
Autonomous UAVs in Critical Infrastructure Inspection: Empirical Evaluation, Reporting Standard, and Implications for Strategic Decisions

Submitted 24/07/26, 1st revision 11/08/26, 2nd revision 22/08/26, accepted 10/09/26

Magdalena Majchrzak¹, Grzegorz Diemientiew², Piotr Niedzielski³,
Krzysztof Dyl⁴, Natalia Drop⁵, Piotr Durajczyk⁶, Ewa Kuzionko-Ochrymiuk⁷,
Jarosław Kozuba⁸, Zbigniew Osadowski⁹, Jacek Marcinkiewicz¹⁰

Abstract:

Purpose: Autonomous unmanned aerial vehicles (UAVs) have become a mature tool for remote sensing data acquisition for critical infrastructure (IC) inspections, particularly in the energy, transportation, bridge infrastructure, water networks and industrial installation sectors. The aim of the article is to synthetically capture the results of empirical research: field inspection missions using RGB, thermal imaging (TIR) and LiDAR sensors, as well as automatic data analytics (deep learning) to detect defects and anomalies.

Design/Methodology/Approach: In the methodological part, a unified scheme for reporting empirical research (mission parameters, method of obtaining the “reference truth”, quality metrics and operational utility indicators) is proposed, as well as a framework for integrating results with crisis management processes (time-to-assessment, false dispositions, repair priorities).

¹Warsaw University of Technology, The College of Economics and Social Sciences, Poland, e-mail: magdalena.majchrzak@pw.edu.pl;

²General Tadeusz Kościuszko Military University of Land Forces, Poland, ORCID: 0000-0002-0229-5004, e-mail: diemientiew@gmail.com;

³University of Szczecin, Poland, ORCID: 0000-0001-5024-4722, e-mail: piotr.niedzielski@usz.edu.pl;

⁴Pomeranian University in Słupsk, Poland, ORCID: 0000-0003-1768-9644, e-mail: k.dyl@wp.pl;

⁵Maritime University of Szczecin, Poland, ORCID: 0000-0002-6148-7815, e-mail: n.drop@am.szczecin.pl;

⁶The same as in 5, ORCID: 0000-0003-1447-973X, e-mail: pdurajczyk@szczecin.uzs.gov.pl;

⁷University of Białystok, Poland, ORCID: 0000-0003-3172-5998, e-mail: e.kuzionko-ochrymiuk@uwb.edu.pl;

⁸Silesian University of Technology, Faculty of Transport and Aerospace Engineering, Poland, e-mail: jaroslaw.kozuba@polsl.pl;

⁹Pomeranian University in Słupsk, Poland, ORCID: 0000-0001-5086-7671, e-mail: zbigniew.osadowski@upsl.edu.pl;

¹⁰Corresponding author, the same as in 7, ORCID: 0000-0003-4701-8155, e-mail: j.marcinkiewicz@uwb.edu.pl;

Findings: A review of representative studies indicates that the effectiveness of component and fault detection depends not only on the selection of the machine learning model, but above all on the quality and standardization of data, environmental conditions, and the method of field validation.

Practical Implications: Comparable reporting enables robust synthesis across sectors and accelerates transfer to practice. Integrating UAV products with GIS/CMMS/SCADA and tracking time-to-assessment and false dispatch supports crisis-management decisions and repair prioritization.

Originality/Value: A minimal, six-block protocol is proposed to standardize reporting of empirical UAV inspections of critical infrastructure.

Keywords: Autonomous UAVs, critical infrastructure management, strategic decision-making, decision support systems, infrastructure inspection, crisis management, remote sensing, operational readiness, deep learning, repair prioritization.

JEL Codes: M15, O32, H54, L90.

Paper type: Review article.

1. Introduction

Critical infrastructure requires continuous monitoring of the technical condition and quick diagnostics after emergencies (failures, extreme weather phenomena, fires, seismic events or hybrid incidents). Classic field inspections can be expensive, risky for personnel and limited by the availability of facilities (heights, danger zones, long-distance linear infrastructure). UAVs allow you to reduce reconnaissance time, increase range, and improve repeatability of measurements with mission planning, photogrammetry, and automatic image analysis.

On a thematic level, this can also be seen in the world of science itself, where special editions are devoted to infrastructure inspections and UAV autonomy emphasize the transition from "data acquisition itself" to systems that include flight planning, processing, analytics and implementation in the practice of infrastructure maintenance. Despite the dynamic development of empirical research on the use of autonomous UAVs in the inspection of critical infrastructure, the literature on the subject shows significant inconsistencies in the reporting of mission parameters, the method of obtaining and verifying the "ground truth" and the selection of quality assessment metrics. Many publications focus on model performance metrics (mAP, IoU, F1), but rarely consider operational metrics and the integration of results into decision support systems.

Thus, there is a lack of a unified approach to reporting that would allow for the comparability of empirical research between critical infrastructure sectors and the assessment of their operational readiness in the context of strategic decisions. As a consequence, the current state of research makes it difficult to both synthetically

assess technological progress and reliably compare the proposed solutions in terms of their application in real conditions.

This gap concerns not only the lack of standardization of reporting, but also the insufficient link between remote sensing results and the operational requirements and decisions of organizations responsible for maintaining critical infrastructure. In response to the identified research gap, this article introduces four essential elements.

First, it proposes an operational definition of empirical research in the area of critical infrastructure inspection using autonomous UAVs, specifying minimum conditions for mission documentation, sensor characterization, and reference data validation.

Secondly, it develops a six-block reporting standard including: mission parameters, sensors used, data processing procedures, method of ground truth verification, model quality metrics and operational utility indicators.

Third, it introduces a semi-quantitative tool for assessing the comparability of empirical publications, allowing for the classification of studies in terms of completeness of reporting and potential operational usefulness.

Fourthly, the article integrates the assessment of remote sensing quality with a decision-making perspective, indicating the conditions under which autonomous UAVs can be a reliable tool to support critical infrastructure management processes.

As a result, the paper shifts the focus of the discussion from the level of the declared effectiveness of algorithms to the level of methodological consistency, operational readiness and the applicability of the results in real operating conditions, indicating at the same time that the standardization of reporting and quality assessment is a condition for rational strategic decisions regarding the choice of technology, allocation of resources and building organizational capacity in the field of the use of autonomous UAVs in critical infrastructure. Based on the above research gap, a hypothesis was formulated that the lack of standardization of empirical UAV research reporting limits their comparability and ability to support strategic decisions in critical infrastructure.

2. Research Methodology

2.1 Definition of "Empirical Study" in the CI Inspection

In the context of this article, empirical works are considered to be works that meet the following in total:

- implementation of UAV missions in field/operational conditions or in a controlled engineering experiment,
- use of at least one remote sensing sensor (RGB, TIR, LiDAR, multispectral),

- validation of results on the basis of 'reference truth' (e.g., engineering inspection, ground measurements, RTK/GNSS, maintenance documentation),
- report of quality metrics (e.g. mAP/IoU/F1, RMSE geometry, location precision) and/or operational metrics (processing time, number of false positives).

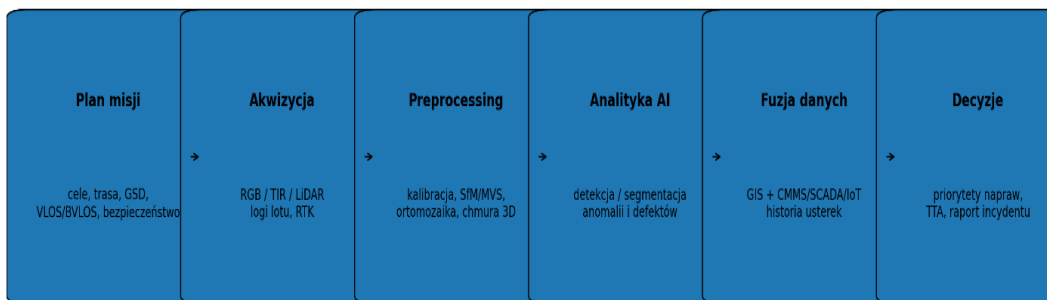
2.2 Minimum Reporting Protocol

In order for the results to be comparable between works, it is recommended to report in 6 blocks:

1. Mission parameters: altitude, speed, GSD, overlap, mode (manual/auto), number of passes;
2. Sensors and calibration: RGB/TIR/LiDAR, camera parameters, synchronization;
3. Preprocessing: georeferencing, photogrammetry (SfM/MVS), point cloud filtration, orthomosaic;
4. Ground truth: method of acquisition, accuracy, uncertainty;
5. Metrics: mAP/IoU/F1 (detection/segmentation), RMSE (3D models), false alarm indicators;
6. Operational usability: time-to-assessment, integration with GIS/CMMS/SCADA, repair priorities.

Figure 1. Protocol six blocks

Rysunek 1. Pipeline inspekcji IK z autonomicznymi UAV (teledetekcja → AI → decyzja).



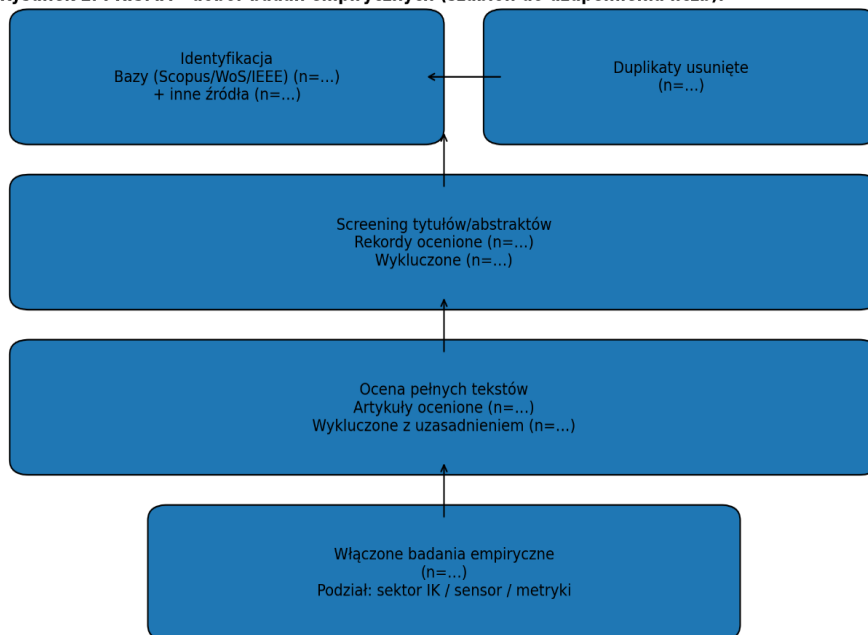
Source: Own study.

Figure 2 shows the PRISMA scheme, a standard way of reporting the literature selection process for review, used here with an emphasis on empirical publications on critical infrastructure inspection using UAVs. The diagram shows the subsequent stages of selection: from the identification of records in bibliographic databases (e.g., Scopus, Web of Science, IEEE Xplore) and possibly in other sources, through the removal of duplicates, to the initial screening of titles and abstracts. Next, the

stage of evaluation of full texts is presented, in which papers that do not meet the empirical criteria are rejected (e.g., lack of field validation, lack of "reference truth", only simulations or lack of quality metrics).

Figure 2. The PRISMA scheme

Rysunek 2. PRISMA - dobór badań empirycznych (szablon do uzupełnienia liczb).



Source: Own study.

The last block of the figure indicates the final number of empirical studies included and suggests their further classification for the synthesis of results (e.g., by critical infrastructure sector, type of sensor, type of analytical task and reported metrics). In practice, such a scheme is intended to show the transparency of the review and enable the reader to assess whether the selection of sources was consistent and whether the conclusions are based on works that have actually been experimentally confirmed in the field.

Figure 3. The framework for conducting and reporting empirical research

Rysunek 3. Ramy badań empirycznych: od misji UAV do metryk i użyteczności kryzysowej.



Source: Own study.

Figure 3 presents the framework for conducting and reporting empirical research in the area of autonomous UAV inspections, arranging the entire process in a logical sequence from the planning of the experiment to the use of the results in operational decisions. At the beginning, the study project was taken into account, i.e. the definition of the critical infrastructure sector, the type of facility and the scenario (routine inspection or action in the event of an incident).

The next step is to design and implement a UAV mission with parameters determining the quality of data (altitude, GSD, photo coverage, autonomous or manual flight mode, use of RTK and registration of flight logs). Next, the figure indicates the selection of sensors (RGB, TIR thermal imaging, LiDAR) and their calibration as a critical element, as each modality detects different classes of defects and has different environmental constraints.

At the same time, the role of "ground truth" obtained from e.g., engineering inspections, NDT methods or ground measurements is emphasized, because without reliable field validation, the results of AI analytics are not comparable or useful for CI operators. Further down the schema, data analytics (DL/ML, transfer learning, validation, and uncertainty) are evaluated, followed by evaluation of the results using both quality metrics typical of computer vision (mAP, IoU, F1) and 3D geometry (RMSE), as well as practical indicators such as processing time, false alarm scale, and Time-to-Assessment, i.e., the time from the raid to obtaining action-ready information.

The framework culminates in decision-making integration, in which the results are plugged into GIS tools and operating systems (e.g., CMMS/SCADA), enabling the prioritization of repairs, the preparation of a report and support for crisis management activities. The whole figure emphasizes that in empirical research in CI inspection, it is crucial not only whether the model "works well", but also whether it is fed with data with known parameters, validated against a reliable reference and provides results in time and in a form useful for operational decisions.

3. Research Results and Discussion

3.1 Linear Energy: RGB/TIR + Deep Learning and UAV-LiDAR

Field studies show that the combination of RGB and thermal imaging can support safe autonomous inspection by detecting wires/objects and assessing anomalies. In the work validated during the inspection mission, very high values of mAP@0.5 were shown for both visible and thermographic images, which is important for near-real-time detection.

At the same time, the importance of LiDAR with UAVs is growing. The empirical direction of "deep learning on point clouds" reduces the need for manual selection of

wire separation thresholds, which has so far been a significant barrier to automation in the practice of maintaining transmission corridors.

In addition, reviews of line extraction and reconstruction methods from LiDAR data organize existing approaches and identify common sources of error (scrub, complex multi-wire scenarios, different scan angles).

3.2 Bridge Infrastructure: UAV Photogrammetry and Object/Deep Learning Analysis

In the area of bridges and viaducts, two empirical currents dominate: (1) UAV photogrammetry (3D models + orthomosaics) as "NDT from the air", (2) automatic detection of defects (cracks, cavities, corrosion, leaks) using OBIA or deep learning methods. Empirical photogrammetric methodology for concrete objects shows that 3D products and object analysis can reliably support the identification of surface degradation.

In maintenance practice, the elements connecting UAVs with infrastructure management systems (BMS/CMMS) and the "digital twin" are also important, which allows for the automation of inspection trajectory planning.

In contrast, empirical work on the inspection of point components (e.g., bolts/connections) shows that vision analysis with UAVs and DL models can automate the assessment of the condition of components that have traditionally required a close approach from personnel.

3.3 Renewable Energy: Thermal Inspections of PV Farms (Hotspots and Defects)

Empirical implementations of UAV-TIR in photovoltaics provide two classes of results:

- (a) thermal anomaly maps (hotspots),
- (b) recommendations of acquisition parameters (e.g., the influence of observation geometry on temperature measurement).

Field studies in PV farm operating conditions show that thermal imaging from UAVs can create temperature orthomosaics and detect faulty modules.

In strictly "AI-ready" terms, segmentation and classification of defects in thermal images from UAVs is being developed as an operational method for large installations.

This is of direct importance for the IC in the context of the continuity of energy supply: fast diagnostics reduce losses and the risk of secondary failures.

3.4 Water Supply Networks and Leaks: UAV-IR and Fusion with Height/Shadow Models

Detecting water leaks is difficult because the thermal signal can be subtle and dependent on conditions (ground, humidity, sunlight, shade). An empirical study based on IR imaging from UAVs shows an approach that integrates the TIR image with additional layers (e.g., height and shadow models) to improve the detection of subsurface leaks.

It is worth noting that the experimental protocol (measurement time, thermal stabilization, field validation) is critical in water inspection, because without it, the risk of false alarms increases.

3.5 Infrastructure Pipelines and Corridors: UAV-NIR + Satellite Data as a Baseline

An interesting class of empirical research is the work building a "baseline" for monitoring infrastructure corridors through the integration of UAV-NIR with satellite data and archival orthophotomaps. This approach is operationally useful because it allows for the detection of changes over time and supports early warning (e.g., technical lane violations).

3.6 Tunnel and Railway Infrastructure: Autonomous Inspections and 3D Modelling

In the case of tunnels and 'tunnel-like' facilities, it is important to combine autonomy with the acquisition of high-quality data (often under conditions of limited connectivity and GNSS). Empirical work demonstrating an aerial robotics solution for the autonomous inspection of such objects shows that integrating a UAV platform into an inspection system can reduce the cost and risk of manual inspections.

The verification and evaluation of the empirical works was carried out in two stages:

In Stage 1, the following boundary criteria were applied: (E1) implementation of the UAV mission in field/experimental conditions, (E2) use of at least one remote sensing sensor, (E3) validation against the "reference truth", (E4) report of quality metrics and/or operational metrics.

Phase 2 assessed the quality of reporting and the comparability of results across the six blocks of the minimum protocol (mission parameters, sensors and calibration, preprocessing, ground truth, quality metrics, operational usability), awarding 0-2 points for each block (0-12 points in total). The result was used to classify the papers into low/medium/high empirical comparability.

Table 1. *Boundary criteria: is this an empirical study?*

ID	Criterion (must be met)	Proof in the article (where?)	Rating (YES/NO)
E1	Execution of UAV missions in the field / operationally or in a controlled engineering experiment	Section / Page	
E2	≥ 1 remote sensing sensor (RGB/TIR/LiDAR/MS) is used	Section / Page	
E3	Validation based on "reference truth" (engineering inspection, RTK/GNSS, NDT, documentation)	Section / Page	
E4	Quality metrics report (mAP/IoU/F1/RMSE...) and/or operational metrics	Section / Page	

Note: If at least one = NO $\rightarrow$ the work does not enter the "empirical" pool.

Source: Own study.

Table 1 presents the boundary criteria for the initial qualification of publications as empirical studies that can be compared in this review. The use of conditions E1–E4 makes it possible to distinguish between conceptual or simulation work and research based on real data obtained during a UAV mission. This ensures that further analysis only includes publications that provide verifiable and interpretable results in the context of critical infrastructure inspections.

Table 2. *Reporting Quality Scale*

Block	What we verify (minimum)	0	1	2	Points	Note (what's missing)
B1 Mission Parameters	altitude, speed, GSD, overlap, flight mode, number of passes/logs	☐	☐	☐		
B2 Sensors and Calibration	Sensor Type + Key Parameters + (if applicable) Synchronization/Calibration	☐	☐	☐		
B3 Preprocessing	georeferencing, SfM/MVS/orthomosaic, cloud filtration, pipeline	☐	☐	☐		
B4 Ground truth	Method of acquisition + Accuracy/Uncertainty + Field validation procedure	☐	☐	☐		
B5 Quality metrics	mAP/IoU/F1/RMSE + job definition (detection/segment/instances)	☐	☐	☐		
B6 Operational usability	time-to-assessment, false dispatch/FP, GIS/CMMS/SCADA integration, repair priorities	☐	☐	☐		

Note: Point scale: 0 points – none / cannot be reproduced, 1 point – partially (incomplete/ambiguous), 2 points – full (replicable, parameters clearly stated).

Source: Own study.

Table 2 is an appropriate tool for assessing the quality of reporting and the comparability of empirical research results. It is built on the basis of six blocks of the "minimum reporting protocol", which organize the key information necessary for replication and evaluation of the operational usefulness of solutions.

A scale of 0–2 points for each block allows for a semi-quantitative assessment of the degree of completeness of reporting (from the lack of data, through a partial description, to a fully replicable description). The cumulative score then allows you to compile publications in a consistent and transparent way.

Table 3. Interpretation of the result

Total (0-12)	Quality class	Practical interpretation
0–4	Low	the result is difficult to compare; no data for replication and deployment
5–8	Average	Partial comparability; limited replicability/operability
9–12	High	High replicability; data and metrics sufficient for synthesis and deployments

Source: Own study.

Table A3 shows how to interpret the total score obtained in Table A2 and translates it into three quality classes: low, medium and high. The proposed score thresholds act as a practical 'interpreter' between the sub-assessment and the synthetic conclusions, facilitating the presentation and discussion of the results of the review. These classes indicate the extent to which a given publication supports replication, interdisciplinary comparisons, and potential implementations in the operational environment.

3.7 Autonomy and Operational Risk and Credibility

In practice, "autonomy" in an IK inspection usually means: route planning, obstacle avoidance, stable data acquisition and partially automatic inference (edge/onboard). In the European operational reality, risk assessment (especially for high-risk flights) is often described in relation to the SORA methodology developed by JARUS.

From the perspective of empirical research, this means that authors should report not only the "result of the model", but also the limitations of the mission (e.g. GNSS availability, obstacles, lighting conditions), because these factors determine the possibility of transferring the method to practice.

3.8 "Empirical Comparability Problem" – Why Metrics and Ground Truth Are Crucial

In the empirical literature, a wide range can be seen:

- definition of defect (what is a crack/corrosion/anomaly),

- the method of obtaining ground truth (inspection, RTK, documentation),
- metrics (sometimes no false positives and processing delays).

Hence the recommendation to report operational indicators in addition to mAP/IoU/F1: time-to-assessment, false dispatch rate and validation costs.

3.9 Integration with Crisis Management: From Data to Decisions

Crisis management needs not so much "pretty pictures" as actionable information: where the problem is, how urgent it is, what resources to send. Works in the field of situational awareness emphasize that UAVs significantly reduce recognition delays and support coordination through near-real-time data.

From a process perspective, this can be embedded in incident management logic (roles, resources, interoperability), which is consistent with ISO guidelines on incident management.

4. Conclusions, Proposals and Recommendations

1. The most "empirically mature" areas of IK inspection with UAVs are: linear power engineering (RGB/TIR/LiDAR), bridge infrastructure (photogrammetry + object analytics/DL) and thermal imaging of PV farms.
2. The effectiveness of AI models is highly dependent on data quality, environmental conditions, and the quality of field validation – so the reporting standard (mission parameters + ground truth + metrics + operational metrics) is just as important as the choice of network architecture.
3. The greatest practical value comes from the integration of remote sensing products into the decision-making process: prioritization of faults, reduction of false dispatches and reduction of time from raid to action.

In the empirical literature on critical infrastructure inspection, there is an increasing thesis that the combination of RGB and thermal imaging in combination with deep learning models provides a high level of security and detection efficiency, and even enables near-real-time operation. In fact, field studies report very high values of mAP@0.5 for both visible and thermographic data, which is an important argument in favor of the operational use of such systems in the inspection of lines and infrastructure objects.

However, in a scientific discussion, it is worth arguing with the simplified interpretation, according to which a high mAP@0.5 in itself determines the readiness of the technology for autonomous implementations.

First, mAP calculated at a threshold of 0.5 is a relatively "mild" metric and may not reveal problems typical of IK inspections, such as the detection of thin, fine or low-contrast objects (e.g., fragments of wires against the background of complex

vegetation or hardware elements with an unfavorable viewing angle). In such cases, the advantage reported on mAP@0.5 does not always hold up to stricter criteria, and the risk of false positives or omissions may be more significant than the "average performance" in benchmark terms.

Second, the argument about the advantage of RGB and TIR fusion must be grounded in remote sensing realities: thermals are highly dependent on environmental conditions and the object's "thermal window", and modality fusion requires synchronization, calibration, and stable georeferencing; Errors in these elements can in practice generate apparent anomalies or uncertain location of the defect.

Third, even if the inference of the model is fast, in the inspection system, "near-real time" is owned by the entire processing chain, not just the neural network: latency is affected by data transmission, video caching, stabilization and geolocation of detection, and conversion of the result into a decision product (e.g., GIS point, risk category, action recommendation).

In other words, the dominant narrative of "real-time detection" should be complemented by an end-to-end perspective that measures time-to-assessment and the costs of false dispositions, not just the effectiveness of classification.

At the same time, the literature is increasingly promoting the UAV-LiDAR approach, in which "deep learning on point clouds" is to reduce the need to manually tune thresholds and cable segmentation rules, which has so far been a barrier to automation in the maintenance of transmission corridors. Indeed, the empirical results of such studies suggest a shift in weight from heuristic methods to models that learn geometric representations.

However, it is worth arguing with the thesis that the elimination of thresholds automatically translates into the elimination of implementation complexity.

First, in the case of LiDAR, the problem of thresholding is largely replaced by sensitivity to the quality and characteristics of the input: cloud density, noise, flight trajectory, filtration and altitude normalization affect whether the model will be stable in other measurement campaigns and with a different sensor. This means the risk of limited domain shift, which in the operating environment may force domain adaptation or expensive model training.

Second, the effectiveness of segmentation or detection in standard metrics (e.g., IoU) does not always reflect the cost of errors relevant to the IK operator. In the inspection of linear infrastructure, a false positive mistake (e.g., a branch classified as a conduit) can be less critical than false negative (overlooking a conduit or incorrectly determining the route), because it is the errors of omissions that can lead to dangerous conclusions about the gauge or the risk of contact with vegetation.

Therefore, in addition to classical metrics, it is necessary to report geometric metrics (e.g., wire axis deviation) and cost-per-error analysis to understand whether the model actually reduces operational risk.

Third, LiDAR's deep learning-based automation doesn't eliminate manual work, but shifts it toward data preparation: labeling wires in point clouds is time-consuming, and label errors and annotation inconsistencies can degrade the model in a way that is difficult to detect at the deployment stage. As a consequence, the promise of "automation" should be verified not only by predictive results, but also by the balance of expenditures (cost of preparing ground truth, label quality control, maintaining the model over time).

In this context, reviews of line extraction and reconstruction methods from LiDAR data, which organize common sources of error (brush, multi-wire scenarios, the impact of scan angles), are a valuable signpost, but they also need critical refinement if they are to become truly "implementation".

First, identifying sources of error at a general level is not yet an operational recommendation: the IK operator needs answers on how the flight parameters and scan configuration reduce the risk of losing wires in brush or in multilayer geometry. Without translating into acquisition and QA protocols (e.g., minimum cloud density, omnidirectional raid strategies, redundancy), reviews remain predominantly descriptive.

Secondly, the problem of "multiwires" is sometimes reported in the literature in a heterogeneous way, because different works define the task differently: some treat cables as one class, others require instance segmentation of each cable. This makes it difficult to compare results and draw generalized conclusions about the advantage of a given method. For this reason, there is a need for standardization: a clear separation of whether the study concerns class or instance segmentation, and reporting consistent metrics in both approaches.

Third, the influence of the scan angle and the geometry of the raid shows that the quality of the result is a property of the system, not just the algorithm. Even the best model can be unstable if the acquisition generates measurement shadows or uneven distribution of points, so the conclusions of the reviews should more often lead to the recommendation of a "better data acquisition protocol" as an equally important element as "better network architecture".

As a result, although empirical research confirms the high effectiveness of detecting infrastructure elements in both RGB/TIR and UAV-LiDAR approaches supported by deep learning, the narratives dominant in the literature about "readiness for autonomy" and "reduction of manual labor" need to be clarified. It becomes crucial to shift the discussion from the effectiveness metric itself to the assessment of end-to-end resilience: from the quality and standardization of acquisitions, through the

reliability of ground truth, to the costs of false positives, the time to assess the situation, and the transferability of models between campaigns and sensors. Such an approach allows for better embedding UAV remote sensing in the realities of critical infrastructure maintenance and crisis management requirements, where it is important not only "whether it detects", but "whether it detects reliably, repeatably and in time".

Abbreviations: *The following abbreviations are used in this manuscript:*

UAV	Unmanned Aerial Vehicle
IK	Critical Infrastructure (Infrastruktura Krytyczna)
RGB	Red-Green-Blue imaging
TIR	Thermal Infrared
LiDAR	Light Detection and Ranging
DL	Deep Learning
ML	Machine Learning
mAP	mean Average Precision
IoU	Intersection over Union
RMSE	Root Mean Square Error
GNSS	Global Navigation Satellite System
RTK	Real-Time Kinematic
SfM	Structure from Motion
ILL	Multi-View Stereo
OBIA	Object-Based Image Analysis
GIS	Geographic Information System
CMMS	Computerized Maintenance Management System
SCADA	Supervisory Control and Data Acquisition
SORA	Specific Operations Risk Assessment
JARUS	Joint Authorities for Rulemaking on Unmanned Systems
NDT	Non-Destructive Testing
PV	Photovoltaics
NIR	Near-Infrared

Appendix A:

Total (0-12)	Quality class	Practical interpretation	Practical interpretation
0-4	Low	the result is difficult to compare; No data to replicate and deploy	the result is difficult to compare; No data to replicate and deploy
5-8	Average	partial comparability; Limited replicability/operability	partial comparability; Limited replicability/operability
Block	What we verify (minimum)	Scale 0-2	Note (what's missing)
B1 Mission	altitude, speed, GSD,	0 none; 1 partially;	

Parameters	overlap, flight mode, number of passes/logs	2 full	
B2 Sensors and Calibration	Sensor Type + Key Parameters + (if applicable) Synchronization/Calibration	0 none; 1 partially; 2 full	
B3 Preprocessing	georeferencing, SfM/MVS/orthomosaic, cloud filtration, pipeline	0 none; 1 partially; 2 full	
B4 Ground truth	Method of acquisition + Accuracy/Uncertainty + Field validation procedure	0 none; 1 partially; 2 full	
B5 Quality metrics	mAP/IoU/F1/RMSE + job definition (detection/segment/instances)	0 none; 1 partially; 2 full	
B6 Operational usability	time-to-assessment, false dispatch/FP, GIS/CMMS/SCADA integration, repair priorities	0 none; 1 partially; 2 full	
ID	Criterion (must be met)	Proof in the article (where?)	Rating (YES/NO)
E1	Execution of UAV missions in the field / operationally or in a controlled engineering experiment	Section / Page	
E2	≥1 remote sensing sensor (RGB/TIR/LiDAR/MS) is used	Section / Page	
E3	Validation based on "reference truth" (engineering inspection, RTK/GNSS, NDT, documentation)	Section / Page	
E4	Quality metrics report (mAP/IoU/F1/RMSE...) and/or operational metrics	Section / Page	

References:

A Survey on Multi-UAV Path Planning: Classification, Algorithms, Open Research Problems, and Future Directions. *Drones* 2025, 9(4), 263.
<https://doi.org/10.3390/drones9040263>.

- A Survey on the Key Technologies of UAV Motion Planning. *Drones* 2025, 9(3), 194. <https://doi.org/10.3390/drones9030194>.
- Amir Farzin Nikkhah *et al.*, n.d. UAV-Based Infrastructure Inspections: A Literature Review and Proposed Framework for AEC+FM, *Robotics* 2026. <https://doi.org/10.48550/arXiv.2601.11665>.
- Automatic Inspection of Bridge Bolts Using Unmanned Aerial Vision and Adaptive Scale Unification-Based Deep Learning, *Remote Sens.*, 2023, 15(2), 328. <https://doi.org/10.3390/rs15020328>.
- Chen Lyu *et al.*, n.d. UAV-based deep learning applications for automated inspection of civil infrastructure, *Sensors* 2025, 25(23), 7118. <https://doi.org/10.3390/s25237118>.
- Design and Experimental Evaluation of an Aerial Solution for Visual Inspection of Tunnel-like Infrastructures, *Remote Sens.*, 2022, 14(1), 195. <https://doi.org/10.3390/rs14010195>.
- Drone Imaging and Sensors for Situational Awareness in Hazardous Environments: A Systematic Review, *J. Sens. Actuator Netw.*, 2025, 14(5), 98. <https://doi.org/10.3390/jsan14050098>.
- Elevation Models, Shadows, and Infrared: Integrating Datasets for Thermographic Leak Detection, *Remote Sens.*, 2025, 17(14), 2399. <https://doi.org/10.3390/rs17142399>.
- Hangfei Zhang *et al.* 2025. Review of the Approaches for Urban Underground Infrastructure Inspections Based on UAVs, *Buildings*, 15(13), 2230. <https://doi.org/10.3390/buildings15132230>.
- ISO 22320:2018. Security and resilience — Emergency management — Guidelines for incident management.
- Lun Quan *et al.*, 2020. Survey of UAV motion planning. <https://doi.org/10.1049/iet-csr.2020.0004>.
- Multi-Component Remote Sensing for Mapping Buried Water Pipelines, *Remote Sens.*, 2025, 17(12), 2109. <https://doi.org/10.3390/rs17122109>.
- Path Planning and Control of a UAV Fleet in Bridge Management Systems. *Remote Sens.*, 2022, 14(8), 1858. <https://doi.org/10.3390/rs14081858>.
- Photovoltaics Plant Fault Detection Using Deep Learning Techniques, *Remote Sens.*, 2022, 14(15), 3728. <https://doi.org/10.3390/rs14153728>.
- Power Line Extraction and Reconstruction Methods from Laser Scanning Data: A Literature Review, *Remote Sens.*, 2023, 15(4), 973. <https://doi.org/10.3390/rs15040973>.
- Sha Zhu *et al.* 2024. A Deep-Learning-Based Method for Extracting an Arbitrary Number of Individual Power Lines from UAV-Mounted Laser Scanning Point Clouds. *Remote Sens.*, 16(2), 393. <https://doi.org/10.3390/rs16020393>.
- Solar Photovoltaic Hotspot Inspection Using Unmanned Aerial Vehicle Thermal Images at a Solar Field in South India, *Remote Sens.*, 2023, 15(7), 1914. <https://doi.org/10.3390/rs15071914>.
- Survey of Path Planning for Aerial Drone Inspection of Multiple Moving Objects, *Drones* 2024, 8(12), 705. <https://doi.org/10.3390/drones8120705>.
- Tiago Santos *et al.*, n.d. UAV Visual and Thermographic Power Line Detection Using Deep Learning, *Sensors* 2024, 24(17), 5678. <https://doi.org/10.3390/s24175678>.
- UAV Photogrammetry for Concrete Bridge Inspection Using Object-Based Image Analysis (OBIA), *Remote Sens.*, 2020, 12(19), 3180. <https://doi.org/10.3390/rs12193180>.
- Xin Peng *et al.*, n.d. An autonomous Unmanned Aerial Vehicle exploration platform with a hierarchical control method for post-disaster infrastructures, *Computing in Civil Engineering* 2023. <https://doi.org/10.1061/9780784485248.002>.