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

摘 要 …… I
ABSTRACT …… III
1 智能光电系统实现与跟踪算法集成验证 …… 1
1.1 引言 …… 1
1.2 智能光电系统整体实现方案 …… 3
1.2.1 系统组成及工作原理 …… 3
1.2.2 边缘计算核心模组 …… 8
1.2.3 软件算法框架 …… 13
1.3 面向边缘计算设备的抗遮挡长时跟踪算法 …… 14
1.4 实验结果与分析 …… 14
1.4.1 基于 Nvidia Jetson 和 RK3588 边缘计算平台的算法部署与优化 …… 14
1.4.2 采用模块化设计的端侧全功能软件框架 …… 14
致谢 …… 15
参考文献 …… 16
攻读学位期间取得的研究成果 …… 26
答辩委员会会议决议 …… 27
常规评阅人名单 …… 28
声明

CONTENTS

ABSTRACT (Chinese) …… I
ABSTRACT (English) …… III
1 …… 1
1.1 Introduction …… 1
1.2 Overall Implementation Scheme of Intelligent Electro-Optical System …… 3
1.2.1 …… 3
1.2.2 …… 8
1.2.3 …… 13
1.3 …… 14
1.4 …… 14
1.4.1 Algorithm Deployment and Optimization on Nvidia Jetson and RK3588 Edge Computing Platform …… 14
1.4.2 Modular Design of End-to-end Full-function Software Framework …… 14
Acknowledgements …… 15
References …… 16
Achievements …… 26
Decision of Defense Committee …… 27
General Reviewers List …… 28
Declarations

1 智能光电系统实现与跟踪算法集成验证

本章将详细介绍基于软硬件协同设计理念的智能光电系统整体实现方案，具体涵盖多光谱传感器采集、基于 Nvidia Jetson 和 RK3588 边缘计算平台的算法部署与优化，以及采用模块化设计的端侧全功能软件框架。为了确保系统中各模块间高效、可靠的数据交换与控制，本章专门设计了配套的轻量级通信协议，定义了传感器、计算单元与上位机之间的标准化数据接口。同时，开发了功能完善的上位机软件，集成了实时状态监控、算法参数调整、结果可视化以及视频与日志记录功能，为系统调试、性能评估和人机交互提供了直观友好的操作界面。其次，针对动态复杂环境中目标易丢失的难题，本章提出了面向边缘计算设备的融合重识别机制与自适应模板更新的抗遮挡长时目标跟踪算法，并将其集成到智能光电系统中，通过实际场景测试验证了其鲁棒性。本章工作不仅验证了算法的工程可行性，也为构建自主智能光电系统提供了从底层硬件、核心算法到交互软件的完整原型参考与实践依据。

1.1 引言

随着低空经济的蓬勃发展和无人机任务复杂度的不断提升，机载光电系统正经历着从“被动观测”到“主动感知”、从“功能单一”到“多任务智能”的全面角色转变。在信息化、网络化、智能化融合发展的背景下，现实环境呈现出高度动态、高度复杂、信息饱和的特征，对环境感知的实时性、准确性、自主性提出了更严格的要求。这一系列需求驱动机载光电系统向智能化方向发展。智能化是指系统具备从原始数据中自主提取信息、在复杂不确定环境下进行稳定推理、并依据任务目标做出及时响应或决策的闭环能力，最终目标是使光电系统成为具备一定“认知”能力的空中智能体，能够在最小化人工干预的情况下，独立完成从搜索、发现、识别、跟踪到评估的完整环节。然而，现有系统距离实现上述理想的、完整的“智能化”，仍存在显著的差距，主要体现在硬件与软件两个层面。

在硬件层面，机载平台，尤其是中小型无人机，对载荷有着严格的尺寸、重量、功耗和成本限制，这导致机载信息处理单元算力与内存带宽有限，与云服务器或大型地面站相比，存在数量级差距。这种有限的计算资源与当前先进的视觉算法对算力的庞大需求构成了直接矛盾，尤其对于参数量庞大、计算密集的 Transformer 模型。许多在实验室标准数据集上表现优异的算法，因其复杂的计算度和巨大的延迟，难以直接部署到机载平台进行实时推理。因此，现有的“智能”往往是一种被严重约束的智能，需要在算法性能与实时性之间做出艰难平衡，导致许多先进的感知能力在端侧无法充分发挥。另外，更深层次的制约源于机载环境严苛的功耗与热管理约束。即便在机载平台中采用了具有高算力的边缘计算模块，其性能也往往受到物理规律的严峻挑战。机载平台的电能供给极为有限。无人机续航直接依赖于机载电池，为所有载荷（包括飞控、

传感器、通信与计算单元）分配固定的功率预算。一个在实验室中峰值功耗可达数十瓦的高性能计算模块，在机载环境下可能因超出功率预算而无法被允许持续运行在峰值状态。设计者通常被迫在其峰值性能和平均功耗之间做出权衡，通过软件或硬件手段设置功耗墙，这意味着硬件在其工作的大部分时间里都无法以其最大标称频率运行。其次，与功耗紧密相关的热管理问题在机载环境中尤为突出。计算芯片在高负载下产生的热量必须被有效散出，以防止芯片结温超过安全阈值导致降频、重启甚至永久损坏。然而，无人机紧凑的机体内通常缺乏空间部署大型主动散热系统（如风扇阵列或液冷回路），主要依赖有限表面积下的被动散热或风冷。在低空低速或悬停状态下，对流散热效率降低，极易导致热量积聚。因此，即使硬件模块在理论上具备高算力，在实际飞行中，其持续运算往往会迅速触发温度保护机制，迫使系统动态降频以控制温度，使得实际可持续的运算性能远低于标称峰值。这种由功耗与热约束导致的性能降级对于实际工程应用具有重大影响。许多在实验室标准数据集上、有充足散热保障的硬件平台上表现优异的复杂算法，一旦置于真实的机载环境，便会因计算单元的瞬时算力下降和频繁降频，而遭遇难以预测的性能波动和延迟激增。算法原本稳定的处理流水线被破坏，实时性指标急剧恶化。因此，现有的机载“智能”往往是一种被双重约束的智能：一是底层硬件资源的绝对不足，二是即使存在的硬件资源也因物理限制而无法全力输出。系统设计必须在算法理论性能、实时性要求、功耗上限与散热能力之间进行多维度的、艰难的平衡，这导致许多先进的感知模型在端侧无法充分发挥其潜力，甚至不得不为保障最基本的运行可靠性而牺牲精度与功能。

在软件与算法层面，现有系统更多地解决了“有无”问题，但与“智能”的本质要求尚有巨大差距。现有功能大多围绕单一、特定的任务（如“对指定类别的目标进行检测或跟踪”）进行优化。其算法流程往往是链式或并行的，检测、跟踪等模块相对独立，信息流动单向。算法模型通常在离线状态下用固定数据集训练完成，一旦部署，其参数和结构就已经固化。面对训练数据未充分涵盖的新环境、新目标类型或新型干扰，系统性能会急剧下降。在特定数据集上取得高精度的算法，在实际复杂的使用环境中，其性能边界往往模糊不清。对于采用深度神经网络的算法，当其在边缘设备部署时，必须借助边缘芯片厂商提供的专用工具链（如 NVIDIA 的 TensorRT、华为的 CANN、瑞芯微 RKNN 等）对模型进行量化、编译与优化，以使其能在特定的异构计算硬件上高效执行。然而，这个过程，往往以牺牲一定的算法精度为代价，导致在标准数据集上评测出的高精度无法无损地转化到实际应用中。从 FP32 到 FP16 或 INT8 的量化，使可表示的数值分辨率急剧下降。对于特征图中表征微弱信号或精细结构的小数值，量化后可能归零或产生较大相对误差，直接影响对小目标、低对比度目标或目标边缘的感知精度。这些工具链为了极致性能，会将多个基础算子（如卷积、批归一化、激活）融合为一个复合算子，并进行内存访问优化。然而，这种融合有时会改变运算的微小数值顺序或精度，可能与原始训练时定义的数学行为存在细微偏差，在网络中被逐层放大，最终导致误检、漏检或跟踪漂移。因此，一个神经网络经过量化部署到机载边缘平台后，其实际性能存在不可避免的损失，这种损失使得算法在实际复杂环境下的性能边界更

加模糊和脆弱。

本章的核心将从纯算法研究与理论分析，转向以实际部署与应用验证为导向的系统工程实践。具体而言，我们首先构建一套完整的软硬件协同设计架构，涵盖从多光谱传感器数据采集、到基于边缘计算设备的算法深度优化部署、再到模块化软件框架实现的全链路集成方案。进而，针对复杂动态场景中最具挑战性的目标持续跟踪问题，本章将提出并集成一种创新的抗遮挡长时跟踪算法，并通过实际场景测试，验证评估其在接近真实条件中的鲁棒性与实用性。最终，本章旨在打通从先进算法到可靠系统产品的关键路径，为智能光电系统的工程化与实战化提供一套经过验证的方法论与实例。

1.2 智能光电系统整体实现方案

1.2.1 系统组成及工作原理

机载光电系统主要由红外分系统、可见光分系统、激光分系统、稳定跟踪分系统、系统控制分系统、图像处理分系统、和结构分系统等组成，其工作原理如图1-1所示。从系统运行角度来说，各分系统是一个有机的整体，通过各分系统之间的协调工作实现相应的功能性能。从物理结构角度上说, 光电系统有的分系统与其他分系统综合为一体，如红外分系统、可见光分系统等光机部件交织在一起, 有的分系统是模块化。

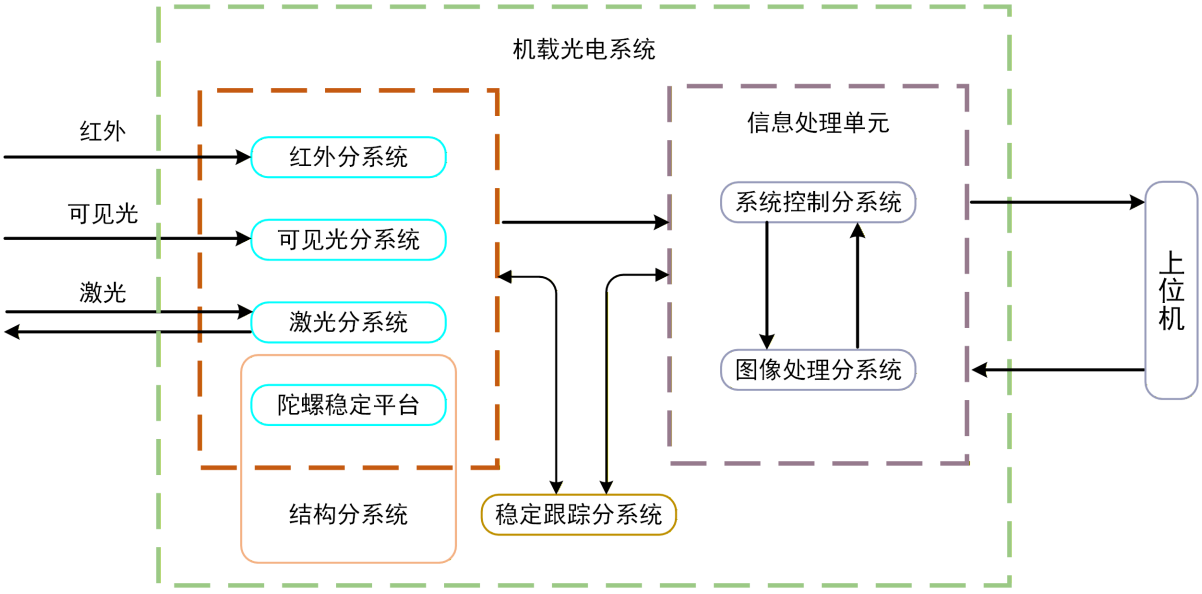


图 1-1 机载光电系统组成框图

1） 红外分系统

红外分系统是机载光电系统实现全天时全天候环境感知能力的主要传感器，由于它利用的是物体 (目标/背景) 自身发射的红外谱段的光线, 可以完全不依赖周边环境的光源，是短波、微光以及激光等不可替代的光学探测通道，通常是光电系统的必备传感器。红外分系统主要由红外光学镜头、支撑光学镜组的精密光机结构、红外探测器、

信息处理部件、视场切换机构、变焦机构等部件组成，如图1-2所示。红外分系统主要功能是对地面景物热辐射信息进行光电转换，输出视频信号给信息处理单元用于后续检测跟踪任务，同时为了适应不同环境下得到便于处理分析的图像，红外分系统具有电子变焦、黑白热切换、图像冻结、增益控制、偏置控制等功能。一般情况下，红外分系统在光电系统中是一个独立的组件，能够方便地进行更换，称为红外热像仪。有时红外光学光机部分与可见光光学、激光光学融合在一起形成一个十分复杂的光机构型，如共光路构型，瞄准吊舱多采用此种构型。

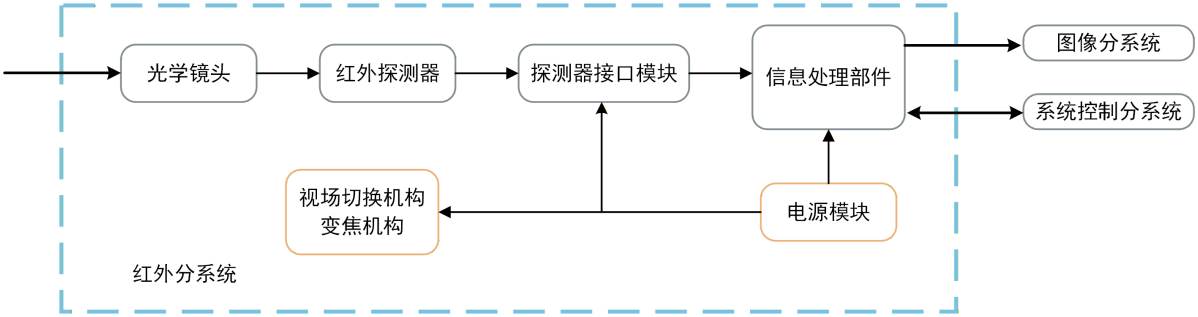


图 1-2 红外分系统组成

红外分系统在设计过程中主要考虑以下几个方面。

（1） 波段选择

红外系统的波长范围确定非常重要，同时工作波长也是系统总体指标之一。在红外分系统设计中，根据大气窗口可以选择中波 (3-5μm) 和长波 (8-12μm) 两个波段，具体波段的选择主要与目标和背景的辐射特性、大气传输特性和系统性能有关。设目标的辐射系数为 $\epsilon(\lambda, T)$, 目标的积分辐射出射度为:

$$M(\lambda_1 - \lambda_2) = \int_{\lambda_1}^{\lambda_2} \epsilon(\lambda, T) M_{\mathrm{eo}}(\lambda, T) \mathrm{d}\lambda \tag{1-1}$$

根据维恩位移定律

$$\lambda_m T = 2897.8 \mu m \cdot K \tag{1-2}$$

由于目标辐射峰值波长与温度成反比，从辐射特性的角度分析，对于目标特征为高温物体的红外系统选用中波波段，对于目标特征为常温物体的红外系统选用长波波段。但红外系统波段的选择还要考虑使用环境，由于空气湿度对长波影响较对中波影响严重，对于海上高温高湿的环境一般选用中波红外系统。从实际应用角度，由于近几年中波红外制冷型探测器灵敏度得到了很大程度的提高，工艺稳定，同时制冷型长波红外探测器较中波红外探测器存在成本高、稳定性较差等问题，因此目前国外研制的光电跟瞄系统采用中波红外的比例很高，在具体设计中红外分系统的波段选择要根据实际系统的使用情况综合考虑。

（2） 光学设计

红外光学镜头是保证系统光学性能的关键部件，在很大程度上影响着红外分系统的性能。在红外探测器确定后，系统总体设计中关于红外发现和识别距离的指标主要依靠光学设计保证。光学设计需要综合考虑视场大小、视场数量、焦距和传递函数等指标的要求，同时也要满足尺寸空间、重量、温度振动等环境条件的约束。

（3） 视场变换

红外分系统一般设计两个以上光学视场，宽视场用于广域搜索潜在目标，窄视场用于分辨识别目标的轮廓细节。视场切换的实现一般采用两种光学构型，第一种采用轴向移动透镜实现视场切换, 第二种采用切入/切出变倍镜组实现视场切换。

（4） 图像增强

由于红外图像动态范围较大，在将其转换为适合人眼观察的模拟图像过程中，容易造成图像细节的缺失，影响人眼的观察效果。红外探测器输出的原始数据一般为 16 位，而显示器显示的图像数据一般为 8 位。现代红外相机通常在内部集成了一系列基础的图像增强与校正功能，优化输出图像的视觉质量，其中以非均匀性校正和坏点补偿为基础，消除由探测器像元响应差异引起的固定模式噪声，确保图像均匀性，同时，对传感器缺陷造成的图像中的亮点或暗点进行检测和补偿，随后，通过基于查找表的动态范围压缩与对比度增强算法，将宽广的原始灰度范围进行智能映射，有选择地拉伸感兴趣温区的对比度以凸显细节，抑制非关键区域的灰度变化。同时，为优化主观视觉效果，相机常提供多种伪彩色编码方案，将灰度温度信息转换为彩色图像，以增强人眼对不同温度分布的辨识能力。此外，实时噪声抑制与电子稳像等辅助功能也常被集成，以进一步提升输出图像的清晰度与稳定性。这一系列嵌入式的处理步骤均在数据输出前完成，为后端分析系统提供实时、清晰、细节丰富的可视化图像。

2） 可见光分系统

可见光探测在机载光电系统中与红外探测一样承担着系统总体的性能指标要求，可见光成像系统获得的图像与人眼观察到的图像一样，非常有利于人的视觉系统识别目标，在光电系统中是必备的传感器。虽然红外探测白天也可以使用，但不能替代可见光探测，两者各有优势。可见光分系统有时是一个独立的组件，常被称为电视摄像机，有时同红外、激光和微光等综合设计在一起。

可见光分系统由光学镜头、调光机构、变倍机构、CCD 探测器和成像处理模块等组成，如图所示。主要功能是收集来自目标/景物的可见光信息，通过光电转换和信号处理，输出标准视频信号用于后续处理。变视场控制机构通过控制透镜组的移动完成视场的切换，调光控制机构通过电子快门和光圈使 CCD 实现在全照度范围的清晰成像。

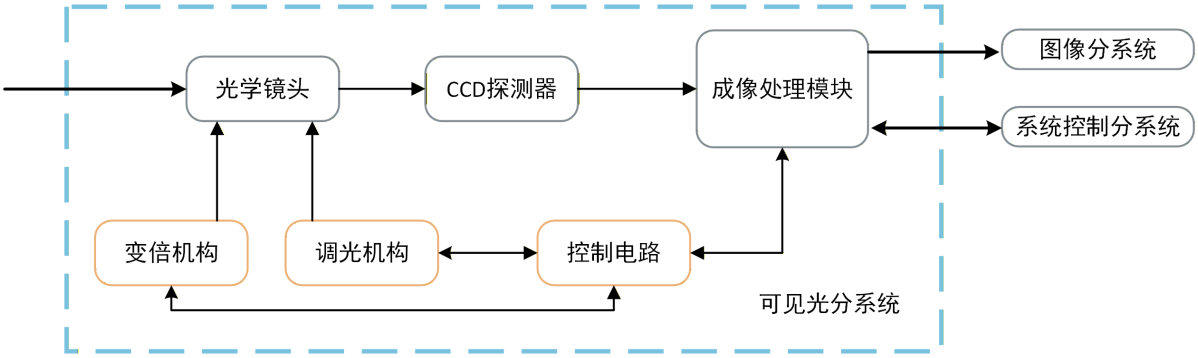


图 1-3 可见光分系统组成

可见光分系统光学镜组的设计与红外光学设计要考虑的问题大致相同，设计方法相似。可见光分系统有时采用单色成像，有时采用彩色成像。目前彩色成像 CCD 非常成熟，在光电系统中应用较为普遍。CCD 探测器由高感光度的半导体材料制成，包含有众多的感光元件，每个感光元件对应一个像素，通过像素敏感光线最终构成一幅完整的图像。随着技术的进步，CCD 探测器的分辨率得到了大幅度的提高，目前 2k 以上的产品已在机载光电系统中得到了应用，这将有效提高可见光分系统的空间分辨率和探测识别距离，也是可见光探测存在的一个必要和优势。

3） 系统控制与图像处理分系统

在传统的光电系统架构中，系统控制分系统与图像处理分系统在功能与硬件上通常是分离的。图像处理分系统作为数据处理的核心，主要负责接收来自可见光、红外等传感器的原始数据流，并执行从基础图像预处理到高级智能分析等一系列计算密集型任务。系统控制分系统负责与飞行控制单元、稳定平台、任务载荷管理器等外部模块进行实时通信，解析并执行来自上位机的任务指令，同时监控各个子系统的状态，并进行统一的调度与协同管理。这两个分系统之间通过高速总线进行数据与指令交互。

随着嵌入式硬件技术的发展，尤其是高性能片上系统与异构计算架构的成熟，上述分离式设计正被高度集成的方案所取代。当前的主流趋势是将这两个分系统的功能，深度融合到单一的边缘计算模组上。如图1-4所示，此类模组（如瑞芯微 RK3588、NVIDIA Jetson 系列）通常采用“CPU+GPU+NPU”的异构设计：CPU 核心负责运行轻量化的操作系统和系统控制分系统的所有逻辑，处理通信、调度与状态管理等任务，GPU 和专用的神经处理单元（NPU）则负责加速图像处理分系统中的深度学习算法，实现实时的目标识别与跟踪。FPGA 常被集成或作为协处理器，用于处理传感器数据接入、预处理等低延迟的流水线操作。

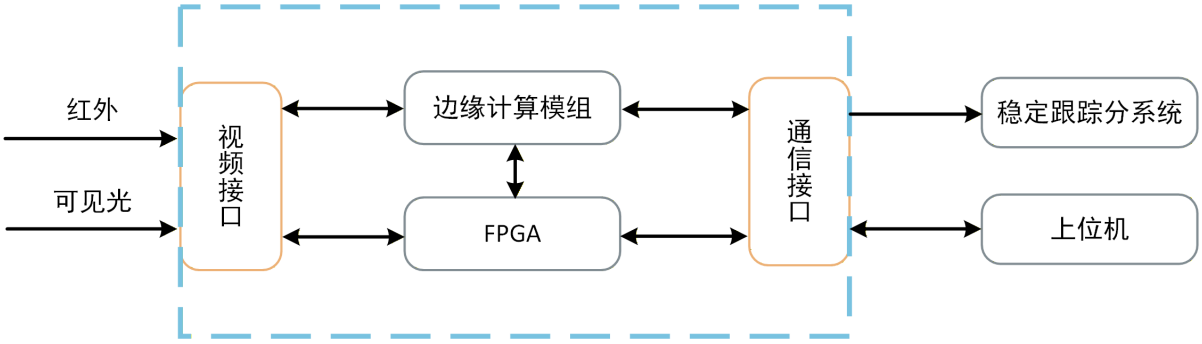


图 1-4 系统控制与图像处理分系统组成

这种硬件层面的集成通过消除分系统间的物理接口和总线传输，极大地降低了系统内部通信延迟，使得控制指令能够更快速地响应图像处理结果，另一方面，硬件资源的统一管理提升了整体效率，并大幅减少了系统的体积、重量和功耗，这对于空间和能源受限的机载平台至关重要。因此，现代智能光电系统的设计，已成为在单一高性能计算平台上，对实时数据流处理、外部交互指令响应及内部系统控制多任务协同优化的系统工程。

4） 稳定跟踪分系统

稳定跟踪分系统的主要功能是利用陀螺构成相对惯性空间稳定的陀螺稳定平台，隔离载机飞行中产生的姿态变化、振动等扰动，为光学传感器提供一个在惯性空间中高度稳定的物理指向基准，并在图像处理分系统的配合下，驱动稳定平台完成对目标的精确跟踪。对于大多数的机载光电系统，为了实现较高的精度要求，其稳定跟踪分系统的组成都是比较复杂的，但是从控制的角度来看，一般是一个或多个伺服控制系统的结合。其组成如图1-5所示。

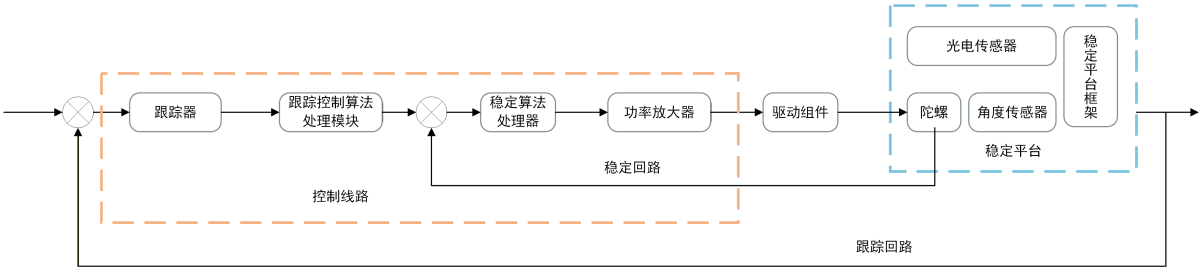


图 1-5 稳定跟踪分系统组成

稳定跟踪分系统是典型的高度集成的机电伺服单元，其硬件架构围绕稳定平台结构这一核心机械骨架构建。该平台通常采用两轴（方位、俯仰）或三轴（增加横滚）框架设计，由高刚度、轻量化的材料精密加工而成，直接承载并固定光电传感器载荷，为其提供物理安装基准和运动自由度。驱动组件负责执行控制指令，输出动力。惯性基准单元，通常为光纤或微机电陀螺，其核心功能是测量平台框架相对于惯性空间的角运动，建立稳定控制的绝对基准。控制线路集成了运行核心控制算法的处理器、指令输出

接口、电机功率放大器及其他状态控制电路，负责处理传感器信号、解算控制参数并驱动电机。角度传感器用于测量框架之间以及框架与载机基座之间的相对角位置。

稳定回路是整个分系统的基础闭环，其核心任务是隔离扰动，保持光电传感器瞄准线的稳定。当载机机动或受外界扰动时，扰动力矩会作用于平台框架，平台框架相对于此时陀螺的惯性基准产生了运动。陀螺测量到这一微小的角速度或位移。该信号经过滤波、解调与放大后，被送入数字控制器，解算出平台偏离预定惯性基准的偏差量，随后，控制器根据此偏差，计算出补偿控制指令。该指令通过功率驱动电路放大，驱动框架轴上的力矩电机产生补偿力矩。这个补偿力矩驱动平台框架朝相反方向运动，从而抵消了外部扰动的影响，使光电传感器的视轴保持动态稳定。

跟踪回路是构建于稳定回路之上的控制闭环，其核心功能是驱动已稳定的视轴，主动、持续地跟随场景中的运动目标。当目标与载机存在相对运动时，其在传感器成像平面上的位置会偏离图像中心。图像跟踪器（通常为图像处理分系统中的专用算法）对视频流进行实时处理，计算出目标相对于图像中心的二维像素偏差，即脱靶量。脱靶量被传输至跟踪控制器，控制器结合当前传感器的焦距、平台姿态等参数，将其转换为使目标重新回归图像中心所需的平台进动角速度或角位移指令。此指令作为新的期望输入，被叠加到稳定回路的控制环路中。在保持自身稳定性的同时，稳定平台在驱动电机的带动下开始执行受控的角运动，驱动视轴朝着目标偏移的方向运动。随着视轴的主动调整，目标在图像中的位置被逐步拉回中心。图像跟踪器持续检测并返回脱靶量，形成一个闭合的视觉伺服闭环。

1.2.2 边缘计算核心模组

随着无人机任务场景向实时化、智能化深度演进，传统依赖于地面站或机载工控机进行后端处理的架构，已无法满足目标实时识别、跟踪及快速自主决策的严苛需求。因此，将高性能算力前移至飞行平台本身，采用边缘计算模组作为机载智能光电系统的信息处理中枢，已成为业内明确的技术发展趋势。边缘计算模组的引入能够提升系统的即时反应的能力同时减轻通信链路带宽压力，提升系统的自主运行能力。这类模组通常采用“CPU+GPU+NPU”的异构计算架构，在有限的尺寸、重量和功耗约束下，兼顾通用计算、图形处理与人工智能加速，为复杂深度学习算法在机载端的实时部署提供了关键硬件基础。本节将选取并深入分析当前业界具有代表性的三款边缘计算核心模组：NVIDIA Jetson 系列、华为昇腾 310 及瑞芯微 RK 系列，从计算架构、AI 算力、能效比及开发生态等方面进行综合对比，为核心计算平台选型提供依据。

1） Nvidia Jetson 系列

Nvidia Jetson 系列是专为边缘人工智能和机器人应用设计的高性能计算平台，能够为嵌入式系统提供高性能、高能效的计算能力，广泛应用于机器人、无人机、智能相机及各类智能系统。Jetson 系列采用“CPU+GPU+NPU”异构计算架构，集成了 GPU 加

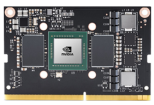
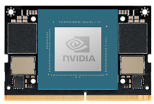
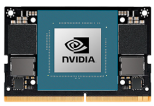
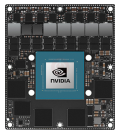
速核心、高性能 ARM CPU 和专用 AI 处理单元，在严格的功耗和尺寸限制下提供强大的算力。Jetson 模组与 JetPack SDK 配套使用，该套件为开发与部署提供全面的软件支持，通过本地实时推理，摆脱对云端服务器的依赖。

早期的 Jetson TX2，Nano 等产品聚焦于在低功耗下实现基本的深度学习推理，为算法原型验证和轻量级应用提供支撑。随后，Xavier 和 Xavier NX 系列通过引入更强大的 CPU-GPU 架构和专门的视觉处理器，支持多路高清视频流实时分析，显著提升了多传感器处理与中等复杂模型推理能力，满足了无人机、机器人对实时环境感知的需求。Jetson Orin 系列使边缘算力提升至数百 TOPS，基于 Ampere 架构的 GPU 与新一代深度学习的结合，使得在边缘设备上运行实时 SLAM、密集点云处理以及复杂目标检测模型等算法成为可能，成为高端无人机与机器人的主流选择。最新发布的 Jetson AGX Thor 是又一次升级，基于专为生成式 AI 设计的 Blackwell GPU 架构，面向下一代具身智能，支持视觉-语言-动作模型在边缘实时运行，将边缘 AI 的关注点从传统的“感知”提升至“感知-决策-控制”闭环，其 AI 算力高达 2070 FP4 TFLOPS，是前代 Orin 平台的 7.5 倍，而能效比也同步提升 3.5 倍，这一升级使得在机载端直接部署多模态大模型、进行复杂的场景推理与任务规划从理论走向工程实践。

Jetson 的核心优势不仅在于硬件，更在于其完整的软硬件生态系统。该生态系统以 JetPack SDK 为软件基石，它提供了一套全面的开发工具链，包含底层驱动程序、加速引擎（如用于 AI 推理的 TensorRT）、计算机视觉工具以及云原生技术支持，这使得应用的部署流程得以简化，并促进了软件在不同项目间的复用。此外，庞大的硬件合作伙伴网络提供了丰富的载板、传感器模块与工业级解决方案，极大地增强了平台在具体应用场景（如无人机光电系统）中的灵活性与集成便利性。由于这种从底层硬件到顶层应用的全栈支持，Jetson 平台已成为支撑边缘智能应用的核心硬件平台，为开发者和行业提供了可扩展的解决方案，为嵌入式系统的智能化发展提供支撑。

更为重要的是，Jetson 平台继承了 Nvidia 在人工智能计算领域的统一架构优势。当前主流的深度学习训练框架（如 PyTorch）其底层硬件加速普遍基于 Nvidia 的 CUDA 并行计算平台进行开发和优化。在服务器使用 Nvidia GPU 训练获得的模型，能够以最小的转换成本和性能损失，直接部署到同样基于 CUDA 和 TensorRT 的 Jetson 边缘平台进行推理。这种“训练-部署”架构的统一性，避免了为不同硬件平台进行繁琐的模型格式转换、重训练或显著的精度调优过程，确保了算法从实验室到嵌入式终端的一致性、高效性和可靠性。相较之下，其他厂商的边缘计算平台往往需要额外的模型转换工具链，在此过程中可能引入兼容性问题、算子不支持或难以量化的精度损失。

表 1-1 Nvidia Jetson 系列核心指标

名称	外观	年份	算力	功率	尺寸	售价
Jetson TX2		2017	1.3 TFLOPS	7.5-20W	70mm×45mm	149$
Jetson Xavier		2018	32 TOPS	10-40W	100mm×87mm	999$
Jetson Nano		2019	0.5 TFLOPS	5-10W	70mm×45mm	129$
Jetson Xavier NX		2020	21 TOPS	10-20W	70mm×4mm	479$
Jetson Orin Nano		2023	40 TOPS	7-15W	70mm×45mm	229$
Jetson Orin NX		2023	100 TOPS	10-25W	70mm×45mm	699$
Jetson AGX Orin		2023	275 TOPS	15-60W	100mm×87mm	999$
Jetson Thor		2025	2070 TOPS	40-130W	100mm×87mm	3199$

2） 华为昇腾 310

华为昇腾（Ascend）310 是华为专为边缘人工智能设计的高性能计算平台，其目标是在严格的功耗和尺寸限制下，为嵌入式设备提供强大的专用 AI 算力，并构建自主可控的软硬件生态。它集成了 NPU、CPU 和图像处理单元，其核心是自研的达芬奇（Da Vinci）架构 3D Cube 矩阵计算单元，该架构采用可扩展的立方体阵列设计，针对 CNN 等深度学习算法的卷积、矩阵运算进行了硬件级优化，能够高效执行 INT8、FP16 等精度运算，每时钟周期可进行 4096 次乘加（Multiply Accumulate）运算，实现了单位功耗下极高的有效算力输出, 能够提供最多 22 TOPS 的 INT8 算力，为机载智能光电系统等对尺寸、功耗和可靠性有严苛要求的领域，提供了一个高度集成化、软硬协同优化的国产化选择。昇腾 310 作为华为全栈自主 AI 技术体系中的关键边缘计算单元，与面向数

据中心的昇腾 910 训练芯片协同，通过统一的 CANN 异构计算平台，构建了从模型开发、训练到边缘推理一体化的国产化技术闭环。

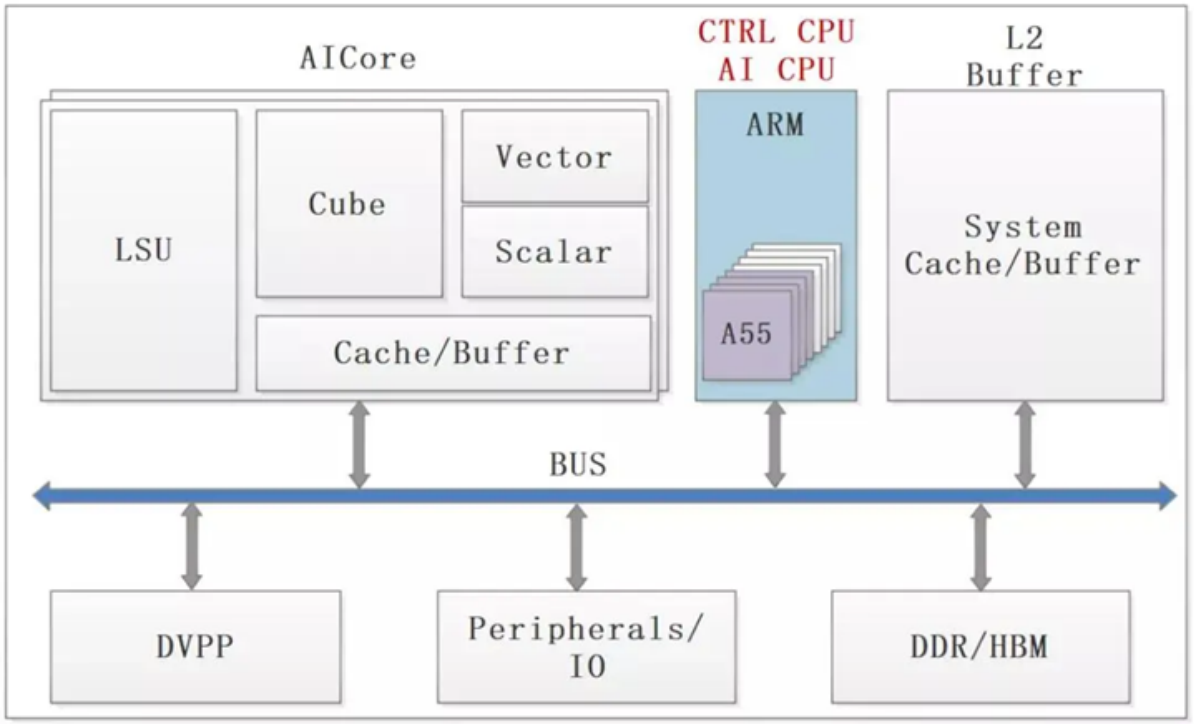


图 1-6 昇腾 310 芯片逻辑结构图

昇腾 310 处理器的核心创新在于达芬奇（Da Vinci）3D Cube 架构，与通用图形处理器（GPU）基于流式多处理器（SM）的 SIMD（单指令多数据）架构不同，达芬奇架构将海量计算单元组织成一个三维立方体矩阵。这种设计使得单条指令能够驱动整个立方体矩阵执行一次完整的乘加运算，从而在硬件层面原生且高效地支持卷积、全连接等神经网络核心操作的张量计算。在执行 ResNet、YOLO 系列等主流视觉模型的推理任务时，能够实现远高于通用 GPU 的硬件利用率和计算密度，为边缘设备提供了显著的单位功耗性能（TOPS/W）。如图1-6所示，昇腾 310 将自研的神经网络处理单元（NPU）、多核 ARM CPU、数字视觉预处理单元（DVPP）、视频编解码器以及内存控制器集成于单一芯片，形成一个完整的 SoC。这种集成化设计减少了外部交互带来的功耗和延迟，DVPP 单元可直接对输入的图像进行缩放、裁剪、格式转换等预处理，将数据以最佳格式传输给 NPU，整个过程在片上完成，最大限度地减少了芯片内外部的数据搬运与格式转换开销，在提升处理速度的同时显著降低了系统功耗与延迟。

为充分发挥其专用硬件潜力，华为构建了贯穿软件栈各层的 CANN（Compute Architecture for Neural Networks）异构计算架构，形成了从底层驱动到上层框架的全栈自主技术体系。CANN 位于操作系统与 AI 框架之间，其核心作用在于充当硬件指令集与主流深度学习模型之间的“编译器”与“优化器”。它通过图编译器（Ascend Graph Compiler）和张量加速引擎（Tensor Boost Engine）等技术，将来自 PyTorch、TensorFlow 等框架的模型进行深度的图融合、算子优化、内存分配以及量化，最终生成高度优化的

离线模型（OM），以实现对计算资源的充分利用。面向应用层，华为提供了 AscendCL（Ascend Computing Language）编程接口与 MindStudio 集成开发环境，通过定义统一的软件接口（API），将底层硬件的具体实现细节封装起来，为开发者提供统一、便捷的 AI 开发与部署工具。必须承认的是，华为昇腾的全栈自主生态在第三方库适配广度、跨平台模型迁移的便捷性以及全球开发者社区活跃度方面，相较于已耕耘数十年的 Nvidia CUDA 生态系统仍存在差距。然而，其战略价值在于构建了一条从芯片、驱动、算子库到框架的完整的自主可控技术链，这种垂直整合模式保障了在关键基础设施与国防安全等敏感领域的供应链安全与技术主权。华为昇腾通过提供一套高性能、自主可控的软硬一体化替代方案，为有明确国产化要求的应用场景提供了坚实的技术基础。

3） 瑞芯微 RK 系列

瑞芯微 RK 系列 SoC 以视频多媒体、图形渲染与接口集成为特长，近年来逐步融入专用 AI 加速与高带宽互连，广泛应用于多媒体终端、工业控制、边缘人工智能与车载信息娱乐等领域。瑞芯微早期在平板电脑、OTT 电视盒及商业显示领域的成功，使其在高清视频编解码、图像信号处理（ISP）及 2D/3D 图形渲染方面积累了行业领先的 IP 核与技术。这一优势被完整继承至后续的 AIoT 芯片中，使得 RK 系列在需要处理多路高清视频流的视觉应用场景中具有天然优势，能够提供较强的多媒体并发能力与丰富 I/O（PCIe、SATA、MIPI、以太网），并通过 RKNN 工具链实现主流深度学习模型的部署。

瑞芯微自 RK3399Pro 开始引入独立的 NPU 模块，并在后续的 RK3568 及旗舰平台 RK3588 中持续升级。其 NPU 算力从早期的 1-3 TOPS 提升至最高 6 TOPS (INT8)，能够流畅运行经优化的 YOLO、MobileNet 等轻量化检测与分类模型，满足大部分基础视觉 AI 任务需求。

RK 系列区别于其他 AI 芯片的显著特征，是其强大的视频处理能力，内置的多核 VPU 支持当今最先进的视频编解码标准，并能在编解码过程中以极低功耗运行。同时，高性能 ISP 支持多路摄像头同时接入，可进行 HDR、3A（自动对焦、曝光、白平衡）等实时图像增强处理，直接输出画质优异的图像供 AI 分析或编码传输。这对于依赖高质量成像的视觉光电系统至关重要。NPU 采用自主研制的架构，支持主流深度学习框架模型转换与部署，其绝对算力无法与 Jetson 系列或昇腾 310 相比，但在成本控制方面具有优异，适合对功耗和预算有严格限制的大规模边缘计算部署场景。

如图1-7所示，旗舰 RK3588 采用“NPU+CPU+GPU”的设计，独立的 6 TOPS 算力 NPU 负责 AI 推理，4 个大核（A76）和 4 个小核（A55）组成的异构 CPU 处理控制与通信任务，Mali-G610 GPU 负责图形渲染与部分并行计算。提供 8 路 MIPI-CSI 摄像头接口、HDMI Tx/Rx、双千兆以太网、SATA 3.0、PCIe 3.0 等接口，这种设计使得单颗 RK3588 即可作为系统的核心枢纽，非常适合空间和载重受限的无人机平台。

RK 系列的软件基于标准 Linux 与 Android，系统层与标准开源社区高度同步，开发

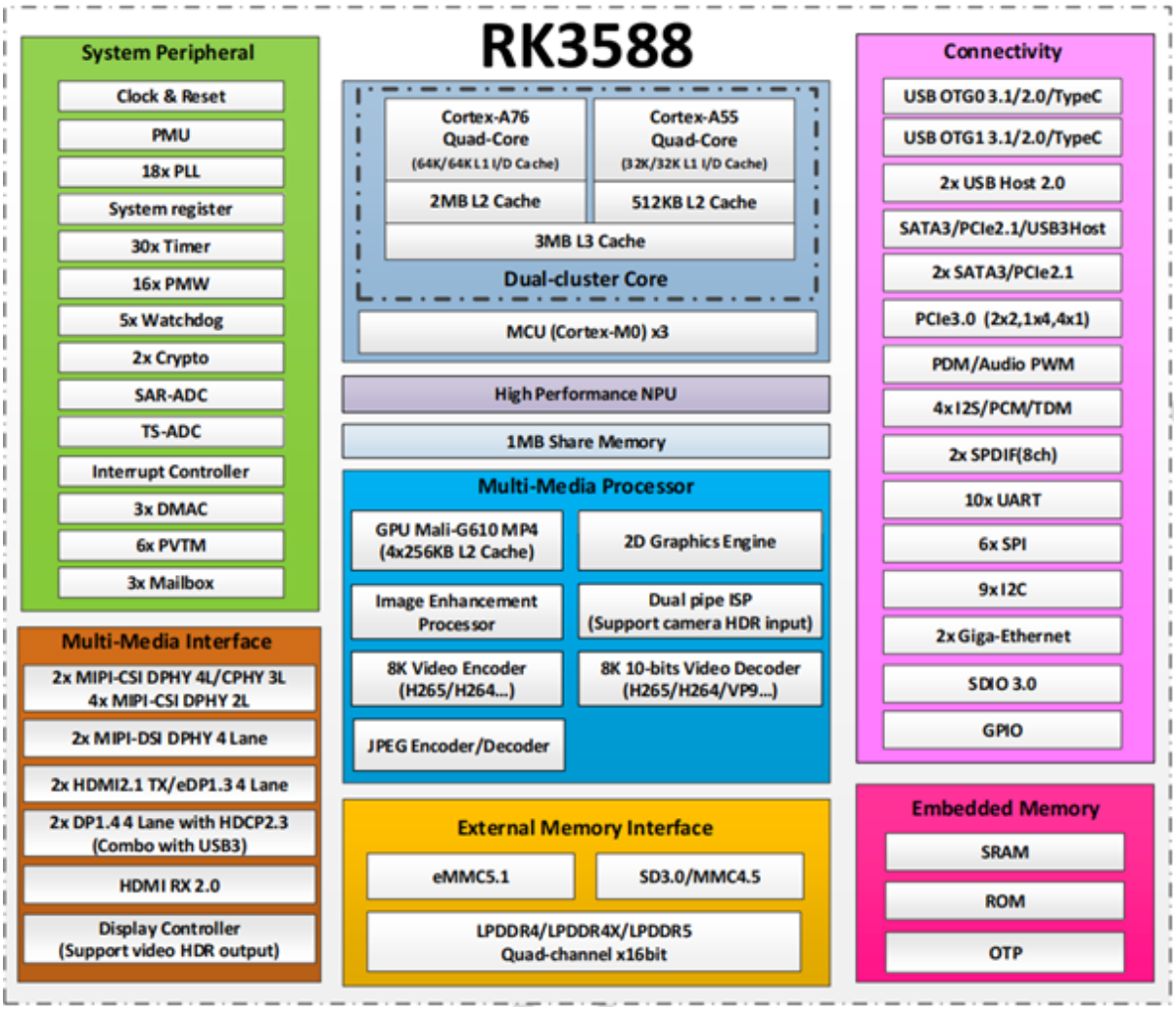


图 1-7 RK3588 组成及外部接口

者可以利用现成的开源库、驱动和中间件，快速构建上层应用程序。瑞芯微提供 RKNN-Toolkit 模型转换与部署工具链，支持将 TensorFlow、PyTorch、ONNX 等格式的模型转换为可在其 NPU 上运行的 RKNN 模型。该工具链包含了量化、性能分析和内存优化等功能，虽然其深度和自动化程度可能不及 Nvidia TensorRT，但对于主流的轻量化模型部署已足够使用。

1.2.3 软件算法框架

本文设计并实现了一套完整的面向机载光电系统的嵌入式智能处理软件框架，不同于将孤立的目标检测或跟踪算法直接部署于嵌入式系统，本文设计的框架提供了面向任务的模块化软件栈。通过引入成熟的软件设计模式，在架构层面强制实现了各功能模块的高内聚与低耦合，将传感器数据采集、智能算法处理、实时通信控制以及人机交互等复杂功能，封装为职责清晰且接口标准的独立模块，再通过统一的服务总线进行协同整合。这种设计不仅确保了整个软件系统的稳定性、高效性与可维护性，更使其

具备了根据实时任务需求，动态重构功能模块的灵活性与适应能力，确保了系统在资源受限的嵌入式平台上高效运行。

1.3 面向边缘计算设备的抗遮挡长时跟踪算法

1.4 实验结果与分析

1.4.1 基于 Nvidia Jetson 和 RK3588 边缘计算平台的算法部署与优化

基于 Nvidia Jetson 和 RK3588 边缘计算平台的算法部署与优化是智能光电系统的核心，包括目标检测算法和目标跟踪算法等。其中，目标检测算法主要用于检测目标，目标跟踪算法主要用于跟踪目标。

1.4.2 采用模块化设计的端侧全功能软件框架

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