XAI-UAV: Explainable Artificial Intelligence Framework for Robust Object Detection in Tactical Unmanned Aerial Vehicles

A CAPSTONE PROJECT REPORT

Submitted in partial fulfillment of the requirement for the award of the Degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE AND ENGINEERING

by

Kana Hakshay Reddy (22BCE9807)
Konakalla Rishitha (22BCE7331)
Appala Pranav Sai (22BCE7558)

Under the Guidance of

Dr. Deepasikha Mishra

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
VIT-AP UNIVERSITY
AMARAVATI- 522237

SEPTEMBER 2025

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 Summary

This project proposes the development of an Explainable Artificial Intelligence (XAI) framework for robust object detection in tactical Unmanned Aerial Vehicles (UAVs). UAVs have become indispensable in modern operations ranging from military reconnaissance and border surveillance to disaster response, environmental monitoring, and urban security. In such contexts, reliability, transparency, and resilience are as critical as accuracy. Although modern deep learning models, particularly YOLO-based architectures and Faster R-CNN, provide strong performance, their inherent *black-box* nature limits interpretability, posing challenges to operator trust, accountability, and compliance with regulatory standards [1][2].

The envisioned XAI-UAV framework will bridge this gap by embedding state-of-the-art object detection models with explainability modules such as Grad-CAM, saliency maps, and SHAP-based methods. These techniques will provide visual heatmaps and feature-based explanations, making model decisions interpretable and transparent. This explainability is especially critical in human-in-the-loop systems, where operators must validate AI-driven outputs under strict time constraints [3][4]. Furthermore, the framework will adopt lightweight detection models (e.g., YOLOv5, YOLOv8, Tiny YOLOv7) optimized for real-time deployment on resource-constrained edge devices such as NVIDIA Jetson Orin/Nano and FPGAs. To balance efficiency and interpretability, advanced model compression methods will be utilized, reducing computational overhead while maintaining reliability.

Beyond efficiency, the system emphasizes robustness against challenges inherent to aerial imagery: low resolution, cluttered backgrounds, variable altitudes, occlusion, adverse weather, and adversarial perturbations. Defense mechanisms against adversarial attacks will be explored to ensure stability in contested environments. Evaluation will include standard detection metrics such as mean Average Precision (mAP) and F1-score, alongside XAI-specific measures such as localization accuracy, faithfulness, stability, and operator usability [3]. Benchmark testing will leverage UAV-focused datasets such as VisDrone and DOTA, ensuring that the framework is validated across diverse and realistic scenarios [3] [5].

The anticipated outcomes include a deployable XAI-UAV software module, an efficient and interpretable detection pipeline, an integrated XAI explanation engine, and a comprehensive

evaluation and benchmarking report. Beyond defense, this framework has potential applications in disaster relief, traffic monitoring, and environmental studies.

Looking forward, this project also lays the foundation for future research directions, including multi-modal explainability (integrating visual, thermal, and sensor data), federated learning for distributed UAV fleets, and fully autonomous yet explainable UAV systems. By uniting accuracy, robustness, efficiency, and interpretability, the XAI-UAV framework aspires to establish a new paradigm of trustworthy AI in aerial platforms, directly advancing national defense capabilities while contributing to the broader vision of responsible, explainable AI for mission-critical applications [4].

1.2 Keywords

Explainable AI (XAI), Object Detection, Unmanned Aerial Vehicles (UAVs), Tactical Systems, Edge AI.

1.3 Objectives of Project

The central objective of this project is to develop an Explainable Artificial Intelligence (XAI) framework that delivers both robust and interpretable object detection capabilities in tactical UAVs. Unlike conventional deep learning models that prioritize raw accuracy while neglecting interpretability, this framework seeks to balance performance, transparency, and efficiency to meet the unique demands of UAV operations. A key objective is to design and train lightweight object detection models such as YOLOv5, YOLOv8, and Tiny YOLOv7 that are computationally optimized for real-time edge deployment on platforms with limited onboard resources [8].

Another objective is the integration of explainability techniques that move beyond post-hoc visualization, towards model-aware explainability that ensures operator comprehension and regulatory accountability [6]. Techniques such as Grad-CAM, LIME, SHAP, and saliency-based frameworks will be adapted to UAV imagery, producing interpretable visual and feature-level explanations that directly support human-in-the-loop decision-making. In parallel, the project will target robustness under operational constraints, including resistance to adversarial perturbations, resilience to low-light conditions, occlusions, cluttered aerial environments, and dynamic weather effects [9].

Furthermore, the framework will establish a dual-layer evaluation methodology: traditional accuracy-based metrics (mAP, F1-score) will be complemented by XAI-specific measures, such as

localization accuracy, stability of explanations, and consistency across environmental variations [10]. In doing so, the project aims to define new evaluation benchmarks for explainable UAV systems, filling a critical gap in both academic research and defence-focused AI deployment. Finally, the project aspires to contribute to the broader scientific discourse on trustworthy AI, providing methodologies that may extend to other mission-critical applications, including healthcare, autonomous driving, and industrial monitoring [11].

1.4 Expected Output and Outcome of The Proposal

The proposed research is expected to deliver a deployable XAI-UAV framework that combines robust object detection with transparent and interpretable decision-making. At its core, the system will provide a software module capable of running on edge devices, integrating both optimized YOLO-based detectors and explanation modules such as Grad-CAM and SHAP, adapted for aerial surveillance conditions [3]. The integration of pruning will ensure that the detection models achieve real-time efficiency while retaining interpretability, a balance that is often difficult to achieve in resource-constrained UAV platforms [3].

Beyond technical outputs, the project will generate a comprehensive benchmarking report, including results on conventional detection metrics such as mAP and F1-score, as well as XAI-specific indicators like localization accuracy, explanation stability, and operator usability. The adoption of UAV-specific datasets such as VisDrone and DOTA will ensure rigorous evaluation under realistic aerial scenarios [12].

The anticipated outcomes extend to operational, scientific, and societal impacts. Operationally, the framework will enhance trust and accountability in AI-assisted UAV systems by making decisions transparent to human operators, thereby improving mission reliability. Scientifically, the project will establish a methodology for integrating interpretability with efficiency, advancing the discourse on trustworthy AI in aerial platforms. Societally, the framework is expected to have implications beyond defence, supporting disaster relief, smart cities, and environmental monitoring, thereby aligning with the growing global demand for responsible AI in safety-critical applications [13].

1.5 Equipment Details

The realization of the XAI-UAV framework requires a synergy of UAV hardware, edge AI devices, and high-performance computing resources. For aerial deployment, quadrotor UAV platforms equipped with high-resolution RGB sensors will be used to capture aerial imagery and validate detection in real-world environments. To address the constraints of real-time deployment, NVIDIA Jetson Orin and Jetson Nano devices will serve as the primary edge hardware due to their proven effectiveness in embedded AI tasks [14]. Additionally, Field Programmable Gate Arrays (FPGAs) will be considered for latency-critical inference, offering energy-efficient acceleration in tactical scenarios [14].

For the training and fine-tuning of deep learning models, GPU-enabled workstations with modern accelerators like NVIDIA RTX A6000/3090 will provide the computational backbone. High-capacity storage infrastructure will be necessary to manage large-scale aerial datasets, including VisDrone, DOTA, and custom UAV-captured datasets. Simulation platforms such as AirSim and Gazebo will be employed for controlled experimentation, enabling the testing of UAV performance in variable conditions before field deployment [7], [15].

To extend robustness studies, supplementary sensors such as thermal and multispectral cameras may be integrated, addressing challenges in night-time or camouflage-heavy environments. Together, this equipment ecosystem ensures that the framework remains scalable, adaptable, and deployable, thereby bridging the gap between theoretical innovation and practical UAV operations [16].

In addition to computation and sensing hardware, the framework will also leverage networking and communication systems to support real-time data transfer between UAVs, ground control stations, and cloud infrastructures. This integration will enable distributed UAV operations, cooperative swarm intelligence, and federated learning approaches where multiple UAVs collaboratively train models without centralizing sensitive data [17], [18]. Such capabilities not only extend the scalability of the framework but also align with emerging requirements in defence, disaster management, and smart city surveillance, where coordinated multi-UAV deployments are increasingly vital.

CHAPTER 2

OTHER TECHNICAL DETAILS

2.1 Review of status of Research and Development

2.1.1 International Status

The international research landscape concerning XAI and robust object detection for UAVs is characterized by rapid advancements, largely propelled by the increasing integration of AI into safety-critical and highly dynamic operational environments. Research efforts span academic institutions, private industry, and defense research organizations, all of which are actively contributing to the foundational principles and practical applications of these technologies.

Global Evolution of Explainable AI for UAV Applications

The international research community has long recognized that the success of deep learning in Unmanned Aerial Vehicles (UAVs) is closely tied to its interpretability. UAVs are deployed in safety-critical contexts such as defense surveillance, disaster management, and autonomous monitoring, where errors can have severe consequences. To address these risks, researchers worldwide have advanced the field of Explainable Artificial Intelligence (XAI), enabling models to provide not only predictions but also rationales behind those predictions.

Early frameworks emerged in the late 2010s when Arrieta et al. [19] published one of the first comprehensive surveys of XAI, proposing a taxonomy of interpretable models and post-hoc methods while stressing the importance of responsibility and trustworthiness in AI systems. This work became a cornerstone reference for subsequent international projects. Building on this foundation, Samek et al. [20] consolidated methods for visualizing neural networks, laying out principles for evaluating explanations in complex deep learning systems. These surveys framed XAI as an indispensable research priority for high-risk domains, including UAV object detection.

International funding initiatives reflect the importance of XAI. The DARPA XAI program in the United States explicitly sought to create AI models that could “explain themselves” to human operators [1]. In Europe, multiple Horizon Europe projects were launched to establish explainability in autonomous systems, highlighting XAI as central to responsible innovation. Meanwhile, Asian researchers, particularly from China and Japan, contributed both algorithms and datasets specifically tailored to UAV imagery, underscoring a global convergence on the need for transparency.

Thus, the evolution of XAI globally has been marked by seminal surveys, government-backed initiatives, and cross-continental collaborations that set the stage for method-driven contributions. In addition, the trend reflects a growing consensus across disciplines including computer vision, human-computer interaction, and cognitive sciences that explainability is not merely a technical enhancement but fundamental requirement for trustworthy deployment of UAV-based intelligent systems across diverse operational environments worldwide.

Contributions of International Researchers in Explainability Methods

International contributions to XAI can be divided into post-hoc interpretability methods and model-intrinsic approaches, but UAV research has largely been dominated by post-hoc methods because of their compatibility with CNN-based object detectors.

A major breakthrough was made by Selvaraju et al. [21], who introduced Grad-CAM in 2017. Grad-CAM produces class activation maps by back-propagating gradients through convolutional layers, highlighting the region's most relevant to a decision. Its low computational overhead made it one of the first explanation methods practical for UAV deployment. Since its publication, Grad-CAM has been extended and widely adopted across aerial object detection studies in VisDrone and UAVDT datasets.

At roughly the same time, Ribeiro et al. [2] proposed LIME (Local Interpretable Model-Agnostic Explanations), which generates local surrogate models around a prediction by perturbing inputs. Though not computationally lightweight, LIME established the principle that interpretability can be model-agnostic, influencing UAV-based detection studies where model flexibility was necessary.

Lundberg and Lee [22] later developed SHAP (SHapley Additive exPlanations), grounded in cooperative game theory. SHAP assigns feature importance values with strong theoretical guarantees of consistency and local accuracy. While computationally expensive, SHAP became a gold standard in research seeking faithful explanations, and international UAV studies applied it primarily in offline evaluation of object detectors.

Other key international contributions include Montavon, Müller, and Lapuschkin [23], who advanced Layer-Wise Relevance Propagation (LRP), a method that attributes prediction relevance to individual neurons. Their work provided a rigorous mathematical framework that has been influential in explainability across computer vision, including UAV datasets.

More recently, Nguyen et al. [10] introduced ODExAI, an evaluation framework designed specifically for object detection explanations. ODExAI formalized three key evaluation dimensions localization accuracy, faithfulness, and efficiency and benchmarked methods such as Grad-CAM, D-RISE, and D-CLOSE. Their study reported that G-CAME achieved 96.13% localization accuracy with only 0.54s runtime, while D-CLOSE reached 0.863 faithfulness at the cost of ~71s runtime, highlighting trade-offs crucial for UAV applications.

Surveys such as those by Arrieta et al. [19] and Samek et al. [20] further consolidated international contributions, establishing a coherent body of knowledge that continues to shape XAI research globally and guide future UAV-based applications.

Advancements in Object Detection and UAV-Specific Adaptations Worldwide

Parallel to explainability, global researchers have transformed UAV capabilities by developing increasingly robust object detection architectures.

The field was revolutionized by Redmon et al. [24], who introduced the YOLO (You Only Look Once) framework. YOLOv1 demonstrated real-time detection by framing object detection as a regression problem. Later iterations, YOLOv2 and YOLOv3, also led by Redmon, improved accuracy and detection of small objects essential for UAV applications.

After Redmon's exit from computer vision research, Bochkovskiy [25] introduced YOLOv4, optimizing architecture and training strategies. YOLOv4 quickly became popular in UAV studies due to its balance of speed and accuracy. Building on this, Glenn Jocher [26] released YOLOv5 and YOLOv8 under an open-source model, enabling widespread experimentation in UAV communities worldwide. These versions offered modularity and scalability, allowing UAV researchers to adapt YOLO for aerial imagery with limited computational resources.

Alongside YOLO, other significant contributions include Lin et al. [27] with RetinaNet, which introduced focal loss to handle class imbalance in detection tasks. This was particularly impactful in UAV datasets where small targets are often underrepresented. Similarly, Liu et al. [28] proposed SSD (Single Shot MultiBox Detector), which emphasized efficiency in single-stage detection, becoming an influential baseline for UAV benchmarks.

Datasets have been another major global contribution. Zhu et al. [29] introduced the VisDrone dataset, which contains thousands of UAV-captured images and videos annotated with small objects in cluttered scenes. Xia et al. [5] developed the DOTA dataset, focusing on oriented object detection with diverse aerial perspectives. Meanwhile, the UAVDT dataset, compiled by Chinese researchers,

provided annotated UAV videos for detection and tracking in urban environments [30]. These datasets enabled international benchmarking and have been cited extensively in both detection and XAI studies.

Specialized UAV adaptations of YOLO, such as SRM-YOLO, were introduced to address UAV challenges like background clutter and scale variation. Researchers worldwide continue to refine these models with context-aware mechanisms and attention modules, bridging gaps between accuracy, explainability, and efficiency.

International Efforts in Robustness, Real-Time Deployment, and Standardization

A significant challenge in UAV AI is balancing interpretability with efficiency and robustness.

On the efficiency side, Han et al. [31] provided seminal work on model compression, including pruning, quantization, and knowledge distillation. Their contributions established the basis for deploying explainable detection models on UAV edge devices such as NVIDIA Jetson. Complementing this, Zhang et al. [32] explored FPGA-based acceleration of CNNs, which has been applied internationally to enable real-time UAV detection.

International research also focused on adversarial robustness. Goodfellow et al. [33] first revealed adversarial examples, small perturbations that drastically alter model outputs. This discovery triggered a wave of studies examining UAV vulnerability to adversarial attacks. Building on this, Xie et al. [34] advanced adversarial training methods that improve model robustness under attack. Recent UAV-specific studies demonstrated adversarial patch and camouflage attacks on aerial detectors, emphasizing the need for explainability as a defensive tool.

Standardization is another global contribution. Nguyen et al. [10] with ODExAI set the foundation for benchmarking explainability in detection, while Samek and Müller [20] emphasized the importance of standardized metrics for faithfulness and localization. Together, these works fostered cross-dataset and cross-method comparisons, advancing global consensus on evaluating explainability and ensuring that UAV XAI systems remain transparent, reliable, and adaptable across evolving real-world.

Synthesis of the International Research Landscape

The international status of XAI in UAV systems is defined by seminal contributions from individual researchers, collaborative benchmarks, and global initiatives. Researchers like Arrieta et al. [19] and Samek et al. [20] framed the field conceptually; Selvaraju et al. [21], Ribeiro et al. [2], and Lundberg & Lee [22] provided widely used explanation tools; Redmon et al. [24], Bochkovskiy [25], and Jocher [26] advanced object detection frameworks; and Zhu et al. [29], Xia et al. [5], and others developed UAV-specific datasets that have become global standards.

Efficiency and robustness contributions from Han et al. [31], Zhang et al. [32], and Xie et al. [34] highlight the international focus on real-world deployment challenges. Evaluation and standardization by Nguyen et al. [10] and Samek & Müller [20] underline the importance of benchmarking in XAI for UAVs.

Collectively, these contributions form a faithful description of the international research status, showing both progress and gaps. Challenges remain in reconciling explainability with adversarial robustness, reducing computational overhead without compromising interpretability, and ensuring standardization across datasets and contexts. However, the trajectory of international research makes clear that XAI in UAVs has matured into a globally recognized, collaborative domain shaped by pioneering researchers and programs across the United States, Europe, and Asia. Furthermore, the increasing emphasis on open-source frameworks, shared datasets, and interdisciplinary collaborations indicates that future advancements will likely emerge from global partnerships rather than isolated efforts. This reinforces the notion that explainability in UAV systems is no longer a regional research trend, but rather a universally acknowledged requirement for safe, transparent, and trustworthy autonomous aerial intelligence.

2.1.2 National Status

India is demonstrating substantial progress in the field of artificial intelligence, particularly in domains directly relevant to unmanned aerial vehicles (UAVs) and object detection. This momentum is driven by a combination of strategic government initiatives, robust academic research, and defence-focused innovation. The National Strategy for Artificial Intelligence (NITI Aayog) [42] has explicitly identified autonomous vehicle technologies, including UAVs, as both an economic opportunity and a critical enabler of national security. Importantly, the policy emphasizes not only investment in assistive AI capabilities such as image recognition and object detection but

also the development of trustworthy and explainable AI (XAI) to enhance safety, transparency, and operational efficiency. These priorities are reflected in the rise of Indian AI startups specializing in UAV-based surveillance, edge-deployed AI systems for video feed monitoring, and explainable computer vision frameworks that support real-time decision-making [53].

At the Indian Institute of Science (IISc), Bangalore, multiple research groups are emerging as leaders in UAV-oriented XAI. Prof. R. Venkatesh Babu, one of India's foremost AI researchers, has made influential contributions to computer vision, adversarial robustness, and interpretable deep learning, all central to ensuring reliable UAV object detection in adversarial environments [45]-[47]. The VISTA Lab at IISc has extended this by benchmarking UAV detection and tracking algorithms while embedding explainability. Its work on "Adding Explainability to Visual Clustering Tendency Assessment" represents one of the first explicit orientations toward XAI in UAV contexts in India [52]. Similarly, Prof. Pradipta Biswas and colleagues in the Department of Design and Manufacturing have pioneered synthetic dataset generation using diffusion models. By reducing reliance on large-scale real-world imagery, this approach enhances training diversity while also enabling explainable performance assessment in UAV data pipelines [41]. Together, these IISc initiatives highlight a deliberate movement toward embedding explainability in UAV research.

The Indian Institutes of Technology (IITs) form another cornerstone of UAV research with growing attention to explainability. At IIT Jodhpur, the DRDO Industry-Academia Centre of Excellence (DIA-CoE) has launched projects such as Futuristic Omni Mobility Drones (OMD) and AI for Information Warfare and War Gaming (AIWG). These efforts aim not only at UAV autonomy but also at explainable decision-support systems for defence operations [35], [38]. At IIT Madras, researchers have designed AI-powered counter-drone systems capable of visually detecting rogue UAVs in challenging conditions, using neural models combined with kernelized correlation filters. The explainability of these systems lies in their ability to provide interpretable visual evidence for detection and interception decisions [151]. At IIT Bombay, the Remote Sensing Laboratory has advanced UAV-based photogrammetry and camouflage detection. Its semi-automatic vehicle and human recognition systems are being extended with explainable object detection mechanisms for security and urban traffic monitoring [36]. Meanwhile, IIT Delhi has explored UAV navigation robustness through Bayesian fault detection and improved YOLOv5 models, incorporating Gabor functions and coordinate attention mechanisms. These architectural improvements enhance interpretability in small-object detection by allowing clearer attribution of

model focus regions [40]. Collectively, these IIT initiatives illustrate an academic ecosystem moving steadily toward transparent UAV autonomy and interpretable aerial object detection.

The Defence Research and Development Organisation (DRDO) remains the central driver of defence-focused UAV research in India. Through its Centres of Excellence, DRDO has strategically directed research in AI and autonomous systems (AIAS), UAV detection, and counter-drone warfare [37], [38]. Operational platforms such as Nishant for battlefield surveillance, and smaller UAVs such as Imperial Eagle and Slybird, already integrate advanced ISR functionalities. Building on this, DRDO is actively pursuing anti-drone technologies with embedded explainability, including detection and tracking systems that can provide interpretable outputs for operators during hostile UAV engagements. These developments position DRDO as both a producer of autonomous UAV platforms and an adopter of explainable AI frameworks in national defence [37], [38].

Despite these advancements, there remains a relative scarcity of systematic XAI integration in Indian UAV literature compared to international counterparts. While IISc's VISTA Lab and IIT-led projects demonstrate early steps toward embedding explainability, the majority of Indian research continues to emphasize autonomy and detection accuracy. This presents a significant opportunity for India to lead in the fusion of explainability with UAV autonomy. Leveraging its strengths in computer vision, established UAV programs, and a robust defence-academic ecosystem, India can position itself at the forefront of trustworthy UAV-AI systems. Embedding XAI principles across national datasets, edge-based UAV platforms, and tactical defence applications would not only align with India's emphasis on self-reliance but also elevate its research to global standards of transparency and accountability [13], [44], [52].

In essence, India's national status in UAV-XAI research is defined by the foresight of NITI Aayog, pioneering academic research at IISc and IITs, defence-oriented innovation under DRDO, and a dynamic startup ecosystem. Researchers such as R. Venkatesh Babu and Pradipta Biswas at IISc, academic groups at IIT Jodhpur, IIT Madras, IIT Bombay, and IIT Delhi, along with DRDO's comprehensive UAV programs, collectively form a fertile ground for integrating explainability. However, this integration remains a largely untapped frontier. Bridging this gap represents both a challenge and an opportunity: to align India's established UAV detection expertise with the global demand for trustworthy, transparent AI in UAV systems, thereby reinforcing India's pursuit of technological self-reliance and defence innovation [35], [36], [39], [37]-[40], [42]-[43], [44], [45]-[48], [50]-[51], [52].

2.1.3 Importance of the proposed project in the context of current status

The deployment of unmanned aerial vehicles (UAVs) has expanded rapidly across domains including surveillance, disaster management, infrastructure inspection, and tactical operations [42], [56], [57]. As UAVs become increasingly autonomous, the role of artificial intelligence (AI), particularly object detection and scene understanding, has grown critical. Modern UAV object detection systems rely heavily on deep learning frameworks, such as YOLOv5, which deliver high detection accuracy but often operate as "black boxes," providing little insight into their decision-making processes [55], [58], [59]. This opacity limits trust in AI, particularly in high-stakes operational contexts where human operators must interpret or validate autonomous decisions.

Explainable AI (XAI) has emerged as a promising solution to this challenge by providing interpretable outputs such as saliency maps, feature attribution scores, and contribution visualizations [54], [28], [10]. These explanations enable operators to understand which aspects of sensor input influence detection outcomes, facilitating trust, error diagnosis, and model refinement [55], [56]. Despite the increasing adoption of XAI in computer vision, its integration into UAV systems remains limited due to the stringent computational and energy constraints of airborne platforms [10], [58]. Real-time object detection on a battery-powered UAV requires balancing model complexity and inference speed, a challenge compounded when incorporating explainability modules. Traditional approaches often prioritize either performance or interpretability, failing to achieve both simultaneously [55], [57].

Our proposed project addresses this gap by combining explainability-guided model optimization with edge AI hardware acceleration. By using pruning frameworks, we identify low-impact network parameters and prune them, resulting in lightweight yet interpretable models suitable for real-time UAV deployment [10], [56], [57]. This approach directly addresses the pressing need for AI systems that are both trustworthy and operationally feasible in resource-constrained aerial platforms. Furthermore, deploying optimized models on FPGAs or Jetson edge modules ensures deterministic low-latency inference, maintaining real-time responsiveness while minimizing energy consumption [58], [59].

Beyond performance, the integration of XAI enhances robustness and security. UAVs operating in contested or complex environments are vulnerable to adversarial attacks, such as physical camouflage patches or input perturbations, which can mislead standard object detectors [61], [37]. By providing interpretable outputs that highlight regions of interest or suspicious input

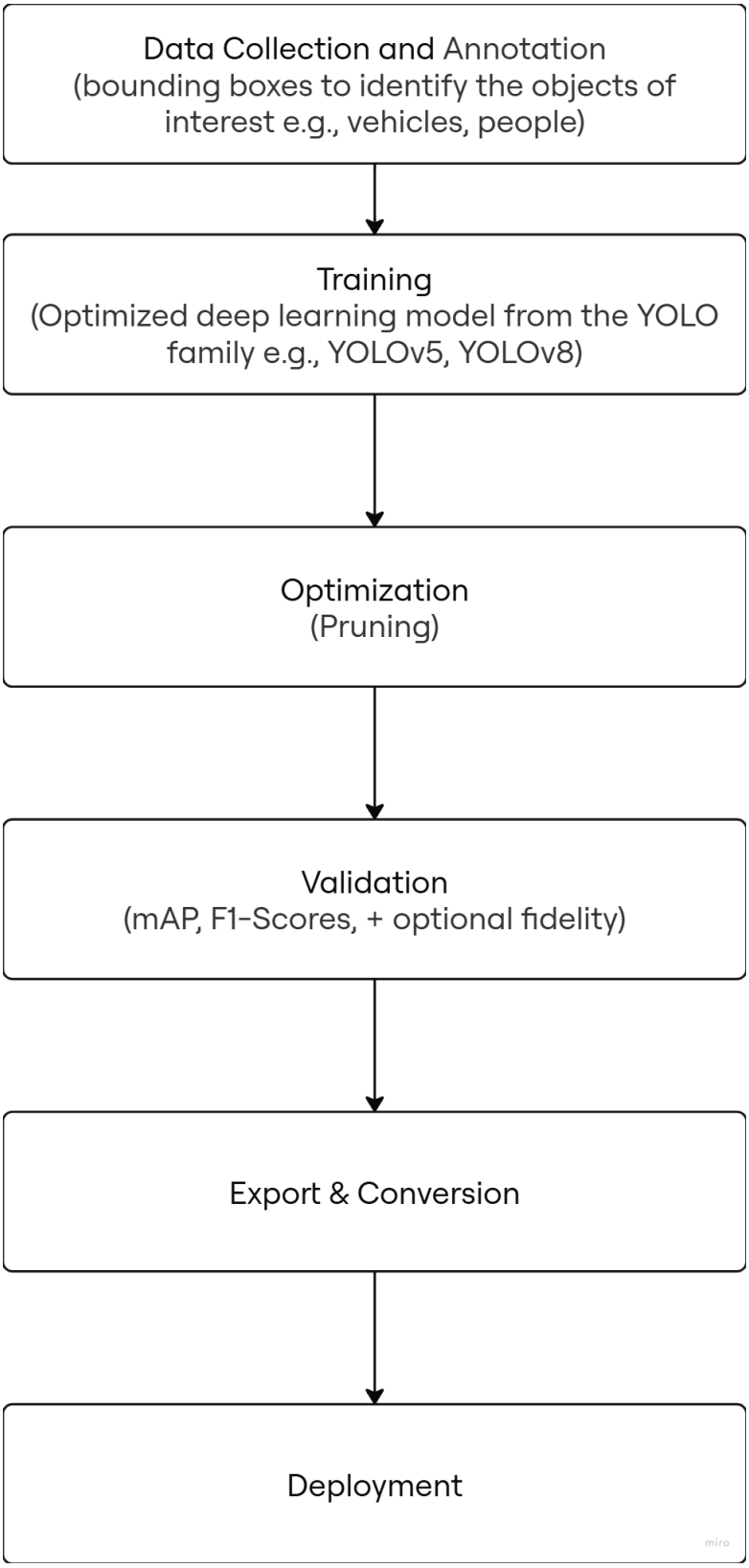
patterns, our framework allows human operators to detect potential threats proactively, thereby reinforcing mission reliability and operational safety [13], [62].

In the broader context, the proposed project aligns with national and global priorities for autonomous systems and AI. Strategies such as India's National Strategy for Artificial Intelligence explicitly emphasize the development of trustworthy AI for defense, surveillance, and autonomous vehicle technologies [42]. Moreover, contemporary research in UAV object detection and XAI highlights the urgent need for frameworks that integrate performance, explainability, and security for practical deployment [54], [28], [60], [13]. By simultaneously addressing these dimensions, our project not only improves detection accuracy and operational safety but also elevates the role of AI from a passive tool to an intelligent teammate capable of reasoning, explaining, and alerting operators about potential vulnerabilities [10], [55], [56], [61], [63].

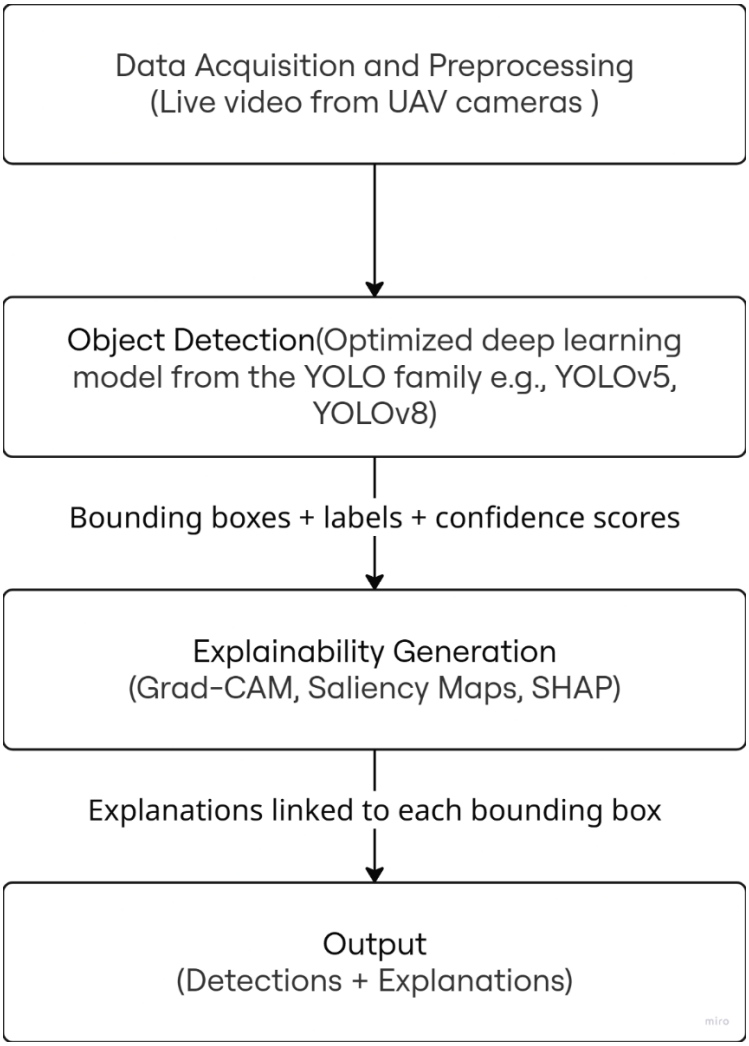
Therefore, the proposed framework represents a significant advance over current UAV AI implementations. It addresses critical gaps in transparency, computational efficiency, and robustness, providing a holistic solution that meets the evolving demands of real-time UAV operations in both civilian and tactical applications [54], [28], [58], [62]. Its successful implementation would set a benchmark for explainable, high-performance, and secure AI systems in airborne platforms, offering a model for future autonomous systems research and deployment [54], [28], [58], [62].

Diagrams

1. Training



2. Deployment



REFERENCES

[1] "Explainable Artificial Intelligence | DARPA," *Darpa.mil*, 2018. https://www.darpa.mil/research/programs/explainable-artificial-intelligence.

[2] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?': Explaining the Predictions of Any Classifier," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, pp. 1135-1144, Aug. 2016, doi: https://doi.org/10.1145/2939672.2939778.

[3] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336-359, Feb. 2020, doi: https://doi.org/10.1007/s11263-019-01228-7.

[4] D. Gunning, E. Vorm, J. Y. Wang, and M. Turek, "DARPA 's explainable AI (XAI) program: A retrospective," *Applied AI Letters*, vol. 2, no. 4, Dec. 2021, doi: https://doi.org/10.1002/ail2.61.

[5] G.-S. Xia *et al.*, "DOTA: A Large-Scale Dataset for Object Detection in Aerial Images," *Thecvf.com*, pp. 3974-3983, 2018, Available: https://openaccess.thecvf.com/content_cvpr_2018/html/Xia_DOTA_A_Large-Scale_CVPR_2018_paper.html

[6] D. Gunning and D. Aha, "DARPA's Explainable Artificial Intelligence (XAI) Program," *AI Magazine*, vol. 40, no. 2, pp. 44-58, Jun. 2019, doi: https://doi.org/10.1609/aimag.v40i2.2850.

[7] microsoft, "microsoft/AirSim," *GitHub*, Aug. 22, 2019. https://github.com/Microsoft/AirSim

[8] M. Hussain, "YOLOv5, YOLOv8 and YOLOv10: The Go-To Detectors for Real-time Vision," *arXiv.org*, Jul. 03, 2024. https://arxiv.org/abs/2407.02988

[9] Sadia Nazim, M. M. Alam, S. S. Rizvi, J. C. Mustapha, S. S. Hussain, and Mazliham Mohd Suud, "Advancing malware imagery classification with explainable deep learning: A state-of-the-art approach using SHAP, LIME and Grad-CAM," *PLoS ONE*, vol. 20, no. 5, pp. e0318542-e0318542, May 2025, doi: https://doi.org/10.1371/journal.pone.0318542.

[10] Nguyen, H. Truong, and H. Cao, "ODExAI: A Comprehensive Object Detection Explainable AI Evaluation," *arXiv.org*, 2025. https://arxiv.org/abs/2504.19249

[11] V. Hassija *et al.*, "Interpreting Black-Box Models: a Review on Explainable Artificial Intelligence," *Cognitive Computation*, vol. 16, no. 1, pp. 45-74, Aug. 2023, doi: https://doi.org/10.1007/s12559-023-10179-8.

[12] S. Zhuang, Y. Hou, and D. Wang, "Towards Efficient Object Detection in Large-Scale UAV Aerial Imagery via Multi-Task Classification," *Drones*, vol. 9, no. 1, pp. 29-29, Jan. 2025, doi: https://doi.org/10.3390/drones9010029.

[13] S. Javaid, M. A. Khan, H. Fahim, B. He, and N. Saeed, "Explainable AI and monocular vision for enhanced UAV navigation in smart cities: prospects and challenges," *Frontiers in Sustainable Cities*, vol. 7, Mar. 2025, doi: https://doi.org/10.3389/frsc.2025.1561404.

[14] J. Black, "Sharpen Your Edge AI and Robotics Skills with the NVIDIA Jetson Nano Developer Kit," *NVIDIA Technical Blog*, Dec. 05, 2022. https://developer.nvidia.com/blog/sharpen-your-edge-ai-and-robotics-skills-with-the-nvidia-jetson-nano-developer-kit

[15] Spheron Network, "Ultimate Guide to the NVIDIA RTX A6000: Key Features and Specs," *Spheron's Blog*, Jun. 23, 2024. https://blog.spheron.network/ultimate-guide-to-the-nvidia-rtx-a6000-key-features-and-specs

[16] "Combining RGB and Thermal Imaging for UAV Surveillance: Enhancing Detection in Complex Environments," *Computer Vision Embedded*, Oct. 21, 2024.

https://cvembedded.com/2024/10/21/combining-rgb-and-thermal-imaging-for-uav-surveillance-enhancing-detection-in-complex-environments/.
[17] Y. Yang, T. Yang, X. Wu, Z. Guo, and B. Hu, "Efficient UAV Swarm-Based Multi-Task Federated Learning with Dynamic Task Knowledge Sharing," *arXiv.org*, 2025. https://arxiv.org/abs/2503.09144 (accessed Aug. 19, 2025).
[18] Y. Qu *et al.*, "Decentralized Federated Learning for UAV Networks: Architecture, Challenges, and Opportunities," *arXiv.org*, 2021. https://arxiv.org/abs/2104.07557 (accessed Aug. 19, 2025).
[19] A. B. Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, Opportunities and Challenges toward Responsible AI," *Information Fusion*, vol. 58, no. 1, pp. 82-115, Jun. 2020, doi: https://doi.org/10.1016/j.inffus.2019.12.012.
[20] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K.-R. Muller, "Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications," *Proceedings of the IEEE*, vol. 109, no. 3, pp. 247-278, Mar. 2021, doi: https://doi.org/10.1109/jproc.2021.3060483.
[21] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 618-626, Oct. 2017, doi: https://doi.org/10.1109/iccv.2017.74.
[22] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *arXiv:1705.07874 [cs, stat]*, Nov. 2017, Available: https://arxiv.org/abs/1705.07874
[23] G. Montavon, S. Lapuschkin, A. Binder, W. Samek, and K.-R. Müller, "Explaining nonlinear classification decisions with deep Taylor decomposition," *Pattern Recognition*, vol. 65, pp. 211-222, May 2017, doi: https://doi.org/10.1016/j.patcog.2016.11.008.
[24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788, 2016, doi: https://doi.org/10.1109/cvpr.2016.91.
[25] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv*, vol. 1, Apr. 2020, Available: https://arxiv.org/abs/2004.10934
[26] G. Jocher, "ultralytics/yolov5," *GitHub*, Aug. 21, 2020. https://github.com/ultralytics/yolov5
[27] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, doi: https://doi.org/10.1109/iccv.2017.324.
[28] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," *Computer Vision - ECCV 2016*, vol. 9905, no. 5, pp. 21-37, 2016, doi: https://doi.org/10.1007/978-3-319-46448-0_2.
[29] "Detection and Tracking Meet Drones Challenge | IEEE Journals & Magazine | IEEE Xplore," *ieeexplore.ieee.org*. https://ieeexplore.ieee.org/document/9573394
[30] D. Du *et al.*, "The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking," *arxiv.org*, Mar. 2018, Available: https://arxiv.org/abs/1804.00518
[31] S. Han, H. Mao, and W. J. Dally, "Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding," *arXiv:1510.00149 [cs]*, Feb. 2016, Available: https://arxiv.org/abs/1510.00149
[32] C. Zhang, P. Li, G. Sun, Y. Guan, B. Xiao, and J. Cong, "Optimizing FPGA-based Accelerator Design for Deep Convolutional Neural Networks," *Proceedings of the 2015 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays - FPGA '15*, 2015, doi: https://doi.org/10.1145/2684746.2689060.
[33] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and Harnessing Adversarial Examples," *arXiv.org*, 2014. https://arxiv.org/abs/1412.6572
[34] C. Xie, J. Wang, Z. Zhang, Y. Zhou, L. Xie, and A. Yuille, "Adversarial Examples for Semantic Segmentation and Object Detection," *arXiv.org*, 2017.https://arxiv.org/abs/1703.08603
[35] "Artificial Intelligence for Information and War Gaming Technologies (AIWT) | DRDO Industry Academia Centre of Excellence IIT Jodhpur (DIA-CoE) | IIT Jodhpur," Iitj.ac.in, 2025. https://iitj.ac.in/dia-coe/en/aiwt (accessed Aug. 20, 2025).

[36] "Un-manned airborne systems (UAS) for remote sensing and photogrammetry | IITBombay," Iitb.ac.in, 2025. https://rnd.iitb.ac.in/node/1274 (accessed Aug. 20, 2025).
[37] "Appl.AI Labs," Appl.ai, 2025. https://appl.ai/projects/adversarialai (accessed Aug. 20, 2025).
[38] "DRDO Industry Academia Centres of Excellence | Defence Research and Development Organisation - DRDO, Ministry of Defence, Government of India," Drdo.gov.in, 2025. https://drdo.gov.in/drdo/adv-tech-center
[39] "Thrust Area DIA-CoE, IIT Jodhpur | Defence Research and Development Organisation - DRDO, Ministry of Defence, Government of India," Drdo.gov.in, 2025. https://www.drdo.gov.in/drdo/thrust-area12 (accessed Aug. 20, 2025).
[40] "Projects - CAPS Lab - IIT Delhi," Iitd.ac.in, 2025. https://web.iitd.ac.in/~kodamana/Projects.html (accessed Aug. 20, 2025).
[41] C. Kumar, "New AI model with synthetic images boosts mixed reality object detection: IISc," The Times of India, Apr. 09, 2025. https://timesofindia.indiatimes.com/india/new-ai-model-with-synthetic-images-boosts-mixed-reality-object-detection-iisc/articleshow/120122125.cms (accessed Aug. 20, 2025).
[42] National Strategy for Artificial Intelligence - NITI Aayog, accessed on August 15, 2025, https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf
[43] EO IR Solutions for UAV Surveillance | Controp Precision Technologies, accessed on August 15, 2025, https://www.controp.com/worlds/uav/
[44] A. Mukhopadhyay, H. Br, P. T. Gaikwad, I. Mukherjee, and P. Biswas, "I-rod: an ensemble of CNNs for object detection in unconstrained road scenarios," Signal Image and Video Processing, vol. 19, no. 1, Nov. 2024, doi: https://doi.org/10.1007/s11760-024-03590-7.
[45] Y. Name, "R. Venkatesh Babu's Homepage," Iisc.ac.in, 2020. https://cds.iisc.ac.in/faculty/venky/ (accessed Aug. 20, 2025).
[46] Md Habibur Rahman, M. Abrar, Md Abdul Aziz, R. Tabassum, J.-I. Baik, and H.-K. Song, "A Comprehensive Survey of Unmanned Aerial Vehicles Detection and Classification Using Machine Learning Approach: Challenges, Solutions, and Future Directions," Remote sensing (Basel), vol. 16, no. 5, pp. 879–879, Mar. 2024, doi: https://doi.org/10.3390/rs16050879.
[47] "R. BABU | Professor | PhD | Indian Institute of Science Bangalore, Bengaluru | IISC | Department of Computational and Data Sciences | Research profile," ResearchGate, 2016. https://www.researchgate.net/profile/R-Babu-6 (accessed Aug. 20, 2025).
[48] "Publications – AI @ IISc," Iisc.ac.in, 2025. https://ai.iisc.ac.in/publications/ (accessed Aug. 20, 2025).
[49] I. Today, "IIT Madras designs AI drones for armed forces to counter and hack 'rogue drones,'" India Today, Mar. 05, 2020. https://www.indiatoday.in/education-today/news/story/iit-madras-designs-ai-drones-for-armed-forces-to-counter-and-hack-rogue-drones-1652813-2020-03-05 (accessed Aug. 20, 2025).
[50] "ModalAI, Inc.," ModalAI, Inc., 2025. https://www.modalai.com/pages/ros-drone (accessed Aug. 20, 2025).
[51] "Ongoing Research Projects in The Institute | Council of Indian Institute of Technology," Iitsystem.ac.in, 2019. https://www.iitsystem.ac.in/mhrdprojects
[52] "VISTA LAB," VISTALABIISC, 2024. https://www.vistalabiisc.com/ (accessed Aug. 20, 2025).
[53] Nidhi Umashankar and K. Sai, "A Comprehensive Study of Artificial Intelligence Applications of Drone," Dec. 2024, doi: https://doi.org/10.31224/4194.
[54] M. Mersha, K. Lam, J. Wood, A. K. AlShami, and J. Kalita, "Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction," Neurocomputing, vol. 599, p. 128111, Sep. 2024, doi: https://doi.org/10.1016/j.neucom.2024.128111.

[55] H. Zhang, F. Shao, X. He, Z. Zhang, Y. Cai, and S. Bi, “Research on Object Detection and Recognition Method for UAV Aerial Images Based on Improved YOLOv5,” Drones, vol. 7, no. 6, p. 402, Jun. 2023, doi: https://doi.org/10.3390/drones7060402.
[56] A. Jain et al., “AI-Enabled Object Detection in UAVs: Challenges, Design Choices, and Research Directions,” IEEE Network, vol. 35, no. 4, pp. 129–135, Jul. 2021, doi: https://doi.org/10.1109/mnet.011.2000643.
[57] B. Yao et al., “SRM-YOLO for Small Object Detection in Remote Sensing Images,” Remote Sensing, vol. 17, no. 12, p. 2099, Jun. 2025, doi: https://doi.org/10.3390/rs17122099.
[58] "A Survey on Real-Time Object Detection on FPGAs," Accessed: Aug. 20, 2025. [Online]. Available: https://www3.diism.unisi.it/~giorgi/papers/Hozhabr25-ia.pdf
[59] NVIDIA, “NVIDIA Jetson AGX Orin,” NVIDIA. https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-orin/
[60] Y. Kniazieva, “Object Detection Evaluation Metric Explained,” labelyourdata.com, Dec. 14, 2023. https://labelyourdata.com/articles/object-detection-metrics
[61] Y. Zhang, J. Qi, K. Bin, H. Wen, X. Tong, and P. Zhong, “Adversarial Patch Attack on Multi-Scale Object Detection for UAV Remote Sensing Images,” 2022, doi: https://doi.org/10.3390/rs14215298.
[62] “Drones AI Software | Artificial Intelligence Software for UAV & UAS,” Unmanned Systems Technology, Oct. 20, 2023. https://www.unmannedsystemstechnology.com/expo/drone-ai-software/
[63] P. J. Phillips et al., “Four Principles of Explainable Artificial Intelligence,” Four Principles of Explainable Artificial Intelligence, Sep. 2021, doi: https://doi.org/10.6028/nist.ir.8312.