

LB/TH/20/2026
TH6209

**INTERSECTION WAYPOINT IDENTIFICATION
FOR VISION BASED INDOOR NAVIGATION OF
NANO-SCALE DRONES**

H.K.A.Y.D. Kodithuwakku

248111U

Master of Science(Major Component of Research)

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

February 2026

INTERSECTION WAYPOINT IDENTIFICATION FOR VISION BASED INDOOR NAVIGATION OF NANO-SCALE DRONES

H.K.A.Y.D. Kodithuwakku

248111U

Dissertation submitted in partial fulfillment of the requirements for the
degree
Master of Science(Major Component of Research)

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

February 2026

DECLARATION

I declare that this is my own work and this Dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date: 18.02.2026

The above candidate has carried out research for the Master of Science(Major Component of Research) Dissertation under our supervision. We confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Dr. Chathuranga Hettiarachchi

Signature of the Supervisor:

Date: 18/02/2026

Name of Supervisor: Dr. Sulochana Sooriyaarachchi

Signature of the Supervisor:

Date: 18/02/2026

ACKNOWLEDGEMENT

This thesis would not have been feasible without the help, direction, and inspiration of numerous people and organizations, for which I am extremely grateful.

First of all, I would like to express my sincere gratitude to my supervisors, Dr. Chathuranga Hettiarachchi and Dr. Sulochana Sooriyaarachchi. Throughout this work, their visionary insights and meticulous dedication to detail improved the work significantly. I am especially grateful for the valuable time they dedicated to guiding me, despite their demanding schedules. Their unwavering faith in my abilities, helpful criticism, and patient coaching inspired me to overcome obstacles and hone my concepts into a solid, unified system.

Additionally, I owe a debt of gratitude to my examiners, Professor Chathura de Silva and Dr. Sunimal Rathnayake. I strengthened my arguments and clarified my contributions as a result of their thorough examination of my study during progress reviews and their insightful inquiries.

I would like to express my sincere gratitude to Senior Professor N.D. Gunawardena, Vice Chancellor of University of Moratuwa and Professor J.M.A. Manatunge, Dean of the Faculty of Engineering for their continued support towards administrative matters. I also gratefully acknowledge Dr. Uthayasanker Thayasivam, Head of the Department of Computer Science and Engineering, University of Moratuwa for his support. I would like to express my appreciation to all of the faculty and administrative staff for their effective management of academic and administrative issues, which enabled me to devote all of my attention to my project. I sincerely appreciate the generous support from the University of Moratuwa Senate Research Committee grant SRC/LT/2024/02.

I gratefully acknowledge Mr. Ignatious Peiris for the technical and moral support. I would also like to thank Mr. Kenul Perera for his efforts in creating video to simulation tool. I would also like to thank Mr. Vihanga Wickramage, Mr. Sandun Dushyantha for the synthetic environment building. I am grateful to my colleagues at the IntelliSense Lab for their technical support, early-stage discussions, and the camaraderie that made the research process both productive and enjoyable.

Personally, I am extremely thankful to my parents and my brother. My biggest source of strength has been their faith in my abilities. I want to express my sincere gratitude to my friends for their support and encouraging remarks.

Lastly, I would like to thank everyone who helped with this research, whether directly or indirectly: the creators of open-source datasets and tools, the peer reviewers whose work influenced my literature review, and the larger academic community whose findings served as the foundation for this thesis.

ABSTRACT

Nano-scale drones represent a promising platform for indoor inspection and search operations in GPS-denied, constrained environments, as their compact form factor enables access to narrow spaces. However, they require lightweight, resource-efficient autonomous navigation algorithms. Vision-based navigation offers an infrastructure-free substitute to GPS and external positioning systems. Simple vision-based navigational algorithms such as ring attractor, optical flow-based approaches, suffer from drift and cumulative error over time. Introducing semantic waypoints provides a mechanism to mitigate such drift.

This work proposes an egocentric, vision-based waypoint navigation framework that formulates waypoint detection as an intersection understanding problem, cast as a concept learning task. We first formulate the intersection identification problem as a multi-class classification. To ensure a model can be trained with limited data and still preserve the conceptual understanding, each intersection is decomposed into its fundamental directional components: left, right, and forward. This novel multi-label classification is advantageous because it can extend to zero-shot inference and complex junction types. Further, we augment the limited real data with synthetic visual data collected from a safer digital twin environment which replicates the real-world scenes. We design five experimental setups in which model weights are initialized from Kaiming He, ImageNet or synthetic pre-trained weights obtained using the Digital Twin dataset.

We show that, (i) Intersections can be used as waypoints for the navigation of nano-drones. (ii) Multi-label classification outperforms multi-class formulation in intersection identification. (iii) Model trained using digital twin data alone can yield an F1 score of 0.6 in real-world predictions, compared to 0.5 from a random classifier. (iv) Introducing limited real-world training data can bring the model up to near-perfect accuracy with faster convergence. (v) The multi-label classification of junctions provides insightful visual explanations ensuring reliability and generalizability of the conceptual framework. Finally, the proposed framework is effective in environments such as libraries and supermarkets where structured aisle naturally exists and GPS is unavailable. This method offers an infrastructure and drift-resistant solution for the navigation of nano-scale drones by leveraging intersections as semantic waypoints.

Keywords: ego-centric vision, explainability, concept learning, nano-scale drones

TABLE OF CONTENTS

Declaration of the Candidate & Supervisor	i
Acknowledgement	ii
Abstract	iii
Table of Contents	iv
List of Figures	vii
List of Tables	x
List of Abbreviations	x
List of Appendices	xii
1 Introduction	1
1.1 UAV Navigation	1
1.2 UAVs taxonomy based on the size	3
1.3 State-of-the-Art Navigation for Nano-Scale Drones: PULP DroNet	4
1.3.1 DroNet on board control framework	5
1.4 Problem Statement, Objectives and Contributions	7
2 Literature Survey	9
2.1 Vision based Intersection Navigation	9
2.1.1 Egocentric Vision	9
2.1.2 Visual Way Point based Navigation	10
2.1.3 Intersection Identification	11
2.2 Performance Metrics Evaluation	12
2.2.1 Hamming Loss	13
2.2.2 Precision	13
2.2.3 Recall	13
2.2.4 F1 score	13
2.2.5 Grad-CAM for interpretability	13
3 Methodology	16
3.1 Hardware Framework	16

3.1.1	UAV Platform	16
3.1.2	Sensor Peripherals	17
3.2	Software Stack	17
3.2.1	Crazyflie PC Client	17
3.2.2	Estimator Module	18
3.2.3	Commander Module	19
3.2.4	Controller Module	20
3.3	Replicating Vision based Obstacle Avoidance Algorithm - DroNet	20
3.3.1	System Prerequisites	20
3.3.2	Flashing and Testing	22
3.3.3	Collision avoidance vs Intelligent Navigation	24
3.4	Intersection Identification Framework	25
3.4.1	Real-World Environmental Setup	27
3.4.2	Digital Twin Environmental Setup	27
3.4.3	Data Collection	28
3.4.4	AI model Creation Strategy A- Multi Class classification	30
3.4.5	Importance of learning orthogonal features	31
3.4.6	AI model Creation Strategy B- Multi Label classification	31
4	Results and Discussion	35
4.1	Multi Class Classification	35
4.2	Multi Label Classification	36
4.2.1	Results of ResNet50	36
4.2.2	Results of MobileNetV2	44
4.3	Explanation of Grad-CAM results	44
4.4	Inference results	47
5	Conclusion and Future Work	49
5.1	Conclusion	49
5.2	Limitations and Future Work	50
	References	52
	Appendix A MobilenetV2 results	58
	Appendix B Grad-CAM ablation study	62

Appendix C	Video to Simulation Tool	64
C.1	Simulation to Real Image Generation	64
C.2	Pipeline for Model Generation	64
Appendix D	Model Quantization	66
Appendix E	Safe Landing Mechanism	67
E.1	Flow Deck working principle	67
E.2	Flow Deck's Sensor readings in different floors	67

LIST OF FIGURES

Figure	Description	Page
Figure 1.1	Taxonomy of vision-based UAV navigation systems, reprinted from[1].	2
Figure 1.2	Drift trajectories of event-only and frame-only navigation methods in loop closing, reprinted from [2].	3
Figure 1.3	PULP Dronet’s data processing framework until the setpoint is communicated to the drone’s commander[3]	6
Figure 1.4	Overall Concept(a) Global coordinate based mission plan (existing) (b) Global coordinate velocity setpoint based mission plan (existing) (c) Egocentric vision-waypoint based mission plan (proposed) (d) Mission plan of the drone (e) Staggered junction - an example of a complex junction which can be inferred zero shot. (f) Orthogonal decomposition of junction vision-waypoints to left-forward-right navigability (g) Validation of conceptual understanding of left-forward-right navigability of the proposed model with Grad-CAM.	7
Figure 3.1	The robotic platform Crazyflie 2.1+, Flow deck and AI-deck with their microcontrollers, adapted from[4]	16
Figure 3.2	User interface of CFClient for connecting to, controlling, and logging data from the Crazyflie.	18
Figure 3.3	General framework of the stabilization structure of the crazyflie with setpoint handling. * This part takes place on the computer through the CFlib for python, so there is also communication protocol in between. It is left out of this schematics for easier understanding(Reprinted from Bitcraze documentation)	19
Figure 3.4	Schematic overview of inputs and outputs of the Kalman Filter, the inputs of our system are highlighted in red(Adapted from Bitcraze Documentation)	19
Figure 3.5	(a)Pathway of the PID controller(Reprinted from Bitcraze Documentation) (b) Proposed controller framework(Adapted from Bitcraze Documentation)	21
Figure 3.6	Overview of the Problem scenario vs Solution scenario	22
Figure 3.7	Three representative examples of drone collisions during onboard execution of the PULP-DroNetV3 model	23
Figure 3.8	Probability of collision and steering angle from (a) seen environments and (b) unseen environments.	24
Figure 3.9	DroNet fails in Decision points	24

Figure 3.10	On Board Processing	25
Figure 3.11	Off Board Processing	25
Figure 3.12	GUI for human-in-the-loop decision-making during intersection-based navigation	26
Figure 3.13	Two different environment configurations in Webots	28
Figure 3.14	Data collected from real-world scenes	29
Figure 3.15	Experimental design workflow of the proposed methodology. Five different experiments were conducted, varying in weight initialization and training procedure. 78% of data used for training, 19% used for validation, 3% used for testing (a)Kaiming He initialization (b)ImageNet initialization (c)Digital twin initialization trained on last two layers (d)Digital twin initialization trained on all layers(e)Digital twin initialization tested on real data, no training using real-world data.	32
Figure 4.1	Training graphs of multi class configuration	36
Figure 4.2	Training Accuracy across different weight initialization methods	37
Figure 4.3	Training Loss across different weight initialization methods	37
Figure 4.4	Validation Accuracy across different weight initialization methods	38
Figure 4.5	Validation loss across different weight initialization methods	38
Figure 4.6	Successful decomposition of Left-Forward-Right mutually exclusive (i.e. orthogonal) conceptual understanding by the digital twin (2-layers) model depicted using the Grad-Cam heatmaps across different junctions. Predicted probabilities are indicated below each example(ResNet50).	41
Figure 4.7	Sequentially generated Grad-CAM heatmaps illustrating label-wise consistency across a navigation sequence. (a) Right-label activation at a right bend; after executing the turn, right-edge features are no longer highlighted. (b) Right-label activation and Re-emergence of right-edge activations as the platform approaches a T-intersection. (c) Right-label activation from right intersection to forward path. (d) Left-label activation indicating leftward navigability at T-intersections	42
Figure 4.8	Comparison of five experiments across Hamming Loss, Precision, Recall, and F1 Score on test data(Resnet50).	43
Figure 4.9	Reconstruction of navigational-affordance maps reprinted from [5] with permission	45
Figure 4.10	VQA Visualizations reprinted from[6]. Copyright © 2017, IEEE	46
Figure 4.11	Example of how human annotates intersections, adapted from the [7]. Copyright © 2019, IEEE. The green box highlights the human perception of a right-bend intersection(red underline).	47
Figure 4.12	Left, Right, and Forward inference probabilities in supermarket and library environments, illustrated using randomly selected snapshots from different timestamps	48

Figure A.1	Training graphs of different configurations for MobileNetV2 model(a) Kaiming He Initialization (b) ImageNet Initialization (c) Synthetic last two (d) Synthetic all	59
Figure A.2	Comparison of four experiments across Hamming Loss, Precision, Recall, and F1 Score on test data(MobileNetV2).	59
Figure B.1	(a)Representation of heatmap attention reduction after masking using white dots in the Masked Original Image. (b) Comparison of how the probability of the label position changes after masking is performed.	63
Figure C.1	Synthetic(image of epoch 1) to Real(image of epoch 30) translation	64
Figure C.2	Video to simulation pipeline	65
Figure C.3	Imported 3D model to Webots showing scaling issues and only one sided aisle construction	65
Figure E.1	Feature tracking with the time(Reprinted from Bitcraze Documentation)	67
Figure E.2	Flow deck sensor values on a Green Mat	68
Figure E.3	Flow deck sensor values on Rigifoam Floor	68
Figure E.4	Flow deck sensor values on Wooden Floor	69
Figure E.5	Flow deck sensor values on Checkered floor	69

LIST OF TABLES

Table	Description	Page
Table 1.1	UAVs taxonomy by vehicle class-size	3
Table 2.1	Comparative synthesis of related literature	15
Table 3.1	Data distribution across training and testing configurations	29
Table 3.2	Training configuration for multi-class and multi-label experiments	30
Table 3.3	Summary of the objectives of the experiments	32
Table 4.1	Grad-CAM visualization results for different intersection types for multi-class configuration (two samples from different locations for each class)	35
Table 4.2	K-Fold performance comparison across experiments A–D	36
Table 4.3	Grad-CAM visualization results for multi-label configuration on the He and ImageNet settings using three different samples (ResNet50)	39
Table 4.4	Grad-CAM visualization results for multi-label configuration on digital twin settings using three different samples (ResNet50)	40
Table 4.5	ImageNet model on our test data vs EgoCart data	43
Table 4.6	Model complexity and resource comparison	48
Table A.1	K-Fold performance comparison across experiments A–D (MobileNetV2)	58
Table A.2	Grad-CAM visualization results for multi-label configuration on the He and ImageNet settings using three different samples (MobileNetV2)	60
Table A.3	Grad-CAM visualization results for multi-label configuration on the digital twin settings using three different samples (MobileNetV2)	61
Table D.1	Performance on test data by ResNet50(last 2 config)	66
Table D.2	Performance on test data by MobileNetV2(last 2 config)	66

LIST OF ABBREVIATIONS

Abbreviation	Description
AVL	Audio-Visual-Language
CfLib	Crazyflie Python Library
CNN	Convolutional Neural Network
COTS	Commercial-off-the-shelf
CRTP	Crazyflie Real Time Protocol
DORY	Deployment ORiented MemorY
EVC	Early Visual Cortex
FLA	Fast Lightweight Autonomy
fMRI	Functional Magnetic Resonance Imaging
GPS	Global Positioning System
IIR	Infinite Impulse Response
IMU	Inertial Measurement Units
INDI	Incremental Nonlinear Dynamic Inversion
IoT	Internet of Things
MCU	Microcontroller Units
NEMO	NEural Minimizer for pyOrch
NUC	Next Unit of Computing
OPA	Occipital Place Area
PCB	Printed Circuit Board
PID	Proportional-Integral-Derivative
PPA	Parahippocampal Place Area
PULP	Parallel Ultra Low Power
RSC	Retrosplenial Complex
SaR	Search and Rescue
SLAM	Simultaneous Localization and Mapping
ToF	Time-of-Flight
TSM	Temporal Shift Module
UAV	Unmanned Aerial Vehicles
VLA	Vision-Language-Action Navigation
VQA	Visual Question Answering
WPN	Waypoint Prediction Networks
YOLO	You Only Look Once

LIST OF APPENDICES

Appendix	Description	Page
Appendix -A	MobilenetV2 results	58
Appendix -B	Grad-CAM ablation study	62
Appendix -C	Video to Simulation Tool	64
Appendix -D	Model Quantization	66
Appendix -E	Safe Landing Mechanism	67