RoofDiffusion: Constructing Roofs from Severely Corrupted Point Data via Diffusion

Kyle Shih-Huang Lo[1], Jörg Peters[1], and Eric Spellman[2]

[1] University of Florida, Gainesville FL 32611, USA
[2] Meta Platforms, Inc.
kyleshihhuanglo@ufl.edu, jorg@cise.ufl.edu, espellman@meta.com

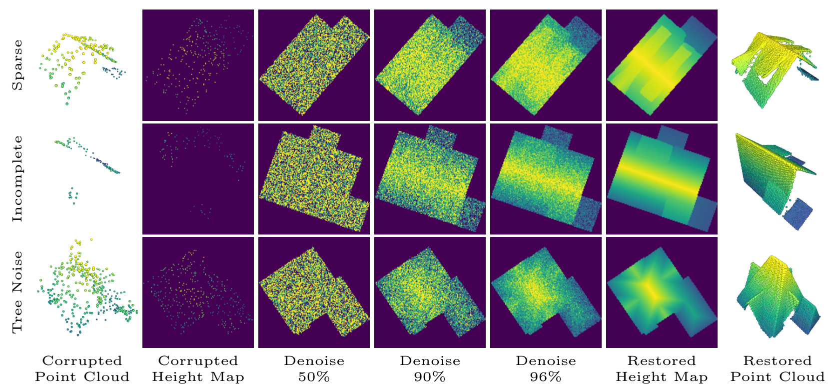
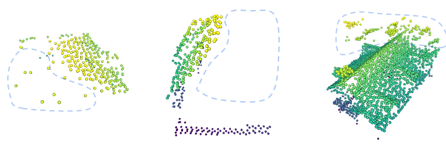


Fig. 1: RoofDiffusion restores height maps of challenging roof geometry, even under conditions of extreme sparsity, regional incompleteness, and noise. The bookend columns of point clouds are 3D views, the other columns are top views of height maps.

Abstract. Accurate completion and denoising of roof height maps are crucial to reconstructing high-quality 3D buildings. Repairing sparse points can enhance low-cost sensor use and reduce UAV flight overlap. RoofDiffusion is a new end-to-end self-supervised diffusion technique for robustly completing, in particular difficult, roof height maps. RoofDiffusion leverages widely-available curated footprints and can so handle up to 99% point sparsity and 80% roof area occlusion (regional incompleteness). A variant, No-FP RoofDiffusion, simultaneously predicts building footprints and heights. Both quantitatively outperform state-of-the-art unguided depth completion and representative inpainting methods for Digital Elevation Models (DEM), on both a roof-specific benchmark and the BuildingNet dataset. Qualitative assessments show the effectiveness of RoofDiffusion for datasets with real-world scans including AHN3, Dales3D, and USGS 3DEP LiDAR. Tested with the leading City3D algorithm, preprocessing height maps with RoofDiffusion noticeably improves 3D building reconstruction. RoofDiffusion is complemented by a new dataset of 13k complex roof geometries, focusing on long-tail issues in remote sensing; a novel simulation of tree occlusion; and a wide variety of large-area roof cut-outs for data augmentation and benchmarking. Code and dataset[3]: github.com/kylelo/RoofDiffusion

[3] Created and released by the University of Florida



(a) Sparse (b) Incomplete (c) Tree Noise

Fig. 2: Types of corrupted roof height maps with real-world scan.

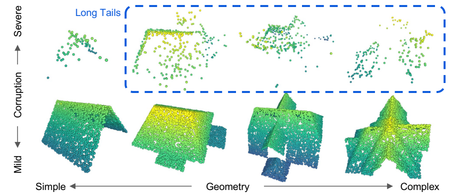


Fig. 3: RoofDiffusion focuses on the "long tail" of hard-to-handle complex, corrupted geometry (inside the dashed blue).

1 Introduction

Digital Surface Models (DSMs), a.k.a. *height maps*, are monochromatic images where the pixel value captures the elevation of features, both natural and artificial. Height maps can be created by rasterizing airborne Light Detection and Ranging (LiDAR) point clouds. Height maps serve as an indispensable data source for reconstructing 3D models of urban buildings. Numerous studies [23, 24, 46, 82] have explored methods for generating compact 3D building models from such data. 3D building models are crucial in a wide range of applications, including navigation [5], urban planning [39], and simulations [28, 48]. With 500M buildings labeled in OpenStreetMap [49], even a 1% failure rate implies that 5M buildings need to be repaired. A close scrutiny of USGS 3DEP LiDAR sampled over Wayne County, MI [69] and Cambridge, MA [70] reveals, respectively, as much as 34% and 50% of height maps corrupted by incompleteness or noise. While height maps are broadly accessible, the following challenges often impede compact and precise building reconstructions.

- **Sparsity.** Factors such as low sensor resolution and poor surface reflectance cause low point density [80]. (Fig. 2a)
- **Incompleteness.** Portions of the roof data can be missing, due to environmental interference, occlusions by taller surrounding objects, or roof substructure when the camera angle is not orthogonal to the ground [66]. (Fig. 2b)
- **Noise.** Intrusions on building footprints, such as trees, can lead to incorrect reconstruction, resulting in artifacts like non-existing dormers. (Fig. 2c). Environmental factors like light and dust can also introduce noise.

Since the reconstruction quality of the 3D models is heavily influenced by the precision and completeness of the underlying height maps, roof repair is crucial, but is under-explored.

RoofDiffusion addresses all three bulleted challenges by building on and adjusting reliable techniques. Unlike typical image inpainting tasks such as [1, 12, 35, 37, 42], available height map pixels can be noisy, and completing sparse data does not fit the standard super-resolution paradigms [19, 29, 38, 88], because missing pixels are usually unevenly distributed. Acknowledging these complexities, we nevertheless, succeed in conceptualizing roof repair as an image restoration task.

Thanks to the efforts of contributors worldwide, OpenStreetMap [49] now increasingly provides vectorized high-precision footprints [20, 24]. Consequently, recent building reconstruction algorithms [2, 20, 24] have incorporated footprints into the reconstruction process. However, these approaches still struggle with sparsity, incompleteness and noise. Footprint-guided RoofDiffusion addresses these long-tail challenges (Fig. 3).

Specifically, we learn a strong prior that accurately approximates the actual distribution of the roof height map. Taking inspiration from the restoration approach in [61] of Joint Photographic Experts Group (JPEG) images, we frame the roof height map restoration challenge as a denoising diffusion process. While JPEG restoration focuses on reverting compressed images to their original state, our strategy seeks to fill in roof heights and remove noise. The following are the major contributions of this paper:

- We introduce RoofDiffusion, a novel method based on conditional Diffusion Probabilistic Model (DPM) [61]. Both footprint-guided RoofDiffusion and a variant, No-FP RoofDiffusion, trained without footprint, allow robust roof height map completion under extreme conditions, surpassing the state-of-the-art depth completion methods [13, 75] and representative DEM inpainting algorithms [3, 27, 65]. The footprint-guided version handles up to 99% missing data points and 80% regional incompleteness while remaining resilient to tree occlusion noise (for some applications, construction from very few data might be flagged for hallucination potential). No-FP RoofDiffusion predicts both footprint and height.
- We propose a novel "tree planting" method for simulating tree occlusion noise, and we introduce multi-Gaussian masking for synthesizing incompleteness in roof height maps. These techniques enable data augmentation for self-supervised learning and benchmark creation for quantitative comparisons.
- We unveil the PoznanRD (Poznan Roof Dataset) with 13k Level of Detail (LoD) 2.2 [4] noise-free roof meshes and height maps. When treated with our noise and incompleteness algorithm, our complex roofs dataset can effectively generate ample training data to address "long tail" challenges, see Fig. 3.

2 Related Works

Reviewing Digital Elevation Model (DEM) inpainting from remote sensing is most relevant, followed by unguided depth completion methods for restoring dense, clean depth maps from sparse LiDAR points. Also, we review denoising diffusion models in computer vision tasks and roof datasets.

DEM Inpainting restores terrain height data missing caused by occlusions, such as mountains obstructing their own opposite sides and areas covered by water. Researchers commonly use Inverse Distance Weighting (IDW) [65], Kriging [43, 57], and Spline fitting [27] for inpainting voids in DEMs. These methods work well for small areas but can fail in larger, complex regions due to the lack of terrain geometry knowledge. To address larger inpainting regions, Delta surface-based approaches [18,41] fill voids with auxiliary DEM patches. Learning-based methods have gained prominence in DEM inpainting due to their superior feature learning capabilities. In particular, several methods [10,11,16,34,53,79,87,90,91] utilize Generative Adversarial Nets (GAN) [17,45] for inpainting by conditioning models on voids to predict filled areas.

Compared to DEM inpainting, roof DSM inpainting deals with higher sparsity, more noise from elements such as trees, and larger areas of incompleteness.

Unguided Depth Completion aims to restore the dense depth map solely from sparse depth data, mainly for urban driving data [72]. The traditional Convolutional Neural Networks (CNNs) [30] often suffer from a mosaic effect of images with sparse pixel values. This issue is addressed by applying convolutions solely to valid pixels using a binary mask [25, 72]. However, a binary mask uniformly weights each pixel, conflicting with the reality that pixels have varying importance. To counter this, a continuous confidence mask was proposed in [14]. Furthermore, a learnable mask is introduced in [13] to address the issue that directly inferring a confidence mask from data may be problematic due to noise. To incorporate semantic learning, the works [40, 83] train the network to additionally predict auxiliary images, including RGB and normal maps. CU-Net [75] employs a two-stage U-Net for coarse-to-fine depth completion.

Most depth completion methods target evenly distributed, sparse depth maps without large missing areas. RoofDiffusion can handle height maps missing more than 80% of the area and still generate inpainting harmonious with the existing roof structure.

Diffusion Models have been proposed in pioneering works [9,21,68] showing remarkable image generation capability. Subsequent research has built upon this foundation by introducing text-conditioned models, enabling text-guided image generation [47, 54, 59, 60, 62, 89]. The technology has further been employed to significantly improve image enhancement tasks like super-resolution [15, 22, 63, 77], inpainting [42, 61, 78], and translation [32, 44, 50, 61]. There are also a few diffusion model-based methods for RGB-guided depth completion [31, 55].

To the best of our knowledge, our work pioneers the use of the diffusion model for DSM completion. We formulate our problem as an image restoration task, leveraging a conditional DPM. Inspired by [26], our approach eliminates the need for a confidence map [13,14,72], directly conditioning the model on the sparse image.

Roof Datasets commonly provide pairs of point clouds and ground truth mesh for 3D reconstruction research. UrbanScene3D [36] and STPLS3D [6] offer real-world LiDAR point clouds along with reconstructed dense triangular meshes. City3D [24] and Building3D [74] provide substantial datasets featuring more compact meshes for Computer-Aided Design (CAD)-like building-reconstruction research. However, in these datasets where ground truth is constructed via 3D reconstruction algorithms, mesh accuracy can be suboptimal due to algorithmic limitations and real-world scanning noise. Government datasets

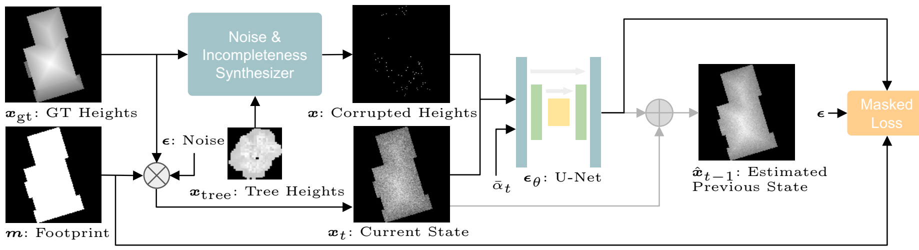


Fig. 4: *Training the RoofDiffusion reverse transition kernel.* GT denotes ground truth, $\bigotimes$ is transformation (3), $\bigoplus$[4] removes predicted noise from x_t by (4). The masked loss is defined by Eq. (5).

like [7, 8] provide compact noise-free meshes but exhibit misalignment with point clouds, often due to modeling simplifications. To ensure alignment between compact meshes and point clouds, BuildingNet [64] generates point clouds directly from compact meshes crafted by artists. However, BuildingNet [64] provides only 2k buildings and contains non-building elements such as people and vehicles. Similarly, we can sample point clouds from datasets with compact meshes, [52, 56, 76], but these have limited roof types.

Considering the limitations in quantity and variety of compact meshes, we propose the PoznanRD, featuring 13k noise-free complex roofs. Besides, existing datasets do not focus on corrupted data, so we introduce a corruption synthesizing approach to generate challenging data.

3 Problem Statement

Given a noisy, sparse, and incomplete roof height map z along with a building footprint m, the goal of RoofDiffusion is to estimate a complete and noise-free height map $\hat{z}$ approximating to the ground truth z_{gt}. Measured from ground level, z, $\hat{z}$, and z_{gt} are single-channel images whose pixel values indicate building heights in meters. In z, zero-valued pixels signify either missing measurements or ground level. For $\hat{z}$ and z_{gt}, zero values represent ground level. Generally, the number of non-zero pixels in $\hat{z}$ and z_{gt} is significantly greater than in z. The mask m delineates the footprint, i.e. equals 1 for pixels belonging to the roof, and 0 otherwise.

Accurate reconstruction of real-world height maps is challenging due to a wide range of building heights, noise, and missing data. While the typical normalization for images or depth scenes of autonomous driving can employ fixed min-max values, the normalization required for roof heights needs to adapt to the wide range of roof heights for each building.

[4] $\hat{x}_{t-1}$ calculation can be ignored during training (the arrows colored as gray).

4 RoofDiffusion

We introduce a diffusion model-based roof completion method to reconstruct $\hat{\boldsymbol{z}}$ from $\boldsymbol{z}$ so that $\hat{\boldsymbol{z}}$ is close to $\boldsymbol{z}_{\text{gt}}$. Section 4.1 presents the conversion of input height maps $\boldsymbol{z}$ to roof-focused height images $\boldsymbol{x}$ in the range $[-1, 1]$ suitable for diffusion models. Section 4.2 applies a diffusion model, conditioned on corrupted normalized height maps $\boldsymbol{x}$ within their building footprint $\boldsymbol{m}$, to predict complete and noise-free height maps $\hat{\boldsymbol{x}}_0$.

4.1 Roof-Focused Height Map Normalization

Diverse building heights and the fixed value range of diffusion models require a careful normalization approach. To focus solely on the roof structure, we identify the lowest roof pixel and subtract this value from the entire height map. Let δ_i be the difference between the maximum and minimum height of the roof structure for i-th building . Then $\underline{\overline{z}}$ is the maximum of the δ_i after removing the largest 1% of δ_i. The normalized input height image is

$$\boldsymbol{x} := \frac{2}{\underline{\overline{z}}}\left(\boldsymbol{z} - \frac{1}{2}\left(\boldsymbol{z}_{\max} + \boldsymbol{z}_{\min}\right)\right) \in [-1, 1], \tag{1}$$

where $\boldsymbol{z}_{\min}$ denotes the smallest nonzero value in $\boldsymbol{z}$ and $\boldsymbol{z}_{\max}$ the largest. Our analysis of 13k buildings shows the cut-off value for $\underline{\overline{z}}$ is 10 meters.

4.2 Height Completion based on a Diffusion Model

Given $\boldsymbol{x}$, our goal is to generate a predicted image, $\hat{\boldsymbol{x}}_0$, that closely approximates the normalized ground truth image, $\boldsymbol{x}_{\text{gt}}$. We conceptualize this height completion task as a diffusion process. Drawing on the work of [21], our approach employs both a *forward* and a *reverse* Markov chain. The forward chain perturbs the ground truth data through noise injection, while the reverse chain utilizes a learnable model, conditioned on the input, to restore the data, see Fig. 4.

The Forward Process transforms a complex distribution $q(\boldsymbol{x}_{\text{gt}})$, the ground truth height maps, into the Gaussian distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. Starting with $\boldsymbol{x}_0 = \boldsymbol{x}_{\text{gt}} \sim q(\boldsymbol{x}_{\text{gt}})$, a transition kernel $q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$ generates a series of random variables $\boldsymbol{x}_t$, $t \in \{1, \ldots, T\}$. The joint distribution of $q(\boldsymbol{x}_1, ..., \boldsymbol{x}_T \mid \boldsymbol{x}_0)$, namely $\prod_{t=1}^{T} q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1})$, is marginalized based on [67] to derive closed-form expression

$$\boldsymbol{x}_t := f_{\text{forward}}\left(t, \boldsymbol{x}_0, \boldsymbol{\epsilon}, \boldsymbol{m}\right) \tag{2}$$

$$= \boldsymbol{m} \odot \left(\sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1 - \bar{\alpha}_t}\boldsymbol{\epsilon}\right) - \boldsymbol{m}', \tag{3}$$

where $\odot$ denotes element-wise multiplication, α_i is a hyperparameter, $\bar{\alpha}_t = \prod_{i=0}^{t} \alpha_i$, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. Subtracting $\boldsymbol{m}'$, the complement mask of $\boldsymbol{m}$, sets pixel values outside the footprint to -1 for two advantages. First, the model can infer the building footprint directly from the noise-injected areas in $\boldsymbol{x}_t$, obviating the need for an additional channel to represent the footprint. Second, non-building heights are prevented from influencing the prediction.

The Reverse Process sequentially removes noise from the data to yield a complete, noise-free version. Specifically, we can draw a noisy image $\boldsymbol{x}_T$ from $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$ at time $t = T$. Then use a reverse transition kernel to recursively eliminate noise until $t = 0$, resulting in the restored height image, $\hat{\boldsymbol{x}}_0$. This transition kernel is parameterized by a learnable model $\boldsymbol{\epsilon}_\theta$. While $\hat{\boldsymbol{x}}_0$ conforms to $q(\boldsymbol{x}_{\mathrm{gt}})$, it may not always correspond to the normalized corrupted height image, $\boldsymbol{x}$. In our case, the objective is to ensure that the repaired roofs seamlessly integrate with pre-existing structures. To ensure a close association between $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}_0$, we condition $\boldsymbol{\epsilon}_\theta$ on $\boldsymbol{x}_t$, $\boldsymbol{x}$, and $\bar{\alpha}_t$, in the spirit of [61]. Each step of the reverse process can be formulated as

$$\hat{\boldsymbol{x}}_{t-1} \leftarrow \frac{1}{\sqrt{\alpha_t}} \left(\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta \right) + \sqrt{1-\alpha_t} \boldsymbol{\epsilon}_t, \tag{4}$$

where α_t and $\bar{\alpha}_t$ are hyperparameters for variance scheduling, and $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$.

Loss Function. For accurate noise prediction at various time steps t we train $\boldsymbol{\epsilon}_\theta$ with L_1 loss restricted to the footprint $\boldsymbol{m}$. For $t \sim \mathcal{U}(1, T)$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$,

$$L := \mathbb{E}_{(\boldsymbol{x}_0, \boldsymbol{x}, \boldsymbol{m}), t, \boldsymbol{\epsilon}} \| \boldsymbol{m} \odot (\boldsymbol{\epsilon} - \tilde{\boldsymbol{\epsilon}}_\theta) \|_1 \tag{5}$$

$$\text{where } \tilde{\boldsymbol{\epsilon}}_\theta := \boldsymbol{\epsilon}_\theta \left(f_{\text{forward}} \left(t, \boldsymbol{x}_0, \boldsymbol{\epsilon}, \boldsymbol{m} \right), \boldsymbol{x}, \bar{\alpha}_t \right). \tag{6}$$

To restore the original scaling in meters for the minimum and maximum extent of a roof, we reverse (1), i.e. multiply $\hat{\boldsymbol{x}}_0$ by $\frac{1}{2}\overline{\boldsymbol{z}}$ and add $\frac{1}{2}(\boldsymbol{z}_{\min} + \boldsymbol{z}_{\max})$.

5 Datasets & Benchmarks

Simulating the real-world requires noisy, sparse, and incomplete height maps. Ground truth requires a high-resolution, noise-free, complete height map; and to tackle long-tail geometric complexity illustrated in Fig. 3, the buildings should be more complex than existing datasets with a LoD 2.2 [4] (LoD 2.2 captures intricate roof details like dormers and gables, that LoD 2.0 overlooks).

To aid machine learning research, we curated a new dataset that not only serves our research needs, but future research needs in the field. We started with 16k compact and high-detail LoD 2.2 roof meshes from the city of Poznan, Poland [8]. To match our focus on complex roof structures (Fig. 3), we re-balanced the dataset by reducing the number of flat roof from 5k to 2k resulting in a new dataset of 13k buildings.

To address the long-tail corruption depicted in Fig. 3, collecting diverse corrupted height maps is essential. However, acquiring real-world corrupted data is challenging due to the difficulty in obtaining ground truths and identifying corrupted height maps. Therefore, we developed a procedure for synthesizing intentionally corrupted height maps suitable for training or benchmarking, as

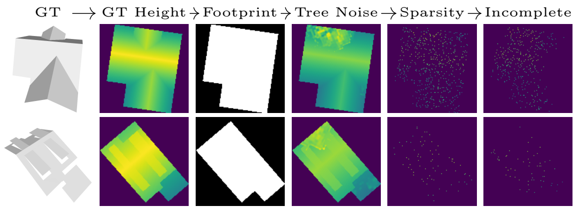


Fig. 5: Examples of synthesizing corrupted height maps from the PoznanRD.

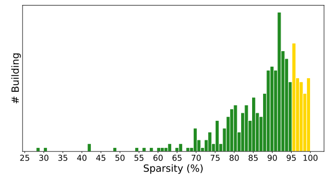


Fig. 6: Sparsity of the real-world USGS 3DEP LiDAR dataset [69, 70].

illustrated in Fig. 5. We first rasterize the roof mesh triangles into ground truth height maps. This process provides complete height maps without missing points and also reconciles real-world captured height maps and artist-generated meshes. Next, footprints are inferred from non-zero pixels in height maps. Then, we simulate real-world conditions by adding synthetic tree noise[5], creating sparsity through random point removal, and injecting incompleteness.

In this paper, "s" denotes Sparsity (%), the ratio of randomly removed to total pixels in the footprint. Similarly, "i" signifies the Incompleteness (%), the proportion of pixels removed due to incompleteness in the footprint.

5.1 Tree Points

In analyzing height maps within a footprint, tree crowns may intrude, causing LiDAR to capture the tree canopy instead of the roof, see Fig. 2c. Thus, we simulate tree intrusion by planting virtual trees around the roof. We collected a real-world database with 1k tree height maps from [8]. Small trees are randomly placed outside the footprint, with random rotations and height adjustments. A "max" operation is used to combine the roof $\boldsymbol{x}$ and tree height map $\boldsymbol{x}_{\text{tree}}$. Algorithm 3 in supplements has the details.

5.2 Incompleteness

To synthesize incompleteness, we employ multiple Gauss masks with various means and variances to indicate missing points. This approach emulates occlusions caused by more intricate objects with soft (Fig. 2a) or hard (Fig. 2b) boundaries, in the spirit of a Gaussian Mixture Model [58]. The rationale behind using Gauss masks is as follows: When multiple Gaussian masks are randomly positioned on the same side of a building, they can completely block one side of the roof point clouds, akin to the ray-tracing-free simulation of occlusion by [84], where the authors remove points distant from the camera. Conversely, when scattered and smaller Gaussian masks are utilized within the roof structure, they effectively simulate the occlusion caused by small roof features. Lastly, a larger variance of Gaussian simulates softer boundaries. The supplement details efficient mask generation for training in Algorithm 1, and mask creation with a specific portion of missing points in Algorithm 2.

[5] Although we focus on tree noise for clarity, we also simulate global Gaussian noise and outlier noise in practice.

5.3 Benchmarks

The above height map synthesizing technique allows us to create benchmarks tailored to various research needs. Specifically, we can adjust different levels of sparsity, incompleteness, and the number of trees introduced. This makes it ideal for both training and evaluating machine learning models geared towards 3D reconstruction and other related fields. We have partitioned the dataset into a training split of 10k samples and a test split of 3k samples. Additionally, for tree-height maps, we provide 766 samples for training and 255 for testing.

6 Experiments

Section 6.1 demonstrates the effectiveness of RoofDiffusion and No-FP RoofDiffusion, trained without footprint, by quantitative evaluation on two datasets with ground truth: our PoznanRD and the BuildingNet [64]. Section 6.2 illustrates the ability of No-FP RoofDiffusion to recover footprints from corrupted height maps. Section 6.3 assesses the 3D mesh reconstruction quality for the City3D [24], comparing results with, and without using RoofDiffusion as a pre-processor. Section 6.4 showcases qualitative outcomes through tests on real-world scans in AHN3 [24], Dales3D [73], and USGS 3DEP LiDAR sampled over Cambridge, MA [70], and Wayne County, MI [69]. Section 6.5 discusses the limitations.

The supplement provides ablation of tree augmentation, sampling step analysis, impact of LiDAR scan pattern, additional qualitative results, and implementation details. Additionally, the supplement shows an extension of RoofDiffusion to be comparable to existing unguided depth completion methods on the KITTI dataset [72].

6.1 Quantitative Evaluation of Height Completion

We examine RoofDiffusion in the four relevant scenarios. Footprints (FP) are derived by converting the ground truth height map into a binary mask.

1. (FP) *Both* sparsity *and* incompleteness on PoznanRD in Tab. 1.
2. (FP) *Either* sparsity *or* incompleteness on BuildingNet [64] in Tab. 1.
3. (No-FP) *Both* sparsity *and* incompleteness in the PoznanRD in Tab. 2.
4. (No-FP) *Either* sparsity *or* incompleteness on BuildingNet [64] in Tab. 2.

PoznanRD Dataset. We conducted tests on 1k randomly selected data from test split of PoznanRD, injected with both global and outlier noise to simulate real-world cases. We follow [33, 81, 85] to emulate global noise by incorporating Gaussian noise into all the normalized data points, with $\sigma_{\text{global}} \sim [0, 0.05]$ for each height map. Also, outliers are introduced by randomly assigning nonzero pixels with random values, as suggested in [86], where the probability for occurrence of an outlier is set to 0.01 %. Additionally, we introduced tree noise in 30% of the cases, involving between 1 to 3 trees, following Sec. 5.1.

BuildingNet [64] contains compact and noise-free building meshes spanning various categories. These models are mostly sourced from 3D artists and often

Methods	PoznanRD								BuildingNet							
	s95 i30		s95 i80		s99 i30		s99 i80		s90 i80		s90 i90		s99 i0		s99.75 i0	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Linear	0.236	0.461	0.687	1.037	0.365	0.631	0.868	1.218	0.654	0.997	0.868	1.224	0.297	0.549	0.528	0.833
Spline	0.278	0.508	0.785	1.198	0.391	0.659	0.888	1.260	0.829	1.288	1.033	1.523	0.330	0.586	0.536	0.845
Nearest	0.288	0.541	0.691	1.089	0.424	0.734	0.866	1.271	0.634	1.044	0.856	1.287	0.347	0.662	0.582	0.961
IDW	0.239	0.449	0.648	0.984	0.377	0.619	0.827	1.172	0.573	0.897	0.808	1.159	0.309	**0.540**	0.537	0.822
P.M. Diff.	0.266	0.473	1.825	2.311	0.523	0.739	3.085	3.548	1.090	1.448	1.775	2.136	0.414	0.706	0.743	0.979
Ours	**0.162**	**0.342**	**0.430**	**0.727**	**0.253**	**0.463**	**0.603**	**0.916**	**0.508**	**0.871**	**0.705**	**1.071**	**0.280**	0.550	**0.447**	**0.751**

Table 1: Evaluation of height map completion on PoznanRD and BuildingNet [64] dataset, measured by Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) in meters. **Bold** represents best outcome.

include additional elements like virtual ground, cars, and trees in the scene. We picked 200 buildings that contain only the building structure, making it easier to infer footprints. We rasterized the meshes in BuildingNet [64] into height maps as ground truths. It is important to note that our tests were conducted using a model trained exclusively on PoznanRD.

(FP) ***Both*** **sparsity** ***and*** **incompleteness.** We evaluated combinations of sparsity at 95% and 99%, along with incompleteness at 30% and 80%, using the method shown in Sec. 5.2. Notably, our sparsity selection is according to USGS 3DEP LiDAR with Quality Level (QL) 2. Note that QL2 only guarantees at least 2 points per square meter and that most of the US territory LiDAR is QL 2 or lower. In fact, it is common to find roof data with extreme sparsity, as shown as yellow bars in Fig. 6. While we noticed several cases whose incompleteness exceeded 95%, the combination of this level of incompleteness with severe sparsity makes the analysis overly difficult and potentially meaningless. Therefore, we set the upper limit for incompleteness at 80%.

In this evaluation, we benchmark against interpolation techniques commonly employed in DEM inpainting, including linear, nearest-neighbor, Spline [27], and IDW [65] interpolation. In particular, we compare with the latest DSM inpainting technique [3] based on Perona-Malik Diffusion (P.M. Diff.) [51]. Table 1 shows that our approach consistently surpasses all these methods in height restoration, regardless of the varying degrees of sparsity and incompleteness.

(FP) ***Either*** **sparsity** ***or*** **incompleteness.** Here, we assess the handling of pure sparsity (s99 i0, s99.75 i0) and incompleteness (s90 i80, s90 i90) independently. We maintain the same noise injection settings as those used for the PoznanRD dataset but exclude tree noise. Table 1 shows RoofDiffusion outperforms most DEM interpolation methods and demonstrates generalizability to unseen BuildingNet [64] datasets. RoofDiffusion demonstrates a stronger performance advantage for the incompleteness task, suggesting particular effectiveness at restoring structural information.

(No-FP) ***Both*** **sparsity** ***and*** **incompleteness.** We compare to the state-of-the-art unguided depth completion methods, pNCNN [13] and CU-Net [75], that we selected for their exclusive use of depth data and hence close relation to height completion tasks. These algorithms were trained on our PoznanRD, using identical data augmentation settings, but no footprint information was used. We

Methods	PoznanRD								BuildingNet							
	s95 i30		s95 i60		s99 i30		s99 i60		s90 i60		s90 i80		s98 i0		s99 i0	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Linear	0.778	1.966	1.330	2.811	1.091	2.542	1.969	3.625	0.568	1.302	1.517	2.484	0.325	0.926	0.426	1.111
pNCNN	1.635	3.016	1.885	3.169	2.012	3.298	2.172	3.378	0.918	1.605	1.185	1.822	0.904	1.632	1.063	1.765
CU-Net	1.246	1.823	1.628	2.244	1.544	2.187	1.923	**2.554**	0.675	1.280	1.641	**2.386**	0.323	**0.717**	0.397	**0.829**
Ours (No-FP)	**0.319**	**1.232**	**0.769**	**2.018**	**0.722**	**1.968**	**1.200**	2.600	**0.509**	**1.228**	**1.501**	2.449	**0.262**	0.803	**0.349**	0.962

Table 2: Evaluation of height map completion *w/o footprint* on the PoznanRD and BuildingNet [64] datasets.

Methods	s99 i30		s99 i80	
	RMSE	#Face	RMSE	#Face
City3D + IDW	0.352	124.40	0.708	105.27
City3D + P.M. Diff	0.577	89.04	3.016	97.47
City3D + Ours	**0.244**	**82.72**	**0.534**	**80.12**

Table 3: Evaluation of point cloud pre-processors for City3D [24] on PoznanRD. City3D tested with GT point cloud achieved 0.104 RMSE and 82.68 average faces.

Methods	s95 i30	s95 i60	s99 i30	s99 i60
Linear	82.18	68.54	73.81	51.04
pNCNN	68.97	66.32	64.80	63.94
CU-Net	82.12	73.88	75.92	69.51
Ours (No-FP)	**92.14**	**81.83**	**83.59**	**75.15**

Table 4: Footprint predictions measured by IoU (%).

adhered to the default hyperparameter configurations specified for each method. To ensure a fair comparison, we also trained No-FP RoofDiffusion, a variant of RoofDiffusion that does not require footprints. We tested reconstruction accuracy on PoznanRD (s95 i30, s95 i60, s99 i30, and s99 i60) utilizing MAE and RMSE. The noise injection settings mirror the experiment in Tab. 1. Since predicting both footprint and roof simultaneously is a more difficult problem, the sparsity and incompleteness selection in Tab. 2 is lower than in Tab. 1.

Table 2 shows that RoofDiffusion achieves the most accurate reconstruction while pNCNN [13] fails to accurately predict the height values. CU-Net [75] tends to over-smooth heights. Linear interpolation struggles to recognise noise and incomplete regions.

(No-FP) *Either* sparsity *or* incompleteness. When tested on the BuildingNet [64] dataset, No-FP RoofDiffusion outperforms linear interpolation, CU-Net [75], and pNCNN [13] in scenarios with dominant sparsity (s98 i0, s99 i0) or incompleteness (s90 i60, s90 i80). While CU-Net's smoother predictions reduce large errors and improve RMSE, notably at s90 i80, s98 i0, and s99 i0 in Tab. 2, CU-Net does not predict sharp height maps well, as indicated by a higher MAE.

6.2 Footprint Recovery

No-FP RoofDiffusion can help footprint recovery. Given a corrupted height map, we can predict the complete heights and infer a binary footprint by assigning a value of one to non-zero pixel values and zero otherwise. Tab. 4 displays Intersection over Union (IoU) between ground truth and the predicted footprint. RoofDiffusion yields the best the footprint recovery on the PoznanRD dataset compared to linear interpolation, CU-Net [75], and pNCNN [13].

6.3 Enhancement of 3D Reconstruction

This subsection demonstrates that using our model as the point cloud preprocessor can boost the accuracy of 3D building reconstruction. We preprocessed point clouds for City3D [24], a leading algorithm for converting point clouds to compact building meshes, by with our method, IDW [65], and Perona-Malik Diffusion [3]. We report the average RMSE of the distance from each ground truth point to the face of the closest reconstructed mesh[6]. To further demonstrate the advantages of using clean geometric details from RoofDiffusion in City3D [24] reconstructions, we evaluate the minimum number of polygonal faces required for the output while maintaining geometry accuracy. Here, we dissolve all the edges with a dihedral angle of less than 5 degrees. A lower count, while maintaining a similar level of reconstruction accuracy, is preferable as it indicates a more compact representation.

As shown in Tab. 3, City3D achieves the lowest point-to-plane distance when utilizing point clouds processed by RoofDiffusion. When using RoofDiffusion as a preprocessor for City3D, the average minimum face count is nearly identical to feeding ground truth point clouds to City3D.

Figure 8 illustrates that using RoofDiffusion as a preprocessing step significantly enhances the reconstruction quality in City3D on PoznanRD. RoofDiffusion enhances gable details (Fig. 8a), complex geometry recovery (Figs. 8b and 8c), and is robust to tree noise (Fig. 8d).

6.4 Qualitative Evaluation of Height Completion

To illustrate the capability of handling the gap between synthetic and real-world scanning conditions, we evaluated our method on datasets featuring real-world LiDAR scans. Due to the absence of ground truth, we adopt a visual comparison. Importantly, to maintain the real-world properties of the scans, we refrain from injecting any additional noise including global, outlier, and tree noise.

Real-world Noise. We artificially remove points from the rasterized height maps to evaluate the resilience of our model to such corruptions, even in the presence of real-world noise and unseen roof geometries in both AHN3 [24] and Dales3D [73]. The 99% sparsity (Figs. 7a and 7c) and 70% incompleteness (Figs. 7b and 7d) are synthesized. We observe that IDW [65] produces a mosaic effect, particularly in incomplete regions, and fails to capture the true geometry of roofs. Meanwhile, Perona-Malik Diffusion [3] yields smoother outcomes but tends to overlook sharp roof details. In contrast, RoofDiffusion effectively yields clean and faithful geometric reconstructions.

Real-world Incompleteness. We collect point clouds exhibiting incompleteness from real-world scans, sourced from the Wayne County dataset [69]. Notably, the point cloud is rated at QL2 [71], which is the most common quality of LiDAR scans across the United States. Figures 7e and 7f show that

[6] We tested City3D without preprocessing but only a few meshes are reconstructed due to the severe sparsity.

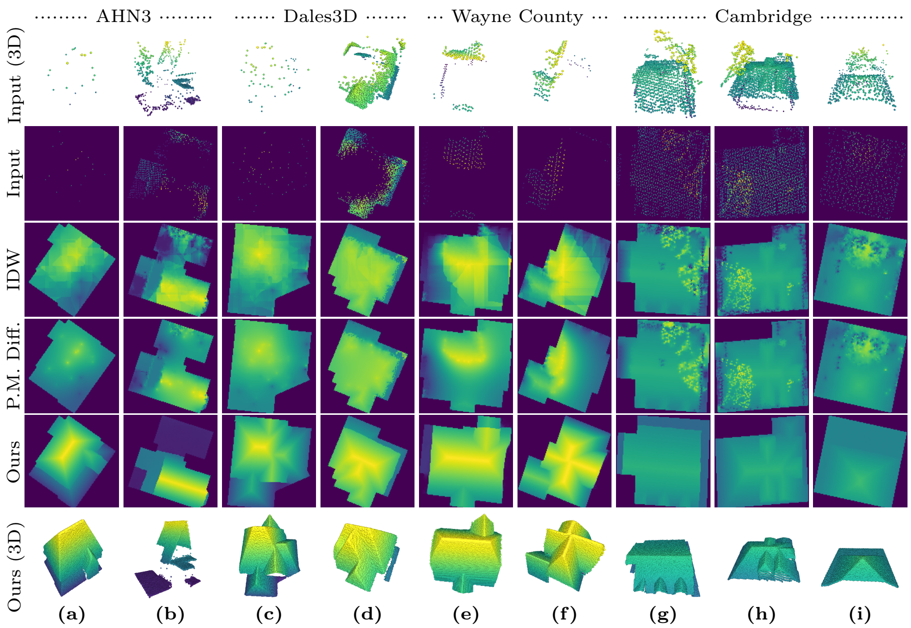


Fig. 7: Evaluation of the completion and denoising of real-world scans.

RoofDiffusion effectively leverages available points in conjunction with footprint information to reconstruct missing parts, even when these parts are distant from the existing points. In contrast, IDW [65] and Perona-Malik Diffusion [3] failed in such scenarios due to a lack of prior information about roof structures. Figures 9a and 9b also shown No-FP RoofDiffusion can recover incompleteness with cleaner and sharper features.

Real-world Tree Noise. We also gather point clouds containing tree noise from the Cambridge dataset [70]. As illustrated in Figs. 7g to 7i, RoofDiffusion demonstrates its efficacy in restoring roof geometry precisely while eliminating the tree points. Moreover, Figs. 9c and 9d highlight that No-FP RoofDiffusion can effectively extract and recover roof geometry from overlapping tree noise.

6.5 Limitations

We selected data containing tree noise from USGS 3DEP LiDAR sampled over Wayne County, MI [69], and Cambridge, MA [70], for analysis. In approximately 36% of these cases, tree noise was mistakenly reconstructed as building structures, see examples in the supplement. This error was due to the resemblance of tree canopies to architectural elements, such as dormers or chimneys, or due to severe occlusion by tree noise obscuring the underlying structures. Footprint-guided RoofDiffusion can encounter challenges due to severe misalignment between the footprint and the height map. This problem also occurs in [24]. Incorporating misalignment into the data augmentation pipeline and relaxing the footprint mask, $\boldsymbol{m}$, during the denoising process warrants further investigation.

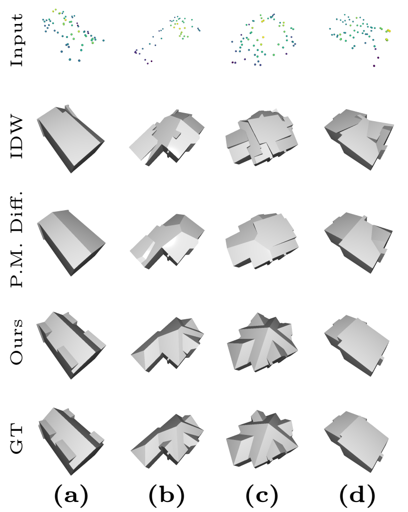


Fig. 8: 3D reconstruction using different height map pre-processors. RoofDiffusion matches the ground truth (GT) meshes generated by feeding GT height maps into City3D [24], except for (b). In (b) RoofDiffusion omits a dormer present in GT due to the lack of data points in that area.

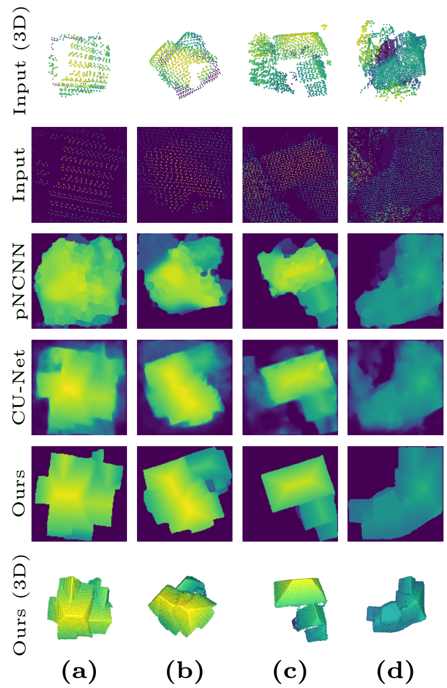


Fig. 9: Evaluation of the completion and denoising in real-world scans without footprints.

Moreover, factors such as floor space, height from the ground, and location are useful indicators of roof geometry but are not yet used in the model. We leave the above as future work.

7 Conclusions

We introduced RoofDiffusion, a diffusion model for roof height map repair. RoofDiffusion is capable of repairing extreme sparsity, incompleteness, and noise. Additionally, we unveiled a comprehensive roof dataset containing over 13k complex geometric roof structures, with complete and a clean ground truth mesh and height map. This dataset can serve as a valuable asset for future long-tail research in remote sensing. To approximate real-world conditions, we also introduced methods for synthesizing tree noise and incomplete shapes. By generating intentionally-corrupted height maps from ground truths, these techniques not only mitigate the absence of ground truth in real-world scans but also facilitate data augmentation. They allow for benchmark customization with varying levels of height map corruption. Our experiments demonstrate the robustness of RoofDiffusion across datasets with real [24, 69, 70, 73] and synthetic scans [64], under diverse corruption conditions. RoofDiffusion outperforms both non-learning [3, 27, 65] and learning-based methods [13, 75] and significantly improves the accuracy of the 3D building reconstruction algorithm [24].

Acknowledgement

We thank Oleksandr Moskalenko and Dr. Zhe Jiang for generously providing computational resources on UF HyperGator. We also extend our appreciation to our colleagues at Meta Platforms for their assistance in processing data and offering valuable feedback throughout this project. We acknowledge the map services and data available from the U.S. Geological Survey, National Geospatial Program, and the Spatial Information System of the City of Poznań, Poland. Additionally, we are grateful for the funding provided by the UF Graduate School Preeminence Award and the UF Informatics Institute Fellowship.

References

1. Anciukevičius, T., Xu, Z., Fisher, M., Henderson, P., Bilen, H., Mitra, N.J., Guerrero, P.: RenderDiffusion: Image diffusion for 3D reconstruction, inpainting and generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12608–12618 (2023) 2
2. Bagheri, H., Schmitt, M., Zhu, X.: Fusion of multi-sensor-derived heights and osm-derived building footprints for urban 3d reconstruction. ISPRS International Journal of Geo-Information **8**(4), 193 (2019) 3
3. Biasutti, P., Aujol, J.F., Brédif, M., Bugeau, A.: Diffusion and inpainting of reflectance and height LiDAR orthoimages. Computer Vision and Image Understanding **179**, 31–40 (2019) 3, 10, 12, 13, 14
4. Biljecki, F., Ledoux, H., Stoter, J.: An improved LOD specification for 3D building models. Computers, Environment and Urban Systems **59**, 25–37 (2016) 3, 7
5. Cappelle, C., El Najjar, M.E., Charpillet, F., Pomorski, D.: Virtual 3D city model for navigation in urban areas. Journal of Intelligent & Robotic Systems **66**, 377–399 (2012) 2
6. Chen, M., Hu, Q., Yu, Z., Thomas, H., Feng, A., Hou, Y., McCullough, K., Ren, F., Soibelman, L.: STPLS3D: A large-scale synthetic and real aerial photogrammetry 3d point cloud dataset. In: Proceedings of the British Machine Vision Conference (2022) 4
7. contributors, H.R.I.: 3D City Models of Helsinki. https://hri.fi/data/dataset/helsingin-3d-kaupunkimalli (2023), accessed: July 1, 2023 5
8. contributors, S.C.P.: Three-dimensional model of Poznan. http://sip.poznan.pl/model3d/ (2023), accessed: July 1, 2023 5, 7, 8
9. Dhariwal, P., Nichol, A.: Diffusion models beat GANs on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021) 4
10. Dong, G., Chen, F., Ren, P.: Filling SRTM void data via conditional adversarial networks. In: IEEE International Geoscience and Remote Sensing Symposium. pp. 7441–7443. IEEE (2018) 4
11. Dong, G., Huang, W., Smith, W.A., Ren, P.: A shadow constrained conditional generative adversarial net for SRTM data restoration. Remote Sensing of Environment **237**, 111602 (2020) 4
12. Dong, Q., Cao, C., Fu, Y.: Incremental transformer structure enhanced image inpainting with masking positional encoding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11358–11368 (2022) 2

13. Eldesokey, A., Felsberg, M., Holmquist, K., Persson, M.: Uncertainty-aware CNNs for depth completion: Uncertainty from beginning to end. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12014–12023 (2020) 3, 4, 10, 11, 14
14. Eldesokey, A., Felsberg, M., Khan, F.S.: Propagating confidences through CNNs for sparse data regression. In: Proceedings of the British Machine Vision Conference (2018) 4
15. Gao, S., Liu, X., Zeng, B., Xu, S., Li, Y., Luo, X., Liu, J., Zhen, X., Zhang, B.: Implicit diffusion models for continuous super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10021–10030 (2023) 4
16. Gavriil, K., Muntingh, G., Barrowclough, O.J.: Void filling of digital elevation models with deep generative models. IEEE Geoscience and Remote Sensing Letters **16**(10), 1645–1649 (2019) 4
17. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014) 4
18. Grohman, G., Kroenung, G., Strebeck, J., et al.: Filling SRTM voids: The Delta surface fill method. Photogrammetric Engineering and Remote Sensing **72**(3), 213–216 (2006) 4
19. Gu, S., Zuo, W., Xie, Q., Meng, D., Feng, X., Zhang, L.: Convolutional sparse coding for image super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1823–1831 (2015) 2
20. He, Y., Liao, W., Hong, H., Huang, X.: High-precision single building model reconstruction based on the registration between osm and dsm from satellite stereos. Remote Sensing **15**(5), 1443 (2023) 3
21. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020) 4, 6
22. Ho, J., Saharia, C., Chan, W., Fleet, D.J., Norouzi, M., Salimans, T.: Cascaded diffusion models for high fidelity image generation. The Journal of Machine Learning Research **23**(1), 2249–2281 (2022) 4
23. Holzmann, T., Maurer, M., Fraundorfer, F., Bischof, H.: Semantically aware urban 3D reconstruction with plane-based regularization. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 468–483 (2018) 2
24. Huang, J., Stoter, J., Peters, R., Nan, L.: City3D: Large-scale building reconstruction from airborne LiDAR point clouds. Remote Sensing **14**(9), 2254 (2022) 2, 3, 4, 9, 11, 12, 13, 14
25. Huang, Z., Fan, J., Cheng, S., Yi, S., Wang, X., Li, H.: HMS-Net: Hierarchical multi-scale sparsity-invariant network for sparse depth completion. IEEE Transactions on Image Processing **29**, 3429–3441 (2019) 4
26. Jaritz, M., De Charette, R., Wirbel, E., Perrotton, X., Nashashibi, F.: Sparse and dense data with CNNs: Depth completion and semantic segmentation. In: 2018 International Conference on 3D Vision. pp. 52–60. IEEE (2018) 4
27. Keys, R.: Cubic convolution interpolation for digital image processing. IEEE transactions on acoustics, speech, and signal processing **29**(6), 1153–1160 (1981) 3, 10, 14
28. Kolbe, T.H., Gröger, G., Plümer, L.: CityGML: Interoperable access to 3D city models. In: Geo-information for disaster management, pp. 883–899. Springer (2005) 2

29. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep Laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 624–632 (2017) 2
30. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. nature **521**(7553), 436–444 (2015) 4
31. Lei, J., Tang, J., Jia, K.: RGBD2: Generative scene synthesis via incremental view inpainting using RGBD diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8422–8434 (2023) 4
32. Li, B., Xue, K., Liu, B., Lai, Y.K.: BBDM: Image-to-image translation with Brownian bridge diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1952–1961 (2023) 4
33. Li, L., Song, N., Sun, F., Liu, X., Wang, R., Yao, J., Cao, S.: Point2Roof: End-to-end 3D building roof modeling from airborne LiDAR point clouds. ISPRS Journal of Photogrammetry and Remote Sensing **193**, 17–28 (2022) 9
34. Li, S., Hu, G., Cheng, X., Xiong, L., Tang, G., Strobl, J.: Integrating topographic knowledge into deep learning for the void-filling of digital elevation models. Remote Sensing of Environment **269**, 112818 (2022) 4
35. Li, W., Lin, Z., Zhou, K., Qi, L., Wang, Y., Jia, J.: MAT: Mask-aware transformer for large hole image inpainting. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10758–10768 (2022) 2
36. Lin, L., Liu, Y., Hu, Y., Yan, X., Xie, K., Huang, H.: Capturing, reconstructing, and simulating: the UrbanScene3D dataset. In: European Conference on Computer Vision. pp. 93–109. Springer (2022) 4
37. Liu, Q., Tan, Z., Chen, D., Chu, Q., Dai, X., Chen, Y., Liu, M., Yuan, L., Yu, N.: Reduce information loss in transformers for pluralistic image inpainting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11347–11357 (2022) 2
38. Liu, T., Cheng, J., Tan, S.: Spectral Bayesian uncertainty for image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18166–18175 (2023) 2
39. Liu, X.: Three-dimensional visualized urban landscape planning and design based on virtual reality technology. IEEE access **8**, 149510–149521 (2020) 2
40. Lu, K., Barnes, N., Anwar, S., Zheng, L.: From depth what can you see? depth completion via auxiliary image reconstruction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11306–11315 (2020) 4
41. Luedeling, E., Siebert, S., Buerkert, A.: Filling the voids in the SRTM elevation model — a TIN-based Delta surface approach. ISPRS Journal of Photogrammetry and Remote Sensing **62**(4), 283–294 (2007) 4
42. Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., Van Gool, L.: RePaint: Inpainting using denoising diffusion probabilistic models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11461–11471 (2022) 2, 4
43. McBratney, A., Webster, R.: Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. Journal of soil Science **37**(4), 617–639 (1986) 3
44. Meng, C., He, Y., Song, Y., Song, J., Wu, J., Zhu, J.Y., Ermon, S.: SDEdit: Guided image synthesis and editing with stochastic differential equations. In: Proceedings of the International Conference on Learning Representations (2021) 4
45. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014) 4

46. Nan, L., Wonka, P.: PolyFit: Polygonal surface reconstruction from point clouds. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2353–2361 (2017) 2
47. Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I., Chen, M.: GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models. In: International Conference on Machine Learning. p. 16784–16804 (2022) 4
48. Ntinas, G.K., Shen, X., Wang, Y., Zhang, G.: Evaluation of CFD turbulence models for simulating external airflow around varied building roof with wind tunnel experiment. In: Building Simulation. vol. 11, pp. 115–123. Springer (2018) 2
49. OpenStreetMap contributors: Planet dump retrieved from https://planet.osm.org. https://www.openstreetmap.org (2017) 2, 3
50. Parmar, G., Kumar Singh, K., Zhang, R., Li, Y., Lu, J., Zhu, J.Y.: Zero-shot image-to-image translation. In: ACM SIGGRAPH 2023 Conference Proceedings. pp. 1–11 (2023) 4
51. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on pattern analysis and machine intelligence **12**(7), 629–639 (1990) 10
52. Qian, Y., Zhang, H., Furukawa, Y.: Roof-GAN: Learning to generate roof geometry and relations for residential houses. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2796–2805 (2021) 5
53. Qiu, Z., Yue, L., Liu, X.: Void filling of digital elevation models with a terrain texture learning model based on generative adversarial networks. Remote Sensing **11**(23), 2829 (2019) 4
54. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., Chen, M.: Hierarchical text-conditional image generation with CLIP latents. arXiv preprint arXiv:2204.06125 **1**(2), 3 (2022) 4
55. Ran, W., Yuan, W., Shibasaki, R.: Few-shot depth completion using denoising diffusion probabilistic model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6558–6566 (2023) 4
56. Ren, J., Zhang, B., Wu, B., Huang, J., Fan, L., Ovsjanikov, M., Wonka, P.: Intuitive and efficient roof modeling for reconstruction and synthesis. ACM Transactions on Graphics **40** (2021) 5
57. Reuter, H.I., Nelson, A., Jarvis, A.: An evaluation of void-filling interpolation methods for SRTM data. International Journal of Geographical Information Science **21**(9), 983–1008 (2007) 3
58. Reynolds, D.A., et al.: Gaussian mixture models. Encyclopedia of biometrics **741**(659-663) (2009) 8
59. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022) 4
60. Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., Aberman, K.: DreamBooth: Fine tuning text-to-image diffusion models for subject-driven generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22500–22510 (2023) 4
61. Saharia, C., Chan, W., Chang, H., Lee, C., Ho, J., Salimans, T., Fleet, D., Norouzi, M.: Palette: Image-to-image diffusion models. In: ACM SIGGRAPH 2022 Conference Proceedings. pp. 1–10 (2022) 3, 4, 7
62. Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E.L., Ghasemipour, K., Gontijo Lopes, R., Karagol Ayan, B., Salimans, T., et al.: Photorealistic text-

to-image diffusion models with deep language understanding. Advances in Neural Information Processing Systems **35**, 36479–36494 (2022) 4
63. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(4), 4713–4726 (2022) 4
64. Selvaraju, P., Nabail, M., Loizou, M., Maslioukova, M., Averkiou, M., Andreou, A., Chaudhuri, S., Kalogerakis, E.: BuildingNet: Learning to label 3D buildings. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10397–10407 (2021) 5, 9, 10, 11, 14
65. Shepard, D.: A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 1968 23rd ACM national conference. pp. 517–524 (1968) 3, 10, 12, 13, 14
66. Shih, P.T.Y., Huang, C.M.: The building shadow problem of airborne LiDAR. The Photogrammetric Record **24**(128), 372–385 (2009) 2
67. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International conference on machine learning. pp. 2256–2265. PMLR (2015) 6
68. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (2021) 4
69. Survey, U.G.: MI Wayne County 2017 A17 345307 LAS. https://rockyweb.usgs.gov/vdelivery/Datasets/Staged/Elevation/LPC/Projects/MI_WayneCounty_2017_A17/ (2017), accessed: July. 1, 2023 2, 8, 9, 12, 13, 14
70. Survey, U.G.: MA CentralEastern 2021 B21 19TCG324693 LAS. https://rockyweb.usgs.gov/vdelivery/Datasets/Staged/Elevation/LPC/Projects/MA_CentralEastern_2021_B21/ (2019), accessed: July. 1, 2023 2, 8, 9, 13, 14
71. Survey, U.G.: Topographic Data Quality Levels (QLs). https://www.usgs.gov/3d-elevation-program/topographic-data-quality-levels-qls (2023), accessed: Oct. 19, 2023 12
72. Uhrig, J., Schneider, N., Schneider, L., Franke, U., Brox, T., Geiger, A.: Sparsity invariant CNNs. In: 2017 international conference on 3D Vision (3DV). pp. 11–20. IEEE (2017) 4, 9
73. Varney, N., Asari, V.K., Graehling, Q.: DALES: A large-scale aerial LiDAR data set for semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 186–187 (2020) 9, 12, 14
74. Wang, R., Huang, S., Yang, H.: Building3D: A urban-scale dataset and benchmarks for learning roof structures from point clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 20076–20086 (2023) 4
75. Wang, Y., Dai, Y., Liu, Q., Yang, P., Sun, J., Li, B.: CU-Net: LiDAR depth-only completion with coupled U-Net. IEEE Robotics and Automation Letters **7**(4), 11476–11483 (2022) 3, 4, 10, 11, 14
76. Wichmann, A., Agoub, A., Kada, M.: ROOFN3D: Deep learning training data for 3D building reconstruction. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **42**, 1191–1198 (2018) 5
77. Wu, C., Wang, D., Bai, Y., Mao, H., Li, Y., Shen, Q.: HSR-Diff: Hyperspectral image super-resolution via conditional diffusion models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7083–7093 (2023) 4
78. Xie, S., Zhang, Z., Lin, Z., Hinz, T., Zhang, K.: SmartBrush: Text and shape guided object inpainting with diffusion model. In: Proceedings of the IEEE/CVF

Conference on Computer Vision and Pattern Recognition. pp. 22428–22437 (2023) 4
79. Yan, L., Tang, X., Zhang, Y.: High accuracy interpolation of DEM using generative adversarial network. Remote Sensing **13**(4), 676 (2021) 4
80. Yan, W.Y.: Airborne LiDAR data artifacts: What we know thus far. IEEE Geoscience and Remote Sensing Magazine (2023) 2
81. Yu, L., Li, X., Fu, C.W., Cohen-Or, D., Heng, P.A.: PU-Net: Point cloud upsampling network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2790–2799 (2018) 9
82. Yu, M., Lafarge, F.: Finding good configurations of planar primitives in unorganized point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6367–6376 (2022) 2
83. Yu, Q., Chu, L., Wu, Q., Pei, L.: Grayscale and normal guided depth completion with a low-cost LiDAR. In: 2021 IEEE International Conference on Image Processing (ICIP). pp. 979–983. IEEE (2021) 4
84. Yu, X., Rao, Y., Wang, Z., Liu, Z., Lu, J., Zhou, J.: PoinTr: Diverse point cloud completion with geometry-aware transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 12498–12507 (2021) 8
85. Yuan, W., Khot, T., Held, D., Mertz, C., Hebert, M.: PCN: Point completion network. In: 2018 international conference on 3D vision (3DV). pp. 728–737. IEEE (2018) 9
86. Zeng, X., Vahdat, A., Williams, F., Gojcic, Z., Litany, O., Fidler, S., Kreis, K.: LION: Latent point diffusion models for 3D shape generation. Advances in Neural Information Processing Systems (2022) 9
87. Zhang, C., Shi, S., Ge, Y., Liu, H., Cui, W.: DEM void filling based on context attention generation model. ISPRS International Journal of Geo-Information **9**(12), 734 (2020) 4
88. Zhang, K., Gool, L.V., Timofte, R.: Deep unfolding network for image super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3217–3226 (2020) 2
89. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3836–3847 (2023) 4
90. Zhou, G., Song, B., Liang, P., Xu, J., Yue, T.: Voids filling of DEM with multiattention generative adversarial network model. Remote Sensing **14**(5), 1206 (2022) 4
91. Zhu, D., Cheng, X., Zhang, F., Yao, X., Gao, Y., Liu, Y.: Spatial interpolation using conditional generative adversarial neural networks. International Journal of Geographical Information Science **34**(4), 735–758 (2020) 4