

Article

Fast 3D Gaussian Reconstruction for Open-Pit Mine Teleoperated Excavation via Monocular-LiDAR Fusion

Lin Bi ^{1,*} , Muqian Tan ¹, Ziyu Zhao ¹ , Jinbo Li ^{1,2} and Xintong Wang ³

¹ School of Resources and Safety Engineering, Central South University, Changsha 410083, China; 235512142@csu.edu.cn (M.T.); 225501023@csu.edu.cn (Z.Z.); 245501017@csu.edu.cn (J.L.)

² School of Mining Engineering and Geology, Xinjiang Institute of Engineering, Urumqi 830023, China

³ College of Information and Intelligence, Hunan Agricultural University, Changsha 410128, China; 1625225378@stu.hunau.edu.cn

* Correspondence: mr.bilin@csu.edu.cn

Abstract

Teleoperated open-pit excavation requires fast and reliable 3D scene modeling under lightweight sensor configurations. To this end, this paper proposes a monocular camera–LiDAR fusion-based fast 3D Gaussian reconstruction method tailored for teleoperated open-pit excavation. The proposed approach uses only two sensors, a monocular camera and LiDAR, and integrates SPNet, a depth completion network, to improve the geometric completeness of the reconstructed scene. It further introduces a stride-aware initialization strategy that leverages the depth–stride correlation to jointly construct the initial Gaussian set and estimate the initial scales. During optimization, scale and color regularization are applied to prevent uncontrolled growth of Gaussians. Experiments in a Carla-simulated open-pit excavation scenario show that, under high-resolution input of 1920×1080 , the proposed method achieves a stable 3D model update rate of approximately 2.5 Hz. The reconstruction quality under training viewpoints reaches PSNR 30.5388, SSIM 0.9161, and LPIPS 0.1333. Compared with 4DTAM and MonoGS, the proposed method achieves better overall reconstruction quality. It also maintains a much higher update rate than 4DTAM and a comparable update rate to MonoGS. Ablation studies further verify the critical contribution of the depth completion module and the stride-aware initialization strategy to the overall reconstruction performance. In addition, preliminary validation on field data further demonstrates the applicability of the proposed method under real-world open-pit excavation-loading conditions. The proposed method generates stable and usable 3D models of rock-pile working face under a lightweight sensor configuration, providing a reliable geometric basis for remote situational awareness and excavation assistance.

Keywords: open-pit mine; teleoperated excavation; 3D Gaussian; monocular reconstruction; scene updating

MSC: 68T45



Academic Editors: Gintautas Dzemyda and Vladimir Balan

Received: 27 February 2026

Revised: 18 March 2026

Accepted: 1 April 2026

Published: 2 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The intelligent transformation of open-pit mining has become a prevailing trend, and the loading stage is increasingly shifting toward remote and teleoperated operation. The core objective of this shift is to ensure operator safety and improve production efficiency [1–4]. Under teleoperated conditions, operators mainly rely on multiple camera video streams to infer spatial relationships. The inherent limitations of two-dimensional

imagery, such as insufficient depth perception, restricted viewpoints, and occlusion, make it difficult to accurately judge the relative positions between the bucket and the rock pile, thereby reducing operational efficiency and increasing safety risks [5–8]. In existing open-pit mine 3D modeling practice, full-site 3D surveys are still typically conducted every 2–3 months. Although local model updates are relatively more frequent, they are still mainly performed on a daily or weekly basis [9]. This still falls short of the operational demand for sustained high-frequency updating in teleoperated excavation. Therefore, teleoperated excavation requires a three-dimensional working-face modeling method with fast update capability and stable results, so as to provide a reliable geometric foundation for situational awareness and subsequent auxiliary functions [10–13].

Existing 3D modeling approaches for open-pit mining environments can be broadly categorized into four classes: methods that acquire point clouds using terrestrial laser scanning or LiDAR and directly use them as the scene representation; methods that further derive surface meshes or volumetric models from point clouds; multi-view image-based reconstruction methods such as Structure-from-Motion and Multi-View Stereo (MVS); and Simultaneous Localization and Mapping (SLAM) methods designed for online incremental mapping [14–16]. Point clouds and their derived mesh representations are susceptible to non-uniform sampling, noise, occlusion, and missing regions, and typically require intensive filtering and hole-filling to yield a stable surface, making it even more difficult to ensure consistent outputs under high-frequency update requirements [17,18]. Volumetric representations facilitate fusion and rendering, but in large-scale open-pit scenarios, they often incur substantial memory and computational overhead, necessitating more sophisticated data structures and engineering optimizations [19,20]. MVS relies on sufficient multi-view coverage and reliable pixel correspondences; it can degrade under weak textures, illumination changes, occlusions, and dynamic disturbances, and its overall computational cost is high, which is unfavorable for the high-refresh modeling demands of teleoperated operations [15,21]. LiDAR SLAM in open-pit mine environments must also contend with dust-induced outliers and noise, interference from dynamic equipment and vehicles, and large terrain variations and scene scale, all of which undermine the stability and robustness of map updates [16,22].

Three-dimensional Gaussian Splatting (3DGS) explicitly models a scene radiance field using a set of optimizable 3D Gaussian primitives and achieves high-quality real-time novel-view synthesis at high resolution via efficient rasterization-based rendering [23]. This explicit point-primitive representation combines the visual fidelity of volumetric rendering with the flexibility of point-cloud-style geometric expression, enabling stable free-viewpoint inspection in interactive interfaces [23]. Moreover, 3DGS can be integrated into camera–LiDAR fusion pipelines, where LiDAR geometric priors are incorporated to improve geometric reliability in large-scale outdoor scenes, which closely matches the typical input modalities available in open-pit mining reconstruction scenarios [24–26]. Meanwhile, 3DGS has begun to appear in digital modeling and visualization studies related to mining and geosciences, further indicating its potential for practical deployment in open-pit mine 3D reconstruction [27–29].

In recent years, 3DGS has gradually evolved from offline modeling of static scenes and novel-view synthesis to online incremental mapping for static environments. Representative work such as MonoGS integrates 3DGS into a SLAM pipeline, enabling continuous updates of a Gaussian map under monocular input and thereby producing a 3D scene representation that can be rendered in real time [30]. However, such online mapping methods are predominantly designed for bounded indoor environments, and their performance often degrades when extended to unbounded outdoor settings [26,31]. Even when LiDAR is incorporated to provide accurate surface points, regions with insufficient

depth supervision may still remain, leading to degraded geometric accuracy [24–26]. To alleviate geometric under-constraint under single-view observation, several studies have introduced depth completion, fusing sparse depth with imagery to produce denser depth estimates, thereby improving scene representation accuracy and enhancing multi-view geometric consistency [32–35]. Meanwhile, to meet the modeling demands of dynamic scenes, 4DTAM further proposes a time-varying modeling approach based on surface Gaussians under RGB-D input, enabling 4D tracking and mapping of dynamic, non-rigid scenes and offering important insights for dynamic-scene modeling [36]. However, because open-pit excavation scenes are large-scale and subject to strong environmental disturbances, commonly used RGB-D cameras struggle to provide reliable depth observations consistently [37,38]. As a result, existing methods are difficult to apply directly to this setting, and further research is still required on reconstruction for open-pit excavation scenarios.

Therefore, we propose a fast 3DGS reconstruction method tailored for teleoperated open-pit excavation, based on monocular camera–LiDAR fusion. The method reconstructs only the excavation region of interest in the scene, mitigates geometric incompleteness caused by point-cloud sparsity and occlusions through depth completion, and employs stride-aware initialization to construct a high-quality initial set of Gaussian primitives, thereby accelerating convergence and improving reconstruction stability. Finally, we validate the proposed method in a Carla-simulated open-pit excavation scenario and further conduct a preliminary validation on field data, demonstrating its reconstruction quality and model update capability under high-resolution inputs as well as its applicability under real-world operating conditions. The remainder of this paper is organized as follows: Section 2 introduces the proposed method; Section 3 presents comparative experiments, ablation results, and preliminary validation on field data; Section 4 analyzes the reconstruction results, ablation results, and field-data results, and discusses limitations; and Section 5 concludes the paper.

2. Approach

This section presents a monocular camera–LiDAR fusion-based 3D Gaussian reconstruction method for teleoperated open-pit excavation, aiming to produce stable and practically usable 3D models of rock-pile working faces under limited observability and online update constraints. In this study, the simulation data are generated in the Carla (version 0.9.12) virtual environment using simulated camera and LiDAR sensors, whereas the field data are collected using an SG3S-ISX031C-GMSL2F-H100 monocular camera (Sensing World, Shenzhen, China) and a JT128 LiDAR (Hesai Technology, Shanghai, China). As illustrated in Figure 1, the method first projects the LiDAR point cloud onto the image plane to obtain sparse depth observations and then applies SPNet-based depth completion to recover finer geometric details of the scene. We then propose a stride-aware initialization strategy that explicitly incorporates the relationship between pixel sampling stride and depth into the initialization process, enabling simultaneous Gaussian point set generation and initial scale assignment within the reconstruction region to improve the quality of the initial representation and accelerate training convergence. Building on this initialization, we jointly optimize Gaussian appearance and scale parameters via differentiable rendering and introduce scale and color regularization to suppress unconstrained expansion, improving optimization stability and rendering consistency. Finally, the reconstructed 3D Gaussian model can be directly deployed for visualization and assistance tasks in teleoperated excavation interfaces, providing reliable geometric support for subsequent situational awareness and operational decision-making.

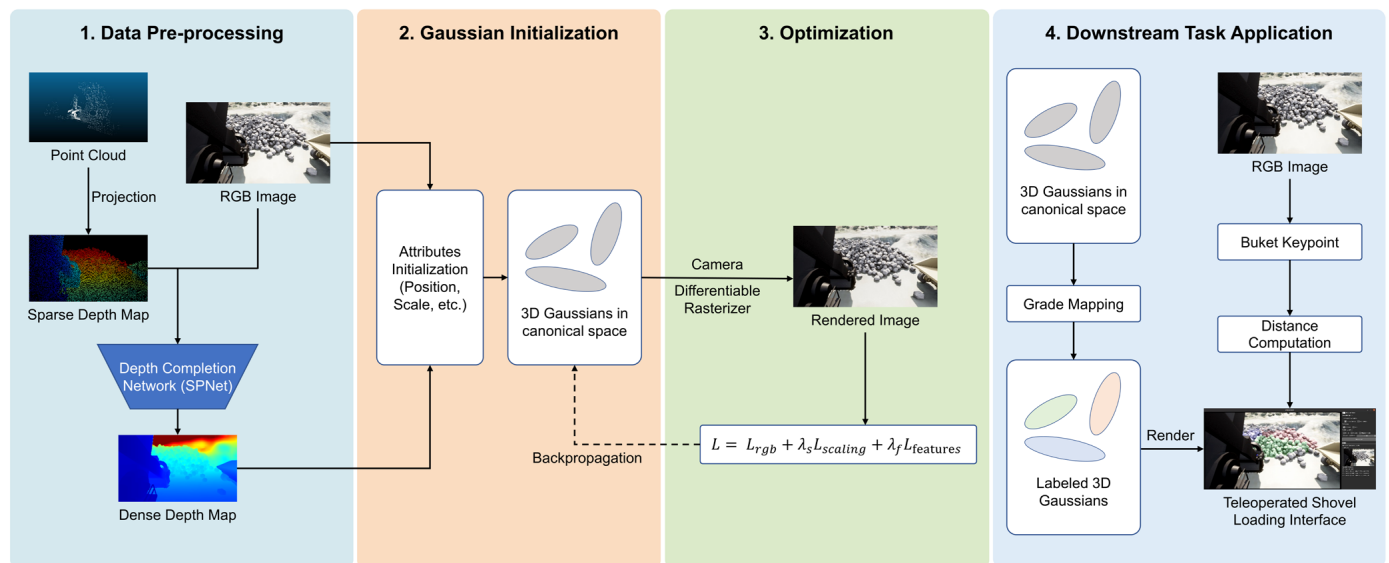


Figure 1. Research framework of this study.

2.1. Fundamental Principles of 3D Gaussian Splatting

Explicitly, 3DGS models a scene radiance field using a set of differentiable 3D Gaussian primitives [23]. For the i -th Gaussian primitive, its state is jointly parameterized by the 3D mean $\mu_i \in \mathbb{R}^3$, the 3D covariance matrix $\Sigma_i \in \mathbb{R}^{3 \times 3}$, the opacity $\alpha_i \in [0, 1)$, and the color c_i :

$$g_i = \{\mu_i, \Sigma_i, c_i, \alpha_i\}, \quad (1)$$

where c_i is represented by spherical-harmonic coefficients to characterize view-dependent radiance color. To ensure that the covariance matrix remains positive semidefinite throughout optimization, 3DGS expresses Σ_i as a composition of rotation and scale:

$$\Sigma_i = R_i S_i S_i^T R_i^T, \quad (2)$$

where $R_i \in \mathbb{R}^{3 \times 3}$ denotes the rotation matrix and $S_i \in \mathbb{R}^{3 \times 3}$ denotes the diagonal scale matrix. The initial scale is obtained by estimating local neighborhood distance statistics for each point via k-nearest-neighbor (KNN) search and mapping them to the Gaussian's initial scale, so that the scale matches the local point-cloud density.

During training, 3DGS uses differentiable rendering to project and composite the current set of Gaussians into an image, constructs a loss function from the reconstruction error between the rendered output and the observed image, and iteratively updates the geometric and appearance parameters of the Gaussians via backpropagation. To enhance the representation of scene structure and fine details, the training process performs adaptive density control on the Gaussian set based on error or gradient signals: primitive density is increased in regions of interest via cloning or splitting, while low-contributing primitives are removed through pruning. This enables the Gaussian set to progressively better fit the scene's geometric and appearance distributions, albeit at the cost of increased computation time.

2.2. Teleoperation-Oriented Initialization with a Stride-Aware Scaling Rule

In open-pit excavation scenarios, point clouds can provide reliable geometric anchors, yet they remain sparse relative to the scene scale, making it difficult to support stable optimization of 3D Gaussians. Therefore, we incorporate a depth completion network constrained by sparse depth observations into the modeling pipeline, using the sparse depth from the point cloud as a metric-scale reference and leveraging RGB images to

regularize geometric structure. This produces image-aligned dense depth maps, thereby improving the accuracy and coverage of 3D Gaussian initialization. In implementation, we adopt SPNet as the depth completion network because it achieves high inference efficiency with a compact model size while maintaining accuracy, exhibits strong cross-scene generalization, and can be deployed directly using pretrained weights without additional fine-tuning [35,39]. In contrast, purely monocular depth estimation often exhibits large variations in metric accuracy in real-world settings [40] and thus fails to meet the high-precision and consistency requirements of modeling excavation scenes; accordingly, it is not considered in this study.

We unify the sparse depth ground truth and the dense depth prediction in the same pixel coordinate system by constructing a uniform grid on the image plane with a fixed pixel stride s , and selecting a representative pixel $p = [u, v]^T$ within each grid cell as the sampling location. We then query the depth at this pixel: if a valid sparse depth derived from point-cloud projection is available at p , it is prioritized as the geometric constraint; otherwise, the dense depth predicted by the depth completion network is used for completion. Because sampling is performed on the image plane with a fixed stride, the initialized point set achieves more uniform spatial coverage, and its size follows $N \propto HW/s^2$, allowing the initialization density and runtime efficiency to be directly controlled via s . For each sampled pixel p with depth d , we form the homogeneous pixel coordinate $\tilde{p} = [u, v, 1]^T$, back-project it into the camera coordinate frame using the camera intrinsics to obtain a 3D point as in Equation (3), and then transform it to the world coordinate frame using the camera extrinsics as in Equation (4).

$$X_c = dK^{-1}\tilde{p}, \quad (3)$$

$$X_w = R_{wc}X_c + t_{wc}, \quad (4)$$

where K denotes the camera intrinsic matrix, and R_{wc} and t_{wc} are the rotation matrix and translation vector from the camera coordinate frame to the world coordinate frame, respectively. The resulting X_w is used as the initial center of a 3D Gaussian, yielding an initialized point set that is strictly aligned with image pixels, uniformly covers the reconstruction region, and has controllable density, thereby providing a stable starting point for subsequent fast optimization.

Because the point cloud used for initialization essentially corresponds to surface samples of the working face, and our reconstruction is performed under a single-view setting where geometric cues are mainly provided by local surface structure, we initialize each 3D Gaussian as a thin sheet that extends along the image plane while being constrained along the viewing direction. The scaling matrix S of a 3D Gaussian is initialized as:

$$S = \text{diag}(s_x, s_y, s_z), \quad (5)$$

$$s_z = \tau \min(s_x, s_y), \quad (6)$$

where s_x and s_y denote the scales along the image-plane directions, s_z denotes the scale along the viewing direction, and $\tau > 0$ is a thickness ratio coefficient set to 0.001 in this work to prevent unreasonable expansion of Gaussians along the viewing direction. The in-plane scale s_x and s_y is jointly determined by the image resolution and the point depth. For a point with depth d , under the pinhole camera model, the spatial footprint of one pixel is approximately proportional to d/f . Accordingly, we set s_x and s_y as the product of this ratio and the sampling stride, and impose lower and upper bounds to ensure numerical stability, as follows:

$$s_x = \text{clip}\left(\frac{d}{f_x}k, s_{\min}, s_{\max}\right), \quad (7)$$

$$s_y = \text{clip}\left(\frac{d}{f_y}k, s_{\min}, s_{\max}\right), \quad (8)$$

where f_x and f_y are the focal lengths along the image x-axes and y-axes, respectively; k is a coefficient proportional to the pixel sampling stride s ; $s_{\min}$ and $s_{\max}$ are the minimum and maximum thresholds for the in-plane scale; and $\text{clip}(\cdot)$ truncates the scale within the prescribed bounds to avoid geometric blurring from overly large scales or insufficient coverage from overly small scales.

To alleviate the under-constrained optimization of appearance parameters in a single-view setting, we represent the color of each 3D Gaussian using zeroth-order spherical harmonics (i.e., a view-independent constant RGB parameter). The Gaussian rotation is initialized to face opposite the camera optical axis. Meanwhile, the opacity of each 3D Gaussian is initialized to a relatively high value. With the reliable geometric initialization, this ensures sufficient rendered contributions in the early stage, preventing weak gradient signals caused by negligible initial contributions and thereby improving optimization stability and accelerating convergence.

During training, because observation constraints are limited under a single-view setting, optimizing solely with a photometric loss can cause 3D Gaussians to reduce local reconstruction error by continuously increasing their scale parameters or amplifying appearance feature magnitudes. This may lead to excessive Gaussian expansion and abnormal accumulation of pixel color contributions, ultimately resulting in unstable artifacts such as blurred renderings and abnormal global brightness. To suppress these degeneracies, we introduce regularization terms on the Gaussian scale parameters and appearance features during optimization to constrain their unconstrained growth and keep both geometry and appearance within a reasonable range. In single-frame optimization, we do not perform adaptive density control and jointly optimize the Gaussian parameters using the following loss function:

$$Loss = L_{rgb} + \lambda_s L_{scale} + \lambda_f L_{feature}, \quad (9)$$

where L_{rgb} denotes the photometric loss, while L_{scale} and $L_{feature}$ are regularization terms on the 3D Gaussian scale and color parameters, respectively. λ_s and λ_f are their corresponding weights, which are set to 10.0 and 1.0 in this work. The terms L_{rgb} , L_{scale} and $L_{feature}$ are defined as follows:

$$L_{rgb} = \frac{1}{3HW} \|\hat{I} - I\|_1, \quad (10)$$

$$L_{scale} = \frac{1}{3N} \sum_{i=1}^N \sum_{d \in \{x,y,z\}} [\max(0, s_{i,d} - s_{i,d}^{(0)} - m_s)]^2, \quad (11)$$

$$L_{feature} = \frac{1}{3N} \sum_{i=1}^N \sum_{c \in \{r,g,b\}} [\max(0, |f_{i,c} - f_{i,c}^{(0)}| - m_f)]^2, \quad (12)$$

where H and W are the image height and width, respectively; $\hat{I}$ and I denote the rendered image and the input image; and $\|\hat{I} - I\|_1$ denotes the element-wise absolute error summed over all pixels and RGB channels. N is the number of Gaussian primitives; $s_{i,d}$ denotes the scale parameter of the i -th Gaussian along dimension d ; $s_{i,d}^{(0)}$ is its initial scale; and m_s is the margin threshold for scale. $f_{i,c}$ denotes the color parameter of the i -th Gaussian in channel c ; $f_{i,c}^{(0)}$ is the corresponding initial color value; and m_f is the margin threshold for color. This design allows parameters to vary freely within a permissible range, while applying a quadratic penalty to deviations beyond the margin, thereby suppressing unconstrained expansion and improving geometric stability.

3. Experiments

3.1. Experimental Setup

The experiments were conducted in an open-pit excavation scenario built on the Carla simulation platform, where camera images (1920×1080), LiDAR point clouds, and camera pose information were recorded synchronously during operation. The sensing suite consists of a monocular RGB camera and a LiDAR sensor, rigidly mounted above the excavator cab with the field of view oriented toward the rock-pile working face. This mounting configuration is consistent with sensor layouts commonly adopted in real-world teleoperated excavation. The sensor mounting location and a representative camera viewpoint are shown in Figure 2. All experiments were performed on a standard desktop computer running Ubuntu 20.04.4 LTS, equipped with an Intel Core i7-7800X CPU, 32 GB DDR4 memory, and an NVIDIA GeForce RTX 3090 GPU with 24 GB of VRAM.



Figure 2. Sensor configuration in the Carla simulation. (a) Sensor mounting positions, where the blue region indicates the camera field of view. (b) Camera viewing direction.

We conduct comparative experiments against 4DTAM [36] and MonoGS [30] to validate the overall performance and further perform ablation studies to separately evaluate the contributions of the depth completion module and the initialization rule to reconstruction quality. We quantitatively evaluate the results using three complementary metrics, namely PSNR, SSIM, and LPIPS, which measure rendering quality in terms of pixel-wise error, structural consistency, and perceptual similarity. Let $\hat{I}$ denote the rendered image, I the ground-truth image, and MAX the maximum possible pixel value. PSNR is defined based on the mean squared error $\text{MSE}(\hat{I}, I)$ as follows:

$$\text{MSE}(\hat{I}, I) = \frac{1}{3HW} \|\hat{I} - I\|_F^2, \quad (13)$$

$$\text{PSNR}(\hat{I}, I) = 10 \log_{10} \frac{\text{MAX}^2}{\text{MSE}(\hat{I}, I)}, \quad (14)$$

PSNR characterizes reconstruction error at the pixel level. SSIM measures the consistency of local luminance, contrast, and structural information, and is defined as:

$$\text{SSIM}(\hat{I}, I) = \frac{(2\mu_{\hat{I}}\mu_I + C_1)(2\sigma_{\hat{I}I} + C_2)}{(\mu_{\hat{I}}^2 + \mu_I^2 + C_1)(\sigma_{\hat{I}}^2 + \sigma_I^2 + C_2)}, \quad (15)$$

where $\mu_{\hat{I}}^2$ and μ_I^2 are local means, $\sigma_{\hat{I}}^2$ and σ_I^2 are local variances, $\sigma_{\hat{I}I}$ is the local covariance, and C_1 and C_2 are stabilizing constants. LPIPS leverages a pretrained AlexNet to extract deep features and computes the feature-space distance between the rendered and ground-truth images, reflecting perceptual differences. For these metrics, higher PSNR and SSIM indicate better quality, whereas lower LPIPS is preferable.

3.2. Comparison with Baselines

For baseline selection, we choose 4DTAM and MonoGS as the reference methods. Under RGB-D input, 4DTAM represents the scene using dynamic surface Gaussians and, within a sliding window, jointly optimizes camera poses and the Gaussians' geometric and appearance parameters via differentiable rendering, yielding a time-varying 3D scene model [36]. MonoGS integrates 3DGS into a SLAM pipeline, enabling continuous updates of a Gaussian map under monocular input and thereby producing a 3D scene representation that can be rendered in real time [30]. Our evaluation focuses on comparing rendering quality under the same observation conditions. Therefore, candidate baselines must (i) operate with single-view input, (ii) provide a renderable 3D representation, and (iii) produce stable reconstructions in dynamic scenes. Among existing monocular reconstruction methods, only a few satisfy these criteria and are directly comparable to our approach; other methods often rely on additional sensors or specific prior assumptions, making fair comparison under our evaluation protocol difficult. Considering both comparability and representativeness, we adopt 4DTAM and MonoGS as the baselines for quantitative and qualitative experiments. It should be noted that 4DTAM is originally designed for RGB-D input, and MonoGS can also operate with RGB-D input, whereas our method takes RGB images, LiDAR point clouds, and ground-truth camera poses as input. To ensure a fair and comparable evaluation, we use a consistent configuration for 4DTAM and MonoGS: (i) we directly use the ground-truth camera poses recorded during data acquisition and disable pose optimization; (ii) we provide depth maps registered to the RGB images as the depth input. Qualitative rendering comparisons are shown in Figure 3, and quantitative results are reported in Figure 4 and Table 1.

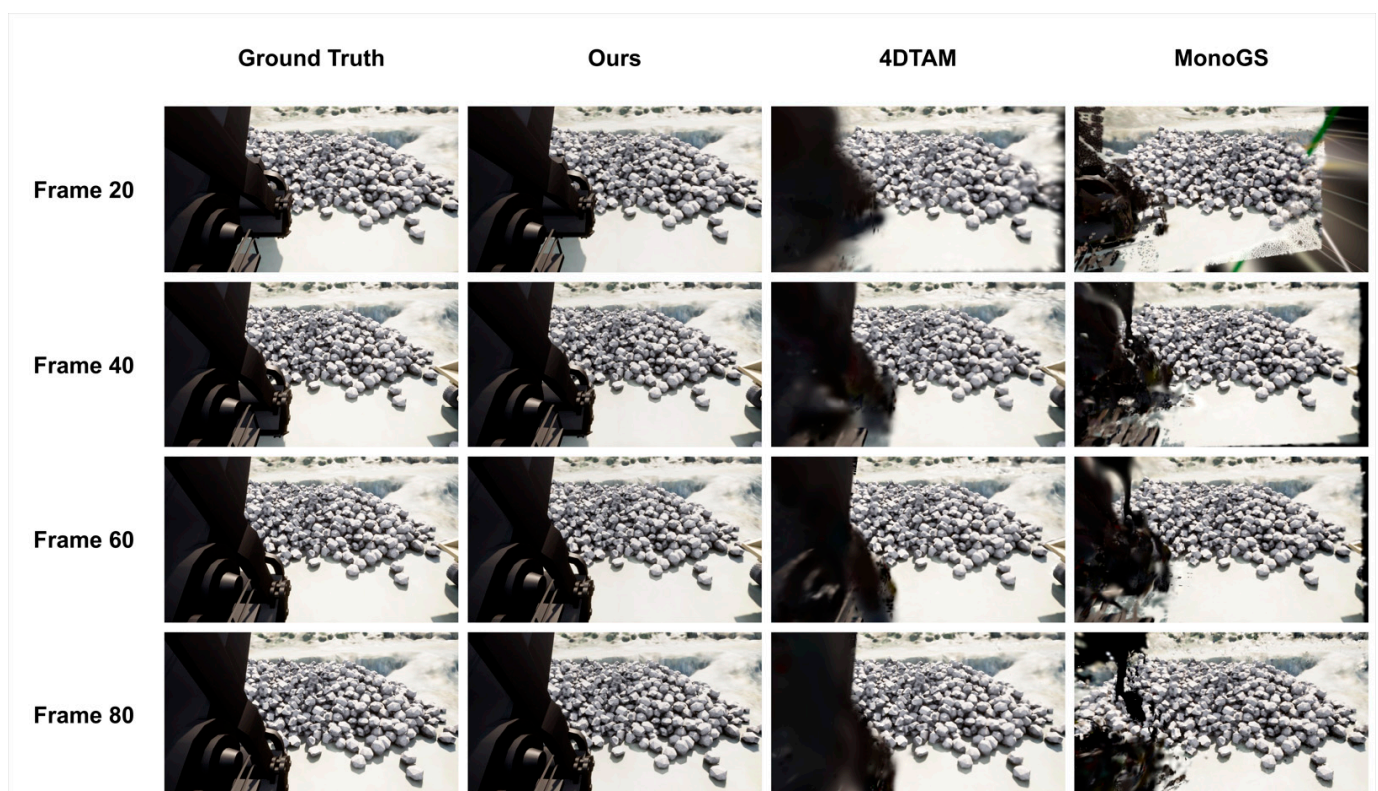


Figure 3. Qualitative comparison with 4DTAM and MonoGS across frames.

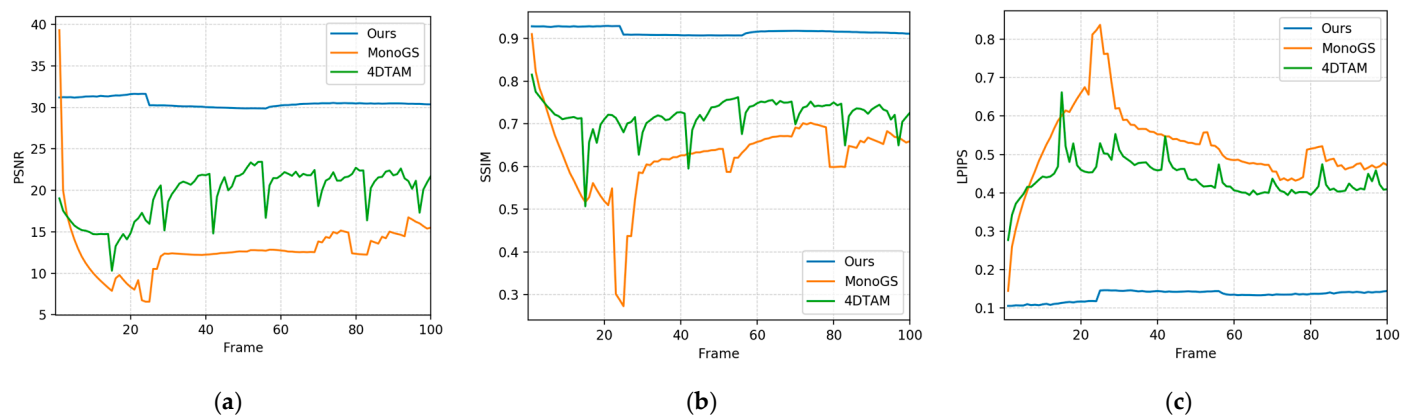


Figure 4. Frame-wise rendering quality metrics comparison among Ours, 4DTAM, and MonoGS. (a) PSNR. (b) SSIM. (c) LPIPS.

Table 1. Average rendering quality and runtime comparison.

Methods	PSNR ¹	SSIM ¹	LPIPS ¹	Update Rate ²
Ours	30.5388	0.9161	0.1333	2.3240
4DTAM	19.6060	0.7215	0.4384	0.1844
MonoGS	12.8296	0.6250	0.5207	2.1908

¹ PSNR, SSIM, and LPIPS are averaged over 100 frames; ² update rate is reported in Hz, indicating the 3D model update frequency.

3.3. Ablation Study

To validate the effectiveness of the key components in our method, we conduct ablation studies on the depth completion module (SPNet) and the stride-aware initialization strategy. All ablation settings follow the same experimental configuration as in Section 3.1, and are evaluated on the same 100-frame sequence using PSNR, SSIM, and LPIPS. Table 2 summarizes the quantitative results under different combinations of these strategies, and Figure 5 presents the corresponding qualitative rendering comparisons.

Table 2. Ablation study on depth completion and stride-aware initialization.

Depth Completion *	Stride-Aware Initialization	Original Initialization	PSNR	SSIM	LPIPS
		✓	8.3732	0.4369	0.8152
	✓		4.2882	0.0891	0.7456
✓		✓	22.3074	0.8550	0.2794
✓	✓		30.5388	0.9161	0.1333

* Depth completion is implemented using SPNet. A check mark (✓) indicates that the corresponding component is enabled, while a blank cell indicates that it is not used.

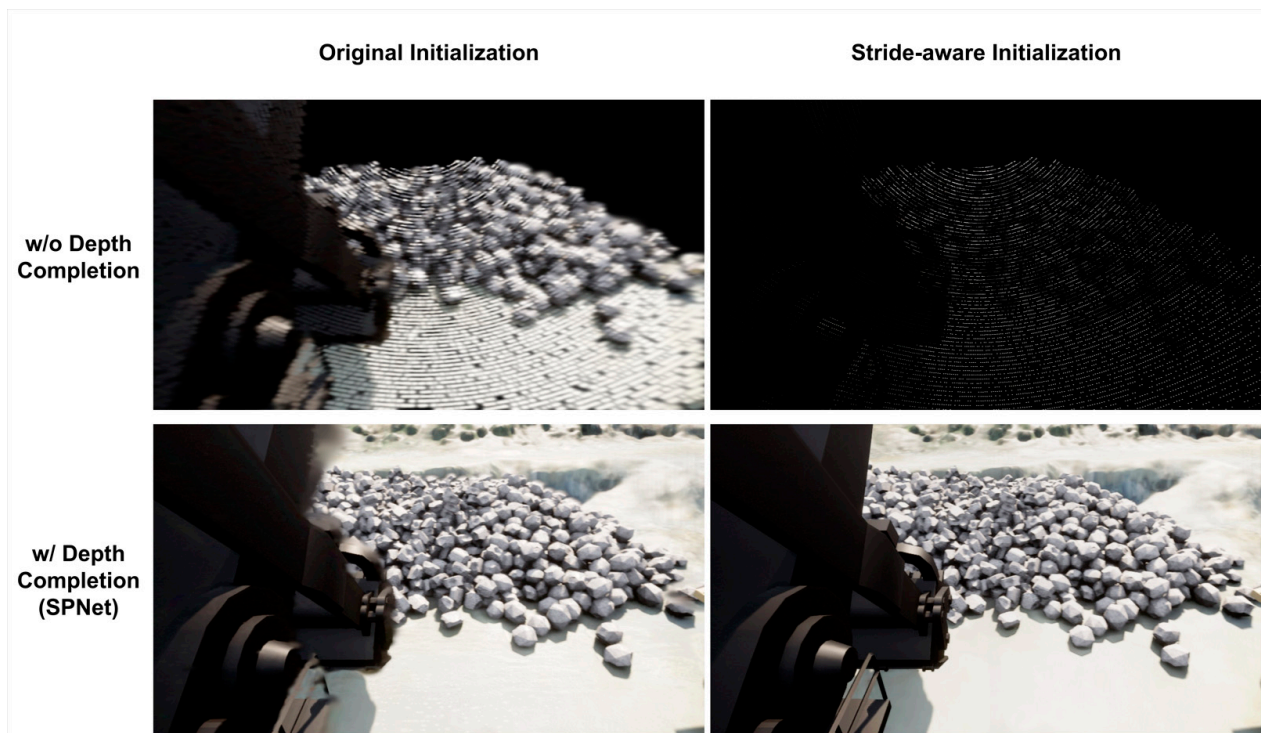


Figure 5. Qualitative ablation on depth completion and stride-aware initialization.

3.4. Application to Teleoperated Excavation

To validate the practical usability of the proposed reconstruction method, we implement a complete teleoperated excavation reconstruction pipeline on a Carla-simulated open-pit excavation dataset, as shown in Figure 6. Based on the reconstructed 3D rock-pile model, we develop illustrative assistive visualization functions for a teleoperation interface, including rule-based rock-pile region annotation and an interactive display of bucket localization. These functions help operators understand the working-face geometry and the relative spatial relationship between the bucket and the rock pile during key stages such as approaching the target, adjusting the pose, completing loading, and preparing for dumping. As illustrated in Figure 6, the reconstructed model provides intuitive 3D situational awareness throughout the teleoperated excavation process.

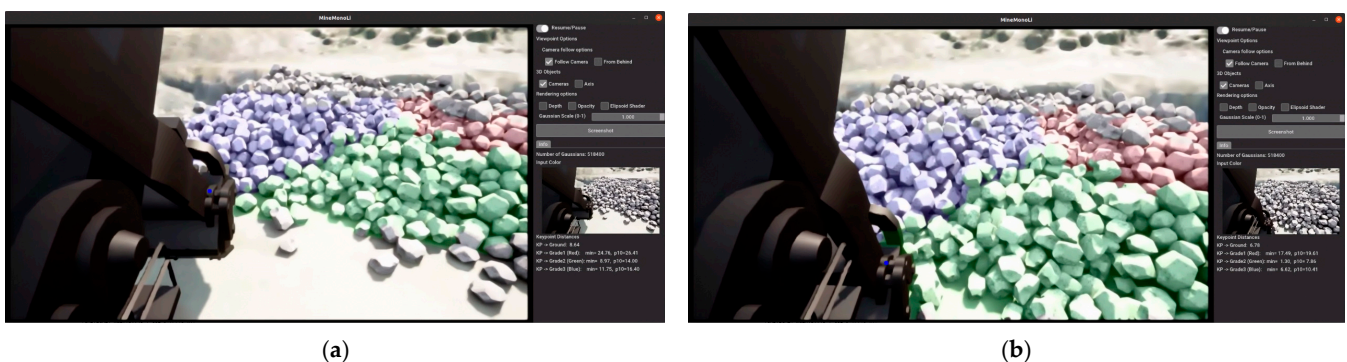


Figure 6. Cont.

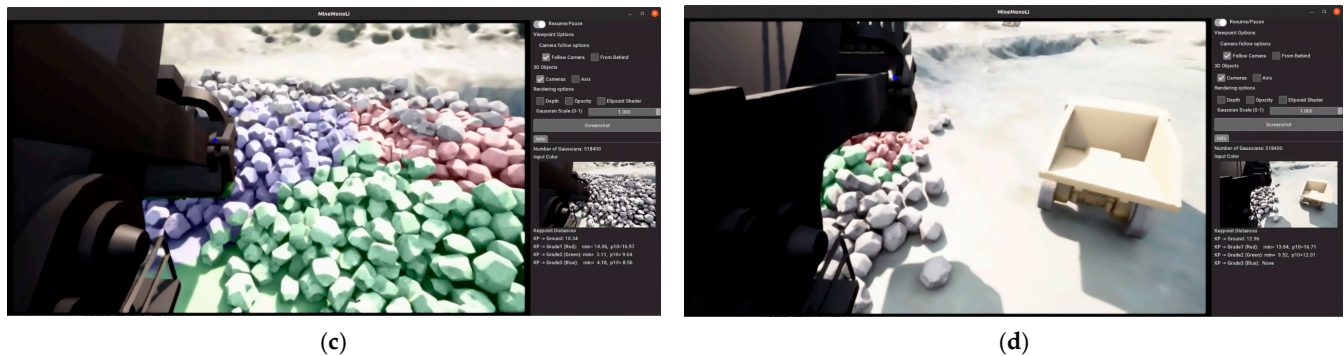


Figure 6. Teleoperation-oriented visualization using the reconstructed rock pile. The interface overlays rule-based region labels and bucket positioning for operator assistance, where different colors indicate different ore-grade regions: (a) approaching the target region, (b) adjusting the bucket pose, (c) after bucket filling, and (d) before unloading.

Moreover, to evaluate the online modeling capability of the proposed method under continuous input, we measure the model update rate on sequences of varying lengths. The results are summarized in Table 3; for each sequence length, the measurement is repeated twice. The system maintains a stable update rate of approximately 2.5 Hz across all sequence lengths, indicating that the proposed method can provide continuous reconstructions for teleoperated excavation.

Table 3. Model update rate on datasets of different lengths.

Dataset Length (Frames)	568	568	1500	1500
Update rate (Hz)	2.48	2.51	2.52	2.51

The achieved update rate of approximately 2.3–2.5 Hz is not yet sufficient for using the reconstructed 3D model as the primary visual feedback channel in direct teleoperation. However, it is still meaningful for auxiliary operator assistance, since the reconstructed model is mainly used to provide supplementary geometric awareness of the working surface rather than replace live video streams. Under this role, the current update frequency can still provide timely structural cues about the latest surface condition. For more direct teleoperation-oriented interaction, a substantially higher refresh rate would be required, with around 10–15 Hz being a more practical target.

3.5. Preliminary Validation on Field Data

To further verify the applicability of the proposed method in real-world open-pit excavation–loading scenarios, we conduct a preliminary test on a segment of a field-data sequence. The sequence contains a total of 1100 image frames, covering both bucket pose variations and continuous changes in the working face. Figure 7 presents representative reconstruction results from the field-data sequence. Starting from frame 100, 1 frame is selected every 100 frames, and the rendered results of nine selected frames are shown to visually illustrate the continuous reconstruction performance of the proposed method under real operating conditions.

To quantitatively evaluate the reconstruction results on the field-data sequence, we further compute the per-frame PSNR, SSIM, and LPIPS metrics and record the average model update frequency. Figure 8 shows the per-frame metric curves over the 1100-frame sequence, and Table 4 reports the corresponding average metric results.



Figure 7. Representative reconstruction results on the field-data sequence.

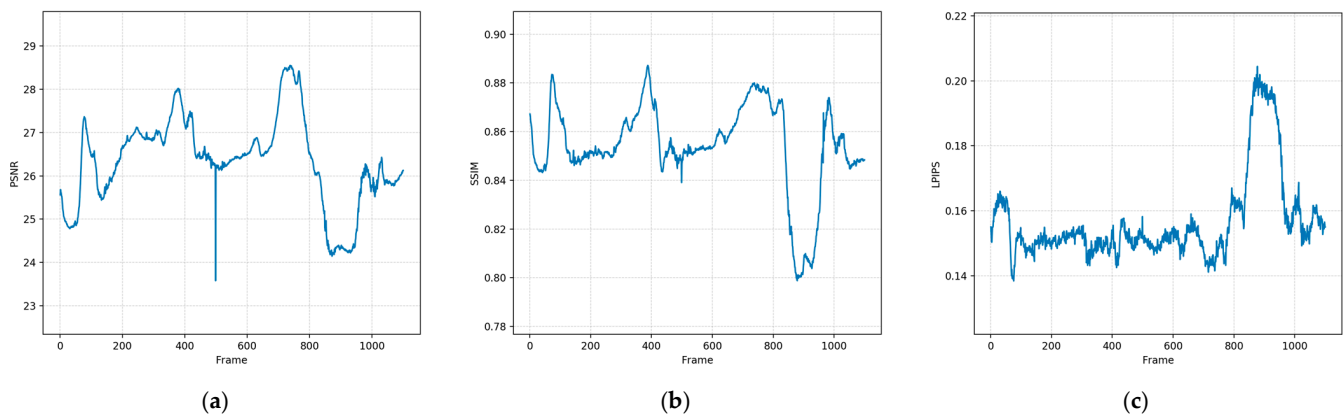


Figure 8. Frame-wise rendering quality metrics on the field-data. (a) PSNR. (b) SSIM. (c) LPIPS.

Table 4. Average reconstruction quality and update rate on the field-data sequence.

Dataset Length (Frames)	PSNR	SSIM	LPIPS	Update Rate (Hz)
1100	26.3389	0.8541	0.1567	1.95

4. Discussion

4.1. Analysis of Reconstruction Results

As shown in Figure 3, we present qualitative comparisons among our method, 4DTAM, and MonoGS at frames 20, 40, 60, and 80. Overall, our method preserves more consistent appearance details across frames and produces reconstructions that are closer to the ground truth. This is especially evident around the rock-pile boundaries and the interfaces between the boom/bucket and the surrounding scene. In contrast, 4DTAM is more prone to localized blurring and dark smearing artifacts in these regions. These artifacts persist across multiple frames and lead to degraded boundary stability and reduced detail fidelity. MonoGS

shows even stronger structural disturbances and local artifacts. More severe ghosting and fragmented regions can be observed around the bucket and along the foreground–background boundaries. As a result, its structural consistency across frames is weaker. Overall, our method preserves scene structure and appearance details more stably under this teleoperated excavation-view sequence.

Figure 4 plots the frame-wise curves of PSNR, SSIM, and LPIPS for the three methods over the 100-frame interval. Our method maintains higher and more stable PSNR and SSIM throughout the sequence. It also consistently achieves lower LPIPS. These results indicate that the rendered images are closer to the ground truth in both pixel-wise and perceptual terms, with smaller inter-frame fluctuations. In contrast, 4DTAM yields lower PSNR and SSIM, higher LPIPS, and larger fluctuations. This suggests that it is more sensitive to occlusions and rapid viewpoint changes under this teleoperation view, and that it has difficulty producing consistent reconstructions. MonoGS performs worse than 4DTAM overall. At the beginning of the sequence, some of its metrics are relatively high. However, as the bucket motion and scene changes become more pronounced, its PSNR and SSIM drop rapidly, while LPIPS rises sharply. In the latter part of the sequence, the curves become relatively stable, but they remain at a lower quality level. This indicates that MonoGS, as a method mainly designed for monocular Gaussian SLAM and online incremental mapping, is more easily affected under the current sequence with obvious motion variation and continuous changes in the working face. Overall, the quantitative trends in Figure 4 are consistent with the qualitative observations in Figure 3.

Table 1 further summarizes the average performance of the three methods. Our method achieves a PSNR of 30.5388, an SSIM of 0.9161, and an LPIPS of 0.1333. Compared with 4DTAM, this corresponds to a PSNR gain of 10.9328 dB, an SSIM increase of 0.1946, and an LPIPS reduction of 0.3051. In addition, our method reaches a model update frequency of 2.3240 Hz, which is about 12.6 times that of 4DTAM at 0.1844 Hz. Compared with MonoGS, our method improves PSNR by 17.7092 dB, increases SSIM by 0.2911, and reduces LPIPS by 0.3874. Its update frequency is also slightly higher than that of MonoGS, with 2.3240 Hz versus 2.1908 Hz. These results show that, compared with 4DTAM, our method achieves both higher reconstruction quality and a much faster update speed. Compared with MonoGS, it delivers a much clearer quality advantage while maintaining a similar update efficiency. Overall, the proposed method provides a better balance between reconstruction fidelity and computational efficiency under the current evaluation setting, and it can offer more stable and timely 3D model feedback for teleoperated excavation.

4.2. Analysis of Ablation Results

Table 2 and Figure 5 illustrate the effects of depth completion (SPNet) and stride-aware initialization on reconstruction quality. The results indicate that depth completion is crucial for stable reconstruction, and stride-aware initialization can further improve detail fidelity and consistency on top of it.

- (1) Depth completion is a major contributor to the performance gains. Without depth completion, PSNR and SSIM remain low, and LPIPS is substantially higher, regardless of the initialization strategy (the first two rows of Table 2). As shown in Figure 5, the rock-pile surface and excavator structures cannot be rendered stably, with evident missing regions and fragmented appearances. This suggests that, when relying only on sparse depth constraints, Gaussian optimization under a limited iteration budget lacks sufficient geometric support, making the reconstruction prone to instability. After incorporating SPNet, PSNR and SSIM increase markedly, while LPIPS decreases substantially (the third and fourth rows of Table 2), indicating that more complete

depth information provides stronger geometric constraints and thereby enhances reconstruction stability and quality.

- (2) Stride-aware initialization further improves local details on top of depth completion. With depth completion enabled, stride-aware initialization further improves performance over the original initialization: PSNR and SSIM increase, while LPIPS decreases (Table 2, third row vs. fourth row). As shown in Figure 5, stride-aware initialization produces sharper contours and clearer textures in the boom/bucket region and on the rock-pile surface. This indicates that combining the more complete geometry provided by depth completion with our initialization strategy yields more reliable and higher-quality reconstructions under a limited optimization budget.

4.3. Analysis of Field-Data Reconstruction Results

As shown in Figure 7, the proposed method can still continuously output 3D reconstruction results of the working face in real-world open-pit excavation-loading scenarios. From the representative rendered results, it can be observed that, under continuous bucket pose changes and ongoing local updates of the working face, the reconstruction can maintain the overall geometric shape of the working face and the relative spatial relationship between the bucket and the working face with reasonable stability. For rock accumulation areas, local slope structures, and the spatial regions near the bucket, the results can provide generally consistent 3D representations. This indicates that, under the lightweight sensing configuration driven by monocular images, LiDAR point clouds, and camera poses, the proposed method is applicable not only to simulation environments but also to continuous scene reconstruction in real excavation-loading conditions.

The frame-wise PSNR, SSIM, and LPIPS curves in Figure 8 show that the overall metrics on the field-data sequence remain within a relatively favorable range, indicating that the proposed method has basic reconstruction stability on long real-world sequences. However, a noticeable local fluctuation can still be observed within a short interval of the sequence. During this period, PSNR and SSIM decrease simultaneously, while LPIPS increases significantly. Based on the field-data conditions, this fluctuation is mainly caused by severe INS drift induced by vibration and jolting during operation, which further increases the camera pose error. Since pose information directly affects the spatial correspondence between images and point clouds, as well as the Gaussian optimization process, pose errors can be propagated to the reconstruction results and lead to short-term quality degradation in this interval. Even so, the curves remain relatively continuous and stable during the rest of the sequence. This suggests that the proposed method still has a certain degree of adaptability to common disturbances in real scenes, while also showing clear sensitivity to pose estimation accuracy.

According to the statistics in Table 4, the average PSNR, SSIM, and LPIPS on the field-data sequence are 26.3389, 0.8541, and 0.1567, respectively, and the average reconstruction update frequency is 1.95 Hz. Compared with the simulation results, both the average reconstruction quality and the update frequency on the field data decrease to some extent, but the overall metrics still remain within a practically usable range. This indicates that the proposed method can still maintain continuous reconstruction and online updating capability under real excavation-loading conditions. The reduced update frequency is mainly caused by the much larger number of LiDAR points in each field-data frame. In the simulation dataset, each frame contains about 30,000 points, whereas in the field dataset, each frame contains about 50,000 points. The higher point density significantly increases the computational cost of point cloud projection, depth-constraint construction, and Gaussian optimization, thereby reducing the overall update speed. In addition, the field scene contains stronger dust interference, illumination variation, and more complex texture

distributions. These factors further increase the difficulty of reconstruction and also affect the update efficiency to a certain extent. Overall, the field-data experiments demonstrate the preliminary practical potential of the proposed method, while also revealing that there is still room for improvement in robustness to pose errors and online reconstruction efficiency under complex operating conditions.

4.4. Limitations and Future Work

(1) The primary limitation of the current algorithm is that the model update speed remains relatively slow. As reported in Table 3, the model update frequency is stable at approximately 2.5 Hz, but further improvements are still required for teleoperated excavation scenarios. The main bottleneck limiting the update speed is the large number of Gaussians required to represent the scene. To preserve the expressive capability for rock-pile details and boundary regions, our method employs a relatively dense initialization sampling strategy to enhance single-view scene representation; however, this also increases the per-frame optimization cost and the overall model update overhead. Notably, simply reducing the number of Gaussians to lower computational cost often degrades detail fidelity and boundary quality, making it difficult to achieve an ideal trade-off between speed and reconstruction quality.

Second, the current system is implemented using a Python (version 3.7.13) multiprocessing architecture, where the reconstruction module and the GUI module run in parallel, and the Gaussian scene data must be transferred across processes to the GUI for rendering after each update. As the number of Gaussians increases, the overhead of serialization, memory copying, and inter-process transmission grows substantially, becoming an additional performance bottleneck that further reduces the model update frequency. The combined impact of this engineering-level communication and visualization overhead and the algorithmic computation cost makes it difficult for the current implementation to achieve higher-frequency online updates.

To further improve the system's online modeling capability, future work will pursue optimizations from both algorithmic and engineering perspectives: (i) at the algorithmic level, we will investigate more efficient representations and initialization strategies to better control the number of Gaussians while preserving reconstruction quality as much as possible; (ii) at the engineering level, we will optimize the multi-process communication and rendering pipeline to reduce cross-process serialization and data-transfer overhead, thereby improving the model update rate.

(2) The 3D representation of the excavator boom and bucket remains incomplete. Due to the sensor placement and the limited viewing angles, the monocular camera and LiDAR primarily observe only one side of the boom/bucket, making it difficult to reconstruct a complete 3D model of this critical component and consequently affecting geometric consistency and visualization quality under novel viewpoints. Future work will consider incorporating additional viewpoints or integrating geometric priors of the boom/bucket for completion, in order to improve model completeness and usability under novel viewpoints.

5. Conclusions

This paper addresses the perception requirements of teleoperated open-pit excavation and proposes a monocular camera–LiDAR fusion-based 3D Gaussian reconstruction method for fast and stable 3D modeling of rock-pile working faces. The method improves geometric completeness under sparse point-cloud observations via depth completion, and introduces a stride-aware initialization strategy to construct an initial Gaussian set with sufficient coverage and reasonable scales, thereby accelerating convergence and enhancing output stability. In the Carla-simulated open-pit excavation scenario, the method achieves

a PSNR of 30.5388, an SSIM of 0.9161, and an LPIPS of 0.1333, with a model update rate of 2.3240 Hz under 1920×1080 input. Compared with 4DTAM and MonoGS, it delivers better overall reconstruction quality, while maintaining a much higher update rate than 4DTAM and a comparable update rate to MonoGS. In the field-data experiment, it reaches an average PSNR of 26.3389, an SSIM of 0.8541, and an LPIPS of 0.1567, with an average update rate of 1.95 Hz . These results indicate that the method can maintain continuous and usable 3D reconstruction not only in simulation but also under real-world open-pit excavation-loading conditions. Using only monocular images, LiDAR point clouds, and camera poses as input, the method provides reliable 3D model support for situational awareness and interactive excavation assistance. Future work will focus on improving model update efficiency by better controlling the number of Gaussians while preserving reconstruction quality, and on enhancing the 3D representation of critical components to mitigate incomplete modeling of the boom and bucket caused by limited viewpoints, thereby improving system usability in real-world teleoperation.

Author Contributions: Conceptualization, M.T. and L.B.; methodology, M.T.; software, M.T.; validation, M.T.; formal analysis, M.T. and L.B.; investigation, M.T. and Z.Z.; resources, L.B.; data curation, Z.Z., J.L. and X.W.; writing—original draft preparation, M.T.; writing—review and editing, L.B. and M.T.; visualization, M.T.; supervision, L.B.; project administration, L.B.; funding acquisition, L.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China, grant number 2023YFC2907305.

Data Availability Statement: The data presented in this study include simulation data and field data. These data are not publicly available due to confidentiality restrictions associated with the real-world open-pit mine site, related map information, and field data acquisition conditions.

Acknowledgments: The authors gratefully acknowledge the support of the funders and colleagues who contributed to this work.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript; or in the decision to publish the results.

References

1. McNab, K.; Garcia-Vasquez, M. *Autonomous and Remote Operation Technologies in Australian Mining*; Centre for Social Responsibility in Mining (CSRM), Sustainable Minerals Institute, The University of Queensland: Brisbane, Australia, 2011.
2. Morinaga, J. *Teleoperation System for Ultra-Large Hydraulic Excavators*; Komatsu: Tokyo, Japan, 2022.
3. Long, M.; Schafrik, S.; Kolapo, P.; Agioutantis, Z.; Sottile, J. Equipment and operations automation in mining: A review. *Machines* **2024**, *12*, 713. [\[CrossRef\]](#)
4. Bodin, U.; Andersson, U.; Dadhich, S.; Uhlin, E.; Marklund, U.; Häggströmff, D. Remote controlled short-cycle loading of bulk material in mining applications. *IFAC-PapersOnLine* **2015**, *48*, 54–59. [\[CrossRef\]](#)
5. Dadhich, S.; Bodin, U.; Sandin, F.; Andersson, U. From tele-remote operation to semi-automated wheel-loader. *Int. J. Electr. Electron. Eng. Telecommun.* **2018**, *7*, 178–182. [\[CrossRef\]](#)
6. Hainsworth, D.W. Teleoperation user interfaces for mining robotics. *Auton. Robot.* **2001**, *11*, 19–28. [\[CrossRef\]](#)
7. Flangas, A.T.; Sattarvand, J.; Dascalu, S.M.; Harris, F.C. Merging live video feeds for remote monitoring of a mining machine. In Proceedings of the 2021 European Symposium on Software Engineering, New York, NY, USA, 19–21 November 2022; pp. 6–13.
8. Sato, R.; Kamezaki, M.; Niuchi, S.; Sugano, S.; Iwata, H. Cognitive untunneling multi-view system for teleoperators of heavy machines based on visual momentum and saliency. *Autom. Constr.* **2020**, *110*, 103047. [\[CrossRef\]](#)
9. Vassena, G.; Clerici, A. Open pit mine 3D mapping by tfs and digital photogrammetry: 3D model update thanks to a slam based approach. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2018**, *42*, 1145–1148. [\[CrossRef\]](#)
10. Kamran-Pishhessari, A.; Moniri-Morad, A.; Sattarvand, J. Applications of 3D reconstruction in virtual reality-based teleoperation: A review in the mining industry. *Technologies* **2024**, *12*, 40. [\[CrossRef\]](#)

11. James, C.A.; Haustein, K.; Bednarz, T.P.; Alem, L.; Caris, C.; Castleden, A. Remote operation of mining equipment using panoramic display systems: Exploring the sense of presence. *Ergon. Open J.* **2011**, *4*, 93–102. [[CrossRef](#)]
12. Tanimoto, T.; Fukano, R.; Shinohara, K.; Kurashiki, K.; Kondo, D.; Yoshinada, H. Research on superimposed terrain model for teleoperation work efficiency. *J. Robot. Mechatron.* **2016**, *28*, 173–184. [[CrossRef](#)]
13. Rafiei, S.; Brunnström, K.; Andersson, J.; Sjöström, M. Investigation of human interaction with an augmented remote operating system for scaling in mining applications. *Qual. User Exp.* **2024**, *9*, 5. [[CrossRef](#)]
14. Kekeç, B.; Bilim, N.; Karakaya, E.; Ghiloufi, D. Applications of terrestrial laser scanning (TLS) in mining: A review. *Türkiye Lidar Derg.* **2021**, *3*, 31–38. [[CrossRef](#)]
15. Stathopoulou, E.K.; Remondino, F. A survey on conventional and learning-based methods for multi-view stereo. *Photogramm. Rec.* **2023**, *38*, 374–407. [[CrossRef](#)]
16. Hui, C. A review of LiDAR-based SLAM techniques for open-pit coal mine environments. *J. Eng. Res. Rep.* **2025**, *27*, 290–302. [[CrossRef](#)]
17. Berger, M.; Tagliasacchi, A.; Seversky, L.M.; Alliez, P.; Guennebaud, G.; Levine, J.A.; Sharf, A.; Silva, C.T. A survey of surface reconstruction from point clouds. *Comput. Graph. Forum* **2017**, *36*, 301–329. [[CrossRef](#)]
18. Wolff, K.; Kim, C.; Zimmer, H.; Schroers, C.; Botsch, M.; Sorkine-Hornung, O.; Sorkine-Hornung, A. Point cloud noise and outlier removal for image-based 3D reconstruction. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 118–127.
19. Nießner, M.; Zollhöfer, M.; Izadi, S.; Stamminger, M. Real-time 3D reconstruction at scale using voxel hashing. *ACM Trans. Graph.* **2013**, *32*, 1–11. [[CrossRef](#)]
20. Dryanovski, I.; Klingensmith, M.; Srinivasa, S.S.; Xiao, J. Large-scale, real-time 3D scene reconstruction on a mobile device. *Auton. Robot.* **2017**, *41*, 1423–1445. [[CrossRef](#)]
21. Wang, F.; Zhu, Q.; Chang, D.; Gao, Q.; Han, J.; Zhang, T.; Hartley, R.; Pollefeys, M. Learning-based multi-view stereo: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2026**, in press.
22. Jiang, X.; Xie, Y.; Na, C.; Yu, W.; Meng, Y. Algorithm for point cloud dust filtering of LiDAR for autonomous vehicles in mining area. *Sustainability* **2024**, *16*, 2827. [[CrossRef](#)]
23. Kerbl, B.; Kopanas, G.; Leimkuehler, T.; Drettakis, G. 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* **2023**, *42*, 131–139. [[CrossRef](#)]
24. Jiang, C.; Gao, R.; Shao, K.; Wang, Y.; Xiong, R.; Zhang, Y. Li-GS: Gaussian splatting with LiDAR incorporated for accurate large-scale reconstruction. *IEEE Robot. Autom. Lett.* **2024**, in press. [[CrossRef](#)]
25. Yan, L.; Song, J.; Xie, H.; Wei, P.; Li, G.; Zhu, L.; Fan, Z.; Gong, S. LiDGS: An efficient 3D reconstruction framework integrating LiDAR point clouds and multi-view images for enhanced geometric fidelity. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *142*, 104730. [[CrossRef](#)]
26. Lang, X.; Li, L.; Wu, C.; Zhao, C.; Liu, L.; Liu, Y.; Lv, J.; Zuo, X. Gaussian-LIC: Real-time photo-realistic SLAM with Gaussian splatting and LiDAR-inertial-camera fusion. In Proceedings of the 2025 IEEE International Conference on Robotics and Automation (ICRA), Atlanta, GA, USA, 19–23 May 2025; pp. 8500–8507.
27. Apopei, A.I. 3D Gaussian splatting in geosciences: A novel high-fidelity approach for digitizing geoheritage from minerals to immersive virtual tours. *Geosciences* **2025**, *15*, 373. [[CrossRef](#)]
28. Richter, F.; Bär, F.; Jung, B. Virtual mineral collections using 3D Gaussian splatting. In Proceedings of the 30th International Conference on 3D Web Technology, New York, NY, USA, 9–10 September 2025; pp. 1–7.
29. Lin, X.; Tao, Y.; Ma, Q.; Yao, W.; Ma, H.; Ma, B. Multidimensional visual perception-enhanced high-precision 3DGS SLAM for underground coal mine robots. *J. China Coal Soc.* **2025**, *50*, 4609–4623.
30. Matsuki, H.; Murai, R.; Kelly, P.H.J.; Davison, A.J. Gaussian splatting SLAM. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 18039–18048.
31. Xin, Z.; Wu, C.; Huang, P.; Zhang, Y.; Mao, Y.; Huang, G. Large-scale Gaussian splatting SLAM. *arXiv* **2025**, arXiv:2505.09915. [[CrossRef](#)]
32. Khan, M.A.U.; Nazir, D.; Pagani, A.; Mokayed, H.; Liwicki, M.; Stricker, D.; Afzal, M.Z. A comprehensive survey of depth completion approaches. *Sensors* **2022**, *22*, 6969. [[CrossRef](#)] [[PubMed](#)]
33. Ma, F.; Karaman, S. Sparse-to-dense: Depth prediction from sparse depth samples and a single image. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 21–25 May 2018; pp. 4796–4803.
34. Qiu, J.; Cui, Z.; Zhang, Y.; Zhang, X.; Liu, S.; Zeng, B.; Pollefeys, M. DeepLiDAR: Deep surface normal guided depth prediction for outdoor scenes from sparse LiDAR data and a single color image. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 3308–3317.
35. Lang, X.; Lv, J.; Tang, K.; Li, L.; Huang, J.; Liu, L.; Liu, Y.; Zuo, X. Gaussian-LIC2: LiDAR-inertial-camera Gaussian splatting SLAM. *arXiv* **2025**, arXiv:2507.04004.

36. Matsuki, H.; Bae, G.; Davison, A.J. 4DTAM: Non-rigid tracking and mapping via dynamic surface Gaussians. In Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 11–15 June 2025; pp. 26921–26932.
37. Liu, Z.; Conde, M.H. Evaluation and compensation of the effect of dirt on time-of-flight 3D imaging systems. *IEEE Sens. J.* **2022**, *22*, 16595–16606. [[CrossRef](#)]
38. Gupta, M.; Yin, Q.; Nayar, S.K. Structured light in sunlight. In Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, 1–8 December 2013; pp. 545–552.
39. Wang, H.; Yang, M.; Zheng, X.; Hua, G. Scale propagation network for generalizable depth completion. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *in press*.
40. Louw, J.M.; Verster, J.; Dickens, J. Assessing Depth Anything V2 monocular depth estimation as a LiDAR alternative in robotics. In *Proceedings of MATEC Web of Conferences*; EDP Sciences: Les Ulis, France, 2025; p. 4002.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.