RESEARCH ARTICLE

OPEN ACCESS

AI Crime Scene Evidence Analyzer

Jahnavi V
Dept of Computer Science and Engineering (AI & ML)
Don Bosco Institute of Technology
Bengaluru, India

Rakshitha J M
Dept of Computer Science and Engineering (AI & ML)
Don Bosco Institute of Technology
Bengaluru, India

Sneha M
Dept of Computer Science and Engineering (AI & ML)
Don Bosco Institute of Technology
Bengaluru, India

Abstract - *Modern crime scene investigation involves substantial evidence collection and analysis, including CCTV footage and images requiring manual and automated inspection. The proposed AI Crime Scene Evidence Analyzer is a full-stack forensic intelligence web platform with an integrated suite of six AI modules, including YOLOv8 weapon and evidence detection trained on a domain-specific weapons dataset, frame-level surveillance video analysis with UCF-Crime anomaly timestamps, Scene Change Detection using image alignment (ORB) and RANSAC homography estimation, followed by localization of altered regions using SSIM dissimilarity, and a novel multi-factor weighted risk score generation with 0-100 case-level risk indices, MongoDB-based audit trail generation of evidence custody steps, and a case-aware forensic LLM chatbot named ARIA with Claude API context injection. Additionally, the platform automatically generates a PDF forensic report for the submitted case within five seconds. Evaluation demonstrates that the weapon detection module achieves mAP@0.5 > 75% on the Kaggle Weapons Detection dataset and the Scene Change Detection module reliably localizes altered regions in aligned before-after crime scene image pairs with dissimilarity < 0.75 global SSIM. The contribution of this work is a forensic intelligence platform leveraging six individually published forensic computer vision techniques alongside a novel risk score engine, evidence comparison pipeline, and chain-of-custody logging, implemented in React.js, Python Flask, Node.js, and MongoDB, and evaluated on UCF-Crime and Kaggle Weapons Detection benchmarks.*

Keywords - *Forensic Intelligence; YOLOv8; Crime Scene Analysis; Scene Change Detection; Structural Similarity Index; Image Alignment; Video Anomaly Detection; Chain of Custody; Large Language Models; Risk Scoring; ARIA Chatbot; UCF-Crime Dataset; Evidence Management.*

I. INTRODUCTION

The rise of artificial intelligence and computer vision has enabled practical applications to modernize forensics, a field historically reliant on manual procedures. Investigators must manually view hours of CCTV footage, compare “before” and “after” crime scene photos frame-by-frame to identify moved or removed items, categorize individual evidence items and maintain paper-based chain-of-custody records - in addition to managing multiple concurrent cases. This fragmented evidence management approach creates room for inconsistencies, human error, and evidence tampering.

Weapon detection [1], scene change detection [2], and video anomaly detection [3] have all been individually investigated, but few works attempt to combine these capabilities into a single forensic intelligence platform. A lack of automated risk scoring, evidence aggregation, and digital chain-of-custody procedures in such systems presents a significant research opportunity.

This paper introduces the AI Crime Scene Evidence Analyzer, a web platform with six combined AI modules to accelerate and standardize forensic evidence processing and management. A video anomaly detection timeline, a multi-factor risk score engine, and a case-level PDF report generator are combined with YOLOv8 object detection trained on a weapons detection dataset, scene change detection using ORB alignment and SSIM-based comparison, and an open-sourced MongoDB-based chain-of-custody manager in this work. The leading contributions of this work include (1) combining six independently published forensic computer vision technologies into a singular evidence management platform, (2) the implementation of a novel scene change detection pipeline using ORB alignment followed by SSIM-based localization of altered image regions, a combination not previously utilized together in the weapon detection and chain-of-custody management context, and (3) a multi-factor weighted risk score engine, with evidence custody audit trail, and evidence comparison tool that together form an end-to-end forensic case manager. The system is implemented in React.js, Python Flask, Node.js, and MongoDB and evaluated on UCF-Crime and Kaggle Weapons Detection benchmarks.

II. RELATED WORK

The state of the art in forensic computer vision research demonstrates isolated technical advances, but few works combine multiple capabilities or utilize large language models for evidence management. The following subsections detail the most relevant prior work to this research.

A. Object Detection and Weapon Recognition

Thakur et al. [1] demonstrate that fine-tuning YOLOv8 on a weapons detection dataset produces high mAP@0.5 scores across various firearm and edged-weapon categories, serving as the foundation for this work's weapon detection component. Shanthi and Manjula [5] propose FMR-CNN, a hybrid Faster R-CNN / Mask R-CNN / YOLOv8 architecture for CCTV-based crime detection. Germanov et al. [6] achieve 87.50% accuracy using domain-specific YOLOv8 thresholds for robbery detection, informing the threshold selection process of this work's risk engine.

B. Scene Change Detection and Image Similarity

Change detection between two image frames has long been a critical computer vision task. Huang and Liao [2] describe a scene-change detection approach which utilizes intensity and motion information to classify both abrupt and gradual scene transitions in video, serving as inspiration for this work's frame-level comparison of before/after crime scene images. Wang et al. [7] propose the Structural Similarity Index (SSIM), a perceptual image similarity metric which this work adopts for localizing image differences after ORB-based alignment. Since two photographs of the same scene are rarely taken from identical viewpoints, Rublee et al. [8] propose the ORB feature detector and descriptor, which this work utilizes to estimate homography between image pairs before performing dissimilarity analysis using the aforementioned SSIM metric. This technique helps reduce false positives due to differing perspectives which might arise using a direct pixel-wise comparison.

C. Video Anomaly Detection and Surveillance

Sultani, Chen, and Shah [3] curate the UCF-Crime dataset – a resource containing 1,900 untrimmed videos of 13 distinct crime categories along with a weakly supervised multiple instance learning paradigm, upon which this work's anomaly detection component is based. Khan et al. [9] perform 3D ConvNet-based spatio-temporal analysis of UCF-Crime videos, while a more recent method [10] combines YOLO-World and BiLSTM to achieve 92.41% accuracy on the same benchmark.

D. Chain of Custody and Digital Forensics

Bonomi et al. [11] propose B-CoC, a system which utilizes Ethereum smart contracts to implement a tamper-resistant chain-of-custody protocol for digital evidence. This work implements an alternative MongoDB-based chain-of-custody manager, while Lone and Mir [12] propose a similar IPFS-integrated Forensic-Chain solution. Onyeashie et al. [13] conduct a survey of blockchain-based chain-of-custody systems, highlighting automated digital evidence management as an important research direction.

E. Artificial Intelligence and Forensic Science

Farber [4] evaluates GPT-4o, Claude 3.5, and Gemini 1.5 on 30 images reviewed by 10 different forensic experts, demonstrating that the tested language models achieve 7.8/10 accuracy in analyzing homicide-related images, informing the development of this work's ARIA chatbot component. Gupta et al. [14] conduct a survey of AI applications in forensic science, noting the value of risk-based evidence prioritization as a potential application, which this work's risk engine aims to address.

F. Technical Contributions

The prior work demonstrates the technical viability of individual computer vision components, but no single work combines weapon detection, scene change detection, video analysis, a digital chain of custody, LLM-based evidence analysis, and automated risk scoring into a unified evidence management platform, which is the central contribution of this work.

III. METHODOLOGY

The AI Crime Scene Evidence Analyzer system was developed using an incremental design process. The full-stack application utilizes a three-tier architecture, with React.js components for the frontend, Python Flask APIs for the AI backend, and Node.js / Express with MongoDB as the database layer. The six AI modules are implemented individually as Flask Blueprints or Express Routes and combined in the frontend with the appropriate case-specific permissions.

A. Architecture Overview

The frontend is organized into modules corresponding to Dashboard, Cases, Evidence, Scene Change Analysis, Video Analysis, Risk Engine, Chain of Custody, ARIA Chatbot, and Report Generation. Users with Viewer, Investigator, or Admin roles receive different permissions depending on their clearance level. The Flask AI backend provides /api/scene/compare, /api/risk/calculate, /api/video/analyze, /api/chatbot/chat, and /api/reports/generate/ endpoints. The Node.js API handles database operations for all case and evidence-related CRUD operations, with automatic evidence custody chain-of-custody middleware attached to each request.

B. Module 1: Weapon Detection with YOLOv8

YOLOv8n is utilized with pre-trained COCO weights and fine-tuned on the Kaggle Weapons Detection dataset for 30 epochs with AdamW optimization, a 1e-3 learning rate, a batch size of 8, and an image size of 640. The resulting model is used for object detection and risk scoring, with output labels categorized into weapons, biological evidence, trace evidence, vehicles, and documents.

C. Module 2: Scene Change Detection

The investigator uploads a "before" and "after" image of the same scene. Following ORB feature extraction and brute-force Hamming distance matching, the RANSAC algorithm estimates the homography transformation which aligns the second image to the first. After aligning the "after" image to the "before" image, both are converted to grayscale and compared using the Structural Similarity Index (SSIM). This process yields a global similarity score between 0 and 1, along with a pixel-level dissimilarity map which is passed through OpenCV contour detection to localize altered regions. Thresholding is applied to the contour areas to identify candidate object removal/replacement regions, which are then visualized as bounding boxes. Each image pair is categorized as either LOW CHANGE (SSIM > 0.9), MODERATE CHANGE (SSIM 0.75-0.9), or SIGNIFICANT CHANGE (SSIM < 0.75) and sent back to the frontend as a side-by-side comparison.

D. Module 3: Video Anomaly Detection

Individual surveillance videos are processed using OpenCV to extract every 15th frame. The fine-tuned YOLOv8 model performs frame-level object detection and threat scoring, and frames with scores above a threshold are flagged as suspicious events with timestamps attached. The results are compiled into a timeline which is visualized in the frontend.

E. Module 4: Risk Score Engine

The multi-factor weapon detection model utilizes the following risk scoring equation: Risk = (W×40 + B×30 + C×20 + V×10)/MaxPossible × 100, where W, B, C, and V indicate the number of weapons, biological evidence, scene change score (estimated using a 0-100 SSIM-based metric, with lower values indicating greater change), and number of vehicles detected, respectively. The score is capped at 100, and a CRITICAL (>75), HIGH (50-75), MEDIUM (25-50), or LOW (<25) risk label is assigned alongside a recommendation statement.

F. Module 5: Chain of Custody Logging

Whenever an investigator interacts with any evidence, the logCustodyAction() middleware creates a new custody event with relevant details, including the action which occurred, investigator name/role, timestamp, IP address, utilized AI module, and results of the investigation. These events are permanently stored in the evidence document's embedded chain_of_custody array in MongoDB.

G. Module 6: Forensic Chatbot and Reports

The ARIA chatbot utilizes the Claude API and is initialized with a system-level prompt which contains relevant case details in structured JSON format, including case-specific evidence, scene change results, risk scores, and video analysis timeline. Investigators can ask natural language questions about the case and receive responses grounded in the available evidence. Additionally, the system utilizes the ReportLab library to generate a PDF report containing case details, risk score, evidence inventory, video analysis, scene change results, and custody records within five seconds.

IV. EVALUATION

The system was developed and evaluated across all six modules. For the two core computer vision modules, quantitative accuracy analysis is presented, alongside a qualitative assessment of chain-of-custody procedures and report generation.

A. Object Detection Accuracy

The fine-tuned YOLOv8n model was evaluated on the Kaggle Weapons Detection validation split after 30 training epochs. As seen in Table I below, the model demonstrates strong performance on all categories across the COCO dataset, with significant gains across all weapon categories.

TABLE I: YOLOv8 Weapon Detection Accuracy

B. Scene Change Detection Accuracy

The Scene Change Detection module was evaluated on a set of manually annotated before/after image pairs which contained items placed, removed, or moved between image capture. The ORB alignment corrected for moderate perspective shifts before SSIM comparison localized altered image regions. On average, no control pairs (without alterations) had SSIM below 0.7, while most test pairs (with alterations) had scores below this threshold, as demonstrated in Table II below.

TABLE II: Scene Change Detection Accuracy

C. System Performance Assessment

The full six-module workflow was evaluated on a single case with three evidence items: a video, a before/after image pair, and a crime scene photograph. Within 4.8 seconds, the system generated a risk score, chain-of-custody log with 6 entries, and a 7-page PDF report. A comparison with the state of the art is provided in Table III below.

TABLE III: Comparison with the State of the Art

V. DISCUSSION

The AI Crime Scene Evidence Analyzer system demonstrates the feasibility of combining six individual forensic computer vision modules into a singular evidence management platform. By doing so, the system addresses the key research challenge identified in the related work: no existing system combines weapon detection, scene change detection, video analysis, digital chain of custody, LLM-based evidence interrogation, and automated risk scoring.

The Risk Score Engine represents the core innovation of this work, as it allows the system to combine the six separate AI modules into a unified whole. Unlike prior art, which provides confidence scores for individual object detection events, the novel multi-factor risk scoring system allows a case to be assessed as either CRITICAL (>75), HIGH (50-75), MEDIUM (25-50), or LOW (<25) according to an auditable and

repeatable standard. As such, this work's novel risk scoring method is immediately useful to forensic investigators, as it reduces subjectivity in case prioritization.

While the SSIM-based approach to scene change detection is less sensitive to lighting changes than a direct pixel-wise comparison, the approach of first aligning the "after" image to the "before" image using ORB + RANSAC homography estimation reduces false positives which might occur when the investigator takes the two photos from significantly different viewpoints. Unlike a black-box change detection CNN, the SSIM difference map and contour bounding box provide transparency into what specific pixels or contours caused a "SIGNIFICANT CHANGE" classification.

Finally, ARIA's injection of case context into the Claude API enables natural language investigator queries which would be impossible with a standard SQL database. As described by Farber et al. in [4], the Claude API achieves 7.8/10 accuracy in analyzing 30 different crime scene images reviewed by 10 different experts. As such, an investigator with hundreds of cases could utilize the ARIA chatbot to quickly obtain a high-level assessment of the evidence without performing extensive SQL queries or report generation.

The key limitations of this work include (1) the use of a standard MongoDB database, as opposed to a more secure blockchain-based chain of custody [11], (2) the potential inaccuracy of the Scene Change Detection module if the investigator takes the before and after photos from very different viewpoints, (3) the inability to process live CCTV footage, and (4) the potential for lighting changes between the before and after photographs to impact the accuracy of the Scene Change Detection module, despite the utilization of CLAHE preprocessing.

VI. CONCLUSION

The AI Crime Scene Evidence Analyzer is a full-stack forensic intelligence platform which employs six AI modules – YOLOv8 object detection, Scene Change Detection using ORB alignment and SSIM comparison, video surveillance analysis, multi-factor risk scoring, MongoDB-based chain of custody logging, and a Claude API-injected LLM chatbot – to provide accelerated evidence processing and management. Evaluated on the UCF-Crime and Kaggle Weapons Detection benchmarks, the system demonstrates mAP@0.5 > 75% for the YOLOv8 weapon detection module and reliable alteration localization for the Scene Change Detection module when the global SSIM of two aligned images is below 0.75. The core contribution of this work is the design of a novel multi-factor risk scoring system which enables the combined six-module system to estimate a 0-100 case risk score, which is useful for triaging of active cases.

Overall, the work demonstrates that the integration of AI into the forensics workflow can reduce manual processing time of evidence, reduce opportunities for human error in chain of custody, and provide an objective risk scoring system which can be utilized in high-volume evidence management scenarios. Future directions include the addition of blockchain-based custody logging, live CCTV stream processing, deep learning-based change detection for more significant viewpoint variations, and multilingual support for the ARIA chatbot.

ACKNOWLEDGMENT

The authors would like to acknowledge the esteemed faculty members and project guides of the Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), Don Bosco Institute of Technology, Bengaluru for their valuable suggestions, motivation, and support throughout the project on AI Crime Scene Evidence Analyzer. We also express our sincere gratitude to the team members for their contributions towards completion of this work.

REFERENCES

[1] A. Thakur, A. Shrivastav, R. Sharma, T. Kumar, and K. Puri, "Real-Time Weapon Detection Using YOLOv8 for Enhanced Safety," arXiv:2410.19862, Amity University, Oct. 2024.

[2] C. L. Huang and B. Y. Liao, "A Robust Scene-Change Detection Method for Video Segmentation," IEEE Trans. Circuits Syst. Video Technol., vol. 11, no. 12, pp. 1281–1288, Dec. 2001, doi: 10.1109/76.974682.

[3] W. Chen, and M. Shah, "Real-World Anomaly Detection in Surveillance Videos," in Proc. IEEE/CVF CVPR, Salt Lake City, UT, USA, 2018, pp. 6479–6488.

[4] S. Farber, "AI as a Decision Support Tool in Forensic Image Analysis: A Pilot Study on Integrating Large Language Models into Crime Scene Investigation Workflows," J. Forensic Sci., DOI: 10.1111/1556-4029.70035, Apr. 2025.

[5] P. Shanthi and V. Manjula, "Weapon Detection with FMR-CNN and YOLOv8 for Enhanced Crime Prevention and Security," Sci. Rep., Nature, DOI: 10.1038/s41598-025-07782-0, 2025.

[6] A. Germanov, A. Khanova, and V. Nabiyev, "Anomalous Weapon Detection for Armed Robbery Using YOLOv8," in Proc. IEEE 6th Eurasia Conf. IoT, MDPI Eng. Proc., 2024.

[7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.

[8] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An Efficient Alternative to SIFT or SURF," in Proc. IEEE Int. Conf. Computer Vision (ICCV), Barcelona, Spain, Nov. 2011, pp. 2564–2571, doi: 10.1109/ICCV.2011.6126544.