Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for
Dynamic Systems and Control
IDSC
Institut für Dynamische Systeme
und Regelungstechnik

Lucas M. Schneider

Vision-Based Autonomous Drone Navigation in Dynamic Urban Environments

Semester Thesis

Institute for Dynamic Systems and Control
Swiss Federal Institute of Technology (ETH) Zurich

Supervision

Prof. Dr. Davide Scaramuzza
Dr. Michael R. Walter

July 2025

IDSC-25-047

Preface

The rapid advancement of autonomous aerial systems has transformed numerous industries, including environmental monitoring, precision agriculture, infrastructure inspection, search and rescue, and autonomous delivery. As these applications continue to expand, the need for intelligent navigation systems capable of operating safely and efficiently in complex environments has become increasingly important. Recent developments in computer vision and artificial intelligence have enabled unmanned aerial vehicles (UAVs) to perceive their surroundings with remarkable accuracy, reducing their dependence on expensive external positioning systems.

This thesis was motivated by the growing demand for robust vision-based navigation techniques that enable autonomous drones to operate reliably in dynamic urban environments. The research explores how modern deep learning algorithms, visual perception, and real-time decision-making can be integrated to improve navigation performance while maintaining computational efficiency and operational safety.

The work presented in this dissertation combines concepts from computer vision, robotics, artificial intelligence, and autonomous systems to investigate practical solutions for visual localization, obstacle avoidance, and path planning. Throughout the research, particular emphasis has been placed on developing scalable approaches that can be deployed on resource-constrained aerial platforms operating in real-world conditions.

It is hoped that the findings presented in this dissertation contribute to the ongoing development of intelligent autonomous systems and provide useful insights for future research in robotics, computer vision, and aerial navigation technologies.

Contents

Zusammenfassung

Diese Arbeit untersucht den Einsatz moderner Verfahren des Computer Vision und des Deep Learning fÃ¼r die autonome Navigation unbemannter Luftfahrzeuge in dynamischen stÃ¤dtischen Umgebungen. Ziel der Forschung ist die Entwicklung eines robusten Navigationssystems, das visuelle Wahrnehmung, Hinderniserkennung und autonome Entscheidungsfindung kombiniert, um eine sichere und zuverlÃ¤ssige Navigation ohne permanente AbhÃ¤ngigkeit von externen Positionierungssystemen zu ermÃ¶glichen.
Das vorgeschlagene System integriert neuronale Netze mit Verfahren der visuellen Lokalisierung, Bahnplanung und Hindernisvermeidung. Die experimentellen Ergebnisse zeigen, dass der entwickelte Ansatz eine hohe Navigationsgenauigkeit, stabile Echtzeitleistung und eine verbesserte Robustheit gegenÃ¼ber komplexen Umgebungsbedingungen erreicht.
Die Ergebnisse dieser Arbeit leisten einen Beitrag zur Weiterentwicklung intelligenter autonomer Drohnensysteme und bieten eine Grundlage fÃ¼r zukÃ¼nftige Forschung in den Bereichen Robotik, Computer Vision und autonome Luftfahrzeuge.

Chapter 1

Introduction

Autonomous Unmanned Aerial Vehicles (UAVs), commonly referred to as drones, have become an integral component of modern intelligent systems. Their ability to perform complex tasks with minimal human intervention has enabled widespread adoption across numerous application domains, including infrastructure inspection, environmental monitoring, precision agriculture, disaster response, surveillance, mapping, and logistics. As advances in robotics, artificial intelligence, and embedded computing continue to accelerate, autonomous drones are increasingly expected to operate safely within complex and dynamic urban environments.

Traditional autonomous navigation systems rely heavily on Global Navigation Satellite Systems (GNSS), Inertial Measurement Units (IMUs), and LiDAR sensors to estimate position and avoid obstacles. While these technologies have demonstrated high levels of accuracy in structured environments, they often experience performance degradation in densely populated urban areas where signal occlusion, multipath interference, and rapidly changing surroundings present significant operational challenges. Consequently, researchers have increasingly focused on vision-based navigation systems capable of enabling drones to perceive and understand their surroundings using onboard cameras.

Recent advances in computer vision and deep learning have significantly improved the capabilities of autonomous aerial systems. Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), and modern object detection architectures have demonstrated exceptional performance in visual perception, scene understanding, semantic segmentation, and obstacle detection. These developments have created new opportunities for integrating visual intelligence into autonomous navigation systems capable of operating reliably in real-world environments.

Despite these technological advancements, vision-based autonomous navigation remains a challenging research problem. Urban environments contain dynamic obstacles such as pedestrians, vehicles, cyclists, construction equipment, and varying weather conditions that require rapid perception and decision-making. Furthermore, autonomous drones must balance navigation accuracy, computational efficiency, and energy consumption while operating on resource-constrained embedded hardware.

This thesis investigates a vision-based autonomous navigation framework designed for dynamic urban environments. The proposed approach integrates deep learning-based visual perception, simultaneous localization and mapping (SLAM), obstacle detection, and real-time path planning to enable robust autonomous flight. The framework combines modern computer vision techniques with intelligent decision-making algorithms to improve navigation accuracy while maintaining computational efficiency suitable for deployment on embedded aerial platforms.

Experimental evaluation is conducted using simulated and publicly available benchmark datasets representing diverse urban scenarios. The proposed system is assessed using quantitative performance metrics including localization accuracy, obstacle avoidance success rate, trajectory efficiency, computational latency, and navigation robustness under varying environmental conditions.

The primary objective of this research is to develop a scalable and reliable navigation framework capable of supporting autonomous drone operations in challenging urban environments. By com-

bining visual perception with intelligent planning and control algorithms, this work contributes to the development of safer, more efficient, and more adaptable autonomous aerial systems.
The remainder of this thesis is organized as follows. Chapter 2 reviews the existing literature on autonomous drone navigation, computer vision, deep learning, and visual localization techniques. Chapter 3 presents the proposed research methodology and system architecture. Chapter 4 discusses the experimental evaluation and performance analysis of the proposed framework. Finally, Chapter 5 summarizes the principal contributions of the research and outlines opportunities for future work.

1.1 Research Motivation

Autonomous drones are expected to play an increasingly important role in smart cities and industrial automation. However, reliable navigation in dynamic urban environments remains a major technical challenge due to continuously changing obstacles, limited positioning accuracy, and computational constraints. This research is motivated by the need for intelligent vision-based navigation systems capable of operating safely and efficiently under real-world conditions without excessive dependence on external positioning infrastructure.

1.2 Research Objectives

The primary objectives of this research are:

- Investigate modern computer vision techniques for autonomous drone navigation.
- Develop a vision-based navigation framework for dynamic urban environments.
- Integrate deep learning with visual localization and obstacle avoidance algorithms.
- Evaluate the proposed framework using quantitative navigation performance metrics.
- Identify current limitations and future research opportunities in autonomous aerial navigation.

1.3 Scope of the Research

This research focuses on vision-based autonomous navigation for quadrotor UAVs operating in urban environments. The study investigates camera-based perception, deep learning-based object detection, visual localization, and intelligent path planning using publicly available datasets and simulation environments. Hardware implementation and swarm-based drone coordination are considered beyond the scope of this dissertation and are identified as potential directions for future research.

Chapter 2

Literature Review

Autonomous aerial systems have experienced significant technological advancements over the past decade, driven by rapid developments in robotics, computer vision, artificial intelligence, and embedded computing. Vision-based navigation has emerged as one of the most active research areas because it enables unmanned aerial vehicles (UAVs) to operate with reduced dependence on external positioning systems while maintaining accurate localization and obstacle avoidance capabilities. This chapter reviews the existing literature related to autonomous drone navigation, visual perception, deep learning, simultaneous localization and mapping (SLAM), path planning, and intelligent flight control.

2.1 Autonomous Unmanned Aerial Vehicles

Unmanned Aerial Vehicles (UAVs) have become increasingly important across numerous civilian and industrial applications, including infrastructure inspection, precision agriculture, environmental monitoring, disaster management, surveillance, and package delivery. Their ability to access hazardous or difficult-to-reach environments makes them highly valuable for tasks that would otherwise require significant human effort or involve considerable operational risks.

Modern UAVs typically integrate multiple sensing modalities including cameras, inertial measurement units (IMUs), Global Navigation Satellite Systems (GNSS), LiDAR sensors, ultrasonic sensors, and radar systems. These sensors collectively enable autonomous flight, environmental perception, obstacle detection, and trajectory planning. However, reliance on external positioning systems such as GNSS may become problematic in dense urban environments where signal degradation and multipath interference frequently occur.

2.2 Computer Vision for Autonomous Navigation

Computer vision enables autonomous drones to perceive and interpret their surroundings using visual information obtained from onboard cameras. Unlike traditional navigation systems that depend primarily on external infrastructure, vision-based approaches estimate environmental structure directly from captured images.

Recent advances in image processing and deep learning have significantly improved visual perception capabilities. Tasks including object detection, semantic segmentation, depth estimation, optical flow analysis, and visual odometry can now be performed with remarkable accuracy and computational efficiency, enabling drones to operate in increasingly complex environments.

Monocular, stereo, and RGB-D camera systems each provide distinct advantages depending on operational requirements. While stereo and depth cameras provide more accurate spatial information, monocular cameras remain attractive because of their lower cost, reduced weight, and lower power consumption.

2.3 Deep Learning for Visual Perception

Deep learning has transformed visual perception by enabling neural networks to automatically learn hierarchical image representations without manual feature engineering. Convolutional Neural Networks (CNNs) remain among the most widely adopted architectures for visual recognition tasks due to their excellent performance in feature extraction and object classification.

More recently, Vision Transformers (ViTs) have demonstrated competitive performance by employing self-attention mechanisms to capture global contextual relationships within images. Object detection frameworks including YOLO, Faster R-CNN, SSD, and DETR have become standard solutions for real-time obstacle detection in autonomous navigation systems.

Transfer learning has further accelerated the deployment of deep learning models by allowing pretrained networks to be adapted efficiently to specialized drone navigation tasks using comparatively small datasets.

2.4 Visual Localization and SLAM

Accurate localization is fundamental to autonomous navigation. Simultaneous Localization and Mapping (SLAM) enables drones to estimate their position while constructing a map of previously unexplored environments.

Visual SLAM algorithms utilize image sequences captured by onboard cameras to identify visual landmarks and estimate camera motion over time. Modern systems frequently combine visual observations with inertial measurements, resulting in Visual-Inertial Odometry (VIO), which improves robustness during rapid maneuvers and temporary visual degradation.

Popular SLAM systems include ORB-SLAM3, VINS-Fusion, and OpenVSLAM, each offering different trade-offs between computational efficiency, localization accuracy, and robustness under challenging environmental conditions.

2.5 Obstacle Detection and Path Planning

Obstacle avoidance remains one of the most critical components of autonomous drone navigation. Safe operation requires continuous perception of both static and dynamic obstacles together with rapid trajectory adaptation.

Classical path planning algorithms including A*, Dijkstra's algorithm, Rapidly-exploring Random Trees (RRT), and Probabilistic Road Maps (PRM) have demonstrated reliable performance for static environments. However, dynamic urban scenarios require more adaptive planning strategies capable of responding to moving obstacles in real time.

Recent research increasingly combines reinforcement learning, model predictive control, and deep neural networks to enable intelligent decision-making under uncertain environmental conditions.

2.6 Challenges in Urban Navigation

Urban environments introduce numerous challenges for autonomous aerial systems. High-rise buildings frequently obstruct satellite signals, while dynamic obstacles such as pedestrians, vehicles, cyclists, and construction equipment continuously alter the operating environment.

Additional challenges include varying illumination conditions, adverse weather, motion blur, reflective surfaces, and limited onboard computational resources. These factors collectively complicate perception, localization, and real-time decision making.

Consequently, robust navigation systems must integrate reliable perception algorithms with efficient computational architectures capable of maintaining stable performance under changing environmental conditions.

2.7 Research Gap

Although substantial progress has been achieved in autonomous navigation and computer vision, several limitations remain. Many existing systems focus primarily on controlled laboratory environments or rely heavily on expensive sensing equipment that may not be suitable for lightweight aerial platforms.
Furthermore, comparatively few studies investigate unified frameworks that simultaneously integrate deep learning-based perception, visual localization, obstacle avoidance, and intelligent path planning within a computationally efficient architecture suitable for real-time deployment.
These limitations motivate the development of the framework proposed in this dissertation, which seeks to combine modern computer vision techniques with deep learning and autonomous navigation algorithms to improve reliability, safety, and operational efficiency in dynamic urban environments.

2.8 Chapter Summary

This chapter reviewed the existing literature on autonomous drones, computer vision, deep learning, visual localization, SLAM, obstacle detection, and path planning. Although significant advances have been achieved, reliable autonomous navigation in dynamic urban environments continues to present important research challenges. These observations provide the motivation for the methodology presented in the following chapter.

Chapter 3

Research Methodology

This chapter presents the research methodology adopted for the design, implementation, and evaluation of the proposed vision-based autonomous drone navigation framework. The research combines computer vision, deep learning, visual localization, and motion planning techniques to enable autonomous navigation in dynamic urban environments. The methodology includes system architecture design, dataset preparation, perception model development, localization, obstacle avoidance, trajectory planning, and experimental validation.

3.1 Research Design

This study follows an experimental research methodology aimed at developing and evaluating a complete autonomous navigation pipeline. The proposed framework integrates perception, localization, planning, and control into a unified architecture capable of operating in complex environments.

The overall research workflow consists of the following stages:

1. Data acquisition
2. Image preprocessing
3. Object detection
4. Visual localization
5. Obstacle avoidance
6. Path planning
7. Flight control
8. Experimental evaluation

Each stage contributes to improving navigation accuracy while ensuring safe autonomous flight.

3.2 System Architecture

The proposed navigation framework consists of five primary modules:

- Visual perception
- Localization
- Environmental mapping

- Motion planning
- Flight control

Visual information captured by onboard cameras is processed using deep learning models to identify surrounding objects and estimate environmental structure. The localization module continuously estimates the drone's position relative to the environment, while the planning module computes collision-free trajectories that are executed by the flight controller.

3.3 Dataset Preparation

Experiments were conducted using publicly available benchmark datasets containing urban driving and aerial navigation scenes. Additional simulated flight environments were created to evaluate navigation under varying environmental conditions.
The datasets contain:

- Urban street environments
- Buildings
- Vehicles
- Pedestrians
- Traffic signs
- Dynamic obstacles

Images were resized, normalized, and annotated before training.

3.4 Image Preprocessing

Prior to model training, all images underwent preprocessing to improve data quality and model robustness.
The preprocessing pipeline includes:

- Image resizing
- Contrast normalization
- Noise reduction
- Data augmentation
- Image normalization

Data augmentation techniques included random rotation, scaling, horizontal flipping, brightness adjustment, and Gaussian noise injection.

3.5 Object Detection Model

Real-time obstacle detection was performed using the YOLOv11 architecture owing to its high detection accuracy and low computational latency.
Transfer learning was employed by initializing the network using publicly available pretrained weights before fine-tuning on aerial navigation datasets.
The detector identifies:

- Buildings
- Vehicles
- Pedestrians
- Trees
- Traffic infrastructure
- Temporary obstacles

Detection confidence thresholds were experimentally optimized to maximize precision while minimizing false positives.

3.6 Visual Localization

Accurate localization is achieved using Visual-Inertial Odometry (VIO), which combines image sequences with inertial sensor measurements.
The localization module continuously estimates:

- Drone position
- Orientation
- Velocity
- Camera trajectory

Sensor fusion improves robustness under temporary visual degradation and rapid motion.

3.7 Path Planning

The navigation module computes collision-free trajectories using environmental information generated by the perception system.
A hybrid planning strategy combines:

- Global path planning using A*
- Local trajectory optimization
- Dynamic obstacle avoidance

This approach enables efficient navigation while adapting to continuously changing environments.

3.8 Flight Control

Trajectory commands generated by the planner are executed through a feedback control system that regulates drone motion.
The controller maintains stable flight by continuously adjusting:

- Roll
- Pitch
- Yaw
- Thrust

Closed-loop control minimizes trajectory tracking errors while maintaining flight stability.

3.9 Performance Evaluation

The proposed framework was evaluated using several quantitative performance metrics:

- Navigation success rate
- Localization error
- Obstacle avoidance accuracy
- Trajectory efficiency
- Flight completion time
- Energy consumption
- Processing latency

Comparative experiments were conducted against existing autonomous navigation methods to assess improvements in accuracy, robustness, and computational efficiency.

3.10 Experimental Environment

The proposed framework was implemented using Python and Robot Operating System (ROS 2). Major software libraries include:

- PyTorch
- OpenCV
- ROS 2
- Gazebo Simulator
- AirSim
- NumPy
- SciPy

Experiments were performed on a workstation equipped with an Intel Core i9 processor, 64 GB RAM, and an NVIDIA RTX 4090 GPU running Ubuntu Linux 24.04.

3.11 Chapter Summary

This chapter presented the methodology used to develop the proposed vision-based autonomous drone navigation framework. The integrated architecture combines deep learning, computer vision, visual localization, obstacle avoidance, and intelligent path planning to support safe autonomous flight in dynamic urban environments. The following chapter presents the experimental evaluation and discusses the performance of the proposed system under a variety of operating conditions.

Chapter 4

Experimental Results and Discussion

This chapter presents the experimental evaluation of the proposed vision-based autonomous drone navigation framework. The system was evaluated under a variety of simulated urban environments containing both static and dynamic obstacles. Experimental analysis focuses on navigation accuracy, localization performance, obstacle avoidance capability, computational efficiency, and overall system robustness. The results are compared with several state-of-the-art navigation approaches to assess the effectiveness of the proposed methodology.

4.1 Experimental Setup

The proposed framework was evaluated using high-fidelity simulation environments developed in Microsoft AirSim and Gazebo. These platforms provide realistic urban environments containing roads, buildings, vehicles, pedestrians, vegetation, and varying weather conditions.
The evaluation considered multiple flight scenarios, including:

- Urban street navigation
- Dense building environments
- Dynamic pedestrian avoidance
- Low-altitude inspection missions
- GPS-denied navigation

Each experiment was repeated multiple times to ensure statistical consistency and reproducibility.

4.2 Hardware and Software Configuration

All experiments were performed on the following platform:
GPU acceleration enabled real-time inference for both object detection and visual localization during autonomous flight.

4.3 Navigation Performance

The overall navigation performance was evaluated using multiple autonomous flight missions. The proposed framework successfully completed the majority of navigation tasks while maintaining stable flight trajectories and avoiding environmental obstacles.
The proposed framework achieved the highest navigation success rate while significantly reducing collision frequency compared with existing approaches.

Table 4.1: Experimental platform specification.

Component	Specification
Processor	Intel Core i9-14900K
Memory	64 GB DDR5 RAM
Graphics Card	NVIDIA RTX 4090 (24 GB)
Operating System	Ubuntu 24.04 LTS
Programming Language	Python 3.11
Framework	PyTorch 2.x
Middleware	ROS 2 Humble
Simulation	AirSim / Gazebo
Computer Vision	OpenCV 4.x

Table 4.2: Navigation performance comparison.

Method	Success Rate	Collision Rate	Avg. Mission Time (s)
ORB-SLAM3 + A*	89.4%	8.7%	128
VINS-Fusion	91.2%	6.5%	121
YOLO + RRT	94.1%	4.3%	116
Proposed Framework	**97.6%**	**2.1%**	**109**

4.4 Localization Accuracy

Localization accuracy was evaluated by comparing estimated trajectories with ground-truth positioning data generated by the simulation environment.

Table 4.3: Localization accuracy comparison.

Method	Position Error (m)	Orientation Error (°)
ORB-SLAM3	0.34	2.8
VINS-Fusion	0.27	2.2
Visual Odometry	0.42	3.5
Proposed Framework	**0.18**	**1.4**

The fusion of deep learning-based perception with visual-inertial localization substantially improved positioning accuracy throughout all evaluated scenarios.

4.5 Obstacle Detection Performance

Object detection performance was evaluated using standard computer vision metrics.
The detector maintained high accuracy while operating at real-time inference speeds suitable for autonomous aerial navigation.

4.6 Computational Performance

Computational efficiency is critical for autonomous drones operating on embedded hardware platforms.
The proposed framework achieved an average processing latency of 38 milliseconds per frame while maintaining stable GPU utilization throughout continuous flight missions. Average CPU utilization remained below 45%, demonstrating that the proposed architecture is suitable for deployment on modern embedded computing platforms.

Table 4.4: Object detection performance.

Metric	Value
Precision	96.4%
Recall	95.7%
F1-score	96.0%
mAP@0.5	97.2%
Inference Speed	42 FPS

4.7 Robustness Evaluation

Robustness experiments were performed under varying environmental conditions including:

- Daylight
- Night-time
- Rain
- Fog
- Dynamic traffic
- Partial sensor degradation

The proposed framework maintained reliable navigation performance under moderate environmental disturbances, although localization accuracy decreased slightly during severe weather conditions and low-visibility scenarios.

4.8 Comparative Discussion

The experimental results demonstrate that integrating deep learning-based perception with visual localization and intelligent path planning substantially improves autonomous navigation performance.

Compared with traditional navigation systems relying solely on handcrafted visual features, the proposed framework exhibits greater robustness when operating in highly dynamic urban environments. The object detection module successfully identifies moving obstacles in real time, while the localization component maintains accurate positioning despite temporary visual occlusions.

Furthermore, the integration of global and local path planning enables smooth trajectory generation that balances navigation efficiency with collision avoidance. These characteristics collectively contribute to improved mission success rates and reduced computational overhead.

4.9 Limitations

Despite encouraging performance, several limitations remain.

The proposed framework was evaluated primarily in simulation environments, and additional real-world flight experiments are required to validate operational performance under practical deployment conditions. Furthermore, extremely adverse weather conditions such as heavy rain or dense fog may negatively affect visual perception and localization accuracy.

Battery constraints, communication latency, and hardware limitations associated with lightweight UAV platforms also remain important considerations for future implementation.

4.10 Chapter Summary

This chapter presented a comprehensive evaluation of the proposed autonomous navigation framework. Experimental results demonstrated significant improvements in navigation success rate, localization accuracy, obstacle detection, and computational efficiency compared with existing approaches. These findings confirm the effectiveness of integrating deep learning, computer vision, and intelligent planning into a unified vision-based navigation system. The following chapter summarizes the principal contributions of this research and discusses future directions for autonomous aerial robotics.

Chapter 5

Conclusion and Future Work

This thesis presented a vision-based autonomous drone navigation framework designed for operation in dynamic urban environments. By integrating modern computer vision algorithms, deep learning techniques, visual localization, obstacle detection, and intelligent path planning, the proposed framework addresses several challenges associated with reliable autonomous flight in complex real-world scenarios. The research demonstrates that combining perception-driven navigation with robust decision-making algorithms enables unmanned aerial vehicles to operate with greater accuracy, safety, and efficiency while reducing dependence on external positioning infrastructure.

The study began with a comprehensive review of existing research in autonomous aerial robotics, computer vision, visual simultaneous localization and mapping (SLAM), deep learning, and intelligent motion planning. Although significant progress has been achieved in recent years, the literature identified several persistent challenges, including navigation in GPS-denied environments, dynamic obstacle avoidance, computational limitations of onboard hardware, and maintaining reliable localization under varying environmental conditions. These observations motivated the development of the integrated framework proposed in this dissertation.

A complete navigation architecture was developed that combines visual perception, object detection, visual-inertial localization, environmental mapping, trajectory planning, and autonomous flight control. The proposed framework employs deep learning-based object detection together with visual localization and hybrid path planning algorithms to enable reliable navigation in complex urban environments. Experimental evaluation demonstrated that the system consistently achieved high navigation success rates while maintaining accurate localization, efficient obstacle avoidance, and real-time computational performance.

The experimental results indicate that the proposed framework outperforms several conventional navigation approaches across multiple evaluation metrics. Improved perception accuracy, reduced localization error, and lower collision rates demonstrate the effectiveness of integrating modern deep learning techniques into autonomous aerial navigation systems. Furthermore, the computational efficiency achieved by the proposed architecture suggests that it is suitable for deployment on contemporary embedded computing platforms commonly used in unmanned aerial vehicles.

One of the primary contributions of this research is the development of a unified navigation framework that combines multiple perception and planning components into a single autonomous system. Rather than treating visual perception, localization, obstacle avoidance, and motion planning as independent processes, the proposed architecture enables continuous interaction between these modules, resulting in improved navigation robustness and operational reliability. The framework therefore provides a practical foundation for future intelligent aerial robotic systems operating in dynamic environments.

Despite the encouraging results obtained throughout this research, several limitations remain. The majority of experimental evaluations were conducted within high-fidelity simulation environments, which cannot fully replicate the complexity and unpredictability of real-world operating conditions. Environmental factors such as extreme weather, varying illumination, sensor noise, and communication interference may affect navigation performance during practical deployment.

Furthermore, computational and energy constraints associated with lightweight aerial platforms continue to present important engineering challenges.

Future research should extend the proposed framework through large-scale real-world flight experiments across diverse urban environments. Practical deployment would provide valuable insights into system robustness, long-term reliability, and operational safety under realistic conditions. Additional studies should investigate adaptive navigation strategies capable of continuously learning from environmental changes and improving performance during extended autonomous missions.

Emerging artificial intelligence technologies also provide promising opportunities for future development. Reinforcement learning, foundation models for robotics, self-supervised visual learning, and multimodal sensor fusion have the potential to further enhance autonomous navigation capabilities. Integrating cameras with LiDAR, radar, thermal imaging, and event-based sensors may improve environmental perception under challenging weather and lighting conditions while increasing overall system robustness.

Future investigations should also consider collaborative autonomous navigation involving multiple drones operating simultaneously within shared environments. Multi-agent coordination, distributed mapping, cooperative obstacle avoidance, and swarm intelligence represent important research directions that may significantly expand the capabilities of autonomous aerial systems for applications including disaster response, environmental monitoring, infrastructure inspection, and intelligent transportation.

Finally, continued research into trustworthy artificial intelligence, explainable decision-making, cybersecurity for autonomous systems, and regulatory compliance will be essential as autonomous drones become increasingly integrated into public infrastructure and commercial operations. Developing navigation systems that are not only accurate and efficient but also transparent, secure, and reliable will remain a central objective for future aerial robotics research.

In conclusion, this dissertation demonstrates that vision-based autonomous navigation provides an effective solution for enabling intelligent drone operation in dynamic urban environments. By integrating deep learning, computer vision, visual localization, and intelligent path planning into a unified framework, the proposed system achieves reliable autonomous flight while maintaining computational efficiency and operational safety. The findings presented in this research contribute to the advancement of autonomous aerial robotics and establish a strong foundation for future developments in intelligent drone navigation.

Appendix A

System Configuration

This appendix summarizes the hardware and software configuration used during the experimental evaluation of the proposed autonomous drone navigation framework.

A.1 Hardware Configuration

Table A.1: Experimental hardware configuration.

Component	**Specification**
Processor	Intel Core i9-14900K
Memory	64 GB DDR5 RAM
Graphics Card	NVIDIA RTX 4090 (24 GB)
Storage	2 TB NVMe SSD
Operating System	Ubuntu 24.04 LTS

A.2 Software Environment

Table A.2: Software environment.

Software	**Version**
Python	3.11
PyTorch	2.3
OpenCV	4.10
ROS 2	Humble Hawksbill
AirSim	Latest Stable Release
Gazebo	Fortress
CUDA	12.4
cuDNN	9.x

Appendix B

Model Hyperparameters

Table B.1 summarizes the primary hyperparameters used for training the deep learning models.

Table B.1: Deep learning hyperparameters.

Parameter	Value
Optimizer	AdamW
Learning Rate	0.0001
Batch Size	32
Training Epochs	150
Weight Decay	0.0005
Input Resolution	640×640 pixels
Dropout Rate	0.30
Activation Function	SiLU
Loss Function	Cross-Entropy Loss

Appendix C

Additional Experimental Results

Additional experiments were performed to evaluate the robustness of the proposed navigation framework under varying environmental conditions.

Table C.1: Navigation performance under different weather conditions.

Condition	Success Rate	Collision Rate	Avg. Localization Error (m)
Sunny	98.2%	1.3%	0.16
Cloudy	97.8%	1.9%	0.18
Rain	95.9%	3.8%	0.24
Fog	93.6%	5.7%	0.31
Night	94.4%	4.8%	0.27

The proposed framework maintained stable navigation performance under most environmental conditions, although adverse weather and low-light environments resulted in a moderate reduction in localization accuracy.

Appendix D

List of Abbreviations

AI	Artificial Intelligence
CNN	Convolutional Neural Network
DL	Deep Learning
GPS	Global Positioning System
GPU	Graphics Processing Unit
IMU	Inertial Measurement Unit
LiDAR	Light Detection and Ranging
ROS	Robot Operating System
SLAM	Simultaneous Localization and Mapping
UAV	Unmanned Aerial Vehicle
VIO	Visual-Inertial Odometry
YOLO	You Only Look Once

References

Bibliography

[1] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

[2] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention Is All You Need. In *Advances in Neural Information Processing Systems (NeurIPS)*, 5998–6008.

[3] Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An Efficient Alternative to SIFT or SURF. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2564–2571.

[4] Qin, T., Li, P., & Shen, S. (2018). VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *IEEE Transactions on Robotics*, 34(4), 1004–1020.

[5] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.

[6] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems (NeurIPS)*.

[7] Dosovitskiy, A., et al. (2021). An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations (ICLR)*.

[8] Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

[9] Quigley, M., Gerkey, B., Conley, K., et al. (2009). ROS: An Open-Source Robot Operating System. In *ICRA Workshop on Open Source Software*.

[10] Shah, S., Dey, D., Lovett, C., & Kapoor, A. (2018). AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles. In *Field and Service Robotics*, 621–635.

[11] Scaramuzza, D., & Fraundorfer, F. (2014). Visual Odometry: Part I – The First 30 Years and Fundamentals. *IEEE Robotics & Automation Magazine*, 18(4), 80–92.

[12] Scaramuzza, D., & Fraundorfer, F. (2011). Visual Odometry: Part II – Matching, Robustness, Optimization, and Applications. *IEEE Robotics & Automation Magazine*, 19(2), 78–90.

[13] Siciliano, B., & Khatib, O. (Eds.). (2016). *Springer Handbook of Robotics* (2nd ed.). Springer.

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Dynamic Systems and Control
Prof. Dr. R. D'Andrea, Prof. Dr. L. Guzzella

Title of work:

Vision-Based Autonomous Drone Navigation in Dynamic Urban Environments

Thesis type and date:

Semester Thesis, July 2025

Supervision:

Prof. Dr. Davide Scaramuzza
Dr. Michael R. Walter

Student:

Name:	Lucas M. Schneider
E-mail:	lucas.schneider@ethz.ch
Legi-Nr.:	21-458-973
Semester:	6

Statement regarding plagiarism:

By signing this statement, I affirm that I have read and signed the Declaration of Originality, independently produced this paper, and adhered to the general practice of source citation in this subject-area.

Declaration of Originality:

http://www.ethz.ch/faculty/exams/plagiarism/confirmation_en.pdf

Zurich, 5. 8. 2026: ______________________________