或成果。

本人如违反上述声明，愿意承担以下责任和后果：

1. 交回学校授予的学位证书；
2. 学校可在相关媒体上对作者本人的行为进行通报；
3. 本人按照学校规定的方式，对因不当取得学位给学校造成的名誉损害，进行公开道歉。
4. 本人负责因论文成果不实产生的法律纠纷。

论文作者（签名）：　　日期：　　年　　月　　日

学位论文独创性声明（2）

本人声明：研究生　　所提交的本篇学位论文已经本人审阅，确系在本人指导下由该生独立完成的研究成果。

本人如违反上述声明，愿意承担以下责任和后果：

1. 学校可在相关媒体上对本人的失察行为进行通报；
2. 本人按照学校规定的方式，对因失察给学校造成的名誉损害，进行公开道歉。
3. 本人接受学校按照有关规定做出的任何处理。

指导教师（签名）：　　日期：　　年　　月　　日

学位论文知识产权权属声明

我们声明，我们提交的学位论文及相关的职务作品，知识产权归属学校。学校享有以任何方式发表、复制、公开阅览、借阅以及申请专利等权利。学位论文作者离校后，或学位论文导师因故离校后，发表或使用学位论文或与该论文直接相关的学术论文或成果时，署名单位仍然为西安交通大学。

论文作者（签名）：　　日期：　　年　　月　　日

指导教师（签名）：　　日期：　　年　　月　　日