

Towards Trustworthy AI-Assisted Drone Evidence Collection

for Security and Digital Forensic Investigations

Core principle
AI confidence $\neq$ evidential trust

Rishabh Choudhary • Student ID: 2434332

Supervisor: Dr. Dipo Dunsin • UWTSD London Campus • September 2026

Trustworthy Drone Evidence Framework (TDEF): integrity • provenance • reliability • robustness • human review

Introduction

- The usage of drones is growing in security, emergency operations and surveillance missions.
- They create video footage, frames, timestamp data and telemetry that may be evidentially relevant.
- AI could be helpful for their analysis, but mere confidence or accuracy doesn't imply the evidential relevance.
- The purpose of the study is to explore how AI-based drone evidence could be evaluated forensically.

Problem

The existing research is very good in individual fields, but lacks in integrated evaluation.

Source data can be intact while AI interpretation is incomplete.

TDEF keeps every trust dimension visible instead of hiding it in one score.

Digital forensics

UAV cybersecurity

AI detection

Human review

Audit trail

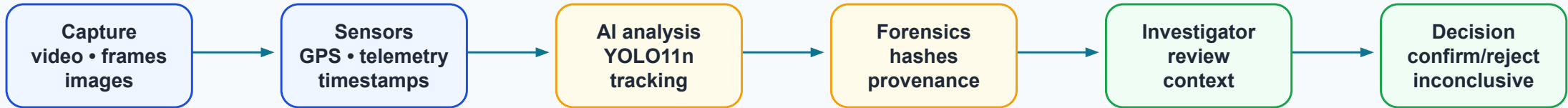
Existing Research Limitations & Identified Gaps

Previous research is mature, but mostly fragmented across separate technical domains.

Research strand	Contribution	Limitation	TDEF response
Drone forensics	Acquisition, preservation and artefact recovery	Limited AI reliability + robustness assessment	Preserves source evidence and adds 8 trust dimensions
UAV cybersecurity	Spoofing, jamming and telemetry threats	Attack-focused; often separated from visual evidence	Uses cross-source cyber-forensic integrity checks
Computer vision benchmarks	Detection, tracking and model performance	Precision/confidence can hide missed objects	Separates reliability and detection completeness
Chain of custody / trustworthy AI	Integrity, traceability and accountability principles	No unified investigator-facing workflow	Builds an explainable TDEF trust profile

Gap codes addressed: G1 workflow fragmentation • G2 AI performance–trust conflation • G3 completeness • G4 robustness • G5 temporal trust • G6 cyber-forensic validation • G7 provenance/audit • G8 multidimensional integration

Why TDEF?



Key issue: each stage adds value, but also adds uncertainty.

- Hash verification ensures byte-level integrity, not sensor veracity.
- Object detection identifies objects, but not whether the scene is complete.
- Tracking gives us algorithmic correlation, but not verified physical identification.
- Telemetry anomalies point to consistency issues, not automatic attack identification.

TDEF considers the trustworthiness of evidence in multi-dimensional forensic terms, rather than one AI assessment.

Methodology

Drone-derived digital evidence
Video • Images • Frames • Telemetry

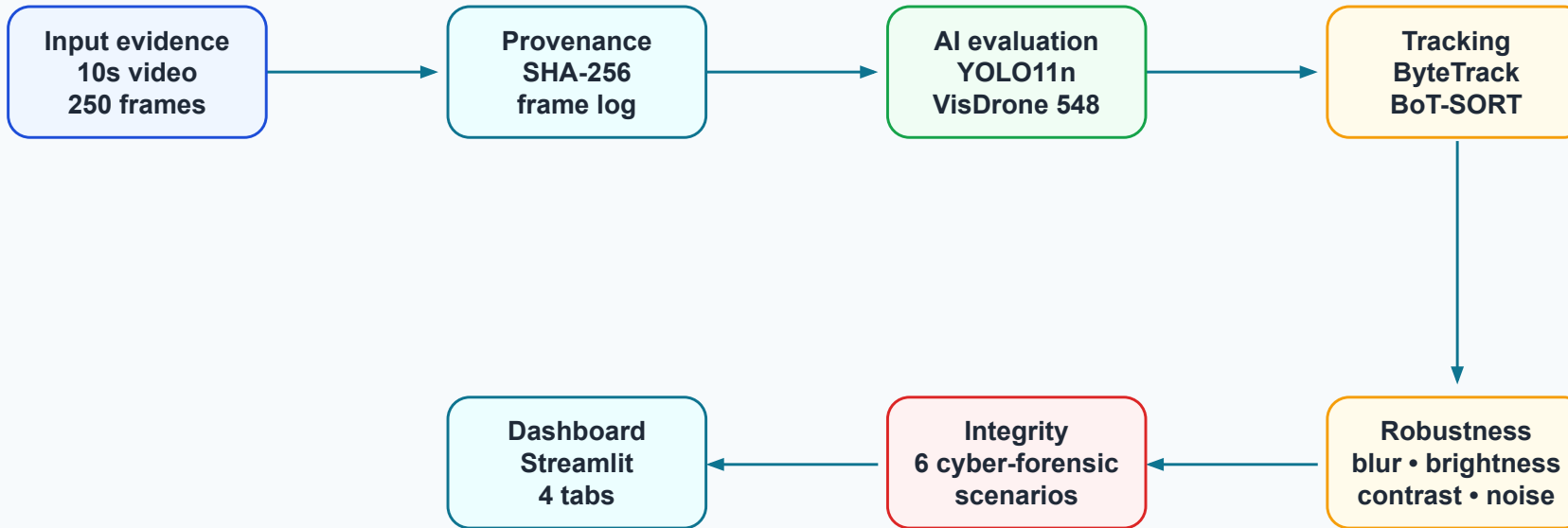
- 1 Evidence acquisition & provenance
Evidence ingestion • SHA-256 hashing • Metadata extraction • Frame provenance
- 2 AI-assisted evidence analysis
Object detection (YOLO11n) • Temporal reconstruction • ByteTrack / BoT-SORT
- 3 Trust & robustness validation
Environmental robustness tests • Cyber-forensic integrity checks • Cross-source consistency
- 4 Multidimensional TDEF assessment
Integrity • Provenance • AI reliability • Completeness • Temporal • Environmental • Cyber status
- 5 Investigator decision support
Explainable trust profile • Human review disposition • Confirm / Reject / Inconclusive

Outcome: An explainable evidence-trust profile with a human re-evaluation stance.

Methodological logic

- 1 Never boil down trust into one arbitrary score.
- 2 Always keep visible the limits as reasons to re-evaluate.
- 3 Support the decision making of investigators and not replace them.

Implementation



- Repository data is provided through GitHub in the development phase.
- Every module provides structured output for the sake of reproducibility.
- The dashboard presents raw outputs of experiments in the explainable forensic workflow.

Investigator-facing implementation: Evidence Overview • AI & Environmental Analysis • Telemetry & Forensic Integrity • Evidence Trust Assessment

Results I — Evidence Overview and Provenance

TDEF Investigator Dashboard

Explainable assessment of AI-assisted drone-derived digital evidence

Evidence Overview AI & Environmental Analysis Telemetry & Forensic Integrity Evidence Trust Assessment

Evidence Overview

This view summarises source evidence, provenance, integrity verification and processing traceability.

Video Frames	Frame Rate	Derived Provenance Frames	Audit Events
250	25.0 FPS	10	19

Integrity and Provenance

Cryptographic Integrity

Status: VERIFIED

Source video and derived-frame hashes are consistent with the recorded provenance information.

Evidence Provenance

Status: VERIFIED

The derived-frame records preserve a consistent source-video and temporal provenance chain.

Source frames

250

Frame rate

25 FPS

Provenance frames

10

- Source video SHA-256 matched recorded provenance.
- 10/10 derived-frame hashes verified.
- No missing provenance frames or frame-hash mismatches.
- Frame timestamps remained monotonic.

Cryptographic Integrity: VERIFIED

Evidence Provenance: VERIFIED

Dashboard evidence overview

Interpretation: the source-to-derived evidence chain was preserved before AI interpretation.

Results II — AI Detection Reliability and Completeness

AI & Environmental Analysis

Ground-truth object-detection, temporal reconstruction and environmental robustness results.

VisDrone Ground-Truth Baseline [↗](#)

Precision	Recall	F1-score	Images Evaluated
70.96%	15.93%	26.02%	548

Precision must not be interpreted without recall. Low recall indicates that the absence of an AI detection cannot establish object absence.

Per-Class Detection Performance

Class	TP	FP	FN	Precision (%)	Recall (%)	F1 (%)
bicycle	2	4	1276	33.33	0.16	0.31
bus	40	61	207	39.6	16.19	22.99
car	4439	1602	9541	73.48	31.75	44.34
motorcycle	60	46	4822	56.6	1.23	2.41
person	970	295	12952	76.68	6.97	12.77
truck	74	278	674	21.02	9.89	13.45

VisDrone evaluation and per-class variation

Images evaluated

548

Precision

70.96%

Recall

15.93%

F1-score

26.02%

Comparable GT objects

35,057

- 7,871 predictions; 5,585 true positives.
- 29,472 false negatives show low detection completeness.
- Per-class results varied strongly across bicycles, vehicles and people.
- Absence of AI detection cannot be treated as object absence.

AI Detection Reliability: LIMITED
Detection Completeness: LOW

Results III – Temporal Reliability and Environmental Robustness

Temporal Reconstruction

Temporal Reconstruction

Tracker	Unique Tracks	Mean Continuity (%)	Median Continuity (%)	High-Continuity Tracks	Fragmented Tracks
ByteTrack	78	91.78	100	62	1
BoT-SORT	75	92.53	100	58	0

Track identifiers represent algorithmic associations and must not be interpreted as verified physical identity.

ByteTrack mean continuity
91.78%

BoT-SORT mean continuity
92.54%

Temporal Reliability: **STRONG**

Track IDs = algorithmic associations, not verified physical identity.

Key point: degradation sometimes increased precision while recall/F1 fell — precision alone can mislead.

Environmental Robustness



Environmental degradation reduced F1 across all tested conditions. Some conditions increased precision while recall decreased, demonstrating why precision alone can produce a misleading impression of reliability.

- Blur: -5.04 pp F1
- Low brightness: -1.85 pp
- Low contrast: -4.11 pp
- Noise: -11.13 pp F1

Environmental Robustness:
DEGRADED

Results IV — Cyber-Forensic Integrity Validation

Telemetry & Forensic Integrity

Controlled cyber-forensic validation of evidence integrity and consistency.

Baseline Status

Controlled Scenarios

Expected Indicators Detected

VERIFIED

6

6/6

The six scenarios are controlled functional validations. The result must not be interpreted as 100% real-world cyberattack detection accuracy.

Controlled Manipulation Results

Scenario	Artifact Type	Forensic Flags	Human Review
Gps Spoofing Simulation	Telemetry Evidence	Telemetry Hash Mismatch, Location Integrity Anomaly	Yes
Timestamp Manipulation	Telemetry Evidence	Telemetry Hash Mismatch, Timestamp Sequence Anomaly	Yes
Telemetry Log Tampering	Telemetry Evidence	Telemetry Hash Mismatch	Yes
Video Telemetry Time Mismatch	Telemetry Evidence	Telemetry Hash Mismatch, Cross Source Timeline Mismatch	Yes
Evidence Hash Mismatch	Derived Visual Evidence	Frame Hash Mismatch	Yes
Flight Path Discontinuity	Telemetry Evidence	Telemetry Hash Mismatch, Location Integrity Anomaly, Trajectory Discontinuity	Yes

Telemetry and forensic integrity dashboard

Baseline status

VERIFIED

Controlled scenarios

6

Expected indicators

6/6

- GPS/location manipulation
- Timestamp manipulation
- Telemetry log tampering
- Cross-source time mismatch
- Derived-frame hash mismatch
- Flight-path discontinuity

Cyber-Forensic Integrity: VERIFIED

Controlled functional validation — not a claim of 100% real-world cyberattack detection accuracy.

TDEF Presentation | Rishabh Choudhary

10

Results V — Integrated TDEF Trust Profile and Human Review

Evidence Trust Assessment

TDEF retains individual forensic trust dimensions rather than collapsing them into an arbitrary numerical score.

Overall TDEF disposition: **HUMAN REVIEW REQUIRED**

Multidimensional Trust Profile

Cryptographic Integrity

Status: VERIFIED

Source video and derived-frame hashes are consistent with the recorded provenance information.

Evidence Provenance

Status: VERIFIED

The derived-frame records preserve a consistent source-video and temporal provenance chain.

AI Detection Reliability

Status: LIMITED

Predictions show substantially greater precision than detection completeness. AI detections may assist investigation but should not be treated as an exhaustive representation of scene content.

Detection Completeness

Status: LOW

More comparable ground-truth objects were missed than correctly matched. Absence of an AI detection must therefore not be interpreted as evidence of object absence.

Temporal Reliability

Status: STRONG

Both evaluated tracking methods achieved mean continuity above the experiment's 90% high-continuity reference level.

Limitation: Tracking identifiers represent algorithmic associations and must not be interpreted as verified object identity.

Environmental Robustness

Status: DEGRADED

Every tested visual degradation reduced F1-score relative to the original imagery. Environmental quality therefore materially affects detection completeness.

Cyber-Forensic Integrity

Status: VERIFIED

The clean cyber-forensic control remained verified. In controlled manipulation scenarios, the framework produced the intended integrity or consistency indicator.

Limitation: Controlled simulations validate framework functionality and explainability; they do not establish real-world cyberattack detection accuracy or attribution.

Auditability

Status: VERIFIED

Major TDEF analysis stages have corresponding structured successful audit events.

Multidimensional trust profile

Dimension	Final status
Cryptographic Integrity	VERIFIED
Evidence Provenance	VERIFIED
AI Detection Reliability	LIMITED
Detection Completeness	LOW
Temporal Reliability	STRONG
Environmental Robustness	DEGRADED
Cyber-Forensic Integrity	VERIFIED
Auditability	VERIFIED

Overall disposition: **HUMAN REVIEW REQUIRED**

- Review triggers:
- AI reliability
 - Detection completeness
 - Environmental robustness

Future Work and Conclusion

Future work

A multimodal forensics dataset for synchronized UAV video, telemetry, timestamps, and inertial data. Extension of flight sequences to evaluate temporal identity switching. Presenting the evidence in a degraded state due to environmental influences, including weather, low-light imaging, motion blur, and compression. An experimental forensics setup for the cyber-forensic UAV was used to simulate inconsistent telemetry and positioning data. Evaluation by an independent investigator using decision thresholding.

Conclusion

TDEF does not claim that AI alone provides evidence trust. Instead, TDEF identifies various features such as integrity, provenance, trustworthiness, completeness, temporal coherence, resilience, cyber-consistency, and audibility. The most important contribution is the development of an explainable forensic process for investigators.

**Trustworthy Drone Evidence
Framework (TDEF)
AI confidence $\neq$ evidential trust**

**Final message
Trustworthy drone evidence requires more
than detection accuracy.**

Thank You

Questions?

Rishabh Choudhary
MSc Cybersecurity and Digital Forensics
University of Wales Trinity Saint David