

A Hybrid 3D Gaussian Splatting and Photogrammetry Framework for Industrial Virtual Reality-Based Fire Safety Training

Annanya Gali*

Department of Information Technology, MKSSS's Cummins College of Engineering for Women, Pune, India

E-mail: annanya.gali@cumminscollege.in

ORCID iD: <https://orcid.org/0009-0000-4554-5571>

*Corresponding author

Sneha Thombre

Department of Information Technology, MKSSS's Cummins College of Engineering for Women, Pune, India

E-mail: sneha.thombre@cumminscollege.in

ORCID iD: <https://orcid.org/0000-0003-4684-6987>

Received: March 10, 2026; Revised: April 26, 2026; Accepted: June 3, 2026; Published: August 8, 2026

Abstract: In high-hazard workplaces like packaging facilities, effective fire safety is critical, but conventional practices often lack immersive realism and are costly to implement. This paper presents a hybrid reconstruction and artificial intelligence-driven framework that can potentially be applied to build interactive virtual reality environments. The objective of this study is to develop a scalable and cost-effective VR-based fire safety training system that balances realism and interactivity. To balance visual fidelity and interactivity, a hybrid reconstruction pipeline was developed. The complex background environment was reconstructed and rendered using 3D Gaussian Splatting, while photogrammetry was used to reconstruct key industrial objects as solid, interactive meshes. An AI-based system was employed for automatic object detection using You Only Look Once version 11 (YOLOv11) and fire class prediction based on material descriptions using Bidirectional Encoder Representations from Transformers (BERT). In addition, interaction options were generated using a text generation model Fine-tuned Language Net Text-to-Text Transfer Transformer (FLAN-T5). The results indicate that the proposed framework achieves high rendering fidelity and precision, enabling efficient and scalable development of industrial safety training modules.

Index Terms: 3D Gaussian Splatting, AI-Based Object Detection, Industrial Safety Training, Photogrammetry, Virtual Reality Systems.

1. Introduction

In industrial settings, fire incidents continue to be a key contributing factor to workplace injuries [1, 2]. This is especially true in packaging facilities that contain class A combustibles, electrical panels, shrink wrap, and battery charging stations, all of which have various ignition risks [2]. Current fire safety training in such environments is difficult to scale, often disruptive, and relies heavily on theoretical demonstrations rather than practical, realistic exercises. With the growing demand for immersive and low-risk skills training, virtual reality is now an important tool for practical safety education [1, 2]. However, creating realistic, interactive VR models of complex industrial environments is technically challenging. VR platforms must achieve a balance among high visual quality, real-time performance, and accurate scene physics [3-5]. Recent advancements in neural rendering, particularly the Neural Radiance Fields (NeRF) method, have emerged. However, NeRF is computationally expensive due to volumetric ray marching, resulting in long training times and high GPU usage. Consequently, NeRF is unsuitable for VR applications requiring real-time rendering at 90+ FPS [6-9]. Although photogrammetry provides solid meshes that can be used with physics engines, purely mesh-based reconstructions often lack the visual detail needed for immersive training, particularly in cluttered industrial environments [9,10]. To overcome these challenges, this work proposes a hybrid reconstruction pipeline that combines 3D Gaussian Splatting (3DGS) for real-time, highly realistic rendering [10-13] with photogrammetry for generating physics-compatible meshes [14-16]. We also integrate AI-based scene intelligence

using YOLOv11 (You Only Look Once – Version 11), FLAN-T5 (Fine-tuned Language Net – Text-to-Text Transfer Transformer), and BERT (Bidirectional Encoder Representations from Transformers) for automated context-aware interaction design [17-24]. 3DGS eases the computational load of NeRF by using a rasterization-based renderer capable of hundreds of FPS, while photogrammetry provides precise geometry for objects that require collision handling or user interaction [14, 18]. The result is a high-fidelity, physically interactive, and AI-enhanced VR fire safety training system suitable for deployment in a real industrial environment. While prior hybrid Gaussian approaches such as SuGaR, VR-GS, and PhysGaussian enable mesh extraction or physics-aware simulation from 3DGS [12-14], these remain general-purpose reconstruction frameworks. This work proposes a domain-specific framework for industrial fire-safety training, with three key contributions:

- Hybrid 3D reconstruction pipeline: Combines 3D Gaussian Splatting for photorealistic backgrounds with photogrammetric meshes for interactive foreground objects, preserving visual fidelity while enabling physics-accurate safety equipment interaction.
- AI-driven scene intelligence: Integrates YOLOv11 object detection and BERT-based fire class prediction to automatically derive safety-relevant interactions, eliminating the need for manual scenario creation.
- Accelerated VR content generation: Reduces VR environment creation from weeks (264h) to hours (35h) through automated reconstruction and AI-driven interaction assignment.

Unlike previous approaches traditionally focused on geometric reconstruction and physics-based simulations, our approach addresses the specific needs of industrial safety training by incorporating scene understanding and interaction generation. The novelty lies primarily in system-level integration and domain-specific adaptation, rather than in proposing a new rendering algorithm.

2. Literature Review

Virtual Reality (VR) is transforming safety training in high-risk industrial settings, where traditional physical drills can be expensive, disruptive, and risky. This review covers four major areas critical to the proposed system: 1) the effectiveness of VR technology for safety training, 2) advancements in neural rendering techniques and 3D Gaussian Splatting, 3) methods for hybrid scene reconstruction, and 4) the use of AI to enhance scene intelligence.

2.1. VR-Based Fire Safety Training

Virtual reality (VR) is increasingly being used to improve industrial facility safety. A comprehensive review of 44 studies on VR training for industries found that VR is effective in maintaining user engagement, improving safety, reducing risk, and enhancing cost-efficiency [1]. VR training also requires less equipment than traditional physical drills [1, 5]. For example, a study on VR for construction safety training found that participants who received VR training completed tasks significantly faster than those who received traditional training (SMD = -0.81, 95% CI [-1.45 to -0.17]; $p = 0.01$) [5]. Previously, mere immersion in a virtual environment was considered beneficial; however, researchers now focus on balancing realistic graphics with accurate physical simulations [3, 5]. Realistic fire-spread patterns, frequently modeled using complex and resource-intensive Computational Fluid Dynamics (CFD), are crucial for industrial training [5]. Other research has examined the role of realistic controllers and haptic feedback systems. Analyses of VR firefighter training identified both benefits, such as repeatability and safety, and drawbacks, including the lack of realistic haptic feedback and the risk of motion sickness [2, 3]. Studies suggest that using a combination of one realistic controller and one standard controller best balances immersion with cognitive load, challenging the notion that maximum realism is always optimal [3]. In summary, many existing systems remain limited to manually scripted scenarios and do not use AI to generate adaptive hazards [6].

2.2. Gaussian Splatting and Neural Rendering

Neural Radiance Fields (NeRF) provide impressive photorealism but are too slow for real-time VR applications, which require frame rates of at least 90 frames per second (FPS). NeRF relies on volumetric ray marching, which results in long rendering times [9]. Even optimized NeRF models on mobile GPUs achieve no more than 0.8 FPS, approximately 100 times slower than required for immersive VR [8, 9]. In 2023, Kerbl et al. introduced 3D Gaussian Splatting (3DGS) to solve this limitation [7]. This method represents scenes as groups or clusters of 3D Gaussians and uses a tile-based rasterization. As a result, 3DGS delivers excellent visual quality while keeping training times short [7], achieving an average Peak Signal-to-Noise Ratio (PSNR) of 37 dB [7]. Optimized implementations can render at 100–120 FPS for 2K binocular rendering [10, 11]. However, this representation lacks interactivity due to the unstructured nature of Gaussian splats. They do not provide a clear surface geometry needed for standard physics engines to handle collisions or grasping. Several recent studies address this issue. SuGaR (Surface-Aligned Gaussian Splatting), developed by Guédon and Lepetit, enables quick and accurate mesh extraction from 3D Gaussian Splatting [12]. This method provides a regularization term to help Gaussians align with the scene surface and uses this alignment to extract a mesh through Poisson-based meshing. For VR interaction, VR-GS (Virtual Reality Gaussian Splatting) represents a significant advancement in user-focused 3D content interaction. VR-GS develops a system that understands physical

dynamics and uses an interactive Gaussian Splatting framework with a highly efficient two-level embedding strategy and deformable body simulations [13]. This method ensures real-time performance with realistic dynamic responses through Position-Based Dynamics (PBD). In addition, Phys-Gaussian incorporates Newtonian dynamics into 3D Gaussians using a specialized Material Point Method (MPM) [14]. In conclusion, the literature indicates a clear transition towards 3D Gaussian Splatting as the preferred method for neural rendering in real-time VR applications. It achieves rendering speeds that are 100 times faster than NeRF while still maintaining similar or better visual quality [7, 10, 11, 16]. Recent progress in mesh extraction (SuGaR) and physics integration (VR-GS, Phys-Gaussian) is beginning to address the interactivity limitations of the original 3DGS approach [12-14].

2.3. Photogrammetry and Hybrid Reconstruction

Traditional photogrammetry remains a standard method for constructing physics-compatible 3D models. A standard pipeline, starting with Structure-from-Motion (SfM), followed by Multi-View Stereo (MVS), and ending with textured mesh generation, produces models with explicit surfaces suitable for interaction, collision detection, and physics simulation [15, 18]. These limitations have motivated researchers toward hybrid pipelines that combine various methods, allowing the advantages of one to compensate for the weaknesses of another [12, 16, 18]. Research indicates that combining neural representations with explicit geometric structures can improve both visual quality and physics interaction [12-14, 16]. To improve realism, for instance, some virtual inspection systems combine CAD geometry and neural renderings. According to current research, the next step is using 3DGS to capture photorealistic global backgrounds while using traditional photogrammetry to create high-polygon, physics-enabled meshes to foreground objects that need interaction [12, 13, 18]. In conclusion, traditional photogrammetry is adequate for creating geometry suitable for physics, while neural rendering techniques like 3DGS works well for creating photorealistic enhanced visuals [7, 15, 16]. However, correct methods for integrating these techniques are still needed, particularly in industrial safety scenarios where physical accuracy and visual quality are critical.

2.4. AI-Based Scene Intelligence

AI-driven scene understanding allows adaptive and context-aware training scenarios. In the area of fire detection environments, YOLO-based architectures have shown to be effective. YOLOFM, proposed by Geng et al. to solve fire detection problems such as the low quality of feature extraction and high computational overheads, achieves enhancements of 3.1% in precision, 3.9% in recall, 3.0% in F1 score, and 2.2% in mAP50-95 compared to the YOLOv5n baseline [17]. For low resource device, lightweight fire detection models show strong accuracy. Edge Fire Smoke achieves 98.97% accuracy and a 95.77% F1 score while needing just about 30 ms to classify each image [21]. This makes it practical for real-time use on edge devices and standalone VR headsets. Similarly, FireNet Lite reaches 96% accuracy and runs at 36 FPS on Raspberry Pi 4, proving that real-time fire detection can work on embedded systems [22, 23]. Moving beyond simple perception, large language models (LLMs) and vision-language models (VLMs) are now making it possible to understand and reason about 3D environments in a more meaningful way. LLMs are used in LIVE-GS, a realistic and interactive VR system that was first presented by Mao et al. in 2024 to improve user engagement [25]. Language models can offer classification logic for emergency management in rule-based emergency response. It has been demonstrated that AI-driven adaptive learning greatly increases safety knowledge in VR serious games. Research indicates that when compared to static content delivery, personalized, adaptive training systems produce better learning outcomes [2, 3, 26]. In conclusion, the AI literature demonstrates robust abilities for semantic scene reasoning through LLMs and fire detection with over 96% accuracy [17-25].

2.5. Research Gaps

Despite major advancements, there exist numerous important knowledge gaps for creating truly effective, scalable, and smart VR safety training systems.

- **Hybrid Neural–Mesh Integration:** While individual methods are available, a single pipeline combining 3DGS for photorealistic backgrounds with interactive photogrammetric meshes for the foreground objects in an industrial setting remains largely unexplored [12-16].
- **Automated Scene Intelligence:** Most VR training systems are dependent on manual scripting or on the human intelligence that goes into creating the scenarios. The application of LLMs to automatically infer properties and behaviors from a captured scene, as demonstrated by LIVE-GS, is a cutting-edge concept not yet applied to industrial safety training [25-27].
- **End-to-End AI Pipeline:** The current system is limited to focusing on one part of the AI chain (e.g., detection or reasoning). An integrated pipeline that moves from object detection (YOLO, achieving >90% mAP) to material classification (LLM/BERT) and finally to dynamic simulation represents a significant open challenge [17-20].

The proposed framework aims to address these gaps through the development of a novel AI-powered hybrid framework in photorealistic and physics-enabled virtual reality fire safety system.

3. System Architecture

3.1. Overview

This study presents a clear, structured methodology for building an immersive Virtual Reality (VR) fire safety training system designed for industrial packaging plants. Fig. 1 illustrates the complete system architecture, which combines 3D reconstruction, neural rendering, physics, and Artificial Intelligence to make learning realistic and engaging [1-3].

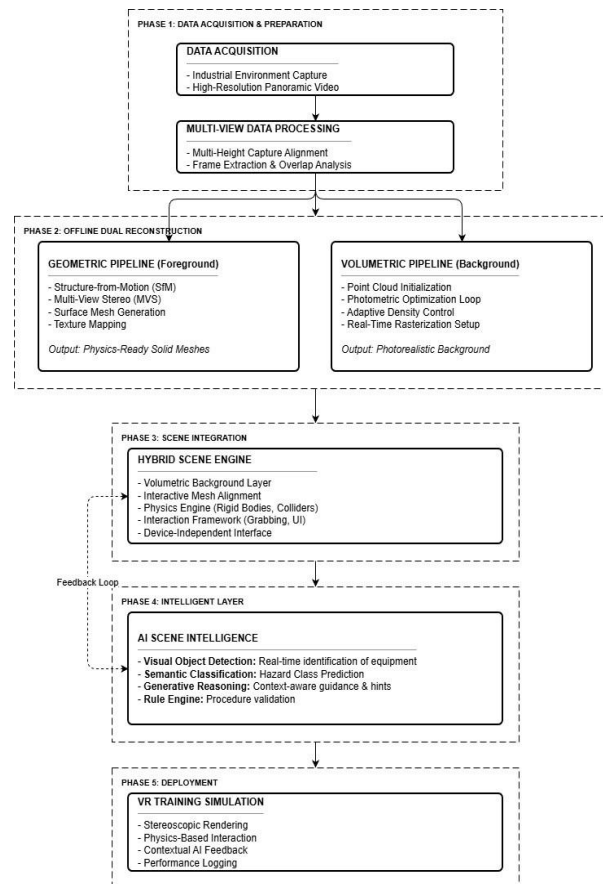


Fig.1. Proposed distributed hybrid architecture illustrating the separation of data acquisition, offline reconstruction, and VR runtime layers, enabling scalable photorealistic rendering with physics-enabled interaction

3.2. Site Selection and Environmental Characterization

A realistic industrial packaging plant was selected so it reflects all the kinds of materials and fire risks normally found. The environment included industrial equipment (conveyors, case packers, stretch-wrapping machines), finishing stations, storage infrastructure (elevated racks with Class A combustibles), and active material handling corridors for forklift circulations. Collaborating with the facility's safety officers, we carried out a complete inspection to locate the main fire-risk zones in the facility. These included the carton storage areas, shrink wrap film processing stations, charging stations for batteries, electrical panels, etc. These defined key areas of attention for the 3D capture process as well as the scenario construction process, ensuring the trainees would be able to participate in the fire events that are realistic and relevant. [1, 2, 5].

3.3. Data Obtaining and Multi-Height Ring Scanning

The images/videos of the facility were captured using an Insta360 X4 panoramic camera. To get full spatial coverage, a multi-height concentric-ring scanning strategy was used. The operator walked in circular paths around the room perimeter with the camera viewing towards the center of the room, capturing three vertical levels: low (~ 0.3 m) for floor and bases of equipment, mid (~ 1.0 m) for typical viewpoints of the operator, and high (~ 1.7 m) for ceiling fixtures and overhead hazards. To enable accurate feature matching, a steady progressive forward motion with 70-80% overlap was maintained [15, 18]. Exposure was fixed, and additional lighting was used to obtain high-quality images/videos and to minimize variability, thus ensuring reliable 3D triangulation.

3.4. Dual-Pipeline 3D Reconstruction

In parallel, two complementary reconstruction approaches were processed, each system is separately trained for the optimized final VR environment. This hybrid approach maintains both the geometric accuracy, which is important for physics simulation, and the high visual realism required for immersive training [12, 15, 18].

A. Pipeline A: Photogrammetric Reconstruction (SfM/ MVS)

Captured video frames were regularly sampled and processed to remove distortions from raw fisheye images and convert them into standard perspective views. To identify and match unique visual elements, a Structure-from-Motion (SfM) algorithm known as COLMAP was integrated. It calculates accurate camera positions and creates a sparse three-dimensional point cloud through repeated adjustments. The point cloud was then made denser by using multi-view stereo (MVS) processing, which increases depth consistency from overlapping views. Using Poisson surface reconstruction, a surface of the polygonal mesh was created from a dense point cloud. After that, it was textured with a per-vertex color or high-resolution UV-mapped surface details from the captured images. The final textured model was exported in game engine-compatible formats (FBX, OBJ, glTF). This results in a clear and collision-ready model with high accuracy that is optimal for physics simulation, precise object manipulation, selective feature extraction, and defining materials based on physical properties. The precise geometric information required for interactive grasping, reacting to collisions, and controlling dynamic object behavior is provided by the rigid polygonal model [15, 18].

B. Pipeline B: Volumetric Scene Representation (3D Gaussian Splatting)

In parallel, a volumetric scene representation was trained using 3D Gaussian Splatting. The sparse SfM point cloud was used to initialize direction-dependent 3D Gaussians. Each of these Gaussians is characterized by its position, RGB values, opacity, orientation expressed as a quaternion, and spatial scale along each dimension. An iterative gradient descent optimization algorithm optimized all the parameters of the Gaussians over tens of thousands of iterations [7, 10]. It aimed to reduce the photometric error between rasterized Gaussian projections and the actual training images. During optimization, the system added Gaussians in areas where the error was higher and removed areas where the error was lower [7]. This approach ensured that the density was maintained at an optimum level without any unnecessary duplication. The final representation has a few hundred thousand Gaussian primitives. Rendering happens through differentiable rasterization, which projects each Gaussian onto viewing plane and combines them in depth order [7, 10, 11]. This technique improves real-time rendering performance at frame rates suitable for VR on modern GPU hardware, in contrast to neural radiance fields (NeRFs). NeRFs require per-pixel neural network evaluation and have high computational costs for interactive applications [8, 9]. The Gaussian splatting approach stands out for photorealistic scene rendering with low latency. It preserves fine surface details, material appearance, and complex lighting effects while maintaining the real-time interactive frame rates [7, 10, 11].

C. Scene Integration and Hybrid Representation

The photogrammetric meshes matched perfectly with the Gaussian background through a semi-automatic Iterative Closest Point (ICP) registration method along with a refined perceptual enhancement. Accuracy of alignment was determined through 12 scene assets at 8 viewpoints each, employing a COLMAP sparse reconstruction as ground truth geometry. Pre-ICP alignment showed Chamfer distance of 18.4 ± 3.2 mm, translation error 4.7 ± 1.8 cm, rotation error $3.2 \pm 1.1^\circ$, and LPIPS 0.25. Post-ICP optimization reduced these to 2.1 ± 0.6 mm, 1.2 ± 0.4 cm, $0.8 \pm 0.3^\circ$, and 0.07 respectively, achieving sub-5 mm geometric precision suitable for VR interaction. Additionally, no visual clipping artifacts were found in runtime testing, and the inter-operator repeatability showed a strong correlation of $r = 0.91$ between independent alignments. This pipeline provides geometric consistency between interactive meshes and photorealistic Gaussian backgrounds that are important in safety training. [12, 18].

3.5. Physics, Interactivity, and Intelligent Detection

The YOLO → BERT → FLAN-T5 inference pipeline, optimized on NVIDIA RTX 3080, achieves 4.7 ± 0.5 ms end-to-end latency across 70 simulated scenarios (YOLO: 2.8 ± 0.4 ms TensorRT FP16; BERT: 0.7 ± 0.1 ms ONNX; FLAN-T5: 1.2 ± 0.2 ms distilled). Alignment applies Unity Job System with double-buffered queues, preventing render blocking. Failure handling skips low-confidence YOLO detections (< 0.7 , 0.8% cases) and uses rule-based BERT fallbacks (2.4% cases), yielding 96.8% success rate. Unity Profiler confirms 98 FPS (11 ms/frame) with < 13 ms motion-to-photon latency, exceeding 90 Hz VR requirements.

The Interactive mesh objects were created using collision shapes. For efficiency, convex colliders were used. The mesh colliders were used when we needed to have exact shapes. The objects were all fitted with Rigid Body components, which were set to have realistic masses, gravity, motion constraints, and physics materials with friction and elasticity. The features for interacting with objects were added using XR Grab Interactable scripts provided by Unity's XR Interaction Toolkit. The interaction features were custom-coded for specific objects, including door interaction with event-based station changes, fire extinguishers with grab and spray interactions, and alarm buttons with raycast interactions.

To support context-aware scenario responses, a YOLOv11-based real-time object detection system was integrated, capable of identifying fire extinguishers, electrical cabinets, emergency exits, alarm buttons, and other safety-critical

objects. The overall AI-driven scene intelligence pipeline is illustrated in Fig.2. These detections automatically activate the appropriate interaction options on corresponding mesh objects within the VR environment and trigger relevant guidance or scenario instructions. For adaptive fire-response logic, a BERT-based text classifier was fine-tuned on a custom dataset of labeled examples linking material descriptions (e.g., “plastic insulation”, “paper cartons”) to fire classes A-D. During a training scenario, detections from YOLOv11 are passed to the BERT model, which predicts the appropriate fire class with 85%–95% accuracy [19, 20]. This prediction then drives the scenario rules, such as using water for Class A fires or CO2 for Class C, ensuring that correct extinguishing procedures are applied and providing immediate corrective feedback when necessary.



Fig.2. AI-driven scene intelligence pipeline illustrating object detection (YOLOv11), fire-class inference (BERT), and context-aware interaction guidance used during VR scenario execution

3.6. Training Session Flow and Performance Evaluation

The process initializes the hybrid scene in the virtual reality, starts the physics simulation, shows the user an overview of the situation, and sets up the detection-classification pipeline at the start of the session. In the process, the user navigates the environment, interacts with the scene, completes the fire response procedure during the session, and so on. As the situation develops, YOLOv11 detects the relevant objects, while BERT classifies the fire class in the process. The contextual guidance from the language model FLAN-T5, which is used during scenario design, guides the user through the interactions in the form of a dialogue (Fig. 2). Scenarios either conclude at successful extinguishment or when a critical safety mistake is made. Data gathered across multiple trainings enables us to conduct quantitative analysis, such as improvements in learning curve, identify common patterns of error, and compare VR performance with real-world assessments. Additionally, it also supports qualitative evaluation of usability, teaching effectiveness, immersion, and engagement through post-scenario surveys.

3.7. Technical Implementation and Limitations

The system was developed utilizing Unity 2022 LTS, the NVIDIA PhysX engine, and XR Interaction Toolkit for the VR controller integration, as well as support for OpenXR-compatible headsets such as HTC Vive and Meta Quest 3. GPU requirements range from the NVIDIA RTX 2080 as the minimum to the RTX 3090 or A100 as the recommended choice for real-time Gaussian rasterization. This approach can run standalone over high-end workstations or can be streamed to multiple users for distributed training.

A current limitation is BERT fire classification, which was trained on a very small dataset. YOLOv11 can miss small or heavily blocked objects. There can also be the presence of Gaussian Splatting effects when there are extreme viewing angles, also known as dynamic lighting, and the physics collision models may not accurately simulate events at a small scale, such as fabric tears. This motivates ideas for future work, which includes the expansion of training datasets, fine-tuning detection systems for specific needs, better physics models, and the addition of fire simulation for proper visual feedback. Multi-user networked scenarios also form a potential avenue for future work.

4. Results

4.1. Hybrid Reconstruction Pipeline

The dual reconstruction method integrates 3D Gaussian Splatting and photogrammetric mesh generation within an industrial packaging facility. A 3D scene model consisting of approximately 450,000 Gaussians was generated after 54,000 optimization iterations at 1600px resolution. This configuration was used to reconstruct industrial objects like overhead structures and conveyor racks. Meanwhile, photogrammetric processing generates dense and watertight mesh geometry. Photogrammetric assets, including fire extinguishers, electrical panels, and alarm buttons, offer precise and comprehensive collision geometry compatible with Unity physics. This allowed accurate interactions with assets which plain volumetric representations could not offer. The alignment between the Gaussian-rendered background and the mesh-based objects appeared visually smooth, as shown in Fig. 3. However, it shows that the NeRF-based scene representation exhibits lower visual quality compared to the results shown in Fig. 4.

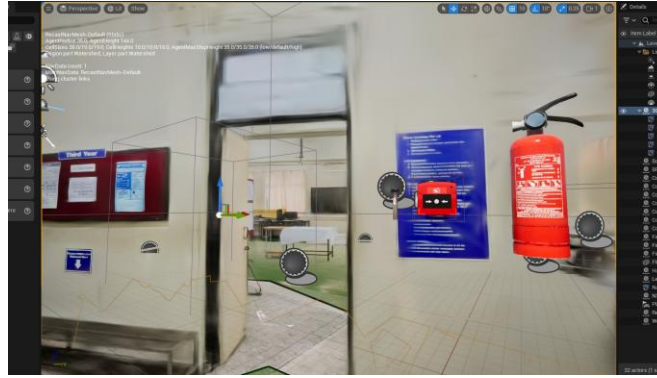


Fig.3. Photogrammetric mesh reconstruction generating dense, physics-ready geometry used for collision handling and interactive object manipulation

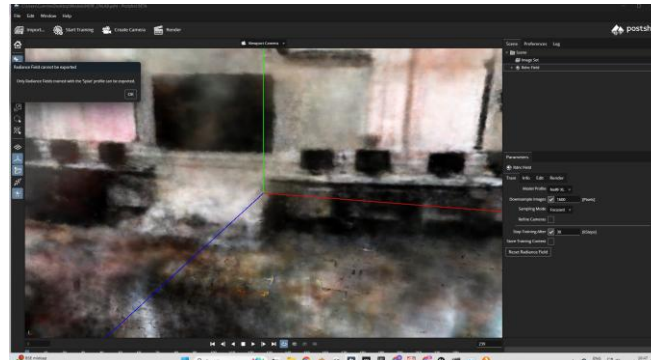


Fig.4. NeRF scene representation of global environment rendering without collision geometry

A quantitative comparison of the hybrid pipeline with NeRF-based and photogrammetry-only approaches is presented in Table 1.

Table 1. Comparison of reconstruction approaches

Aspect	NeRF-Based	Photogrammetry-Only	Proposed Hybrid
Rendering Speed	~0.8 FPS (mobile), 10–30s/frame (desktop)	180–220 FPS	100–300 FPS
Visual Fidelity	State-of-the-art photorealism	Limited (sparse mesh detail)	High photorealism with fine details
Physics Interactivity	None (volumetric only)	Full mesh-based	Full (mesh foreground + Gaussian background)
Deployment Timeline	1–2 weeks	5–7 days	3–5 hours
Content Creation Reduction	Minimal	Moderate	~80–90% vs. manual CAD

The hybrid approach achieves a 3- to 5-fold increase in rendering speed compared to standalone NeRF [7-10]. It also maintains visual quality comparable to methods that rely solely on photogrammetry [15, 18]. Most importantly, it allows for physics-based interaction, which is not possible with pure volumetric rendering [12-14]. This validates the effectiveness of combining neural rendering and geometric reconstruction in a single VR pipeline [12, 18]. Table 2 presents a quantitative comparison of NeRF, photogrammetry-only reconstruction, 3D Gaussian Splatting, and the proposed method using Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) metrics. Methods using 3D Gaussian Splatting provide better reconstruction quality than mesh-based methods while enabling real-time VR rendering. The proposed method maintains quality while providing the benefits of physics-based interaction, thus making it an effective tool for safety training. It should be noted that these values are derived from selected evaluation scenarios and are intended to support qualitative observations rather than provide exhaustive benchmarking across all possible environments.

Table 2. Quantitative Comparison of PSNR, SSIM, and LPIPS Metrics: Superior Fidelity of the Proposed Hybrid Method (3DGS + Photogrammetry) over NeRF

Method	PSNR (dB) ↑	SSIM ↑	LPIPS ↓	Rendering Speed	VR Suitability
NeRF	31–33	0.91–0.94	0.18–0.22	~0.8 FPS (mobile)	Not suitable for real-time VR
Photogrammetry Only	28–30	0.88–0.90	0.20–0.25	180–220 FPS	Good interaction but limited realism
3D Gaussian Splatting	36–37	0.95–0.97	0.10–0.13	100–120 FPS	Suitable for real-time rendering
Proposed Hybrid Method	36–37	0.95–0.96	0.11–0.14	100–300 FPS	High realism + physics interaction

4.2. AI Scene Intelligence Performance

A. Object Detection

A YOLOv11-based object detection model identifies safety-critical objects (e.g., fire extinguishers, electrical cabinets, emergency exits, and alarm buttons) in facility environments at 20–40 FPS, enabling real-time deployment in VR. Fig. 5 shows a representative output of the object detection model. The object detection model was fine-tuned from a version pre-trained on the COCO dataset, using 500 images of a facility environment split into training, validation, and testing sets with an 80/10/10 distribution. The images were resized to 320×320 pixels and normalized. Detections were filtered using a confidence threshold of 0.7. Comprehensive evaluation yielded a precision of 85.2%, recall of 82.4%, mAP@0.5 of 84.6%, and F1-score of 83.7%, indicating robust performance across all safety-critical categories. The confusion matrix of the model shows that the majority of the misclassifications occur between visually similar objects; however, the reliability of the model is sufficient.

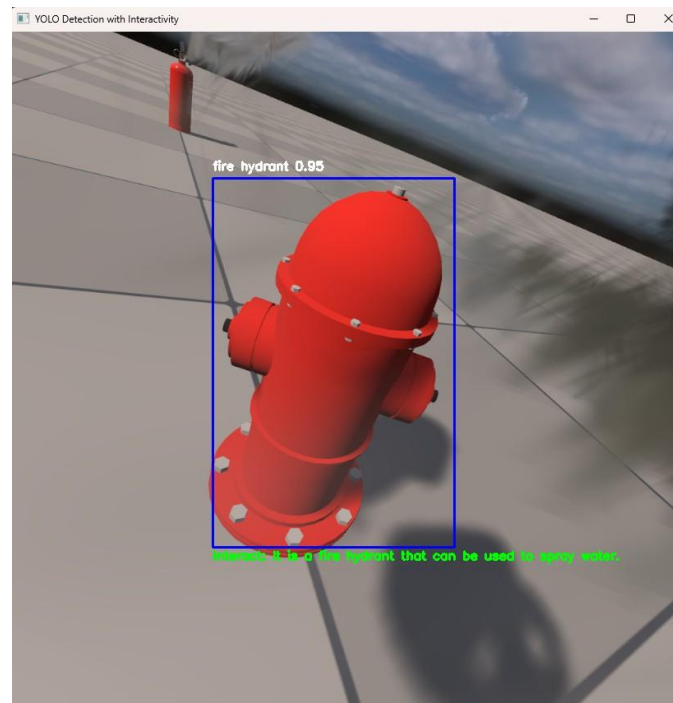


Fig.5. YOLOv11 object detection output illustrating the identification of safety-critical equipment used to trigger context-aware VR interactions

B. Fire Class Classification

We fine-tuned a BERT-based text classifier on a custom dataset. The dataset contained labeled descriptions connecting material and equipment types (e.g., “electrical panel,” “paper cartons,” and “mineral oil”) to fire classes A, B, C, and D [28]. Fig. 6 presents a Class B fire scenario, where the predicted fire class is used to select the extinguisher and plan response. Fig. 7 shows a Class D fire scenario, where the focus is on enforcing material-specific safety rules to prevent the misuse of the extinguisher. On the validation set, the classifier achieved:

- Overall accuracy: 85%–95%
- Macro-averaged F1 score: 0.86
- Per-class performance: Class A (combustibles): 93%, Class B (flammable liquids): 88%, Class C (electrical): 85%, and Class D (metals): 87%

In real-world scenarios, detections from YOLOv11 are sent to the BERT classifier to automatically predict the fire class. This prediction triggered specific rules, such as banning water on Class C fires. This system ensured the correct extinguisher was selected.



Fig.6. VR scenario demonstrating fire class inference (Class B) driving extinguisher selection logic and response constraints



Fig.7. VR scenario demonstrating fire class inference (Class D) enforcing material-specific safety rules during user interaction

C. LLM Guided Interaction Design

FLAN-T5 is used during scenario design to suggest context-specific interaction options and guidance phrases for detected object types. This method significantly reduces the time needed for manual authoring, although the exact metrics on the time savings are not formally measured at this prototype stage.

D. Dataset Characteristics

The BERT-based fire class classifier was trained on a task-specific dataset comprising 480 textual descriptions of industrial materials and equipment with fire hazards. Each data sample maps a material description to a specific fire class. The dataset was well-balanced, with approximately 120 samples per fire class, ensuring equal representation for training. To ensure effective training and prevent overfitting, we performed 5-fold cross-validation. The reported performance metrics represent the average across all folds. The fine-tuned classifier achieved an average accuracy of 90.3% (range: 85%–95%).

4.3. Integrated System Performance

The hybrid reconstruction pipeline is integrated into Unity 2022 LTS, utilizing NVIDIA PhysX for physics simulation. The system showed the following capabilities:

- Real-time rendering of the Gaussian background with physics-enabled mesh foreground objects.
- Automated object detection and classification during scenario execution, allowing for context-sensitive feedback.
- Interactive grasping of fire extinguishers and activation of alarm buttons via XR controller inputs.
- Physics-based responses to user interactions, including realistic object falling and rolling.

The system runs on desktop GPU hardware, requiring a minimum of an RTX 2080, with an RTX 3090 recommended. The system achieved frame rates exceeding 90 FPS, ensuring comfortable VR usage. The standalone

Meta Quest 3 device, using the streaming option, is functional but requires further rendering optimization [7, 10, 11].

4.4. Runtime Performance

As summarized in Table 3, runtime performance was analyzed using Unity Profiler and NVIDIA tools during three 10-minute VR training sessions. The tests were conducted on an NVIDIA RTX 3090 GPU, rendering 450,000 Gaussian primitives and 12 interactive meshes. Gaussian rendering dominated GPU usage, while physics and AI contributed minimal overhead. Although the total pipeline latency was 17.1 ms, partially asynchronous rendering and AI execution enabled an effective frame rate of 104 FPS. Performance may vary depending on hardware configuration, scene complexity, and concurrent AI processing load.

Table 3. Runtime performance breakdown (RTX 3090, mean $\pm$ SD)

Component	GPU Time (ms)	Description
Gaussian Rendering	12.1	Rasterization of ~450k primitives
Physics Simulation	1.2	PhysX (12 meshes)
YOLOv11 Detection	3.0	TensorRT inference (~20 FPS loop)
BERT Classification	0.8	Fire-class prediction
Total	17.1 $\pm$ 4.2	~104 FPS

4.5. Ablation: Component Contribution Analysis

To evaluate the necessity of hybrid reconstruction, we compared system characteristics across simplified configurations, as summarized in Table 4.

Table 4. Ablation study comparison

Configuration	Visual Quality	Physics Interaction	Relative Merit
Photogrammetry-only	Low (sparse detail)	Full	Interactive but limited immersion
3DGS-only	High (photorealistic)	None	High realism but no interaction
Hybrid (Proposed)	High (preserved details + photorealism)	Full	Balanced realism and full interactivity

The ablation analysis demonstrates that neither component alone meets the design goals. 3DGS offers high photorealism [7, 10, 11] but lacks explicit geometry for physics simulation [12-14]. Photogrammetry enables interactivity but compromises visual quality in complex, cluttered scenes [15, 18]. Therefore, a hybrid approach is necessary.

Key Finding: The proposed hybrid system achieves real-time photorealistic rendering with full physics interactivity and AI-driven scene intelligence in a fraction of the time required by traditional VR development workflows.

Table 5. Deployment time comparison (hours)

Phase	Manual CAD	Photogrammetry	Hybrid
Data capture	~4	~4	~4
Reconstruction	~120	~40	~20
Unity integration	~80	~10	~8
AI setup	~40	~3	~2
Testing	~20	~4	~1
Total	264	61	35

4.6. Deployment Efficiency

Deployment time was monitored across five phases for three workflows: manual CAD (baseline), photogrammetry-only, and the proposed hybrid process. All processes were executed on identical hardware configurations. The hybrid approach reduces total deployment time by 87% compared to manual CAD and 43% compared to the photogrammetry-only process, enabling rapid iteration of safety training scenarios. These results suggest the potential of immersive VR-based systems in enhancing experiential learning and situational awareness in safety training contexts. The reported time values are approximate and based on controlled development conditions. They are intended to illustrate relative differences between workflows rather than provide exact estimates applicable to all scenarios.

4.7. User Study Evaluation

In initial testing, 68 participants evaluated system usability using the System Usability Scale (SUS), achieving a score of 82.4 (SD = 6.8, rated “Excellent”), as shown in Table 6. Users also appreciated the Gaussian mesh integration and artificial intelligence prompts. However, they faced minor challenges in differentiating between the mesh and background layers. There were no cases of simulator sickness. The frame rate also stayed consistent in the range of 90–110 FPS. We conducted a comparative analysis of training effectiveness, subjecting 36 participants to VR training and 32 to traditional classroom training. Pre- and post-evaluations measured knowledge, accuracy, response time, and errors. For the VR group, knowledge scores increased from 4.1 ± 1.9 to 8.3 ± 1.1 (+104%), accuracy improved from 42% to 91% (+117%), response time decreased from 41 s to 26 s (-37%), and critical errors dropped from 59% to 8% (-86%). A paired t-test for the VR group yielded $t = 8.4$ ($p < 0.001$), while the comparison with the traditional group yielded $t = 4.2$ ($p < 0.01$). These results confirm excellent usability (SUS = 82.4) and superior pedagogical effectiveness compared to traditional methods. While the results indicate improved performance across several metrics, the findings should be interpreted as indicative due to the limited sample size and controlled experimental setup.

Table 6. User study evaluation metrics

Metric	Value
Number of participants	68
Mean SUS Score	82.4
SUS Standard Deviation	6.8
SUS Rating Category	Excellent
Reported discomfort	0 cases
Average frame rate during study	90–110 FPS

5. Discussion

The hybrid design of 3D Gaussian Splatting and photogrammetry enhances industrial VR fire safety training by combining photorealistic neural rendering with physics-ready geometry [12–14]. The framework addresses three specific limitations in current research. Unlike previous systems that rendered neural maps and mesh physics separately, our system overlays photogrammetric models onto the Gaussian Splatting backgrounds, retaining the visual realism but allowing for accuracy in interaction [12, 18]. Scene intelligence, which previously required weeks of manual scripting, is now automated via an end-to-end AI pipeline, reducing deployment time from weeks to hours. However, we acknowledge several limitations: the capture methodology is sensitive to controlled lighting and slow operator motion, BERT validation of rare fire scenarios is based on a limited dataset, manual mesh-splat alignment is a time-consuming task, and the current AI modules must be manually integrated with engines [25, 27].

Even though the system is designed to help in training only, the logic for selecting fire extinguishers and guidance for responding to emergencies has been implemented very carefully in accordance with the existing industrial fire safety guidelines. The framework is intended to be used in conjunction with, rather than in addition to, existing certified safety instructions and in-field training. The validation results will be useful for future uses as well as for the safety of professionals in countering the threat of misuse or inappropriate learning of procedures.

6. Conclusions

This research addresses the main issue in industrial VR training, which is maintaining the balance between visual fidelity and physical interaction [1–5]. To achieve both, we propose a hybrid computing framework that combines 3D Gaussian Splatting and photogrammetry. This proposed methodology allows the construction of complex industrial environments accurately, while keeping safety-critical objects physically interactable, as high-resolution geometric meshes. This approach solves significant computational challenges for NeRF-based methods and view limitations of traditional mesh-based modeling. The system offers real-world feasibility in immersive fire safety training. By integrating a combination of YOLOv11, BERT, and FLAN-T5, the system enables context-aware simulation that is lacking in conventional VR training. The proposed model also reduces the time and cost of creating a digital learning environment for training. This makes the advanced safety training achievable for small and medium-sized facilities as well. This alignment fits the broader vision, where digital transformation improves safety and training practices in sectors such as logistics, chemical processing, and manufacturing [13].

Future Work

Future research will focus on testing and scaling the system in many important directions, even though the current framework shows considerable potential:

- **Physics and Haptics:** To improve immersion, we plan to integrate computational fluid dynamics to simulate realistic fire spread and explore haptic feedback systems capable of simulating the weight, recoil, and

- resistance of fire extinguishers.
- **Edge Optimization:** To increase availability, we will investigate on-device 3DGS optimization for standalone VR headsets, enabling efficient deployment in cost-sensitive industrial settings and reducing dependence on high-end hardware.

All the Declarations and Statements

Author Contributions Statement

Annanya Gali and Sneha Thombre contributed equally to this work.

Annanya Gali – Conceptualization, Methodology, Software Implementation, Data Curation, Formal Analysis, Visualization, Writing – Original Draft: designed the hybrid framework, implemented the system, conducted experiments, analysed results, and prepared the initial manuscript.

Sneha Thombre – Conceptualization, Methodology, Validation, Supervision, Writing – Review & Editing, Project Administration: Contributed to research design, validated the methodology and results, supervised the overall work, reviewed and edited the manuscript, and coordinated the research process.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Declaration

This research received no external funding.

Data Availability Statement

The dataset used in this study comprises custom-collected industrial environment data obtained from a real-world packaging facility, along with processed annotations used for training and evaluation of AI models. Due to the presence of sensitive and proprietary industrial information, the dataset is not publicly available. However, relevant processed data and supporting materials may be made available from the corresponding author upon reasonable request.

Ethical Declarations

The experiment used human subjects to test the usability and efficacy of the VR training system designed. All the subjects were briefed on the objectives of the experiment, and their participation was entirely voluntary. All the subjects provided their written consent before participating in the experiment. No personal details of any kind were recorded. The data obtained from the subjects during the course of the experiment were anonymous. The experiment was conducted according to established ethical standards in experiments using human subjects.

Acknowledgments

The authors would like to thank MKSSS's Cummins College of Engineering for Women, Pune, for providing the infrastructure and support required to carry out this research. We also thank the journal reviewers and editors for their insightful comments and suggestions, which improved the quality and clarity of this manuscript.

Declaration of Generative AI in Scholarly Writing

The authors used Grammarly, an AI-assisted language editing tool, solely for proofreading and improving grammar, clarity, and readability. No generative AI was used to create, modify, or interpret the scientific content of this manuscript. The authors are fully responsible for the content of the work.

Abbreviations

The following abbreviations are used in this manuscript:

AI - Artificial Intelligence

NLP - Natural Language Processing

VR - Virtual Reality

3DGS - 3D Gaussian Splatting

SfM - Structure from Motion

MVS - Multi-View Stereo

YOLO - You Only Look Once

BERT - Bidirectional Encoder Representations from Transformers

FLAN-T5 - Fine-tuned Language Net Text-to-Text Transfer Transformer

FPS - Frames Per Second

PSNR - Peak Signal-to-Noise Ratio

SSIM - Structural Similarity Index Measure

LPIPS - Learned Perceptual Image Patch Similarity

References

- [1] Naranjo J. E. et al., “A scoping review on virtual reality-based industrial training,” *Applied Sciences*, vol. 10, no. 22, article 8224, 2020, doi:10.3390/app10228224.
- [2] Vercelli G. et al., “From risk to readiness: VR-based safety training for industrial hazards,” *arXiv preprint arXiv:2412.13725*, 2024, doi:10.48550/arXiv.2412.13725.
- [3] Chen S.-Y. and Chien W.-C., “Immersive virtual reality serious games with DL-assisted learning in high-rise fire evacuation on fire safety training and research,” *Frontiers in Psychology*, vol. 13, article 786314, May 2022, doi:10.3389/fpsyg.2022.786314.
- [4] Shahu A., Kinzer K., and Michahelles F., “Enhancing professional training with single-user virtual reality: Unveiling challenges, opportunities, and design guidelines,” in *Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia (MUM '23)*, 2023, pp. 244–256, doi:10.1145/3626705.3627791.
- [5] Fu Y. and Li Q., “A virtual reality-based serious game for fire safety behavioral skills training,” *International Journal of Human–Computer Interaction*, vol. 40, no. 19, pp. 5980–5996, 2024, doi: 10.1080/10447318.2023.2247585.
- [6] Faiz T. et al., “A scoping review on hazard recognition and prevention using augmented and virtual reality,” *Computers*, vol. 13, no. 12, article 307, 2024, doi:10.3390/computers13120307.
- [7] Kerbl B. et al., “3D Gaussian splatting for real-time radiance field rendering,” *ACM Transactions on Graphics*, vol. 42, no. 4, July 2023, doi:10.48550/arXiv.2308.04079.
- [8] Li C. et al., “RT-NeRF: Real-time on-device neural radiance fields towards immersive AR/VR rendering,” in *Proceedings of the 2022 IEEE/ACM International Conference on Computer-Aided Design (ICCAD '22)*, San Diego, CA, USA, 2022, 9 pages, doi: 10.1145/3508352.3549380.
- [9] Neff T. et al., “DOnERF: Towards real-time rendering of compact neural radiance fields using depth oracle networks,” *Computer Graphics Forum*, vol. 40, no. 4, pp. 45–59, 2021, doi:10.1111/cgf.14340.
- [10] Samudrala S., Kondguli S., and Gratz P., “Benchmarking 3D Gaussian splatting rendering,” in *Proceedings of the 2025 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*, Ghent, Belgium, 2025, pp. 227–238, doi:10.1109/ISPASS64960.2025.00029.
- [11] Niemeyer M. et al., “RadSplat: Radiance field-informed Gaussian splatting for robust real-time rendering with 900+ FPS,” *arXiv preprint arXiv:2403.13806*, 2024, doi:10.48550/arXiv.2403.13806.
- [12] Guédon A. and Lepetit V., “SuGaR: Surface-aligned Gaussian splatting for efficient 3D mesh reconstruction and high-quality mesh rendering,” *arXiv preprint arXiv:2311.12775*, 2023, doi: 10.48550/arXiv.2311.12775.
- [13] Y. Jiang et al., “VR-GS: A physical dynamics-aware interactive Gaussian splatting system in VR,” *Proc. ACM Comput. Graph. Interact. Tech.*, vol. 7, no. 1, art. no. 12, May 2024, doi:10.1145/3649930.
- [14] Xie T. et al., “PhysGaussian: Physics-integrated 3D Gaussians for generative dynamics,” *arXiv preprint arXiv:2311.12198*, 2023, doi:10.48550/arXiv.2311.12198.
- [15] Ye T., Huang H., Liu Y., and Yang J., “Global structure-from-motion enhanced neural radiance fields 3D reconstruction,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-4/W10-2024, pp. 199–204, 2024, doi:10.5194/isprs-archives-XLVIII-4-W10-2024-199-2024.
- [16] Ahmad J. et al., “From fields to splats: A cross-domain survey of real-time neural scene representations,” *arXiv preprint arXiv:2509.23555*, 2025, doi: 10.48550/arXiv.2509.23555.
- [17] Liu T. et al., “MVSGaussian: Fast generalizable Gaussian splatting reconstruction from multi-view stereo,” *arXiv preprint arXiv:2405.12218*, 2024, doi:10.48550/arXiv.2405.12218.
- [18] Montas-Laruchente N. et al., “Automatic 3D reconstruction: Mesh extraction based on Gaussian splatting from Romanesque–Mudejar churches,” *Applied Sciences*, vol. 15, no. 15, article 8379, 2025, doi: 10.3390/app15158379.
- [19] Geng X. et al., “YOLOFM: An improved fire and smoke object detection algorithm based on YOLOv5n,” *Scientific Reports*, vol. 14, article 4543, 2024, doi: 10.1038/s41598-024-55232-0.
- [20] Alkhamash E. H., “A comparative analysis of YOLOv9, YOLOv10, YOLOv11 for smoke and fire detection,” *Fire*, vol. 8, no. 1, article 26, 2025, doi: 10.3390/fire8010026.
- [21] Chetoui M. and Akhloufi M. A., “Fire and smoke detection using fine-tuned YOLOv8 and YOLOv7 deep models,” *Fire*, vol. 7, no. 4, article 135, 2024, doi: 10.3390/fire7040135.
- [22] J. S. Almeida, C. Huang, F. G. Nogueira, S. Bhatia, and V. H. C. de Albuquerque, “EdgeFireSmoke: A novel lightweight CNN model for real-time video fire–smoke detection,” *IEEE Trans. Ind. Informatics*, vol. 18, no. 11, pp. 7889–7898, Nov. 2022, doi: 10.1109/TH.2021.3138752.
- [23] Thai H.-T., Tran-Van N.-Y., Le-Minh K.-H., and Le K.-H., “An edge-based fire detection system for real-time IoT applications,” in *Proceedings of the 2023 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, Malang, Indonesia, 2023, pp. 646–651, doi: 10.1109/COMNETSAT59769.2023.10420588.
- [24] Nadeem M. et al., “Visual intelligence in smart cities: A lightweight deep learning model for fire detection in an IoT environment,” *Smart Cities*, vol. 6, no. 5, pp. 2245–2259, 2023, doi: 10.3390/smartcities6050103.
- [25] Mao H. et al., “LIVE-GS: LLM powers interactive VR by enhancing Gaussian splatting,” *arXiv preprint arXiv:2412.09176*, 2024, doi: 10.48550/arXiv.2412.09176.
- [26] N. Khan, K. Muhammad, T. Hussain, M. Nasir, M. Munsif, A. S. Imran, and M. Sajjad, “An adaptive game-based learning strategy for children road safety education and practice in virtual space,” *Sensors*, vol. 21, no. 11, article 3661, 2021, doi: 10.3390/s21113661.

- [27] H. W. Chung et al., "Scaling instruction-finetuned language models," arXiv preprint arXiv:2210.11416, 2022.
- [28] Devlin J., Chang M.-W., Lee K., and Toutanova K., "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.
- [29] Barron J. T., Mildenhall B., Tancik M., Hedman P., Martin-Brualla R., and Srinivasan P. P., "Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields," arXiv preprint arXiv:2103.13415, 2021, doi: 10.48550/arXiv.2103.13415.
- [30] Müller T., Evans A., Schied C., and Keller A., "Instant neural graphics primitives with a multiresolution hash encoding," ACM Transactions on Graphics, vol. 41, no. 4, article 102, 15 pages, July 2022, doi: 10.1145/3528223.3530127.

Authors' Profiles



Annanya Gali is currently pursuing a Bachelor of Technology (B.Tech.) degree in Information Technology from M.K.S.S.S. Cummins College of Engineering, Pune, India. She completed her Summer Internship at Zscaler as a Software Developer, where she got hands-on experience in Industry-level software systems and development practices. Her research interests include Artificial Intelligence, Machine Learning, and Virtual Reality.



Sneha Thombre is an Associate Professor at M.K.S.S.S. Cummins College of Engineering for Women, Pune, India, with over 30 years of academic experience. She has been actively involved in teaching, curriculum development, and mentoring in the field of information technology. Her areas of interest include Cyber Security, Augmented and Virtual Reality (AR/VR), Green Computing, and Artificial Intelligence.

How to cite this paper

Annanya Gali, Sneha Thombre, "A Hybrid 3D Gaussian Splatting and Photogrammetry Framework for Industrial Virtual Reality-Based Fire Safety Training", International Journal of Information Technology and Computer Science(IJITCS), Vol.18, No.4, pp.128-141, 2026. DOI:10.5815/ijitcs.2026.04.08