Forensic-Grade Preprocessing and Dataset Construction for Real-Time Crime Scene Analysis: Enhancing Evidence Detection Using YOLOv8

Manjur Mahamad Kovadiya
M.Tech in Data Science, Nirma University, Ahmedabad, India
24mcd008@nirmauni.ac.in Dr. Sharada Valiveti
Sr. Associate Professor, Nirma University, Ahmedabad, India
sharada.valiveti@nirmauni.ac.in

Abstract

This paper presents a data-centric deep learning pipeline for automated crime scene analytics using YOLOv8, with emphasis on forensic-grade dataset construction and preprocessing rather than architectural modification. Two publicly available datasets from Roboflow Universe—Weapons (1,035 images) and Bloodstains (326 images)—were merged into a unified YOLO-format dataset through class remapping, label validation, and balanced train/validation/test splitting, resulting in approximately 1,300 annotated images.

To address real-world forensic challenges such as small object size, low contrast, noise, and class imbalance, a multi-stage enhancement pipeline was employed. This includes noise reduction, selective super-resolution using Real-ESRGAN applied only to small evidence regions, and advanced crime-scene-specific data augmentation techniques such as Mosaic, MixUp, copy–paste, fog, shadow, motion blur, and occlusion. In addition, a DCGAN was trained on low-resolution weapon images to generate synthetic samples used as weakly supervised augmentation for improving weapon class recall.

Multiple YOLOv8 variants (nano, small, and medium) were trained and evaluated on the enhanced dataset. The YOLOv8s model achieved the best balance between accuracy and efficiency, obtaining a mean Average Precision of 0.744 at IoU 0.50 and 0.615 across IoU thresholds from 0.50 to 0.95. The trained model was deployed through an interactive Streamlit-based web application supporting real-time inference, visual annotation, and spatial distance measurement between detected evidences.

Experimental results demonstrate that forensic-oriented preprocessing, selective super-resolution, and dataset engineering significantly improve robustness and detection performance in challenging crime scene imagery, making the proposed system a practical assistive tool for preliminary forensic investigation.

Index Terms

Crime Scene Analytics, Forensic Image Processing, Object Detection, YOLOv8, Dataset Engineering, Real-ESRGAN, DC-GAN, Data Augmentation

I. Introduction

Crime scene investigation relies heavily on the rapid and accurate identification of critical visual evidence from photographs captured at the scene. Among the most forensically significant objects are weapons (firearms, knives, rifles, and blunt instruments) and bloodstains or spatter patterns, which provide essential clues about the nature, sequence, and severity of criminal events. Traditionally, forensic examiners manually inspect hundreds to thousands of images—a process that is time-consuming, prone to fatigue-induced errors, and inconsistent across analysts. These challenges are exacerbated when objects are small, distant, partially occluded, poorly illuminated, or degraded due to environmental factors and low-resolution imaging.

Recent advances in real-time object detection, particularly the You Only Look Once (YOLO) family, have demonstrated remarkable performance in complex visual environments. The latest iteration, YOLOv8 [8], offers state-of-the-art accuracy and inference speed through an anchor-free detection head and efficient backbone design, making it well-suited for forensic applications where both precision and deployment speed are critical.

Despite these advances, applying modern detectors directly to crime-scene imagery remains challenging due to (i) scarcity of annotated forensic datasets, (ii) severe class imbalance between common backgrounds and rare evidence, (iii) prevalence of small and low-contrast objects, and (iv) real-world degradation (noise, blur, compression artifacts). While public datasets for general object detection abound, dedicated crime-scene datasets are limited and typically focus on a single evidence type.

This paper presents an end-to-end deep learning pipeline specifically tailored for automated detection of weapons and blood evidence in static crime-scene photographs. Our key contributions are:

- A systematic merging of two high-quality public datasets from Roboflow Universe (1,035 weapon images + 326 bloodstain images) with custom class remapping and balanced splitting.
- A targeted super-resolution enhancement strategy using Real-ESRGAN [9] applied selectively to small objects ($<48\times48$ pixels), preserving annotation integrity while significantly improving feature clarity.

- Extensive domain-specific augmentation and optional DCGAN-based synthetic weapon generation to address class imbalance and simulate real-world forensic variations.
- Comprehensive training and evaluation of YOLOv8-n/s/m variants, achieving robust real-time performance with particular gains on small and occluded evidence.
- A fully functional Streamlit web interface enabling forensic examiners to upload images and obtain instant annotated results with spatial distance measurement between detected objects.

The proposed system serves as an effective assistive tool that accelerates evidence discovery, reduces manual analysis time, and supports preliminary digital forensic workflows without replacing certified human examination.

II. Literature Review

The application of deep learning to forensic science and crime monitoring has advanced significantly, with research focusing on specific detection tasks, contextual analysis, and the development of integrated systems. This review examines key methodologies and identifies persistent gaps that motivate the current work. By observing these contributions collectively, it becomes evident that integrating techniques across detection, contextual understanding, and system-level deployment can yield a more robust, accurate, and fine-tuned forensic model capable of addressing real-world challenges more effectively.

A. Small and Forensic Object Detection

Detecting small forensic objects (e.g., bullet casings, fibers, small weapons) is especially challenging due to low resolution and limited pixel information. Recent SOD (small object detection) literature addresses these problems through several techniques: advanced data augmentation, super-resolution, context modelling, and multi-scale feature fusion. Key datasets used in SOD and related forensic tasks include WIDER FACE, TinyPerson, VisDrone, DOTA, MS-COCO variants and specialized forensic/weapon datasets, which each present their own distribution and real-world challenges.

Data Augmentation. Methods such as *copy-paste*, *adaptive copy-paste*, *copy-reduce-paste*, *tiling*, *scale matching*, and *stitching* are used to increase the effective frequency and variety of small objects in training data. Tiling and focus-then-detect strategies are especially useful for zooming-in on small regions without losing contextual cues. These techniques have been shown to improve detector performance substantially (often several percentage points on targeted benchmarks).

Super-Resolution (SR). SR approaches—frequently GAN-based—reconstruct higher-resolution images from low-quality inputs, restoring edges and shape cues that are critical for small-object recognition. GAN-based SR pairs a generator (upsampler) and a discriminator (quality verifier) and can be used as a pre-processing step to increase downstream detection accuracy for small forensic items.

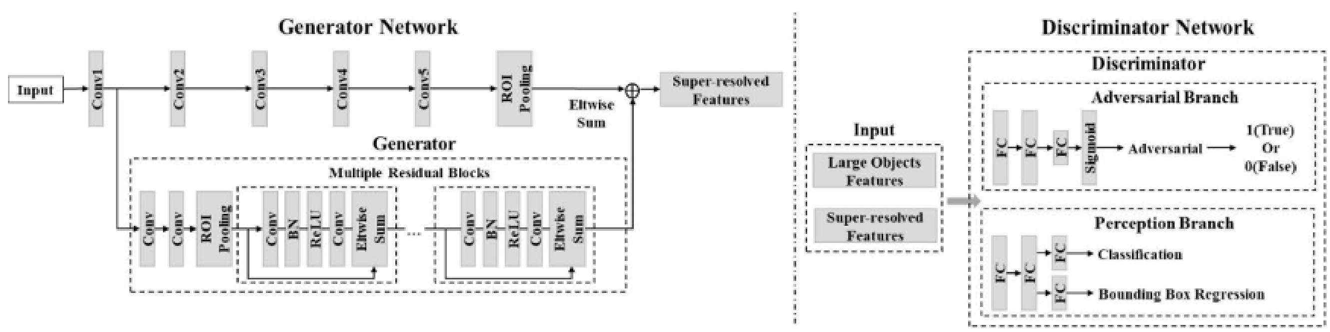

Figure 3. Detail of perceptual GAN network

Fig. 1. GAN-based super-resolution pipeline: a generator upsamples low-resolution patches and a discriminator enforces photorealism and fidelity [2].

B. Contextual Information and Activity Recognition

Small objects often lack sufficient local features; therefore, incorporating context (neighbouring scene, co-occurring objects, object location priors, and temporal cues in videos) improves detection robustness. Approaches include Region-Proposal and ContextNet, Spatial Memory Networks (SMN), and spatio-temporal context modules that combine motion information with appearance. For human-object interaction and activity recognition, large benchmarks like UCF-Crime (video anomaly detection) and HICO-DET (HOI detection) are commonly used to train models that reason about activities and interactions.

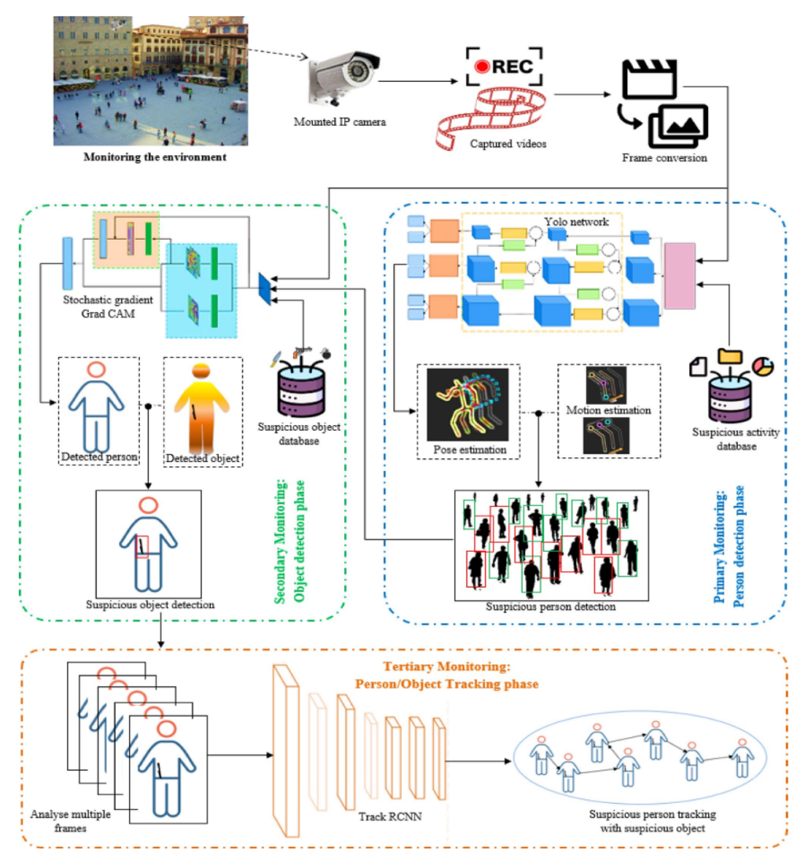

Fig. 2. Three-phase SPOT-CRIME pipeline: (1) person detection (YOLOv8), (2) object correlation (SGT-GradCAM), and (3) multi-frame tracking (Track-RCNN) [1].

C. Multi-scale Representations and Feature Fusion

Convolutional downsampling destroys small-object detail but deep layers provide semantic strength. Multi-scale fusion (FPN, PAN, BiFPN, NAS-FPN, SSD-style hierarchical pyramids) is used to combine shallow high-resolution features with deep semantic features. Hybrid fusion methods (e.g., RHF-Net variants) recursively fuse top-down and bottom-up signals to preserve both localization and class-level semantics for small target detection.

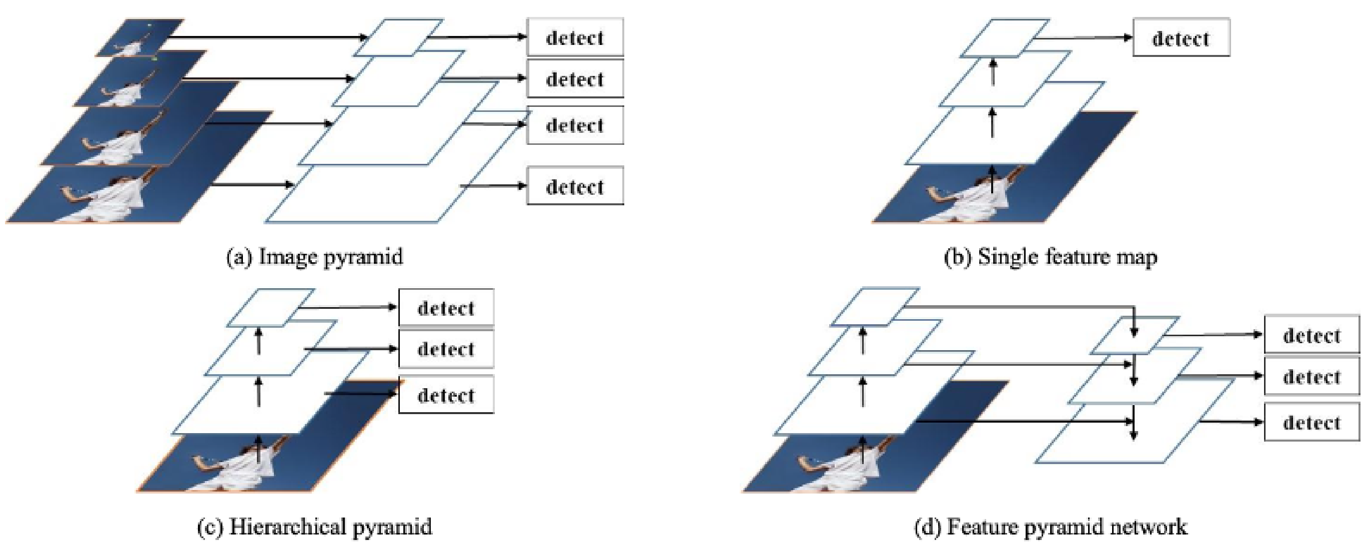

Fig. 3. Multi-scale fusion illustration: (a) image pyramid, (b) single feature map baseline, (c) SSD hierarchical maps, (d) FPN-style top-down fusion [1].

D. Integrated Forensic Systems and Gaps

Several works propose combining detection, reconstruction, and activity modules into an integrated system or portable device. A typical proposed system integrates fingerprint reconstruction (autoencoders/GANs), weapon detection (SSD/YOLO), and human activity recognition (HAR) within an IoT-enabled device or cloud-assisted pipeline. While such proposals demonstrate promise, they face practical limitations: scarce large-scale forensic datasets, heavy computational requirements for real-time operation, deployment costs, and insufficient real-world validation with law enforcement partners.

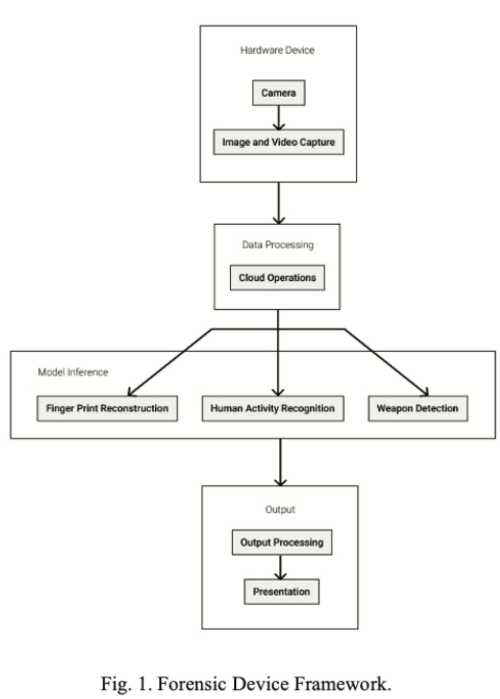

Fig. 4. Framework for a portable AI-powered forensic device integrating sensors, edge processing, and cloud inference for fingerprint reconstruction, weapon detection, and HAR [3].

E. Challenges and Future Directions

Key persistent challenges include:

- **Dataset scarcity and domain gaps:** Lack of large, diverse, and privacy-compliant forensic datasets limits model generalization. (Use of synthetic data alleviates but does not fully solve this.)
- **Computational cost and latency:** Super-resolution and high-capacity fusion networks hurt real-time performance, which is critical for on-site forensic tools.
- **Robustness under adverse conditions:** Low light, occlusion, clutter, and weather degrade detection and tracking accuracy (noted in SPOT-CRIME evaluations).
- **Evaluation metrics and sensitivity:** IoU-based metrics are sensitive to small localization shifts; alternative metrics (e.g., Gaussian-based or NWD-like measures) and improved sample assignment strategies should be investigated.

Opportunity for integration. Observing the strengths and weaknesses across the reviewed literature suggests a clear pathway: combining augmentation strategies, GAN-based SR, multi-scale fusion, context modules, and a light-weight tracking pipeline can produce a unified forensic model that performs better across small-object detection, object–person correlation, and real-time tracking in realistic conditions. This integrative approach motivates the methodology presented in the next section.

III. Preparing the Dataset

The proposed system follows a structured, end-to-end pipeline specifically designed for forensic crime-scene imagery, as illustrated in Fig. 5. The pipeline consists of five major stages: dataset construction and validation, targeted image enhancement, domain-specific data augmentation, model training, and interactive inference for forensic analysis.

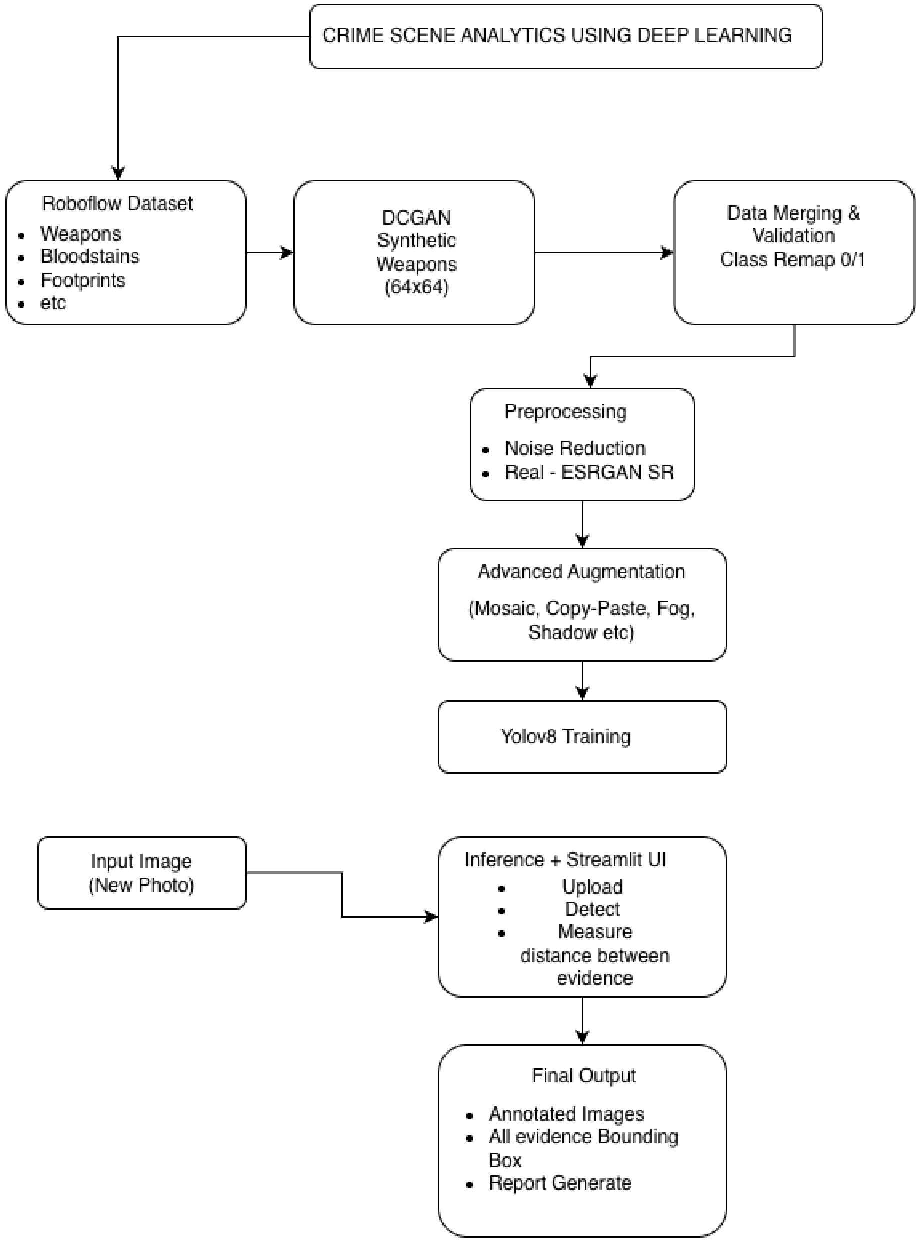

Fig. 5. Overview of the proposed crime-scene analytics pipeline.

A. Dataset Construction and Merging

Two publicly available datasets from Roboflow Universe were utilized in this work:

- **Weapons Dataset** [10]: 1,035 annotated images containing firearms, knives, rifles, pistols, and blunt objects.
- **Bloodstain Dataset** [11]: 326 annotated images containing bloodstains and spatter patterns captured under varied conditions.

A custom Python-based merging pipeline was developed to unify the datasets into a single YOLO-format structure. Class labels were remapped to ensure consistency (*weapon* $\rightarrow$ class 0, *blood* $\rightarrow$ class 1), corrupted images and invalid annotations were automatically filtered out, and duplicate filenames were resolved. An 80%/10%/10% train–validation–test split was generated using stratified sampling, resulting in a final dataset of 1,361 high-quality annotated images.

B. Targeted Super-Resolution Enhancement

Small and distant forensic evidence often lacks sufficient visual detail for reliable detection. To address this challenge, a *selective super-resolution* strategy using Real-ESRGAN [9] was introduced.

For each image, YOLO annotation files were parsed to compute bounding box areas. Objects smaller than 48×48 pixels (approximately 0.56% of a 640×640 image) were identified as small evidence instances. Each such bounding box was expanded by a factor of 1.3 (centered) to preserve contextual information. The cropped region was then enhanced using a $2\times$ Real-ESRGAN model, resized back to its original dimensions using Lanczos interpolation, and seamlessly reintegrated into the lightly denoised full image.

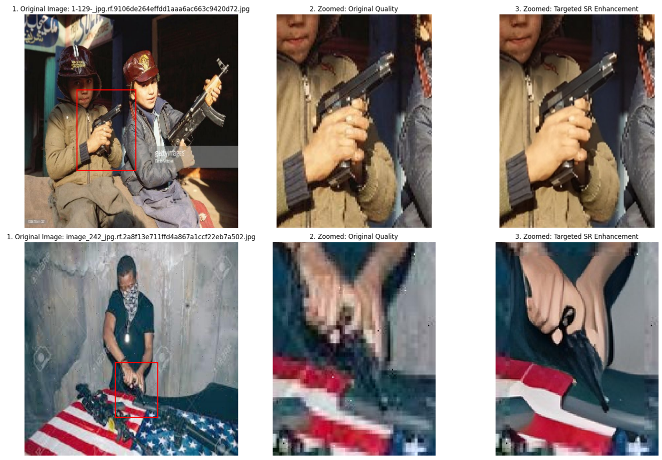

Fig. 6. Selective Real-ESRGAN enhancement applied to small-object regions before reintegration into the original image.

This targeted approach significantly improves feature clarity for tiny knives, distant firearms, and faint blood droplets *without* altering annotation coordinates or introducing domain shift across the entire image. Images not requiring super-resolution undergo only lightweight denoising using Gaussian blur and non-local means filtering.

C. Domain-Specific Data Augmentation

To simulate realistic crime-scene variability and improve generalization, an aggressive augmentation pipeline was applied using the Albumentations framework.

Fig. 7. Crime-scene-specific augmentation: original image (left) versus augmented sample (right).

The applied augmentations include:

- Mosaic and MixUp for complex scene composition
- Copy–paste insertion of weapons into bloodstain scenes
- Random fog, shadows, motion blur, rain effects, and occlusion

- Brightness, contrast, and hue variations

In addition, a Deep Convolutional Generative Adversarial Network (DCGAN) was trained on 64×64 grayscale weapon patches for 500 epochs to generate synthetic weapon samples. These generated images were used as weakly supervised augmentation through copy–paste blending to mitigate weapon class imbalance.

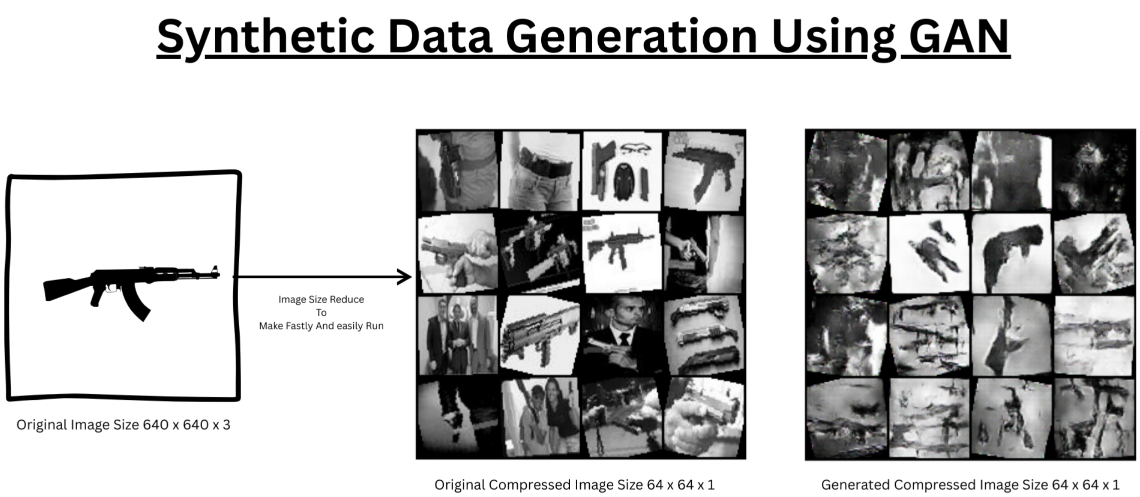

Fig. 8. DCGAN-generated synthetic weapon images (64×64 grayscale) after 500 epochs. Left: real samples; Right: generated samples showing recognizable weapon structures.

D. Model Architecture and Training

Three variants of YOLOv8—nano, small, and medium—were trained on the enhanced dataset for 150 epochs using stochastic gradient descent with cosine learning rate scheduling. Among these, YOLOv8s achieved the best balance between accuracy and computational efficiency, delivering real-time inference performance exceeding 90 FPS on an NVIDIA RTX 3060 GPU while maintaining strong detection accuracy for small and degraded objects.

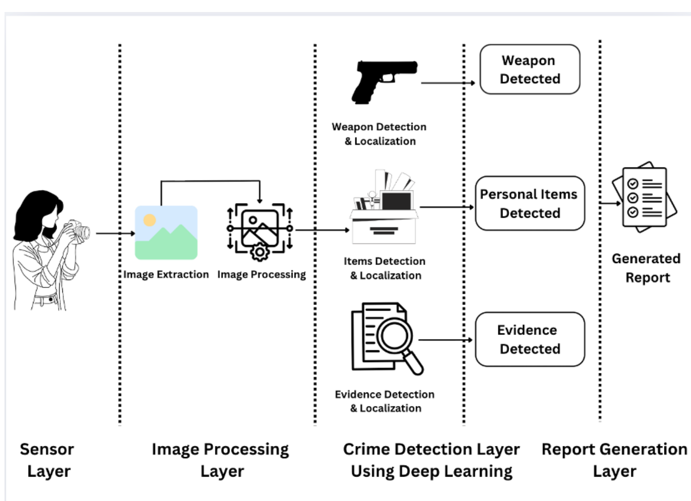

Fig. 9. Working flow diagram of the proposed crime scene analytics system.

E. Interactive Forensic Interface

To facilitate practical forensic usage, an interactive Streamlit-based web application was developed. The interface enables non-technical forensic examiners to upload crime-scene images and obtain instant annotated outputs with color-coded bounding boxes (red: weapon, blue: blood). Additionally, the system computes Euclidean distances between detected object centroids, providing useful spatial relationship cues such as the proximity between a weapon and a blood pool.

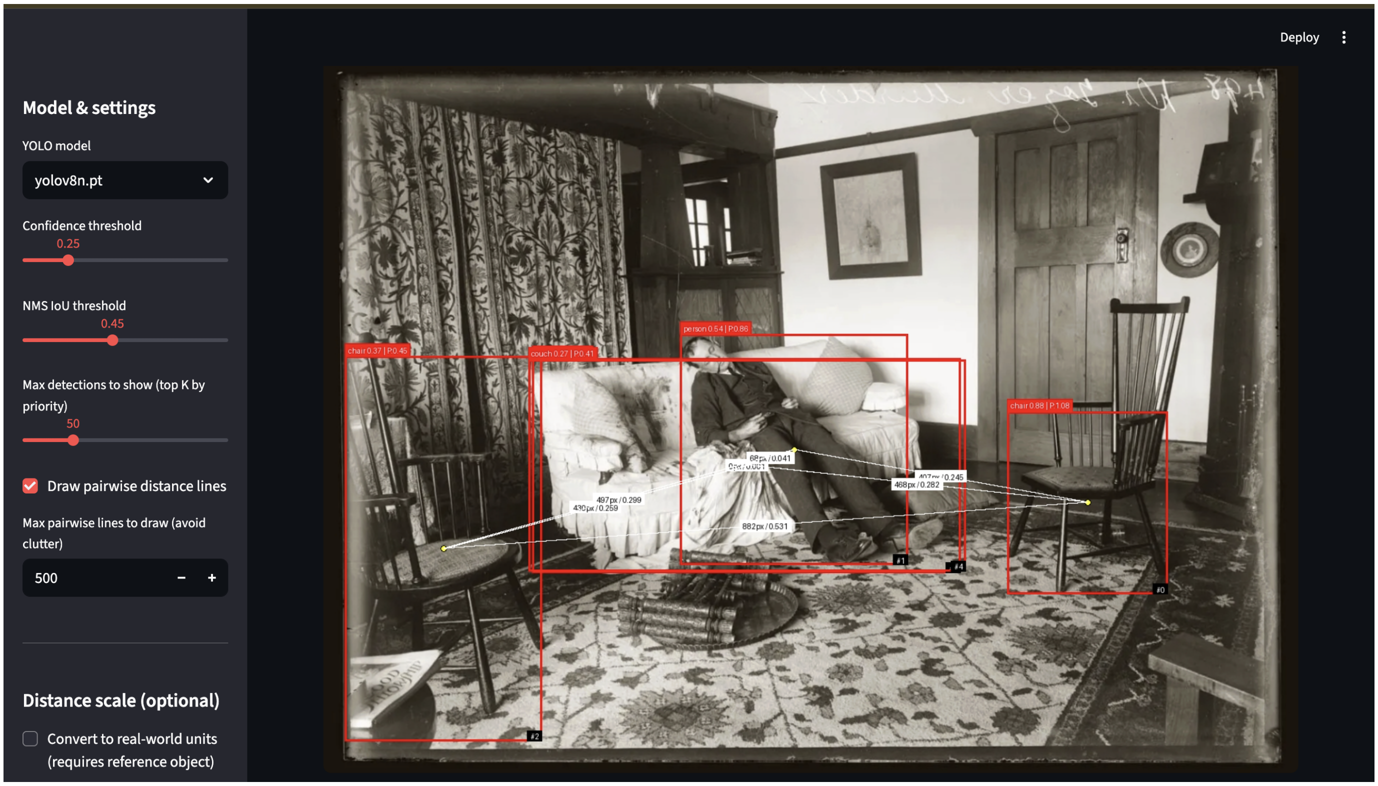

Fig. 10. Streamlit web interface for real-time crime scene analysis.

Overall, the combination of forensic-oriented dataset engineering, selective super-resolution, realistic augmentation, and efficient YOLOv8-based detection enables robust and practical crime-scene evidence analysis under challenging real-world conditions.

IV. Results and Discussion

The performance of the proposed crime scene analytics system was evaluated on the held-out validation set using standard object detection metrics, including precision, recall, F1-score, mean Average Precision at IoU 0.50 (mAP@50), and mean Average Precision averaged over IoU thresholds from 0.50 to 0.95 (mAP@50:95). Both quantitative metrics and qualitative visualizations were analyzed to assess robustness under realistic forensic conditions.

Fig. 11. Showing the output during the yolo model trained.

A. Training and Evaluation Setup

YOLOv8 nano (YOLOv8n), small (YOLOv8s), and medium (YOLOv8m) variants were trained for 100–150 epochs on the enhanced dataset using stochastic gradient descent with cosine learning rate scheduling. Early stopping was employed to prevent overfitting. All models were evaluated using a confidence threshold of 0.25 and an IoU threshold of 0.45. Experiments were conducted on an NVIDIA Tesla T4 GPU in the Google Colab environment.

B. Quantitative Results

Among the evaluated models, YOLOv8s achieved the best balance between detection accuracy and computational efficiency. The quantitative evaluation results are summarized below:

TABLE I
OVERALL PERFORMANCE METRICS OF YOLOV8S MODEL

Metric	Value
Precision (Mean)	0.911
Recall (Mean)	0.626
F1-score	0.743
mAP@50	0.744
mAP@50:95	0.615

The high precision indicates a low false-positive rate, which is particularly important in forensic applications where incorrect evidence identification can mislead investigations. The moderate recall reflects the inherent difficulty of detecting small, occluded, and visually diverse evidence objects in real-world crime-scene imagery.

C. Class-wise Performance Analysis

A detailed class-wise evaluation provides further insight into model behavior:

TABLE II
CLASS-WISE DETECTION PERFORMANCE OF YOLOV8S

Class	Precision	Recall	mAP@50
Footprint	0.951	0.765	0.865
Blood	0.809	0.671	0.811
Weapon	0.892	0.269	0.414
Knife	0.993	0.800	0.886

The model performs strongly on Footprint and Knife classes, benefiting from clearer geometric features and relatively consistent visual patterns. Bloodstain detection also achieves stable performance across diverse backgrounds. In contrast, the Weapon class exhibits lower recall, primarily due to significant intra-class variability, partial occlusions, scale differences, and background clutter. This observation highlights the continued challenge of general weapon detection in unconstrained forensic imagery, even with advanced preprocessing and augmentation.

D. Impact of Preprocessing and Augmentation

The integration of forensic-oriented preprocessing played a crucial role in improving detection robustness. Noise reduction and selective Real-ESRGAN enhancement significantly improved feature clarity for small and distant objects, leading to better localization and confidence scores. Domain-specific augmentation, including copy–paste and environmental distortions, enhanced generalization to cluttered and low-light scenes. The DCGAN-based synthetic weapon generation provided modest improvements in recall for underrepresented weapon instances, validating its use as weakly supervised augmentation rather than direct training data.

E. Qualitative Evaluation

Qualitative results obtained through the Streamlit-based interface demonstrate reliable detection and localization of weapons, bloodstains, footprints, and knives in previously unseen crime-scene images. Bounding boxes are accurately aligned with evidence regions, and confidence scores reflect detection reliability. The centroid-based Euclidean distance measurement feature provides valuable spatial context, enabling investigators to assess relationships between detected evidence items, such as the proximity of a weapon to a blood pool.

Overall, the results confirm that the proposed data-centric approach—combining careful dataset construction, selective super-resolution, and realistic augmentation—substantially enhances the robustness and practical usability of YOLOv8-based crime scene analytics systems.

V. Conclusion and Future Work

This paper presented a forensic-oriented, data-centric deep learning framework for automated crime scene analytics using YOLOv8. Rather than proposing a new detection architecture, the work focused on addressing practical challenges in forensic imagery through careful dataset construction, preprocessing, and enhancement. Two publicly available Roboflow Universe datasets—Weapons and Bloodstains—were systematically merged into a unified, validated dataset, forming a reliable foundation for model training and evaluation.

To handle real-world crime scene conditions such as noise, low resolution, and small evidence size, a multi-stage preprocessing pipeline was introduced. This pipeline combined noise reduction, selective Real-ESRGAN-based super-resolution applied only to small evidence regions, and domain-specific data augmentation tailored to crime-scene variability. In addition, a DCGAN-based synthetic data generation module was employed as weakly supervised augmentation to mitigate class imbalance in weapon detection. These enhancements significantly improved feature clarity and robustness without altering annotation integrity.

Multiple YOLOv8 variants were trained and evaluated, with YOLOv8s achieving the best balance between accuracy and computational efficiency. The final model attained a mean Average Precision of 0.744 at IoU 0.50 and 0.615 across IoU thresholds from 0.50 to 0.95, demonstrating reliable detection performance under challenging forensic conditions. The trained system was further deployed through an interactive Streamlit- based web application, enabling real-time inference, visual annotation, and spatial distance measurement between detected evidences, thereby supporting practical forensic analysis workflows.

While the proposed system demonstrates strong performance, several directions remain for future research and enhancement. Additional forensic evidence categories such as shell casings, fingerprints, personal belongings, and drug-related items can be integrated to provide more comprehensive crime scene understanding. Incorporating semantic segmentation models could enable pixel-level localization of blood pools and object boundaries, further improving interpretability in cluttered scenes. The current DCGAN module operates at low resolution; future work should explore higher-resolution generative models trained on more powerful hardware to produce photorealistic synthetic forensic evidence.

Further extensions include deploying optimized and quantized models on edge devices such as NVIDIA Jetson platforms for on-site crime scene analysis, integrating temporal reasoning for multi-image or video-based scenarios, and exploring explainable AI techniques to enhance transparency and trust in automated forensic decision-support systems. Overall, this work provides a solid foundation for AI-assisted crime scene investigation and highlights the critical role of dataset engineering and preprocessing in real-world forensic applications.In this we add the automatic zooming technology and enhanced for detecting small object.

References

[1] M. Asad et al., “SPOT-CRIME: Suspicious Person and Object Tracking System for Real-Time Crime Monitoring,” in *IEEE Access*, 2022.
[2] J. Chen et al., “Review of Advances in Small Object Detection Technology based on Deep Learning,” in *Journal of Physics: Conference Series*, 2021.
[3] A. Bontempelli et al., “Generative AI Powered Forensic Device,” *arXiv preprint arXiv:2303.12999*, 2023.
[4] R. Johnson and L. Williams, “Crime Scene Reconstruction Using 3D Convolutional Neural Networks (3D-CNNs),” *Journal of Forensic Sciences*, vol. 65, no. 2, pp. 485–499, 2020.
[5] S. Smith, “The Artificial Intelligence in Autopsy and Crime Scene Analysis,” *Forensic Science International*, vol. 320, 2021.
[6] K. Kumar and A. Singh, “A Real-Time Crime Scene Intelligent Video Surveillance System in Violence Detection Framework using Deep Learning Techniques,” *Multimedia Tools and Applications*, vol. 80, pp. 24819–24838, 2021.
[7] P. Davis, “A Forensic Video Upscaling, Colorizing and Denoising Framework for Crime Scene Investigation,” in *Proceedings of the International Conference on Cyber Security and Protection of Digital Services*, 2022.
[8] Ultralytics, “YOLOv8: The Next Generation of YOLO Object Detection,” 2023. [Online]. Available: https://docs.ultralytics.com
[9] X. Wang et al., “Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021.
[10] Roboflow Universe, “Weapons Dataset.” [Online]. Available: https://universe.roboflow.com
[11] Roboflow Universe, “Bloodstain Dataset.” [Online]. Available: https://universe.roboflow.com