Article | Received 4 February 2025; Accepted 2 July 2025; Published 22 July 2025
https://doi.org/10.55092/sc20250019

Revolutionising construction site simulations with automated 3D segmentation and mesh construction

Ping Chai[1,2], Lei Hou[1,2,*], Xianwan Lo[3], Guomin (Kevin) Zhang[1,2], Haosen Chen[1,2] and Yang Zou[4]

[1] School of Engineering, RMIT University, Melbourne, Victoria 3000, Australia
[2] Centre for Future Construction, RMIT University, Melbourne, Victoria 3000, Australia
[3] Faculty of Engineering and IT, The University of Melbourne, Parkville, Victoria 3010, Australia
[4] Department of Civil and Environmental Engineering, University of Auckland, Auckland 1010, New Zealand

* Correspondence author; E-mail: lei.hou@rmit.edu.au.

Highlights:

- The 3-Dimensional Reconstruction, Integration, Segmentation, and Editing (3D-RISE) model enables dynamic construction site simulations, integrating 3D reconstruction, segmentation, and mesh refinement for site layout planning (SLP) and safety training applications.
- Implements 3D Gaussian Splatting (3DGS) for real-time, computationally efficient modelling, transforming captured scenes into realistic and editable virtual environments.
- Combines advanced tools like "Segment Any 3D Gaussians (SAGA)" for precise automated segmentation and "Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering (SuGaR)" for detailed mesh reconstruction, enabling flexible customisation of 3D scene components.
- Results demonstrate high-quality outputs with realistic visualisation and seamless integration of segmented objects, full models, and online 3D elements into dynamic virtual environments.

Abstract: The construction industry requires dynamic and realistic site simulations for effective site layout planning (SLP) and safety training. This paper introduces 3-Dimensional Reconstruction, Integration, Segmentation, and Editing (3D-RISE), a novel workflow integrating 3D Gaussian Splatting (3DGS), Segment Any 3D Gaussians (SAGA), and Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering (SuGaR) to create customisable and high-quality 3D construction scenes. The workflow leverages advanced segmentation and mesh reconstruction techniques to generate editable models while maintaining scene realism. Evaluation metrics from the initial 3DGS outputs across three image datasets (namely, steel structure, excavator, and random model) demonstrate the effectiveness of the pipeline, achieving peak signal-to-noise ratio (PSNR) values of 22.984, 35.254, and 25.854, respectively, after 30K iterations. The structural similarity index measure (SSIM) scores ranged from 0.783 to 0.951, highlighting the workflow's ability to generate visually accurate outputs.

Despite its robust capabilities, 3D-RISE has limitations, including reliance on datasets with 360-degree coverage, the inability to directly modify 3DGS-rendered scenes in Unreal Engine (UE) version 5.3, and a high demand for computational power. Running 3D-RISE on GPUs with lower specifications significantly increases processing time, requiring optimisation through downsampling or lower iteration counts, which may affect output quality. Future work focuses on integrating generative artificial intelligence (AI) to generate 3D-ready models from single images, reducing dataset requirements and computational overhead while improving scalability.

Keywords: SLP; safety; 3D-RISE; 3D reconstruction; segmentation; mesh reconstruction; 3DGS; SAGA; SuGaR

1. Introduction

Construction sites are inherently dynamic environments, with each site and project location possessing unique properties that are independent of others [1]. Successful construction site management can lead to smooth project progress, improved site safety, enhanced visualisation and communication for both stakeholders and workers, effective resources and cost optimisation [2]. Historically, construction site management has often required on-site presence for manual safety equipment inspections, overseeing construction site layout planning (SLP), and managing logistics. Such traditional management practices have proven time-consuming, increased labour costs, and ineffective construction site data collection [3,4]. This inefficiency underscores the importance of having readily accessible construction site data without the need for physical site visits [5]. Having readily accessible construction site information can enhance communication, improve safety training, and facilitate better decision-making [6,7].

However, the availability of comprehensive construction site datasets is scarce due to the proprietary nature and liability concerns associated with each company's projects [8]. Numerous researchers have attempted to bridge the gap of limited construction datasets by conducting data augmentation on construction datasets. Most existing research on construction site datasets tends to focus on specific on-site construction resources, such as workers and construction objects [9,10], rather than capturing the entire construction site scene (Figure 1). A holistic view of the construction site enhances understanding of the spatial layout, interrelationships, and dynamic interactions between all on-site elements [11]. For example, Jiang *et al.* [12] optimised crane placement and hoisting operations through a holistic planning approach. Kim *et al.* [13] modelled a 2-Dimensional (2D) construction site plan in Unity [14] to enable a deep reinforcement learning model for realistic worker pathfinding and behavior simulation. Zhu *et al.* [15] developed a digital replica of a construction site layout, integrating essential components and locations to aid site logistics, safety training, and planning coordination in virtual environments.

Recent advancements in 3-Dimensional (3D) reconstruction techniques, such as computer vision, 3D laser scanning, simultaneous localisation and mapping, and rapid image modelling, have made it easier to acquire detailed on-site data. Unlike traditional surveying methods such as total stations and GPS receivers, which often require extensive manual effort, time and experience, these modern techniques allow for the accurate replication of complex real-world environments with high fidelity. However, the drawback of ensuring a realistic 3D construction scene is the challenge when it comes to editing or manipulating within the captured scene, without disrupting its realism. Bang *et al.* [16]

proposed an image inpainting method that uses a cut-and-paste technique to expand construction datasets. While this approach successfully created different construction scenarios by inserting various elements into existing scenes, it significantly diminished the immersion and realism of the modified scenes. This limitation poses challenges in construction SLP, where dynamic adjustments are frequently necessary.

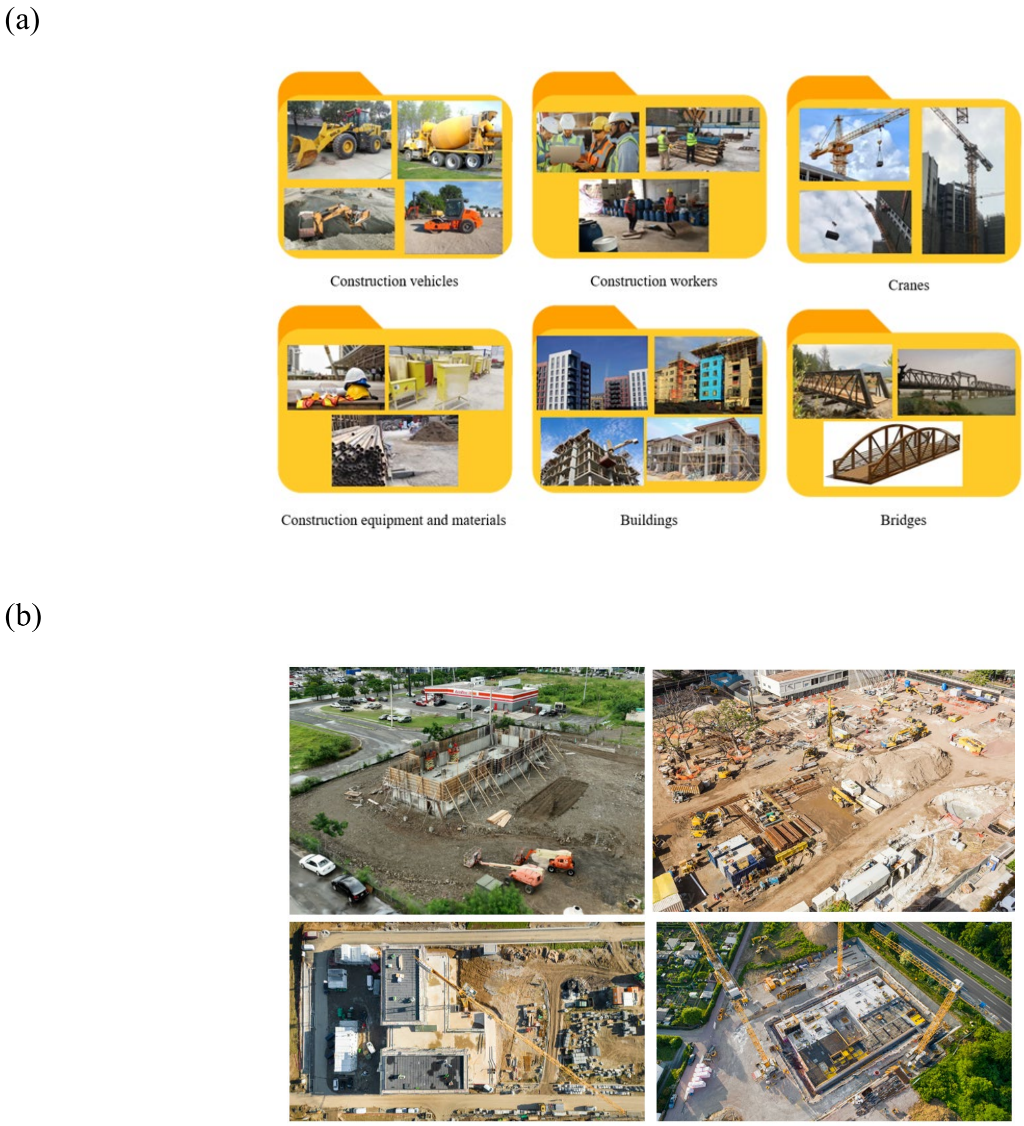

Figure 1. Two categories of construction datasets: **(a)** Micro-level construction datasets: datasets that focus on specific on-site construction resources and elements; **(b)** Macro-level construction datasets: datasets that capture the entire construction scene.

The ability to edit and experiment with different configurations within an existing construction scene is crucial for dynamic SLP [17], which is essential at various stages of the construction process. Dynamic SLP accounts for factors such as the arrival times of materials and equipment, as well as their specific locations during different phases of the project. The capability to adjust and modify elements within a construction scene, while preserving realism, is critical for enhancing both project execution and stakeholder communication. While realistic visualisation is particularly valuable for training simulations, enabling workers to engage with lifelike scenarios, its role extends far beyond safety training. High-quality visualisation facilitates easy understanding and interpretation, allowing stakeholders to comprehend spatial layouts, equipment placement, and construction progress more effectively [18]. This improved clarity reduces misunderstandings and aligns project peers with critical decisions, ensuring smoother collaboration. Moreover, immersive visualisation provides users with an in-depth experience through various interfaces, screens, and devices. By offering multi-sensory feedback, these tools allow users to perceive experiences that closely resemble reality. Feedbacks delivered through sensory channels enhances the sense of control and realism, which in turn improves users' efficiency in interpreting data and making informed decisions in construction planning [19].

This paper proposes 3-Dimensional Reconstruction, Integration, Segmentation, and Editing (3D-RISE), a workflow to enable the manipulation of construction scenes. The proposed workflow demonstrates the potential to create diverse variations in site configurations, supporting dynamic planning and realistic training simulations. By allowing users to experiment with site layouts and optimise safety and logistics in a flexible virtual environment, the integration of these advanced techniques aims to revolutionise construction site management and training.

The technical tools used in this study are individual components that exist independently. The novelty of this research lies in the integration of these tools into a unified and practical workflow focusing on the construction site layout synthesis domain. Our research on the topic of customisable construction site layout synthesis represents one of the earlier attempts to combine these components cohesively to enable editable, photorealistic 3D reconstruction scenes that support dynamic SLP and virtual safety training applications. While most existing works utilise 3D reconstruction for static visualisation, our approach goes further by demonstrating how segmentation and mesh refinement for large-scale SLP can be incorporated into a pipeline that enables object-level editing and scene compositing. This capability significantly enhances customisability, reusability, and realism in training simulations and planning environments.

The manuscript is structured as follows: Section 2 reviews the related works, providing context and background for the study. Section 3 outlines the methodology used to evaluate the proposed 3D-RISE workflow. Section 4 discusses the results, including insights into the workflow's performance, limitations, and applications, and highlights potential future work to address current challenges. Finally, Section 5 concludes the study by summarising key findings and their implications.

Table 1 provides the abbreviations and corresponding definitions used in this study.

Table 1. Abbreviations and definitions.

Abbreviations	Definitions
3D-RISE	3-Dimensional Reconstruction, Integration, Segmentation, and Editing
3D	3-Dimensional
2D	2-Dimensional
3DGS	3D Gaussian Splatting
AI	Artificial intelligence
FPS	Frames per second
NeRF	Neural Radiance Field
OpenMVS	Open Multi-View Stereo
SAGA	Segment Any 3D Gaussians
SLP	Site layout planning
SuGaR	Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering
UE	Unreal Engine

2. Related

2.1. Advancements of 3D reconstruction techniques in construction field

The field of 3D reconstruction in the construction industry has seen significant advancements over the decades, moving from manual, labour-intensive methods to more automated, data-driven solutions. As illustrated in Figure 2, these techniques can be broadly categorised into traditional, geometry-based, and deep learning-based methods. In the past, there was no comprehensive mathematical-based optimisation approach that could help in capturing the site into a 3D scene [17]. Traditional methods, such as manual hands-on approach and the use of early-designed total stations, dominated early construction site planning. While these approaches provided essential data for site layout, they were often time-consuming, prone to human error, and lacked the flexibility required for modern, dynamic construction projects [20].

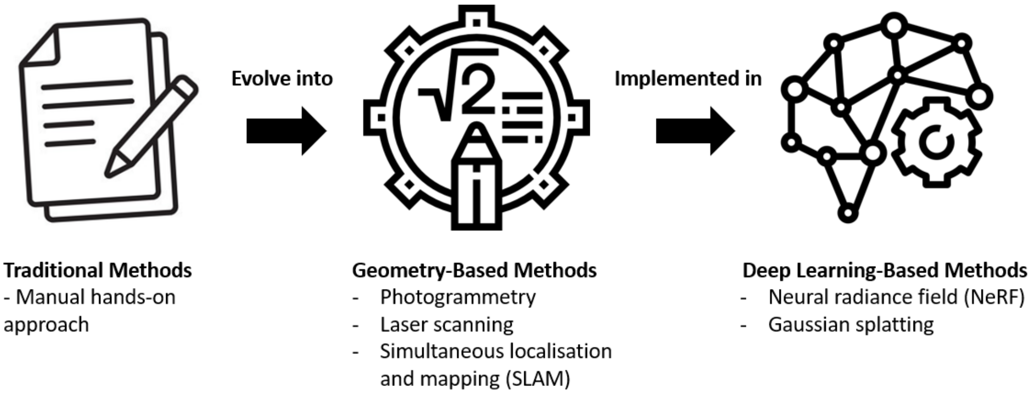

Figure 2. Timeline of the progression of various 3D reconstruction techniques.

Currently, there are two techniques used for image-based 3D scene reconstruction in the construction field, namely geometry-based methods and deep learning-based methods [21]. Geometry-based methods are computational techniques that rely on mathematical principles of geometry to infer and reconstruct 3D shapes and structures from 2D images or spatial data. Photogrammetry, laser scanning, structured light, Time of Flight and simultaneous localisation and mapping are popular applications used for 3D map surveying [22], autonomous driving [23], and smart cities [24]. In the construction domain, structural health monitoring is one of the applications in which geometry-based 3D reconstruction is widely used. Researchers have utilised the photogrammetry technique by stitching drone images to generate a 3D model for structural inspections. Chen *et al.* [25] compared different geometry-based 3D reconstruction techniques to evaluate the efficiency of performing bridge inspection. Aside from the structural health monitoring sector, geometry-based 3D reconstruction is used in conducting the 3D mapping process. Hu *et al.* [26] integrated robot motion control and pathfinding with simultaneous localisation and mapping to reconstruct a 3D point cloud to overcome laborious and time-consuming manual work. Laser scans are also used to generate accurate spatial representations of buildings quickly, which can help support disaster management in a multi-layered complex buildings [27]. Despite extensive applications, geometry-based methods faced challenges with having to handle complex geometries or dynamic lighting conditions, which can affect the overall performance [21]. Su *et al.* [28] commented that laser scanning is not only costly and inconvenient, but the process of image-based 3D reconstruction performances is easily influenced by environmental conditions. Cai *et al.* [29] commented on the limitations of these 3D reconstruction methods as they may encounter errors in image orientation to texture mapping, which can accumulate leading to inaccurate results. Additionally, while geometry-based methods are effective in reconstructing simple objects and scenes, they struggle when trying to model meshes of complex geometries and details [30].

Recently, deep learning-based methods have emerged as a powerful alternative for 3D reconstruction by leveraging artificial neural networks to infer 3D structures from 2D images. Most deep learning-based approaches have evolved into hybrid models by incorporating geometry-based techniques, often relying on geometry-based preprocessing techniques such as photogrammetry to generate camera poses for dataset preparation. As a result, deep learning-based methods are not entirely independent, but rather build upon the foundational outputs of geometry-based techniques. In the 3D reconstruction process, unlike geometry-based methods that explicitly calculate depth and 3D geometry, deep learning models learn the underlying patterns in the data through large datasets and trainable parameters. Su *et al.* [28] combined the geometry-based laser scanning 3D reconstruction method with deep learning model to optimise low-light images at the pixel level and to achieve better 3D reconstruction performance. As construction site scene tends to be more complex, deep learning-based methods are able to capture complex scene geometries with high-fidelity [31]. Similarly, these more advanced 3D reconstruction techniques serve similar purposes in the construction domain, such as structural health monitoring and 3D mapping [32], offering improved performance and overcoming the limitations of geometric-based methods, particularly in handling complex geometric constraints.

While the intended problem-solving solutions are achieved, most research work has not ventured beyond this scope. Existing studies primarily utilise 3D reconstruction techniques for enhanced visualisation to aid respective fields, but these visualisations tend to be static, offering limited variability and flexibility to address broader limitations. Jiang *et al.* [12] applied 3D reconstruction techniques to aid in hoisting facility layout planning, improving efficiency by predicting optimal crane placement for better project progress. The 3D reconstruction process is primarily used to extract the site's dimensions and the relative positions of key elements, allowing easier measurements to determine the optimal crane position on the site. However, the inability to alter components within the captured scene presents an opportunity for future work to explore dynamic scene manipulation for more flexible planning. Similarly, Lian *et al.* [33] implemented 3D reconstruction techniques in virtual fire scene simulations for firefighting training, enhancing user experience. Yet, this approach is limited by the lack of diverse scenarios, reducing its adaptability for more comprehensive training. In both cases, scene manipulation could overcome these limitations. A 3D reconstruction technique followed by dynamic scene manipulation allows for greater variation in both construction SLP and safety training. This approach offers valuable datasets for model training, improves layout planning, and provides diverse scenarios for more effective worker safety training, enhancing the overall flexibility and usefulness of 3D reconstruction in these domains.

2.2. 3D modelling for real-time scene

3D Gaussian Splatting (3DGS) is one of the deep learning-based methods for image-based 3D scene reconstruction. Kerbl *et al.* [34] designed 3DGS to address the limitations of Neural Radiance Fields (NeRF) [31], another deep learning-based 3D reconstruction method. While NeRF is highly effective in generating photorealistic 3D scenes from sparse image inputs, it requires significant computational resources and longer rendering times, making it less suitable for real-time applications. In contrast, 3DGS provides a more computationally efficient alternative by representing 3D scenes as a collection of Gaussian functions. This approach enables faster processing and real-time rendering while maintaining high-quality visual fidelity. Additionally, the rendering output from 3DGS can be seamlessly integrated with post-processing operations such as scene editing and animations, whereas the efficient rasterisation and high rendering quality allow the support of implementation with virtual reality [35]. The application of virtual reality in construction field is vast, such as enhancing worker safety, simulation performance evaluation, and remote inspection [36]. With the integration of 3DGS, it offers the potential to create highly realistic and interactive virtual environments to enhance training simulations. The breakdown step-by-step of 3DGS is shown in Figure 3.

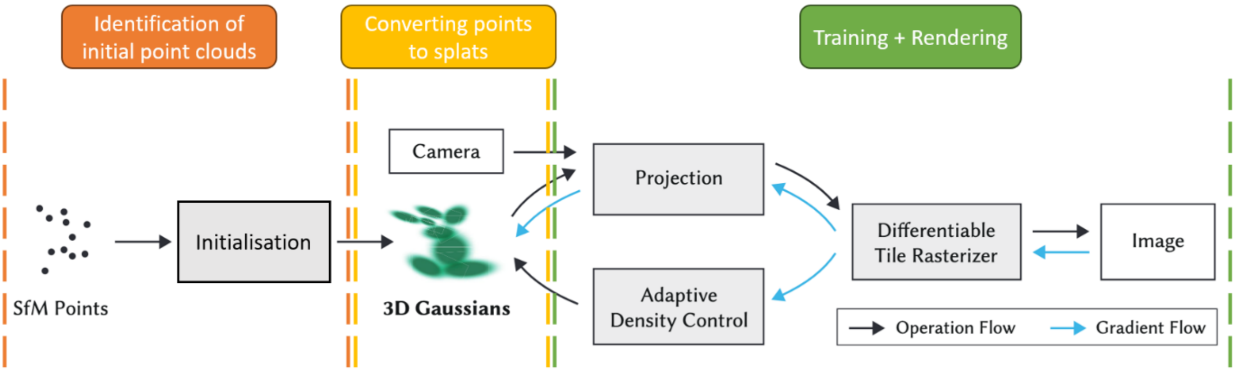

Figure 3. Optimisation process of 3DGS.

When image collections are completed and ready to be processed using 3DGS, the initial point clouds from images are identified using the structure-from-motion method. COLMAP [37] is a technique used in structure-from-motion method that takes a set of 2D images captured from different viewpoints and estimates the 3D structure of the scene by analysing the motion between these viewpoints. This step establishes the basic geometry of the scene by determining camera poses and sparse 3D points. Next, the sparse point clouds are then used to generate a set of 3D Gaussian ellipsoids, known as splats, where each splat represents a local region in the scene. In contrast to NeRF, this conversion significantly reduces computational power by replacing dense point clouds or complex mesh representations with a compact and continuous representation at the cost of reduced precision. Gaussian splats efficiently approximate the geometry and attributes of the scene, allowing for smooth and realistic rendering without the need for high-density polygons or intensive geometric calculations.

Lastly, the Gaussian ellipsoids are rendered into a 2D image from a new viewpoint. By using learnable 3D Gaussian ellipsoid functions to represent scenes, realistic 2D images can be synthesised to allow smooth rendering transitions between Gaussian ellipsoids to avoid visible artifacts [29]. The function can be defined as:

$$C = \sum_{i \in N} c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j) \tag{1}$$

Where N represents all the Gaussian splats along the ray, c_i represents the colour intensity of the i-th splat, α_i is the transparency or opacity of the i-th splat, while the latter half of the equation accounts for the occlusion caused by splats closer to the camera along the ray. Nearer splats contribute first, with their opacity determining how much of the farther splats remain visible. If earlier splats are fully opaque $(\alpha_j = 1)$, the i-th splat's contribution is completely blocked, whereas if earlier splats are fully transparent $(\alpha_j = 0)$, the i-th splat contributes fully. The function describes how the final rendered colour C of a pixel is computed in the 3DGS rendering process. It accounts for the contributions of multiple overlapping Gaussian splats along a ray as seen from a particular camera viewpoint, which enables 3DGS to achieve smooth and realistic rendering by calculating the contributions of overlapping splats along a viewing ray.

2.3. Dynamic scene editing through 3D segmentation and mesh reconstruction

2D image segmentation is a deep learning-based computer vision technique that enables advanced applications in scene understanding, medical image analysis, robotic perception, video surveillance, and augmented reality by partitioning images into meaningful regions or objects for precise interpretation and interaction [38]. In the construction industry, image segmentation is widely used for structural cracks and defects maintenance [39], and worker's health safety monitoring [40]. Kirillov *et al.* [41] have developed a model called Segment Anything that excels in generalising across a wide range of 2D image segmentation-related tasks, making it highly adaptable for applications mentioned beforehand.

In our research, the objective is to enable the generation and extraction of 3D objects from captured frames to facilitate dynamic SLP. While 2D segmentation excels at analysing static or still images, it faces challenges in extracting 3D geometric elements from objects viewed at multiple angles. This limitation hinders its application in tasks requiring spatial understanding across a three-dimensional scene. By contrast, 3D segmentation allows for accurate identification and isolation of objects in a volumetric space, enabling greater flexibility and precision in dynamic environments.

To overcome the static nature of 2D segmentation, many researchers have proposed extending Segment Anything's capabilities to 3D segmentation by integrating it with radiance fields, such as NeRF [42–44]. Cen *et al.* [45] presented a 3D segmentation method called Segment Any 3D GAussians (SAGA) that can segment corresponding 3D targets in the realm of 3DGS. These integrations have achieved significant success in extracting 3D geometric information, allowing for seamless 3D object segmentation across multiple viewpoints. This advancement bridges the gap between static image analysis and dynamic 3D scene understanding, paving the way for innovative applications in construction site simulations, virtual planning, and safety monitoring.

Another method for achieving dynamic SLP is through mesh reconstruction from rendered 3D scenes. A mesh-based representation provides a structured framework that enables powerful tools for editing, sculpting, animating, and relighting the scene. While 3DGS excels at producing realistic renderings of scenes, it poses challenges when it comes to extracting explicit surface geometry from the rendered data. Guédon *et al.* [46] introduced an approach that combines a regularisation term with Poisson surface reconstruction [47] to refine and bind splats to the surface of the mesh, effectively creating a structured and editable 3D representation. This process bridges the gap between visually compelling Gaussian splats and practical mesh-based models, enabling the generation of high-quality meshes that faithfully represent the underlying scene geometry.

A ready-mesh representation of the site layout provides the flexibility to modify and edit the scene such as object removal or insertion virtually as project requirements evolve over time. Project managers can adjust object placements, pathways, or structural elements directly in the virtual environment, ensuring smoother project progress and more efficient SLP. This eliminates the need for frequent physical inspections or testing, reducing costs and saving time while maintaining precision and adaptability.

By enabling real-time modifications and seamless integration of project updates, 3D segmentation and mesh reconstruction methods offer a robust solution for dynamic construction site simulations and planning. These approaches not only enhance visualisation but also support the creation of flexible and editable virtual environments, ultimately improving decision-making and operational efficiency throughout the project lifecycle.

2.4. Research gaps

Despite significant advancements in construction site simulations and 3D modelling techniques, several research gaps remain. Existing studies have primarily focused on either static visualisations or isolated aspects of site management, lacking an integrated approach that supports dynamic scene editing and diverse site configurations. Moreover, while 3D reconstruction techniques such as photogrammetry and NeRF have demonstrated success in generating realistic 3D scenes, they often fall short in enabling seamless customization and interaction with scene elements. The inability to modify or adapt reconstructed scenes limits their applicability in dynamic SLP and training simulations. Additionally, segmentation and mesh reconstruction processes remain labour-intensive and computationally demanding, with current methods often requiring extensive manual refinement to achieve clean, application-ready outputs. This is particularly evident in scenarios where background artifacts or noisy data interfere with the quality of the reconstructed scene.

3. Methodology

Our study aims to develop and implement 3D-RISE, a novel workflow designed to facilitate dynamic SLP and realistic construction safety training simulations. Unlike previous approaches, 3D-RISE is the first to successfully integrate a combination of advanced techniques: 3DGS, automated segmentation, and mesh reconstruction into a cohesive and functional pipeline. This unique combination enables users to edit and experiment with different configurations within a self-selected construction scene in a game environment engine, addressing key challenges in dynamic SLP. The proposed workflow provides a new approach to conducting essential tasks such as efficient project execution, optimised safety protocols, and logistical planning. By leveraging these advanced techniques, 3D-RISE not only preserves scene realism but also allows seamless modification of scene elements in a virtual environment. This adaptability supports various stages of the construction process, including planning, training, and safety analysis. The workflow's successful implementation demonstrates its potential to address limitations in existing methods, such as the inability to dynamically edit 3D-reconstructed scenes or combine segmented components with full site models without disrupting the realism. Figure 4 illustrates the 3D-RISE workflow overview.

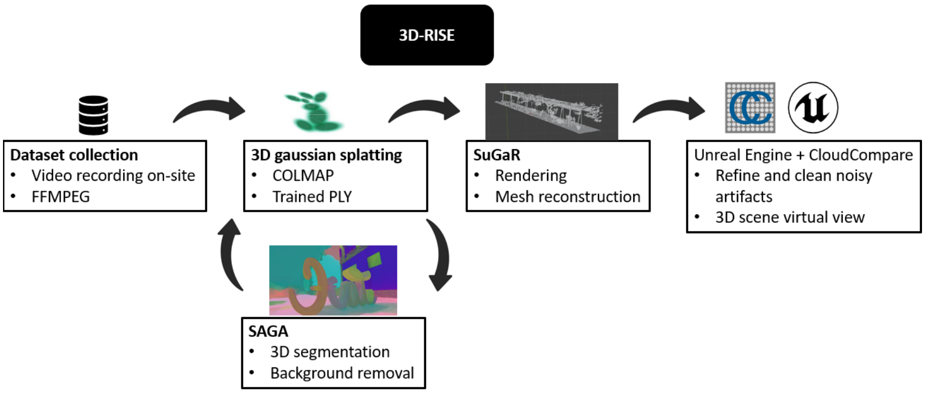

Figure 4. Overview of the 3D-RISE workflow.

3.1. Dataset collection

The first step in implementing the 3D-RISE workflow involves capturing real-world construction scenes or objects or interest to be used in the 3D scene. The capturing process is conducted in clear daylight with stable lighting conditions to minimise visual noise and ensure consistency in image quality across the dataset. Videos are recorded using either a standard smartphone camera or drone, ensuring that the target scene is captured from multiple angles. In our implementation, we used an iPhone 12 Pro Max and a DJI Phantom 2 drone for dataset collection. To reduce motion blur and ensure sufficient frame overlap during capture, we maintained a walking speed of approximately 0.09–0.12 m/s. Higher rotation or walking speeds during data collection (0.164 m/s and above) can reduce capture time but tend to increase motion blur and reduce frame overlap, thereby negatively affecting feature matching in structure-from-motion pipelines. This speed was chosen based on prior photogrammetry studies [48,49], and helped maximise reconstruction accuracy while keeping the footage stable. The capture duration parameters vary depending on the complexity and physical size of the object, as larger or more detailed objects require longer capture times to ensure sufficient coverage from multiple viewpoints. These videos are then processed using FFMPEG version 7.1 [50], which extracts individual frames per second (FPS) for use in subsequent reconstruction steps.

To evaluate the effectiveness of the proposed 3D-RISE workflow, three datasets were prepared, each representing a unique object and environment: an excavator, a steel structure, and a random model (Figure 5). These datasets were selected to assess the versatility and applicability of 3D-RISE across varied object types, detailed complexity, and surrounding environmental conditions. The capturing method varied across locations depending on the rotation speed and recording duration. The total capture duration was 52 seconds for the excavator, 21 seconds for the steel structure, and 109 seconds for the random model. Frames were extracted at 2FPS for both the excavator and random model, and 3 FPS for the steel structure, resulting in 103, 65, and 219 images respectively.

We selected a frame range between 50 and 300 images pre dataset to ensure a practical balance between rendering quality and computational efficiency during 3DGS. Using too few images may compromise reconstruction detail and spatial accuracy, while using too many significantly increases training time without proportional gains in output quality [51]. This range allowed us to systematically evaluate how dataset size influences the overall performance and fidelity of the 3D-RISE pipeline. The results of this evaluation are discussed in Section 4.

An excavator on construction site A steel structure on gravel surface A random model on the street

Figure 5. Locations chosen for the 3D-RISE workflow.

3.2. 3D Gaussian Splatting

The extracted frames from each dataset are processed using 3DGS to generate a preliminary 3D scene. As a preprocessing step, COLMAP is first used to determine the spatial location of each frame by reconstructing camera poses and generating a sparse point cloud. This information provides the necessary spatial structure for subsequent 3DGS training.

3DGS then takes the COLMAP output to produce a high-fidelity, photorealistic 3D representation of the scene, saved in the .ply format. This serves as the foundational representation used for later processing, including segmentation and mesh reconstruction. The training process is configured to run for 30K iterations, which typically requires between 1 to 2 hours per dataset depending on the number of input frames.

Due to the high computational demand of 3DGS, segmentation, and mesh reconstruction, all experiments are conducted on a high-performance workstation equipped with the following specifications:

- CPU: Cascade Lake P-8275CL
- RAM: 1152 GiB
- GPU: NVIDIA A100 Tensor Core GPU
- Storage: 500 GB SSD

The trained .ply output from 3DGS acts as the branching point in the 3D-RISE workflow, where it is either passed into a segmentation process, or directly into mesh reconstruction for full-scene modelling. The detailed refinement pipeline is illustrated in Figure 6.

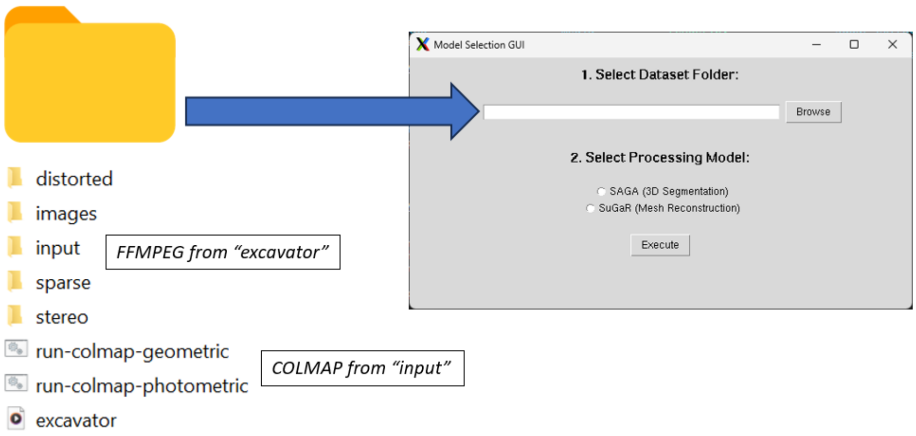

Figure 6. Data structure for 3D-RISE workflow testing.

3.3. Segmentation and mesh reconstruction

To refine the 3D scene generated by 3DGS, the 3D-RISE workflow branches into two pathways depending on the desired output:

(1) SAGA + SuGaR: Datasets are first segmented to extract target objects, which are then reconstructed into detailed 3D meshes.

(2) SuGaR only: Full-scene .ply files are directly reconstructed into mesh models without segmentation.

Both outputs are used to simulate different configurations of a dynamic construction site. Specifically, each dataset alternates between being the main scene (unsegmented) and a modular component (segmented) inserted into another scene. For instance, the unsegmented excavator dataset serves as the base environment, while the segmented steel structure is placed within it. This process is rotated across all datasets, allowing us to evaluate the usability and spatial harmony of integrating modular objects into reconstructed scenes. Figure 7 shows the breakdown of these two methods:

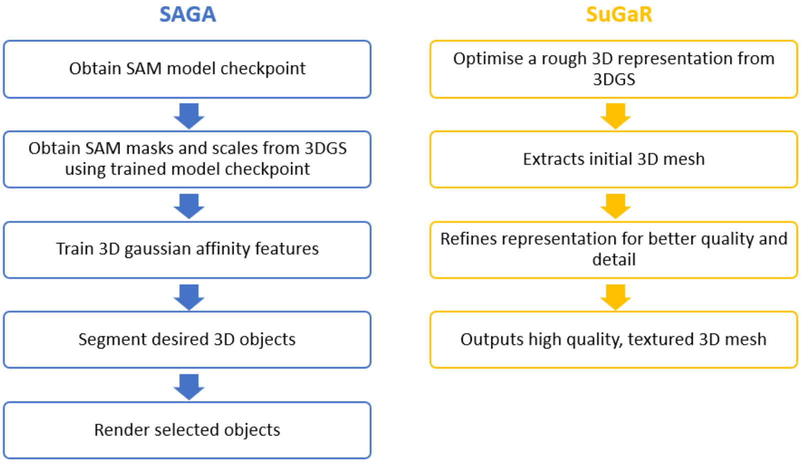

Figure 7. Workflow process for the segmentation and mesh reconstruction method.

3.3.1. SuGaR

The .ply file produced by 3DGS is used as input to SuGaR to generate a refined, high-quality meshed 3D model. The process begins by optimising the initial 3D representation to extract a coarse mesh. This mesh is then refined to improve surface detail and realism through iterative optimisation. SuGaR employs three regularisation techniques:

- dn_consistency: Maintains smoothness by ensuring consistency between neighbouring Gaussians.
- density: Balances point distribution to prevent sparsity or overcrowding.
- Sdf (Signed Distance Function): Refines surface accuracy by minimising geometric error.

For our implementation, we selected the “density” regularisation technique, which yielded smoother and more balanced meshes. A high-poly mesh setting was used, generating models with approximately 1 million vertices and one Gaussian per triangle, which is ideal for high-fidelity simulations. The refinement process ran for 15 K iterations and took approximately 2–4 hours per dataset. This approach provides a detailed and realistic model suitable for dynamic construction site applications.

However, applying SuGaR directly to multiple .ply files from different datasets can lead to complications. Each full-scene .ply includes its own background and combining them may result in overlapping elements and cluttered layouts. Instead of enhancing realism, this can introduce disorder and reduce usability (Figure 8).

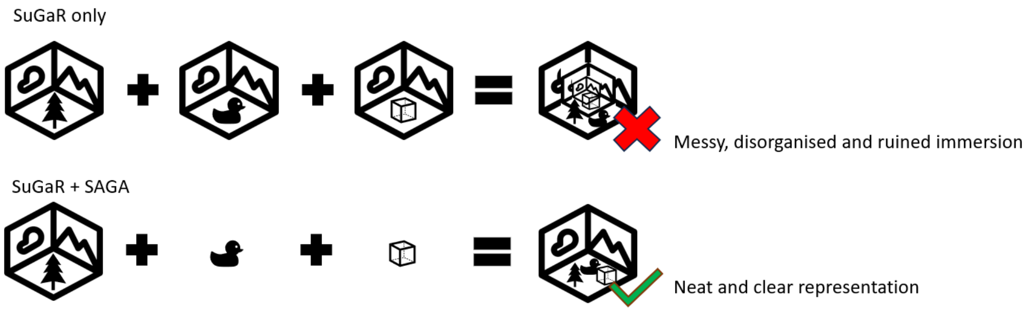

Figure 8. 3D scene outcomes with or without the use of SAGA.

3.3.2. SAGA

To prevent this, we apply SAGA to isolate specific objects and remove background elements before mesh reconstruction. This segmentation process begins by extracting mask scales from the 3DGS-trained scene and image dataset. SAGA uses this information to learn 3D Gaussian affinity features, enabling precise segmentation. Through an interactive GUI, users can select the object of interest and apply a score threshold (typically between 0.75 to 0.90) to control the intensity of the segmentation. Higher thresholds reduce false positives but may increase the risk of over-segmentation (Figure 9).

Once segmentation is complete, the system saves the output as a binary mask or bitmap of the selected Gaussians. These masks are then used to update the original image dataset, creating a new set of frames that contain only the segmented object. This updated dataset is reprocessed in 3DGS to generate a new .ply file focused solely on the isolated component. The result is a cleaner, more modular asset that integrates seamlessly into larger 3D environments, enhancing scene realism and organisation. The segmentation results are analysed further in Section 4.

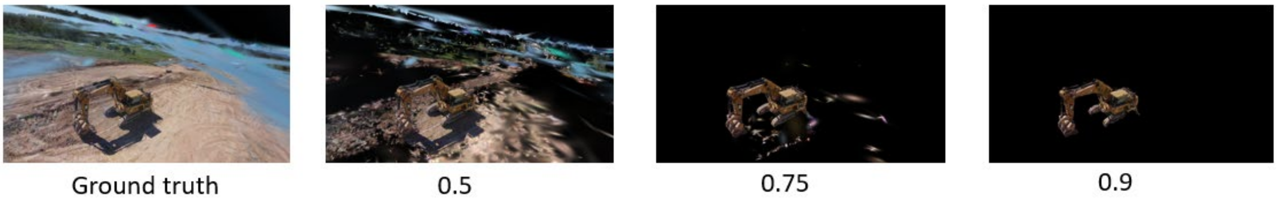

Figure 9. Score threshold (0–1) for the segmentation process using the excavator dataset.

3.4. Refinement and final assembly

The final stage of the 3D-RISE workflow involves refining the generated meshes and assembling them into a dynamic and editable virtual scene. After completing the mesh reconstruction and segmentation processes, the outputs are first cleaned using CloudCompare [52], an open-source 3D point cloud and mesh processing software optimised for handling large-scale datasets. CloudCompare is one of the several widely used software tools for 3D point cloud processing, alongside alternatives such as MeshLab, RealityCapture, Autodesk ReCap, and Trimble RealWorks. It is particularly popular among researchers due to its open-source nature, user-friendly visual interface, and robust segmentation and editing capabilities [53]. Our study involves 3D reconstruction from point clouds as well as visualisation within a game engine environment. CloudCompare was chosen for its reliable support of a wide variety of file formats [54], enabling a smooth transition from point cloud processing to real-time rendering platforms that support those formats. Additionally, this workflow has also been adopted previously where CloudCompare was used to convert processed 3D data into formats compatible with game engines such as Unity [55].

This refinement step focuses on removing mesh noise, cropping unnecessary finer elements, and improving the overall geometric quality of the models. This process ensures that the final 3D assets are clean, accurate, and free from artifacts, making them suitable for integration into realistic construction layouts. As shown in Figure 10, CloudCompare significantly enhances the quality and organisation of the scene. Two excavator examples are illustrated: one using the unsegmented dataset; another incorporating segmentation. The comparison highlights the improvement in cleanliness and usability of the mesh after refinement.

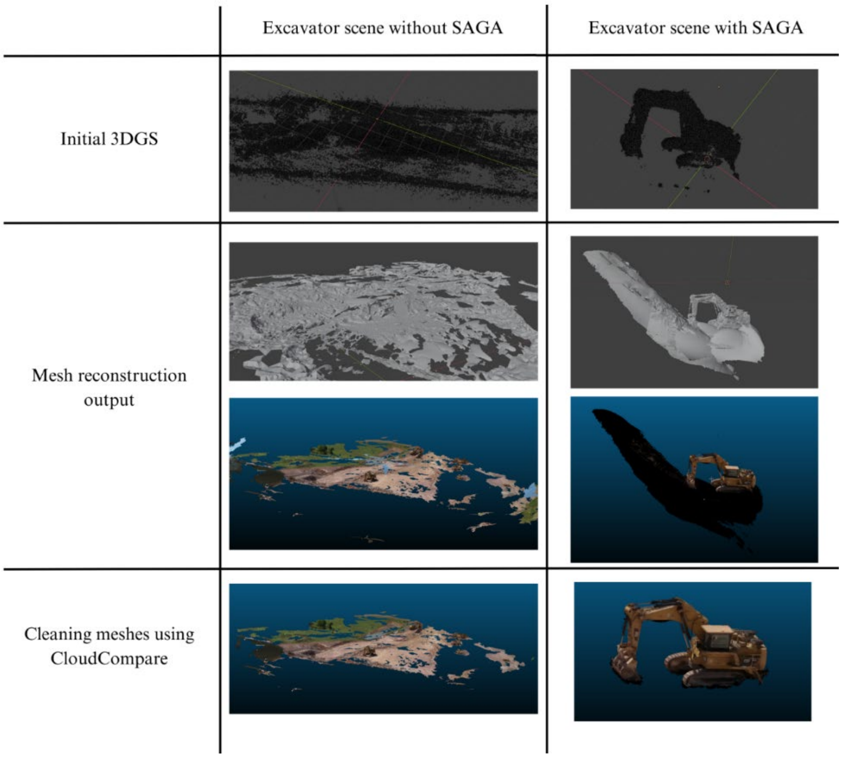

Figure 10. Evolution of mesh outputs for excavator scene with and without segmentation.

Once the models are cleaned, they are imported into UE version 5.3 [56], a widely used real-time 3D creation tool for developing immersive virtual environments. UE is used in this study to construct a dynamic SLP environment where users can interactively test scene configurations. Figure 11 illustrates the full workflow within UE.

To support seamless integration of 3DGS-rendered models, we utilise the XVERSE 3DGS UE Plugin (version 1.15-ue5.3) [57]. This plugin enables direct compatibility with .ply files generated by 3DGS and allows them to be visualised and managed within UE. Prior to importing the models, an initial level setup is performed to ensure visual realism. This includes configuring scene elements such as a landscape, point and directional lights, exponential height fog, sky atmosphere, sky light, and volumetric clouds (Figure 12).

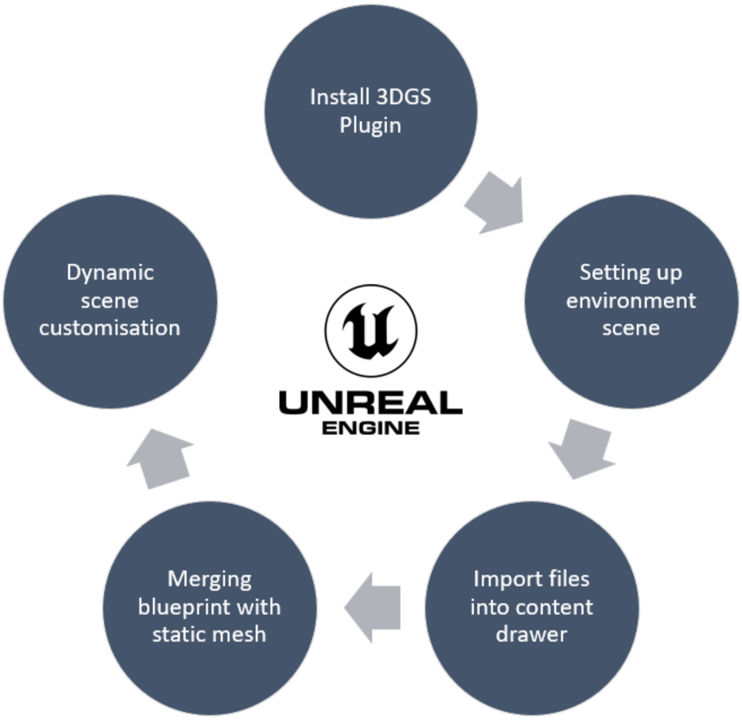

Figure 11. Workflow in UE 5.3.

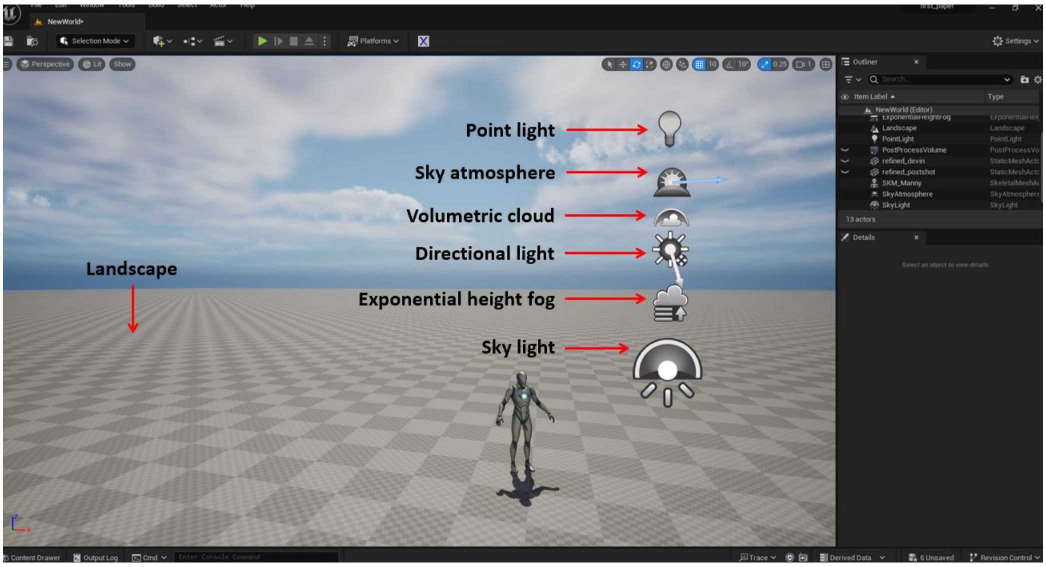

Figure 12. Initial UE setup.

Model importation follows two paths: the 3DGS .ply files are imported via the XVERSE plugin and registered as blueprints, while the .obj files generated through SuGaR are brought in directly via standard drag-and-drop as static meshes. Since blueprint imports do not retain material, texture, or collision properties, they are manually merged with their corresponding static meshes to achieve proper visual and interactive functionality (Figure 13).

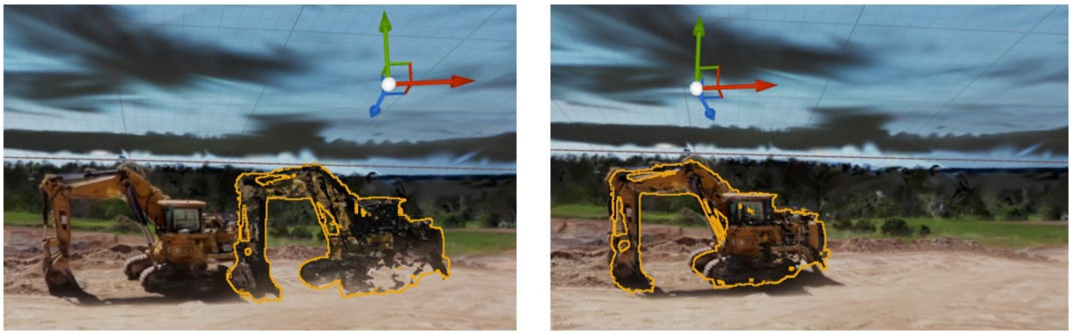

Figure 13. Aligning static mesh into blueprint.

Once blueprints and meshes are paired, the virtual environment becomes fully editable. Objects can be repositioned, lighting can be adjusted, and additional models from 3D-RISE can be added to the scene. Figure 14 presents the final scene setup, which uses the excavator environment as the main context, into which the segmented steel structure and random model meshes are inserted. This scene also includes online assets to enrich the environment and demonstrate extensibility. The resulting environment not only achieves high visual fidelity but also supports full interactivity and real-time modification, making it ideal for dynamic SLP. Users can easily edit, rearrange, or augment the virtual construction site to simulate various scenarios. This adaptability is central to 3D-RISE's contribution, enabling practitioners to explore configuration alternatives and optimise construction planning workflows in an efficient, immersive way.

Figure 14. Final representation of excavator scene in UE.

4. Discussion

4.1. Initial 3DGS testing results

The evaluation of the 3D-RISE workflow highlights the interplay between hardware capabilities, training iterations, and dataset size in determining the performance and quality of the reconstructed 3D scenes. As we are utilising NVIDIA A100 Tensor Core GPU when running our workflow, the running time was significantly reduced when testing the 3D-RISE workflow. The high computational power of the A100 facilitated faster processing across all stages, including 3DGS, SAGA, and SuGaR. We have conducted similar testing on NVIDIA GeForce RTX 3090 GPU resulting in an additional 1 to 2 hours per dataset depending on the number of frames in the dataset. To optimise performance on a lower-end GPU, downsampling the input images may be necessary to reduce computational load, or reducing the iteration count to speed up the processing time. Table 2 presents the evaluation metrics of the initial 3DGS output after 30K iterations, providing a benchmark for the quality and performance of the 3D reconstruction process. These metrics include training time, FPS, SSIM, PSNR, LPIPS, loss, and memory usage, offering a comprehensive overview of the system’s performance.

From Table 2, the quality of the 3D-RISE’s output is heavily influenced by the number of input images extracted during the initial video capture. For example, the excavator dataset (103 images) achieved the highest PSNR (35.254) and SSIM (0.951), showcasing the benefits of adequate image coverage in creating realistic and detailed reconstructions. In contrast, the steel structure dataset (65 images) had the lowest scores for PSNR (22.984) and SSIM (0.783), indicating that fewer images limit the 3D-RISE’s ability to capture finer details and achieve high reconstruction quality. The random model dataset (219 images), while having the most images, demonstrated a more moderate improvement in quality, with a PSNR of 25.854 and SSIM of 0.869. This suggests that while a higher image count generally improves quality, the complexity of the dataset and diminishing returns on additional frames also play a role. Additionally, the random model dataset required the most computational resources, with the highest memory usage (5.231 GB) and the longest training time (1h 16m 55s).

Table 2. Metric results for initial 3D-RISE testing.

Metric **Dataset**	**Number of images used**	**Train**[a]	**FPS**[b]	***SSIM*↑**[c]	***PSNR*↑**[d] ***7K***	***PSNR*↑ *30K***	***LPIPS*↓**[e]	***Loss***[f]	**Memory**[g]
Steel structure	65	32 m 16 s	15.49	0.783	23.106	22.984	0.251	0.018	2.253 GB
Excavator	103	39 m 47 s	12.57	0.951	33.397	35.254	0.103	0.014	1.845 GB
Random model	219	1h 16 m 55 s	6.50	0.869	24.711	25.854	0.183	0.042	5.231 GB

[a] Reflects the total time required to complete 30K iterations, which varies depending on the GPU used.

[b] Indicates the number of iterations processed per second during rendering.

[c] Measures the perceptual similarity between rendered images and ground truth, with higher values indicating better quality.

[d] Evaluates the accuracy of image reconstruction, with higher values reflecting reduced noise and distortion. We have separated into PSNR at 7K iterations *versus* 30K iteration to show a clear improvement in output quality with extended training.

[e] Assesses the perceptual similarity of image patches, with lower score indicating higher fidelity.

[f] Tracks the error during training, where a lower loss value indicates effective learning and improved reconstruction accuracy.

[g] Indicates the GPU memory consumption during processing, an important factor for scalability.

The metrics reveal that while larger datasets and higher iterations generally produce better visual quality, they also demand significantly more computational resources. Balancing dataset size and training iterations with available hardware resources are crucial to achieve optimal performance and quality for 3D scene outputs.

To further validate the computational efficiency and reconstruction quality of 3D-RISE, we conducted a comparative analysis against a representative geometry-based method, Open Multi-View

Stereo (OpenMVS) [58]. We render novel views from reconstructed meshes from the two methods and evaluate PSNR and SSIM against ground truth images (Figure 15). Figure 16 and Table 3 show the comparison results from both methods.

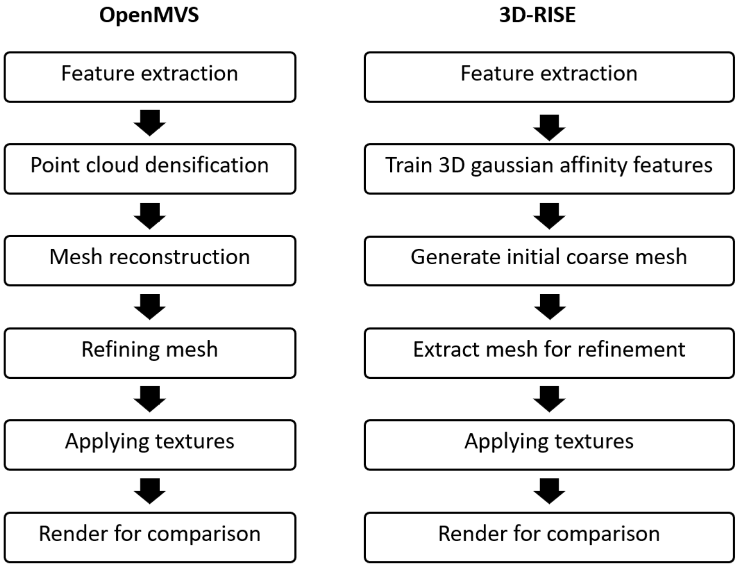

Figure 15. Evaluation process between geometry-based methods and deep-learning based methods.

While OpenMVS demonstrated faster reconstruction times due to the absence of deep neural network training, it consistently produced lower PSNR and SSIM values compared to 3D-RISE. This indicates reduced visual accuracy and detail fidelity in the output. In contrast, the 3D-RISE workflow, which integrates deep learning techniques, delivered significantly higher reconstruction quality, albeit at the cost of increased computational time and GPU resource usage.

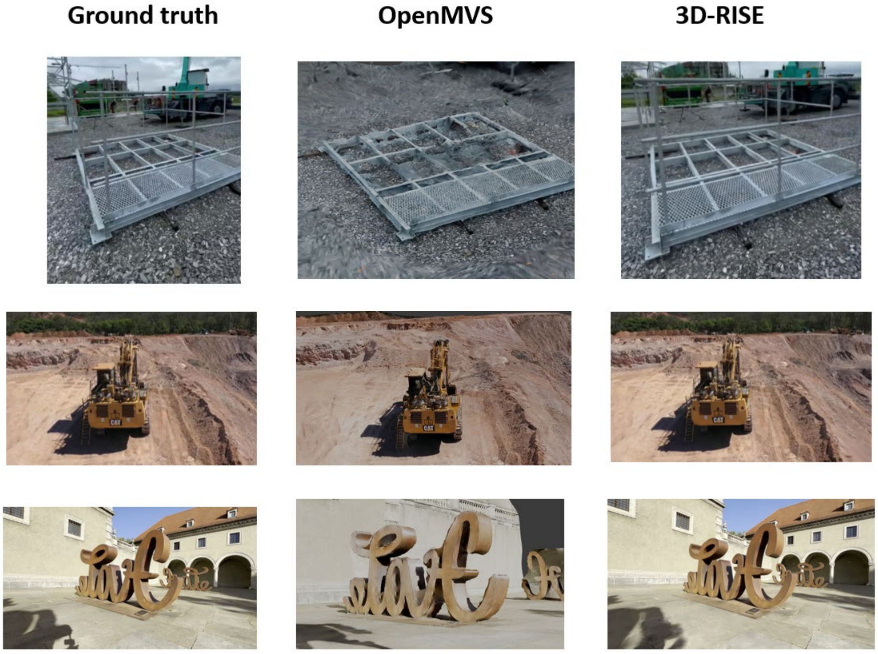

Figure 16. Visual comparison between ground truth images and outputs from both 3D reconstruction methods.

Table 3. Metric results comparison between OpenMVS and 3D-RISE.

Metric Dataset	Train	*SSIM*↑	*PSNR*↑	Train	*SSIM*↑	*PSNR*↑
Methods	**OpenMVS**			**3D-RISE**		
Steel structure	12 m 27 s	0.183	11.230	32 m 16 s	0.783	22.984
Excavator	22 m 05 s	0.317	13.260	39 m 47 s	0.951	32.254
Random model	40 m 35 s	0.392	10.040	1 h 16 m 55 s	0.869	25.854

4.2. SAGA results and segmentation quality

The segmentation results produced by SAGA across the three different scenes (Excavator, steel structure, and random model) are illustrated in Figure 17. In this figure, the red-colored outlines represent the segmented objects identified during the process, while the blue-colored regions denote background elements that were not selected for segmentation. These visualisations highlight the effectiveness of SAGA in isolating target objects from complex scenes while removing unwanted background elements.

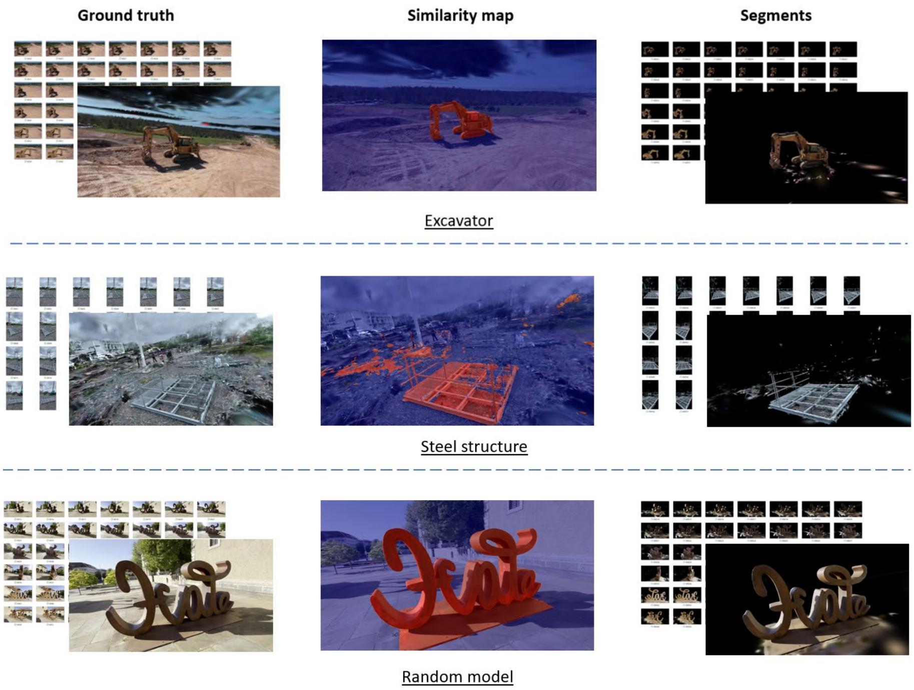

Figure 17. Qualitative results of SAGA across three different scenes.

Overall, the results demonstrate that SAGA performed well in segmenting desired objects across the scenes. However, the effectiveness of segmentation was influenced by the number of input images in each dataset. For instance, the steel structure scene, which had the lowest number of images (65 images), exhibited challenges in identifying and extracting the desired object due to the low resolution and blurriness in the 3D scene generated by 3DGS. As seen in Figure 17, the segmentation results struggled to accurately capture the desired object, often requiring additional manual cleaning after running the model to refine the output, which is shown in Figure 18.

As explained in Figure 9, it is evident that selecting an appropriate score threshold value is important to achieve a good quality segmentation. In our experiments, thresholds ranging from 0.75 to 0.90 were used depending on the scene being segmented. While a higher threshold generally yields better segmentation results by reducing false positives, it also increases the risk of over-segmentation, where irrelevant parts of the scene are mistakenly included as part of the object. This sensitivity to over-segmentation necessitated manual inspection of the segmented frames to ensure quality and consistency. Irregular outputs resulting from over-segmentation could cause issues during reprocessing in COLMAP, as inaccuracies in the updated segmented dataset may disrupt the generation of a refined segmented 3D scene before proceeding with SuGaR.

4.3. SuGaR results and mesh refinement

Figure 18 showcases the result of SuGaR process across three different scenes, comparing the raw mesh outputs from SAGA to the refined meshes after processing with CloudCompare. This figure highlights the effectiveness of SuGaR in generating detailed meshes and the critical role of CloudCompare in refining these outputs.

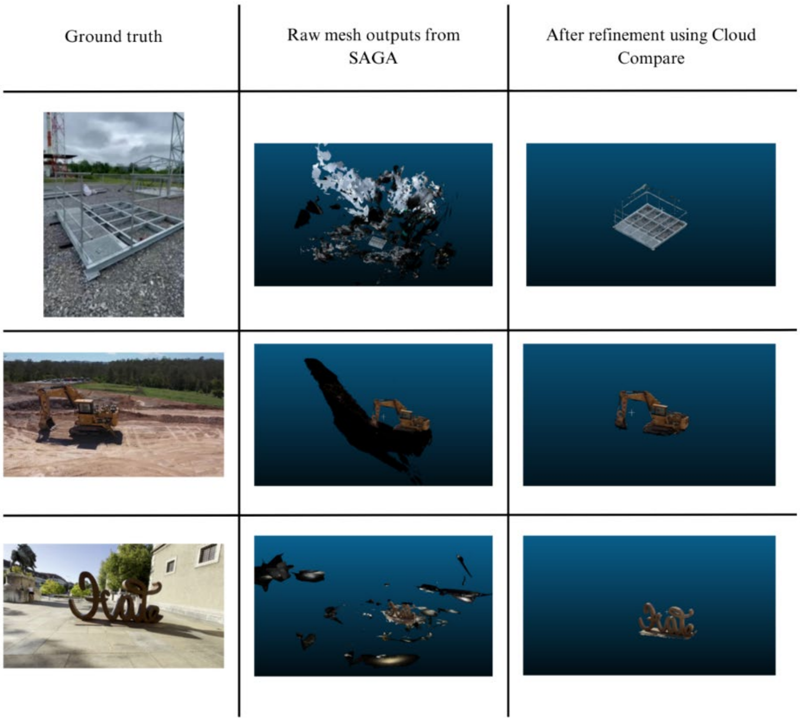

Figure 18. Comparing mesh outputs with different scenes.

Among the three scenes, the steel structure scene produced the noisiest raw mesh, requiring the most extensive cleaning. This is primarily due to the lower number of images in the dataset, which results in less accurate reconstruction and a higher presence of artifacts. Despite these challenges, CloudCompare proved to be an efficient tool for mesh refinement, offering an intuitive interface that requires minimal effort to clean and refine the mesh outputs.

While the SuGaR process successfully generated meshes for all scenes, it is important to note that the output meshes were not perfect and required additional refinement. This highlights the inherent limitations of the 3D-RISE workflow, where further cleaning and adjustments are necessary to prepare the meshes for practical applications, such as dynamic SLP or virtual simulations. Hence, all SuGaR output models, whether full models or segmented models, require additional cleaning to ensure they are ready to use and import in UE. While it is possible to manually crop out unwanted meshes to obtain a segmented result using only SuGaR and CloudCompare, the integration of SAGA significantly simplifies the cleaning process when creating clean, segmented meshes. SAGA effectively removes most of the background and noise during segmentation, providing a much cleaner initial mesh and reducing the effort needed during refinement. For full models that do not involve SAGA, cleaning artifacts and refining the mesh is still necessary. However, CloudCompare remains an efficient and user-friendly tool, making it straightforward to achieve the desired result regardless of the workflow.

4.4. Final construction scene outputs

The integration of all models generated by 3D-RISE workflow into UE showcases the ability to create dynamic and immersive virtual environments. Figure 19 presents the results from 3D-RISE worfklow, comparing the ground truth scenes with their modified versions. Each individual scene serves as a foundation to import and integrate models from other datasets, demonstrating the flexibility of the approach. For example, the excavator scene can incorporate segmented elements such as the steel structure and random model, or vice versa. Similarly, other scenes can be customised with elements from different datasets, creating varied and dynamic setups without altering the original foundation. Additionally, our work supports the inclusion of online 3D models for example construction-related 3D models, which can complement the real-life 3D models generated by 3D-RISE. This capability expands the creative possibilities for constructing virtual environments, enabling more complex and detailed scene configurations.

Figure 19. Transformation of real scenarios into virtual customisation scenes using 3D-RISE.

From the results, the level setup in UE (Figure 12) includes features such as lighting and shadow casting, allowing realistic and immersive visualisation outputs. The ability to accurately render shadows on mesh models further elevates the quality of the scenes, making them suitable for applications such as SLP and construction simulations.

4.5. Practical applications

A notable advantage of this approach is the freedom to modify scenes in UE, including editing terrain, adjusting object placements, and augmenting the datasets to create diverse and holistic construction site scenarios. This flexibility allows users to dynamically adapt the environment to meet specific project needs, providing valuable tools for experimentation and planning. The practical applications of the proposed method are twofold, optimising SLP, and enhancing construction safety.

In the context of SLP, most current SLP methods optimise construction facilities from the 2D plane space, which makes large-scale environments that require the 3D spaces to determine the safety difficult to guarantee. One of the examples is hoisting and tower crane placements in the SLP space [12]. Having the ability to utilise real-time 3D spatial information as a constraint for determining the placement of the space allows for more accurate and context-aware decision making. The reconstructed 3D stie can help ensure that tower cranes are positioned with adequate clearance from nearby structures and within optimal operating ranges.

For safety training, the proposed model potentially enhances the realism and flexibility of virtual reality environments. Traditional virtual reality-based construction safety programs often rely on fixed scenarios [59], limiting their adaptability to real-world site conditions. With the ability to capture, segment, and reconstruct actual construction scenes, 3D-RISE enables the generation of immersive, editable training environments tailored to specific project phases and site configurations. This allows workers to train in virtual scenes that closely mirror their actual work environment during safety training to reduce risks and injuries.

However, one limitation of 3D-RISE is that while the mesh models that are imported into the 3DGS-rendered scenes can be customised, 3DGS-rendered scenes itself cannot be directly modified in UE. To overcome this, it is recommended to capture an empty environment, such as a vacant construction site, and then import all relevant models generated by 3D-RISE or sourced online to build and customise the desired scene as needed.

4.6. Future work

While the 3D-RISE demonstrates significant advancements in creating customisable and realistic construction scenes, it relies heavily on datasets that capture a full 360-degree view of the environment. This dependency on comprehensive datasets poses challenges when the input data is insufficient or when capturing a complete scene is impractical.

A promising direction for future work is to leverage generative AI to address this limitation. With the rapid advancements in generative AI, it may become feasible to generate 3D-ready models from a single image or a limited dataset [60]. This approach could significantly reduce the need for extensive data collection while maintaining the quality and realism required for dynamic scene applications. By training generative models to extrapolate and reconstruct missing details, users could create fully interactive and editable 3D scenes from minimal input data, enabling greater flexibility and scalability.

Integrating generative AI into the 3D-RISE workflow would not only streamline the workflow creation process but also expand its applicability to scenarios where complete scene data is unavailable. This would mark a significant step toward making dynamic SLP and construction simulations more accessible and efficient.

5. Conclusion

3-Dimensional Reconstruction, Integration, Segmentation, and Editing (3D-RISE), as its name suggests, achieves a comprehensive and efficient workflow for creating dynamic and customisable construction scenes. By integrating advanced techniques such as 3D reconstruction, segmentation, and mesh refinement, 3D-RISE offers a robust pipeline for applications such as SLP and construction safety training simulations. The proposal of 3D-RISE demonstrates the successful integration of 3DGS, SAGA, and SuGaR into a unified pipeline, overcoming challenges faced in construction by enabling realistic, editable, and versatile virtual environments. Furthermore, 3D-RISE offers a cost-effective and accessible solution, requiring only standard video input to achieve scene modification and customisation, ultimately helping the construction industry improve communication, enhance safety training, and optimise SLP workflows. The ability to import models into UE further enhances the workflow, allowing users to modify terrain, adjust object placements, and add external elements for a realistic and interactive virtual environment.

While limitations exist, such as the inability to directly edit 3DGS-rendered scenes, these can be addressed using mesh-based models and custom scene creation workflows. The evaluation results demonstrate that 3D-RISE achieves high-quality outputs while maintaining flexibility and adaptability. However, future work aims to explore the integration of generative AI, enabling the generation of 3D-ready models from minimal input data such as a single image. This advancement could significantly enhance the 3D-RISE scalability and applicability, marking a step forward in the field of virtual construction simulations.

Ultimately, 3D-RISE provides a comprehensive and efficient approach to revolutionising construction site simulations and plannings, offering immense potential for improving safety, efficiency, and decision-making in the construction industry.

Acknowledgement

The authors thank RMIT RACE Hub for supporting this research by providing access to a high-performance NVIDIA A100 Tensor Core GPU workstation. This advanced computational resource enabled smooth and seamless testing of the 3D-RISE model, effectively meeting the project’s high computational demands. The generous provision of this hardware was crucial to the successful completion of the work.

Author’s contribution

Ping Chai: writing—review and editing, writing—original draft, validation, resources, project administration, methodology, formal analysis, data curation, investigation, conceptualization, resources. Lei Hou: writing—review and editing, supervision, data curation, resources. Xian Wan Lo: methodology, formal analysis, data curation, investigation, resources. Guomin (Kevin) Zhang: writing—review and editing, supervision, data curation. Haosen Chen: writing—review and editing, data curation, resources. Yang Zou: writing—review and editing, resources. All authors have read and agreed to the published version of the manuscript.

Conflicts of interests

The authors declare that they have no conflicts of interest.

References

[1] Jaafar K, Elbarkouky R, Kennedy J. Construction site layout optimization model considering cost and safety in a dynamic environment. *Asian J. Civ. Eng.* 2021, 22(2):297−312.

[2] Williams T. Identifying success factors in construction projects: a case study. *Project Manage. J.* 2016, 47(1):97–112.

[3] Parsamehr M, Perera U, Dodanwala T, Perera P, Ruparathna R. A review of construction management challenges and BIM-based solutions: perspectives from the schedule, cost, quality, and safety management. *Asian J. Civ. Eng.* 2023, 24(1):353–389.

[4] Musarat M, Alaloul W, Zainuddin S, Qureshi A, Maqsoom A. Digitalization in malaysian construction industry: awareness, challenges and opportunities. *Results Eng.* 2024, 21:102013.

[5] Nakanishi Y, Kaneta T, Nishino S. A review of monitoring construction equipment in support of construction project management. *Front. Built Environ.* 2022, 7:632593.

[6] Janjic T. Why is data management important in construction industry? 2024. Available: https://bexelmanager.com/why-is-data-management-important-in-construction-industry/ (accessed on 22 July 2024).

[7] Kiziltas S, Akinci B. The need for prompt schedule update by utilizing reality capture technologies: a case study. In *Construction Research Congress 2005: Broadening Perspectives*, San Diego, USA, April 5–7, 2005, pp. 1–10.

[8] Cheng X, Jia M, He J. A large-scale dataset of buildings and construction sites. *Comput.-Aided Civ. Infrastruct. Eng.* 2024, 39(9):1390–1406.

[9] Lin T, Maire M, Belongie S, Bourdev L, Girshick R, *et al.* Microsoft COCO: common objects in context. In *Computer vision–ECCV 2014: 13th European conference*, Zurich, Switzerland, September 6–12, 2014, pp. 740–755.

[10] Xiao B, Kang S. Development of an image data set of construction machines for deep learning object detection. *J. Comput. Civ. Eng.* 2021, 35(2):05020005.

[11] Salah M, Khallaf R, Elbeltagi E, Wefki H. Construction site layout planning: a social network analysis. *Buildings* 2023, 13(10):2637.

[12] Jiang W, Zhou Y, Ding L, Zhou C, Ning X. UAV-based 3D reconstruction for hoist site mapping and layout planning in petrochemical construction. *Autom. Constr.* 2020, 113:103137.

[13] Kim M, Ham Y, Koo C, Kim T. Simulating travel paths of construction site workers via deep reinforcement learning considering their spatial cognition and wayfinding behavior. *Autom. Constr.* 2023, 147:104715.

[14] Juliani A, Berges VP, Teng E, Cohen A, Harper J, *et al.* Unity: a general platform for intelligent agents. *arXiv* 2020, arXiv:1809.02627.

[15] Zhu Z, Cheng J, Jeelani I, Gheisari M, Issa RRA. Virtual onboarding: construction site orientation for new employees in a VR environment. In *Construction Research Congress 2024*, Las Vegas, USA, March 20–24, 2024, pp. 11–19.

[16] Bang S, Baek F, Park S, Kim W, Kim H. Image augmentation to improve construction resource detection using generative adversarial networks, cut-and-paste, and image transformation techniques. *Autom. Constr.* 2020, 115:103198.

[17] Hawarneh AA, Bendak S, Ghanim F. Construction site layout planning problem: past, present and future. *Expert Syst. Appl.* 2021, 168:114247.

[18] Pevec K, Pučko Z. On-site visualization using BIM and extended reality. In *Proceedings of the 2024 European Conference on Computing in Construction*, Chania, Greece, July 15–17, 2024, pp. 1–8.

[19] Zhang Y, Wang Z, Zhang J, Shan G, Tian D. A survey of immersive visualization: focus on perception and interaction. *Vis. Inform.* 2023, 7:22–35.

[20] Boddie J. Evolution of surveying: total station's historical development. 2024. Available: https://www.mysurveyingdirect.com/blogs/surveying/total-stations-history (accessed on 9 September 2024).

[21] Cui D, Wang W, Hu W, Peng J, Zhao Y, *et al.* 3D reconstruction of building structures incorporating neural radiation fields and geometric constraints. *Autom. Constr.* 2024, 165:105517.

[22] Tang P, Vick S, Chen J, Paal SG. Surveying, geomatics, and 3D reconstruction. In *Infrastructure Computer Vision*, 1st ed. Oxford: Butterworth-Heinemann, 2020. pp. 13–64.

[23] Ma X, Wang Z, Li H, Zhang P, Ouyang W, *et al.* Accurate monocular 3D object detection via color-embedded 3d reconstruction for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, South Korea, October 27–November 2, 2019, pp. 6851–6860.

[24] Danilina N, Slepnev M, Chebotarev S. Smart city: automatic reconstruction of 3D building models to support urban development and planning. In *VI International Scientific Conference "Integration, Partnership and Innovation in Construction Science and Education" (IPICSE-2018)*, Moscow, Russia, October 11–13, 2018, p. 03047.

[25] Chen S, Laefer DF, Mangina E, Zolanvari SMI, Byrne J. UAV bridge inspection through evaluated 3D reconstructions. *J. Bridge Eng.* 2019, 24(4):05019001.

[26] Hu D, Gan VJL, Yin C. Robot-assisted mobile scanning for automated 3D reconstruction and point cloud semantic segmentation of building interiors. *Autom. Constr.* 2023, 152:104949.

[27] Nikoohemat S, Diakité AA, Zlatanova S, Vosselman G. Indoor 3D reconstruction from point clouds for optimal routing in complex buildings to support disaster management. *Autom. Constr.* 2020, 113:103109.

[28] Su Y, Wang J, Wang X, Hu L, Yao Y, *et al.* Zero-reference deep learning for low-light image enhancement of underground utilities 3D reconstruction. *Autom. Constr.* 2023, 152:104930.

[29] Cai Z, Yang J, Wang T, Huang H, Guo Y. 3D reconstruction of buildings based on 3D Gaussian Splatting. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* 2024, 48:37–43.

[30] Romanoni A, Fiorenti D, Matteucci M. Mesh-based 3D textured urban mapping. *arXiv* 2017, arXiv:1708.05543.

[31] Mildenhall B, Srinivasan PP, Tancik M, Barron JT, Ramamoorthi R, *et al.* NeRF: representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 2021, 65(1):99–106.

[32] Yu Z, Shen Y, Zhang Y, Xiang Y. Automatic crack detection and 3D reconstruction of structural appearance using underwater wall-climbing robot. *Autom. Constr.* 2024, 160:105322.

[33] Lian H, Liu K, Cao R, Fei Z, Wen X, *et al.* Integration of 3D Gaussian Splatting and neural radiance fields in virtual reality fire fighting. *Remote Sens.* 2024, 16(13):2448.

[34] Kerbl B, Kopanas G, Leimkuehler T, Drettakis G. 3D Gaussian Splatting for real-time radiance field rendering. *ACM Trans. Graph.* 2023, 42:1–14.

[35] Jiang Y, Yu C, Xie T, Li X, Feng Y, *et al.* VR-GS: a physical dynamics-aware interactive gaussian splatting system in virtual reality. In *ACM SIGGRAPH 2024 Conference Papers*, Hamburg, Germany, November 13–16, 2024, p. 1.

[36] Ahmed S, Hossain M, Hoque I, A brief discussion on augmented reality and virtual reality in construction industry. *J. Syst. Manag. Sci.* 2017, 7(3):1–33.

[37] Schonberger JL, Frahm JM. Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, June 26–July 1, 2016, pp. 4104–4113.

[38] Minaee S, Boykov YY, Porikli F, Plaza AJ, Kehtarnavaz N, *et al.* Image segmentation using deep learning: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 2021, 44(7):3523–3542.

[39] Kheradmandi N, Mehranfar V. A critical review and comparative study on image segmentation-based techniques for pavement crack detection. *Constr. Build. Mater.* 2022, 321:126162.

[40] Seo J, Han S, Lee S, Kim H, Computer vision techniques for construction safety and health monitoring. *Adv. Eng. Inf.* 2015, 29(2):239–251.

[41] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, *et al*. Segment anything. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, October 2–6, 2023, pp. 3992–4003.

[42] Cen J, Zhou Z, Fang J, Yang C, Shen W, *et al.* Segment anything in 3D with NeRFs. *Adv. Neural Inf. Process. Syst.* 2023, 36:25971–25990.

[43] Chen X, Tang J, Wan D, Wang J, Zeng G. Interactive segment anything NeRF with feature imitation. *arXiv* 2023, arXiv:2305.16233.

[44] Fan Z, Wang P, Jiang Y, Gong X, Xu D, *et al.* NeRF-SOS: any-view self-supervised object segmentation on complex scenes. *arXiv* 2022, arXiv:2209.08776.

[45] Cen J, Fang J, Yang C, Xie L, Zhang X, *et al.* Segment any 3D Gaussians. *arXiv* 2025, arXiv:2312.00860.

[46] Guédon A, Lepetit V. SuGaR: surface-aligned Gaussian Splatting for efficient 3D mesh reconstruction and high-quality mesh rendering. *arXiv* 2023, arXiv:2311.12775.

[47] Kazhdan M, Bolitho M, Hoppe H. Poisson surface reconstruction. In *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*, Goslar, Germany, June 26–28, 2006, pp. 61–70.

[48] Rangelov D, Waanders S, Waanders K, van Keulen M, Miltchev R. Impact of data capture methods on 3D reconstruction with Gaussian Splatting. *J. Imaging* 2025, 11:65.

[49] Koschel A, Müller C, Reiterer A. Selection of key frames for 3D reconstruction in real time, *Algorithms* 2021, 14(11):303.

[50] Tomar S. Converting video formats with FFmpeg. *Linux J.* 2006, 146:10.

[51] Rangelov D, Waanders S, Waanders K, van Keulen M, Miltchev R. Impact of camera settings on 3D reconstruction quality: insights from NeRF and Gaussian Splatting. *Sensors* 2024, 24(23):7594.

[52] CloudCompare: 3D point cloud and mesh processing software. 2024. Available: https://www.cloudcompare.org/ (accessed on 22 December 2024).

[53] Li J, Huang Q, Wang X, Xi B, Duan J, *et al.* A method for the 3D reconstruction of landscape trees in the leafless stage. *Remote Sens.* 2025, 17(8):1473.

[54] Antova G, Peev I. Comparison on commercial and free software for point cloud processing. *Inż. Miner.* 2024, 1(1):511–519.

[55] Sepulveda J, Capps J, Johnson K, Parada C, Garcia A, *et al.* Photogrammetric modeling of subterranean features through three-dimensional software analysis. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* 2020, 44:95–98.

[56] Epic Games, Unreal Engine. 2024. Available: https://www.unrealengine.com (accessed on 23 December 2024).

[57] XVERSE 3D-GS UE Plugin. 2023. Available: https://github.com/xverse-engine/XV3DGS-UEPlugin (accessed on 5 December 2024).

[58] OpenMVS: open Multi-View Stereo reconstruction library. 2020. Available: https://github.com/cdcseacave/openMVS (accessed on 27 June 2025).

[59] Gupta A, Varghese K. Scenario-based construction safety training platform using virtual reality. In *37th International Symposium on Automation and Robotics in Construction (ISARC 2020)*, Kitakyushu, Japan, October 27–28, 2020, pp. 1099–1106.

[60] Chai P, Hou L, Zhang G, Tushar Q, Zou Y. Generative adversarial networks in construction applications. *Autom. Constr.* 2024, 159:105265.