

Self-Supervised Large Scale Point Cloud Completion for Archaeological Site Restoration

Aocheng Li¹, James R. Zimmer-Dauphinee², Rajesh Kalyanam¹,
Ian Lindsay¹, Parker VanValkenburgh³, Steven Wernke², Daniel Aliaga¹
¹Purdue University, ²Vanderbilt University, ³Brown University

li3922@purdue.edu, james.r.zimmer-dauphinee@vanderbilt.edu, rkalyana@purdue.edu
ilindsay@purdue.edu, parker.vanvalkenburgh@brown.edu, s.wernke@vanderbilt.edu, aliaga@purdue.edu



Figure 1. Restoration results of applying our method on a large structure consisting of 15 million points from Mawchu Llacta in Peru. The structure is completed and displayed over the original site for visualization.

Abstract

Point cloud completion helps restore partial incomplete point clouds suffering occlusions. Current self-supervised methods fail to give high fidelity completion for large objects with missing surfaces and unbalanced distribution of available points. In this paper, we present a novel method for restoring large-scale point clouds with limited and unbalanced ground-truth. Using rough boundary annotations for a region of interest, we project the original point clouds into a multiple-center-of-projection (MCOP) image, where fragments are projected to images of 5 channels (RGB, depth, and rotation). Completion of the original point cloud is reduced to inpainting the missing pixels in the MCOP images. Due to lack of complete structures and an unbalanced distribution of existing parts, we develop a self-supervised scheme which learns to infill the MCOP image with points resembling existing "complete" patches. Special losses are applied to further enhance the regularity and consistency of completed MCOP images, which is mapped back to 3D to form final restoration. Extensive experiments demonstrate the superiority of our method in completing 600+ incom-

plete and unbalanced archaeological structures in Peru.

1. Introduction

Point clouds are important for a wide range of vision applications including autonomous driving, real-estate walk-throughs, pre/post-disaster understanding, and digital archaeology and historic preservation (e.g., [6, 13, 26, 35]). Effective algorithms for learning and understanding point clouds have been developed and validated on synthetic clean shapes [27, 39, 45]. However, for modeling large outdoor real world objects raw point clouds are sparse, noisy, and incomplete. The incompleteness stems from either occlusion (e.g., samples are behind a wall or tree) or missing surfaces (e.g., archaeological remains, post-disaster), which results in different probability distributions of point samples. Hence, point cloud completion for datasets that include significant missing surfaces/areas of large objects, and also lack ground-truth data, is a critical and necessary task. Such 3D reconstructions are vital to understanding the structuring of spaces within archaeological settlements and for modeling social dynamics in the past.

To complete point clouds, prior works [4, 5, 16, 24, 37, 38] have attempted to fill-in the partial point clouds using supervised, unsupervised, or self-supervised learning methods. Supervised learning rely on the presence of ground-truth and labeled input-output pairs. Unsupervised learning omit labeled input-output pairs but still expect ground truth (e.g., examples of complete models spanning the distribution space). Self-supervised learning does not assume ground truth and generates supervision signals itself (e.g., additional self-masking [5, 24], consistency of multi-view partial inputs [16], or similarity of curvatures [38]) from fragmented data like KITTI [11] and MatterPort3D [3]. Self-supervision promises to be a good methodology to complete datasets of large point cloud with significant missing data and no ground truth (of the missing surfaces).

In this work, we present a self-supervised point-cloud completion approach for datasets targeted to archaeological settings with a significant amount of missing surfaces due to weathering, erosion, physical stress and human activities [29], which would otherwise be extremely hard to restore physically to their original appearance. As opposed to prior methods (e.g., [5, 16, 24, 38]), we are able to infer point clouds varying from 50k to 2M points, with up to 95% missing surfaces, and to produce colors/textures. Our approach relaxes the typical uniform-probability assumption of missing points and instead extracts a probability from the partially provided data. This benefits point completion for more general, complex, and large-scale scenarios.

Our method stores points clouds using multi-center-of-projection (MCOP) images and using patch-based self-supervised inpainting. Instead of processing the point cloud in full 3D, our approach uses an MCOP image to combine points observed from multiple viewpoints along a path. Such images are used to train a network using adversarial loss of sampled patches against complete patches to generate completed patches. The image-based representation scales well to point-cloud completion for millions of points (e.g., see Fig. 1 and Fig. 8). Further [35], we use Poisson reconstruction [17] on the completed point cloud to produce a 3D visualization. Our experiments show that our method not only out-performs previous self-supervised point cloud completion methods, but also yields finer geometric details than current self-supervised image inpainting methods for similar (archaeological) settings and produces high-quality completion of models with color data, even when far less than half of the point samples are available.

Our main contributions include the following:

- (1) we provide a novel multiple-center-of-projection (MCOP) representation for point cloud data which unlocks the potential for large scale and high quality point cloud completion of extensive structures,
- (2) we describe a patch-based point-cloud completion algorithm for a 3D scene encoded into a MCOP image repre-

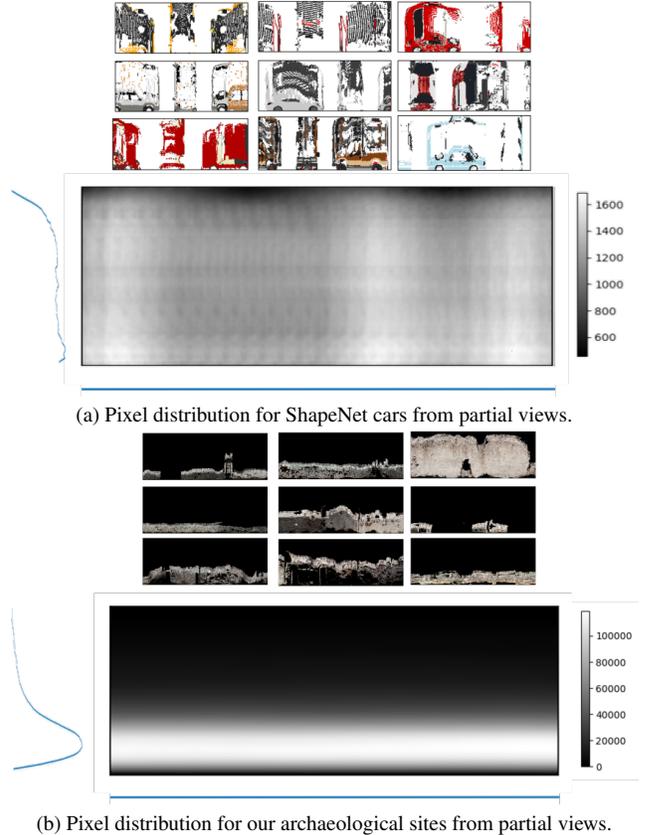


Figure 2. Comparison of distribution of available points for incomplete point clouds suffering from occlusion/partial view and missing surfaces. The distributions are quite different and that of our outdoor sites is clearly vertically unbalanced.

- senting millions of points,
- (3) using (1) and (2) we construct a self-supervised point cloud completion scheme despite spatially unbalanced point distribution and lack of ground truth data (as in the case, at least, in outdoor archaeological settings), and
- (4) we provide a large-scale 3D point-cloud dataset for improving future methods having to complete acquisitions with an extensive spatial range, millions of RGB points, and high-levels of incompleteness.

2. Related Works

2.1. Supervised/Unsupervised Completion

Supervised point cloud completion generally requires complete and clean point clouds at training time. Further, it is assumed provided ground-truth data spans the distribution of target objects. Voxel-based methods [8, 40] benefit from 3D convolution but are challenged with processing speed and resolution. Point-based (e.g., [20, 27, 41, 43]) can adapt better but neither method works well when ground-truth data are not abundantly provided. Unsupervised point cloud

completion does not rely on paired point clouds. Chen et al. [4] proposed an early unpaired point cloud completion method using auto-encoders. Other methods (e.g., [2, 22, 37, 44]) have made further improvements. However, this approach still relies on abundant complete point clouds during training and typically fails for uncommon, extensive and/or highly incomplete objects (e.g., [9, 19]).

2.2. Self-supervised Completion

For some real world scenarios, such as urban environments and archaeological sites, complete samples are costly or unavailable. Self-supervised methods are specifically designed to complete partial point clouds under such settings. Mittal et al. [24] proposed the first learning-based self-supervised method, PointPnCNet, which randomly masked out small regions during training in order to predict both the removed and the originally missing points. Hong et al. [16] developed a closed-loop point cloud completion system named ACL-SPC based on self consistency. Cui et al. [5] partitioned the partial point clouds into three disjoint groups, which served to improve performance. Wu et al. [38] introduced a Pattern Retrieval Network to further find similarities and subsequently densify the completion.

These methods pave the way for point cloud completion without ground truth. Nevertheless, they still exhibit several limitations: 1) the size of the point cloud is limited (i.e., 10k or less); 2) the methods generally assume missing data is caused by a limited field of view, which aggregates to a uniform distribution of all the available points. For object completion in outdoor and archaeological settings, the missing data is usually caused by weathering, erosion, physical stress, and activities. This exhibits a heavily unbalanced distribution, which misguides prior methods to overcompletion/undercompletion (see Section 4); 3) the methods usually lack support for color, which is highly important for downstream visualization and analysis applications. Our approach tackles all three of these prior limitations.

2.3. Computational Archaeology

Computing is essential to archaeology and to cultural heritage[33]. The Digital Michelangelo project [18] paved the way for large-scale archaeological object digitization. Later works addressed completion often by texture/image inpainting [30], finding and exploiting symmetries [19, 23], and/or a priori object databases [12, 14, 15, 34].

Despite the general success on reassembly of small scale archaeological objects, current methods heavily rely on prior knowledge like different types of symmetries and scanned complete shapes. This does not hold true for large-scale building structures, which also come with almost no available ground truth. However, this is the scenario our work addresses.

3. Method

For completing a target object, we first project its point cloud P_{in} into 2D using our MCOP representation (Section 3.1). Then, completing the original point cloud is reduced to inpainting the MCOP image I with an inpainter C (Section 3.2), after which the completed shape P_{out} is produced by reprojection R of the infilled MCOP image $P_{out} = R(C(I))$. Finally, Poisson reconstruction [17] is applied to form a water-tight mesh. An illustration of the whole pipeline is given in Fig. 3.

3.1. Multiple Center of Projection Images

To exploit convolutional deep networks ability to process millions of pixels, we construct a single MCOP image $I \in R^{W \times H \times 5}$ to store the raw point cloud data as seen from multiple viewpoints. The object can be represented as an MCOP image in different ways based on the trajectory and pose of the virtual cameras. For a faithful reconstruction of the original point cloud, we argue that by 1) moving the camera along a trajectory around/through the scene and 2) setting the camera to look perpendicular to the trajectory well preserves the scene’s fine details. Hence in our work, a virtual camera flies around a target scene/object with its viewing orientation set perpendicular to the flying direction. The pixel-wide columns from each captured image (i.e., "slits") are concatenated to form the MCOP. Note that the MCOP projection is not necessarily conformal [7].

The MCOP image has 5-channels so that each column/slit samples the color and distance to both "side" and "top" of the target object. Unlike the original MCOP representation [28] which captures the side and top into different columns, we include a (viewing) rotation angle so that both parts are continuously placed into a single slit (Fig. 4d).

The rotation channel also enables shape manipulation. As shown in supplemental, during generation the user can specify different rotations to essentially change the wall height or overall shape, enabling restoring and completing the archaeological structures based on optional simple user guidance.

Formally, we represent the total process as

$$I, M = MCOP(P_{in}, \mathbf{C}) \quad (1)$$

where $\mathbf{C} = \{C_j | j \in [1, H]\}$ are the cameras used for each column of pixels, $I \in R^{W \times H \times 5}$ is the MCOP image, and $M \in R^{W \times H}$ is a binary mask defined by

$$M_{ij} = \mathbb{1}[I_{ij4} < \infty] \quad (2)$$

which is zero for pixels which do not intersect any point by camera C_j at position (i, j) .

3.2. Self-supervised MCOP Completion

Using MCOP image I , the completion for an object fragment is achieved through infilling the missing pixels. We

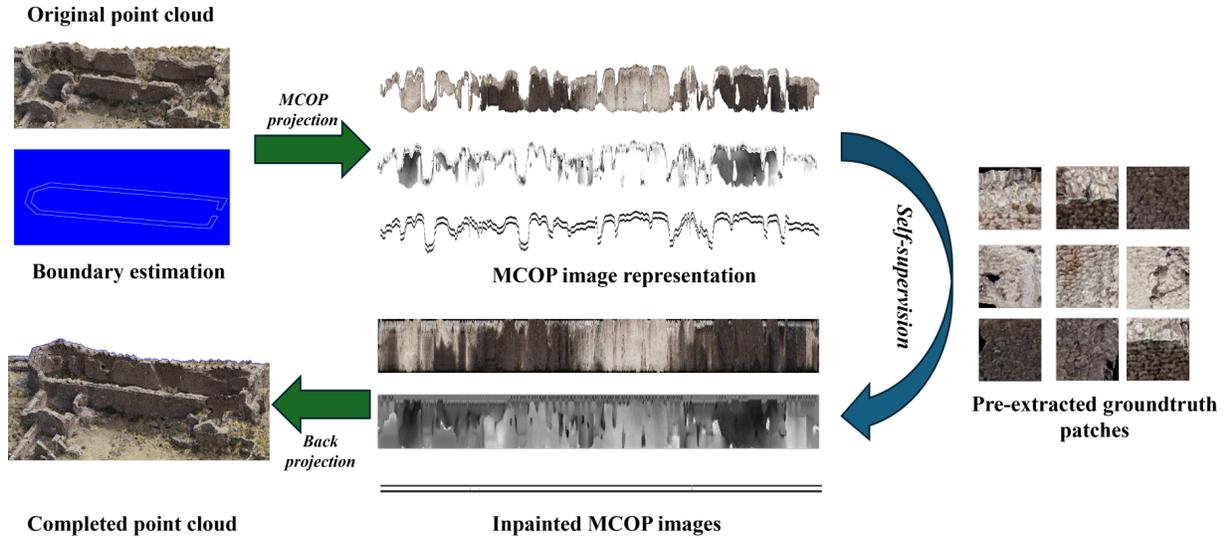


Figure 3. General pipeline of our MCOP based point cloud completion method. The input point cloud is first projected onto a 2D image with 5 channels (RGBD + rotation) using our MCOP representation. During training, a random sampling function f_θ is applied on the MCOP image to extract a group of local windows, which are trained against mostly complete local patches extracted from the dataset of all structures in an adversarial way. For inference, the MCOP image and an outline of the desired shape (e.g., wall height) is passed to our inpainting network adapted from [36]. The final completed point cloud is obtained by remapping the inpainted MCOP image back into 3D.

tackle the scenario of high-levels of incompleteness (e.g.,

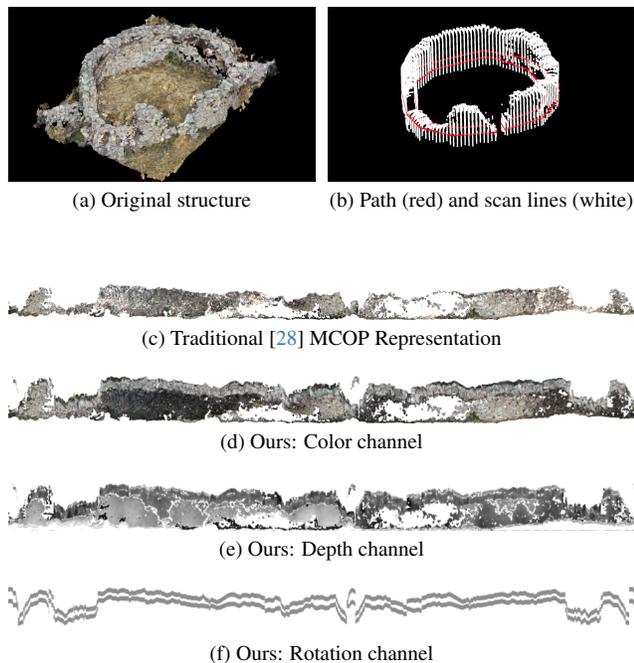


Figure 4. MCOP [28] representation. For (a), we sample along a circular path while MCOP camera captures scan lines perpendicular to the path (b). Instead of the traditional MCOP image (c), we capture a 5 channel image (d, e, f).

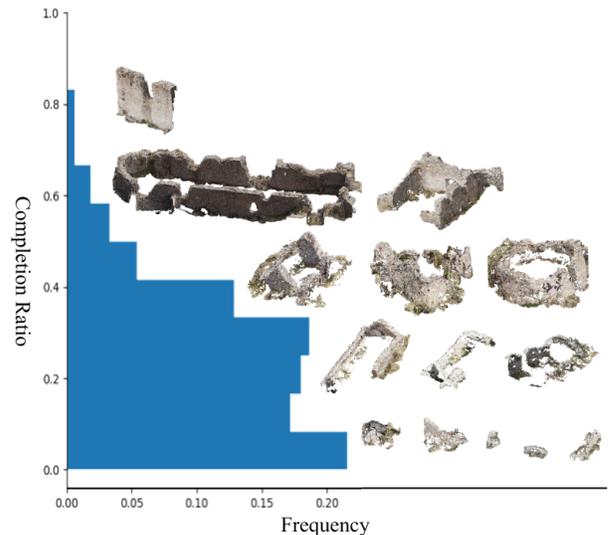


Figure 5. Histogram of random point cloud samples from the Mawchu Llacta dataset, with high completion ratios at the top to low-completion ratios at the bottom.

over 90% of the walls are missing over 90% of pixels) and lack of ground truth. This results in the sampled point distribution exhibiting a non-uniform distribution (e.g., see Fig. 2b and Fig. 5 – most points are missing in the top). These problematic features still manifest in the MCOP images, for which we resort to a self-supervision mechanism which

finds "signals" in subsets of the MCOP images to use in self-supervision during training.

Image generation and completion in self- and unsupervised approaches have been addressed in the literature [1, 21, 25, 42]. In these approaches, given partial observation of MCOP images $\mathbf{D} = \{I_1, I_2, \dots, I_n\}$ and corresponding mask dataset $\mathbf{M} = \{M_1, M_2, \dots, M_n\}$ with $I_i, M_i = MCOP(P_i, \mathbf{C}), i = 1, \dots, n$, an observation function $f_{\mathbf{M}} : R^{W \times H \times 5} \rightarrow R^{W \times H \times 5}$

$$f_{\mathbf{M}}(x) = M_k \otimes x \quad (3)$$

with M_k drawn at random from M would be applied on the completer output O to generate a masked sample of the completion with the masked pixels in x filled with a given constant value (i.e., 0 in our case). In [25], the completer C is then trained adversarially to complete images from \mathbf{D} to $\mathbf{D}' = \{I'_1, I'_2, \dots, I'_n\}$ through

$$I'_i = C(I_i) \quad (4)$$

by making the masked observation of $\mathbf{D}'_{\mathbf{M}} = \{f_{\mathbf{M}}(I'_1), f_{\mathbf{M}}(I'_2), \dots, f_{\mathbf{M}}(I'_n)\}$ to be indistinguishable from \mathbf{D} .

However, directly applying the above would fail under our targeted scenario where the distribution of available pixels in the image is heavily unbalanced (Fig. 2b). Because few fragments have the upper part of the wall, the completer C will learn to escape from adversarial penalty by simply filling the upper part with a default value, as shown in Fig. 10b. It will not be heavily penalized under the masked observation of $\mathbf{D}'_{\mathbf{M}}$.

Instead, to provide effective supervision on both pixels from bottom and top, we apply the following patch-based observation function f_w

$$f_w(I, x, y) = I[x : x + w, y : y + w] \quad (5)$$

where x and y are the coordinates for the upper left corner of the window and w is the fixed size of the patches. Despite the sparseness and unbalanced data in the MCOP images, there are smaller windowed samples of uniform completeness. By carefully selecting the observation window size w , we extract a new observation dataset D_w composed of all such patches under observation f_w . One choice would be f_{64} with windows of no less than 95% completeness, whose average distribution will be (nearly) uniform, and enable the self-supervised strategy in [25, 42].

We empirically search for the optimal patch size w and construct the patch ground truth dataset \mathbf{D}_w based on the data distribution. For each MCOP image I , we use a high resolution image inpainting network derived from LaMa [36] to obtain the output O through $O = C(I)$. To encourage the completion to be genuine, a random patch o observed by

$$o = f_w(O, x_i, y_i) \quad (6)$$

with x_i and y_i sampled from the image coordinates uniformly should be in the distribution modeled by \mathbf{D}_w . The observation patches are fed into the discriminator D adapted from [31] to train adversarially against random patches drawn from D_w to minimize the adversarial loss

$$L_{adv} = \mathbb{E}_{I' \sim D'}[(1 - \log(D(f_w(C(I'))))] \quad (7)$$

while discriminator D is optimized by

$$L_D = \mathbb{E}_{I \sim D_w}[D(I)] + \mathbb{E}_{I' \sim D'}[(1 - \log(D(f_w(C(I'))))] \quad (8)$$

Regularity Terms. To further improve the quality of the generated MCOP images, additional loss terms are applied. To drive completion to reproduce the input MCOP image properly, consistency loss is applied on the known pixels as

$$L_{cons} = \|(O - I) \otimes M\|_2 = \|(C(I) - I) \otimes M\|_2 \quad (9)$$

which encourages the completion to reproduce existing MCOP regions. To correctly reproject O to a closed 3D object, each column of pixels in the rotation channel of C should sum up to the angle camera rotated, giving

$$L_{reg} = \left(\sum_{j=1}^h O_{ij} - \pi \right)^2 \quad (10)$$

Appearance Consistency Terms. Due to the relative limited field of view of observation patches compared to the original image size, additional effort has to be taken to ensure the consistency of patterns and textures at a larger receptive field. Inspired by texture synthesis works [10], we use the pretrained VGG19 network to extract deep features on adjacent patch pairs randomly extracted from completion O and minimize their correlations of features in order to yield a similar appearance. The term is defined as

$$L_{sim} = \sum_{ij} \|G(f_w(O, x_1, y_1)) - G(f_w(O, x_2, y_2))\|_2 \quad (11)$$

where G computes the Gram matrix of deep features. The optimization goal is defined as

$$L = L_{adv} + \lambda_{cons}L_{cons} + \lambda_{reg}L_{reg} + \lambda_{sim}L_{sim} \quad (12)$$

where λ_{cons} , λ_{reg} and λ_{sim} are weights controlling the importance of different terms.

3.3. Implementation Details

Based on the anticipated maximum height of structures, we generally divide the objects to complete into resolutions of $R^{256 \times W \times 5}$ and $R^{384 \times W \times 5}$, which maximizes the pixel utility in the MCOP image for completion of 4 and 6 meter maximum height structures. Two different types of rotation

channels are also preconstructed by making the rotation at the expected height and fed jointly with the original RGBD channels from the MCOP image. Noticing the influence of f_w on different resolutions, we apply D_{64} and D_{96} separately on the above two distributions. To mitigate for efficient training with abundant self-supervision signals, we precomputed the necessary patch statistics and store the desired patches as an efficient way of accessing D_w .

To train the high resolution image inpainter C , we use the proposed self-supervised scheme on MCOP images with a batch size of 4 for 1 million iterations. We use the Adam optimizer with an initial learning rate of 0.0003 and decreasing by 2% every 600 iterations. The model is trained on 4 NVIDIA V100 GPU_s with NV-LINK for 3 days.

With the model properly trained, we feed the MCOP images of the target objects into C to get the completed MCOP representation O , which is back projected into a 3D point cloud. To mitigate for minor issues and facilitate downstream processing/visualization, we apply Poisson reconstruction on the output to obtain a watertight mesh with colors as texture and normals computed per pixel using finite differences with the 4 neighboring pixels in MCOP.

4. Experiments

4.1. Dataset Construction

We conduct qualitative and quantitative experiments on hundreds of point cloud datasets of buildings from three real-world archaeological sites, namely Mawchu Lacta, Huamanmarca, and Kuelap in Peru which suffer from heavy erosion and are missing a significant portion. Each site is reconstructed from multi-view high resolution UAV images as in [32], consisting of billions of points in total. Point clouds for individual walls/structures are extracted following coarse annotations from archaeologists in our team. To accommodate for efficient learning while not missing out important texture and geometric details, we sample the original point clouds with the proposed improved MCOP at a resolution of 2 centimeters, at which the tallest and biggest structure (as shown in Fig. 1) are represented within an MCOP image of 384×9353 . Since the original point clouds exhibit visible shadows due to lighting or self-occlusion in the input stage, we manually annotate out the shadowed regions in the MCOP images and "deshadow" them by either histogram matching to unshadowed regions or masking out the corresponding regions. Some important statistics for the datasets are listed in supplemental. We name the datasets *Mawchu Raw*, *Huaman Raw* and *Kuelap Raw*.

4.2. Quantitative Comparisons

Due to lack of ground truth, we randomly remove sets of points from *Mawchu Raw* to synthesize a more incomplete dataset *Mawchu Incomplete*. The removal is implemented

by applying random masks from D_{128} with completeness ratio in between 20% to 80% at random positions in the original MCOP images. As shown in Fig. 7, the masked out regions could serve as the ground truth for quantitative study when we train on *Mawchu Incomplete* only and evaluate against the masked out regions in *Mawchu Raw*. To fully evaluate the reconstruction quality both in inpainting and point cloud form, we measure the overall RGBD differences of the masked regions, patch/distribution similarity, and Chamfer distance of reconstructed point cloud.

We compare our method on the aforementioned dataset against several state-of-the-art methods. These methods include self-supervised point-cloud approaches (ACL-SPC [16] and P2C [5]), self-supervised image inpainting methods (UAIR [25] and SSII [42]) on the MCOP images (i.e., MCOP+UAIR and MCOP+SSII). We also compare against MCOP+SSII(Patch) which uses image-patch self-supervision. ACL-SPC and P2C do not produce colors so we omit those metrics for them.

	Feature Type	Texture	Geometry
MAE↓	ACL-SPC	N/A	0.172
	P2C	N/A	0.195
	MCOP+SSII	0.206	0.143
	MCOP+UAIR	0.259	0.149
	MCOP+SSII(Patch)	0.209	0.145
	Ours	0.189	0.136
SSIM↑	ACL-SPC	N/A	0.085
	P2C	N/A	0.048
	MCOP+SSII	0.236	0.291
	MCOP+UAIR	0.173	0.261
	MCOP+SSII(Patch)	0.213	0.278
	Ours	0.246	0.353
PSNR↑	ACL-SPC	N/A	15.06
	P2C	N/A	13.52
	MCOP+SSII	11.79	17.18
	MCOP+UAIR	12.37	16.54
	MCOP+SSII(Patch)	11.65	17.10
	Ours	14.28	17.87
FID↓	ACL-SPC	N/A	211.69
	P2C	N/A	331.10
	MCOP+SSII	203.85	184.84
	MCOP+UAIR	192.45	158.79
	MCOP+SSII(Patch)	179.05	169.06
	Ours	90.81	128.41

Table 1. Quantitative comparison of completion quality using image-level metrics. Our approach has the *best* performance (shown in bold).

We first report metrics in image inpainting as shown in Table 1. For each completed point cloud, we project it into 2D using our proposed MCOP representation and perform image patch evaluation with the commonly used metrics of

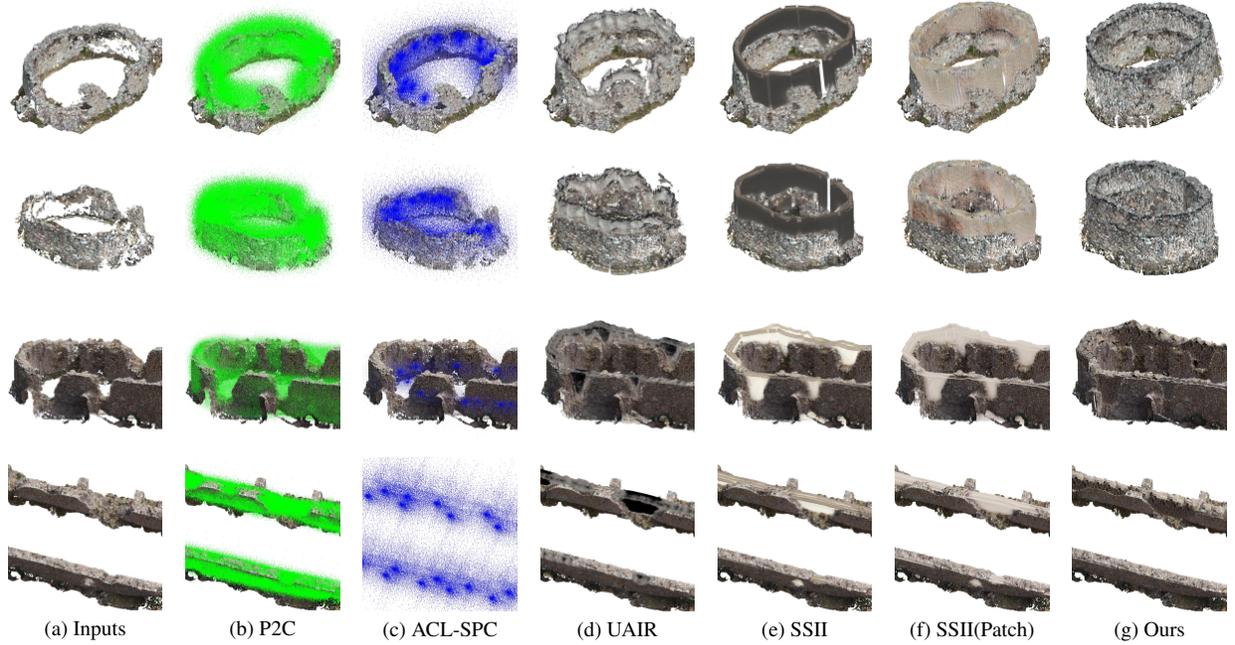


Figure 6. Point cloud completion results of different methods on *Mawchu Raw*. We overlay completion onto input. More visualization results can be found in the supplemental.

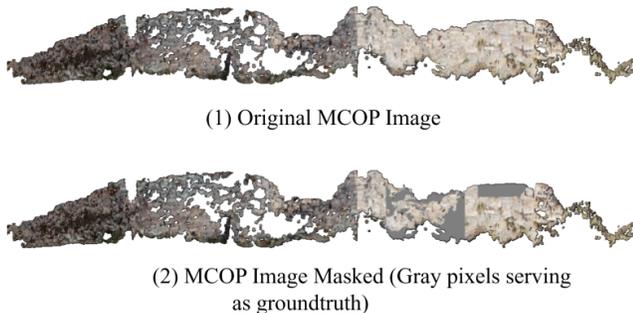


Figure 7. Demonstration of the construction of *Mawchu Incomplete* used in evaluation. Using the original MCOP image, we randomly apply small masks from D_{128} onto random locations producing an incomplete MCOP image.

MAE, SSIM, PSNR and FID based on 500 pairs of randomly chosen windows covering the masked out regions. Our method achieves the best performance in both geometry and texture of predicted point cloud in all metrics.

We further report the completion quality using the same point cloud metrics as in [16], where we compute the Chamfer distance of prediction O and ground-truth part G in the masked regions of *Mawchu Incomplete* as follows:

$$CD(O, G) = \frac{1}{N_o} \sum_{p \in O} \min_{q \in G} \|p - q\|_2 + \frac{1}{N_g} \sum_{q \in G} \min_{p \in O} \|q - p\|_2 \quad (13)$$

Method	P↓	C↓	CD↓
ACL-SPC	91.07	35.00	126.07
P2C	11.33	12.24	23.57
MCOP+SSII	5.20	4.58	9.79
MCOP+UAIR	2.90	6.29	9.20
MCOP+SSII(Patch)	4.27	4.91	9.18
Ours	<u>3.80</u>	4.42	8.22

Table 2. Quantitative comparison of completion quality using point cloud metrics: Chamfer distance (CD), precision (P), and coverage (C) as used in [16] and in centimeters. Our method is best or second best in all metrics.

where N_o and N_g are the number of points in O and G . The first and second term are denoted separately as Precision P and C following [16]. We achieve best or second best in all metrics.

4.3. Ablation Study

MCOP representation. From Table. 1 and 2, methods using MCOP as their base representation generally achieve better performance by a large margin as compared to fully point clouds based ones. The behavior is also interpretable in Fig. 6, where P2C and ACL-SPC generate noisy and non-uniform points around the input, whereas MCOP efficiently regularizes the output point cloud to distribute along given boundary.

Patch-level self supervision. We demonstrate using patch-

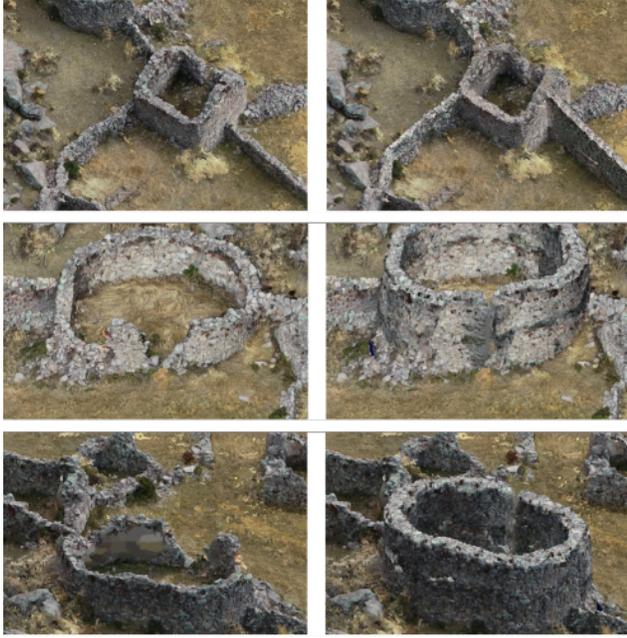


Figure 8. Example completion results from Mawchu Llacta and Huamanmarca. The roof in the fourth row is generated with additional synthesis for visualization purposes only. Completion results on Kuelap are included in the supplemental.

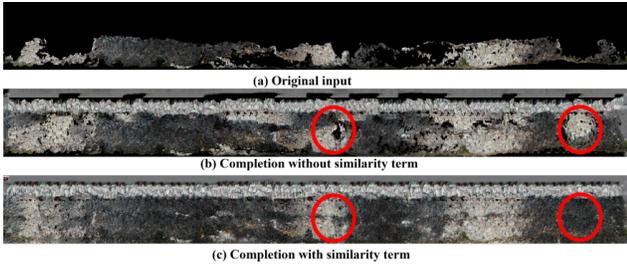


Figure 9. Demonstration of need for texture similarity term.

level self-supervision in Fig. 10. Given (a), directly applying UAIR on the completed patches would yield significantly smaller supervision at the higher pixels compared to the rest because of the unbalanced distribution of available pixels (b), for which the completer simply learns to escape adversarial penalty by infilling default pixel values at those positions. SSII also suffers from the unbalanced distribution and tends to produce blurry completions at pixels far from known regions (c). Our patch-based scheme successfully

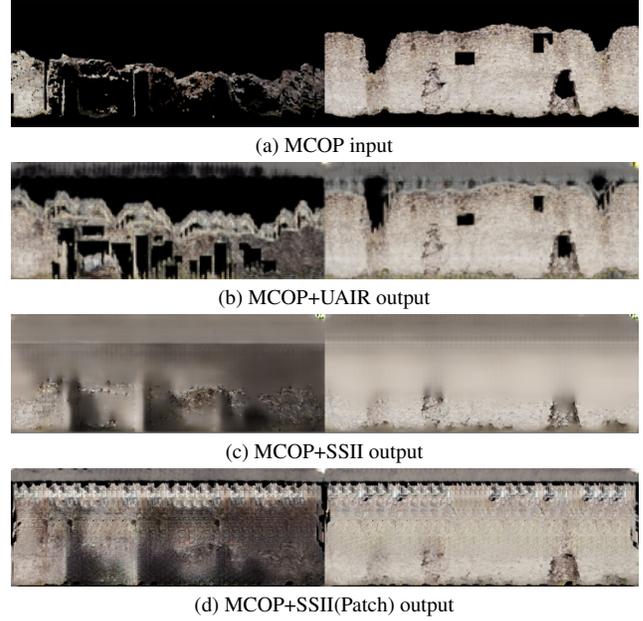


Figure 10. Demonstration of the necessity of applying patch based self supervision for inpainting our MCOP images.

overcomes the unbalanced distribution and helps produce a sharp completion everywhere (d).

Texture similarity. As shown in Fig. 9, completion without texture similarity constraints produces inconsistent horizontal and vertical patterns at a receptive field larger than the observation window, leading to visual inconsistencies as highlighted in red circles. By incorporating the texture similarity term, the completer is able to maintain consistent completion in both directions at the whole image level.

5. Conclusions and Future Work

We propose the first self-supervised point cloud completion pipeline that handles large scale partial point cloud points suffering from missing surfaces, which is common for fragments in archaeological sites. By embedding the original structure in 2D using multiple-center-of-projection (MCOP) images, the approach can handle point clouds of up to millions of points. Patch-based self-supervision with texture similarity terms helps overcome the unbalanced distribution of available pixels resulting from missing surfaces. Results demonstrate our method performs best quantitatively and qualitatively. Nevertheless, our work relies on prior annotation for objects to collect high fidelity MCOP images, which could influence the performance and quality of point cloud completion. As future work, we would like to extend our method to 1) be fully automated to complete input fragments; 2) complete point clouds of general objects at high resolution, and 3) improve shadow elimination.

Acknowledgement This research is partially funded by NSF Grant 2107096, 1835739 and 2411273.

References

- [1] Ashish Bora, Eric Price, and Alexandros G. Dimakis. AmbientGAN: Generative models from lossy measurements. In *International Conference on Learning Representations*, 2018. 5
- [2] Yingjie Cai, Kwan-Yee Lin, Chao Zhang, Qiang Wang, Xiaogang Wang, and Hongsheng Li. Learning a structured latent space for unsupervised point cloud completion, 2022. 3
- [3] Angel Chang, Angela Dai, Thomas Funkhouser, Maciej Halber, Matthias Niessner, Manolis Savva, Shuran Song, Andy Zeng, and Yinda Zhang. Matterport3d: Learning from rgb-d data in indoor environments. *International Conference on 3D Vision (3DV)*, 2017. 2
- [4] Xuelin Chen, Baoquan Chen, and Niloy J Mitra. Unpaired point cloud completion on real scans using adversarial training. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2020. 2, 3
- [5] Ruikai Cui, Shi Qiu, Saeed Anwar, Jiawei Liu, Chaoyue Xing, Jing Zhang, and Nick Barnes. P2c: Self-supervised point cloud completion from single partial clouds, 2023. 2, 3, 6
- [6] Yaodong Cui, Ren Chen, Wenbo Chu, Long Chen, Daxin Tian, Ying Li, and Dongpu Cao. Deep learning for image and point cloud fusion in autonomous driving: A review. *IEEE Transactions on Intelligent Transportation Systems*, 23(2): 722–739, 2022. 1
- [7] Sean Curry and A. Rod Gover. An introduction to conformal geometry and tractor calculus, with a view to applications in general relativity, 2015. 3
- [8] Angela Dai, Charles Ruizhongtai Qi, and Matthias Nießner. Shape completion using 3d-encoder-predictor cnns and shape synthesis, 2017. 2
- [9] Aref Enayati, Luca Palmieri, Sebastiano Vascon, Marcello Pelillo, and Sinem Aslan. Semantic motif segmentation of archaeological fresco fragments. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1717–1725, 2023. 3
- [10] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Texture synthesis using convolutional neural networks, 2015. 5
- [11] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012. 2
- [12] Robert Gregor, Ivan Sipiran, Georgios Papaioannou, Tobias Schreck, Anthousis Andreadis, and Pavlos Mavridis. Towards automated 3d reconstruction of defective cultural heritage objects. In *Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage*, page 135–144, Goslar, DEU, 2014. Eurographics Association. 3
- [13] Lei Han, Tian Zheng, Yinheng Zhu, Lan Xu, and Lu Fang. Live semantic 3d perception for immersive augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):2012–2022, 2020. 1
- [14] Gur Harary, Ayellet Tal, and Eitan Grinspun. Feature-preserving surface completion using four points. *Computer Graphics Forum*, 33(5):45–54, 2014. 3
- [15] Renato Hermoza and Ivan Sipiran. 3d reconstruction of incomplete archaeological objects using a generative adversarial network, 2018. 3
- [16] Sangmin Hong, Mohsen Yavartanoo, Reyhaneh Neshatavar, and Kyoung Mu Lee. Acl-spc: Adaptive closed-loop system for self-supervised point cloud completion. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2, 3, 6, 7
- [17] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, 2006. 2, 3
- [18] Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, Jonathan Shade, and Duane Fulk. *The Digital Michelangelo Project: 3D Scanning of Large Statues*. Association for Computing Machinery, New York, NY, USA, 1 edition, 2023. 3
- [19] Er Li, Xiaopeng Zhang, and Yanyun Chen. Symmetry based chinese ancient architecture reconstruction from incomplete point cloud. In *2014 5th International Conference on Digital Home*, pages 157–161, 2014. 3
- [20] Shanshan Li, Pan Gao, Xiaoyang Tan, and Mingqiang Wei. Proxyformer: Proxy alignment assisted point cloud completion with missing part sensitive transformer, 2023. 2
- [21] Steven Cheng-Xian Li, Bo Jiang, and Benjamin Marlin. Learning from incomplete data with generative adversarial networks. In *International Conference on Learning Representations*, 2019. 5
- [22] Changfeng Ma, Yinuo Chen, Pengxiao Guo, Jie Guo, Chongjun Wang, and Yanwen Guo. Symmetric shape-preserving autoencoder for unsupervised real scene point cloud completion. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13560–13569, 2023. 3
- [23] P. Mavridis, I. Sipiran, A. Andreadis, and G. Papaioannou. Object completion using k-sparse optimization. *Computer Graphics Forum*, 34(7):13–21, 2015. 3
- [24] Himangi Mittal, Brian Okorn, Arpit Jangid, and David Held. Self-supervised point cloud completion via inpainting. *arXiv preprint arXiv:2111.10701*, 2021. 2, 3
- [25] Arthur Pajot, Emmanuel de Bezenac, and Patrick Gallinari. Unsupervised adversarial image reconstruction. In *International Conference on Learning Representations*, 2019. 5, 6
- [26] Gabyong Park, Antonis Argyros, Juyoung Lee, and Woon-tack Woo. 3d hand tracking in the presence of excessive motion blur. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):1891–1901, 2020. 1
- [27] Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space, 2017. 1, 2
- [28] Paul Rademacher and Gary Bishop. Multiple-center-of-projection images. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*,

- page 199–206, New York, NY, USA, 1998. Association for Computing Machinery. [3](#), [4](#)
- [29] Avi Resler, Reuven Yeshurun, Filipe Natalio, and Raja Giryes. A deep-learning model for predictive archaeology and archaeological community detection. *Humanities and Social Sciences Communications*, 8, 2021. [2](#)
- [30] M. S. Sagirolu and A. Erçil. A Texture Based Approach to Reconstruction of Archaeological Finds. In *The 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST*. The Eurographics Association, 2005. [3](#)
- [31] Axel Sauer, Kashyap Chitta, Jens Müller, and Andreas Geiger. Projected gans converge faster. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. [5](#)
- [32] Young-Ha Shin, Sang-Yeop Shin, Heidar Rastiveis, Yi-Ting Cheng, Tian Zhou, Jidong Liu, Chunxi Zhao, Günder Varinlioğlu, Nicholas K. Rauh, Sorin Adam Matei, and Ayman Habib. Uav-based remote sensing for detection and visualization of partially-exposed underground structures in complex archaeological sites. *Remote Sensing*, 15(7), 2023. [6](#)
- [33] Ivan Sipiran. The role of computing in the study of latin american cultural heritage. *Commun. ACM*, 67(8):58–62, 2024. [3](#)
- [34] Ivan Sipiran, Alexis Mendoza, Alexander Apaza, and Cristian Lopez. Data-driven restoration of digital archaeological pottery with point cloud analysis. *International Journal of Computer Vision*, 130(9):2149–2165, 2022. [3](#)
- [35] Andreas Stasinakis, Elisavet Chatzilari, Spiros Nikolopoulos, Ioannis Kompatsiaris, Dimitrios Karolidis, Annareta Touloumtzidou, Nikolaos Papas, and Dimitrios Tzetzis. A hybrid 3d object auto-completion approach with self-supervised data augmentation for fragments of archaeological objects. *Journal of Cultural Heritage*, 56:138–148, 2022. [1](#), [2](#)
- [36] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions. *arXiv preprint arXiv:2109.07161*, 2021. [4](#), [5](#)
- [37] Xin Wen, Zhizhong Han, Yan-Pei Cao, Pengfei Wan, Wen Zheng, and Yu-Shen Liu. Cycle4completion: Unpaired point cloud completion using cycle transformation with missing region coding. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13075–13084, 2021. [2](#), [3](#)
- [38] Lintai Wu, Xianjing Cheng, Junhui Hou, Yong Xu, and Huanqiang Zeng. Self-supervised 3d point cloud completion via multi-view adversarial learning, 2024. [2](#), [3](#)
- [39] Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, and Hengshuang Zhao. Point transformer v3: Simpler, faster, stronger, 2024. [1](#)
- [40] Haozhe Xie, Hongxun Yao, Shangchen Zhou, Jiageng Mao, Shengping Zhang, and Wenxiu Sun. Grnet: Gridding residual network for dense point cloud completion, 2020. [2](#)
- [41] Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation, 2018. [2](#)
- [42] Sriram Yenamandra, Ansh Khurana, Rohit Jena, and Suyash P. Awate. Learning image inpainting from incomplete images using self-supervision. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 10390–10397, 2021. [5](#), [6](#)
- [43] Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. Pcn: Point completion network, 2019. [2](#)
- [44] Junzhe Zhang, Xinyi Chen, Zhongang Cai, Liang Pan, Haiyu Zhao, Shuai Yi, Chai Kiat Yeo, Bo Dai, and Chen Change Loy. Unsupervised 3d shape completion through gan inversion, 2021. [3](#)
- [45] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip Torr, and Vladlen Koltun. Point transformer, 2021. [1](#)