

Article

Large-Scale Point Cloud Completion Through Registration and Fusion of Object-Level Reconstructions

Taiming He, Yixuan Fang * , Keyuan Li  and Lu Yang

School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; 18908008810@189.cn (T.H.); yanglu@uestc.edu.cn (L.Y.)

* Correspondence: 202444060110@std.uestc.edu.cn

Abstract

Existing 3D reconstruction algorithms commonly struggle with modeling specific local objects within large-scale scenes, often resulting in a lack of local detail and incomplete geometric structures. While current mainstream point cloud completion methods can restore these missing structures to some degree, they are fundamentally based on generative in-filling, a process that relies on geometric priors learned from large-scale datasets. Consequently, the physical realism and geometric accuracy of the results cannot be guaranteed. To address these limitations, this paper proposes a novel, data-driven framework for point cloud completion. Our core method involves the high-precision, heterogeneous data registration and seamless fusion of an object-level point cloud—reconstructed with high-fidelity appearance and geometry by our optimized Neural Radiance Fields (NeRF) framework—with our target large-scale scene point cloud. By using high-precision, physically based data as a strong prior for geometric completion, we offer an alternative route to conventional generative completion methods. Concurrently, we employ unsupervised evaluation metrics to assess the intrinsic quality of the final results. This work provides a robust and high-fidelity solution to the problem of completing local objects within large-scale scenes. Evaluated on our self-constructed UAV-Recon dataset, the proposed method achieved a Structural Plausibility ≥ 0.995 , Geometric Smoothness ≤ 0.19 , and Distribution Uniformity ≈ 1.2 , offering a robust solution for the high-fidelity completion of local objects within large-scale scenes.

Keywords: point cloud completion; 3D reconstruction; point cloud registration; large-scale

1. Introduction

Three-dimensional reconstruction technology has made remarkable progress, serving as a bridge between the physical and digital worlds. Point cloud data, as the core output, have found widespread application in fields such as autonomous driving, robotics navigation, and cultural heritage digitization. Large-scale scene reconstruction has matured, with mainstream methods like Visual Odometry (VO) [1] and Simultaneous Localization and Mapping (SLAM) [2–5] enabling real-time modeling, and offline techniques like Structure from Motion (SfM) and Multi-View Stereo (MVS) constructing high-precision, dense 3D models. Emerging deep learning methods such as Neural Radiance Fields (NeRF) [6–9] and 3D Gaussian Splatting have further expanded the data acquisition avenues.

However, despite improvements in global reconstruction quality, existing methods still struggle to simultaneously meet the demands for both expansive coverage and high-fidelity detail when modeling specific local objects. This fundamental conflict is primarily due to a



Academic Editor: Christos Bouras

Received: 7 December 2025

Revised: 25 December 2025

Accepted: 2 January 2026

Published: 5 January 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.

trade-off: large-scale 3D reconstruction algorithms, while striving for broad coverage and computational efficiency, often sacrifice the reconstruction quality of local objects. These methods typically rely on sparse point cloud representations or low-resolution scan data, which can lead to the smoothing or loss of geometric details. Crucially, due to insufficient sampling or occlusions, local objects in large-scale reconstructions often contain point cloud holes or incomplete geometric structures. These geometric discontinuities render the reconstruction results unsuitable for tasks that demand precise dimensions or complete topological information, such as industrial inspection or robotic grasping.

The development of point cloud completion has provided solutions to this problem. Traditional methods, such as Poisson Surface Reconstruction, relied on established geometric priors to fill missing parts. With the rise of deep learning, completion was revolutionized. Pioneering works like PointNet [10,11] and subsequent encoder–decoder architectures like PCN (Point Completion Network) [12] achieved effective completion by learning features from point sets. For large scenes, methods like ScanComplete [13] integrated semantic and geometric information to fill missing parts.

Despite these advancements, the prevailing algorithmic paradigms in deep learning completion fundamentally rely on speculative inference to infill geometric voids. These models are generative, learning statistical geometric priors from large-scale (often synthetic) datasets to predict and generate geometry for unobserved regions.

The core issue lies in the lack of strong physical realism constraints within this paradigm. The unconstrained generation process is highly sensitive to geometric biases, potentially fabricating structures that appear geometrically plausible but do not physically exist. This problem has been demonstrated in various practical applications. For instance, Mezghanni et al. (2021) [14] pointed out that traditional deep learning generative models often produce “hallucinated” structures when processing complex objects. These geometries may comply with visual geometric constraints but fail to adhere to physical realities, such as structural stability or connectivity, indicating that the generative models do not effectively incorporate physical realism. Furthermore, Li et al. (2025) [15] highlighted a critical failure point in existing self-supervised methods, particularly when applied to real-world, large-scale scenarios such as archeological site restoration. They demonstrated that these methods often fail to achieve high-fidelity completion for large objects with significant missing surfaces and unbalanced point distribution, necessitating specific mechanisms—such as their proposed multi-center-of-projection (MCOP) representation and consistency losses—to ensure the regularity and consistency of the generated geometry.

The prevalence of these issues indicates that generative models lacking physical constraints are unable to provide sufficient geometric accuracy and reliability, particularly in scenarios demanding high fidelity and reliability. As a result, the applicability of these techniques is significantly limited in such contexts.

Furthermore, the standard supervised evaluation approach measures performance by calculating geometric distances (e.g., Chamfer Distance) between a generated model and a “Ground Truth” reference. However, the “Ground Truth” itself is often flawed due to errors inherent in the 3D reconstruction process. For example, Aiger et al. (2008) [16] and Fan et al. (2017) [17] demonstrated that reconstructed models may contain noise, misalignment, or incomplete data, which can distort the reference model. As a result, these evaluations primarily measure how well a generated model replicates the geometric features of the reference, but they do not assess the physical accuracy of the generated geometry. This limitation is particularly critical in applications requiring high physical fidelity, as the “Ground Truth” may not fully reflect real-world conditions.

In summary, within the modern fields of computer vision and 3-D reconstruction, modeling specific local objects embedded in large-scale scenes remains a formidable chal-

lenge. Existing reconstruction algorithms often fail to capture these local details, yielding results that lack geometric completeness and physical realism. To address this issue, we present a novel data-driven framework that couples high-precision registration and fusion techniques with an optimized Neural Radiance Field (NeRF) reconstruction pipeline. Our approach enhances fine-scale object geometry while guaranteeing compatibility with the overall large-scale scene structure, offering an effective solution for high-fidelity point-cloud completion. The proposed workflow is illustrated in Figure 1.

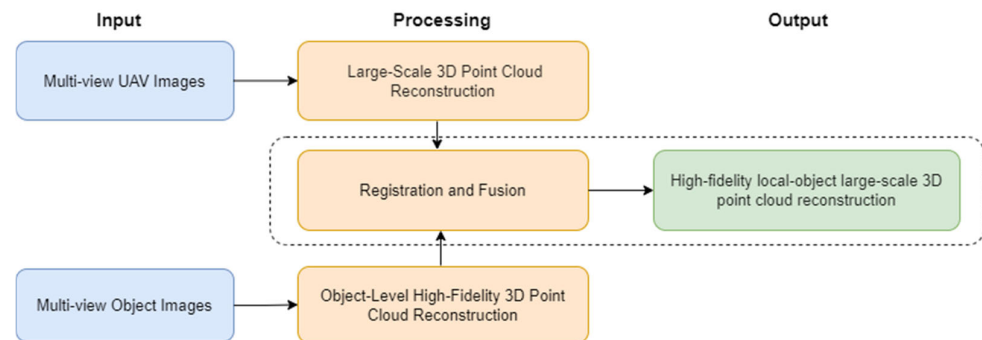


Figure 1. The workflow of this study.

Our core methodology was built on high-precision heterogeneous data registration and seamless fusion. Specifically, we first leveraged an optimized Neural Radiance Fields (NeRF) framework to reconstruct a standalone, object-level point cloud with high-fidelity appearance and geometry from multi-view 2D images. NeRF was chosen for its unparalleled ability to capture complex geometry and intricate photometric details of real-world objects. Subsequently, this high-precision point cloud was aligned with the target large-scale scene point cloud, which is in a global coordinate system but has geometric deficiencies. This registration process must resolve significant discrepancies between the two datasets in terms of coordinate systems, scale, rotation, and point density. Conceptually, the key innovation is the introduction of high-precision, physically-grounded observation data as a “strong prior” for geometric completion. Unlike conventional methods that rely on statistical averages as weak priors, our approach employs a high-fidelity, instance-specific model, derived from real physical observations, to serve as a strong prior. This distinction repositions the completion task from a speculative inferential process to an evidence-based data fusion process using authentic, precise geometric information, thereby pioneering a distinct technical trajectory in point cloud completion. Aligned with this methodology, we used a suite of unsupervised evaluation metrics to assess the intrinsic geometric quality of the final results, ensuring that the evaluation also focused on physical realism. All in all, this work provides an innovative, robust, and evidence-based solution for completing local objects within large-scale scenes.

2. Related Works

Point cloud completion is a key technology for solving the problem of incomplete 3D data, and its development is closely tied to the progress of 3D perception technologies. Before the rise of deep learning, point cloud completion primarily relied on traditional geometric methods. The core idea of this period was to use a point cloud’s own geometric properties and mathematical models to fill in missing parts through rigorous derivation. Researchers proposed a series of classic algorithms aimed at reconstructing a continuous geometric surface from incomplete data. For example, Poisson Surface Reconstruction, proposed by Kazhdan, Bolitho, and Hoppe in 2006 [18], became a landmark in this field. This method cleverly treats a point cloud with normals as a vector field, and by solving

the Poisson equation, it reconstructs an implicit surface that best matches this vector field, thus naturally and smoothly filling large holes in the point cloud. Additionally, Hoppe and his colleagues' 1992 [19] concept of "Point Set Surfaces", which uses local quadric surface fitting to infer surface normals and geometric structures, is a concept often used for local point cloud interpolation and completion. Other methods are borrowed from mathematical morphology in image processing, such as dilation and erosion, to fill small gaps and discontinuities in point clouds. While these traditional methods demonstrated some effectiveness in handling simple, locally smooth holes, their inherent limitation was a high dependence on established geometric priors, making them difficult to apply to complex, irregular, or topologically diverse holes.

With the rise of deep learning in computer vision, point cloud completion saw a revolutionary breakthrough. Researchers began to treat point clouds as a special data format that could be processed directly by neural networks and designed new network architectures for the completion task. The pioneering work of this era was PointNet, proposed by Charles R. Qi et al. in 2017 [10]. It was the first to show that a neural network could directly process an unordered set of points, and by learning the independent features of each point and aggregating global information with a max-pooling layer, it laid the foundation for deep learning on point clouds. Based on this, the point cloud completion task was redefined as an encoder–decoder architecture. Networks like PCN (Point Completion Network) [12], for instance, use an encoder to extract high-level semantic features from an incomplete point cloud and then use a decoder to generate a complete point cloud from these features. By training on large-scale datasets such as ShapeNet [20], these data-driven methods can learn and generalize common geometric patterns of objects, thereby recovering missing structures to a certain extent. In recent years, the development trend of point cloud completion technology has become more diverse and in-depth. In terms of model architecture, researchers have begun to explore more refined feature learning and generation methods, such as introducing attention mechanisms to capture the global context information of a point cloud or using Generative Adversarial Networks (GANs) [21] to produce more realistic and detailed completion results. On the application front, the focus has also expanded from simple object completion to large-scale scene point cloud completion. Methods like ScanComplete [13] have started to integrate semantic information into the completion process, allowing algorithms to better understand the scene context and effectively fill in missing parts of large-scale environments.

Despite the significant advancements brought to the field of point cloud completion by the aforementioned pioneering works (e.g., PointNet, PCN, ScanComplete), the core mechanism of the prevailing algorithmic paradigms still falls within the scope of speculative inference. These methods are fundamentally generative models that are trained on (typically large-scale and synthetic) datasets to learn a comprehensive set of statistical geometric priors concerning object shapes. When presented with an incomplete input, the algorithm leverages these learned priors to "speculate" and "infill" the geometric structures it deems most probable.

The very nature of this pattern-learning approach dictates that while the completion results may be macroscopically similar to the training data, they often lack real-world constraints for the specific physical instance. Since the model generates a "statistically likely" shape rather than a "physically necessary" one, the geometric accuracy and physical realism of the output cannot be guaranteed. This deficiency makes the algorithm highly susceptible to hallucinating structures that are geometrically plausible yet factually non-existent in the real world. This inherent limitation becomes particularly salient when processing objects with complex topologies, atypical geometric morphologies, or those that deviate significantly from the training distribution. In such cases, the model's "speculations"

often fail, severely restricting the practical utility of these completion techniques in critical applications that demand high fidelity and reliability, such as industrial quality inspection, cultural heritage digitization, and robotic interaction.

3. Materials and Methods

3.1. Large-Scale 3D Reconstruction Using COLMAP

For our large-scale point cloud reconstruction, we employed the classic COLMAP pipeline, a general-purpose Structure from Motion [22–24] and Multi-View Stereo [1,25,26] framework designed to automatically generate high-precision 3D models from unordered images. The reconstruction proceeds in two main stages: Sparse Point Cloud generation (SfM) and Dense Point Cloud generation (MVS).

The reconstruction process is executed in two primary stages:

- **SfM Stage:** This stage recovers the camera poses and the sparse 3D structure of the scene. The core optimization is achieved through Bundle Adjustment, which jointly refines all camera poses and 3D point coordinates by minimizing the total reprojection error across all views.
- **MVS Stage:** Following the accurate estimation of camera poses by SfM, the MVS stage generates a dense point cloud by computing and fusing depth maps from multiple neighboring views.

3.2. High-Fidelity Object Reconstruction with an Enhanced NeRF Model

NeRF (Neural Radiance Fields) represents a static scene as a continuous 5D function, where a 3D spatial location (x, y, z) and a 2D viewing direction (θ, ϕ) are mapped to a volume density and a view-dependent emitted color (radiance). The color of any given pixel is synthesized by integrating this volume density and color along the corresponding camera ray. This entire process, known as volume rendering, is fully differentiable, enabling novel view synthesis from arbitrary viewpoints. At its core, NeRF uses a Multi-Layer Perceptron (MLP) [8,27] to implicitly encode the scene's geometry and appearance. Requiring only a sparse set of input images with known camera poses, the network is trained in a self-supervised manner by minimizing the photometric error between the rendered images and the ground-truth images. This optimization process ultimately recovers a highly detailed and multi-view consistent 3D representation of the scene.

Unlike explicit meshes, NeRF encodes a continuous, differentiable density field over the entire 3-D volume, allowing gradients to flow through empty space and enabling sub-voxel accuracy during optimization. This volumetric continuity guarantees that the surface inferred from any ray-bundle is view-consistent by construction: the same 3-D coordinate always returns the same density, so silhouettes and depth maps rendered from arbitrary viewpoints automatically satisfy multi-view coherency without extra post-processing. Consequently, the object-level point cloud extracted by marching along optimal depth rays inherits this cross-view agreement, providing a physically-plausible prior that can be safely fused into the large-scale scene.

To achieve high-fidelity, object-level reconstruction, we propose a geometric-enhanced Neural Radiance Fields (NeRF) model. Unlike a standard NeRF that learns a scene implicitly from scratch, our method deeply fuses external geometric and regional constraints as powerful guiding signals. Specifically, we first jointly estimated initial depth maps and camera poses through a self-supervised learning framework while simultaneously generating precise object masks using the SAMv2 model. During training, these depth maps provide the model with direct geometric constraints, while the object masks strictly confine the loss calculation to the target region, effectively eliminating background interference. Building on this foundation, the model internally employs advanced techniques

such as multi-resolution hash encoding and spherical harmonics to efficiently capture fine geometric details and complex appearance effects. Through this combination of multi-level guidance and an efficient architecture, our model ultimately generates high-fidelity 3D objects that possess both geometric accuracy and appearance realism, laying a solid foundation for future high-level interactive editing and world model construction. This method differs from the standard NeRF implementation in the following key aspects:

- Incorporated depth maps as a geometric prior. In a standard NeRF, geometry is inferred through photometric consistency across multiple views, with the volume density (σ) integrated along each ray and the network optimized by minimizing photometric loss. However, this indirect approach can lead to scale ambiguity and local minima. In contrast, our method uses a self-supervised depth network to generate dense, metric-scale depth maps for each frame. Through a four-stage iterative pipeline, the depth rendered by the NeRF is used to refine the depth network, creating a feedback loop between depth estimation and NeRF optimization. This enables simultaneous convergence of absolute scale, relative geometry, and fine details, effectively eliminating the “floater” artifacts and geometric drift typical in standard NeRF.
- Multi-resolution Hash Encoding. Multi-resolution hash encoding efficiently represents 3D space using a multi-level feature grid, making it ideal for NeRF rendering. This approach captures both the overall shape (with low-resolution grids) and fine surface details (with high-resolution grids). Our method enhances this by adaptively optimizing the hash encoding using depth maps and object masks during training. In regions with sharp object edges or depth changes, high-resolution features are updated more strongly to capture precise boundaries. In contrast, low-resolution features maintain overall shape stability, preventing overfitting in areas with sparse views. This “local detail, global stability” approach works in synergy with depth and mask constraints, ensuring a unified segmentation boundary, depth, and NeRF density surface, leading to efficient, high-fidelity 3D object reconstruction.
- The model is supervised by a joint loss function with three components. A standard NeRF relies solely on pixel color for supervision, without explicit constraints for object boundaries or geometry. Our method introduces a composite loss function with three components. The first is an absolute depth loss, using the depth maps as supervision to constrain NeRF’s learning of object geometry. The second is a relative depth loss, which enforces correct depth ordering by sampling pixel pairs within the same instance region that are sufficiently separated in image space. The third component is an instance mask loss, which uses a binary mask from SAMv2 to isolate the object from the background, ensuring that the loss is computed only for the target region and sharpening its boundary. These three loss components provide strong supervision with explicit geometric semantics, transforming the weak RGB-only supervision into a robust model that aligns the reconstructed surface with the true edges at a sub-pixel level.
- We leveraged the Segment Anything v2 (SAMv2) model to obtain accurate object masks for NeRF reconstruction. Instead of treating SAMv2 as a static pre-processor, we executed its frozen image-encoder and mask-decoder once per frame before NeRF optimization began, producing binary masks that are kept fixed throughout training. These masks are then used to pixel-wise weight the photometric, depth, and density losses, so gradients flow only into the NeRF MLP and its hash encoding while SAMv2 parameters remain untouched. This lightweight strategy keeps the entire pipeline fully differentiable, yet enables the NeRF to focus on the object region and achieves precise alignment between the mask boundary and the reconstructed surface.

3.3. Evaluation Metrics

To rigorously assess the performance of our completion framework, we implemented a comprehensive evaluation strategy organized into two principal domains: Registration Quality Metrics and Completion Quality Metrics. The former quantifies the alignment accuracy between the large-scale point cloud and the NeRF-reconstructed object model, while the latter evaluates the intrinsic geometric fidelity of the final fused model.

3.3.1. Registration Quality Metrics

To evaluate the quality of the alignment between the large-scale point cloud and the NeRF-reconstructed point cloud, we utilized three key metrics: Fitness, Inlier RMSE, and Normal Consistency.

- Fitness measures how well the source point cloud aligns with the target point cloud after applying a transformation. It calculates the ratio of inlier correspondences (points that align well) to the total number of points in the target cloud. A higher Fitness score indicates better alignment.
- RMSE calculates the root mean square error of the Euclidean distances between inlier points in the two clouds. It quantifies the accuracy of the alignment, where a lower RMSE value signifies better precision in matching the point clouds.
- Normal Consistency evaluates the smoothness and coherence of the surface by examining the alignment of normal vectors (directional vectors perpendicular to the surface) of neighboring points. A higher score reflects a smoother and more geometrically consistent surface, with fewer artifacts.

Together, these metrics help assess both the alignment quality and intrinsic geometric properties of the point cloud, ensuring that the reconstructed model is accurate and geometrically coherent.

3.3.2. Completion Quality Metrics

This study incorporated a set of unsupervised evaluation metrics to measure the intrinsic geometric quality of the completed point cloud, with a strong focus on physical realism. These metrics—Geometric Smoothness, Distribution Uniformity, and Structural Plausibility—are essential for ensuring that the generated 3D objects exhibit both fine geometric details and an accurate physical appearance.

Geometric Smoothness assesses the local surface smoothness, ensuring that the completed geometry transitions naturally and seamlessly with the surrounding environment. This metric helps avoid artifacts such as jagged edges or abrupt transitions, which could detract from the physical realism of the model. Distribution Uniformity evaluates how evenly the points are distributed within the completed point cloud, preventing issues such as clustering or voids that may negatively impact downstream tasks like meshing and feature extraction. Finally, Structural Plausibility checks whether the completed geometry aligns with expected structural patterns, helping to ensure that the object maintains its intended shape and is geometrically consistent.

- Geometric Smoothness is designed to quantify the local geometric fidelity of the surface constituted by the new point set generated by the completion algorithm relative to the original point cloud. An ideal completion algorithm ought to generate smooth and natural geometry that transitions seamlessly with the adjacent original surface, rather than introducing noisy or jagged surface artifacts. Consequently, this metric measures the roughness of the completed surface, serving as a key proxy for its physical realism. We employ Surface Variation as the core computational method, a technique based on the covariance analysis of local neighborhoods. For each point within the completed

region, we construct the covariance matrix from the set of points within its spherical neighborhood of radius r . The eigenvalues of this matrix describe the variance of the neighborhood's point distribution along the three principal directions. Among these, the smallest eigenvalue, λ_0 , corresponds to the variance along the normal direction of the point's local tangent plane. The Surface Variation is thus defined as the normalized smallest eigenvalue, formulated as follows:

$$\sigma(p_i) = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} \quad (1)$$

The value of the Surface Variation metric ranges from 0 to 1/3. A lower value signifies that the point's local neighborhood is highly coplanar, thereby indicating that the completed surface is smoother and more physically plausible.

- Distribution uniformity is used to assess the quality of the spatial distribution of a completed point cloud. The resultant point set from a superior completion algorithm should be uniformly distributed to avoid the formation of unnatural point clusters or sparse voids in localized regions, as such distributional artifacts can adversely affect subsequent downstream tasks like meshing and feature extraction. To quantify the distribution uniformity of the point cloud, we compute the Local Density Variance (LDV) metric and further derives the Normalized Uniformity Error (*NUE*), thereby eliminating scene-scale and sampling-resolution biases and enabling comparable assessment across different objects and datasets.

$$\sigma_\rho^2 = \frac{1}{N} \sum_{i=1}^N (\rho_i - \bar{\rho})^2, \quad \rho_i = \frac{k}{\pi r_{i,k}^2} \quad (2)$$

$$|NUE| = \frac{\sigma_{incomplete}}{\sigma_{complete}} \quad (3)$$

In the formula, $\rho_i = \frac{k}{\pi r_{i,k}^2}$ treats the disk enclosed by the k -th nearest neighbour of point i as the local neighbourhood, defining the density at that point as the ratio of the fixed point count k to the disk area; the smaller the radius $r_{i,k}$, the more compact the area and the higher the density. The variance of all these local densities is then computed as σ_ρ^2 , yielding the density fluctuation across the entire point cloud. A smaller σ_ρ^2 indicates smaller density differences among neighbourhoods and thus a more uniform spatial distribution.

We adopt the Normalized Uniformity Error (*NUE*) as the final metric. Defined as the ratio of the local-density variance of the completed region, $\sigma_{complete}$, to that of the original large-scale scene, $\sigma_{incomplete}$, this dimensionless value directly reflects how consistently the completed point cloud fluctuates in local density compared with its surroundings. Empirically, $NUE \leq 1.0$ indicates that the completed area is more uniform than the scene; $1.0 < NUE \leq 1.3$ is perceived as a seamless transition; and $NUE > 2.0$ signals noticeable density stripes, calling for parameter retuning or additional sampling. Through this normalization equation, quantitative uniformity comparison across different objects and datasets is achieved.

- To assess the Structural Plausibility of the completed geometry, we introduce the Primitive Fitting Score metric. Methodologically, this is achieved by iteratively applying a robust estimation algorithm, such as RANSAC (Random Sample Consensus), to the set of newly generated points, P_{new} , to segment instances of multiple geometric primitives. We define $P_{primitive}$ as the union of all points identified as inliers to any of the successfully detected primitives, where $P_{primitive} \subseteq P_{new}$. The Primitive Fitting Score is then defined as the ratio of the cardinality of this inlier set to the total number of newly generated points. Its core formula is:

$$\text{Primitive Fitting Score} = \frac{|P_{\text{primitive}}|}{|P_{\text{new}}|} \quad (4)$$

Consequently, a higher score (ranging from 0 to 1, with values closer to 1.0 being better) provides strong evidence that the completion is dominated by regular, interpretable geometric structures. This indicates that the algorithm has successfully restored the object's expected shape with high structural fidelity.

Physical-realism-driven completion demands metrics that go beyond “geometric closeness” to reveal whether the infilled shape is mechanically valid and structurally explainable. Chamfer Distance, by design, only measures the average point-to-point displacement against a potentially flawed “ground-truth” scan; it therefore rewards trivial point-copying and penalizes any departure from the reference, even when the reference itself is incomplete or noisy. In contrast, the proposed unsupervised triplet—Geometric Smoothness, Distribution Uniformity, and Structural Plausibility—directly encodes first-order physical priors (continuity, isotropic sampling, piecewise planarity) that are invariant to absolute coordinate errors and that have been shown to correlate better with human judgements of “realistic shape” in recent metrology and archeological studies [14–16].

4. Experiments

This section presents a comprehensive experimental validation and performance evaluation of our proposed point cloud completion framework. The core objective of these experiments was to systematically assess the efficacy of the framework in fusing heterogeneous 3D data, repairing incomplete geometry and enhancing local detail fidelity through a series of quantitative and qualitative analyses.

4.1. Datasets

To address the notable gap in publicly available benchmarks for 3D reconstruction, particularly the scarcity of datasets that concurrently provide both expansive outdoor scenes and high-resolution, object-centric imagery from an unmanned aerial vehicle (UAV) perspective, we introduced UAV-Recon. This novel, large-scale dataset was acquired utilizing a professional-grade DJI Matrice 300 RTK UAV platform, which is manufactured by DJI (Shenzhen, China), ensuring high-precision geospatial metadata for all captures. UAV-Recon is specifically designed to fill the aforementioned data void, featuring a diverse range of environments that span from structured urban landscapes to unstructured natural terrains. In total, the dataset comprises over 10,000 high-resolution (4 K) RGB images, providing the rich visual information necessary for developing and evaluating high-fidelity 3D reconstruction algorithms.

4.2. Detailed Experimental Setup

The experimental software requirements and environment configuration of this paper are shown in Table 1.

This study proposes a point cloud completion framework integrating registration and fusion techniques, designed to complete a local object's point cloud within a large-scale scene by leveraging a high-fidelity, object-level NeRF model. After separately performing large-scale scene reconstruction and object-level NeRF modeling, the framework accurately registers the two point clouds, which originate from disparate sources and exhibit significant discrepancies in their coordinate systems, scale, and background noise. Through a subsequent series of processes including data fusion, the incomplete target point cloud is efficiently completed using the high-fidelity NeRF reconstruction as a reference. The entire pipeline is divided into the following three core stages.

Table 1. Software and environment configuration.

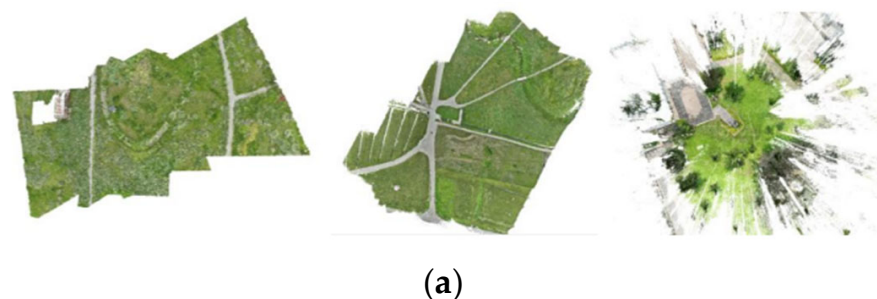
Item	Configuration
Interpreter	Python 3.12.2
Dependencies	NumPy 2.2.5, open3d 0.19.0, OpenCV-python 4.12.0.88, torch 2.12
Operating System	Ubuntu 20.04 LTS
Training Platform	PyTorch 2.4.0
Hardware Configuration	GPU NVIDIA RTX 4090
Number of Training Epochs	60
Optimizer	Adam
Data Processing Batch Size	24
Learning Rate Schedule	WarmUp for the first 2 epochs, followed by cosine annealing, maximum learning rate 0.001

4.2.1. Stage 1: Large-Scale Scene Reconstruction Using COLMAP

In our experiments, we employed COLMAP for large-scale scene reconstruction. The process was initiated with our collected UAV dataset as input, after which we leveraged GPU acceleration for feature extraction and matching to establish robust two-dimensional correspondences between multiple views. For feature matching, we set the nearest-neighbor ratio test threshold to 0.7 and the maximum L2 descriptor distance to 0.7, with both mutual best match consistency checks and guided re-matching enabled.

Following the matching stage, an incremental sparse reconstruction was performed, which iteratively recovers camera poses and the sparse 3D structure of the scene from these correspondences. Throughout this process, Bundle Adjustment (BA) was repeatedly executed to jointly optimize all camera parameters and 3D point locations, thereby minimizing reprojection error and ensuring global geometric consistency. For this optimization, the maximum number of BA iterations was set to 100, with 200 iterations for the linear solver, while the convergence and gradient tolerance thresholds were set to 0.

After obtaining a globally optimized sparse model with precise intrinsic and extrinsic camera parameters, we proceeded with the dense reconstruction module. This module computes depth maps using a patch-match stereo algorithm—with the maximum image size set to 2000 pixels and the number of views loaded to the GPU simultaneously set to 8—which are then fused into a unified dense point cloud. This pipeline yields a high-quality large-scale reconstruction result, which served as the baseline for subsequent performance comparisons. The reconstruction results of the large-scale scene are illustrated in Figure 2.

**Figure 2.** Cont.

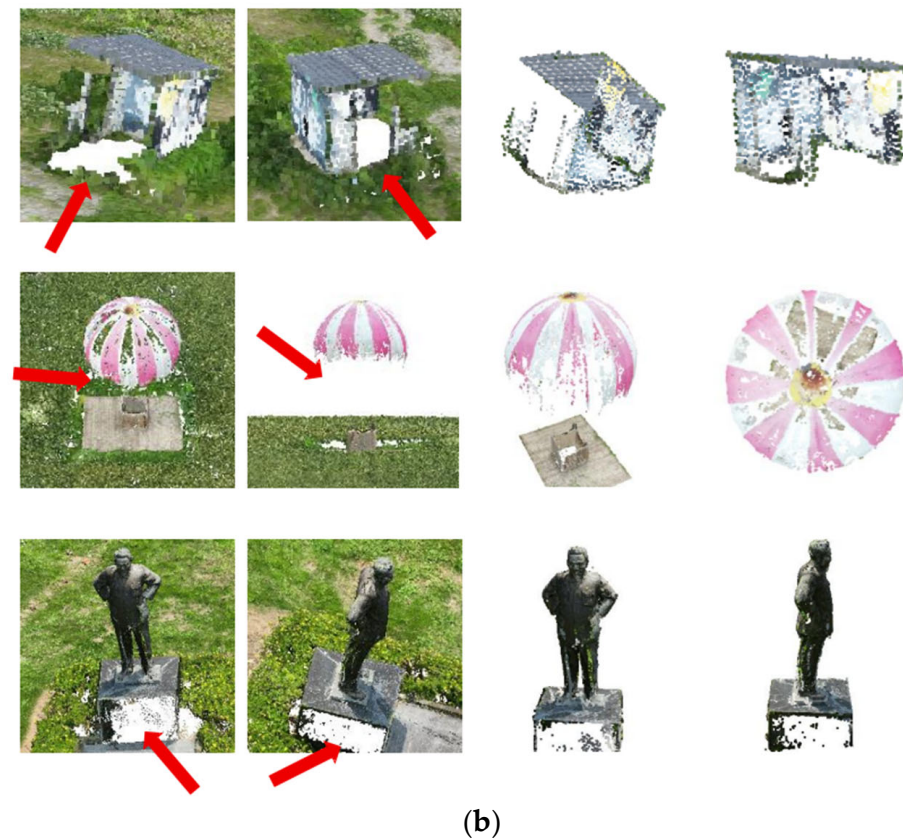


Figure 2. Baseline large-scale 3D reconstruction results using COLMAP. (a) Presents two complete models of the reconstructed large-scale scenes: the left and center images belong to Scene 1, while the right image is from Scene 2. All views cover broad environmental features of the entire region (fields, roads, etc.) and the objects of interest (house, balloon and sculpture). (b) Shows multi-view renderings of the local objects—house, balloon and sculpture—extracted from the large-scale models (house and balloon from Scene 1, sculpture from Scene 2). The images on the left show the reconstruction results of local objects in the complete scene, with the red arrow pointing to the part with obvious geometric defects. The images on the right crop out their 3D models for a more detailed observation. The significant structural incompleteness and surface holes of these local objects are clearly observable, demonstrating the limitations of large-scale reconstruction for fine-grained object details.

4.2.2. Stage 2: Object-Level Reconstruction with an Enhanced NeRF Model

To achieve high-fidelity object-level reconstruction, we designed and implemented a four-stage iterative optimization training scheme. This scheme establishes a mutually reinforcing loop by cyclically optimizing depth estimation, camera pose, and the NeRF model. The entire training pipeline was implemented in the PyTorch framework and iterated for a total of 60 epochs, utilizing the Adam optimizer. The learning rate was managed by a schedule that incorporates a WarmUp phase for the first two epochs, followed by cosine annealing, with a maximum learning rate of 0.001. A data processing batch size of 24 was used throughout the training.

Specifically, the first stage consists of a preliminary self-supervised depth and pose pre-estimation, where the input monocular video sequence undergoes initial training to generate initial depth maps for each frame and the relative camera poses between them, which serve as the initial geometric priors for the subsequent reconstruction. The second stage is a depth-guided initial NeRF training, which utilizes the depth maps and pose information generated in the first stage to train a preliminary object-level NeRF model. The optimization objective for this model is to jointly minimize a composite loss function that includes a photometric loss based on the original RGB images, a depth contrastive loss

that uses the initial depth maps as a supervisory signal to constrain the geometry, and a density loss to accelerate training and reduce background interference. The third stage involves NeRF feedback-based optimization of depth and pose, where the NeRF model trained in Stage 2 is used to render new, more accurate depth maps, which are then fused with the results from the first stage. This more reliable depth information is in turn used to optimize the depth and pose estimation networks, thereby yielding stronger geometric constraints. The fourth stage is the final object-level NeRF optimization, where the refined, high-reliability depth and pose information from Stage 3 is used to retrain the NeRF model. This stage also employs a joint optimization of photometric, depth contrastive, and density losses, but the stronger geometric constraints ensure that the model can generate a high-quality object-level 3D representation with sharper boundaries and a more accurate shape. This mutually reinforcing iterative strategy effectively overcomes the inherent limitations of monocular input, ultimately producing a controllable 3D object model that possesses both fine geometric details and an accurate appearance. The high-fidelity object-level reconstruction results are illustrated in Figure 3.



Figure 3. This figure illustrates the high-fidelity object reconstruction results produced by our proposed four-stage enhanced NeRF pipeline. Multi-view renderings of the house (**top row**), balloon (**middle row**) and sculpture (**bottom row**) are presented. These outcomes validate that our method can capture fine geometric structures and accurate surface textures from monocular video input, providing a crucial foundation for the subsequent point cloud completion stage.

4.2.3. Stage 3: Point Cloud Completion via Registration and Fusion

The primary challenge is to resolve the discrepancies between two 3D point clouds generated from different reconstruction sources to enable their precise registration and subsequent analysis. The point cloud data used in this study originated from two distinct reconstruction methods: a point cloud from a large-scale scene reconstruction and an object-level point cloud reconstructed via a Neural Radiance Field (NeRF). The object-level NeRF reconstruction contains only the target object but exists in an arbitrary coordinate system and scale that is inconsistent with the real-world. Conversely, the large-scale scene reconstruction includes the target object amidst significant background clutter, operates in a different coordinate frame from the NeRF model, and exhibits a substantial scale disparity.

These fundamental differences preclude any direct registration, fusion, or comparison between the two point clouds. The workflow of this stage is illustrated in Figure 4.

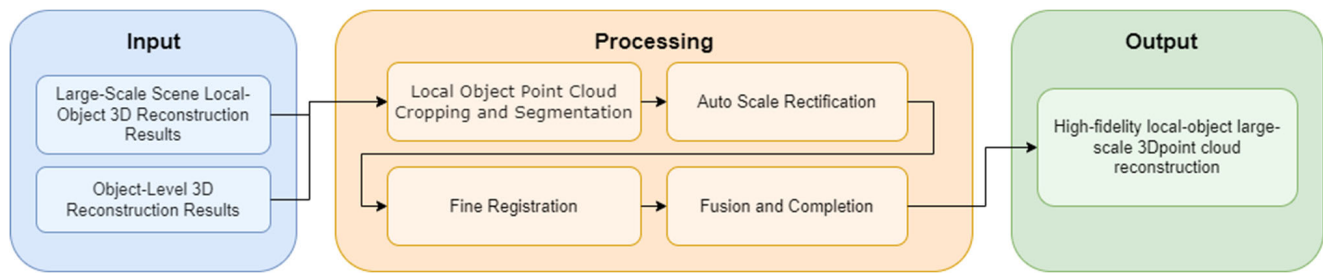


Figure 4. The workflow of registration and fusion.

We first segmented the large-scale point cloud to extract the local object of interest. This is necessary due to challenges such as topological ambiguity when objects are close to the background, difficulty in using simple geometric volumes to represent complex objects, and ineffective filtering for non-planar, non-axis-aligned backgrounds in real-world scans. To address these issues, the operator selects a few fiducial points on the target based on prior knowledge. The system then calculates the minimal Axis-Aligned Bounding Box (AABB) around these points and performs a volumetric crop. This hybrid method ensures accurate segmentation while remaining flexible for complex cases, resulting in a clean point cloud subset for the next fine registration step.

A major challenge in this study was registering two point clouds from different reconstruction methods: a large-scale scene and an object-level NeRF model. The main issue is the significant scale difference between the two datasets. Ignoring this scale gap makes any registration algorithm based on geometric correspondence ineffective. Point-to-point methods like the Iterative Closest Point (ICP) algorithm fail to converge because they cannot identify valid nearest-neighbor pairs across different scales. Even if coarse alignment is possible, the scale conflict prevents precise geometric alignment, resulting in high Root Mean Square Error (RMSE), which is unsuitable for high-fidelity fusion. This problem also affects feature-based methods (e.g., FPFH, RANSAC), since their geometric descriptors are scale-sensitive. As a result, without robust feature correspondences, registration between the point clouds fails.

To eliminate scale drift caused by different reconstruction methods, the algorithm first extracts FPFH (Fast Point Feature Histograms) [28] features from both point clouds and establishes initial correspondences within a RANSAC (Random Sample Consensus) [29–31] framework. Subsequently, by solving a rigid transformation incorporating a scale variable (using Singular Value Decomposition, SVD), a global scale factor is robustly estimated through a ratio-based method, achieving effective scale unification.

Specifically, the scale rectification pipeline begins with centroid normalization to separate translational offsets from rotation and scaling factors. After centering the correspondence sets at the origin, the scale factor s is determined by calculating the ratio of their Root Mean Square (RMS) distances, accurately recovering the scaling parameter for subsequent point cloud normalization. The formula for computing the scale factor s is as follows:

$$s = \sqrt{\frac{\frac{1}{N_o} \sum_{i=1}^{N_o} \|P_i^o - c_o\|^2}{\frac{1}{N_s} \sum_{j=1}^{N_s} \|P_j^s - c_s\|^2}} \quad (5)$$

In the equation, N_o and N_s denote the numbers of points in the source point cloud and the target point cloud, P_i^o and P_j^s are their corresponding 3D coordinates, c_o and c_s are the centroids of the two point clouds, and s is the scale factor to be determined so that the

RMS distances of both sets are equal. After scaling, P_{scaled}^o represents the source point cloud translated to the target centroid and aligned to the same scale as the target point cloud. The scale-unification formula is given as follows.

$$P_{scaled}^o = \{s(P_i^o - c_o) + c_s\}_{i=1}^{N_o} \quad (6)$$

After normalizing both point clouds to a common scale, we achieved high-precision alignment by combining coarse and fine registration. First, 33-dimensional FPFH descriptors were computed on a 4% diagonal voxel-down-sampled copy; RANSAC was then used to establish correspondences with a distance threshold of $1.5 \times \text{voxel}$ and a convergence criterion of 4,000,000 samples or 500 inner loops, yielding the initial rigid transform T_0 . Subsequently, point-to-point Iterative Closest Point (ICP) refined the alignment with a nearest-neighbor distance threshold set to $0.4 \times \text{voxel}$ and a maximum of 35 iterations, ensuring robust convergence on complex object geometries.

From a qualitative perspective, Figure 5 visually summarizes the final output of our complete point-cloud completion framework that integrates registration and fusion. Multi-view renderings of the high-fidelity, object-level 3D models for three targets—house, balloon, and sculpture—are presented after completion and re-integration into the large-scale scene. To emphasize the superiority of our approach, these results were compared with the initial, incomplete geometries extracted from the baseline COLMAP reconstruction. The comparison revealed a fundamental quality improvement: the completed models exhibited excellent structural integrity, successfully filling the large geometric voids caused by view-dependent occlusions or inherent algorithmic limitations. Surface continuity was also substantially enhanced; the prevalent surface defects in the original geometry were effectively repaired, restoring a smooth and detail-rich surface morphology.

This visual evidence provides clear support for the effectiveness of our pipeline. It demonstrates that robust geometric completion can be achieved by precisely aligning and fusing a high-fidelity, standalone, object-level model from NeRF with its geometrically-deficient counterpart from the large-scale model, which resides in the global coordinate system. This method not only compensates for the shortcomings of traditional large-scale reconstruction pipelines in fine-grained modeling but also generates an enhanced large-scale scene model of significantly greater research value.

For a comprehensive quantitative evaluation of our proposed framework, we analyzed both the accuracy of the intermediate registration stage and the intrinsic quality of the final completed point cloud. Table 2 details the performance metrics for the high-precision registration process, where high Fitness and low Inlier RMSE scores validate the precise alignment between the heterogeneous data sources. Following this, Table 3 presents a suite of unsupervised metrics used to assess the final model's quality, evaluating its Geometric Smoothness, Distribution Uniformity, and Structural Plausibility to quantify the physical realism and coherence of the completed geometry.

Table 2. Registration evaluation metrics between the source object and the target object.

Evaluation Metrics	House	Balloon	Sculpture
Fitness (0-1) ↑	0.9302	0.9916	0.9207
Inlier RMSE (0-1) ↓	0.0852	0.0096	0.0348
Normal Consistency (0-1) ↑	0.6023	0.6210	0.7948



Figure 5. This figure presents the final results of our point-cloud completion pipeline achieved through registration and fusion. Multi-view renderings of the three target objects—house (**top**), balloon (**middle**), and sculpture (**bottom**)—are shown; for each object, the upper row displays the model before completion, while the lower row shows the same model after completion. Comparing the two, it is evident that the completed models recovered the geometric information missing in the initial fragments extracted from the large-scale scene, whether the loss was caused by occlusion or by inherent limitations of the reconstruction algorithm. Surface holes were effectively filled, leading to a marked increase in fidelity and overall quality. These results confirm the effectiveness of our registration workflow, which compensates for the well-known inability of large-scale reconstructions to capture fine local detail, and yields models of significantly higher practical and scientific value.

Table 3. Quantitative registration results for target objects.

Evaluation Metrics	House	Balloon	Sculpture
Geometric Smoothness (0-1) ↓	0.1597	0.1897	0.1219
Distribution Uniformity (1.0–1.3) ↓	1.17	1.29	1.21
Structural Plausibility (0-1) ↑	0.9984	0.9954	0.9975

The quantitative results presented in Tables 2 and 3 provide a strong validation for our proposed point cloud completion framework from two key aspects: the precision of the intermediate registration stage and the intrinsic quality of the final fused model. The upward-pointing arrow in the table indicates that the higher the value within the range, the better the effect. The downward-pointing arrow indicates that the lower the value within the range, the better the effect.

Table 2 reports the metrics of the high-precision registration stage. The Fitness scores for House, Balloon, and Sculpture were all close to unity (0.9302, 0.9916, and 0.9207, respectively), confirming that a comprehensive geometric correspondence between the object-level NeRF reconstructions and the large-scale sparse geometry had been successfully established. The low Inlier RMSE values further substantiate this: the balloon error of only 0.0096 m demonstrates that our scale rectification and alignment procedure is accurate and reliable. The Normal Consistency values (0.6023, 0.6210, and 0.7948) indicate a high degree of agreement in surface orientation between the two point clouds, providing a crucial prerequisite for subsequent seamless, artifact-free fusion. Collectively, these metrics validate that the proposed pipeline achieves high-fidelity spatial alignment of the heterogeneous data sources.

Table 3 presents an unsupervised assessment of the intrinsic quality of the completed models. The Structural Plausibility scores for the three objects were 0.9984, 0.9954 and 0.9975—almost unity—demonstrating that the generated geometry was structurally coherent and followed expected priors such as piecewise planarity, rather than the unstructured or noisy infill typical of purely inferential methods. Distribution Uniformity, measured by NUE, was close to 1 for all cases (1.17, 1.29 and 1.21), indicating highly regular point sampling on the continuous surfaces. Finally, the low Geometric Smoothness values confirm that the completed surfaces were virtually free of high-frequency noise and artifacts.

In summary, the effectiveness and superiority of our proposed point-cloud completion framework were thoroughly cross-validated by both qualitative visual evidence and quantitative metrics. Qualitatively, the multi-view renderings in Figure 5 show a clear quality improvement over the initial COLMAP models; structural voids and widespread surface defects were successfully repaired, restoring a smooth, detailed, and geometrically complete morphology for the house and balloon. This visual assessment is strongly supported by the quantitative analyses in Tables 2 and 3. The registration metrics in Table 2 confirm the reliability of our alignment pipeline: exceptionally high Fitness scores (0.9302 and 0.9916) and extremely low Inlier RMSE values demonstrate sub-centimeter, full-coverage matching between heterogeneous data sources. Moreover, the unsupervised metrics in Table 3 validate the intrinsic quality of the final output: near-perfect Structural Plausibility scores prove that the completion is structurally coherent rather than speculative noise, while excellent Distribution Uniformity and Geometric Smoothness indicate a regularly sampled, artifact-free surface of superior overall quality.

4.3. Ablation Experiments

To further validate the effectiveness of our completion framework, we conducted a series of ablation experiments on the original dataset to evaluate the contribution of each key component. Specifically, we removed (i) the cropping-and-segmentation step that extracts the local object from the large-scale model before registration, and (ii) the scale-rectification step that unifies the scales of the two point clouds, and then compared the respective outcomes of the registration stage. The results are illustrated in Table 4.

Table 4 reports the registration metrics for the three objects (House, Balloon, and Sculpture) under the above conditions. The upward-pointing arrow in the table indicates that the higher the value within the range, the better the effect. The downward-pointing

arrow indicates that the lower the value within the range, the better the effect. It can be seen that both ablated settings suffered a dramatic drop in Fitness (maximum 0.5876, average < 0.4), while the Inlier RMSE was generally higher than 0.18 m and Normal Consistency was below 0.6. In particular, when scale rectification was discarded, the Fitness of Balloon plunged to 0.0820 and the RMSE approached 0.18 m, indicating that cross-scale correspondences could hardly be established and even a coarse registration failed.

Table 4. Registration evaluation metrics between the source object and the target object (ablation experiments).

Evaluation Metrics	Object	Without Scale Rectification	Without Cropping or Segmentation
Fitness (0-1) ↑	House	0.3083	0.4129
	Balloon	0.0820	0.5876
	Sculpture	0.3085	0.3621
Inlier RMSE (0-1) ↓	House	0.2614	0.4424
	Balloon	0.1810	0.1804
	Sculpture	0.3724	0.0655
Normal Consistency (0-1) ↑	House	0.4448	0.3056
	Balloon	0.2196	0.5695
	Sculpture	0.4834	0.7373

Under either of these two conditions, the subsequent fusion step becomes invalid and meaningless, and no usable completed model can be generated. These results further demonstrate the necessity of each module in our pipeline and confirm that the proposed design significantly boosts the overall completion performance.

5. Discussion

This study addresses the inherent difficulties in large-scale point cloud processing by proposing an interactive algorithm for point cloud completion based on high-precision, object-level 3D reconstruction. The core idea is to precisely register and seamlessly fuse a fine-grained object model—generated by an optimized Neural Radiance Fields (NeRF) framework—with a large-scale scene point cloud in a global coordinate system. Our contributions are summarized as follows:

- In this work, we propose a novel algorithm that leverages high-fidelity, object-level reconstructions to markedly boost the accuracy of local objects within large-scale scenes. The core idea is to introduce a completion algorithm founded on high-fidelity data rather than on purely algorithmic inference. Specifically, an improved Neural Radiance Fields (NeRF) model was employed to generate metrically accurate, physically plausible data that were subsequently used to complete the missing geometry. This innovation guarantees that the resulting point cloud simultaneously attains superior geometric precision and physical realism, thereby satisfying the stringent accuracy demands of downstream tasks.

From a scientific perspective, our contribution lies in tightly coupling large-scale scene reconstruction with object-level reconstruction, offering the field a new viewpoint and methodology. By fusing high-resolution object data into the global scene, the algorithm preserves fine local details without sacrificing an accurate description of overall geometry. This fusion not only elevates reconstruction accuracy but also enriches the level of detail, laying a solid foundation for understanding and analyzing complex environments.

Moreover, the proposed algorithm exhibits strong generalizability. Relying on high-

fidelity, physically based data as a powerful prior, the completion method performed consistently across diverse scenes and object types. Such cross-domain adaptability opens broad prospects for future applications—ranging from virtual and augmented reality to smart manufacturing—and underscores the long-term research value of our approach.

- **Introducing Complementary Unsupervised Evaluation Metrics:** We critically reviewed the limitations of existing geometry-based evaluation metrics and proposed a suite of complementary, unsupervised metrics (e.g., assessing geometric smoothness, uniformity, and structural plausibility). This provides a more scientific and comprehensive perspective for evaluating the physical realism and functional usability of point cloud completion results.

Nevertheless, we also acknowledge certain limitations in the current framework. The initial segmentation of local objects from the large-scale scene relies on manual cropping, which, while effective, reduces the full automation of the pipeline. Additionally, the quality of the NeRF reconstruction is highly dependent on the quality and quantity of the input multi-view images for the specific object. Future work could focus on automating the segmentation process, for instance, by integrating semantic segmentation or object detection networks. Further research could also explore extending this fusion paradigm to dynamic scenes or incorporating other high-fidelity reconstruction techniques beyond NeRF.

6. Conclusions

This study introduced a novel, data-driven framework for achieving the high-fidelity point cloud completion of local objects within large-scale scenes. We identified the inherent limitations of both standard large-scale reconstruction pipelines, which result in incomplete local geometries, and deep learning-based completion methods, which rely on speculative inference.

Our proposed solution successfully overcomes these challenges by precisely registering and seamlessly fusing a high-fidelity, object-level model generated by an optimized NeRF framework with an incomplete large-scale point cloud. The experimental results demonstrate that our method effectively restores missing geometry, fills surface voids, and generates a coherent final 3D model that is enhanced in both detail and physical realism. Evaluated on our self-constructed UAV-Recon dataset, the proposed method achieved a Structural Plausibility ≥ 0.995 , Geometric Smoothness ≤ 0.19 , and Distribution Uniformity ≈ 1.2 . By leveraging real observation data as a strong geometric prior, and supported by a more comprehensive evaluation methodology, this work provides a robust and well-founded solution to a pressing challenge in 3D reconstruction, offering a higher-value data foundation for downstream applications that require both large-scale context and high-fidelity local detail.

Author Contributions: Conceptualization, Y.F. and L.Y.; methodology, Y.F. and L.Y.; software, Y.F.; validation, Y.F. and T.H.; formal analysis, Y.F.; investigation, K.L. and Y.F.; resources, K.L.; data curation, Y.F.; writing—original draft preparation, Y.F.; writing—review and editing, Y.F., K.L. and L.Y.; visualization, Y.F.; supervision, T.H. and L.Y.; project administration, T.H. and L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are not publicly available due to ongoing use in another study but are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Nistér, D.; Naroditsky, O.; Bergen, J. Visual odometry for ground vehicle applications. *J. Field Robot.* **2006**, *23*, 3–20. [\[CrossRef\]](#)
2. Mur-Artal, R.; Tardós, J.D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262. [\[CrossRef\]](#)
3. Sucar, E.; Liu, S.; Ortiz, J.; Davison, A.J. iMAP: Implicit mapping and positioning in real-time. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 6229–6238. [\[CrossRef\]](#)
4. Bloesch, M.; Czarnowski, J.; Clark, R.; Leutenegger, S.; Davison, A.J. Codeslam—Learning a compact, optimisable representation for dense visual slam. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2560–2568. [\[CrossRef\]](#)
5. Guo, Z.C.; Forbes, J.R.; Barfoot, T.D. Marginalizing and Conditioning Gaussians onto Linear Approximations of Smooth Manifolds with Applications in Robotics. In Proceedings of the 2025 IEEE International Conference on Robotics and Automation (ICRA), Atlanta, GA, USA, 19–23 May 2025; pp. 2606–2612. [\[CrossRef\]](#)
6. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *arXiv* **2020**, arXiv:2003.08934. [\[CrossRef\]](#)
7. Chen, A.; Xu, Z.; Geiger, A.; Yu, J.; Su, H. TensorRF: Tensorial radiance fields. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 333–350. [\[CrossRef\]](#)
8. Fridovich-Keil, S.; Yu, A.; Tancik, M.; Chen, Q.; Recht, B.; Kanazawa, A. Plenoxels: Radiance fields without neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5501–5510. [\[CrossRef\]](#)
9. Yang, Y.; Zhang, S.; Huang, Z.; Zhang, Y.; Tan, M. Cross-ray neural radiance fields for novel-view synthesis from unconstrained image collections. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 15901–15911. [\[CrossRef\]](#)
10. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J.J.I. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [\[CrossRef\]](#)
11. Qian, G.; Li, Y.; Peng, H.; Mai, J.; Hammoud, H.; Elhoseiny, M.; Ghanem, B. Pointnext: Revisiting pointnet++ with improved training and scaling strategies. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 23192–23204. [\[CrossRef\]](#)
12. Yuan, W.; Khot, T.; Held, D.; Mertz, C.; Hebert, M.J.I. PCN: Point Completion Network. In Proceedings of the 2018 International Conference on 3D Vision (3DV), Verona, Italy, 5–8 September 2018. [\[CrossRef\]](#)
13. Dai, A.; Ritchie, D.; Bokeloh, M.; Reed, S.; Sturm, J.; Niener, M. ScanComplete: Large-Scale Scene Completion and Semantic Segmentation for 3D Scans. *arXiv* **2017**, arXiv:1712.10215. [\[CrossRef\]](#)
14. Mezghanni, M.; Boulkenafed, M.; Lieutier, A.; Ovsjanikov, M. Physically-aware generative network for 3d shape modeling. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 9330–9341. [\[CrossRef\]](#)
15. Li, A.; Zimmer-Dauphinee, J.R.; Kalyanam, R.; Lindsay, I.; VanValkenburgh, P.; Wernke, S.; Aliaga, D. Self-Supervised Large Scale Point Cloud Completion for Archaeological Site Restoration. In Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville, TN, USA, 10–17 June 2025; pp. 11759–11768. [\[CrossRef\]](#)
16. Aiger, D.; Mitra, N.J.; Cohen-Or, D. 4-points congruent sets for robust pairwise surface registration. *ACM Trans. Graph. (TOG)* **2008**, *27*, 1–10. [\[CrossRef\]](#)
17. Fan, H.; Su, H.; Guibas, L. A Point Set Generation Network for 3D Object Reconstruction from a Single Image. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
18. Kazhdan, M.; Bolitho, M.; Hoppe, H. Poisson surface reconstruction. In Proceedings of the Fourth Eurographics Symposium on Geometry Processing, Cagliari Sardinia, Italy, 26–28 June 2006. [\[CrossRef\]](#)
19. Hoppe, H.; Deroose, T.; Duchamp, T.; McDonald, J.; Stuetzle, W. Surface reconstruction from unorganized points. *ACM SIGGRAPH Comput. Graph.* **1992**, *26*, 71–78. [\[CrossRef\]](#)
20. Chang, A.X.; Funkhouser, T.; Guibas, L.; Hanrahan, P.; Huang, Q.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; et al. ShapeNet: An Information-Rich 3D Model Repository. *arXiv* **2015**, arXiv:1512.03012. [\[CrossRef\]](#)
21. Achlioptas, P.; Diamanti, O.; Mitliagkas, I.; Guibas, L. Learning Representations and Generative Models for 3D Point Clouds. *arXiv* **2017**, arXiv:1707.02392. [\[CrossRef\]](#)

22. Smith, R.; Self, M.; Cheeseman, P.J.M.I.; Recognition, P. Estimating Uncertain Spatial Relationships in Robotics. *Mach. Intell. Pattern Recognit.* **1988**, *5*, 435–461. [[CrossRef](#)]
23. Cui, H.; Shen, S.; Gao, W.; Wang, Z. Progressive large-scale structure-from-motion with orthogonal MSTs. In Proceedings of the 2018 International Conference on 3D Vision (3DV), Verona, Italy, 5–8 September 2018; pp. 79–88. [[CrossRef](#)]
24. Zhu, S.; Zhang, R.; Zhou, L.; Shen, T.; Fang, T.; Tan, P.; Quan, L. Very large-scale global sfm by distributed motion averaging. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4568–4577. [[CrossRef](#)]
25. Ding, Y.; Zhu, Q.; Liu, X.; Yuan, W.; Zhang, H.; Zhang, C. Kd-mvs: Knowledge distillation based self-supervised learning for multi-view stereo. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; pp. 630–646. [[CrossRef](#)]
26. Zhu, Q.; Min, C.; Wei, Z.; Chen, Y.; Wang, G. Deep learning for multi-view stereo via plane sweep: A survey. *arXiv* **2021**, arXiv:2106.15328. [[CrossRef](#)]
27. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
28. Rusu, R.B.; Blodow, N.; Beetz, M. Fast point feature histograms (FPFH) for 3D registration. In Proceedings of the 2009 IEEE International Conference on Robotics and Automation, Kobe, Japan, 12–17 May 2009; pp. 3212–3217. [[CrossRef](#)]
29. Fischler, M.A.; Bolles, R.C.J.C.o.t.A. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [[CrossRef](#)]
30. Shi, P.; Yan, S.; Xiao, Y.; Liu, X.; Zhang, Y.; Li, J.; Letters, A. RANSAC back to SOTA: A two-stage consensus filtering for real-time 3D registration. *IEEE Robot. Autom. Lett.* **2024**, *9*, 11881–11888. [[CrossRef](#)]
31. Fan, Y.; Zhang, Q.; Tang, Y.; Liu, S.; Han, H. Blitz-SLAM: A semantic SLAM in dynamic environments. *Pattern Recognit.* **2022**, *121*, 108225. [[CrossRef](#)]

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.