本科毕业设计（论文）开题报告

课题名称：面向无人机的 **3D** 目标检测算法

学 员 姓 名：赖宇　　学　　号：202102001020

首次任职专业：无　　学历教育专业：人工智能与大数据

命 题 学 院：计算机学院　　年　　级：2021 级

指 导 教 员：邓明堂　　职　　称：副研究员

所 属 单 位：国防科大计算机学院学员一大队学员五队

国防科技大学教育训练部制

主要内容：

一、课题名称、来源、选题依据；

二、本课题国内外研究现状及发展趋势；

三、课题在理论与实践上的意义；

四、课题需要解决的关键理论问题和实际问题；

五、课题研究的基本方法、实验方案及技术路线的可行性论证；

六、开展研究应具备的条件及已具备的条件，并估计在进行论文工作中可能遇到的困难与问题和解决措施；

七、论文研究的进展计划；

八、课题所需器材、设备清单。

编排打印要求：

（1）采用 A4（21cm×29.7cm）白色复印纸，单面黑字打印。上下左右各侧的页边距均为 3cm；缺省文档网格：字号为小 4 号，中文为宋体，英文和阿拉伯数字为 Times New Roman，每页 30 行，每行 36 字；页脚距边界为 2.5cm，页码置于页脚、居中，采用小 5 号阿拉伯数字从 1 开始连续编排，封面不编页码。

（2）报告正文最多可设四级标题，字体均为黑体，第一级标题字号为 4 号，其余各级标题为小 4 号；标题序号第一级用“一、”、“二、”……，第二级用“（一）”、“（二）”……，第三级用“1.”、“2.”……，第四级用“（1）”、“（2）”……，分别按序连续编排。

（3）正文插图、表格中的文字字号均为 5 号。

一、 面向无人机的 3D 目标检测算法

无人机（Unmanned Aerial Vehicles）是指无人驾驶的飞行机器。近年来，由于其卓越的机动性，无人机被广泛部署在交通监控、精准农业、灾害管理和野生动物监控等领域。与传统路侧端监的控摄像头相比，无人机提供更高的效率和适应性，对推进计算机视觉（Computer Vision）的众多应用至关重要。在这些应用中，鲁棒的对象检测和跟踪对于无人机的有效部署至关重要。然而，现有的无人机应用模型主要是针对传统的 2D 感知任务设计的，例如 LAM-YOLO[1] 和 Drone-TOOD[2]，这限制了需要对环境进行 3D 理解的实际应用的发展。在基于视觉的机器人系统中，3D 感知扮演着重要角色，使其能够处理 2D 感知无法胜任的复杂任务。虽然对于无人机来说，3D 视觉仍然是相对较新的技术，但它提供了在 3D 环境中捕获对象的完整维度数据的能力。

随着自动驾驶技术的发展，研究人员开始关注 3D 视觉下多角度的目标检测技术，出现了 DERE3D[3]、PETR[4]、PETRv2[5]、BEVFormer[6] 等车载 3D 多视角目标检测模型。然而，目前尚未出现专注于无人机的 3D 多视角目标检测模型。同时，针对无人机 3D 多视角目标检测任务的数据集 UAV3D 日前发布，经过测试，先前的 3D 多视角目标检测模型在该数据集上表现均不理想，可能的原因在于车载模型的泛化性问题，这说明当前的研究结果并不适用于无人机领域。因此，有必要研究无人机的 3D 多视角目标检测模型，以提升无人机对于 3D 环境的理解能力。

二、 国内外研究现状及发展趋势

本项目主要研究面向无人机的 3D 目标检测算法。我们将介绍无人机、目标检测算法、3D 目标检测等几个方面的相关工作。

（一） 无人机与深度学习

无人机因价格便宜、使用方便、对人员安全以及操作人员培训简单而越来越受欢迎。[7] 这些优势，加上其的分辨率和强大的跟踪特性，促使它们在各种环境中的使用越来越多。无人机已被用于环境监测，包括空气污染、地表温度、洪水危险、森林火灾、道路表面损坏、地形监测、行人交通监测和灾害疏散。[8] 例如，许多人因为可以通过移动设备控制的先进产品而提高了生活水平。汽车技术通过提供关于交通的最新和精确信息来帮助驾驶员。无人机有多种规格、尺寸和配置。它们被归类为四大类别：固定翼、混合固定翼、单旋翼和多旋翼，同时考虑旋翼的数量。固定翼无人机适合于航空测量和绘图，因为它们稳定且续航时间长。混合固定翼无人机结合了自动化和手动滑翔，提供了可操作性和效率之间的平衡。单旋翼无人机虽然更复杂且成本更高，但为特定任务（如详细的地形测量）提供了卓越的精确度。最后，多旋翼无人机，尤其是四旋翼无人机，因其敏捷性、垂直起降能力和常用于

监控和航空摄影应用而受到高度重视。多旋翼无人机可以是三旋翼、四旋翼、六旋翼或八旋翼。[9]

得益于各种技术和方法的发展，人工智能（Artificial Intelligence）近几十年来经历了显著的进步。机器学习（Machine Learning）技术因其能够通过与环境的互动提供决策自主性而脱颖而出。机器学习是不断发展的数据科学领域的重要组成部分。通过使用统计方法，模型被训练以在各种基于数据挖掘的项目中进行分类或预测，从而发现关键信息。在这一演变的背景下，深度学习（Deep Learning）作为机器学习中的先进技术而出现，它由具有多层的人工神经网络组成。这些深层结构使我们能够处理复杂任务，如识别图像中的模式或理解自然语言。存在多种 ML 算法，根据一些基本标准对它们进行预先分类是有益的。其中一个最重要和最显著的标准与算法的训练方式有关。在这种分类中，区分了四种主要方法：监督学习、无监督学习以及强化学习。[9]

计算机硬件和软件的最新发展使得人工智能成为几乎所有与工程相关的研究领域的关键组成部分。人工智能是解决那些没有明确答案或传统方法需要大量人为干预的难题的强大工具。人工智能与传统认知算法之间的一个显著区别在于，人工智能可以自动提取特征。这取代了昂贵的手工特征工程。无人机的许多问题，如高功率/能源消耗和实时需求，也是边缘计算和边缘人工智能的优点，如低能耗和低延迟。深度学习和无人机（Unmaned Aerial Vehicles）有潜力彻底改变交通部门的交通监控。[10]

（二） 基于 RGB 图像的 3D 目标检测算法

2D 目标检测一定程度上促进了 3D 目标检测的发展。3D 目标检测方法可以分为单视角和多视角两个方面。[11]

1. 基于单视角的 3D 目标检测方法

这些 3D 目标检测算法在思想上与 2D 目标检测算法最为接近，仅以单目/立体图像作为输入来预测 3D 目标实例。[11] 一般分为三种：基于模板匹配的方法、基于几何属性的方法和基于伪激光雷达的方法。

（1） 基于模板匹配的方法

这种方法的主要思想是通过穷尽采样执行 2D/3D 匹配并将 3D 候选区与作为代表模板进行评分。典型例子就是早期由 Chen 提出的 3DOP[12]，其对输入的立体图像估计深度，并将图像平面上的像素级坐标投影到三维空间来估计点云。3DOP 将候选区生成的问题定义为 Markov 随机场（MRF）的能量最小化问题，该问题需要对势函数进行精心设计。获取 3D 目标的候选框后，3DOP 使用 FastR-CNN[13] 方案来回归预测目标位置。 在可能的情况下，汽车仅配备一个摄像头，Chen 随后提

出了 Mono3D[14] 来通过单目相机达到 PAR 性能。与 3DOP 不同的是，Mono3D 不计算深度信息，而是滑动窗口从 3D 空间采样 3D 候选区。当对对象所在地平面与图像平面正交时，可以减少搜索的工作量。3DOP 或 Mono3D 输出制定类别的候选区，这需要为每个类别单独设计方案。然而过度依赖工程和专业领域知识使得模型在复杂场景下的通用性比较差。

（2） 基于几何特性的方法

这种方法并不需要大量的候选区来提高召回率，而是直接从精确的 2D 边界框开始，直接从经验观察中获得几何特性来粗略的估计 3D 姿态。

Mousavian 等人提出的 Deeo3DBox[15] 利用几何特性，让 3D 角的透视投影至少紧密贴在 2D 边界框的一侧。Li 等人提出了 GS3D[16]，仅使用单目 RGB 图像来检测完整的 3D 实例，而并没有引入额外的数据。GS3D 基于 Faster RCNN[17] 的框架，额外添加了一个名为 2D+O 的子网络，用于预测 2D 边界框和观测的方向。接着 GS3D 获得粗略的 3D 框，称为引导。最后，在被送 3D 子网络进一步细化之前，从 2D 框和 3D 框上选择 3 个可见面进行融合来解决表征模糊的问题。尽管 GS3D 有了显著的性能提升，超过了现有的基于单目图像的方法，但其依赖于经验知识，而并不能保证是准确的，且很容易受到对象范围和大小的影响。

（3） 基于伪激光雷达的方法

这些方法首先进行深度估计，然后求助于现有的基于点云的方法。

Xu 等人提出了 MF3D[18],该算法对图像特征和伪激光雷达进行多层次的融合。具体来说，MF3D 首先通过独立的单目深度估计模块计算视差，以获得伪激光雷达。同时，采用标准的 2D 候选区域生成网络，将 RGB 图像与由视差图获得的转换前视图特征融合作为输入。随着 2D 候选区域的获得，RGB 图像和伪激光雷达的特征通过串联融合，以进一步细化。最近，Weng 提出了 Mono3D-PLiDAR[19],通过单目深度估计将输入图像升级成 3D 相机坐标，即伪激光雷达点(例如,DORN[20])。然后是 3D 目标检测模型 Frustum PointNets[21] 应用于伪激光雷达。Weng 等人揭示了伪激光雷达由于单目深度估计的误差，会产生大量的噪声。体现在两个方面：激光雷达点的局部不对准和深度伪影的问题。为了克服前者，Mono3D-PLiDAR 使用 2D-3D 边界框一致性损失（BBCL)来有监督的训练。为了缓解后者问题，Mono3D-PLiDAR 采用 Mask R-CNN[22] 预测的实例掩码代替 2D 边界框来减少截锥内不相关的点。

2. **基于多视角的 3D 目标检测方法。**

这些方法首先将多幅图像转换成前视图或鸟瞰图(BEV)展示，在网格中密集以利用 CNN 和标准 2D 检测方案。

Wang 等人提出了 DETR3D[3]。该算法是 DETR 在三维目标检测领域的延伸，它通过几何反投影和相机变换矩阵将二维特征提取与三维目标预测联系起来，实现了无需密集深度估计的三维目标检测。DETR3D 将多视图检测问题转化为集合到集合的预测任务，通过预设的 object queries 和神经网络解码出三维空间中的参考点，再将这些点反投影到二维特征图上，通过双线性插值采样特征值，最终通过多头注意力机制和 Transformer[23] 解码器来优化 queries 并预测边界框和类别。然而 DETR3D 存在预测参考点不准确、无法从全局角度进行表示学习等缺点，为此，Liu 和 Wang 等人提出了 PETR[4]。PETR 通过引入 3D 坐标生成器和 3D 位置编码器，将三维坐标的位置信息编码为图像特征，从而实现多视图三维目标检测。PETR 首先在三维空间中初始化一组均匀分布的 anchor points，然后通过 MLP 网络生成初始对象查询。与 DETR3D 不同，PETR 先预设三维坐标再编码到 query，这样做避免了在图像平面找不到对应点的问题，并实现了在三维特征空间中的训练。之后 Liu 等人提出了 PETRv2[5]，PETRv2 在 PETR 的基础上增加了时间建模，通过将前一帧的特征与当前帧的特征融合来捕获时间线索，简化了速度预测，并实现了不同帧目标位置的时间对齐。PETRv2 通过特征引导的位置编码器将图像特征和三维位置信息结合，隐式引入了视觉先验，提高了模型的性能。

Huang 等人提出的 BEVDet[24] 是另一种高性能的多相机三维目标检测方法，它在鸟瞰图（BEV）空间中进行目标检测。BEVDet 面临过拟合问题，因为它在 BEV 空间下过度拟合。为了解决这个问题，BEVDet 应用了定制的数据增强策略和尺度 NMS（Scale-NMS），以提高模型的泛化能力和检测性能。 之后 Huang 等人提出了 BEVDet4D[25]，BEVDet4D 通过将前一帧的特征与当前帧中的相应特征融合来捕获时间线索，简化了速度预测，并实现了不同帧目标位置的时间对齐。Li 等人提出的 BEVFormer[6] 利用可变形注意力机制设计了空间交叉注意力和时间自注意力，分别从跨摄像机视图的感兴趣区域提取空间特征和循环融合历史 BEV 信息，从而实现对三维场景的理解和目标检测。

三、 项目需要解决的关键理论问题和实际问题

（一） 如何设计适用无人机 3D 目标检测领域的模型

由于目前尚无专门针对无人机 3D 目标检测领域的模型，本项目在前期主要参考车载 3D 目标检测模型 DETR3D[3]，设计端到端的目标检测模型。模型先使用 CNN 对输入图像提取特征，并使用 Transformer[23] 的编码器（Encoder）进行处理，并使用可以学习的特定查询序列（Query Tokens）作为解码器（Decoder）的输入，将解码的结果使用两个全连接网络（Full Connected Network）进行处理得到最终的结果。

这一端到端框架的优点是减少了非极大值抑制（NMS）等后期处理，极大地提升了运算速度。

（二） 如何针对无人机目标检测领域的特点来细化改进模型以提升模型精度

相较于车载3D目标检测，无人机3D目标检测的UAV3D[26]数据集存在视角移动频繁、目标密度大、目标模糊、目标小等特点。针对该特点，可以设计专门的检测头或引入上下文信息（如目标周围的背景信息），提升小目标的检测精度。还可以在CNN部分使用FPN（多尺度特征金字塔）以捕捉不同尺度的目标特征。除此外，通过替换损失函数也能引导模型更加关注小目标，提升对小目标的检测精度。

（三） 如何在提升模型精度的同时降低计算复杂度

无人机平台的算力有限，在保证精度的同时降低模型的计算复杂度有着重要的研究价值。StreamPETR研究发现，现有模型如BEV时序模型使用BEV特征导致其对于运动物体检测精度较低，而DETR类型的模型需要与多帧图像进行计算来获取时序相关信息，导致计算量翻倍。设计特征来高效地提取特征并避免多帧图像运算是提升模型精度并降低计算复杂度的关键。

四、 项目研究的基本方法、实验方案及技术路线的可行性分析

（一） 项目研究的基本方法和实验方案

1. 研究方法：

（1）文献阅读：阅读3D目标检测的相关文献，掌握3D目标检测的基本原理和常用方法；阅读无人机感知的相关文献，掌握基于无人机目标检测的最新方法和最新进展。

（2） 代码实现：实验复现现有的3D目标检测模型。在此基础上，设计适用UAV3D数据集的无人机3D目标检测模型，确定并实现本项目要使用的模型与方法。

（3）结果分析：采用实验法和定量分析法对本项目提出的模型方法进行验证，采用具体的量化指标对模型进行评估。

（4） 论文撰写：将本项目得到的结果进行归纳总结，并对研究问题、采用的模型和结果分析进行总结记录。

2. 实施方案：

本项目主要分为两个部分研究面向无人机的3D目标检测算法。第一部分主要是基于UAV3D数据集[26]的进行3D目标检测模型的初步设计，并训练验证不同模

型架构的可行性；第二部分是基于第一部分的最优模型架构进行进一步优化创新，并将模型尝试应用于无人机目标定位等领域。

（1） 基于 UAD3D 数据集的模型训练

UAV3D 数据集是专为无人机 (UAV) 平台的 3D 感知任务而设计[26]，包含 1,000 个场景（700 个训练场景、150 个验证场景和 150 个测试场景）、50 万张 RGB 图像和 330 万个 3D 框。本项目基于 UAV3D 数据集训练设计的模型架构，并使用 mAP、NDS、mATE、mASE、mAOE 等指标测试，选择综合性能最佳的模型架构。

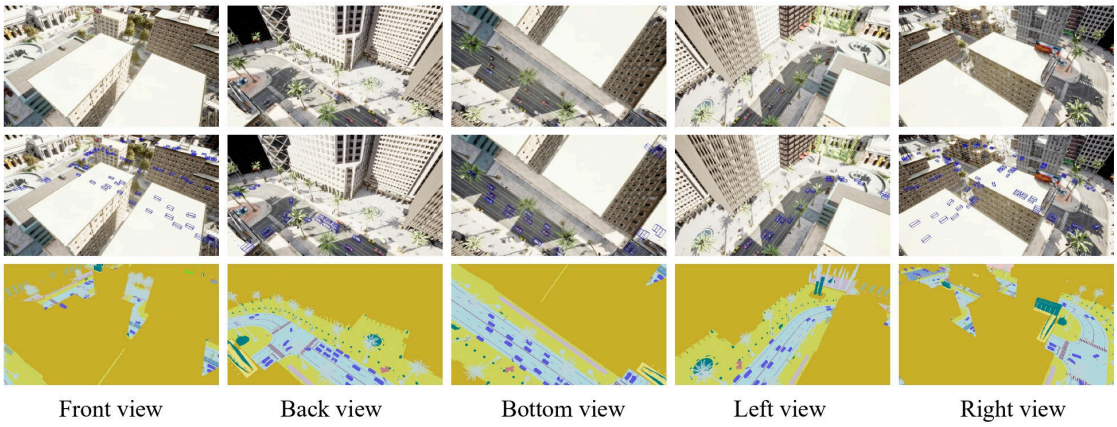

图 1　UAD3D 示意图

（2） 模型改进与应用

基于第一部分的模型架构的缺点进行针对性修改，使用消融实验等方法逐步验证每个改进模块的效果，确定其对模型性能的贡献。评估模块是否可以独立发挥作用，或者需要与其他模块协同工作。通过实验找到最优的模块组合，简化模型结构并提升性能。最后尝试将模型应用于无人机目标定位等领域。

（二） 技术路线的可行性分析

项目开展初期对研究现状进行了调研，总结了当前的研究的难点。同时此前也积累了有关目标检测的理论知识与编程经验。已完成了大量相关文献的阅读工作，在实验过程中遇到的相关问题也有一定的解决基础。总的来说，本项目总体研究方向清晰、研究前景广阔，且具有足够的理论、技术和工具积累，因此是可行的。

五、 研究条件

（一） 开展研究应具备的条件

软件条件： Python、Pytorch

硬件条件： 服务器、显卡计算资源

知识储备：熟悉 3D 目标识别的研究现状，熟悉 Python 编程，Pytorch 深度学习框架，熟悉 Linux 操作系统

（二） 开展研究已经具备的条件

已具备软硬件条件和相应知识储备。

（三） 可能遇到的困难

1. 3D 多视角检测模型的设计以及实现

2. 算法需要多次迭代修改才能有效

（四） 解决措施

及时查阅文献，广泛搜集资料，积极与导师，同学交流。

六、 工作计划

起讫日期	主要完成研究内容	预期成果
2024 年 11 月-2025 年 1 月	阅读相关论文文献，准备开题	了解相关背景，确定项目方向
2025 年 1 月-2025 年 3 月	分析数据特点，构建模型的输入方式	完成基础的模型构建
2025 年 3 月-2025 年 4 月	完成基准实验，并设计实验方案	得到基准实验数据，完成实验方案设计方案
2025 年 4 月-2025 年 5 月	完成面向无人机 3D 目标检测的实验和结果分析	完成最终模型构建并得到最终实验数据
2025 年 5 月-2025 年 7 月	整理研究成果，撰写毕业论文，准备答辩	完成毕业论文

七、 器材设备清单

Intel(R) Xeon(R) Gold 6226R CPU @ 2.90GHz 16 核 * 2

NVIDA RTX 3090 * 8

八、 参考文献

[1] ZHENG Y, JING Y, ZHAO J, 等. LAM-YOLO: Drones-based Small Object Detection on Lighting-Occlusion Attention Mechanism YOLO[J]2024:

[2] OU K, DONG C, LIU X, 等. Drone-TOOD: A Lightweight Task-Aligned Object Detection Algorithm for Vehicle Detection in UAV Images[J]IEEE Access, 2024: 41999-42016

[3] WANG Y, GUIZILINI V, ZHANG T, 等. DETR3D: 3D Object Detection from Multi-view Images via 3D-to-2D Queries[J]2021:

[4] LIU Y, WANG T, ZHANG X, 等. PETR: Position Embedding Transformation for Multi-view 3D Object Detection[C]//European Conference on Computer Vision2022

[5] LIU Y, YAN J, JIA F, 等. PETRv2: A Unified Framework for 3D Perception from Multi-Camera Images[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV)2023

[6] LI Z, WANG W, LI H, 等. BEVFormer: Learning Bird's-Eye-View Representation from Multi-camera Images via Spatiotemporal Transformers[C]//European Conference on Computer Vision2022

[7] MANFREDA S, MCCABE M F, MILLER P E, 等. On the Use of Unmanned Aerial Systems for Environmental Monitoring.[J]Remote Sensing, 2018(4): 641

[8] BARBEDO J G A. A Review on the Use of Unmanned Aerial Vehicles and Imaging Sensors for Monitoring and Assessing Plant Stresses[J]Drones, 2019(2): 40

[9] CABALLERO-MARTIN D, LOPEZ-GUEDE J M, ESTEVEZ J, 等. Artificial Intelligence Applied to Drone Control: A State of the Art[J]Drones, 2024(7): 296

[10] PAL O K, SHOVON M S H, SHIN M F M & …Jungpil. In-depth review of AI-enabled unmanned aerial vehicles: trends, vision, and challenges[J]Discover Artificial Intelligence, 2024(1): 1-24

[11] QIAN R, LAI X, LI X. 3D Object Detection for Autonomous Driving: A Survey[J]Pattern Recognition, 2022: 108796

[12] CHEN X, KUNDU K, ZHU Y, 等. 3D object proposals for accurate object class detection[C]// NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 12015

[13] GIRSHICK R. Fast R-CNN(Conference Paper)[J]Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448

[14] CHEN X XZ (Chen, KUNDU K K (Kundu, ZHANG Z ZY (Zhang, 等. Monocular 3D Object Detection for Autonomous Driving[J]2016 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2016: 2147-2156

[15] MOUSAVIAN A, ANGUELOV D, FLYNN J, 等. 3D bounding box estimation using deep learning and geometry[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)2017

[16] LI B, OUYANG W, SHENG L, 等. GS3D: An Efficient 3D Object Detection Framework for Autonomous Driving[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)2019

[17] REN S SQ (Ren, HE K KM (He, GIRSHICK R R (Girshick, 等. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks(Article)[J]IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017(6): 1137-1149

[18] XU B B (Xu, CHEN Z ZZ (Chen, IEEE. Multi-level Fusion Based 3D Object Detection from Monocular Images(Conference Paper)[J]Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018: 2345-2353

[19] WENG X, KITANI K. Monocular 3D Object Detection with Pseudo-LiDAR Point Cloud[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)2019

[20] FU H, GONG M, WANG C, 等. Deep Ordinal Regression Network for Monocular Depth Estimation.[J]Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, 2018: 2002-2011

[21] QI C R, LIU W, WU C, 等. Frustum PointNets for 3D Object Detection from RGB-D Data[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition2018

[22] HE K k, GKIOXARI G g, DOLLAR P p, 等. Mask R-CNN.[J]IEEE Transactions on Pattern Analysis & Machine Intelligence, 2020(2): 386-397

[23] VASWANI A, SHAZEER N, PARMAR N, 等. Attention Is All You Need[J]Learning, 2017:

[24] HUANG J, HUANG G, ZHU Z, 等. BEVDet: High-performance Multi-camera 3D Object Detection in Bird-Eye-View[J]2022:

[25] HUANG J, HUANG G. BEVDet4D: Exploit Temporal Cues in Multi-camera 3D Object Detection[J]2022:

[26] YE H, SUNDERRAMAN R, JI S. UAV3D: A Large-scale 3D Perception Benchmark for Unmanned Aerial Vehicles[J]2024:

指导教员审核意见：

较好明确了毕业设计的计划和目标，研究项目具有较好的现实意义，制定了可行的实施方案，正确评估了研究过程中可能遇到的问题和相应的解决方案，文献调研充分，同意开题。

签名：邓明堂　　2025 年 1 月 14 日

教研室（研究室、实验室）意见：

同意

领导签名：刘英子　　2025 年 1 月 14 日

系（研究所、重点实验室）意见：

同意

领导签名：邓明堂　　2025 年 1 月 14 日

学院教学科研处（教务处）意见：

同意开题

（公章）

2025 年 1 月 14 日

注：开题报告由学员撰写，答辩结束交指导教员。