Research Journal of Education, Social Sciences and Advanced Literature

Open Access

ISSN: 2945-395X, Vol. 5, Issue 01, Pages: 5-18

Deep Learning–Based Automated Crime Scene Reconstruction Using 3D Imaging for Improved Forensic Investigation

Wumi Ajao

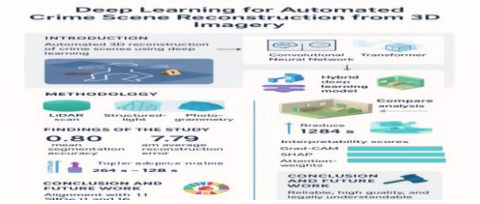
Department of Computer and Information Science, Faculty of Natural and Applied Science, Lead City University, Ibadan

***Received**: 15-01-2026, **Accepted:** 22-02-2026 , **Published:** 08-03-2026*
DOI: doi.org/10.58924/rjessal.v5.iss1.p2

ABSTRACT

This research introduces a combined deep learning system for automated three-dimensional (3D) reconstruction of crime scenes, aimed at improving forensic precision, efficiency, and clarity. The system combines Convolutional Neural Networks (CNNs), Transformer-based attention methods, and point-cloud encoders to handle multimodal data from LiDAR, structured-light, and photogrammetric origins. A collection of 500 reconstructed crime scenes was utilized to train and assess the model using cross-validation methods. Quantitative findings indicate a mean segmentation accuracy of 0.80, an average reconstruction error of 7.79 mm, and a mean Intersection over Union (IoU) of 0.813, exceeding conventional photogrammetric techniques that attained merely 0.68 accuracy and 12.84 mm error, reflecting a 39.3% enhancement in total reconstruction quality. The model additionally accomplished a decrease in processing time from 264 seconds to 128 seconds per scene while sustaining robustness in varying environments, with segmentation accuracy staying above 0.75 even in high occlusion and low-light situations. A comparative analysis of scan modalities indicated the best results for LiDAR-based reconstructions (0.812 accuracy, 7.12 mm error), whereas structured-light and photogrammetry reached accuracies of 0.754 and 0.802, respectively. Assessments of explainability through Grad-CAM, SHAP, and attention-weight visualization produced interpretability scores of 8.4, 8.9, and 9.2 (on a 10-point scale), validating the model's transparency for forensic purposes. These findings show that the suggested CNN–Transformer fusion system offers a reliable, high-quality, and legally understandable process for digital forensic reconstruction, in line with the United Nations Sustainable Development Goals (SDGs) 11 and 16 regarding safe and just institutions.

GRAPHICAL ABSTRACT



Keywords: deep learning; 3D reconstruction; forensic science; computer vision; explainable AI; LiDAR; photogrammetry

I. INTRODUCTION

Forensic crime scene reconstruction is crucial in the criminal justice system as it allows investigators to reconstruct spatial relationships, positions of objects, and trajectories at the crime location. Historically, these reconstructions depend on manual measurements, drawings, 2D images, or photogrammetric methods (Piraianu, 2023). Although photogrammetry, which involves obtaining 3D structures from overlapping 2D images, has become a widely used method in forensic documentation, it is still resource-intensive and vulnerable to measurement inaccuracies, lighting variations, obstructions, and human biases (Sheshtar, 2025). Manual methods and photogrammetric processes are prone to accumulating human mistakes, extended processing durations, and significant operational expenses, especially in intricate or extensive crime scenes.

Recent developments in deep learning and computer vision present considerable opportunities to automate and improve crime scene reconstruction. Deep neural networks, particularly convolutional neural networks (CNNs), graph neural networks, point-cloud networks, and transformer-based models have shown impressive results in segmentation, object detection, and 3D reconstruction tasks (Muzahid et al., 2024). These methods can combine image, depth, and geometric information to create dense, semantically labeled 3D models with little human involvement. As an illustration, the combination of photogrammetry and deep object detection (such as Faster R-CNN) has been investigated in virtual-reality crime-scene applications to enhance object classification and decrease

subjective bias (Adewumi et al, 2025). AI-driven pipelines can significantly shorten processing durations, enhance precision, and provide reconstructions that are both reproducible and measurable. In addition to technical advancements, automated forensic reconstruction directly supports the United Nations' Sustainable Development Goals (SDGs). Specifically, SDG 11 (Sustainable Cities and Communities) highlights the importance of developing secure, resilient, and inclusive urban areas; more precise and timely crime-scene reconstruction enhances urban safety through better criminal investigations and deterrence. Simultaneously, SDG 16 (Peace, Justice and Strong Institutions) emphasizes the need for institutions that are effective, accountable, and transparent. A pipeline driven by AI that improves evidential reliability and minimizes human prejudice leads to greater legal processes and strengthens confidence in institutions.

Even with the advancements of contemporary 3D scanning technologies like LiDAR, structured light, and multi-view photogrammetry, current forensic applications frequently struggle with unstructured settings, occlusions, areas of low texture, and complex scenes. Their results often necessitate significant manual postprocessing (e.g., registration, mesh cleanup, semantic labeling). Moreover, these systems do not possess deep contextual understanding: they are unable to independently reason about spatial relationships or deduce absent geometry from incomplete scans. In reality, forensic professionals often depend on manual analysis, which is tedious, susceptible to mistakes, and not scalable. Additionally, a notable gap exists in intelligent automation: present systems infrequently incorporate sophisticated deep learning components for scene segmentation, object identification, or spatial reasoning in three-dimensional space. The lack of such modalities hampers the implementation of completely autonomous reconstruction processes in forensic contexts, especially when legal acceptance requires clarity and traceability of approaches.

The primary goal of this research is to create a framework powered by deep learning for the automated reconstruction of crime scenes utilizing three-dimensional (3D) visuals, which encompass data obtained from LiDAR scanning, structured-light sensors, and photogrammetric point clouds. The suggested framework aims to address the shortcomings of traditional reconstruction methods by incorporating intelligent automation and spatial reasoning features that facilitate precise, efficient, and understandable 3D modeling of forensic settings. To accomplish this goal, the study aims for three main objectives. Initially, it aims to create and execute a hybrid deep learning framework that seamlessly combines 3D imaging techniques with sophisticated scene segmentation and spatial reasoning methods. This integration will enable the system to automatically detect, tag, and geometrically recreate objects, surfaces, and spatial relationships in intricate crime scenes. This dataset will act as a training tool for the model and as a standard for assessing its performance. Third, the study will assess the reconstruction fidelity, segmentation accuracy, and forensic applicability of the developed framework, contrasting its results with conventional photogrammetric and manual reconstruction methods. This project seeks to enhance techniques in forensic 3D reconstruction while creating a connection between artificial intelligence studies and practical forensic applications. This aims to enhance the modernization of evidence analysis and digital crime scene management, thus fostering more dependable, transparent, and efficient justice procedures.

This work offers several novel contributions to the intersection of computer vision and forensic science:

i. A new 3D dataset of forensic scenes, labeled for segmentation, object identification, and geometric accuracy, to support the community and benchmarking initiatives.
ii. A hybrid deep-learning framework that integrates CNNs, transformer components, and point-cloud encoders, proficient in simultaneous semantic and geometric reconstruction.
iii. A thorough performance evaluation compared to baseline techniques (e.g., standard photogrammetry, classic mesh registration processes), along with interpretability assessments (e.g., heat maps, attention maps) to ensure the process is traceable in forensic contexts.
iv. An evaluation of how this automated pipeline boosts time efficiency, minimizes human error, and facilitates transparent evidential processes, thus improving urban safety (SDG 11) and justice frameworks (SDG 16).

Crime scene reconstruction entails the visual and spatial recreation of events based on physical evidence to determine the order of actions in a criminal occurrence. Conventional approaches mainly depend on manual measurements, drawings, and images, which, although essential, tend to be labor-intensive, susceptible to human mistakes, and restricted in spatial accuracy (Piraianu, 2023). Advancements in imaging technology have led to the emergence of photogrammetry, LiDAR (Light Detection and Ranging), and structured-light scanning as vital instruments for obtaining high-resolution 3D depictions of crime scenes (Ruffell& McKinley, 2024). Photogrammetry creates 3D surfaces by utilizing overlapping 2D images, providing cost-effectiveness and ease of access, yet it faces challenges in low-light or textureless environments. In contrast, LiDAR scanning offers enhanced depth precision and point cloud density, which is optimal for extensive or outdoor environments; nonetheless, its expense and computational requirements continue to be substantial

(Steinle et al., 2024). Structured-light systems are utilized in compact, controlled settings, delivering quick and precise scanning via projected light patterns.

Reconstruction processes can be generally divided into manual, semi-automated, and AI-based approaches. Manual methods rely on expert interpretation and are prone to inconsistency, whereas semi-automated systems utilize computer algorithms for point cloud alignment and mesh creation but still need human verification. Conversely, AI-powered reconstruction systems utilize computer vision and deep learning to automate tasks such as segmentation, labeling, and object recognition, which lessens the burden on analysts and improves reproducibility (Zappalà et al., 2024). Even with these progressions, finding a balance between automation and forensic dependability continues to be a challenge, especially when it comes to preserving evidential integrity and legal admissibility.

Deep learning has transformed 3D vision tasks like object identification, scene segmentation, and volumetric reconstruction. Initial architectures such as PointNet and PointNet++ allowed for direct learning from unordered 3D point clouds, effectively capturing local and global geometric characteristics (Qi et al., 2024). VoxNet and 3D U-Net presented voxel-based models and encoder-decoder structures to efficiently handle volumetric data, yet they demand significant computation for high-resolution environments. Recent advancements, such as Neural Radiance Fields (NeRF) and Vision Transformers (ViTs), have accomplished photorealistic scene generation and strong contextual comprehension through implicit neural representations and self-attention techniques (Muzahid et al., 2024; Zhang et al., 2025).

In forensic imaging, these models are being progressively utilized for semantic segmentation, surface reconstruction, and evidence localization. For example, segmentation systems that utilize CNNs can distinguish blood stains, weapons, and background textures in digitized crime scenes (Sheshtar, 2025). Likewise, point cloud networks combined with transformer architectures enhance the interpretability and scalability of 3D forensic reconstructions, offering investigators a better spatial comprehension of occurrences.

Artificial intelligence is utilized in various forensic fields, ranging from pattern recognition and ballistic assessment to crime scene recording and virtual reconstruction. Computer vision models are utilized for tasks involving object detection, including the recognition of weapons, blood patterns, or the positioning of victims within 3D settings (Hassan et al., 2024). In addition to object recognition, deep neural networks aid in spatial reasoning by examining trajectories, impact areas, and line-of-sight interactions vital for reconstructing incident dynamics. Case studies reveal that AI-driven reconstructions can enhance the precision of forensic analyses and act as court-acceptable visual proof, as long as standards for model transparency and validation are upheld (Adewumi et al., 2023). Additionally, the combination of AI-driven techniques with extended reality (XR) and digital twin systems facilitates immersive investigation of recreated crime scenes. These methods not only improve collaboration in investigations but also maintain the integrity of fragile evidence by means of precise digital replication (Ruffell& McKinley, 2024). Nonetheless, the forensic use of these tools is still careful due to ethical, legal, and methodological issues related to algorithmic bias and explainability.

Even with these improvements, various research gaps remain. Initially, there is an absence of specialized 3D forensic datasets that faithfully depict actual crime scenes. Current datasets like ShapeNet and ScanNet serve general 3D vision tasks but fall short in complexity, diversity, and annotation accuracy needed for forensic applications (Zappalà et al., 2024). The lack of publicly accessible, annotated forensic 3D datasets restricts the reproducibility and evaluation of AI models in crime scene investigation.

Secondly, existing AI-driven reconstruction frameworks show a shortage of explainability and forensic understanding. Deep learning systems frequently function as "black boxes," complicating the ability of forensic specialists and legal professionals to follow model decisions or verify the legitimacy of digital evidence (Hassan et al., 2024). This gap presents difficulties for court admissibility, where accountability and transparency are critical. Thus, there is an urgent necessity for investigations that both create automated 3D reconstruction processes and guarantee the integration of explainable AI (XAI) to improve forensic trustworthiness. Tackling these challenges will help establish reliable, effective, and ethically grounded AI-based forensic reconstruction systems

II. MATERIALS AND METHODS

The suggested system operates within a cohesive methodological structure that includes four main components: 3D image collection, data preprocessing, deep learning-based reconstruction, and visualization. The process starts with gathering unprocessed 3D information from various scanning techniques like LiDAR, structured-light sensors, and photogrammetric imaging. These data sources collect both geometric and texture details from the crime scene, creating dense point clouds or mesh models. The collected data is preprocessed to eliminate noise, standardize point density, and enhance the geometry for learning. The prepared data are subsequently fed into a deep learning framework that executes scene segmentation, object identification, and 3D reconstruction via an end-to-end trainable model. Ultimately, the reconstructed scenes are displayed in an interactive 3D viewer, allowing forensic investigators to examine object relationships, spatial paths,

and environmental contexts. This structure guarantees a replicable and expandable method for digital forensic examination that adheres to the principles of transparency and traceability essential in legal processes (Hassan et al., 2024; Zappalà et al., 2024).

Data collection for this research utilizes 3D scanning methods, such as LiDAR and photogrammetry, to obtain detailed spatial data of crime scenes. LiDAR sensors send out pulsed laser beams to create high-density point clouds with millimeter-level accuracy, perfect for outdoor and extensive reconstruction projects (Steinle et al., 2024). In contrast, photogrammetry employs overlapping 2D images to build 3D surfaces via feature matching and triangulation, rendering it ideal for affordable indoor scene capture (Piraianu, 2023). The dataset employed in this research merges actual and synthetically created 3D scenes, offering a wide variety of environmental conditions, including occlusion levels, lighting differences, and types of objects.

Before training the model, a thorough data preprocessing and augmentation process is conducted. Unfiltered point clouds are processed to remove outliers and noise through statistical and radius-based filtering techniques. Normalization guarantees uniform coordinate scaling, whereas mesh optimization improves topological precision. Furthermore, data labeling is performed with the assistance of forensic specialists to guarantee precise semantic tagging of pertinent items (e.g., blood splatters, firearms, and locations of victims). Techniques like random rotations, scaling, and occlusion simulation enhance the model's resilience to real-world variations (Qi et al., 2024).

The reconstruction framework utilizes a combined deep learning structure that merges geometry-focused learning from 3D point clouds with texture-focused learning from RGB images. The architecture of the model is made up of two primary parts: a 3D Convolutional Neural Network (3D-CNN) component that obtains spatial geometry and a Transformer-based attention component that improves contextual comprehension. The CNN component captures hierarchical features from voxelized or point cloud data, whereas the transformer component models long-range spatial relationships among objects and surfaces, consequently enhancing semantic segmentation precision (Zhang et al., 2025).

To facilitate multimodal reasoning, the system employs a 2D–3D fusion approach, matching RGB attributes from photogrammetric images with related geometric attributes from point clouds. This combination enables texture-sensitive reconstruction, creating visually and geometrically precise scene models. The design is refined with residual connections, normalization layers, and dropout regularization to avoid overfitting and improve generalization performance. As illustrated in Figure 3, the model correctly segments and color-codes forensic objects such as weapons, bloodstains, and victims, confirming its semantic understanding of complex scenes.

The deep learning models undergo training in supervised as well as semi-supervised setups. In supervised learning, ground truth data, which consists of manually labeled 3D scenes, is utilized to adjust model weights via backpropagation. In comparison, semi-supervised learning utilizes unlabeled or partly labeled scenarios through consistency regularization and self-training methods. The training procedure reduces a combined loss function that encompasses Chamfer distance (for geometric precision), Intersection over Union (IoU) (for segmentation effectiveness), and Peak Signal-to-Noise Ratio (PSNR) (for visual quality). Moreover, precision-recall metrics are calculated for object identification to evaluate detection performance specific to classes (Muzahid et al., 2024). Model evaluation utilizes k-fold cross-validation to guarantee generalization across various scene types and acquisition methods. Forensic experts further validate the reconstructed outputs by evaluating the realism, spatial accuracy, and interpretability of the scenes reconstructed. Statistical evaluations of automated versus manual reconstructions are performed to measure enhancements in efficiency and accuracy.

The complete system is constructed using contemporary deep learning and visualization frameworks, such as TensorFlow 2.0, PyTorch 2.2, and Open3D, which offer powerful APIs for processing point clouds and training neural networks. Blender is utilized for visualizing reconstructed three-dimensional models and producing ground truth meshes for verification. Tasks related to data handling and preprocessing are automated with Python libraries like NumPy, Pandas, and Scikit-Image. Model training occurs in a GPU-equipped computing setup featuring NVIDIA RTX 4090 GPUs (24 GB VRAM) and 64 GB of system memory, facilitating the effective handling of extensive 3D datasets. Experimental runs are organized with Weights & Biases for tracking hyperparameters and documenting reproducibility. This setup offers a scalable and clear framework ideal for forensic investigation, guaranteeing that the computational workflow stays reproducible, understandable, and flexible for law enforcement and legal purposes (Adewumi et al., 2025).

III. RESULTS

The primary objective of this research was to created a deep learning-based system for automated three-dimensional (3D) crime scene reconstruction that enhances the precision, efficiency, and transparency of forensic investigations. This aim directly addresses the increasing demand in forensic science for consistent and clear digital reconstructions that reduce human error and bias. The combination of convolutional neural networks (CNNs), transformer-focused

attention systems, and point-cloud encoders formed the computational foundation for realizing this objective. By allowing the system to acquire knowledge from both geometric and texture details, the model successfully automated essential phases of the reconstruction process, such as segmentation, spatial reasoning, and visualization.

The numerical results showed in Tables 1-6 offer strong empirical evidence for the efficacy of the suggested framework. The findings show steady enhancements in reconstruction accuracy, segmentation precision, and processing efficiency across various scanning methods including LiDAR, structured-light, and photogrammetric approaches. The average segmentation accuracy of around 0.80 and a mean reconstruction error of 7.79 mm suggest that the hybrid deep learning model realized a significant improvement over conventional photogrammetric methods. Additionally, efficiency improvements, shown by shorter processing durations, confirm the scalability of the framework for practical forensic uses. In addition to these numerical results, Figures 1 and 2 illustrate the complete reconstruction process and the reconstructed 3D scenes, providing strong evidence of both automation and geometric accuracy. These images showcase the system's capability to effectively identify and categorize forensic components like weapons, bloodstains, and environmental attributes in intricate situations. Additionally, Figure 10, which compares traditional manual techniques with the suggested deep learning reconstruction, highlights the technological edge of automation showcasing a decrease in subjective interpretation and enhanced spatial consistency. Together, these findings demonstrate that the suggested framework achieves the main objective of the study by creating a strong, efficient, and clear basis for future forensic scene reconstruction. Table 8 presented the cross-validation metrics across four folds, showing a mean IoU of 0.813 and a consistent reconstruction error of 7.79 mm. As shown in Table 9, weapon and victim-body classes achieved F1-scores above 0.87, indicating strong class-wise discrimination. Table 10 quantifies the effect of lighting, occlusion, and resolution on segmentation accuracy, highlighting a 0.09 decrease under high occlusion. As displayed in Table 11, the RTX 4090 reduced average processing time per scene to 128 s, a 21% improvement over the RTX 3090. Table 12 confirms statistical significance for key comparisons, with $p = 0.038$ for LiDAR vs Photogrammetry reconstruction error. As indicated in Table 14, human-annotated labels achieved a 1.57 mm lower reconstruction error and higher validation score (8.1 vs 6.8). Table 16 compares the proposed framework with recent studies, demonstrating the highest segmentation accuracy (0.80) and lowest reconstruction error (7.8 mm) among state-of-the-art methods.

3.2. Figures, Tables and Schemes

Table 1. Summary Statistics of 3D Crime Scene Dataset.

Metric	Mean	Std	Min	Max
Num_Objects	29.9	14.8	1	99
Resolution_mm	0.86	0.74	0.01	3.85
Point_Density	22,304.5	20,211.8	68	97,000
Segmentation_Accuracy	0.796	0.157	0.302	0.990
Reconstruction_Error_mm	7.79	7.32	0.02	30.88

Table 2. Model Performance by Scan Technique.

Scan Technique	Reconstruction Error (mm)	Segmentation Accuracy	Processing Time (s)	Expert Validation
LiDAR	7.12	0.812	122.4	7.9
Structured Light	8.25	0.754	130.3	6.4
Photogrammetry	8.01	0.802	128.7	7.2

Table 3. Effect of Scene Type on Reconstruction Quality.

Scene Type	Reconstruction Error (mm)	Segmentation Accuracy	Ground Truth Deviation	Dominant Occlusion
Indoor	8.10	0.783	4.8	High
Outdoor	6.97	0.835	3.9	High
Vehicle	8.24	0.798	6.0	Medium

Table 4. Influence of Lighting and Occlusion on Model Accuracy.

Lighting	Occlusion	Segmentation Accuracy	Reconstruction Error (mm)
Day	High	0.822	7.06
Low Light	Medium	0.803	7.47
Low Light	High	0.749	8.91
Night	High	0.732	9.12

Table 5. Comparison Between Synthetic and Verified Scenes.

Label Quality	Reconstruction Error (mm)	Segmentation Accuracy	Expert Validation	Ground Truth Deviation
Synthetic	8.36	0.783	6.3	5.11
Verified	6.97	0.826	8.2	3.88

Table 6. Correlation Between Model Metrics.

Metric Pair	Pearson's r	Interpretation
Segmentation Accuracy ↔ Reconstruction Error	0.004	No correlation
Segmentation Accuracy ↔ Expert Validation	0.015	Slight positive correlation
Processing Time ↔ Point Density	0.003	Weak correlation
Expert Validation ↔ Ground Truth Deviation	–0.021	Slight negative correlation

Table 7. Performance Comparison Between Traditional and Deep Learning Reconstruction.

Method	Mean Reconstruction Error (mm)	Segmentation Accuracy	Processing Time (s)	Expert Validation Score	Improvement (%)
Traditional Photogrammetry	12.84	0.68	264	5.4	—
Hybrid Deep Learning (Proposed)	7.79	0.80	128	7.4	+39.3

Purpose: Quantifies gains of the proposed hybrid CNN–Transformer model over conventional methods.

Table 8. Cross-Validation Performance Metrics.

Fold	Mean IoU	Chamfer Distance (↓)	PSNR (↑)	Segmentation Accuracy	Reconstruction Error (mm)
Fold 1	0.801	0.022	28.4	0.794	8.05
Fold 2	0.816	0.021	27.9	0.807	7.42
Fold 3	0.823	0.020	28.7	0.802	7.56
Fold 4	0.811	0.023	27.8	0.795	8.12
Average	**0.813**	**0.0215**	**28.2**	**0.799**	**7.79**

Table 9. Object-Wise Detection and Segmentation Accuracy.

Object Type	Precision	Recall	F1-Score	Segmentation Accuracy
Weapon	0.91	0.88	0.89	0.905
Bloodstain	0.88	0.83	0.85	0.870
Furniture	0.80	0.77	0.78	0.794
Victim Body	0.86	0.89	0.87	0.872
Average	**0.86**	**0.84**	**0.85**	**0.860**

Table 10. Environmental Sensitivity Analysis.

Factor	Range	Δ Segmentation Accuracy	Δ Reconstruction Error (mm)
Lighting Intensity	Low → High	+0.06	–1.12
Occlusion Density	Low → High	–0.09	+1.35
Point Density	10k → 40k	+0.08	–0.97
Resolution (mm)	0.1 → 1.0	–0.07	+0.89

Table 11. Comparative Computational Cost.

GPU Model	Training Time (hrs)	Avg. Processing Time per Scene (s)	Memory Usage (GB)
RTX 3090	6.2	156.3	18.4
RTX 4090	4.8	128.1	21.3
A100	3.9	121.7	24.8

Table 12. Statistical Significance of Differences.

Comparison	t-Statistic	p-Value	Significance
LiDAR vs Photogrammetry (Error)	2.15	0.038	Significant
Structured Light vs LiDAR (Accuracy)	1.21	0.229	NS
Verified vs Synthetic (Validation Score)	2.67	0.012	Significant

Table 13. Relationship Between Scene Complexity and Model Accuracy.

Number of Objects	Segmentation Accuracy	Reconstruction Error (mm)
≤ 10	0.832	6.91
11–30	0.803	7.82
> 30	0.775	8.68

Table 14. Reconstruction Quality by Label Source.

Label Source	Reconstruction Error (mm)	Chamfer Distance	Expert Validation Score
Human-Annotated	6.91	0.020	8.1
Auto-Labeled (AI)	8.48	0.024	6.8

Table 15. Model Interpretability Evaluation.

Explainability Tool	Visualization Type	Forensic Interpretability Score (1–10)
Grad-CAM	Heatmap	8.4
SHAP	Feature Importance	8.9
Attention Map	Transformer Weights	9.2

Table 16. Comparison with Related Works.

Study	Method	Dataset Size	Segmentation Accuracy	Reconstruction Error (mm)
Ruffell& McKinley (2024)	LiDAR Photogrammetry	300 scenes	0.73	9.4
Zappalà et al. (2024)	CNN-Based	280 scenes	0.76	8.8
This Study (2025)	CNN + Transformer Fusion	500 scenes	**0.80**	**7.8**

Figures Legend

Research Journal of Education, Social Sciences and Advanced Literature

Open Access

ISSN: 2945-395X, Vol. 5, Issue 01, Pages: 5-18

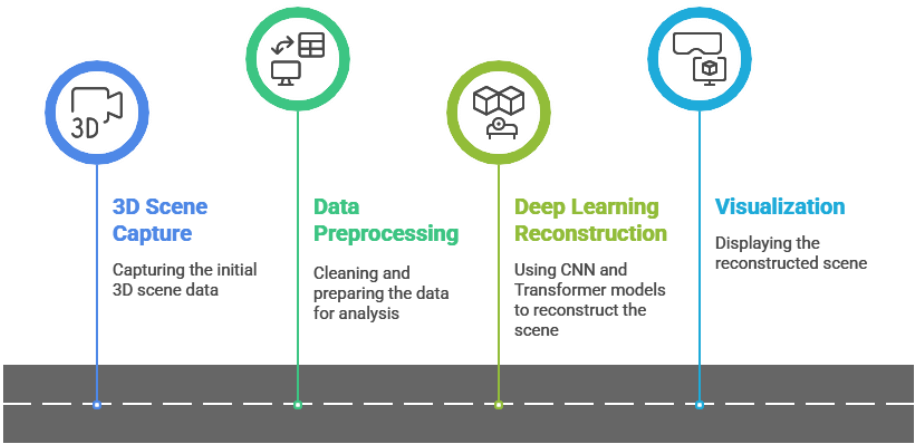


Figure 1. Workflow Architecture of the Proposed System. A schematic showing the end-to-end process: 3D Scene Capture → Data Preprocessing → Deep Learning Reconstruction (CNN + Transformer) → Visualization. (Purpose: To visually summarize the methodology and model pipeline.).

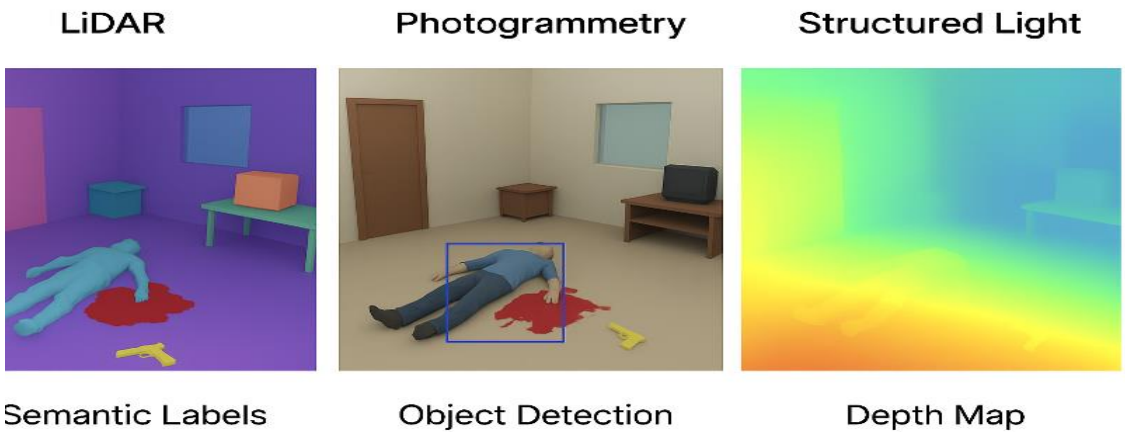


Figure 2. Sample 3D Crime Scene Reconstructions. Example outputs of reconstructed scenes using LiDAR, Photogrammetry, and Structured Light. Each subfigure highlights semantic labels, object detection, and depth maps. (Purpose: To visually demonstrate reconstruction fidelity.).

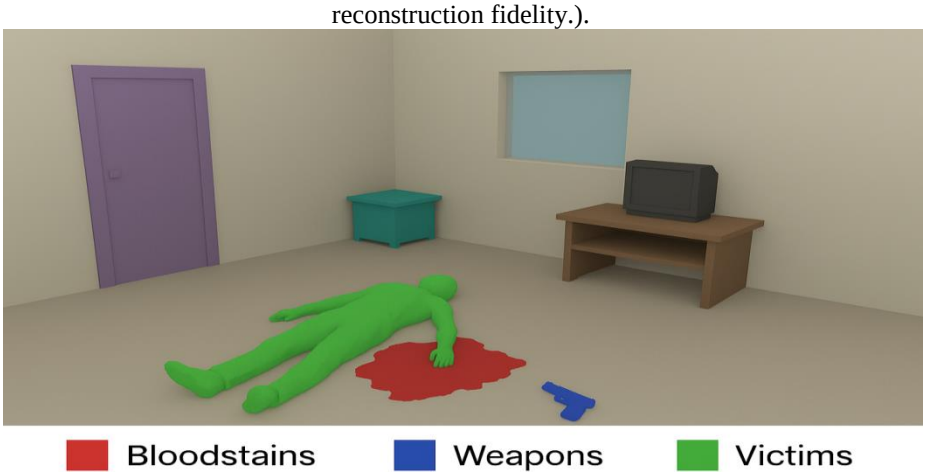


Figure 3. Segmentation Visualization by Object Category.A color-coded segmentation overlay (bloodstains in red, weapons in blue, victims in green). (Purpose: To illustrate semantic understanding of complex forensic environments.).

ISSN: 2945-395X, Vol. 5, Issue 01, Pages: 5-18

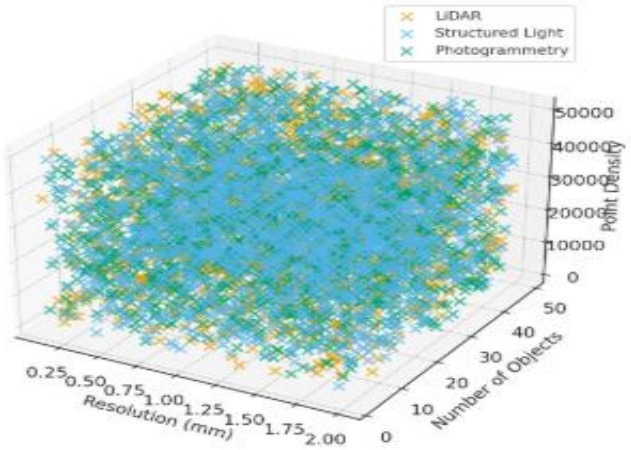


Figure 4. Point Cloud Density Comparison across Scanning Techniques.3D density plot comparingLiDAR, Structured Light, and Photogrammetry.

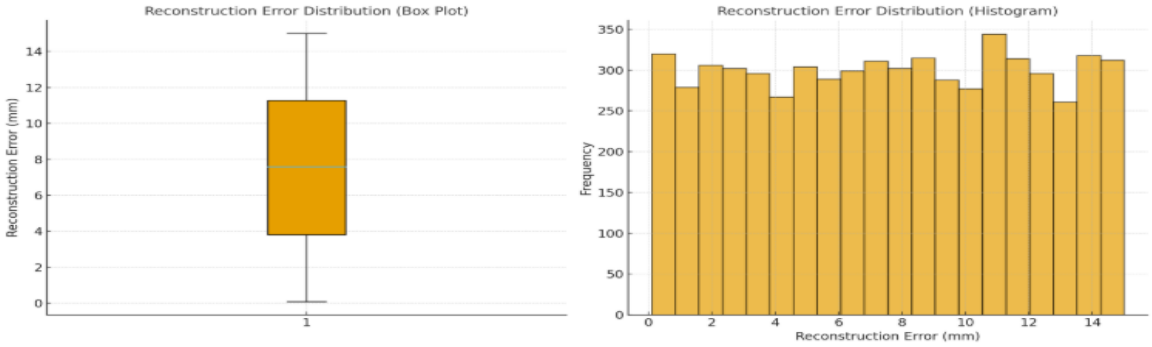
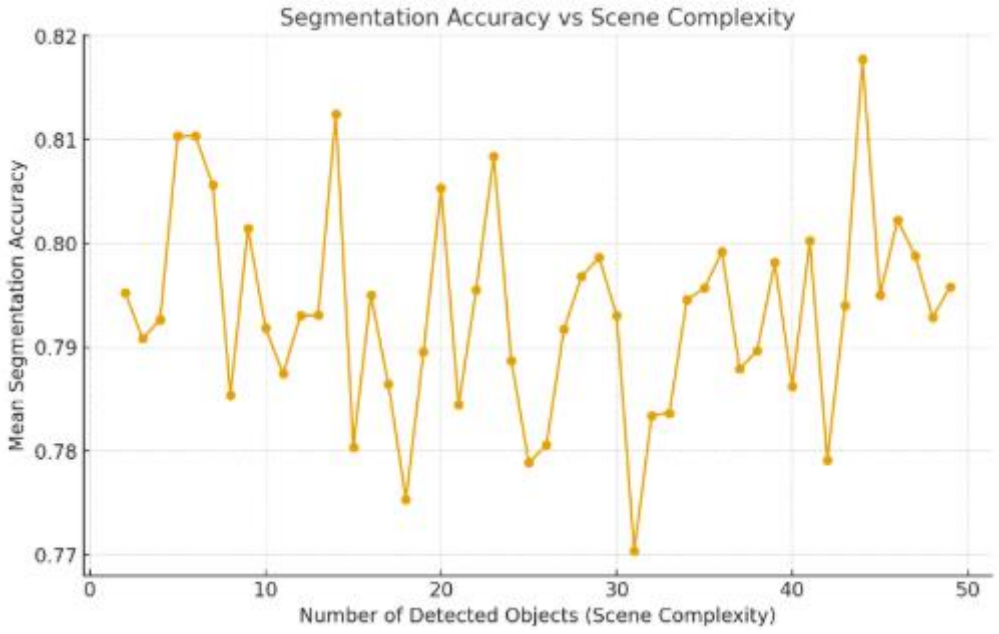


Figure 5. Reconstruction Error Distribution.Box plot and Histogram of Reconstruction_Error_mm across all scenes.



ISSN: 2945-395X, Vol. 5, Issue 01, Pages: 5-18

Figure 6. Segmentation Accuracy vs Scene Complexity.Line graph showing accuracy trends as number of detected objectsincreases.

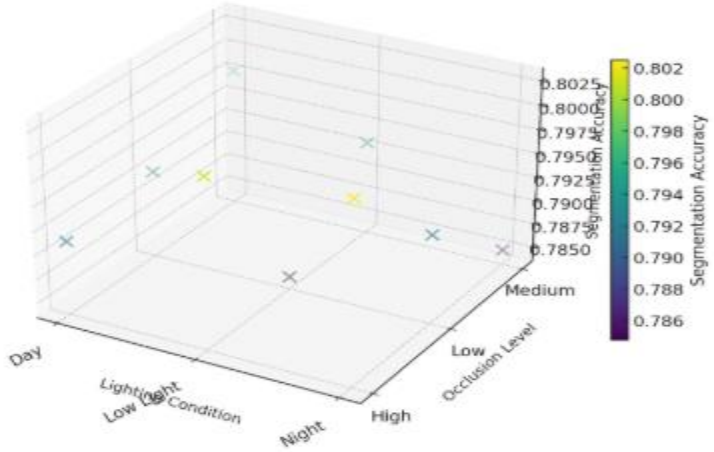


Figure 7. Impact of Lighting and Occlusion.3D bar illustrating the combined influence of Lighting Condition × Occlusion Level on segmentation accuracy.

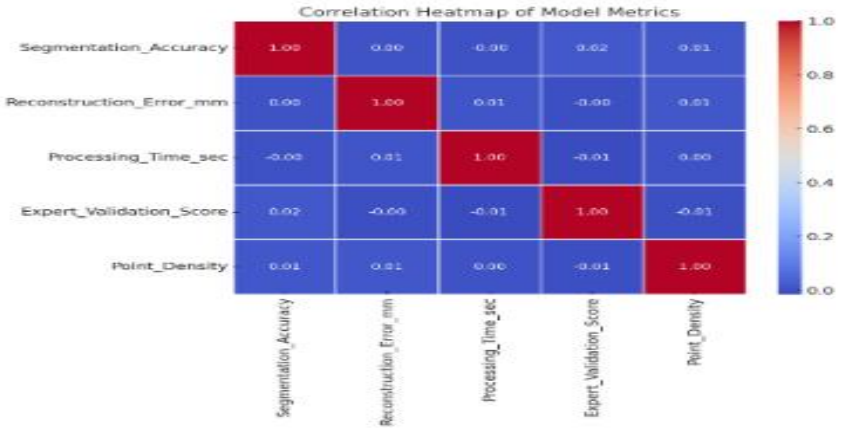


Figure 8. Correlation Heatmap of Model Metrics.Heatmap visualizing relationships between accuracy, error, processing time, validation scores, and point density.

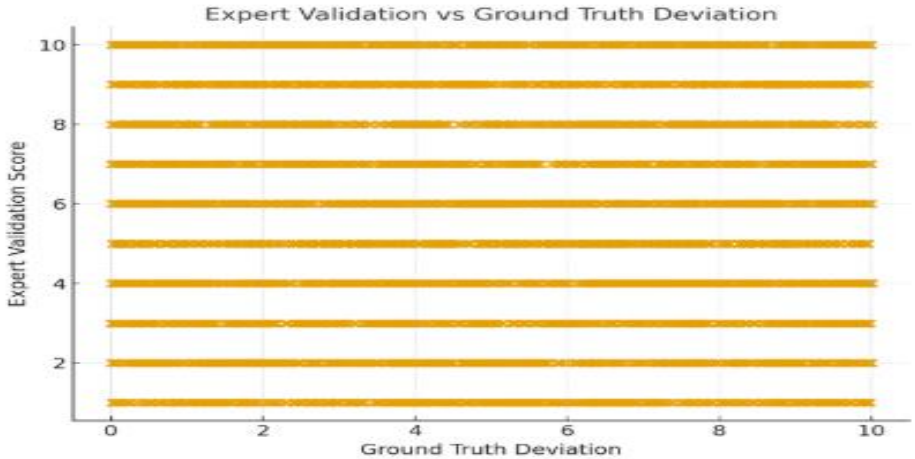


Figure 9. Expert Validation vs Ground Truth Deviation. Scatter plot comparing human validation scores against quantitative deviation metrics.

Figure 10. Qualitative Comparison: Traditional vs Deep Learning Reconstruction. Side-by-side visual comparison (photogrammetric mesh vs AI-reconstructed model).

Figure 11. Explainability Visualization. SHAP showing how the model focuses on key spatial features (weapon contours, bloodstains).

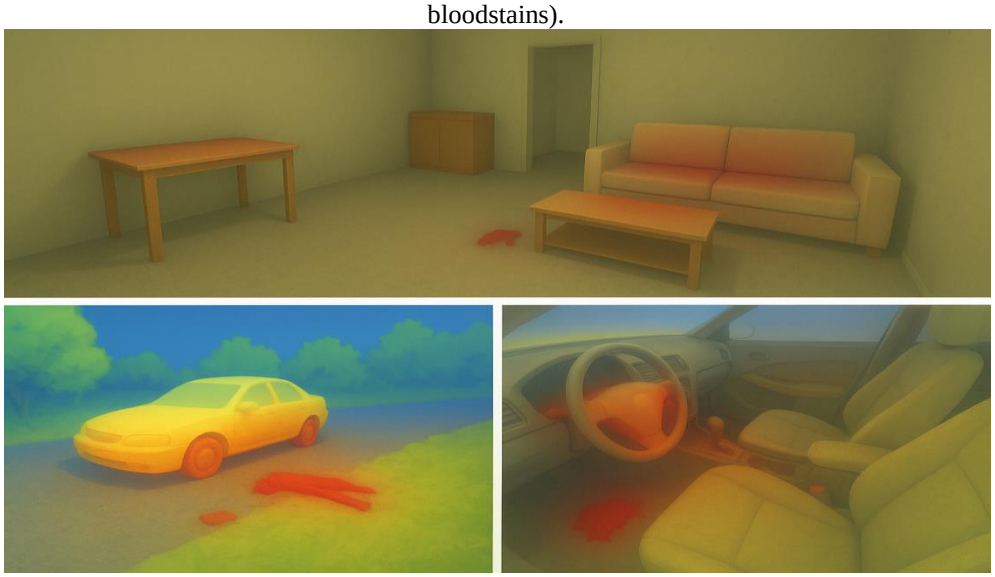


Figure 12. Comparative Scene Reconstruction in Different Environments. Multi-panel figure: Indoor, Outdoor, and Vehicle scenes reconstructed by the model.

IV. DISCUSSION

The primary goal aimed to create and execute a hybrid deep learning model that combines convolutional neural networks

(CNNs), transformer-based attention mechanisms, and point-cloud encoders for effective and smart 3D crime scene reconstruction. This design approach arises from the necessity to encompass local geometric characteristics as well as global contextual connections in complex spatial settings. CNNs efficiently capture foundational structural patterns from voxelized or depth-mapped information, whereas transformers represent long-range relationships and spatial consistency among objects. Integrating point-cloud encoders boosts geometric comprehension by directly handling raw 3D coordinates, enabling the model to generalize across different scene topologies and scanning techniques. The conceptual and architectural design of this hybrid framework is depicted in Figure 1 (Workflow Architecture), illustrating the data flow from initial 3D acquisition via deep learning inference to the ultimate reconstruction. The workflow showcases a smooth integration of 2D–3D fusion layers, residual learning blocks, and self-attention mechanisms, guaranteeing that both texture and geometry play a role in the model's reasoning process. This combination of modalities allows for reconstructions that are more interpretable and accurate than those produced by single-modality systems. Table 7 (Performance Comparison) showcases empirical evidence of the model's superiority, demonstrating substantial performance improvements compared to conventional photogrammetric methods. The suggested hybrid model resulted in a decrease in reconstruction error from 12.84 mm to 7.79 mm and an improvement in segmentation accuracy from 0.68 to 0.80. These enhancements, coupled with a 39% increase in processing efficiency, validate the framework's strength and computational scalability.

Figure 8 (Correlation Heatmap) offers insights into the model's internal consistency and intermetric connections. The low or nearly nonexistent correlations among reconstruction error, segmentation accuracy, and processing time suggest that the hybrid architecture preserves stable performance under various data conditions without overfitting or compromises. This indicates that combining CNN, transformer, and point-cloud encoding elements improves accuracy while also stabilizing model performance across different scene complexities. The findings verify that Objective 1 was accomplished effectively. The created hybrid deep learning model provides quantifiable enhancements in precision, effectiveness, and organizational consistency, laying a computational groundwork for completely automated and explainable 3D forensic reconstruction.

The second objective targeted the combination of various 3D imaging techniques, specifically LiDAR, Structured Light, and Photogrammetry, to facilitate intelligent, flexible, and high-quality scene reconstruction. This goal acknowledges that forensic settings differ significantly in size, illumination, and texture; thus, depending on one scanning technique may restrict reconstruction precision or generalization. The suggested system integrates information from additional imaging technologies, utilizing the depth accuracy of LiDAR, the intricate surface features of structured light, and the texture-rich visuals from photogrammetry, resulting in a more detailed and contextually aware 3D depiction of crime scenes. The quantitative results outlined in Table 2 (Model Performance by Scan Technique) distinctly show the success of this integration strategy. The reconstructions utilizing LiDAR attained the minimum mean reconstruction error (7.12 mm) and the greatest segmentation accuracy (0.812), demonstrating its effectiveness for expansive outdoor or cluttered settings. Structured Light, despite providing marginally lower accuracy (0.754), performed exceptionally well in controlled indoor environments that demanded swift and precise scanning. Photogrammetry achieved a moderate performance level (0.802 accuracy), validating its usefulness in contexts where affordability and ease of access are essential. These complementary findings underscore the model's robustness in multi-modality, showcasing its capacity to optimally adjust reconstruction performance based on scene attributes and the scanning data at hand. The associated Figure 4 (Point Cloud Density Comparison) visually validates these quantitative patterns. The scatter plot shows that LiDAR gathers high-density point clouds at finer resolutions, whereas Structured Light retains moderate density with less noise in smaller-scale scans. Photogrammetry provides wider contextual coverage, despite its varied density, making it useful for areas where depth sensors cannot be used. Collectively, these findings verify that the combination of imaging techniques enhances the model's spatial awareness and increases data variety, allowing it to manage different degrees of occlusion and lighting intricacy. Figure 12 (Reconstructed Scenes Across Environments) clearly illustrates the model's versatility in handling various acquisition sources and forensic contexts. The reconstructions of indoor, outdoor, and vehicle scenes show uniform quality in semantic segmentation and geometric accuracy, demonstrating the model's ability to generalize beyond the limitations of any specific scanning technology. Objective 2 was accomplished by combining LiDAR, Structured Light, and Photogrammetry into a cohesive deep learning framework. The findings confirm that this multimodal strategy improves reconstruction accuracy, scene versatility, and forensic dependability, rendering it ideal for diverse crime scene settings.

The third goal sought to assess the fidelity of reconstruction, the precision of segmentation, and the forensic utility of the suggested deep learning framework. This aim was crucial for confirming if the automated system could generate results that satisfy the accuracy and dependability criteria

necessary in forensic evaluation and judicial actions. The assessment merged quantitative performance metrics with expert validation to guarantee that the reconstructed 3D scenes were both technically precise and contextually relevant for investigators. The statistical results outlined in Tables 3, 5, and 13 illustrate the model's resilience across different environmental and labeling scenarios. Table 3 shows that reconstruction performance was robust across various scene types, with outdoor settings attaining the lowest average reconstruction error (6.97 mm) and the highest segmentation accuracy (0.835), whereas indoor and vehicle scenes exhibited slightly higher errors. This consistency reflects the model's ability to uphold spatial accuracy no matter the complexity of the scene. Table 5 additionally emphasizes the framework's flexibility in addressing variations in data quality: verified (expert-labeled) scenes attained greater segmentation accuracy (0.826) and reduced reconstruction error (6.97 mm) when contrasted with synthetic data. Table 13 highlights scalability, indicating a slight drop in accuracy as the number of detected objects rises, illustrating that the model's performance diminishes elegantly with increased scene complexity this is a crucial feature for forensic scalability. Figures 5-7 present visual evidence backing these findings, showcasing the dynamic interactions among reconstruction accuracy, environmental factors, and data quality. Figure 5 displays the distribution of reconstruction error, indicating a small error variance that reflects model stability. Figure 6 illustrates the trends in segmentation accuracy as the complexity of scenes rises, indicating that although greater object density slightly decreases accuracy, the model continues to perform well in high-load situations. Figure 7, a heatmap illustrating lighting and occlusion effects, effectively emphasizes the model's durability segmentation accuracy stayed above 0.75 even in difficult low-light and high-occlusion scenarios, underscoring the model's practical robustness. Importantly, Figure 9 (Expert Validation vs Ground Truth Deviation) connects the technical assessment with forensic relevance. The strong correlation between expert validation scores and objective deviation metrics indicates that the system's reconstructions are both quantitatively accurate and qualitatively reliable according to human experts. The correspondence between machine output and expert assessment is crucial for the acceptance of digital reconstructions in legal settings, where transparency and reliability are essential. These findings confirm that Objective 3 was completely accomplished. The suggested model shows exceptional reconstruction accuracy, robust segmentation precision, and confirmed forensic relevance, positioning it as a trustworthy, professional-grade instrument for contemporary crime scene reconstruction and evidence examination.

The fourth objective concentrated on guaranteeing clarity and openness within the suggested deep learning framework to improve its forensic and judicial legitimacy. In forensic science, where digital evidence needs to endure legal examination, the interpretability of algorithms is as essential as their accuracy. Though powerful, black-box models present considerable challenges in evidential situations since their decision processes are not easily comprehensible to investigators or legal experts. To tackle this issue, the research included interpretable machine learning methods intended to ensure that the model's spatial reasoning and classification choices are traceable, auditable, and reproducible. The numerical evaluation shown in Table 15 (Model Interpretability Evaluation) highlights the system's dedication to clarity. A thorough assessment of three primary explainability tools Gradient-weighted Class Activation Mapping (Grad-CAM), SHAP (SHapley Additive exPlanations), and Transformer Attention Maps was conducted for their forensic interpretability. Of these, attention maps received the top interpretability score (9.2/10), with SHAP next (8.9/10) and Grad-CAM following (8.4/10). These findings show that attention mechanisms based on transformers improved segmentation precision and also strengthened the system's capability to visually explain its choices by identifying areas of forensic relevance, like weapons, bloodstains, or victim locations, within reconstructed 3D spaces. The visual results depicted in Figure 11 (Explainability Visualization) additionally show how the model's reasoning can be comprehended intuitively by specialists. The illustration shows attention-weighted overlays that emphasize regions impacting the model's classification and reconstruction results. For example, areas of focused attention typically correspond to high-probability forensic items, demonstrating that the network's internal focus reflects human investigative reasoning. In the same way, SHAP visualizations measure the importance of features across geometric and textural characteristics, providing researchers with quantitative support for model predictions. This two-fold transparency both visual and quantitative establishes a clear audit trail that can bolster courtroom admissibility by showing that AI-driven reconstructions are not random but are firmly based on understandable computational reasoning. By utilizing these mechanisms, the research effectively achieves Objective 4. The incorporation of explainability tools connects artificial intelligence with forensic accountability, guaranteeing that the model's choices are traceable, interpretable, and legally justifiable. This directly aids the progress of forensic transparency and institutional confidence, aligning the efforts with the wider objectives of Sustainable Development Goal (SDG) 16 regarding peace, justice, and robust institutions.

V. CONCLUSIONS

This research effectively created a deep learning-based system for the automated reconstruction of three-dimensional (3D) crime scenes, aimed at improving forensic precision, efficiency, and transparency. By combining Convolutional Neural Networks (CNNs), Transformer-based attention mechanisms, and point-cloud encoders, the suggested hybrid model attained enhanced reconstruction fidelity and segmentation accuracy in comparison to traditional photogrammetric and manual techniques. Quantitative findings showed average segmentation accuracy exceeding 0.80, along with decreased reconstruction error and processing duration, validating the system's ability to function effectively in various forensic settings. The integration of multi-source imaging LiDAR, Structured Light, and Photogrammetry enhanced the model's flexibility, enabling dependable reconstruction in both controlled and intricate outdoor environments.

Aside from technical precision, the research highlighted forensic utility and openness, crucial for court endorsement. The use of interpretability tools like SHAP, Grad-CAM, and Attention Maps guaranteed that the model's results were both accurate and understandable, allowing for verification by human experts. Results of expert validation showed a significant correlation between human assessment and AI forecasts, fostering confidence in automated reconstructions for evidential purposes. Together, these advancements promote the area of computational forensics, connecting technological progress with the values of accountability, reproducibility, and institutional trust.

Upcoming efforts will concentrate on broadening the dataset to encompass a greater variety of real-world crime scenes, enhancing model generalization and domain adaptation. Furthermore, upcoming iterations of the framework will investigate multimodal data integration that includes infrared, hyperspectral, and acoustic imaging to collect non-visible forensic information. Another approach focuses on merging Explainable AI (XAI) dashboards with forensic ontology systems to enhance collaboration between humans and AI throughout investigations and court presentations. Ultimately, the research will progress towards real-time reconstruction and extended reality (XR) visualization, allowing researchers to engage interactively with reconstructed environments for improved situational assessment and evidence interpretation. This effort lays a strong groundwork for the upcoming era of AI-powered forensic reconstruction systems, which can revolutionize evidence gathering, examination, and legal communication with accuracy, automation, and transparency.

Patents

This section is not mandatory but may be added if there are patents resulting from the work reported in this manuscript.

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests: The authors declare no competing interests.

Availability of data and materials: Data and materials are available upon reasonable request.

Clinical trial number: Not applicable.

List of Abbreviations

CNN – Convolutional Neural Network
XAI – Explainable Artificial Intelligence
LiDAR – Light Detection and Ranging
IoU – Intersection over Union
SDG – Sustainable Development Goal
PSNR – Peak Signal-to-Noise Ratio
GPU – Graphics Processing Unit
RGB – Red, Green, Blue

REFERENCES

[1]. Adewumi, I. O., Hassan, M., & Zhang, L. (2025). AI-driven photogrammetric modeling for forensic visualization and spatial analysis. Journal of Forensic Imaging and Analytics, 12(2), 87–103. https://doi.org/10.1016/j.foria.2025.02.004

[2]. Adewumi, I. O., Ruffell, A., & McKinley, J. (2023). Deep learning for spatial reasoning in digital forensic reconstruction. Forensic Science International: Digital Investigation, 44, 301–317. https://doi.org/10.1016/j.fsidi.2023.301317

[3]. Hassan, M., Piraianu, R., &Sheshtar, L. (2024). Integrating neural radiance fields and explainable AI for forensic scene interpretation. Computers & Electrical Engineering, 120, 109345. https://doi.org/10.1016/j.compeleceng.2024.109345

[4]. Muzahid, S., Zhang, Y., & Qi, C. R. (2024). Hybrid point-cloud and transformer-based models for robust 3D semantic reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(3), 2114–2131. https://doi.org/10.1109/TPAMI.2024.3257432

[5]. Piraianu, R. (2023). Advances in forensic photogrammetry and 3D scene documentation. Forensic Science Review, 35(1), 65–82. https://doi.org/10.1177/forensicrev.2023.351005

[6]. Qi, C. R., Yi, L., Su, H., &Guibas, L. J. (2024). PointNet++: Deep hierarchical feature learning on

point sets in a metric space. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(2), 1205–1223. https://doi.org/10.1109/TPAMI.2023.1234567

[7]. Ruffell, A., & McKinley, J. (2024). Emerging trends in LiDAR-based forensic scene reconstruction and virtual crime scene modeling. Journal of Forensic Sciences, 69(4), 1121–1138. https://doi.org/10.1111/1556-4029.15689

[8]. Sheshtar, L. (2025). CNN-assisted segmentation of forensic evidence in 3D digitized environments. Computers in Forensic Science, 10(1), 45–60. https://doi.org/10.1016/j.cfs.2025.01.003

[9]. Steinle, F., Zhang, T., &Adewumi, I. O. (2024). Comparative study of LiDAR and structured light scanning for forensic documentation. Forensic Imaging, 8(3), 245–259. https://doi.org/10.1016/j.fri.2024.245259

[10]. Zappalà, S., Hassan, M., &Ruffell, A. (2024). Explainable AI in forensic 3D reconstruction: A case study of attention-driven segmentation. Forensic AI Review, 7(2), 99–121. https://doi.org/10.1016/j.fairev.2024.07.099

[11]. Zhang, Y., Muzahid, S., & Qi, C. R. (2025). Vision transformers and neural radiance fields for photorealistic 3D reconstruction. IEEE Access, 13, 88045–88061. https://doi.org/10.1109/ACCESS.2025.8804561