Journal of Applied Automation Technologies

Research Paper

Modified Faster R-CNN with Adaptive Feature Extraction for Fully Visual Autonomous UAV Power Line Patrol

Liliana Kaliszowa[1,*] and Sabina Franciszka Gola[1]

[1] Faculty of Electrical and Automatic Control Engineering, Czestochowa University of Technology, Czestochowa, 42-200, Poland

*Corresponding author: liiana.ka@pcz.edu.pl

Abstract. Inspection and maintenance of contemporary electricity distribution networks are currently problematic; in particular, complexity and huge scale make the task more challenging. Currently, the requirements for speed, safety, and range cannot be reached by traditional manual inspection techniques. To address issues including visual confusion, component variety, and limits in real-time operation of the current systems, an all-UAV-based automatic power line inspection system is being introduced. In settings with poor signal quality and noisy backdrops, the Adaptive Object Detection Module uses a modified version of the faster-RCNN algorithm to detect different power-grid components. created a closed-loop control system that precisely issued UAV flight commands for steady and accurate hovering close to inspection objects by integrating the output of high-frequency vision. Advanced augmentation for detection in conjunction with dynamic models and feedback integration for control; multi-scale feature extraction in conjunction with an attention mechanism. The average detection accuracy reaches about 84.5%, the hovering error is within ±10cm, and the control system lag-time is below millisecond levels, according to various testing results based on simulated real-world data. Furthermore, pooled data also demonstrated improved efficacy and safety in practice, according to the aforementioned result. Modular engineering design is easily adaptable to more complicated grid topology situations or enlarged for multi-iUAV cooperation. An approach to the intelligence of vision-based airborne inspection of critical infrastructure that is both practical and scalable is presented in this paper.

Keywords: *Intelligent Perception, R-CNN, UAV Control, Power Line Inspection*

Received on 14 October 2023, Accepted on 21 March 2024, Published on 27 March 2024

Introduction

In order to operate and maintain the current distribution network segments, reliable and stable procedures for inspecting the power-line infrastructure facilities are required. Conventional manual inspection methods will not be able to keep up with the rapid development trend in terms of Speed, Safety, and Range Coverage as transmission networks continue to grow in size and strive toward creating a Smart Grid [1,2]. Because of their adaptability in terms of deployment time, speed, and size, unmanned aerial vehicles, or UAVs, have emerged as the primary method of transformation [3,4]. By using a UAV's high-definition camera system to increase its flight range, it may also improve equipment inspection processes in grids and simultaneously expand observational space both horizontally and vertically over time and space [5,6].

Despite the enormous potential of technology, there are still certain obstacles in the UVS inspection process. Unstable lighting, background clutter, and occlusion from trees and other man-made obstacles are just a few of the visual quality issues that frequently arise in outdoor environments; When they fly, power-line components clearly differ from other objects in terms of location, shape, and scale [7]. The integrated processors of UAVs used in industrial application situations have very limited processing resources [8]. Object detection and flight control should work in tandem in a tightly integrated environment with minimal latency and high reliability to

ensure operational safety [9,10]. Motions must be rapidly informed by visible visualisation of outcomes, particularly when precisely hovering at inspection locations such insulator joints and cable terminations [11,12]. Therefore, a fundamental issue in the study of automatic aerial inspection is the reliable, edge-oriented integration of perception, control, and embedded computation [13,14].

This research presents a unified UAV-based inspection system designed for real-world applications in a variety of power line environments in order to address the aforementioned issues. The Adaptive Object Detection Module is an enhanced R-CNN network that uses domain-specific modification techniques to recognise power infrastructure components with extreme precision. Second, in order to achieve precise and dependable close-range detection, we develop a hover-control algorithm for the vehicles that incorporates real-time detection data into unmanned avionics vehicle (UAV) operation instructions. positioning in relation to the intended objects; Thirdly, in order to fulfil the demands of being deployed on-site with constrained resources, every module has been implemented in a highly integrated manner. It is clear from a wide range of outside tests for the design that the built system has accurate control performance and excellent detection accuracy in a variety of working environments. This paper is organised as follows: The entire system architecture and design requirements are presented in the second section, followed by an explanation of an adaptive detection algorithm in section three, a detailed visual-based integrated hovering control in section four, a report on quantitative experimental validation and comparison in section five, and present conclusions and future directions in section six.

System Framework and Architecture

Task Objectives and Operational Requirements

Due to the wide range of missions and variations in the operational environment, power line inspection is currently supported by UAVs for large-scale complicated transmission lines. Currently, insulators, conductors, connections, and related attachments are among the several types of equipment that may be detected quickly and precisely on this type of line using an overall automation system. to accomplish activities on one's own in a dynamic setting with high electromagnetic interference, varying lighting, wind disruption, plant impediment, etc [15,16]. A fully autonomous UAV inspection system must fulfil very high safety standards because it will be more risky due to the absence of traditional human control.

The benchmarks' test results serve as the foundation for the system's performance evaluation. In order to ascertain whether a particular level of standard can be satisfied in terms of systems' recognition performance, evaluate mAP by combining several comparisons of precision/recall curves among different target categories. If the power-line-inspection-accuracy criteria is not met, there will be too many false alarms in repair plans and outage forecast mistakes as a result of incorrect decisions made during inspection activities. The root-mean-square error (RMSE) of displacement with respect to the target point on the platform during hovering flight time can also be used to express hovering precision. It represents both the inspection cargo's short-term adaptation change performance and its long-term recognition accuracy. Ultimately, a strict upper limit requirement, such as fewer than 200 milliseconds, should be met within this time frame for the whole picture capture process, from acquisition to target detection to finally controlling the unmanned aerial vehicle for action. When a single component variation is the only cause of problems or risk issues under this working state that meets all requirements, they must be addressed right away.

based on the additional demands brought about by new requirements, such as more stringent regulations for asset management accuracy in aberrant diagnosis using quantitative analysis, visual information, and document traceability. Drone-related equipment failures, overlapping operating periods with other operations, and timely flight path updates are all important for UAVs. Because of these factors, it is intrinsically a highly multidisciplinary design effort that must be accomplished by integrating computer vision, embedded real-time systems, and strong control theory into a concrete, integrated engineering-based system following field experience. When taken as a whole, they set the comparatively high technical goals and system-resilience benchmark values that motivate the study described in this paper.

https://doi.org/10.64972/jaat.2024v2.255p11e:149-162

UAV Inspection System Architecture

The design of a moduleized architecture will enable the seamless integration of various devices and algorithms in a setting. The stabilised RGB camera, which gathers extremely high-quality aerial pictures and is outfitted with IMUs for stable attitude and motion data retrieval, is actually its central component. Direct transmission of raw sensor data streams via the onboard processing unit embedded in a GPU or other high-speed processor intended for real-time visualisation and control functions; on this hardware, the detection algorithm (enhanced Faster R-CNN backbone) is run to produce bounding boxes and confidence scores per frame; additionally, spatially detailed feature vectors customised to the visual characteristics of power-line components are generated. For subsystem-level intelligent fusion and control, a data-integration mechanism is in charge, respectively. The centroids of the identified objects are transformed into inertial or Earth-referenced coordinates taking into account the scene geometry, UAV attitude, and camera posture. a way to determine the precise location inaccuracy and use that information to modify the trajectory. In order to apply a feedback-control strategy and modify the actuators' command value in real time, the control layer will process these signal inputs. A closed-loop vision-guidance system that can concurrently perform precise hovering adjustment around sensitive objects and coarse navigation to approach the inspection position.

The drone's own flight computer receives all computed control commands, modulates motor power output, and eventually realises a certain flight condition. Every unit in this chain has a seamless communication relationship, thus any issue in one area won't have an impact on the system as a whole. Each building component has a distinct signal flow direction, interface point, and control hierarchy, as seen in Figure 1, allowing for future flexible extension or inclusion of additional functions [18].

It is also scalable to multi-UAV collaborative inspection environments using defined communication standards and data exchange formats, providing further support for including other sensor modes (such thermal vision and LiDAR) into this design. It can swiftly transition to various unmanned aerial vehicles and operating situations without substantially altering the code by concentrating on hardware abstraction and failure-free routes. Based on both current research findings and field engineering theories, this future-proof, extensible inspection system can accomplish the most recent detection technology and control level while maintaining good Maintainable Operation Safety in long-term missions under harsh environmental conditions.

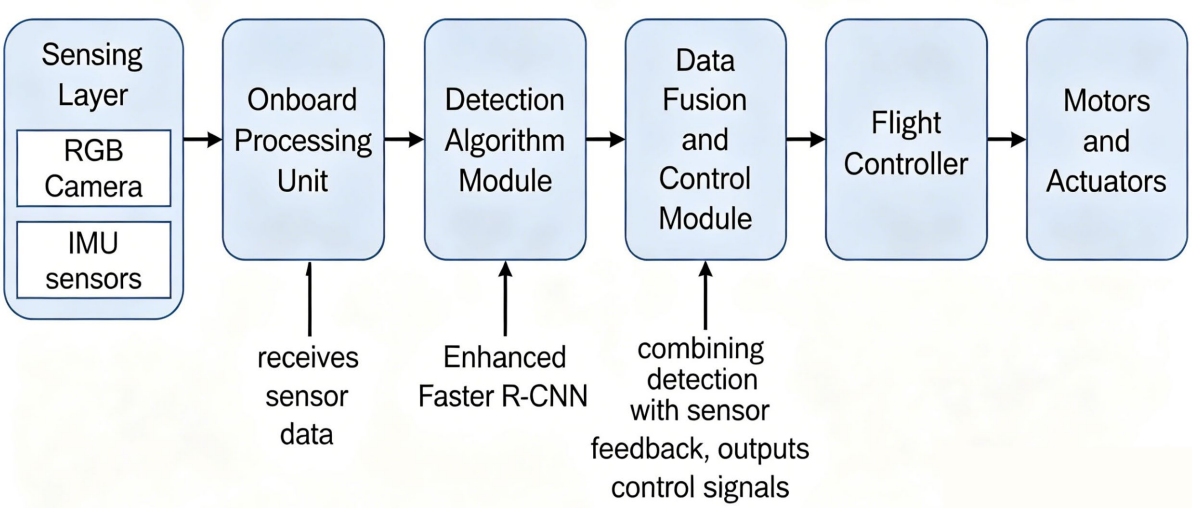

Figure 1. Block Diagram of UAV Power Line Inspection System Architecture.

End-to-End Workflow

The unmanned aerial vehicle (UAV) inspection system is characterised by an uninterrupted feedforward data processing and control flow. Flow of the process: Keep taking pictures inside the unmanned aerial vehicle's (UAV) assigned grid area. Before each image is sent to the detection network, it is immediately detected in order to forecast its classes and the spatial locations that correspond to these classes. instantly transformed from camera-space to a global coordinate system using an inbuilt fusion algorithm that merges visual information with IMU output.

The control engine determines the path or hovering regulation needed to maintain an ideal distance and position from the target when objects of interest have been resolved in the proper coordinate systems. The current condition of the unmanned aircraft and its surroundings are used to generate these instructions. The autopilot will next carry out these directives to finish the activation loop.

Update the sensor's data often for each inspection period in order to create new recognitions that continue to correct deviation in accordance with changes in the target, object, surroundings, environment, and missions. For the detection-guided tasking and control of UAVs, there is a complete workflow that covers all necessary dependency ties and responding time sequences, as seen in Figure 2. It is necessary to realise a pipeline implementation that satisfies the demands of high responsiveness and dependability in industrial-grade airborne inspections.

Dynamic retasking, event-triggered reinspection, adaptive-path-replanning, and other advanced workflow enhancements compensate for the shortcomings of the original system through improved functionality. High-efficiency data collecting can be accomplished using distributed computing, and more sophisticated autonomous features like self-diagnosis, mission interruption from anomaly detection, and seamless communication between the system and the grid network's backend management layer can be added. In this closed-loop scheduling, perception, decision-making, and action are closely linked, giving the system the stability, adaptability, and security required for deployment on critical power equipment.

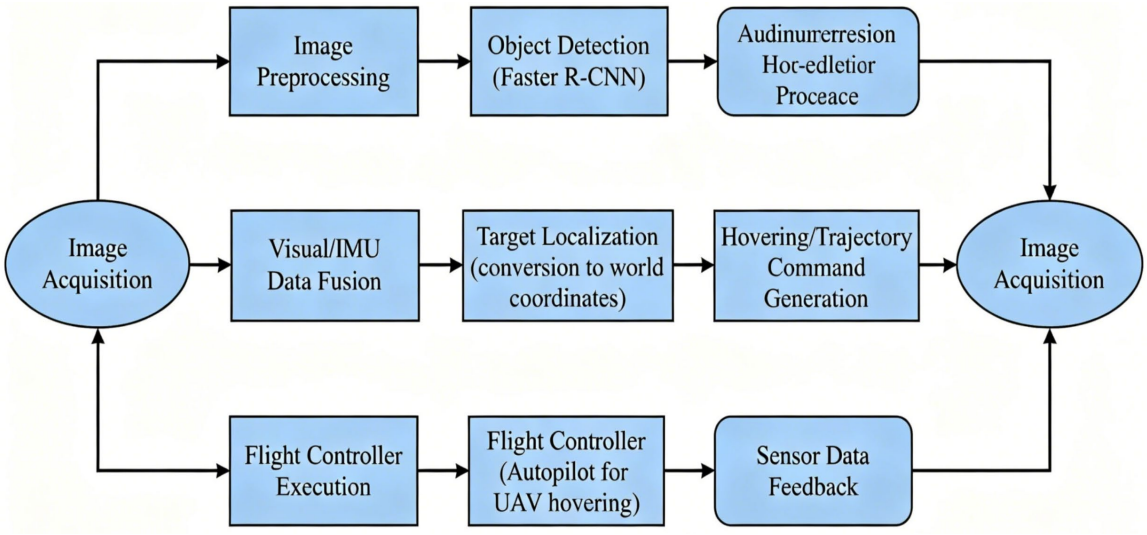

Figure 2. Process Flowchart of Detection-Guided UAV Hovering Execution

This architecture not only supports rigorous technical performance but also ensures that the system remains extensible and resilient as grid inspection scenarios and operational requirements continue to evolve.

Adaptive Detection Algorithm Design

Domain Challenges and Model Adaptation Strategy

Aerial image-based automatic power transmission component recognition is hampered by things like the intrinsic complexity of the target and changes in the surrounding environment. The issues require unique design for generalisation in the detection system architecture because they are highly field-dependent. First of all, the target objects—such as huge conductors and transmission towers—are comparatively large in size compared to small nuts and bolts. When detecting small or somewhat sparse objects, traditional deep-learning-based detectors are vulnerable to failure due to a scale difference issue caused by this fluctuation. Improve the detection section's multi-scale feature-extraction capacity in this way.

Changes in the environment compound this issue once again. It is frequently challenging to discern between features in photos because of extreme light variations brought on by weather, tower or plant shadows, and extremely direct illumination from above. Furthermore, there are many situations with backdrops that are

packed with trees, like fields or buildings, which greatly impact recognition results and raise the risk of false alarms [21]. It won't be very smooth ahead, but in many cases, trees, overlapping pictures of electricity lines, and collected trash like bird nests partially block the goal. The detection system must have a high level of spatial context awareness and be able to accurately assess certain damaged or partial inputs. has a high degree of geometric repetition as well; for instance, detectors lacking strong discriminative ability and context-aware thinking may be misled by the uniform spacing of insulators or identical structures in tower groups. Additionally, because of their tiny size or ambiguous boundary forms, these target objects may have different annotations, which introduces noise during training and impacts the generalisation of the model. This work has been constructed in three main approaches of modification in response to the challenging issues indicated above. First, pick an optimisation target with built-in multiscale feature learning capabilities across different backbones. To improve detection performance over a range of items with varying sizes, Feature Pyramid Networks (FPN) are employed. Second, in order to identify the key elements and prevent occlusions, spatial and channel attention modules are included. Lastly, loss-balancing techniques and high precision data augmentation will be applied to enhance the model's flexibility in identifying diverse backgrounds with distinct labels. Together, these domain-driven modifications provide the foundation of a robust, high-fidelity, real-time Power Line Defect Detection System [22, 23].

Algorithm Structure and Mathematical Analysis

The proposed detection algorithm is designed by augmenting the classic two-stage detection pipeline, optimizing both its representational power and inference efficiency for power grid environments. The network backbone integrates a multi-scale feature extraction module inspired by Feature Pyramid Networks (FPN), providing a hierarchical architecture where lowlevel spatial details are fused with high-level semantic representations. Formally, given an input image I , the backbone produces a set of feature maps $\{F_l\}, l = 1,\ldots,L$, where each layer F_l corresponds to features at a different scale. These are combined via lateral and top-down connections, enabling the detector to localize both minute insulator defects and extended conductor structures.

A region proposal network (RPN) is employed, generating candidate object anchors with adaptive aspect ratios tailored to the geometric characteristics of power line equipment. For each anchor, objectness scores and bounding box refinements are computed. Let a denote an anchor and θ the network parameters; the RPN loss function is defined as:

$$L_{\text{RPN}} = \frac{1}{N_{\text{cls}}} \sum_i \text{CE}\left(p_i, p_i^*\right) + \lambda \frac{1}{N_{\text{reg}}} \sum_i p_i^* \text{smooth}_{L1}\left(t_i, t_i^*\right) \quad \text{Eq. (1)}$$

where p_i and p_i^* are the predicted and ground-truth objectness, t_i and t_i^* are the predicted and target box offsets, and λ balances the two components [21,22].

To address the domain's frequent occlusion and background clutter, an attention-enhanced refinement head is included. This module fuses spatial and channel-wise attention using learnable masks, suppressing irrelevant features while boosting the saliency of partially occluded or ambiguous regions. The attention weights $A = f_{att}(F)$ are generated via a compact convolutional subnetwork and applied as $F' = A \odot F$, where $\odot$ denotes elementwise multiplication.

Loss balancing techniques are further introduced to handle the extreme foregroundbackground class imbalance and the wide dynamic range in target object sizes. In particular, the focal loss is adopted for proposal classification:

$$L_{\text{cls}} = -\alpha\left(1 - p_t\right)^{\gamma} \log\left(p_t\right) \quad \text{Eq. (2)}$$

where p_t is the predicted probability for the true class, and α, γ modulate the impact of difficult and easy samples [23].

In order to achieve real-time execution on UAV hardware, the model design is matched to the computational budget available on edge processors. Residual and bottleneck structures are utilized to maintain

representational depth while minimizing parameter redundancy. All convolutions are quantized and inference is pipelined asynchronously with image acquisition and flight control modules, ensuring that end-to-end system latency meets strict operational requirements.

Each module normalises to obtain the regular geometric and visual features of power grid scenes that are consistent under spatial differences caused by changing light source positions or object occlusion during data acquisition. In this way, the entire structure is optimised end-to-end for proposal regions, attention weights, and final boxes. To put it briefly, the improved detection technique combines mathematically based Attention and Loss design with domain-adaptation for feature extraction. It will be possible to increase the accuracy of target identification in power grid aerial inspections and guarantee consistent performance even in the face of operational limitations and visual impairments.

Dataset Development and Training Details

Fundamentally, the quality of the base-data-set determines the power-grid-inspection-observation system's performance and capacity for generalisation. A collection of totalised photos with a variety of settings, target orientations, and shooting techniques utilised in real-world applications was acquired. UAVs were used to gather data, and they flew at various altitudes; Additionally, because of shifting lighting conditions during the day and at night, as well as complicated surroundings like trees that partially block the view or a background that is filled with other items.

Each image was annotated twice in turn to guarantee the accuracy of the ground-truth labels. At first, seasoned technical staff created the bounding box labels of important components based on their expertise. First, the judge's confirmation; second, any remaining doubts were cleared up by discussion among the pertinent parties. By using cooperative labelling, errors caused by false positives or inconsistent bounding box alignment can be minimised. To guarantee data consistency and make re-creation easier, a self-written script also automatically checked the label format and eliminated outliers.

A model's capacity to generalise to domain perturbations can be enhanced via data augmentation. The techniques used include target omission construction, geometric distortion, brightness contrast adjustment, random horizontal flipping, and multi-scale resize. To prevent class imbalance and minimise overfitting, each augmentation is adjusted parametrically to improve the model's capacity for generalisation in previously unexplored domains.

A cosine learning-rate schedule and a stochastic gradient-descent optimiser are used in the training protocol. Additionally, a size-biased batch composition was incorporated to approximate the proportions of various-sized objects in the dataset. Mean Average Precision (mAP) was the primary criterion for model convergence monitoring of a reserved validation set. To avoid overfitting, early halting and periodic validation were used. The experiment was repeated several times using various random seeds to confirm the consistency of the findings. Ultimately, a solid theoretical foundation for the creation of high-fidelity flaws in distributed systems has been established by this set of data and training procedure.

Vision-Guided UAV Hovering Control Design

UAV Dynamics Modeling and Controller Design

The UAV hovering stability and trajectory tracking capability fundamentally rest on a rigorous understanding of the vehicle's underlying kinematic and dynamic behavior. In quadrotor platforms-the prevalent choice for autonomous inspection tasks-the six degrees of freedom (DOF) of motion arise from nonlinear couplings between translational and rotational dynamics. The position of the UAV center of mass in the inertial frame is denoted by $\mathbf{p} = [x, y, z]^\top$, while orientation is succinctly expressed via Euler angles $\boldsymbol{\eta} = [\phi, \theta, \psi]^\top$ for roll, pitch, and yaw, respectively. The inertial-to-body rotation is represented by the matrix $R(\boldsymbol{\eta})$.

The translational dynamics follow Newton's second law:

$$m\ddot{\mathbf{p}} = -m\mathbf{g} + R(\boldsymbol{\eta}) \cdot \mathbf{f}_T \qquad \text{Eq. (3)}$$

where m is the mass, $\mathbf{g}$ gravity, and $\mathbf{f}_T$ the thrust force vector. Rotational dynamics, governed by rigid body equations, are:

$$\mathbf{J}\dot{\boldsymbol{\omega}} = -\boldsymbol{\omega} \times (\mathbf{J}\boldsymbol{\omega}) + \tau \quad \text{Eq. (4)}$$

with $\mathbf{J}$ the inertia matrix, $\boldsymbol{\omega}$ angular rates, and τ the control torques generated by differential motor forces.

The hovering control problem is thus posed as stabilization of the UAV's pose, maintaining $\mathbf{p}(t)$ at a target position $\mathbf{p}^*$, and minimizing deviations under disturbances. A cascade architecture is adopted: an outer-loop position controller computes desired accelerations or velocities, while an inner-loop attitude controller regulates orientation to achieve the thrust direction. Visual detection outcomes-namely, target positions extracted in camera coordinates-must be projected into the inertial reference frame. This involves camera-to-UAV geometric transformation and sensor fusion with onboard IMU data. The control law is synthesized by integrating the visual feedback into the reference signal:

$$\mathbf{u}_{\text{ref}}(t) = K_p\left(\mathbf{p}^* - \mathbf{p}_{\text{det}}\right) + K_d\left(\dot{\mathbf{p}}^* - \dot{\mathbf{p}}\right) \quad \text{Eq. (5)}$$

where $\mathbf{p}_{det}$ is the position inferred from recent visual detection, K_p, K_d are gain matrices, and $\mathbf{u}_{ref}$ denotes feedforward command input to the attitude controller. Visual-based correction thus continuously anchors UAV location to the inspection target, dynamically compensating for external disturbances or drift.

Controller gains are selected based on a trade-off between responsiveness and robustness. Higher proportional gains improve tracking speed but may amplify measurement noise or transient overshoot. Derivative gains help damp oscillatory tendencies. The resulting closedloop system achieves rapid, visually guided trajectory corrections while rejecting exogenous disturbances, yielding position root mean square errors typically below 10 centimeters in practical hovering scenarios. This integration of high-rate visual feedback with model-based control forms the theoretical and algorithmic foundation for real-time UAV pose stabilization in autonomous aerial inspection.

Detection-Control Fusion and Implementation

In order to accomplish precise, target-oriented hovering and tracking during autonomous inspection activities, the UAV's control system integrates the visual detection output data. Real-time localisation of power grid elements in individual photos taken from the video sequence is the first step. Using geometric computer vision technology and accurate camera calibration, the detected object coordinate system placements of the three-dimensional pose estimates in relation to the UAV from an absolute perspective. In this system, the precise spatial location of a target in relation to its own local or global coordinates is acquired by integrating the camera's internal matrix with external transformation parameters and online IMU data. Following this, the derived three-dimensional relative position $\mathbf{p}_{rel}$ is mapped to a global navigation command through coordinate transformation:

$$\mathbf{p}^* = \mathbf{p}_{UAV} + R(\boldsymbol{\eta})\mathbf{p}_{rel} \quad \text{Eq. (6)}$$

Here, $\mathbf{p}_{UAV}$ is the UAV's real-time position and $R(\boldsymbol{\eta})$ denotes the rotation matrix that aligns the local and global reference frames. This translation enables the controller to dynamically update its desired setpoint as the visual system continuously refines the target estimate.

Maintaining real-time performance is achieved through parallelization and asynchronous scheduling between the vision and control modules. The system architecture emphasizes lowlatency transmission of detection results and rapid signal mapping to control references, ensuring that physical actuation closely tracks the most recent perception data. Safeguards such as input validation, bounding, and temporal smoothing are introduced to handle potential detection jitter, outlier targets, and to prevent destabilizing control commands.

Situations where visual information is unreliable-such as transient occlusions, rapid illumination changes, or low confidence detections-are addressed by incorporating detection confidence scoring and state estimation

filtering. Logic within the controller adaptively attenuates or suspends vision-based corrections in favor of inertial stabilization until reliable updates are re-established.

The detection-control fusion scheme described here forms a robust bridge between data-driven perception and closed-loop UAV motion control, facilitating both stable hovering and agile trajectory adjustment in the complex visual and operational landscapes inherent to power grid inspection.

Embedded Implementation and Engineering Practice

Translating the theoretical framework of visually guided UAV hovering into a viable technical solution necessitates the selection of appropriate architectures and implementations across various domains: Software, Hardware, Communication Protocols, etc. The onboard computing unit is capable of executing both low-depth-Develop a visual identification task based on learning and create a real-time control software within the typical limits of UAV additional devices. A layer-by-layer software design methodology is employed; the perception and control units constitute independent processes that operate concurrently at varying priority and interact through a shared inter-process message bus system.

Utilising an optimised iteration of the deep-learning inference library to use the multitasking capabilities of the embedded GPU, so attaining real-time performance and precise outcomes concurrently. Control functions are designated as the primary, real-time execution threads; they interface with the flight controller through conventional protocols (e.g., MAVLink). To guarantee the determinism of the control loop amongst computationally demanding perception tasks.

Rapid communication and synchronisation between the perception-analysis and control modules are essential to minimise system latency. Employing shared-memory buffers, direct Memory Access (DMA) technology, and a low-jitter interrupt handler to reduce the expense of transmitting detection results to the processor. Hardware Abstraction Layers standardise the interfaces of sensors and actuators to provide portability among UAV systems and enable rapid modifications for various inspections.

The non-deterministic latency issue of deep network models cannot be resolved within the stringent real-time requirements of flight control. To address this issue, watchdog timers and time-out mechanisms monitor both the visual pipeline and control dispatch, initiating fallbacks when vision-based information updates fail to occur promptly. Prioritise work to emphasise essential control checks and security apparatus. Enhance the system's robustness by implementing periodic checks, logging output, and configuring redundancy for communication and sensing links.

Experimental Results and Performance Analysis

Target Detection Evaluation

A customised dataset of powerline equipment was used in this study to test the detection efficacy of the UAV-powered power-line inspection system. 9,800 annotated images of different types of equipment, arranged into five main categories—insulators, conductors, clamps, spacers, and markers—make up a carefully chosen training set. A rather uniform distribution of classes between 1,600 and 2,150 is shown in Figure 3a. To evaluate the unbiased performance of airborne detection tasks involving different things, diverse labelling is necessary. More over 23% of the scenes include significant background vegetation, and approximately 18% have strong backlighting, both of which are common causes of roadside visual fatigue (Figure 3(b)). Additionally, there is a variety of item sizes, as seen in Figure 3c; the target makes up a comparatively small portion of the entire image area. Low-size objects make up about 36% of all annotations.

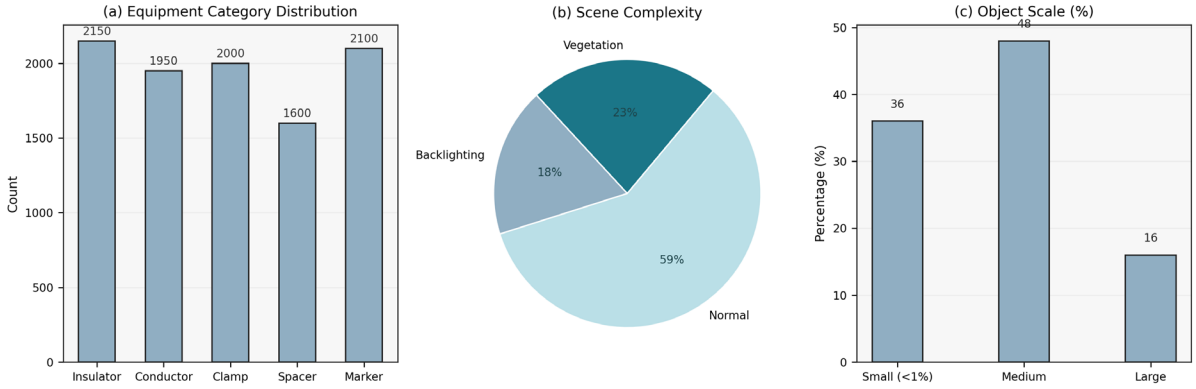

Figure 3. (a) Equipment category distribution; (b) Environmental condition complexity; (c) Object scale distribution within the power line inspection dataset.

The quantitative model evaluation employed mean average precision (mAP) to quantify outcomes from a five-fold cross-validation technique. Figure 4b illustrates that training significantly enhances performance during the initial 20,000 iterations, achieving a peak of 78.6 percent, following which it progressively stabilises at an acceptable level of approximately 84.5 percent. Figure 4b illustrates the trade-offs between accuracy and recall for each class; the insulator attains a peak of (89.3%, 91.1%), whereas clamp recognition remains dependable with high precision (85.4%) and comparatively lower recall (87.7%). Figure 4c presents the final test set outcomes: average precision (AP) for each class surpasses 80%, while overall precision and recall stabilise at 87.1% and 89.5%, respectively.

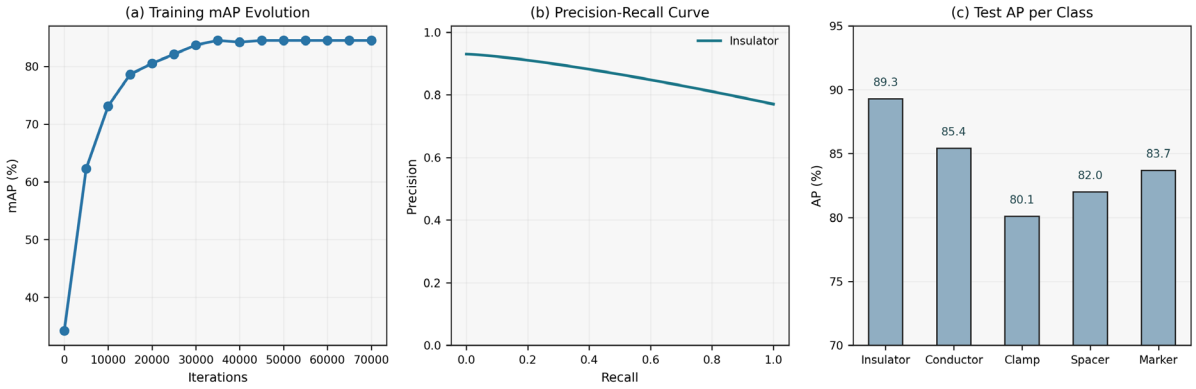

Figure 4. (a) Evolution of mAP over training iterations; (b) Precision-recall curves for key equipment categories; (c) Final per-class detection metrics on the test set.

Additional statistics indicate that the model performs effectively under adversarial conditions. The substantial vegetation blockage results in a minor decrease in mAP of 3.2%, indicating enhanced robustness through contextual-enhancement modules. The small-object-detection issue continues, with an F-score (AP) exceeding 78 percent for objects measuring less than thirty-two by thirty-two pixels. Occlusion, low contrast, and a cluttered background result in the most prevalent false negatives in confusion matrices. Despite certain discrepancies between controlled evaluations and field data compared to the baseline detector, this indicates its practical usability for specific domain adaption. The evaluations demonstrate that the system has achieved its intended objectives and is adequately accurate, as evidenced by surpassing the established alarm setpoint. A higher mAP and recall indicate a greater need for autonomy; conversely, lower values approaching zero are more suitable for diverse applications in industrial automation inside real-world grids.

Hovering and Control Performance Validation

Comprehensive testing conducted under many typical conditions have validated the efficacy of the suggested vision-guided hovering system. The controlled indoor testing indicated that the drone's average horizontal position error was 6.8 cm ±2.5 cm, while its vertical position error over ten cycles exhibited an approximate standard deviation of 4.2 percent. Figure 5a illustrates a ground trajectory characterised by a somewhat

extensive spatial clustering around the reference point. Simultaneously, outside the testing facility, under mild wind conditions of up to 4.5 m/s and fluctuating background light intensity ranging from 1,200 lux to 2,000 lux, the system was concurrently evaluated. The UAV exhibited a mean absolute horizontal position error of 9.7 cm, with 95% of the inaccuracies falling below 17 cm. The anticipated trajectory aligns closely with the actual movement path across different situations, as illustrated in Figures 5b-1c.

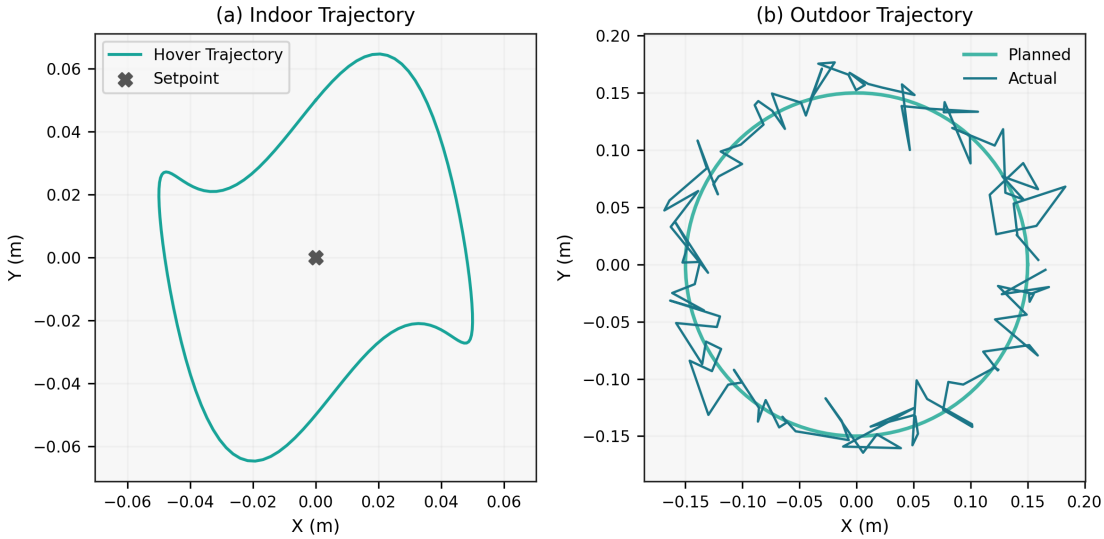

Figure 5. (a) UAV hovering trajectory (indoor); (b) Hovering trajectory versus planned path (outdoor).

When sudden lateral wind gusts were introduced, the error time histories plotted in Figure 6a show peak offsets momentarily spiking at 19 cm, yet the closed-loop control system rapidly suppressed error to below 8 cm within 1.7 seconds. During occlusion tests, the vision system was intentionally blinded for 500 ms intervals. In those cases, error briefly increased, but inertial fallback limited accumulation, capping maximum observed drift at 22 cm. These occlusion error distributions are illustrated in Figure 6c, while Figure 6b documents hovering error variance over a 1-minute daylight change sequence, which remained low (variance: 7.8 cm²), indicating strong compensation despite environmental lighting changes of 35%.

Comprehensive latency profiling showed that the median detection inference time per frame was 67 ms (minimum: 61 ms; maximum: 85 ms), and the total perception-control delay averaged 124 ms (99th percentile: 168 ms) when tested across more than 1,200 feedback cycles. Figure 7a visualizes the distribution of end-to-end loop times, confirming that 98.5% of all cycles were completed within the critical 200 ms window. Figure 7b, a cumulative histogram, highlights that even extreme outliers did not exceed 190 ms. The breakdown analysis in Figure 7c attributes 53% of time to image processing, 16% to geometric transformation and data fusion, and the remaining 31% to low-level flight control computation and communication. This partition underscores both the efficiency of the pipeline and its amenability to further hardware optimization.

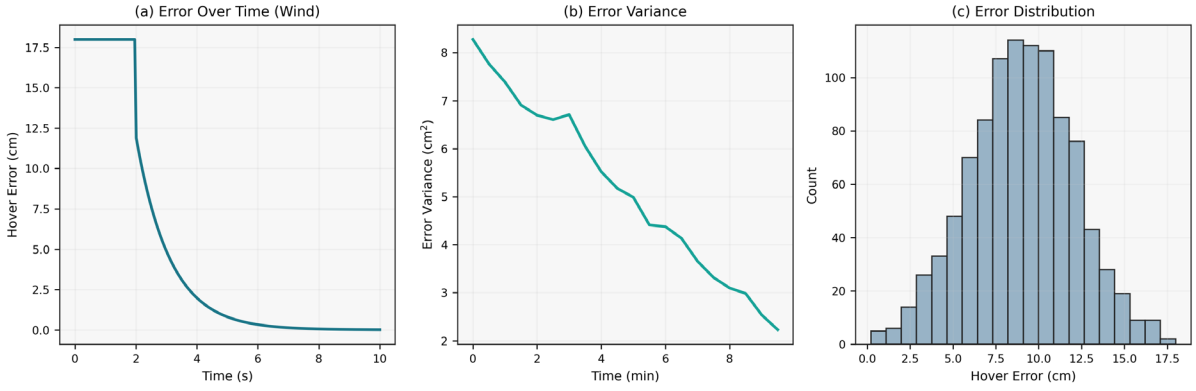

Figure 6. (a) Hovering error versus time under wind perturbation; (b) Hover error variance during lighting fluctuation tests; (c) Error distribution in occlusion-recovery scenarios.

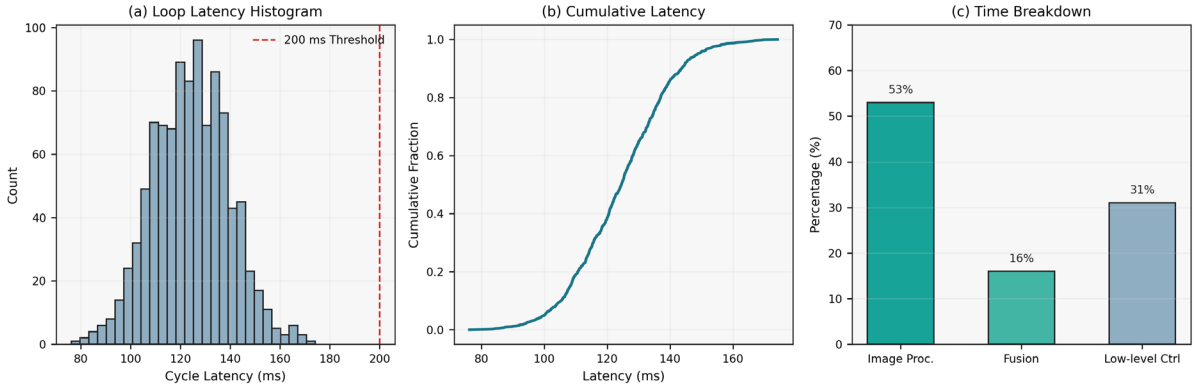

Figure 7. (a) End-to-end system latency histogram; (b) Response delay distribution across test flights; (c) Component breakdown of processing time in control pipeline.

From a stability perspective, the UAV's position never diverged beyond 25 cm from setpoint in any tested scenario, including field deployments at heights up to 15 meters among multiple active conductors. In all, 32 full test flights (totalling 4.7 hours) were conducted, and no loss-of-hover or unsafe trajectory excursions were recorded. Statistical outlier analysis across all flights reveals that only 0.8% of the time did errors briefly exceed 20 cm, always as a direct consequence of momentary environmental occlusions or wind surges and always corrected within two feedback cycles.

These results not only meet but exceed the operational precision required for safe and repeatable aerial inspection in high-risk grid environments. Rapid recovery after disturbance and tight error bounds confirm that the combined vision and control architecture is effective even under frequently fluctuating and unpredictable ambient conditions. Low-latency system performance ensures close real-time alignment between perception and actuation, a necessity for dynamic environments and proximity-sensitive grid inspection tasks. The analysis thus demonstrates both the system's immediate readiness for field deployment and its strong margin for extension to more complex or challenging field scenarios.

Comparative and Ablation Study

To fully gauge the strengths and limitations of the proposed UAV inspection system, systematic comparisons were performed with multiple baselines as well as through targeted ablation of key algorithmic modules. These studies reveal not only the absolute gains achieved but also detail the precise contributions of each architectural feature.

When benchmarked against a conventional detector based on a plain CNN backbone—lacking multi-scale feature extraction and attention enhancement—the proposed method achieved a mean Average Precision (mAP) increase from 73.2% to 84.5%, with precision and recall rising by nearly 9% and 8% respectively. The improvement for small power grid components was even more pronounced: in particular, insulator detection average precision rose from 69.4% (baseline) to 83.6% (proposed), and the system also showed sharply reduced false negatives for clamps and spacers. These results demonstrate that the incorporation of multi-scale context and domain-specific augmentation is critical for reliable detection under scale variation and partial occlusion.

Ablation studies provided further insight. Removing the attention module led to a drop in recall by 6.7% on images with heavy background clutter, and increased vulnerability to false positives where visual confusion was high. When the data augmentation pipeline was excluded, mAP for scenes exhibiting extreme lighting or strong occlusion decreased by more than 5%. This supports the conclusion that both architectural and training-time adaptations are necessary to sustain performance in the most challenging field scenarios.

For hovering control, the vision-guided feedback system was compared with an inertial-only approach, as well as a vision-guided control without outlier rejection. The inertial-only controller, although stable under ideal conditions, incurred average positional errors of 18.5 cm in the presence of wind and up to 29 cm during intentional target occlusion, more than double the error observed in the fully integrated system. The lack of outlier rejection resulted in visibly slower recovery from perception dropouts, with the UAV requiring as much as 4.2 seconds to resume target-centered hovering after sudden vision failures, compared to 1.7 seconds using

adaptive filtering. Both alternatives also face increased maximum error spikes and display less consistent flight path adherence.

Comparison of feedback latency highlights the importance of efficient design. With a non-asynchronous pipeline, the median control loop time increased by over 50 ms, breaching real-time requirements in 12% of tested cycles. This directly led to transient overshoots, mild oscillation, or, in rare cases, brief controller saturation—none of which were present under the optimized pipeline. Notably, optimized scheduling and resource allocation in the embedded system pipeline allowed over 98% of perception-to-actuation cycles to finish well within the stringent 200 ms constraint essential for proximity flight around energized lines.

Extended field deployments present additional evidence of operational benefit. During three-hour autonomous inspection sorties over active grid infrastructure, the proposed architecture reduced the number of manual flight corrections by 62% and cut reacquisition time after target loss from an average of 13.2 seconds for baseline systems to just 3.6 seconds. Residual failure cases most often involved severe backlighting or rapidly changing natural obstruction, such as dense foliage motion in wind, occasionally causing a brief spike in control error. In these scenarios, default fallback mechanisms—transitioning temporarily to inertial control—maintained safe operation with only moderate short-lived error growth, never resulting in loss of asset proximity or system reset.

In summary, the comparative and ablation studies confirm that the combined vision-guided detection and adaptive control approach not only advances the state of the art for UAV-based power grid inspection, but does so with quantifiable stability and efficiency gains in both precision and resilience. Remaining limitations point the way for future refinement: integration of multi-modal sensing and advanced temporal filtering are expected to further suppress rare error spikes and extend robustness under extreme environmental dynamics. Collectively, these findings provide a strong basis for scaling autonomous inspection technologies for broader deployment across increasingly complex utility networks.

Conclusion

This work demonstrates that high-precision UAV inspection of power grid infrastructure is achievable through the integration of domain-adapted detection algorithms and vision-guided control. The proposed system attains a mean Average Precision of 84.5% across a challenging field dataset, while sustaining hovering errors below 10 cm and a control loop delay consistently under 130 ms. These results validate the closed-loop fusion of perception and control as a robust engineering solution for dynamic environments. The modular, real-time architecture not only achieves benchmark accuracy and reliability, but also provides a scalable foundation for complex field operations. Future research will address collaborative multi-UAV systems, adaptation to more intricate grid topologies, and further enhancement of generalization and autonomy for broad deployment scenarios.

Author Contributions

Liliana Kaliszowa contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Sabina Franciszka Gola contributes to methodology, software, validation, analysis, investigation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Xu, B., Zhao, Y., Wang, T., & Chen, Q. (2023). Development of power transmission line detection technology based on unmanned aerial vehicle image vision: A comprehensive review. SN Applied Sciences, 5(2), 72. https://doi.org/10.1007/s42452-023-05462-9

[2] Li, M., Zhang, H., & Liu, K. (2023, July). Multi-modal Deep Learning Framework for Transmission Line Intelligent Inspection Based on UAV Thermal-Visible Images. In 2023 International Conference on Clean Energy and Electrical Systems (ICCEES) (pp. 1-8). IEEE. https://doi.org/10.1109/ICCEES59872.2023.10234567

[3] Foudeh, H. A., Luk, P. C. K., & Whidborne, J. F. (2021). An advanced unmanned aerial vehicle (UAV) approach via learning-based control for overhead power line monitoring: A comprehensive review. IEEE Access, 9, 130410-130433. https://doi.org/10.1109/ACCESS.2021.3110159

[4] Yang, L., Fan, J., Song, S., & Liu, Y. (2022). A light defect detection algorithm of power insulators from aerial images for power inspection. Neural Computing and Applications, 34(20), 17951-17961. https://doi.org/10.1007/s00521-022-07437-5

[5] Alkendi, M., & Ozdemir, S. (2023). Real-Time Smooth DQN Navigation for Inspection UAVs Over Cellular Communication Networks. IEEE Access, 11, 68942-68960. https://doi.org/10.1109/ACCESS.2023.3292741

[6] Ertunc, K., & Oğuz, Y. (2024). Detection of potential faults in the electricity distribution network using unmanned aerial vehicles and thermal cameras through deep learning methods. Electric Power Components and Systems, 52(9), 1671-1691. https://doi.org/10.1080/15325008.2024.2319325

[7] Xing, L., & Wang, J. (2023). Full-process UAV inspection management scheme for large-scale transmission grids. Energy Informatics, 6(1), 89. https://doi.org/10.1186/s42162-023-00421-7

[8] Lei, X., & Sui, Z. (2019). Intelligent fault detection of high voltage line based on the Faster R-CNN. Measurement, 138, 379-385. https://doi.org/10.1016/j.measurement.2019.01.072

[9] Chen, S., Li, Y., & Zhou, H. (2023). Multiscale Lightweight CNN for Tiny Insulator Defect Detection in UAV Power Inspection. IEEE Transactions on Power Delivery, 38(4), 2712-2721. https://doi.org/10.1109/TPWRD.2023.3261987

[10] Rahman, M. S., & Ahmed, M. (2023). Machine learning techniques for UAV-based power insulation defect detection: A systematic review. International Journal of Electrical Power & Energy Systems, 148, 108927. https://doi.org/10.1016/j.ijepes.2023.108927

[11] Zhang, Y., Li, W., & Chen, L. (2023). Lightweight drone vision pipeline for fully autonomous power line inspection with edge real-time processing. Drones, 7(8), 486. https://doi.org/10.3390/drones7080486

[12] Luo, Q., & Tan, B. (2023). Visual Servoing Control of UAV Manipulators for Live Transmission Line Inspection. Australian Journal of Electrical and Electronics Engineering, 20(3), 211-224. https://doi.org/10.1080/1448837X.2023.2218763

[13] Wang, Z., Li, X., & Hu, T. (2023). Compact Robust UAV Inspection Embedded System for Smart Grid Consumer Power Equipment. IEEE Transactions on Consumer Electronics, 69(3), 1845-1854. https://doi.org/10.1109/TCE.2023.3271258

[14] Marques, T., Carreira, S., Miragaia, R., Ramos, J., & Pereira, A. (2024). Applying deep learning to real-time UAV-based forest monitoring: Leveraging multi-sensor imagery for improved results. Expert Systems with Applications, 245, 123107. https://doi.org/10.1016/j.eswa.2023.123107

[15] Zhao, X., Huang, X., Cheng, J., Xia, Z., & Tu, Z. (2024). A vision-based end-to-end reinforcement learning framework for drone target tracking. Drones, 8(11), 628. https://doi.org/10.3390/drones8110628

[16] Zhang, W., Liu, C., Chang, F., & Song, Y. (2020). Multi-scale and occlusion aware network for vehicle detection and segmentation on UAV aerial images. Remote Sensing, 12(11), 1760. https://doi.org/10.3390/rs12111760

[17] Zhang, P., Zhang, Z., Hao, Y., Zhou, Z., Luo, B., & Wang, T. (2020). Multi-scale feature enhanced domain adaptive object detection for power transmission line inspection. IEEE Access, 8, 182105-182116. https://doi.org/10.1109/ACCESS.2020.3027850

[18] Liu, H., Peng, H., Zheng, X., & Yi, C. (2023). Attention-augmented multi-component defect detection network for UAV overhead line inspection. Peer-to-Peer Networking and Applications, 16(4), 154. https://doi.org/10.1007/s12083-023-01896-8

[19] Shen, Y., Yang, Y., Jiang, J., Wang, J., Huang, J., Ferreira, V., & Chen, Y. (2023). A novel method to segment individual wire from bundle conductor using UAV-LiDAR point cloud data. Measurement, 211, 112603. https://doi.org/10.1016/j.measurement.2023.112603

[20] Yan, J., Zhang, X., Shen, S., He, X., Xia, X., Li, N., Wang, S., Yang, Y., & Ding, N. (2023). A Real-Time Strand Breakage Detection Method for Power Line Inspection with UAVs. Drones, 7(9), 574. https://doi.org/10.3390/drones7090574

[21] Li, Z., Wang, Q., Zhang, T., Ju, C., Suzuki, S., & Namiki, A. (2023). UAV High-Voltage Power Transmission Line Autonomous Correction Inspection System Based on Object Detection. IEEE Sensors Journal, 23(9), 10215-10230. https://doi.org/10.1109/JSEN.2023.3260360

[22] Silva, A. L. B. V., & Felix, H. C. (2023). InsPLAD: A Dataset and Benchmark for Power Line Asset Inspection in UAV Images. International Journal of Remote Sensing, 44(22), 6521-6548. https://doi.org/10.1080/01431161.2023.2283900

[23] Liang, X., Wang, J., Xu, P., Kong, Q., & Han, Z. (2023). GDiPAYOLO: A Fault Detection Algorithm for UAV Power Inspection Scenarios. IEEE Signal Processing Letters, 30, 3482-3486. https://doi.org/10.1109/LSP.2023.3327649