

# GARF-CH: Generalizable Flow-Matched 3D Reassembly of Fragmented Cultural Heritage Point Clouds

Emanuele Balloni<sup>1</sup>, Marina Paolanti<sup>2</sup>, Emanuele Frontoni<sup>2</sup>, Stefano Mereu<sup>3</sup>, Ferdinando Cannella<sup>3</sup>, Roberto Pierdicca<sup>1</sup>

<sup>1</sup> Department of Construction, Civil Engineering and Architecture (DICEA), Università Politecnica delle Marche,  
Via Brecce Bianche, 12, 60131, Ancona, Italy

<sup>2</sup> Department of Political Sciences, Communication and International Relations, University of Macerata,  
Via Don Minzoni, 22A, 62100, Macerata, Italy

<sup>3</sup> Industrial Robotics Facility, Italian Institute of Technology, Via Greto di Cornigliano, 6/R, 16152, Genoa, Italy

**Keywords:** Cultural Heritage, 3D Point Clouds, Fracture Reassembly, Restoration, Artificial Intelligence, Deep Learning

## Abstract

Digital restoration of cultural heritage (CH) assets is essential for preserving historical structures fragmented by time, environmental degradation, or catastrophic events. Although learning-based approaches have demonstrated impressive results in 3D fragment reassembly for everyday objects and small artifacts, their application to large-scale CH structures is largely unexplored. To overcome this gap, this study examines the feasibility and reliability of neural point cloud reassembly methods in complex CH scenarios. Building upon the state-of-the-art GARF framework, we introduce GARF-CH, a parameter-efficient adaptation of GARF tailored to architectural-scale heritage data. We have built a novel fractured dataset derived from the ArCH dataset by generating realistic fractures on large-scale CH point clouds. This enables a systematic evaluation of these methods in the CH domain. Extensive quantitative and qualitative experiments on 15 real-world CH structures show that GARF-CH reduces rotation error by  $1.76^\circ$ , translation error by  $0.3 \times 10^{-2}$ , and Chamfer Distance by  $0.71 \times 10^{-3}$  compared to the original GARF model, while maintaining the same part accuracy (83.33%). Furthermore, we assess robustness under missing-fragment conditions and show that the proposed approach maintains coherent reconstructions even with incomplete data. These results highlight the potential and limitations of learning-based reassembly for large-scale CH restoration and establish a foundation for future research on scalable, robust, and autonomous digital reconstruction pipelines for complex CH environments.

## 1. Introduction

Cultural heritage (CH) restoration is essential for preserving the material evidence of human history, identity, and collective memory (Basu et al., 2023; Pietroni and Ferdani, 2021; Li et al., 2025a). Artifacts, monuments, and architectural structures embody irreplaceable historical, artistic, and social value. However, they are threatened by natural disasters, environmental degradation, armed conflicts, and human-induced damage (Li, 2024; Cacciotti et al., 2021). In this context, digitalizing CH assets has become a vital strategy for long-term preservation, documentation, and dissemination (Storeide et al., 2023; Silitina et al., 2024; Balloni et al., 2024). High-fidelity digital representations not only safeguard fragile or partially lost heritage, but also support restoration planning, research, and public accessibility. This ensures that cultural memory can be transmitted across generations, even when physical preservation is no longer feasible. Among digital acquisition methods, 3D point clouds have become the standard for capturing the geometry of CH objects and environments thanks to advances in laser scanning, photogrammetry, and multi-view reconstruction (Sanchez-Aparicio et al., 2023; Solina, 2018; Llabani and Abazaj, 2024; Pierdicca et al., 2020). Point clouds provide accurate, dense, metric descriptions of surfaces, making them ideal for documenting complex shapes and large-scale structures. However, many CH assets are not preserved in their original integrity; collapses, fractures, and missing parts are common, especially in archaeological and architectural contexts. Consequently, digitally restoring and reassembling fragmented point clouds has become a key challenge in computational CH.

Recent advances in computational geometry and artificial in-

telligence, deep learning in particular, have significantly improved point cloud processing and restoration. Neural methods have demonstrated remarkable capabilities in shape completion, fragment alignment, and reassembly by learning geometric priors from data and modeling complex spatial relationships, achieving strong performance on everyday objects and synthetic fracture scenarios (Li et al., 2024; Xu et al., 2024; Liu et al., 2021). This growing interest has also extended to the CH domain, where researchers have begun to explore fragment reassembly for small-scale artifacts, such as pottery, sculptures, and archaeological fragments, typically in controlled or object-centric settings (Li et al., 2025b; Zhang et al., 2018; Villegas-Suarez et al., 2023). Despite these advances, reassembling large-scale CH structures, such as architectural remains, monuments, and complex scenes, remains largely unexplored. Reassembling of this large structural elements and architectural fragments is as critical as restoring individual artifacts, as it can enable a more comprehensive understanding of historical sites and built environments. A primary challenge in this field is the absence of benchmark datasets and methodologies tailored to large-scale CH structures fragment reassembly. Existing datasets and learning-based approaches focus on small-scale artifacts or everyday objects (Sellán et al., 2022; Lamb et al., 2023; Tsemelis et al., 2024; Li et al., 2025b), not capturing the geometric complexity and environmental context that characterize real-world CH scenes. A promising starting point for addressing this gap is the ArCH dataset (Matrone et al., 2020), which provides high-quality point clouds of CH scenes and structures. Although ArCH was not originally designed for fragment reassembly, its richness and diversity make it a valuable resource for exploring the feasibility of learning-based reassembly in CH

contexts.

Building on this foundation, we investigate whether state-of-the-art neural reassembly models can generalize from everyday objects and small artifacts to complex CH structures. To this end, we adopt GARF (Generalizable 3D reAssembly framework for Real-world Fractures) (Li et al., 2025b), a state-of-the-art neural model for 3D fragment reassembly. GARF is designed to handle real-world fracture patterns and generalize across object categories. In this study, we assess the applicability of GARF to CH structures by creating a fractured dataset from ArCH scenes and conducting extensive quantitative and qualitative experiments on 15 unique CH structure scenarios. Additionally, we introduce GARF-CH, a fine-tuned version of the original GARF model with Low-Rank Adaptation (LoRA) (Hu et al., 2022), to assess its ability to generalize to the different geometric and contextual complexity of CH environments. Furthermore, we also investigate the robustness of the model in the presence of missing fragments, a common and often unavoidable situation in CH restoration due to erosion, loss, or incomplete artifacts availability. The goal is to provide further insight into the practical usability of these learning-based reassembly methods in real-world CH scenarios.

The main contributions of this work are as follows:

- The construction of a novel fractured CH dataset based on the ArCH dataset. This dataset is specifically designed to support the evaluation of point cloud reassembly for CH structures
- The implementation and systematic evaluation of the GARF framework on large-scale CH scenes, to provide insight into its strengths and limitations in this domain
- The implementation of a LoRA-based fine-tuning strategy on the proposed dataset. This strategy was accompanied by a comprehensive evaluation and comparison to analyze the generalization and adaptation capabilities for CH structure reassembly
- A robustness analysis under missing-fragment conditions, assessing the model's behavior when a fragment is absent, evaluating its practical applicability in real-world CH contexts.

The rest of this paper is organized as follows: Section 2 reviews the related work on 3D fragment reassembly, with particular emphasis on learning-based approaches and their applications in CH contexts. Section 3 describes the proposed methodology, including the construction of the fractured CH dataset, the development of the GARF-CH framework through LoRA-based fine-tuning, and the experimental protocols adopted for evaluation. Section 4 presents the quantitative and qualitative experimental results, along with an analysis of robustness under missing-fragment conditions. Finally, Section 5 concludes the paper and discusses current limitations and directions for future research.

## 2. Related works

The reassembly of fragmented objects has attracted increasing attention in computer vision and graphics, particularly in CH scenarios where collapses, erosion, and excavation processes

frequently result in incomplete or fractured artifacts and structures. Although dense point clouds and mesh representations are now standard for digitizing CH assets, fragment reassembly remains challenging, as it requires reasoning about fragment correspondences, fracture interfaces, and relative poses under incomplete, noisy, and highly irregular observations (Sellán et al., 2022; Zhang et al., 2018).

Early approaches mainly focused on small-scale artifacts such as pottery or utensils, relying on handcrafted geometric descriptors, fracture surface matching, template guidance, or combinatorial optimization strategies (Zhang et al., 2015, 2018; Liu et al., 2021). While effective in controlled settings, these methods were often sensitive to noise and showed limited robustness and scalability when applied to irregular real-world fracture patterns or complex multi-fragment scenarios. More recent learning-based methods aim to overcome these limitations by learning geometric priors directly from data. Neural approaches for fragment matching and alignment have demonstrated improved robustness compared to purely geometric pipelines, particularly for irregular shapes and partial observations (Villegas-Suarez et al., 2023). Diffusion-based approaches, such as FragmentDiff (Xu et al., 2024), frame reassembly as a stochastic generative process, while other learning-based methods address pose estimation under arbitrary fragment orientations (Xu et al., 2025). Geometry-aware and annotation-free strategies have also been proposed for scenarios where semantic labels are unavailable, including archaeological reassembly settings (Jia et al., 2025). Despite these advances, most existing methods remain object-centric and have not been evaluated on large-scale CH structures, characterized by significant scale variation, acquisition noise, and complex spatial context. Moreover, several CH-oriented systems still rely on interactive or semi-automatic workflows, limiting their applicability in fully autonomous pipelines (Ye et al., 2022). We address these limitations by shifting the focus from small-scale artifacts to the reassembly of large-scale CH structures. Rather than proposing a new reassembly architecture from scratch, we investigate the adaptability of a state-of-the-art neural reassembly framework (Li et al., 2025b) to this underexplored domain. Our approach enables automatic reassembly, systematically evaluates robustness under complete and incomplete conditions, and provides the first benchmark analysis of neural reassembly methods applied to large-scale CH structures.

In this context, benchmark datasets such as Breaking Bad (Sellán et al., 2022), Fantastic Breaks (Lamb et al., 2023), and RePAIR (Tsesmelis et al., 2024) have enabled supervised training and systematic evaluation of reassembly methods by providing paired fractured and complete object representations. However, these benchmarks predominantly focus on everyday objects or small artifacts and typically rely on synthetic or controlled fracture processes. Therefore, there is a lack of benchmarks explicitly designed for fragment reassembly in large-scale CH scenarios. To overcome this, we leverage the ArCH dataset (Matrone et al., 2020), which provides high-quality point clouds of large CH structures, but is not designed to model fragmentation processes or reassembly. Starting from this dataset, we designed and implemented a pipeline for the construction of a dataset for fragment reassembly.

## 3. Materials and methods

This work aims to evaluate the feasibility and reliability of learning-based 3D reassembly methods for large-scale CH

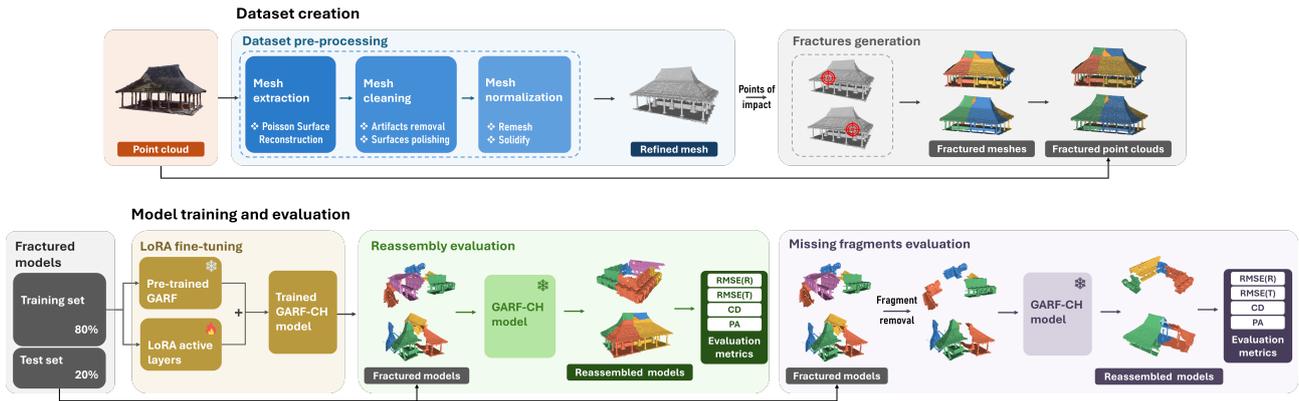


Figure 1. Overview of the proposed pipeline. Starting from ArCH dataset point clouds, we create a novel fracture dataset with both meshes and point clouds. Then, the GARF model is fine-tuned with the novel dataset and evaluated in both complete and incomplete fragmentation scenarios.

structures, with particular focus on generalization to complex scenes and resilience to missing fragments. To this end, we have developed a novel pipeline, reported in Fig. 1, which includes the construction of a novel fractured dataset derived from real-world CH point clouds (Sec. 3.2), the introduction of GARF-CH, an adaptation of the GARF neural reassembly model with LoRA-based fine-tuning (Sec. 3.3.1), and a systematically evaluation of the model’s performance under both complete (Sec. 3.3.2) and incomplete (Sec. 3.3.3) fragmentation scenarios, in comparison to the original GARF model.

### 3.1 Preliminaries

**GARF** In this work, we expand on GARF (Li et al., 2025b), a state-of-the-art, learning-based approach to 3D fragment reassembly. GARF is designed to address the challenge of generalizing reassembly models trained on synthetic data to real-world fracture scenarios, in which breakage patterns may be complex and varied. GARF combines two key components at its core: fracture-aware pretraining and flow matching on the  $SE(3)$  manifold. During pretraining, the model learns to identify and encode local geometric features around fracture surfaces from isolated fragments. This fracture-aware representation enables the model to generalize across unseen object shapes and varying fracture types. For the reassembly task, GARF uses a flow matching paradigm to estimate precise six-degree-of-freedom (6-DoF) transformations (rotation and translation) for each fragment. Flow matching treats the pose estimation problem as a continuous transformation between an initial noisy distribution and the target aligned configuration on the manifold of rigid body motions ( $SE(3)$ ), thereby enabling robust alignment across diverse geometric configurations. To improve robustness during inference, GARF introduces a one-step preassembly stage. In this stage, a single pass of the flow matching process provides a strong initial pose estimate, improving alignment quality, especially for unseen objects or varying numbers of fragments.

**LoRA** LoRA is a parameter-efficient fine-tuning (PEFT) technique that was originally developed for large neural networks. In LoRA, a subset of model weights is augmented with low-rank, trainable components, while the original weights remain frozen. This approach allows for efficient domain adaptation with a minimal number of additional parameters.

Formally, given a linear transformation with weight matrix  $W \in \mathbb{R}^{d_{out} \times d_{in}}$ , LoRA represents the weight update  $\Delta W$  as the

product of two low-rank matrices  $A \in \mathbb{R}^{r \times d_{in}}$  and  $B \in \mathbb{R}^{d_{out} \times r}$ , such that:

$$W' = W + BA, \quad (1)$$

where  $r \ll \min(d_{in}, d_{out})$  is the chosen rank. During fine-tuning, only the parameters of  $A$  and  $B$  are optimized, while the original weights  $W$  remain fixed. This significantly reduces the number of trainable parameters and limits overfitting, which is particularly beneficial when adapting to domain-specific datasets of limited size.

In GARF, LoRA is employed to adapt the pretrained reassembly model to new fracture domains without modifying the full set of network weights. LoRA modules are applied to the flow matching network responsible for predicting rigid transformations in the  $SE(3)$  space. More specifically, LoRA adapters are injected into the self-attention and global attention layers of the final Transformer block of the Point Encoder while unfreezing the MLP heads for pose prediction. As a result, fine-tuning primarily affects how fragment features are mapped to pose updates, without altering the underlying geometric representations learned during pretraining.

### 3.2 Dataset creation

Our dataset construction pipeline begins with the ArCH dataset (Matrone et al., 2020), a publicly available benchmark designed specifically for CH documentation and scene understanding. It provides high-quality 3D point clouds of architectural and archaeological environments, including scenes that are part (or a candidate) of the UNESCO World Heritage List<sup>1</sup>. These point clouds capture a wide range of CH structures that are characterized by large scale, geometric complexity, and real-world acquisition noise. Some examples are provided in Fig. 2. The dataset includes multiple scenes with varying architectural styles and levels of structural integrity, making it suitable for exploring fragment reassembly in CH contexts. We selected the most appropriate scenes for our task, resulting in 15 scenes total.

The ArCH dataset is provided in point cloud format. GARF, on the other hand, also requires watertight mesh representations during training to identify fracture surfaces and label fracture points on corresponding point clouds. In particular, mesh geometry is exploited to derive ground-truth fracture interfaces and

<sup>1</sup> <https://whc.unesco.org/en/list/>

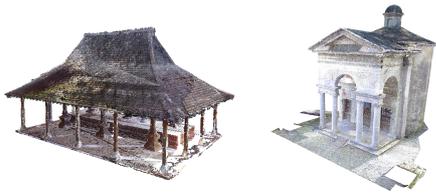


Figure 2. Examples of point clouds in the original ArCH dataset.

supervision signals for the neural reassembly process. For this reason, we have developed a dedicated pre-processing pipeline that generates clean, normalized mesh representations from the original point clouds (see Sec. 3.2.1). After generating and refining the meshes, we simulate realistic fractures on the meshes and their corresponding point clouds to create a novel dataset of large-scale CH fragments (Sec. 3.2.2). We subsequently use this fractured dataset for LoRA-based fine-tuning of GARF and for evaluation under different experimental conditions, including missing-fragment scenarios.

**3.2.1 Dataset pre-processing** We performed three pre-processing steps to convert the ArCH point clouds into mesh representations suitable for fracture simulation and model training.

**Mesh extraction** First, we performed mesh extraction. Meshes are reconstructed from the original point clouds using Poisson Surface Reconstruction (Kazhdan et al., 2006). This method generates continuous surface representations from noisy and incomplete point samples, making it ideal for large-scale CH scenes acquired through real-world scanning processes. The reconstruction parameters are tuned to balance surface fidelity and topological stability, as reported in Table 1. In particular, the octree depth controls the spatial resolution of the reconstructed surface, while the minimum number of samples per node regularizes sparsely sampled regions. The interpolation weight balances adherence to the input points and global surface smoothness, while confidence weighting and pre-cleaning are disabled as preprocessing is handled upstream.

Parameter	Value
Octree Depth	10
Min. Samples per Node	1.5
Interpolation Weight	4
Confidence Weighting	Disabled
Pre-Clean	Disabled

Table 1. Poisson Surface Reconstruction parameters used for mesh extraction.

**Mesh cleaning** Reconstructed meshes often contain artifacts, such as spurious components, surface noise, and non-manifold elements, which are introduced during reconstruction. These artifacts are removed through a combination of connected-component filtering, hole filling, and surface smoothing. This step ensures that the resulting meshes exhibit coherent geometry and are suitable for subsequent fracture simulation.

**Mesh normalization** Finally, to improve geometric consistency and robustness during fracture generation, the cleaned meshes undergo remeshing and solidification. Remeshing enforces a more uniform triangle distribution across the surface,

and solidification converts thin or open surfaces into volumetric, watertight meshes. This normalization step is critical for producing stable and physically plausible fracture surfaces.

**3.2.2 Fractures generation** Once the refined meshes have been obtained, realistic fractures are generated to simulate the fragmentation of CH structures. This process is based on Breaking Good (Sellán et al., 2023), which is a physically-inspired approach designed to produce realistic fracture patterns by modeling material stress propagation and crack formation. Breaking Good simulates fractures by iteratively applying stress fields to volumetric meshes, which are first normalized within a unit cube to ensure scale-consistent fracture behavior. This results in irregular, non-planar fracture surfaces that closely resemble real-world breakage. Unlike simple geometric cuts, this approach produces fracture interfaces with complex morphologies, which are crucial for training and evaluating learning-based reassembly methods. Notably, the original GARF framework also adopts Breaking Good, ensuring methodological consistency and compatibility with the model’s design assumptions.

In our pipeline, we fracture each normalized mesh using the Breaking Good methodology and then apply the same fracture transformations to the corresponding original point clouds. We generate two distinct fractured instances for each original mesh by simulating a point of impact on the left and right sides of the object, respectively. This strategy increases dataset diversity by introducing different fracture configurations for the same underlying structure while maintaining geometric consistency. Depending on the geometry and scale of the scene, the fracture process naturally produces a variable number of fragments. To ensure compatibility with the GARF framework while preserving scene-specific fragmentation characteristics, we adopt a controlled fragment merging strategy. Specifically, we merge the smallest fragments iteratively, since GARF prioritizes reassembling larger structural components. Through this process, we obtain a variable number of resulting fragments ranging from two to seven per fractured instance. This approach enables us to retain the most informative large-scale fragments while reducing fragmentation noise introduced by very small pieces, which are less relevant to large-scale CH structures. At the same time, this approach preserves realistic breakage patterns and reflects the variability observed across different CH scenes. The resulting paired data, comprising fractured meshes and corresponding fragmented point clouds with known fragment identities and fracture interfaces, form the basis of our novel, large-scale CH fracture dataset. This dataset has then been leveraged to fine-tune the original GARF model through a LoRA-based approach.

### 3.3 Model training and evaluation

**3.3.1 LoRA fine-tuning** We use our generated fractured CH dataset to fine-tune the original GARF model with a LoRA strategy. We split the dataset into 80% for the training set and 20% for the test set. We begin with publicly available pre-trained GARF weights, that were trained using synthetic and everyday object fracture datasets. These weights exhibit strong generalization capabilities across object categories. However, due to significant geometric, scale, and structural differences between everyday objects and CH scenes, directly applying the pretrained model to CH data may result in suboptimal performance. To address this domain gap while preserving the generalizable fracture priors learned during pretraining, we use a

parameter-efficient, LoRA-based fine-tuning approach. During fine-tuning, we keep the weights of the original GARF backbone and flow matching network frozen and inject LoRA adapters into the linear layers of the flow matching module that predict rigid transformations in the SE(3) space. Only the parameters introduced by the LoRA modules are optimized, which significantly reduces the number of trainable parameters and mitigates the overfitting risks associated with the limited size of the CH dataset. The training samples consist of fragmented point clouds and meshes, as described in Section 3.2. Following the original GARF training formulation, the model is fine-tuned to predict the relative pose of each fragment with respect to a canonical reference frame. To increase robustness and generalization, we use the two fracture configurations generated for each structure (left-side and right-side impact simulations) for implicit data augmentation. This exposes the model to diverse fracture topologies and fragment arrangements for the same underlying geometry, encouraging invariance to impact location and fragmentation patterns.

**3.3.2 Reassembly evaluation** To evaluate the effectiveness of the proposed LoRA fine-tuning method for CH structures, we compare our fine-tuned GARF-CH model with the original, pretrained (non fine-tuned) model using the test set from our generated dataset. The models outputs are quantitatively compared using metrics that follow the evaluation protocols adopted in the original GARF paper and are widely adopted in the literature (Lu et al., 2023; Wang et al., 2024), facilitating consistent comparison and benchmarking.

**Rotation and Translation Errors** We compute the root mean square error for both rotation (RMSE(R)) and translation (RMSE(T)) between the predicted and ground-truth fragment poses. Rotation error measures the angular difference (in degrees) between the predicted and ground-truth orientations. Translation error quantifies the Euclidean distance (in the same units used for the point clouds) between the predicted and ground-truth fragment positions. Lower RMSE values indicate better alignment accuracy. These metrics directly assess the model’s ability to recover the correct 6-DoF transformations of individual fragments.

**Chamfer Distance** Chamfer distance (CD) measures the geometric fidelity of the assembled object relative to the ground truth object. Given two point sets representing the assembled and ground-truth surfaces, the CD computes the average distance from each point in one set to the other, and vice versa. A lower CD indicates closer geometric alignment between the reconstructed and reference objects, capturing both global and local discrepancies in shape reconstruction.

**Part Accuracy** Part accuracy (PA) is the proportion of fragments that are correctly placed in the final assembly. A fragment is considered correctly assembled if the CD between the fragment and its corresponding ground-truth position, once transformed by the predicted pose, is below a predefined threshold (0.01 in our case). The PA metric, expressed as a percentage, emphasizes pose precision and the combined effects of rotation and translation on assembly quality. Higher PA indicates more reliable reassembly performance.

These metrics provide an overall evaluation of the quality of the reassembly in terms of pose estimation and geometric reconstruction. This enables a fair and rigorous comparison between the fine-tuned and baseline GARF models on CH structures.

**3.3.3 Missing fragments evaluation** In CH contexts, objects and structures are often incomplete due to erosion, damage, or loss of material. To evaluate the robustness of reassembly models under these conditions, we perform a further analysis, assessing the pretrained and LoRA-fine-tuned GARF models in a missing-fragment scenario. For each test instance, one fragment is randomly removed before reassembly. Then, the model is asked to reconstruct the object using only the remaining fragments, with no explicit indication of the missing part. This setting introduces additional ambiguity in fragment alignment and more accurately reflects real-world CH restoration challenges. We evaluate reassembly performances using the same metrics described in Section 3.3.2, computed on the partial reconstruction.

## 4. Experiments and results

### 4.1 Experimental settings

All experiments were conducted on a workstation running Ubuntu 22.04, equipped with an Intel Core i7-11700K CPU and an NVIDIA GeForce RTX 3090 GPU with 24 GB of VRAM. All training, inference, and evaluation stages were executed on a single GPU. The fragmentation stage was performed following Breaking Good, with a dedicated Python 3.8 environment. Subsequent assembly, evaluation, and fine-tuning were carried out using the GARF framework, leveraging Python 3.11 with CUDA 12.4 and Pytorch 2.5.1. The complete software stack and training configuration are summarized in Table 2.

Category	Specification
Python (GARF)	3.11.14
Python (Breaking Good)	3.8.20
PyTorch	2.5.1
PyTorch Lightning	2.5.0
CUDA	12.4
GPU driver	NVIDIA 580.95.05
Numerical libraries	NumPy, SciPy
Geometry processing	Open3D
Fine-tuning method	LoRA (PEFT)
Training precision	Mixed precision (FP16)
Training epochs	100 (checkpoint at epoch 86)
Initialization strategy	One-step initialization

Table 2. Software stack and training configuration used in the experiments.

Fine-tuning was performed using Low-Rank Adaptation (LoRA), keeping the feature extractor frozen and updating only the denoiser parameters. Training was run for 100 epochs, using an Early Stopping approach based on the aforementioned metrics. Based on this, epoch 86 was selected for all evaluations, as it provided the best trade-off between rotation error, translation error, and CD. All experiments employed one-step initialization during inference.

### 4.2 Results

Table 3 reports the quantitative comparison between the original GARF model and our LoRA-fine-tuned variant, GARF-CH, on the proposed large-scale CH fracture dataset.

Overall, the results demonstrate that GARF-CH outperforms the base GARF model consistently across all pose- and geometry-related metrics. GARF-CH achieves notably lower

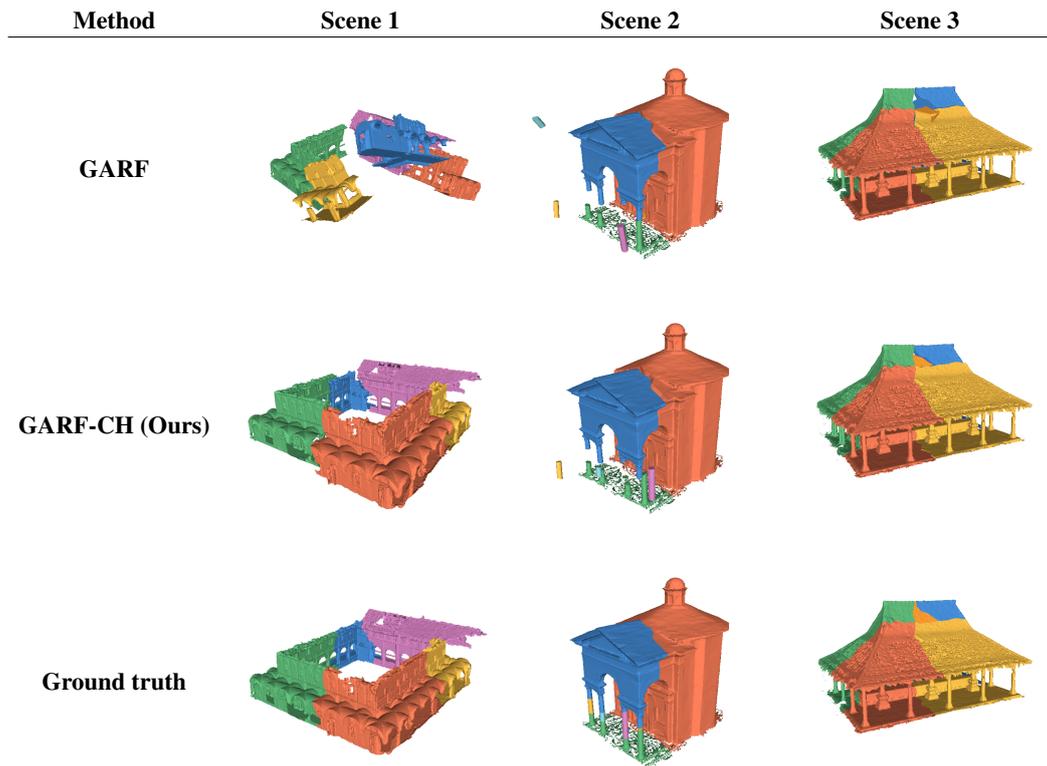


Figure 3. Qualitative results of the proposed approach, comparing the original GARF model, our LoRA fine-tuned version (GARF-CH) and the ground truth.

Method	RMSE(R) ↓ degree	RMSE(T) ↓ $\times 10^{-2}$	CD ↓ $\times 10^{-3}$	PA ↑ %
GARF	13.71	1.88	2.38	<b>83.33%</b>
Ours	<b>11.95</b>	<b>1.58</b>	<b>1.67</b>	<b>83.33%</b>

Table 3. Quantitative comparison between GARF and our proposed approach on the novel dataset. RMSE(T) and CD values are computed within a unit cube.

rotational error ( $11.95^\circ$  vs  $13.71^\circ$ ) and translational error (0.0158 vs 0.0188), indicating more accurate estimation of fragment poses in large-scale heritage scenarios. Similarly, the CD is substantially reduced (0.00167 vs 0.00238), reflecting improved geometric consistency in the reconstructed assemblies. Both approaches achieve the same placement accuracy (83.33%), suggesting that the original GARF framework already exhibits generalization beyond object-scale settings. However, the consistent improvements in pose estimation and geometric reconstruction highlight the necessity of domain adaptation to address the unique characteristics of large-scale CH data effectively. The performance gains achieved by GARF-CH are due to its LoRA-based fine-tuning strategy. This strategy allows the original model to be adapted to large-scale heritage fragments. These fragments are characterized by increased size, structural complexity, and variability. At the same time, the strategy preserves the core reasoning and reassembly capabilities of GARF. These results demonstrate that, although GARF provides a strong baseline, lightweight fine-tuning is necessary to improve its predictions and bridge the gap between object-centric and monument-scale fracture reassembly.

Table 3 reports qualitative reassembly results for three representative CH scenes, comparing the original GARF model with

our GARF-CH, and the corresponding ground truth. For Scene 1, GARF-CH achieves substantially more accurate reassembly than the base GARF model. While the base model struggles to correctly align multiple fragments and exhibits noticeable rotational and translational inconsistencies, the fine-tuned version produces a coherent reconstruction that closely matches the ground truth. In Scene 2, both GARF and GARF-CH successfully reconstruct the dominant, larger fragments, indicating that the base model already exhibits a certain degree of generalization. However, both methods have difficulty precisely aligning smaller fragments, which remain partially misaligned. Nevertheless, GARF-CH alignment results closer that the original GARF alignment even for the smaller fragments. This suggests that fine-grained fragment interaction remains challenging in highly complex scenes, though some progress can be seen. For Scene 3, both approaches yield visually plausible reconstructions that are largely consistent with the ground truth. Notably, GARF-CH correctly reassembles an additional small ceiling fragment misaligned or omitted in the GARF reconstruction. This highlights the improved robustness of the fine-tuned model when dealing with smaller yet structurally relevant components. Overall, the visual results confirm the quantitative values, demonstrating that, while the original GARF model shows promise in generalizing to CH data, the proposed GARF-CH results in more accurate and complete reassembles, especially in challenging, large-scale scenes and when dealing with smaller fragments.

#### 4.3 Missing fragments results

Table 4 shows the quantitative performance of GARF-CH when reassembling complete objects and objects with one missing fragment.

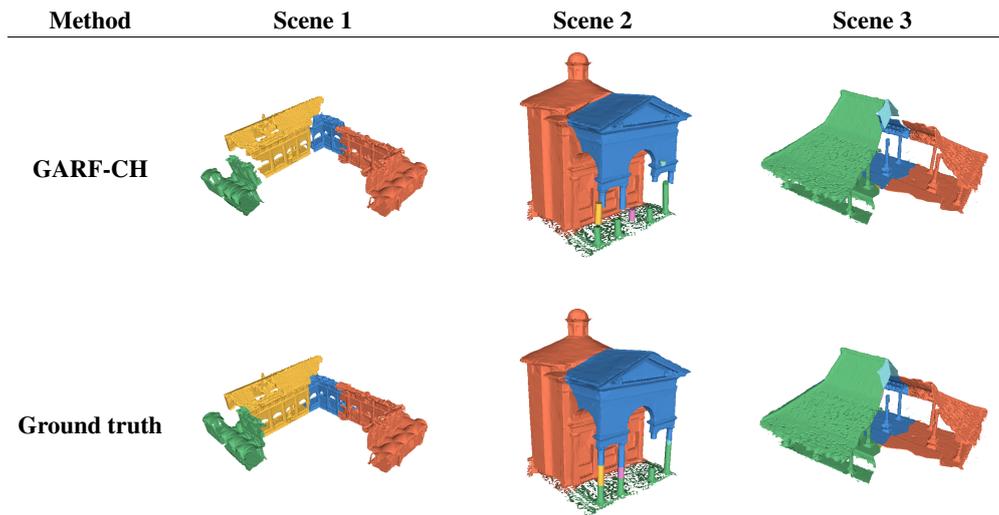


Figure 4. Qualitative results of the GARF-CH reassembly with a missing fragment.

Method	RMSE(R) ↓ degree	RMSE(T) ↓ $\times 10^{-2}$	CD ↓ $\times 10^{-3}$	PA ↑ %
Complete	11.95	<b>1.58</b>	<b>1.67</b>	83.33%
Missing fragment	<b>11.19</b>	2.99	7.61	<b>86.67%</b>

Table 4. Quantitative comparison between the complete and incomplete reassembly.

This reflects a common scenario in CH applications, where parts of a structure are lost or unavailable. The results show that the rotation error remains largely unchanged between the two settings, indicating that the model can estimate consistent global orientations for the remaining fragments, even without complete geometric information. In contrast, the translation error and CD increase in the missing-fragment scenario, as expected, due to reduced spatial constraints and ambiguity from incomplete geometry. Interestingly, PA increases when a fragment is missing. This can be explained by the smaller number of fragments that must be correctly placed, which simplifies the combinatorial aspect of the reassembly task while still requiring accurate alignment of the main structural components. The qualitative results for the missing-fragment scenario, shown in Fig. 4, further support the quantitative findings. Despite the absence of one fragment, GARF-CH can generally produce visually coherent reassemblies, correctly positioning the remaining major components and preserving the overall structural layout. Minor inconsistencies are present, but, nevertheless, the reconstructed scenes remain plausible and structurally consistent. Taken together, these results suggest that GARF-CH exhibits robust behavior under incomplete conditions, supporting its suitability for real-world CH scenarios in which missing fragments are common.

## 5. Conclusions and future works

In this work, we investigated the applicability of learning-based 3D fragment reassembly methods to large-scale CH structures, a scenario that remains largely underexplored in the literature. Starting with the ArCH dataset, we created a new fractured dataset derived from real-world CH point clouds. This new dataset is designed to support the systematic evaluation of fragment reassembly at an architectural scale. Based on

the state-of-the-art GARF framework, we analyzed its generalization capabilities in this challenging domain. We also proposed GARF-CH, a lightweight, parameter-efficient adaptation strategy of the original model using LoRA fine-tuning. Extensive quantitative and qualitative experiments demonstrated that, while the original GARF model exhibits a certain degree of robustness when applied to CH data, domain adaptation is crucial to addressing the geometric complexity, scale variability, and acquisition noise characteristic of heritage structures. Furthermore, robustness analysis under missing-fragment conditions highlighted the approach’s practical suitability for real-world CH restoration scenarios where incomplete data is the norm. Several areas could be explored in the future. One primary extension would be to expand the proposed dataset to include a larger number of CH structures with a wider variety of architectural typologies, historical periods, materials, and levels of degradation, as smaller datasets limit generalization. Large public repositories of digitized heritage assets, such as the recently released Global Digital Heritage collection<sup>2</sup>, are valuable sources for increasing dataset diversity and improving model generalization. Beyond expanding the dataset, future research could investigate multi-fragment loss scenarios, progressive or hierarchical reassembly strategies for very large scenes, and integrating semantic or structural priors to guide reassembly in ambiguous cases. Finally, combining fragment reassembly with complementary restoration tasks, such as point cloud completion or semantic segmentation, could enhance the reliability and usability of learning-based approaches in real-world CH conservation workflows.

## References

- Balloni, E., Ceka, D., Pierdicca, R., Paolanti, M., Mancini, A., Zingaretti, P., 2024. Comparative assessment of Neural Rendering methods for the 3D reconstruction of complex heritage sites in the inner areas of the Marche region-Italy. *Digital Applications in Archaeology and Cultural Heritage*, 35, e00371.
- Basu, A., Paul, S., Ghosh, S., Das, S., Chanda, B., Bhagvati, C., Snares, V., 2023. Digital restoration of cultural heritage with data-driven computing: A survey. *IEEE Access*, 11, 53939–53977.

<sup>2</sup> <https://sketchfab.com/GlobalDigitalHeritage>

- Cacciotti, R., Kaiser, A., Sardella, A., De Nuntii, P., Drdácĳ, M., Hanus, C., Bonazza, A., 2021. Climate change-induced disasters and cultural heritage: Optimizing management strategies in Central Europe. *Climate Risk Management*, 32, 100301.
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W. et al., 2022. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2), 3.
- Jia, C., Ren, Y., Wang, Z., Zhang, Y., 2025. GeoAssemble: A Geometry-Aware Hierarchical Method for Point Cloud-Based Multi-Fragment Assembly. *Sensors*, 25(21), 6533.
- Kazhdan, M., Bolitho, M., Hoppe, H., 2006. Poisson surface reconstruction. *Proceedings of the fourth Eurographics symposium on Geometry processing*, 7number 4.
- Lamb, N., Palmer, C., Molloy, B., Banerjee, S., Banerjee, N. K., 2023. Fantastic breaks: A dataset of paired 3d scans of real-world broken objects and their complete counterparts. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4681–4691.
- Li, A., Zimmer-Dauphinee, J. R., Kalyanam, R., Lindsay, I., VanValkenburgh, P., Wernke, S., Aliaga, D., 2025a. Self-supervised large scale point cloud completion for archaeological site restoration. *Proceedings of the Computer Vision and Pattern Recognition Conference*, 11759–11768.
- Li, J., Guo, S., Wang, L., Han, S., 2024. CompleteDT: Point cloud completion with information-perception transformers. *Neurocomputing*, 592, 127790.
- Li, M., 2024. Disaster risk management of cultural heritage: A global scale analysis of characteristics, multiple hazards, lessons learned from historical disasters, and issues in current DRR measures in world heritage sites. *International Journal of Disaster Risk Reduction*, 110, 104633.
- Li, S., Jiang, Z., Chen, G., Xu, C., Tan, S., Wang, X., Fang, I., Zyskowski, K., McPherron, S. P., Iovita, R. et al., 2025b. Garf: Learning generalizable 3d reassembly for real-world fractures. *arXiv preprint arXiv:2504.05400*.
- Liu, B., Wang, M., Niu, X., Wang, S., Zhang, S., Zhang, J., 2021. A fragment fracture surface segmentation method based on learning of local geometric features on margins used for automatic utensil reassembly. *Computer-Aided Design*, 132, 102963.
- Llabani, A., Abazaj, F., 2024. 3D documentation of cultural heritage using terrestrial laser scanning. *Journal of Applied Engineering Science*, 22(2), 267–271.
- Lu, J., Sun, Y., Huang, Q., 2023. Jigsaw: Learning to assemble multiple fractured objects. *Advances in Neural Information Processing Systems*, 36, 14969–14986.
- Matrone, F., Lingua, A., Pierdicca, R., Malinverni, E. S., Paolanti, M., Grilli, E., Remondino, F., Murtiyoso, A., Landes, T., 2020. A benchmark for large-scale heritage point cloud semantic segmentation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 1419–1426.
- Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E. S., Frontoni, E., Lingua, A. M., 2020. Point cloud semantic segmentation using a deep learning framework for cultural heritage. *Remote Sensing*, 12(6), 1005.
- Pietroni, E., Ferdani, D., 2021. Virtual restoration and virtual reconstruction in cultural heritage: Terminology, methodologies, visual representation techniques and cognitive models. *Information*, 12(4), 167.
- Sanchez-Aparicio, L. J., del Blanco-Garcia, F. L., Mencias-Carrizosa, D., Villanueva-Llaurado, P., Aira-Zunzunegui, J. R., Sanz-Arauz, D., Pierdicca, R., Pinilla-Melo, J., Garcia-Gago, J., 2023. Detection of damage in heritage constructions based on 3D point clouds. A systematic review. *Journal of Building Engineering*, 77, 107440.
- Sellán, S., Chen, Y.-C., Wu, Z., Garg, A., Jacobson, A., 2022. Breaking bad: A dataset for geometric fracture and reassembly. *Advances in Neural Information Processing Systems*, 35, 38885–38898.
- Sellán, S., Luong, J., Mattos Da Silva, L., Ramakrishnan, A., Yang, Y., Jacobson, A., 2023. Breaking good: Fracture modes for realtime destruction. *ACM Transactions on Graphics*, 42(1), 1–12.
- Siliutina, I., Tytar, O., Barbash, M., Petrenko, N., Yepyk, L., 2024. Cultural preservation and digital heritage: challenges and opportunities. *Amazonia Investiga*, 13(75), 262–273.
- Solina, F., 2018. 3D Point Clouds for Documentation, Analysis and Promotion of Cultural Heritage Artifacts.
- Storeide, M. S. B., George, S., Sole, A., Hardeberg, J. Y., 2023. Standardization of digitized heritage: a review of implementations of 3D in cultural heritage. *Heritage Science*, 11(1), 1–22.
- Tsesmelis, T., Palmieri, L., Khoroshiltseva, M., Islam, A., Elkin, G., Shahar, O. I., Scarpellini, G., Fiorini, S., Ohayon, Y., Alali, N. et al., 2024. Re-assembling the past: The RePAIR dataset and benchmark for real world 2D and 3D puzzle solving. *Advances in Neural Information Processing Systems*, 37, 30076–30105.
- Villegas-Suarez, A. M., Lopez, C., Sipiran, I., 2023. Matchmakernet: Enabling fragment matching for cultural heritage analysis. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1632–1641.
- Wang, Z., Chen, J., Furukawa, Y., 2024. Puzzlefusion++: Auto-agglomerative 3d fracture assembly by denoise and verify. *arXiv preprint arXiv:2406.00259*.
- Xu, Q.-C., Cao, Y.-P., Cheng, W., Mu, T.-J., Shan, Y., Yang, Y.-L., Hu, S.-M., 2025. High-accuracy fractured object reassembly under arbitrary poses. *International Conference on Computational Visual Media*, Springer, 194–215.
- Xu, Q.-C., Chen, H.-X., Hua, J., Zhan, X., Yang, Y.-L., Mu, T.-J., 2024. Fragmentdiff: A diffusion model for fractured object assembly. *SIGGRAPH Asia 2024 Conference Papers*, 1–12.
- Ye, S., Zhu-Tian, C., Chu, X., Li, K., Luo, J., Li, Y., Geng, G., Wu, Y., 2022. PuzzleFixer: A visual reassembly system for immersive fragments restoration. *IEEE transactions on visualization and computer graphics*, 29(1), 429–439.
- Zhang, K., Yu, W., Manhein, M., Waggenspack, W., Li, X., 2015. 3d fragment reassembly using integrated template guidance and fracture-region matching. *Proceedings of the IEEE international conference on computer vision*, 2138–2146.
- Zhang, Y., Li, K., Chen, X., Zhang, S., Geng, G., 2018. A multi feature fusion method for reassembly of 3D cultural heritage artifacts. *Journal of Cultural Heritage*, 33, 191–200.