



## EG-ICE 2025 GLASGOW

# A Conditional Diffusion Model for Bridge Point Cloud Repair

YUNPING FANG<sup>1</sup> JELENA NINIC<sup>1</sup>

<sup>1</sup>Department of Civil Engineering, University of Birmingham, Birmingham, UK

### ABSTRACT

Accurate and complete point cloud data are essential for the digital reconstruction of large-scale infrastructure. However, point clouds acquired in real-world scenarios are often incomplete due to occlusions, sensor limitations, and environmental constraints. This study extends a conditional denoising diffusion probabilistic model to address bridge point cloud completion. A point cloud encoder is designed to extract latent shape representations from incomplete inputs, which are then used to guide the generative process of the diffusion model. The method enables the production of high-fidelity, structurally consistent, and resolution-flexible point clouds. In addition, a dataset generation strategy is introduced to simulate typical point cloud defects encountered during scanning. Experiments on real-world bridge data validates the effectiveness of the proposed approach in addressing complex and large-scale point cloud completion tasks. Compared to the baseline, the modified encoder achieves consistent improvements across multiple structural components, with Chamfer Distance (CD) reduced by up to 23.2% and Earth Mover's Distance (EMD) by up to 54.0%, indicating enhanced geometric accuracy and structural integrity.

### KEYWORDS:

Point cloud completion; Diffusion models; Bridge reconstruction; Conditional generation.

## 1. INTRODUCTION

Point clouds represent a core enabler for the digitalisation of infrastructure and are widely employed in structural health monitoring, modelling, and asset management. With the advancement of point cloud-based 3D reconstruction techniques, reverse engineering using point clouds to generate BIM models has emerged as an effective approach for asset management. By providing rich geometric and semantic information, 3D point clouds enhance the level of detail in BIM models and reduce reliance on construction drawings and conventional surveying methods.

However, raw point clouds acquired through LiDAR or photogrammetry are often incomplete due to occlusions, sensor limitations, and environmental constraints. In bridge structures, missing data frequently occur in areas with complex geometries—such as the undersides of decks, joints, and intricate connections—where direct scanning is difficult (Fang et al., 2025). This incompleteness presents significant challenges for downstream tasks, including bridge information modelling (BIM), structural health monitoring (SHM) (He et al., 2022), and finite element analysis (FEA). Consequently, point cloud completion has become

a critical step in the digital reconstruction of infrastructure assets. In response, various learning-based methods have been developed to address the issue of incompleteness by inferring missing structural elements. Among them, deep learning approaches have shown considerable promise in predicting and completing missing bridge components with high fidelity (Matono and Nishio, 2024). By learning spatial patterns and structural correlations from data, these models can produce realistic and topologically consistent point cloud reconstructions.

Nevertheless, the inherent structural complexity and diversity of infrastructure, coupled with the limited availability of open-source training datasets, significantly constrain the generalisation capabilities of such models when applied to large-scale and geometrically complex scenarios.

To address these challenges, this study focuses on point cloud completion for infrastructure. The main contributions of this study are as follows:

- i. A conditional diffusion model is refined and applied to the reconstruction of bridge point clouds, enabling the generation of complete and structurally consistent point clouds from partial inputs.
- ii. A shape encoder is integrated to extract latent representations from incomplete point clouds, providing conditional guidance for the generative process and enhancing the structural fidelity of the outputs.
- iii. The proposed method is validated on real-world bridge cases, demonstrating its effectiveness in completing large-scale, complex point clouds typical of infrastructure scenarios.

## 2. RELATED WORK

Existing point cloud completion methods require paired data comprising incomplete point clouds and their corresponding ground truth completions for training. These models are predominantly trained by minimising similarity metrics between the predicted and ground truth point clouds, most commonly using Chamfer Distance (CD) (Fan et al., 2017) or Earth Mover’s Distance (EMD) (Rubner et al., 2000). Representative models such as PCN (Yuan et al., 2018), MSN (Liu et al., 2020), PF-net (Huang et al., 2020) and SnowflakeNet (Xiang et al., 2021) follow this paradigm. While these approaches have achieved reasonable performance, they suffer from inherent limitations.

CD is sensitive to local geometric outliers, which can result in uneven point distributions in the generated point cloud. Although EMD provides more accurate alignment, its high computational cost limits its scalability for training. Furthermore,

these models typically generate a fixed number of points, making it difficult to adapt to scenarios that demand variable or high-density outputs, particularly in the context of detailed infrastructure reconstruction.

Compared with conventional models trained using Chamfer Distance (CD) or Earth Mover’s Distance (EMD), diffusion-based methods are capable of generating point clouds with more uniform density and allow flexible control over the number of generated points, owing to their strong ability to model complex data distributions.

Denosing Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020) have been proposed and have achieved remarkable success in image generation. Similarly, their effectiveness has also been demonstrated in the domain of point cloud processing (Luo and Hu, 2021; Lyu et al., 2021; Melas-Kyriazi et al., 2023; Zhou et al., 2021). With the integration of conditional information, condition-guided diffusion models have shown impressive performance across a range of downstream tasks, such as image synthesis and shape completion.

For example, Luo et al. formulated point cloud completion as a conditional generation task (Luo and Hu, 2021), in which the shape of the object serves as the condition for generating the corresponding complete point cloud. This approach successfully applied diffusion models to the point cloud completion problem

However, existing studies have primarily focused on public datasets containing simple synthetic objects, and have demonstrated completion capabilities only on small-scale items such as chairs, sofas, or lamps. The application of diffusion models to large-scale objects in real-world scenarios, such as bridges, remains largely unexplored and faces considerable limitations.

This gap can be attributed to several challenges. Firstly, outdoor scene point clouds differ significantly from synthetic data, often containing millions of points and suffering from uneven density, noise, and occlusions, making completion considerably more difficult. Secondly, the structural diversity of large-scale infrastructure necessitates extensive and varied training data. However, such datasets are difficult to obtain, particularly due to access restrictions and privacy concerns, which often prevent their public release. This lack of open large-scale infrastructure datasets severely limits the training and evaluation of models for real-world applications.

The largest publicly available real-world bridge point cloud dataset to date is the Cambridgeshire Bridge Dataset (Lu et al., 2018). This dataset comprises point clouds of 10 highway reinforced concrete (RC) bridges acquired using terrestrial

laser scanning (TLS). However, the limited number of bridges is insufficient for training data-hungry neural networks. Furthermore, the raw scans contain various defects and lack accompanying design drawings, making it impossible to reconstruct accurate ground truth representations.

For these reasons, many researchers have turned to synthetic datasets as substitutes for real-world data. One notable method is the Heidelberg LiDAR Operations Simulator (HELIOS) (Winiwarter et al., 2022), which simulates the ray tracing process of laser scanners. By specifying scanner coordinates, scanning parameters, and model positioning, HELIOS can generate realistic TLS-like point clouds. However, this approach is computationally expensive and time-consuming (Esmorís et al., 2022), with the generation of a point cloud sample taking approximately 20 minutes, rendering it unsuitable for large-scale dataset generation. To address these limitations and enable efficient acquisition of multiple samples, most existing studies adopt a more practical approach by preparing 3D models (e.g., in .obj format) in building information modelling (BIM) software such as Revit, and then uniformly sampling points on the model surfaces using a mesh-based sampling method (Shi et al., 2024; Yang et al., 2022). The synthetic dataset in this study is also constructed following this widely used strategy.

In response to these limitations, this study introduces a novel data generation framework specifically designed for large-scale infrastructure, and investigates, for the first time, the feasibility of applying conditional diffusion models to bridge point cloud completion.

### 3. DATASET PREPARATION

The dataset employed in this study comprises two real-world bridge point clouds for case studies, along with a synthetic dataset containing the fundamental structural components of the bridge. The two real-world bridges, Polyfytos Bridge and the Aqueduct Bridge, are scanned using terrestrial laser scanning, as shown in Figure 1. The included components cover both geometrically simple elements (such as piers and bridge caps) and more complex objects (such as I-shaped girders and decks with varying heights), representing the most common types found in reinforced concrete highway bridges.

Polyfytos bridge dataset is acquired from the Polyfytos Bridge in Greece, which is a simply supported reinforced concrete girder bridge. The structure is arranged with three main I-shaped girders positioned in parallel, and the piers are designed with octagonal cross-sections (Figure 2a). Following initial denoising and segmentation, each

span was partitioned into a block, and each block was further subdivided into five components: bridge deck, girders, transverse diaphragms, bridge caps, and piers.

The Aqueduct Bridge dataset is collected from a simply supported highway bridge located in Birmingham, United Kingdom. The bridge carries a canal along its central span, allowing for boat passage, while pedestrian walkways are located on either side. The point cloud is segmented into two main components: the pier and the deck. The pier features a composite cross-section consisting of two semicircles and a rectangle. The deck exhibits a more complex geometry, characterised by a smoothly varying height profile in the side view and a grooved polygonal cross-section (Figure 2b).



Figure 1: Real photos of Polyfytos Bridge (left) and Aqueduct Bridge (right).

As the collected point cloud data exhibited incompleteness and defects (shown in Figure 2, a complete point cloud representation of the bridge was reconstructed based on the original design drawings, which was subsequently used as the ground truth for evaluation purposes.

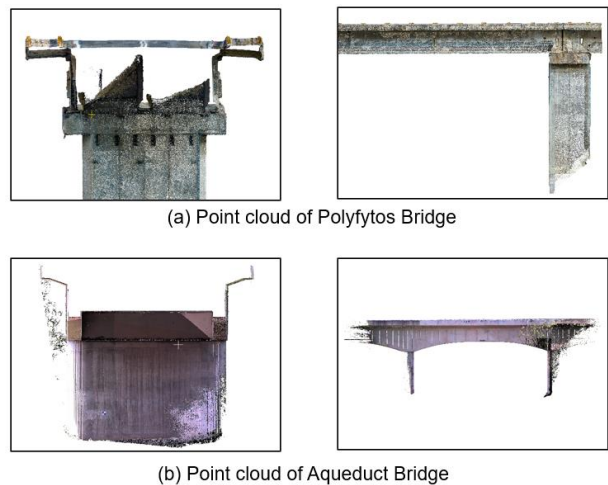


Figure 2: Incomplete bridge point cloud: cross-sectional view (left) and side view (right).

The synthetic dataset is generated following a systematic strategy. Geometric models of various bridge components are first constructed in Revit based on engineering drawings. Complete point clouds are then created by uniformly sampling

points on the surfaces of these geometric entities. To simulate incompleteness, a virtual sphere with a predefined radius is randomly moved within the bounding box of each complete point cloud, with the process implemented in a Python script. The intersection between the sphere and the complete point cloud is then computed via Boolean operations to remove the points within the sphere, thereby generating the incomplete point cloud.

To further simulate defects that may occur during real-world scanning processes, additional degradation strategies are introduced beyond the Boolean-based occlusion simulation. Uneven point density is modelled by computing the intersection between a second virtual sphere and the bounding box, within which 40% of the points are randomly replaced by (0, 0, 0), representing missing or corrupted data. In addition, noise is introduced by randomly selecting 10% of the points and replacing them with points sampled uniformly at random from within the bounding space. The overall process of generating the synthetic point cloud dataset is illustrated in Figure 3.

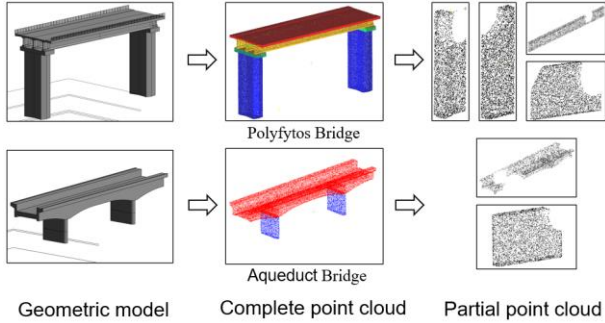


Figure 3: Workflow for generating the synthetic point cloud dataset.

In this study, the synthetic dataset includes 7 typical components of a girder bridge. For each component, 500 samples are generated, resulting in a total of 3,500 samples. Each sample consists of one complete point cloud and 24 incomplete variations, with each point cloud containing 5,000 points.

#### 4. METHODOLOGY

In this study, point cloud completion is formulated as a conditional generation task, where the incomplete point cloud serves as the conditional input. A point cloud encoder is employed to extract a latent representation  $z$ , which captures the underlying object shape. Since the incomplete and complete point clouds share the same overall structure, the latent code  $z$  is used as a condition to guide the generative process. The conditional denoising diffusion probabilistic model is then utilised to generate the corresponding complete

point cloud. The overall completion pipeline is illustrated in Figure 4.

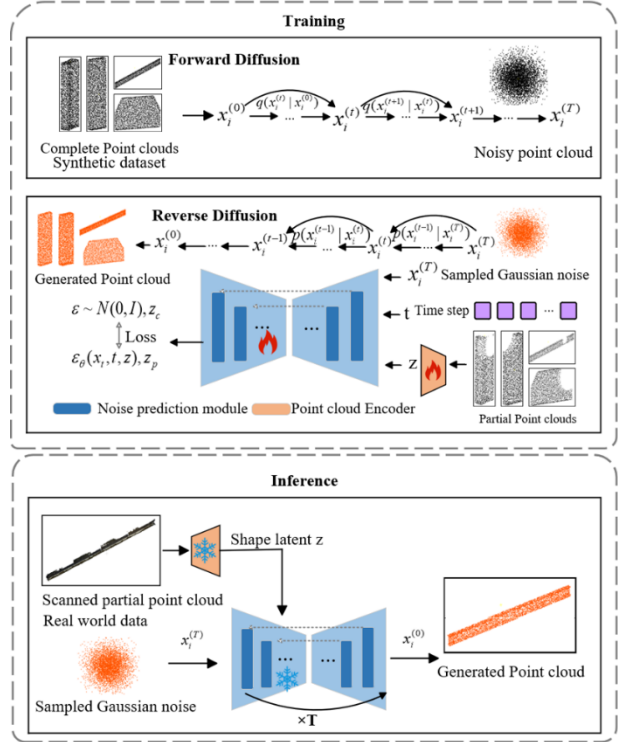


Figure 4 : Point cloud completion pipeline.

Section 4.1 introduces the theoretical foundations of the proposed model. Section 4.2 presents the design of the point cloud encoder. Section 4.3 details the implementation of the forward diffusion process, while Section 4.4 describes the network architecture used to predict noise during the reverse denoising process.

#### 4.1 Theoretical foundations of conditional denoising diffusion probabilistic models

This work adopts the theoretical formulation of the point cloud diffusion probabilistic model (DPM) (Luo and Hu, 2021) as the foundation. The DPM framework interprets the distribution of point clouds analogously to the physical process of diffusion. Specifically, noise is incrementally added to the point cloud at each time step  $t$ . As  $t$  increases, the structure of the point cloud progressively deteriorates until it becomes indistinguishable from random Gaussian noise distributed in space. Conversely, as  $t$  decreases, the point cloud gradually recovers the shape of the original object from noise. These two processes correspond to the forward diffusion and reverse denoising steps, respectively. In both steps, the probability distribution of the point cloud can be modelled as a function of the time step  $t$  and the noise level  $\epsilon$ .

**The forward process.** In the denoising diffusion probabilistic model (DPM) framework, the forward

process is denoted by  $q$ , which defines the gradual corruption of data by noise. Operating at the level of individual points, the DPM assumes independent noise perturbations for each point. At every time step  $t$ , Gaussian noise  $\varepsilon$ , sampled from a standard normal distribution  $\mathcal{N}(0, I)$ , is added to the input point cloud according to Equation (1). Specifically, the forward process is defined as:

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t | \sqrt{1 - \beta_t} x_{t-1}, \beta_t I) \quad (1)$$

Where  $t = 1, \dots, T$ ,  $T$  denotes the total number of diffusion steps. In the forward process, a predefined variance schedule  $\beta_1 \dots \beta_T$  is used to determine the noise level added at each time step.

It can further be derived that the point cloud at any time step  $t$  follows a distribution that can be expressed as a combination of the original clean point cloud  $x_0$  and Gaussian noise  $\varepsilon$ , as shown in Equation (2):

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \varepsilon \quad (2)$$

Where  $\bar{\alpha}_t$  represents the remaining proportion of the original image after  $t$  steps.

**The reverse process.** In the reverse denoising process, the denoising diffusion probabilistic model (DPM) learns to approximate the reverse-time dynamics, denoted by  $p$ , which iteratively removes noise to recover the original data. This reverse process constitutes the generative mechanism of the diffusion framework. At each time step  $t$ , the model takes as input the noisy point cloud  $x_t$  and predicts the noise  $\hat{\varepsilon}_\theta$  that was added during the corresponding step of the forward diffusion process, where  $\theta$  represents the parameters of the neural network.

In addition, a point cloud encoder is employed to compress the clean point cloud  $x_0$  into a latent representation  $\mathbf{z}$ , which serves as conditional information to guide the generation process. During training, both the complete and partial point clouds are provided to the encoder to learn the robust conditional representation. During inference, only the partial point cloud is input to the encoder to generate the shape latent for guiding the completion process.

Given  $x_t$ ,  $\mathbf{z}$ , and  $\hat{\varepsilon}_\theta$ , the distribution of the point cloud at the previous time step  $t - 1$  can be updated as Equation (3):

$$p_\theta(x_{t-1} | x_t, \mathbf{z}) = \frac{1}{\sqrt{\alpha_t}} \left( x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \hat{\varepsilon}_\theta(x_t, t, \mathbf{z}) \right) + \sqrt{\beta_t} \varepsilon \quad (3)$$

Where  $\alpha_t = 1 - \beta_t$  denotes the noise retention coefficient at time step  $t$ , and  $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$  is the cumulative product of noise retention terms from step 1 to  $t$ , which controls the total amount of noise added up to time  $t$ .

**Training.** During training, the network is optimised to minimise the difference between the

predicted noise  $\hat{\varepsilon}_\theta$  and the true noise  $\varepsilon$  added during the forward diffusion process, typically using a mean squared error (MSE) loss.

$$\mathcal{L}_{\text{simple}} = E_{x_0, \varepsilon, t} [|\varepsilon - \hat{\varepsilon}_\theta(x_t, t, \mathbf{z})|_2^2] \quad (4)$$

In contrast to conventional point cloud completion approaches that are predominantly trained by minimising geometric distances—such as Chamfer Distance (CD)—between predicted and ground truth point clouds, the diffusion-based method adopts a fundamentally different training paradigm. The model is trained by minimising a simple mean squared error (MSE) loss between the predicted and true noise values added during the forward diffusion process, thereby enabling it to implicitly learn the underlying data distribution. This approach avoids the limitations of CD-based optimisation, which often results in uneven point densities and poor generalisation in complex geometries.

Since the input during inference is an incomplete point cloud, it is essential for the encoder to correctly identify the underlying shape category to enable accurate completion. To achieve this, during training, both the complete point cloud and its corresponding incomplete version are passed through the encoder to obtain two latent representations  $z_c$  for the complete point cloud and  $z_p$  for the partial input. To encourage consistency between these two representations, their cosine similarity is incorporated into the overall loss function as an additional regularisation term. The resulting training objective is defined as follows in Equation (5):

$$\mathcal{L} = \mathcal{L}_{\text{simple}} + \lambda \cdot (1 - \cos(z_c, z_p)) \quad (5)$$

The weighting coefficient  $\lambda$  is introduced to balance the reconstruction loss and the latent alignment term. In our experiments, we set  $\lambda=0.08$ , which is found to provide a good trade-off between shape consistency and noise prediction accuracy. Furthermore, the generation process is conditioned on a latent shape embedding extracted from the input point cloud, which guides the reverse diffusion trajectory towards semantically and structurally consistent reconstructions. Unlike traditional models that are constrained to producing a fixed number of output points, the diffusion framework offers the flexibility to generate point clouds of arbitrary resolution. This makes it particularly suitable for large-scale infrastructure scenarios, where detailed and high-density reconstructions are often required.

## 4.2 Latent Shape Encoder

Most existing approaches for point cloud feature extraction are based on the PointNet architecture (Charles et al., 2017), which utilises multilayer perceptron (MLP) to independently encode each point into a high-dimensional feature space. For

instance, a point cloud of shape  $(N,3)$  is typically processed by an MLP that maps each point’s three-dimensional coordinates into a  $C$ -dimensional feature vector ( $C>3$ ). A symmetric aggregation function, such as max-pooling or average pooling, is then applied across all  $N$  points to obtain a single global latent representation of the shape, resulting in a feature vector of shape  $(1, C)$ .

For comparison, a PointNet-based fully connected encoder is adopted as the baseline, as illustrated in Figure 5.

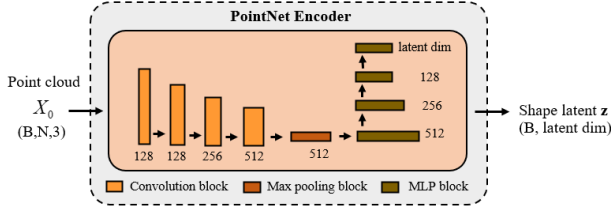


Figure 5: Detailed architecture of PointNet based encoder (baseline).

This encoder extracts global shape features by sequentially mapping the input point cloud through a series of 1D convolutional and MLP blocks, with output dimensions of 128, 256, 512, and 256, respectively.

While effective for capturing global shape information, such encoders neglect the local geometric relationships between points, which are crucial for fine-grained tasks such as point cloud completion. As a result, they struggle to represent complex geometries with multiple distinct edges, such as I-shaped bridge components, often leading to ambiguous representations and misinterpretation of structural details. Completing missing regions requires detailed local context to infer the structural continuity of partially observed shapes. Thus, to effectively capture local latent shape representations from unordered 3D point clouds, we adopt a hierarchical encoder based on the PointCNN framework (Li et al., 2018). PointCNN overcomes the limitation of point cloud unorderedness by enabling convolution-like operations to be applied directly to 3D point sets. It introduces the XConv layer, which learns a  $K \times K$  transformation matrix to reorder the input points within each local neighbourhood. This learned transformation effectively imposes a canonical order, allowing feature aggregation over neighbouring points in a consistent and learnable manner.

By adjusting the receptive field size (the kernel size and dilation rate), the model can capture features at multiple spatial resolutions. Compared with PointNet++ (Qi et al., 2017) and other hierarchical architectures, PointCNN has

demonstrated superior performance in capturing both local and global geometric information.

Unlike the conventional PointCNN architecture, which progressively propagates features through stacked layers, our encoder independently extracts features at each receptive field level and aggregates them at the feature fusion stage. This design enhances the representation of local geometric structures by preserving and combining information from multiple spatial resolutions more explicitly. These modifications enable the simultaneous learning of both fine-grained local structures and global shape context. The encoder consists of four stacked XConv layers, each operating on a downsampled version of the input point cloud to progressively capture geometric features at increasing spatial resolutions. To enhance structural representation, we apply global average pooling after each XConv layer to extract intermediate features from distinct receptive fields. These features are later concatenated and fused through a lightweight MLP to form a comprehensive latent representation.

Specifically, it consists of four stacked XConv layers with output dimensions of 48, 96, 192, and 384, employing increasing kernel sizes of 8, 12, 16, and 16, respectively. Our encoder applies global average pooling after each XConv layer to extract intermediate feature vectors  $F_i$ , each corresponding to a distinct receptive field and resolution level. These vectors are then concatenated and passed through a modified multi-layer perceptron (MLP) with progressively decreasing dimensions ( $720 \rightarrow 384 \rightarrow 256 \rightarrow 128$ ) to obtain the final latent shape representation  $z$ . This hierarchical aggregation and dimensionality reduction enable the encoder to integrate local-to-global information more effectively, and are tailored to support conditional shape completion tasks. The detailed architecture of the proposed encoder is illustrated in Figure 6.

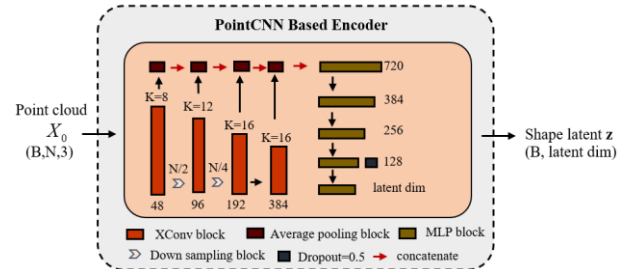


Figure 6: Detailed architecture of modified PointCNN encoder.

The quantitative comparison between the proposed encoder and the baseline will be presented in Section 5 to evaluate their effectiveness in the point cloud completion task.

### 4.3 Implementation of the forward diffusion process

The created synthetic dataset containing five types of bridge components is used as the input for the forward diffusion process. According to Equation (2), point clouds with noise  $x_t$  are sampled at each time step  $t$ . Figure 7 illustrates the forward diffusion results using a pier component as an example, showing the progressive degradation of the point cloud with increasing noise levels over time steps.



Figure 7: Results of the forward diffusion of the bridge pier.

The purpose of the forward process is to simulate the gradual corruption of the data distribution by noise, and to calculate two key parameters  $\beta$  and  $\bar{\alpha}_t$  for reverse diffusion process. The noise schedule  $\beta_t \in (0,1)$  controls the amount of noise injected at each time step  $t$ . The cumulative product  $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$  determines the proportion of the original signal retained at time step  $t$ , and thus defines the balance between the preserved structure and the added noise throughout the diffusion process.

In this study, the total number of diffusion steps is set to  $T=1000$ . The noise schedule  $\beta_t$  is linearly increased from  $1 \times 10^{-4}$  to 0.02 over the course of the diffusion process, progressively controlling the degree of noise added at each time step.

### 4.4 Noise predict network architecture

Inspired by the architecture of the original DPM model, we design the main noise prediction network using a residual pyramid-based feature extraction structure equipped with gated modules. Unlike the standard pyramid feature extraction structure, fully connected layers are replaced with 1D convolutional layers to reduce computational cost in our model, as weight sharing in convolution enables more efficient processing of large-scale point clouds. This allows for more computational resources to be allocated to the encoder, which enhances the ability of the model to capture fine-grained local structures. Moreover, the convolutional design accommodates variable-sized inputs and supports higher point densities. The detailed architecture of the pyramid network is illustrated in Figure 8.

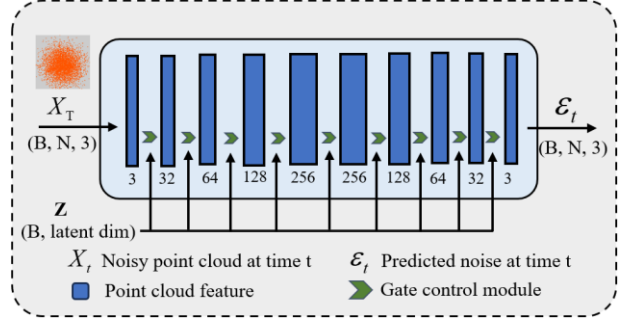


Figure 8: Pyramid-Based Noise Prediction Network.

The noise prediction network takes as input a set of spatially random points sampled from a standard Gaussian distribution, a randomly selected time step  $t \in (0, T)$ , and the shape latent code  $z$  obtained from the PointCNN-based encoder. The network is trained to predict the noise  $\hat{\epsilon}_t$  corresponding to time step  $t$ . The predicted noise shares the same shape as the input point cloud. Specifically, the spatial coordinates of the random points are progressively transformed from a 3D space to a 256-dimensional latent space and then projected back to 3 dimensions. Therefore, the predicted noise can be intuitively understood as the displacement required to transform the spatially random points into the target shape, i.e., the error between the random input and the ground truth point cloud.

The latent code  $z$  serves as a condition to guide the generation process, enabling the model to progressively transform noise into a coherent and complete object shape for point cloud completion. This conditional information is injected into the network through a gated feature modulation module, as illustrated in Figure 9.

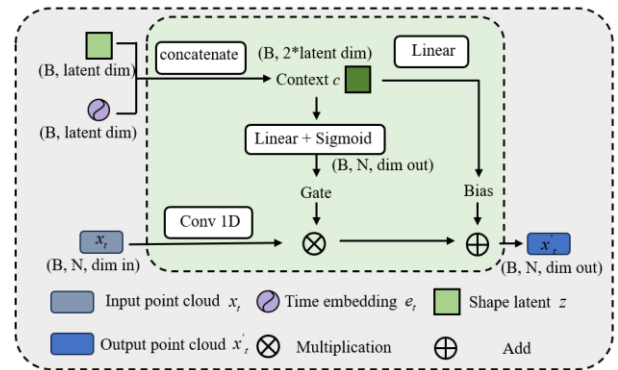


Figure 9: Gated Feature Modulation Module.

Specifically, the latent shape code  $z$  is concatenated with the time step  $t$  and passed through two parallel linear layers to produce a gating vector and a bias vector; the latter is generated without an additional bias term. Meanwhile, the input noisy point cloud  $x_t$  is processed by a 1D

convolutional layer to match the dimensionality of the gating and bias vectors.

$$y = (\text{Conv1D}(x_t)) \odot \text{Gate}(z, t) + \text{Bias}(z, t) \quad (6)$$

As shown in Equation (6), the output of the module is obtained by applying element-wise multiplication between the transformed input and the gate, followed by the addition of the bias.

Following the training strategy adopted in DPM, we randomly sample the time step  $t \in (0, T)$  during each training iteration to improve efficiency and ensure temporal diversity. Once the predicted noise  $\hat{\epsilon}_\theta$  is obtained, the point cloud at the previous time step  $x_{t-1}$  can be sampled using the denoising equation defined in Equation (3). This reverse sampling process is iteratively performed from the sampled  $t$  down to  $t=0$ , progressively refining the point cloud until a final clean output  $x_0$  is generated.

## 5. RESULTS

The synthetic dataset for Polyfytos Bridge is used for training. The dataset is split into training, validation, and test sets with a ratio of 7:2:1. Following standard practices in denoising diffusion probabilistic models, we set the number of diffusion steps to  $T = 1000$ , and use a learning rate of 0.002 and a batch size of 16 with the Adam optimiser. Due to the simplicity of the loss function and the stable convergence behaviour of diffusion models, the training process exhibits rapid initial convergence, as illustrated in Figure 10. Nevertheless, the model is trained for 2000 epochs to ensure sufficient learning of fine-grained geometric structures under the conditional setting. Although the mean squared error (MSE) loss stabilises early as a result of averaging across diffusion timesteps, extended training is essential for improving the fidelity and consistency of the generated point clouds, particularly in tasks involving partial-to-complete shape recovery.

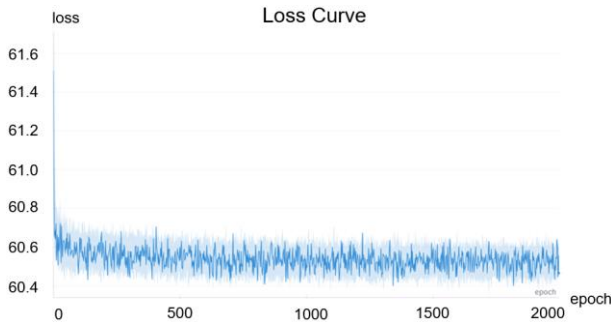


Figure 10: Loss curve of the conditional diffusion model.

All experiments are performed on an NVIDIA A100 GPU with 40 GB memory, and the total training time is approximately 8 hours.

Subsequently, real-world scanned point clouds are used as input for case study evaluation. In both the training and inference phases, all input point clouds are normalised to ensure consistency and improve learning stability. The completion results for the five types of components are presented in Figure 11.

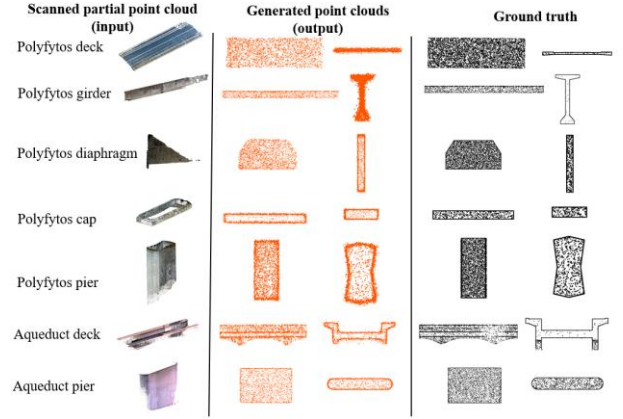


Figure 11: Comparison between generated point clouds and ground truth for different bridge components.

Overall, the model is capable of inferring the underlying shape from incomplete inputs and successfully reconstructs complete point clouds with uniform density. For geometrically complex components, such as I-shaped girders, the network effectively preserves local structural details. Importantly, as a generative model, it is capable of producing an arbitrary number of points, thereby eliminating the need for post-processing steps such as upsampling commonly required by traditional completion models. Although some sharp corners may not be fully generated, the reconstructed shapes remain globally and locally consistent.

To evaluate the quality of the completed point clouds, Chamfer Distance (CD) and Earth Mover's Distance (EMD) are employed as evaluation metrics. The corresponding formulations are given below:

$$\text{CD}(P, Q) = \frac{1}{|P|} \sum_{x \in P} \min_{y \in Q} |x - y|_2 + \frac{1}{|Q|} \sum_{y \in Q} \min_{x \in P} |x - y|_2 \quad (7)$$

$$\text{EMD}(P, Q) = \min_{\phi: P \rightarrow Q} \frac{1}{|P|} \sum_{x \in P} |x - \phi(x)|_2 \quad (8)$$

Where  $P$  denotes the generated point cloud and  $Q$  denotes the corresponding ground truth point cloud. Given the high computational cost of EMD, Farthest Point Sampling (FPS) is applied to uniformly downsample both the ground truth and the generated point clouds to 2048 points prior to the computation of CD and EMD. The quantitative results are summarised in Table 1. For both metrics, lower values indicate better alignment between the

predicted and ground truth point clouds, reflecting higher completion accuracy.

Table 1: Ablation Experiment.

Metrics (*10 <sup>-3</sup> )	Baseline		Modified Encoder	
	CD	EMD	CD	EMD
deck 1	1.925	0.893	<b>1.864</b>	<b>0.848</b>
girder	1.332	1.227	<b>1.322</b>	<b>0.898</b>
pier 1	4.747	2.251	<b>4.202</b>	<b>1.697</b>
diaphragm	2.514	0.768	<b>2.294</b>	<b>0.764</b>
cap	8.691	5.625	<b>7.731</b>	<b>4.298</b>
pier 2	1.968	0.794	<b>1.512</b>	<b>0.365</b>
deck 2	2.879	1.288	<b>2.651</b>	<b>0.769</b>

The results show that our modified PointCNN encoder consistently outperforms the baseline across all component categories in both metrics. For components with relatively uniform geometry, such as the diaphragm and cap, the proposed encoder yields more noticeable improvements in both CD and EMD, demonstrating its effectiveness in capturing consistent local structures. In contrast, for components with strongly directional or elongated shapes, such as girders and decks, the improvements are less pronounced, likely due to the structural extremity along a particular axis.

Nevertheless, it is worth noting that for geometrically complex components, the modified encoder still achieves performance gains. This suggests that the model is capable of effectively balancing global shape reconstruction with the preservation of fine-grained local features.

To evaluate the effect of the hyperparameter  $\lambda$  in the loss function, a sensitivity analysis is conducted (Table 2).

Table 2: Sensitivity analysis of the loss hyperparameter  $\lambda$  with respect to Earth Mover’s Distance (EMD) and Chamfer Distance (CD).

Hyperparameter $\lambda$	Component: Bridge Pier	
	CD (*10 <sup>-3</sup> )	EMD (*10 <sup>-3</sup> )
0.05	4.397	2.581
<b>0.08</b>	4.202	<b>1.697</b>
0.1	4.537	2.094
0.15	4.073	2.570

The results indicate that when  $\lambda$  is set to 0.08, the Earth Mover’s Distance (EMD) reaches its minimum value, while the Chamfer Distance (CD) is also close to its lowest value. Therefore,  $\lambda = 0.08$  is considered a reasonable and effective setting.

## 6. CONCLUSION

This study presents a conditional diffusion-based framework for the completion of 3D bridge point clouds, addressing the challenges posed by

incomplete and irregular real-world data. By formulating the task as a conditional generation problem and leveraging a point cloud encoder to guide the diffusion process, the proposed method demonstrates strong generative capability, producing high-fidelity and structurally consistent point clouds. In addition, a dataset construction strategy is introduced to simulate realistic scanning defects, providing a practical basis for training and evaluation. Experimental results on real bridge data confirm the effectiveness and flexibility of the approach, particularly in handling large-scale infrastructure with complex geometries. This work lays the foundation for further integration of diffusion models into automated digital reconstruction pipelines for civil infrastructure.

While the proposed method demonstrates potential, further development is required to improve its accuracy. Moreover, due to the limited availability of publicly accessible bridge datasets, the current study is restricted to commonly encountered simply supported concrete girder bridges. This limitation reduces the generalisability of the model, meaning that additional training is required when adapting it to other bridge types. Furthermore, the generation of geometrically complex components, such as decks with varying heights, often lacks sufficient detail. Compared to the ground truth, the generated point clouds tend to lack sharp edges and clearly defined corners.

Future research will aim to improve the model’s accuracy and generalisability by expanding the dataset to include a broader range of bridge types, such as continuous bridges. Additionally, attention mechanisms may be integrated into the model to improve the capture of both local and global features. Improvements to the loss function will also be explored, including the incorporation of engineering prior knowledge, with the aim of increasing the structural fidelity and detail of the generated point clouds.

## REFERENCES

- Charles, R.Q., Su, H., Kaichun, M., Guibas, L.J., 2017. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, pp. 77–85. <https://doi.org/10.1109/CVPR.2017.16>
- Esmoris, A.M., Yermo, M., Weiser, H., Winiwarter, L., Höfle, B., Rivera, F.F., 2022. Virtual LiDAR Simulation as a High Performance Computing Challenge: Toward HPC HELIOS++. IEEE Access 10, 105052–105073. <https://doi.org/10.1109/ACCESS.2022.3211072>
- Fan, H., Su, H., Guibas, L., 2017. A Point Set Generation Network for 3D Object Reconstruction from a Single Image, in: 2017 IEEE Conference on Computer Vision

- and Pattern Recognition (CVPR). Presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, pp. 2463–2471. <https://doi.org/10.1109/CVPR.2017.264>
- Fang, Y., Mitoulis, S.-A., Boddice, D., Yu, J., Ninic, J., 2025. Scan-to-BIM-to-Sim: Automated reconstruction of digital and simulation models from point clouds with applications on bridges. *Results in Engineering* 25, 104289. <https://doi.org/10.1016/j.rineng.2025.104289>
- He, Z., Li, W., Salehi, H., Zhang, H., Zhou, H., Jiao, P., 2022. Integrated structural health monitoring in bridge engineering. *Automation in Construction* 136, 104168. <https://doi.org/10.1016/j.autcon.2022.104168>
- Ho, J., Jain, A., Abbeel, P., 2020. Denoising Diffusion Probabilistic Models, in: *Advances in Neural Information Processing Systems*. Curran Associates, Inc., pp. 6840–6851.
- Huang, Z., Yu, Y., Xu, J., Ni, F., Le, X., 2020. PF-Net: Point Fractal Network for 3D Point Cloud Completion, in: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Presented at the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Seattle, WA, USA, pp. 7659–7667. <https://doi.org/10.1109/CVPR42600.2020.00768>
- Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B., 2018. PointCNN: Convolution On X-Transformed Points, in: *Advances in Neural Information Processing Systems*. Curran Associates, Inc.
- Liu, M., Sheng, L., Yang, S., Shao, J., Hu, S.-M., 2020. Morphing and Sampling Network for Dense Point Cloud Completion. *AAAI* 34, 11596–11603. <https://doi.org/10.1609/aaai.v34i07.6827>
- Lu, R., Brilakis, I., Middleton, C., 2018. Detection of Structural Components in Point Clouds of Existing RC Bridges. <https://doi.org/10.1111/mice.12407>
- Luo, S., Hu, W., 2021. Diffusion Probabilistic Models for 3D Point Cloud Generation, in: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Presented at the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Nashville, TN, USA, pp. 2836–2844. <https://doi.org/10.1109/CVPR46437.2021.00286>
- Lyu, Z., Kong, Z., Xu, X., Pan, L., Lin, D., 2021. A Conditional Point Diffusion-Refinement Paradigm for 3D Point Cloud Completion. Presented at the International Conference on Learning Representations.
- Matono, G., Nishio, M., 2024. Component-level point cloud completion of bridge structures using deep learning. *Comput.-Aided Civ. Infrastruct. Eng.* 39, 2581–2595. <https://doi.org/10.1111/mice.13218>
- Melas-Kyriazi, L., Rupprecht, C., Vedaldi, A., 2023. PC2: Projection-Conditioned Point Cloud Diffusion for Single-Image 3D Reconstruction, in: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Presented at the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Vancouver, BC, Canada, pp. 12923–12932. <https://doi.org/10.1109/CVPR52729.2023.01242>
- Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space, in: *Advances in Neural Information Processing Systems*. Curran Associates, Inc.
- Rubner, Y., Tomasi, C., Guibas, L.J., 2000. The Earth Mover's Distance as a Metric for Image Retrieval. *International Journal of Computer Vision* 40, 99–121. <https://doi.org/10.1023/A:1026543900054>
- Shi, M., Kim, H., Narazaki, Y., 2024. Development of large-scale synthetic 3D point cloud datasets for vision-based bridge structural condition assessment. *Advances in Structural Engineering* 27, 2901–2928. <https://doi.org/10.1177/13694332241260077>
- Winiwarter, L., Esmoris Pena, A.M., Weiser, H., Anders, K., Martínez Sánchez, J., Searle, M., Höfle, B., 2022. Virtual laser scanning with HELIOS++: A novel take on ray tracing-based simulation of topographic full-waveform 3D laser scanning. *Remote Sensing of Environment* 269, 112772. <https://doi.org/10.1016/j.rse.2021.112772>
- Xiang, P., Wen, X., Liu, Y.-S., Cao, Y.-P., Wan, P., Zheng, W., Han, Z., 2021. SnowflakeNet: Point Cloud Completion by Snowflake Point Deconvolution with Skip-Transformer, in: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Presented at the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Montreal, QC, Canada, pp. 5479–5489. <https://doi.org/10.1109/ICCV48922.2021.00545>
- Yang, X., del Rey Castillo, E., Zou, Y., Wotherspoon, L., Tan, Y., 2022. Automated semantic segmentation of bridge components from large-scale point clouds using a weighted superpoint graph. *Automation in Construction* 142, 104519. <https://doi.org/10.1016/j.autcon.2022.104519>
- Yuan, W., Khot, T., Held, D., Mertz, C., Hebert, M., 2018. PCN: Point Completion Network, in: *2018 International Conference on 3D Vision (3DV)*. Presented at the 2018 International Conference on 3D Vision (3DV), pp. 728–737. <https://doi.org/10.1109/3DV.2018.00088>
- Zhou, L., Du, Y., Wu, J., 2021. 3D Shape Generation and Completion through Point-Voxel Diffusion, in: *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Presented at the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Montreal, QC, Canada, pp. 5806–5815. <https://doi.org/10.1109/ICCV48922.2021.00577>