

PUGeo-Net: A Geometry-centric Network for 3D Point Cloud Upsampling

Yue Qian¹, Junhui Hou¹, Sam Kwong¹, and Ying He²

¹ Department of Computer Science, City University of Hong Kong

² School of Computer Science and Engineering, Nanyang Technological University
yueqian4-c@mycityu.edu.hk, jh.hou@cityu.edu.hk (corresponding author),
cssamk@cityu.edu.hk, yhe@ntu.edu.sg

Abstract. In this paper, we propose a novel deep neural network based method, called PUGeo-Net, for upsampling 3D point clouds. PUGeo-Net incorporates discrete differential geometry into deep learning elegantly by learning the first and second fundamental forms that are able to fully represent the local geometry unique up to rigid motion. Specifically, we encode the first fundamental form in a 3×3 linear transformation matrix \mathbf{T} for each input point. Such a matrix approximates the augmented Jacobian matrix of a local parameterization that encodes the intrinsic information and builds a one-to-one correspondence between the 2D parametric domain and the 3D tangent plane, so that we can lift the adaptively distributed 2D samples learned from the input to 3D space. After that, we use the learned second fundamental form to compute a normal displacement for each generated sample and project it to the curved surface. As a by-product, PUGeo-Net can compute normals for the original and generated points, which is highly desired for surface reconstruction algorithms. We evaluate PUGeo-Net on a wide range of 3D models with sharp features and rich geometric details and observe that PUGeo-Net consistently outperforms state-of-the-art methods in terms of both accuracy and efficiency for upsampling factor $4 \sim 16$. We also verify the geometry-centric nature of PUGeo-Net quantitatively. In addition, PUGeo-Net can handle noisy and non-uniformly distributed inputs well, validating its robustness. The code is publicly available at <https://github.com/ninaqy/PUGeo>.

Keywords: Point clouds, Deep learning, Discrete differential geometry, Upsampling, Local parameterization, Surface reconstruction

1 Introduction

Three-dimensional (3D) point clouds, as the raw representation of 3D data, are used in a wide range of applications, such as 3D immersive telepresence [34], 3D city reconstruction [24], [32], cultural heritage reconstruction [47], [3], geophysical information systems [35], [33], autonomous driving [6], [27], simultaneous localization and mapping [10], [7], and virtual/augmented reality [15], [41], just to name a few. Though recent years have witnessed great progress on the 3D

sensing technology [13], [22], it is still costly and time-consuming to obtain dense and highly detailed point clouds, which are beneficial to the subsequent applications. Therefore, amendment is required to speed up the deployment of such data modality. In this paper, instead of relying on the development of hardware, we are interested in the problem of computational based point cloud upsampling: given a sparse, low-resolution point cloud, generate a uniform and dense point cloud with a typical computational method to faithfully represent the underlying surface. Since the problem is the 3D counterpart of image super-resolution [25], [53], a typical idea is to borrow the powerful techniques from the image processing community. However, due to the unordered and irregular nature of point clouds, such an extension is far from trivial, especially when the underlying surface has complex geometry.

The existing methods for point cloud upsampling can be roughly classified into two categories: optimization-based methods and deep learning based methods. The optimization methods [2], [30], [19], [36], [20] usually fit local geometry and work well for smooth surfaces with less features. However, these methods struggle with multi-scale structure preservation. The deep learning methods can effectively learn structures from data. Representative methods are PU-Net [50], EC-Net [49] and MPU [44]. See Section 2 for details. Though these deep learning methods produce better results than the optimization based methods, they are heavily motivated by the techniques in the image domain and takes little consideration of the geometries of the input shape. As a result, various artifacts can be observed in their results. It is also worth noting that all the existing deep learning methods generate points only, none of them is able to estimate the normals of the original and generated points.

In this paper, we propose a novel network, called PUGeo-Net, to overcome the limitations in the existing deep learning methods. Our method learns a local parameterization for each point and its normal direction. In contrast to the existing neural network based methods that generate new points in the abstract feature space and map the samples to the surface using decoder, PUGeo-Net performs the sampling operations in a pure geometric way. Specifically, it first generates the samples in the 2D parametric domain and then lifts them to 3D space using a linear transformation. Finally, it projects the points on the tangent plane onto the curved surface by computing a normal displacement for each generated point via the learned second fundamental form. Through extensive evaluation on commonly used and new metrics, we show that PUGeo-Net consistently outperforms the state-of-the-art in terms of accuracy and efficiency for upsampling factors $4 \sim 16\times$. It is also worth noting that PUGeo-Net is the first neural network that can generate dense point clouds with accurate normals, which are highly desired by the existing surface reconstruction algorithms.

The main contributions of this paper are summarized as follows.

1. We propose PUGeo-Net, a novel geometric-centric neural network, which carries out a sequence of geometric operations, such as computing the first-order approximation of local parameterization, adaptive sampling in the

parametric domain, lifting the samples to the tangent plane, and projection to the curved surface.

2. PUGeo-Net is the first upsampling network that can jointly generate coordinates and normals for the densified point clouds. The normals benefit many downstream applications, such as surface reconstruction and shape analysis.
3. We interpret PUGeo-Net using the local theory of surfaces in differential geometry. Quantitative verification confirms our interpretation.
4. We evaluate PUGeo-Net on both synthetic and real-world models and show that PUGeo-Net significantly outperforms the state-of-the-art methods in terms of accuracy and efficiency for all upsampling factors.
5. PUGeo-Net can handle noisy and non-uniformly distributed point clouds as well as the real scanned data by the LiDAR sensor very well, validating its robustness and practicality.

2 Related Work

Optimization based methods. Alexa *et al.* [2] interpolated points of Voronoi diagram, which is computed in the local tangent space. Lipman *et al.* developed a method based on locally optimal projection operator (LOP) [30]. It is a parametrization-free method for point resampling and surface reconstruction. Subsequently, the improved weighted LOP and continuous LOP were developed by Huang *et al.* [19] and Preiner *et al.* [36] respectively. These methods assume that points are sampling from smooth surfaces, which degrades upsampling quality towards sharp edges and corners. Huang *et al.* [20] presented an edge-aware (EAR) approach which can effectively preserve the sharp features. With given normal information, EAR algorithm first resamples points away from edges, then progressively upsamples points to approach the edge singularities. However, the performance of EAR heavily depends on the given normal information and parameter tuning. In conclusion, point cloud upsampling methods based on geometric priors either assume insufficient hypotheses or require additional attributes.

Deep learning based methods. The deep learning based upsampling methods first extract point-wise feature via point clouds CNN. The lack of point order and regular structure impede the extension of powerful CNN to point clouds. Instead of converting point clouds to other data representations like volumetric grids [31], [39], [46] or graphs [26], [43], recently the point-wise 3D CNN [37], [45], [38], [23], [29], [42] successfully achieved state-of-the-art performance for various tasks. Yu *et al.* pioneered PU-Net[50], the first deep learning algorithm for point cloud upsampling. It adopts PointNet++ [38] to extract point features and expands features by multi-branch MLPs. It optimizes a joint reconstruction and repulsion loss function to generate point clouds with uniform density. PU-Net surpasses the previous optimization based approaches for point cloud upsampling. However, as it does not consider the spatial relations among the points, there is no guarantee that the generated samples are uniform. The follow-up work, EC-Net [49], adopts a joint loss of point-to-edge distance, which

can effectively preserve sharp edges. EC-Net requires labelling the training data with annotated edge and surface information, which is tedious to obtain. Wang *et al.* [44] proposed a patch-based progressive upsampling method (MPU). Their method can successfully apply to large upsampling factor, say $16\times$. Inspired by the image super-resolution techniques, they trained a cascade of upsampling networks to progressively upsample to the desired factor, with the subnet only deals with $2\times$ case. MPU replicates the point-wise features and separates them by appending a 1D code $\{-1, 1\}$, which does not consider the local geometry. MPU requires a careful step-by-step training, which is not flexible and fails to gain a large upsampling factor model directly. Since each subnet upsizes the model by a factor 2, MPU only works for upsampling factor which is a power of 2. PUGeo-Net distinguishes itself from the other deep learning method from its geometry-centric nature. See Sec. 4 for quantitative comparisons and detailed discussions. Recently, Li *et al.* [28] proposed PU-GAN which introduces an adversarial framework to train the upsampling generator. Again, PU-GAN fails to examine the geometry properties of point clouds. Their ablation studies also verify the performance improvement mainly comes from the introduction of the discriminator.

3 Proposed Method

3.1 Overview

Motivation. In our paper, the input is a point cloud sampled from a 3D surface of arbitrary geometry and topology. Given a sparse point cloud $\mathcal{X} = \{\mathbf{x}_i \in \mathbb{R}^{3 \times 1}\}_{i=1}^N$ with N points and the user-specified upsampling factor R , we aim to generate a dense, uniformly distributed point cloud $\mathcal{X}_R = \{\mathbf{x}_i^r \in \mathbb{R}^{3 \times 1}\}_{i,r=1}^{N,R}$, which contains more geometric details and can approximate the underlying surface well. Similar to other patch-based approaches, we first partition the input sparse point cloud into patches via the farthest point sampling algorithm, each of which has M points, and PUGeo-Net processes the patches separately.

As mentioned previously, the existing deep learning based methods are heavily built upon the techniques in 2D image domain, which generate new samples by replicating feature vectors in the abstract feature space, and thus the performance is limited. Moreover, due to little consideration of shape geometry, none of them can compute normals, which play a key role in surface reconstruction. In contrast, our method is motivated by parameterization-based surface resampling.

Local surface parameterization. It is known that parameterization techniques depend heavily on the topology of the surface. There are two types of parameterization, namely local parameterization and global parameterization. The former deals with a topological disk (i.e., a genus-0 surface with 1 boundary) [16]. The latter works on surfaces of arbitrary topology by computing canonical homology basis, through which the surface is cutting into a topological disk, which is then mapped to a 2D domain [12]. Global constraints are required in order

to ensure the parameters are continuous across the cuts [4]. The Fundamental Theorem of the Local Theory of Surfaces states that the local neighborhood of a point on a regular surface can be completely determined by the first and second fundamental forms, unique up to rigid motion (see [5], Chapter 4). Therefore, instead of computing and learning a *global* parameterization which is expensive, our key idea is to learn a *local* parameterization for each point.

Let us parameterize a local neighborhood of point \mathbf{x}_i to a 2D domain via a differential map $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ so that $\Phi(0,0) = \mathbf{x}_i$ (see Fig. 1). The Jacobian matrix $\mathbf{J}_\Phi = [\Phi_u, \Phi_v]$ provides the best first-order approximation of the map Φ : $\Phi(u, v) = \Phi(0, 0) + [\Phi_u, \Phi_v] \cdot (u, v)^\top + O(u^2 + v^2)$, where Φ_u and Φ_v are the tangent vectors, which define the first fundamental form. The normal of point \mathbf{x}_i can be computed by the cross product $\mathbf{n}_i = \Phi_u(0, 0) \times \Phi_v(0, 0)$.

It is easy to verify that the point $\hat{\mathbf{x}} \triangleq \mathbf{x}_i + \mathbf{J}_\Phi \cdot (u, v)^\top$ is on the tangent plane of \mathbf{x}_i , since $(\hat{\mathbf{x}} - \mathbf{x}_i) \cdot \mathbf{n}_i = 0$. In our method, we use the augmented Jacobian matrix $\mathbf{T} = [\Phi_u, \Phi_v, \Phi_u \times \Phi_v]$ to compute the normal $\mathbf{n}_i = \mathbf{T} \cdot (0, 0, 1)^\top$ and the point $\hat{\mathbf{x}} = \mathbf{x}_i + \mathbf{T} \cdot (u, v, 0)^\top$. Matrix \mathbf{T} is of full rank if the surface is regular at \mathbf{x}_i . Furthermore, the distance between \mathbf{x} and $\hat{\mathbf{x}}$ is $\|\mathbf{x} - \hat{\mathbf{x}}\| = \frac{\kappa_1 u^2 + \kappa_2 v^2}{2} + O(u^3 + v^3)$, where κ_1 and κ_2 are the principal curvatures at $\Phi(0, 0)$, which are the eigenvalues of the second fundamental form.

Algorithmic pipeline. As shown in Fig. 2(a), given an input sparse 3D point cloud, PUGeo-Net proceeds as follows: it first generates new samples $\{(u_i^r, v_i^r)\}_{r=1}^R$ in the 2D parametric domain. Then it computes the normal $\mathbf{n}_i = \mathbf{T}_i \cdot (0, 0, 1)^\top$. After that, it maps each generated 2D sample (u_i, v_i) to the tangent plane of \mathbf{x}_i by $\hat{\mathbf{x}}_i^r = \mathbf{T}_i \cdot (u_i^r, v_i^r, 0)^\top + \mathbf{x}_i$. Finally, it projects $\hat{\mathbf{x}}_i^r$ to the curved 3D surface by computing a displacement δ_i^r along the normal direction. Fig. 2(b) illustrates the network architecture of PUGeo-Net, which consists of hierarchical feature extraction and re-calibration (Sec. 3.2), parameterization-based point expansion (Sec. 3.3) and local shape approximation (Sec. 3.4). We adopt a joint loss function to guide the prediction of vertex coordinates and normals (Sec 3.5).

3.2 Hierarchical Feature Learning and Recalibration

We apply DGCNN [45] - the widely used point cloud backbone network - to extract hierarchical point-wise features, which are able to encode both local and global intrinsic geometry information of an input patch.

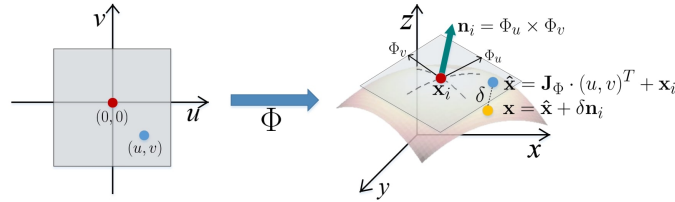


Fig. 1: Surface parameterization and local shape approximation.

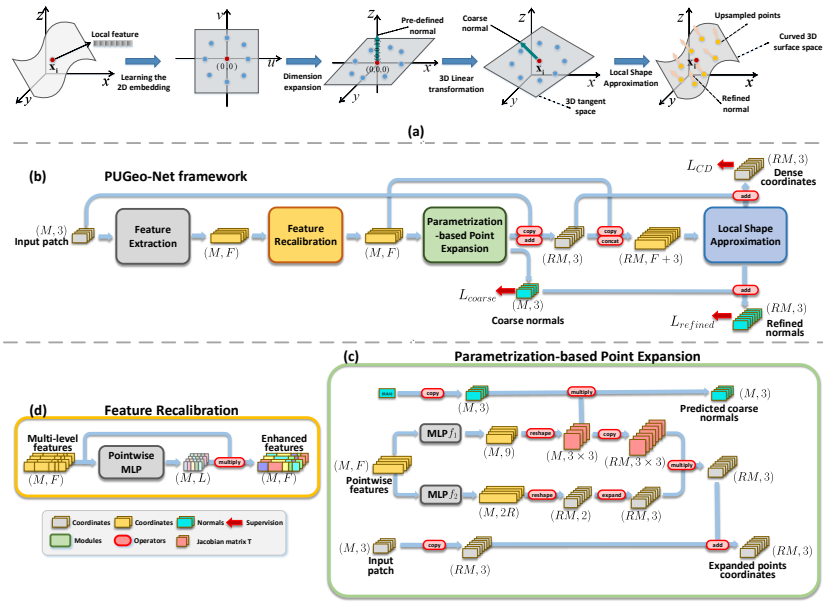


Fig. 2: (a) Illustration the algorithmic pipeline of PUGeo-Net. (b) The end-to-end network structure of PUGeo-Net. The detailed architectures of (c) point expansion and (d) feature recalibration modules

The hierarchical feature learning module extracts features from low- to high-levels. Intuitively speaking, as the receptive fields increase, skip-connection [40], [14], [18], a widely-used technique in 2D vision task for improving the feature quality and the convergence speed, can help preserve details in all levels. To this end, as illustrated in Fig. 2(c), instead of concatenating the obtained features directly, we perform feature re-calibration by a self-gating unit [17], [52] to enhance them, which is computationally efficient.

Let $\mathbf{c}_i^l \in \mathbb{R}^{F_l \times 1}$ be the extracted feature for point \mathbf{x}_i at the l -th level ($l = 1, \dots, L$), where F_l is the feature length. We first concatenate the features of all L layers, i.e., $\hat{\mathbf{c}}_i = \text{Concat}(\mathbf{c}_i^1, \dots, \mathbf{c}_i^L) \in \mathbb{R}^F$, where $F = \sum_{l=1}^L F_l$ and $\text{Concat}(\cdot)$ stands for the concatenation operator. The direct concatenate feature is passed to a small MLP $h_r(\cdot)$ to obtain the logits $\mathbf{a}_i = (a_i^1, a_i^2, \dots, a_i^L)$, i.e.,

$$\mathbf{a}_i = h_r(\hat{\mathbf{c}}_i), \quad (1)$$

which are further fed to a softmax layer to produce the recalibration weights $\mathbf{w}_i = (w_i^1, w_i^2, \dots, w_i^L)$ with

$$w_i^l = e^{a_i^l} / \sum_{k=1}^L e^{a_i^k}. \quad (2)$$

Finally, the recalibrated multi-scale features are represented as the weighted concatenation:

$$\mathbf{c}_i = \text{Concat}(w_i^1 \cdot \mathbf{c}_i^1, w_i^2 \cdot \mathbf{c}_i^2, \dots, \hat{a}_i^L \cdot \mathbf{c}_i^L). \quad (3)$$

3.3 Parameterization-based Point Expansion

In this module, we expand the input sparse point cloud R times to generate a coarse dense point cloud as well as the corresponding coarse normals by regressing the obtained multi-scale features. Specifically, the expansion process is composed of two steps, i.e., learning an adaptive sampling in the 2D parametric domain and then projecting it onto the 3D tangent space by a learned linear transformation.

Adaptive sampling in the 2D parametric domain. For each point \mathbf{x}_i , we apply an MLP $f_1(\cdot)$ to its local surface feature \mathbf{c}_i to reconstruct the 2D coordinates (u_i^r, v_i^r) of R sampled points, i.e.,

$$\{(u_i^r, v_i^r) | r = 1, 2, \dots, R\} = f_1(\mathbf{c}_i). \quad (4)$$

With the aid of its local surface information encoded in \mathbf{c}_i , it is expected that the self-adjusted 2D parametric domain maximizes the uniformity over the underlying surface.

Remark. Our sampling strategy is fundamentally different from the existing deep learning methods. PU-Net generates new samples by replicating features in the feature space, and feed the duplicated features into independent multi-branch MLPs. It adopts an additional repulsion loss to regularize uniformity of the generated points. MPU also replicates features in the feature space. It appends additional code +1 and -1 to the duplicated feature copies in order to separate them. Neither PU-Net nor MPU considers the spatial correlation among the generated points. In contrast, our method expands points in the 2D parametric domain and then lifts them to the tangent plane, hereby in a more geometric-centric manner. By viewing the problem in the mesh parametrization sense, we can also regard appending 1D code in MPU as a *predefined* 1D parametric domain. Moreover, the predefined 2D regular grid is also adopted by other deep learning based methods for processing 3D point clouds, e.g., FoldingNet [48], PPF-FoldNet [9] and PCN [51]. Although the predefined 2D grid is regularly distributed in 2D domain, it does not imply the transformed points are uniformly distributed on the underlying 3D surface.

Prediction of the linear transformation. For each point \mathbf{x}_i , we also predict a linear transformation matrix $\mathbf{T}_i \in \mathbb{R}^{3 \times 3}$ from the local surface feature \mathbf{c}_i , i.e.,

$$\mathbf{T}_i = f_2(\mathbf{c}_i), \quad (5)$$

where $f_2(\cdot)$ denotes an MLP. Multiplying \mathbf{T}_i to the previously learned 2D samples $\{(u_i^r, v_i^r)\}_{r=1}^R$ lifts the points to the tangent plane of \mathbf{x}_i

$$\hat{\mathbf{x}}_i^r = (\hat{x}_i^r, \hat{y}_i^r, \hat{z}_i^r)^\top = \mathbf{T}_i \cdot (u_i^r, v_i^r, 0)^\top + \mathbf{x}_i. \quad (6)$$

Prediction of the coarse normal. As aforementioned, normals of points play an key role in surface reconstruction. In this module, we first estimate a coarse normal, i.e., the normal $\mathbf{n}_i \in \mathbb{R}^{3 \times 1}$ of the *tangent plane* of each input point, which are shared by all points on it. Specifically, we multiply the linear transformation matrix \mathbf{T}_i to the predefined normal $(0, 0, 1)$ which is perpendicular to the 2D parametric domain:

$$\mathbf{n}_i = \mathbf{T}_i \cdot (0, 0, 1)^\top. \quad (7)$$

3.4 Updating Samples via Local Shape Approximation

Since the samples $\hat{\mathcal{X}}_R = \{\hat{\mathbf{x}}_i^r\}_{i,r=1}^{M,R}$ are on the tangent plane, we need to warp them to the curved surface and update their normals. Specifically, we move each sample $\hat{\mathbf{x}}_i^r$ along the normal \mathbf{n}_i with a distance $\delta_i^r = \frac{\kappa_1(u_i^r)^2 + \kappa_2(v_i^r)^2}{2}$. As mentioned in Sec. 3.1, this distance provides the second-order approximation of the local geometry of \mathbf{x}_i . We compute the distance δ_i^r by regressing the point-wise features concatenated with their coarse coordinates, i.e.,

$$\delta_i^r = f_3(\text{Concat}(\hat{\mathbf{x}}_i^r, \mathbf{c}_i)), \quad (8)$$

where $f_3(\cdot)$ is for the process of an MLP. Then we compute the sample coordinates as

$$\mathbf{x}_i^r = (x_i^r, y_i^r, z_i^r)^\top = \hat{\mathbf{x}}_i^r + \mathbf{T}_i \cdot (0, 0, \delta_i^r)^\top. \quad (9)$$

We update the normals in a similar fashion: a normal offset $\Delta \mathbf{n}_i^r \in \mathbb{R}^{3 \times 1}$ for point \mathbf{x}_i^r is regressed as

$$\Delta \mathbf{n}_i^r = f_4(\text{Concat}(\hat{\mathbf{x}}_i^r, \mathbf{c}_i)), \quad (10)$$

which is further added to the corresponding coarse normal, leading to

$$\mathbf{n}_i^r = \Delta \mathbf{n}_i^r + \mathbf{n}_i, \quad (11)$$

where $f_4(\cdot)$ is the process of an MLP.

3.5 Joint Loss Optimization

As PUGeo-Net aims to deal with the regression of both coordinates and unoriented normals of points, we design a joint loss to train it end-to-end. Specifically, let $\mathcal{Y}_R = \{\mathbf{y}_k\}_{k=1}^{RM}$ with RM points be the groundtruth of \mathcal{X}_R . During training, we adopt the Chamfer distance (CD) to measure the coordinate error between the \mathcal{X}_R and \mathcal{Y}_R , i.e.,

$$L_{CD} = \frac{1}{RM} \left(\sum_{\mathbf{x}_i^r \in \mathcal{X}_R} \|\mathbf{x}_i^r - \phi(\mathbf{x}_i^r)\|_2 + \sum_{\mathbf{y}_k \in \mathcal{Y}_R} \|\mathbf{y}_k - \psi(\mathbf{y}_k)\|_2 \right),$$

where $\phi(\mathbf{x}_i^r) = \arg \min_{\mathbf{y}_k \in \mathcal{Y}_R} \|\mathbf{x}_i^r - \mathbf{y}_k\|_2$, $\psi(\mathbf{y}_k) = \arg \min_{\mathbf{x}_i^r \in \mathcal{X}_R} \|\mathbf{x}_i^r - \mathbf{y}_k\|_2$, and $\|\cdot\|_2$ is the ℓ_2 norm of a vector.

For the normal part, denote $\tilde{\mathcal{N}} = \{\tilde{\mathbf{n}}_i\}_{i=1}^M$ and $\overline{\mathcal{N}}_R = \{\overline{\mathbf{n}}_k\}_{k=1}^{RM}$ the ground truth of the coarse normal \mathcal{N} and the accurate normal \mathcal{N}_R , respectively. During training, we consider the errors between \mathcal{N} and $\tilde{\mathcal{N}}$ and between \mathcal{N}_R and $\overline{\mathcal{N}}_R$ simultaneously, i.e.,

$$L_{coarse}(\mathcal{N}, \tilde{\mathcal{N}}) = \sum_{i=1}^M L(\mathbf{n}_i, \tilde{\mathbf{n}}_i), \quad L_{refined}(\mathcal{N}_R, \overline{\mathcal{N}}_R) = \sum_{i=1}^M \sum_{r=1}^R L(\mathbf{n}_i^r, \overline{\mathbf{n}}_{\phi(\mathbf{x}_i^r)}), \quad (12)$$

where $L(\mathbf{n}_i, \tilde{\mathbf{n}}_i) = \max \{\|\mathbf{n}_i - \tilde{\mathbf{n}}_i\|_2^2, \|\mathbf{n}_i + \tilde{\mathbf{n}}_i\|_2^2\}$ measures the unoriented difference between two normals, and $\phi(\cdot)$ is used to build the unknown correspondence between \mathcal{N}_R and $\overline{\mathcal{N}}_R$. Finally, the joint loss function is written as

$$L_{total} = \alpha L_{CD} + \beta L_{coarse} + \gamma L_{refined}, \quad (13)$$

where α , β , and γ are three positive parameters. It is worth noting that our method does not require repulsion loss which is required by PU-Net and EC-Net, since the module for learning the parametric domain is capable of densifying point clouds with uniform distribution.

4 Experimental Results

4.1 Experiment Settings

Datasets. Following previous works, we selected 90 high-resolution 3D mesh models from Sketchfab [1] to construct the training dataset and 13 for the testing dataset. Specifically, given the 3D meshes, we employed the Poisson disk sampling [8] to generate \mathcal{X} , \mathcal{Y}_R , $\tilde{\mathcal{N}}$, and $\overline{\mathcal{N}}$ with $N = 5000$ and $R = 4, 8, 12$ and 16. A point cloud was randomly cropped into patches each of $M = 256$ points. To *fairly* compare different methods, we adopted identical data augmentations settings, including random scaling, rotation and point perturbation. During the testing process, clean test data were used. Also notice that the normals of sparse inputs are not needed during testing.

Implementation details. We empirically set the values of the three parameters α , β , and γ in the joint loss function to 100, 1, and 1, respectively. We used the Adam algorithm with the learning rate equal to 0.001. We trained the network with the mini-batch of size 8 for 800 epochs via the TensorFlow platform. *The code will be publicly available later.*

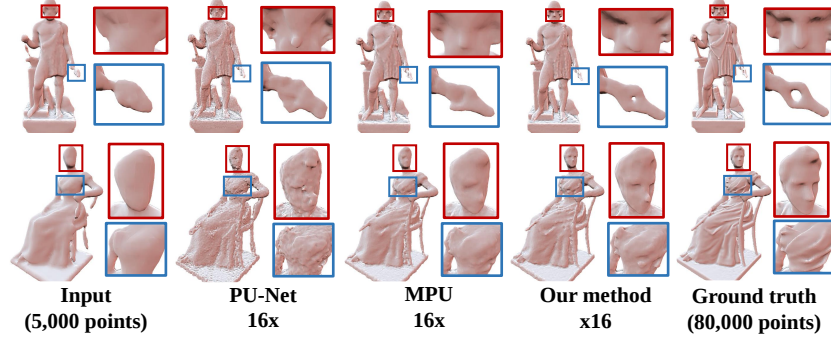
Evaluation metrics. To quantitatively evaluate the performance of different methods, we considered four commonly-used evaluation metrics, i.e., Chamfer distance (CD), Hausdorff distance (HD), point-to-surface distance (P2F), and Jensen-Shannon divergence (JSD). For these four metrics, the lower, the better. For all methods under comparison, we applied the metrics on the whole shape.

We also propose a new approach to quantitatively measure the quality of the generated point clouds. Instead of conducting the comparison between the generated point clouds and the corresponding groundtruth ones directly, we first performed surface reconstruction [21]. For the methods that cannot generate

Table 1: Results of quantitative comparisons. Values are the average of 13 testing models

R	Method	Network size	CD (10^{-2})	HD (10^{-2})	JSDP2F (10^{-2})	me2F (10^{-3})	std (10^{-3})	CD# (10^{-2})	HD# (10^{-2})	JSD# (10^{-2})
4x	EAR [20]	-	0.919	5.414	4.047	3.672	5.592	1.022	6.753	7.445
	PU-Net [50]	10.1 MB	0.658	1.003	0.950	1.532	1.215	0.648	5.850	4.264
	MPU [44]	92.5 MB	0.573	1.073	0.614	0.808	0.809	0.647	5.493	4.259
	PUGeo-Net	26.6 MB	0.558	0.934	0.444	0.617	0.714	0.639	5.471	3.928
8x	EAR [20]	-	-	-	-	-	-	-	-	-
	PU-Net [50]	14.9 MB	0.549	1.314	1.087	1.822	1.427	0.594	5.770	3.847
	MPU [44]	92.5 MB	0.447	1.222	0.511	0.956	0.972	0.593	5.723	3.754
	PUGeo-Net	26.6 MB	0.419	0.998	0.354	0.647	0.752	0.549	5.232	3.465
12x	EAR [20]	-	-	-	-	-	-	-	-	-
	PU-Net [50]	19.7 MB	0.434	0.960	0.663	1.298	1.139	0.573	6.056	3.811
	MPU [44]	-	-	-	-	-	-	-	-	-
	PUGeo-Net	26.7 MB	0.362	0.978	0.325	0.663	0.744	0.533	5.255	3.322
16x	EAR [20]	-	-	-	-	-	-	-	-	-
	PU-Net [50]	24.5 MB	0.482	1.457	1.165	2.092	1.659	0.588	6.330	3.744
	MPU [44]	92.5 MB	0.344	1.355	0.478	0.926	1.029	0.573	5.923	3.630
	PUGeo-Net	26.7 MB	0.323	1.011	0.357	0.694	0.808	0.524	5.267	3.279

CD#, HD#, JSD#: these 3 metrics are used to measure the distance between dense point clouds sampled from reconstructed surfaces and ground truth meshes.

Fig. 3: Visual comparisons for scanned 3D models. Each input sparse 3D point cloud has $N = 5000$ points and upsampled by a factor $R = 16$.

normals principal component analysis (PCA) was adopted to predict normals. Then we densely sampled 200,000 points from reconstructed surface. CD, HD and JSD between the densely sampled points from reconstructed surface and the groundtruth mesh were finally computed for measuring the surface reconstruction quality. Such new measurements are denoted as CD#, HD# and JSD#. Table 1 shows the average results of 13 testing models, where we can observe that PUGeo-Net can achieve the best performance for *all* upsample factors in terms of *all* metrics.

4.2 Comparison with State-of-the-Art Methods

We compared PUGeo-Net with three methods, i.e., optimization based EAR [20], and two state-of-the-art deep learning based methods, i.e., PU-Net [50] and MPU [44]. For fair comparisons, we retrained PU-Net and MPU with the same dataset as ours. Notice that EAR fails to process the task with R greater than 4,

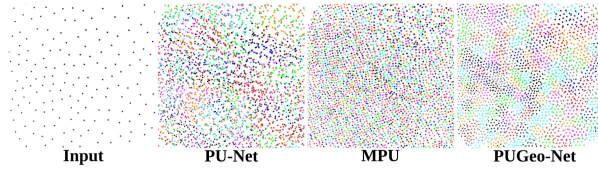


Fig. 4: Visual comparison of the distribution of generated 2D points with upsampling factor $R = 16$

Table 2: Verification of the effectiveness of our normal prediction. Here, the upsampling ratio R is 8. PCA-* indicates the normal prediction by PCA with various numbers of neighborhoods

Methods	CD#	HD#	JSD#	Methods	CD#	HD#	JSD#
PCA-10	0.586	5.837	3.903	PCA-15	0.577	5.893	3.789
PCA-25	0.575	5.823	3.668	PCA-35	0.553	5.457	3.502
PCA-45	0.568	5.746	3.673	PU-Net-M	0.678	6.002	4.139
PUGeo-Net	0.549	5.232	3.464				

due to the huge memory consumption, and MPU can work only for tasks with R in the powers of 2, due to its natural cascaded structure. Note that the primary EAR, PU-Net and MPU cannot predict normals.

Visual comparisons. The superiority of PUGeo-Net is also visually demonstrated. We compared the reconstructed surfaces from the input sparse point clouds and the generated dense point clouds by different methods. Note that the surfaces were reconstructed via the same method as [21], in which the parameters “depth” and “minimum number of samples” were set as 9 and 1, respectively. For PU-Net and MPU which fail to predict normals, we adopted PCA normal estimation with the neighbours equal to 16. Here we took the task with $R = 16$ as an example. Some parts highlighted in red and blue boxes are zoomed in for a close look. From Fig. 3, it can be observed that after performing upsampling the surfaces by PUGeo-Net present more geometric details and the best geometry structures, especially for the highly detailed parts with complex geometry, and they are closest to the groundtruth ones. *See the supplementary material for more visual results.*

Efficiency comparisons. As shown in Table 1, the network size of PUGeo-Net is fixed and much smaller than that of MPU. Due to the deficiency of the independent multi-branch design, the network size of PU-Net grows linearly with the the upsample factor increasing, and is comparable to ours when $R = 16$. Moreover, the running times of PU-Net, MPU and PUGeo-Net are 7.5, 30.3 and 9.6 seconds for upsampling 1000 patches with $R = 16$ on 2080Ti GPU.

Comparison of the distribution of generated points. In Fig. 4, we visualized a point cloud patch which was upsampled with 16 times by different methods. As PUGeo-Net captures the local structure of a point cloud elegantly in a geometry-centric manner, such that the upsampled points are uniformly distributed in the form of clusters. Using PUGeo-Net, the points generated from the same source point \mathbf{x}_i are uniformly distributed in the local neighborhood \mathbf{x}_i ,

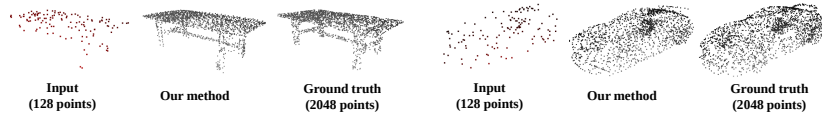


Fig. 5: $16\times$ upsampling results on non-uniformly distributed point clouds.

which justifies our parameterization-based sampling strategy. PU-Net and MPU do not have such a feature. We also observe that our generated points are more uniform than theirs both locally and globally.

4.3 Effectiveness of Normal Prediction

Moreover, we also modified PU-Net, denoted as PU-Net-M, to predict coordinates and normals jointly by changing the neuron number of the last layer to 6 from 3. PU-Net-M was trained with the same training dataset as ours.

The quantitative results are shown in Table 2, where we can see that (1) the surfaces reconstructed with the normals by PUGeo-Net produces the smallest errors for all the three metrics; (2) the number of neighborhoods in PCA based normal prediction is a heuristic parameter and influences the final surface quality seriously; and (3) the PU-Net-M achieves the worst performance, indicating that a naive design without considering the geometry characteristics does not make sense.

4.4 Robustness Analysis

We also evaluated PUGeo-Net with non-uniform, noisy and real scanned data to demonstrate its robustness.

Non-uniform data. As illustrated in Fig. 5, the data from ShapeNet [46] were adopted for evaluation, where 128 points of each point cloud were randomly sampled without the guarantee of the uniformity. Here we took the upsampling task $R = 16$ as an example. From Fig. 5, it can be observed that PUGeo-Net can successfully upsample such non-uniform data to dense point clouds which

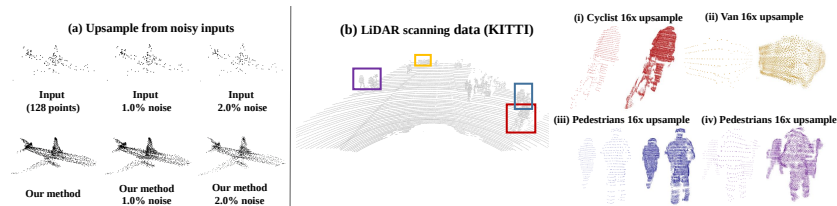


Fig. 6: (a) $16\times$ upsampling results on non-uniform point clouds with various levels of Gaussian noise. (b) $16\times$ upsampling results on real scanned KITTI

Table 3: Ablation study. **Feature recalibration**: concatenate multiscale feature without the recalibration module. **Normal prediction**: only regress coordinates of points. **Learned adaptive 2D sampling**: use a predefined 2D regular grid as the parametric domain. **Linear transformation**: regress coordinates and normals by non-linear MLPs directly. **Coarse to fine**: directly regress coordinates and normals

Networks	CD	HD	JSDP2F	me	std	CD#	HD#	JSD#
Feature recalibration	0.325	1.016	0.371	0.725	0.802	0.542	5.654	3.425
Normal prediction	0.331	2.232	0.427	0.785	0.973	0.563	5.884	3.565
Learned adaptive 2D sampling	0.326	1.374	0.407	0.701	0.811	0.552	5.758	3.456
Linear transformation	0.394	1.005	1.627	0.719	0.720	1.855	11.479	9.841
Coarse to fine	0.330	1.087	0.431	0.746	0.748	0.534	5.241	3.348
Full model	0.323	1.011	0.357	0.694	0.808	0.524	5.267	3.279

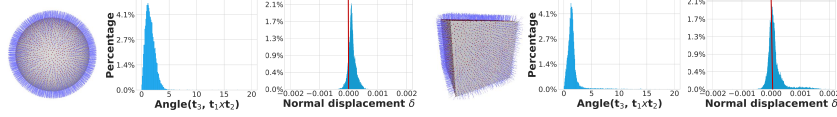


Fig. 7: Statistical analysis of the predicted transformation matrix $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3] \in \mathbb{R}^{3 \times 3}$ and normal displacement δ , which can be used to fully reconstruct the local geometry.

are very close to the ground truth ones, such that the robustness of PUGeo-Net against non-uniformity is validated.

Noisy data. We further added Gaussian noise to the non-uniformly distributed point clouds from ShapeNet, leading to a challenging application scene for evaluation, and various noise levels applying to each dimension in 3D Euclidean space were tested. From Fig. 6, we can observe our proposed algorithm still works very on such challenging data, convincingly validating its robustness against noise.

Real scanned data. Finally, we evaluated PUGeo-Net with real scanned data by the LiDAR sensor [11]. Real scanned data contain noise, outliers, and occlusions. Moreover, the density of real scanned point clouds varies with the distance between the object and the sensor. As shown in Fig. 6, we can see our PUGeo-Net can produce dense point clouds with richer geometric details.

4.5 Ablation Study

We conducted an ablation study towards our model to evaluate the contribution and effectiveness of each module. Table 3 shows the quantitative results for PUGeo-Net with certain module removed. Here we took the task with $R = 8$ as an example, and similar results can be observed for other upsampling factors.

From Table 3, we can conclude that (1) directly regressing the coordinates and normals of points by simply using MLPs instead of the linear transformation decreases the upsampling performance significantly, demonstrating the superiority of our geometry-centric design; (2) the joint regression of normals

and coordinates are better than that of only coordinates; and (3) the other novel modules, including feature recalibration, adaptive 2D sampling, and the coarse to fine manner, all contribute to the final performance.

To demonstrate the geometric-centric nature of PUGeo-Net, we examined the accuracy of the linear matrix \mathbf{T} and the normal displacement δ for a unit sphere and a unit cube, where the ground-truths are available. We use angle θ to measure the difference of vectors \mathbf{t}_3 and $\mathbf{t}_1 \times \mathbf{t}_2$, where $\mathbf{t}_i \in \mathbb{R}^{3 \times 1}$ ($i = 1, 2, 3$) is the i -th column of \mathbf{T} . As Fig. 7 shows, the angle θ is small with the majority less than 3 degrees, indicating high similarity between the predicted matrix \mathbf{T} and the analytic Jacobian matrix. For the unit sphere model, we observe that the normal displacements δ spread in a narrow range, since the local neighborhood of \mathbf{x}_i is small and the projected distance from a neighbor to the tangent plane of \mathbf{x}_i is small. For the unit cube model, the majority of the displacements are close to zero, since most of the points lie on the faces of the cube which coincide with their tangent planes. On the other hand, δ s spread in a relatively wide range due to the points on the sharp edges, which produce large normal displacement.

5 Conclusion and Future Work

We presented PUGeo-Net, a novel deep learning based framework for 3D point cloud upsampling. As the first deep neural network constructed in a geometry centric manner, PUGeo-Net has 3 features that distinguish itself from the other methods which are largely motivated by image super-resolution techniques. First, PUGeo-Net explicitly learns the first and second fundamental forms to fully recover the local geometry unique up to rigid motion; second, it adaptively generates new samples (also learned from data) and can preserve sharp features and geometric details well; third, as a by-product, it can compute normals of the input points and generated new samples, which make it an ideal pre-processing tool for the existing surface reconstruction algorithms. Extensive evaluation shows PUGeo-Net outperforms the state-of-the-art deep learning methods for 3D point cloud upsampling in terms of accuracy and efficiency.

PUGeo-Net not only brings new perspectives to the well-studied problem, but also links discrete differential geometry and deep learning in a more elegant way. In the near future, we will apply PUGeo-Net to more challenging application scenarios (e.g., incomplete dataset) and develop an end-to-end network for surface reconstruction. Since PUGeo-Net explicitly learns the local geometry via the first and second fundamental forms, we believe it has the potential for a wide range 3D processing tasks that require local geometry computation and analysis, including feature-preserving simplification, denoising, and compression.

Acknowledgement. This work was supported in part by the Natural Science Foundation of China under Grant 61871342, and in part by the Hong Kong Research Grants Council under grants 9048123 (CityU 21211518) and 9042955 (CityU 11202320).

References

1. Sketchfab. <https://sketchfab.com>
2. Alexa, M., Behr, J., Cohen-Or, D., Fleishman, S., Levin, D., Silva, C.T.: Computing and rendering point set surfaces. *IEEE Transactions on Visualization and Computer Graphics* **9**(1), 3–15 (2003)
3. Bolognesi, M., Furini, A., Russo, V., Pellegrinelli, A., Russo, P.: Testing the low-cost rpaas potential in 3d cultural heritage reconstruction. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* (2015)
4. Campen, M., Bommers, D., Kobbelt, L.: Quantized global parametrization. *ACM Transactions on Graphics (TOG)* **34**(6), 192:1–192:12 (2015)
5. do Carmo, M.: *Differential Geometry of Curves and Surfaces*. Prentice Hall
6. Chen, X., Ma, H., Wan, J., Li, B., Xia, T.: Multi-view 3d object detection network for autonomous driving. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1907–1915 (2017)
7. Cole, D.M., Newman, P.M.: Using laser range data for 3d slam in outdoor environments. In: *Proceedings 2006 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 1556–1563. IEEE (2006)
8. Corsini, M., Cignoni, P., Scopigno, R.: Efficient and flexible sampling with blue noise properties of triangular meshes. *IEEE Transactions on Visualization and Computer Graphics* **18**(6), 914–924 (2012)
9. Deng, H., Birdal, T., Ilic, S.: Ppfnet: Global context aware local features for robust 3d point matching. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 195–205 (2018)
10. Fioraio, N., Konolige, K.: Realtime visual and point cloud slam. In: *Proceedings of the RGB-D Workshop on Advanced Reasoning with Depth Cameras at Robotics: Science and Systems Conference (RSS)*. vol. 27 (2011)
11. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)* (2013)
12. Gu, X., Yau, S.: Global conformal parameterization. In: *First Eurographics Symposium on Geometry Processing*, June 23–25, 2003. pp. 127–137 (2003)
13. Hakala, T., Suomalainen, J., Kaasalainen, S., Chen, Y.: Full waveform hyperspectral lidar for terrestrial laser scanning. *Optics Express* **20**(7), 7119–7127 (2012)
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 770–778 (2016)
15. Held, R., Gupta, A., Curless, B., Agrawala, M.: 3d puppetry: a kinect-based interface for 3d animation. In: *UIST*. pp. 423–434. Citeseer (2012)
16. Hormann, K., Greiner, G.: Mips: An efficient global parametrization method. *Curve and Surface Design: Saint-Malo* **2000**, 10 (11 2012)
17. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 7132–7141 (2018)
18. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 4700–4708 (2017)
19. Huang, H., Li, D., Zhang, H., Ascher, U., Cohen-Or, D.: Consolidation of unorganized point clouds for surface reconstruction. *ACM Transactions on Graphics (TOG)* **28**(5), 176 (2009)

20. Huang, H., Wu, S., Gong, M., Cohen-Or, D., Ascher, U., Zhang, H.R.: Edge-aware point set resampling. *ACM Transactions on Graphics (TOG)* **32**(1), 9 (2013)
21. Kazhdan, M., Hoppe, H.: Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)* **32**(3), 29 (2013)
22. Kimoto, K., Asada, N., Mori, T., Hara, Y., Ohya, A., et al.: Development of small size 3d lidar. In: 2014 IEEE International Conference on Robotics and Automation (ICRA). pp. 4620–4626. IEEE (2014)
23. Komarichev, A., Zhong, Z., Hua, J.: A-cnn: Annularly convolutional neural networks on point clouds. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 7421–7430 (2019)
24. Lafarge, F., Mallet, C.: Creating large-scale city models from 3d-point clouds: a robust approach with hybrid representation. *International Journal of Computer Vision* **99**(1), 69–85 (2012)
25. 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)
26. Landrieu, L., Simonovsky, M.: Large-scale point cloud semantic segmentation with superpoint graphs. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 4558–4567 (2018)
27. Li, B.: 3d fully convolutional network for vehicle detection in point cloud. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 1513–1518. IEEE (2017)
28. Li, R., Li, X., Fu, C.W., Cohen-Or, D., Heng, P.A.: Pu-gan: a point cloud upsampling adversarial network. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 7203–7212 (2019)
29. Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B.: Pointcnn: Convolution on x-transformed points. In: *Advances in Neural Information Processing Systems*. pp. 820–830 (2018)
30. Lipman, Y., Cohen-Or, D., Levin, D., Tal-Ezer, H.: Parameterization-free projection for geometry reconstruction. In: *ACM Transactions on Graphics (TOG)*. vol. 26, p. 22. ACM (2007)
31. Maturana, D., Scherer, S.: Voxnet: A 3d convolutional neural network for real-time object recognition. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 922–928. IEEE (2015)
32. Musialski, P., Wonka, P., Aliaga, D.G., Wimmer, M., Van Gool, L., Purgathofer, W.: A survey of urban reconstruction. In: *Computer Graphics Forum*. vol. 32, pp. 146–177. Wiley Online Library (2013)
33. Nie, S., Wang, C., Dong, P., Xi, X., Luo, S., Zhou, H.: Estimating leaf area index of maize using airborne discrete-return lidar data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **9**(7), 3259–3266 (2016)
34. Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., Kim, D., Davidson, P.L., Khamis, S., Dou, M., et al.: Holoportation: Virtual 3d teleportation in real-time. In: *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. pp. 741–754. ACM (2016)
35. Paine, J.G., Caudle, T.L., Andrews, J.R.: Shoreline and sand storage dynamics from annual airborne lidar surveys, texas gulf coast. *Journal of Coastal Research* **33**(3), 487–506 (2016)
36. Preiner, R., Mattausch, O., Arikan, M., Pajarola, R., Wimmer, M.: Continuous projection for fast l1 reconstruction. *ACM Transactions on Graphics (TOG)* **33**(4), 47–1 (2014)

37. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 652–660 (2017)
38. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: *Advances in Neural Information Processing Systems*. pp. 5099–5108 (2017)
39. Riegler, G., Osman Ulusoy, A., Geiger, A.: Octnet: Learning deep 3d representations at high resolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3577–3586 (2017)
40. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-assisted Intervention*. pp. 234–241. Springer (2015)
41. Santana, J.M., Wendel, J., Trujillo, A., Suárez, J.P., Simons, A., Koch, A.: Multimodal location based services—semantic 3d city data as virtual and augmented reality. In: *Progress in Location-based Services 2016*, pp. 329–353. Springer (2017)
42. Tatarchenko, M., Park, J., Koltun, V., Zhou, Q.Y.: Tangent convolutions for dense prediction in 3d. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3887–3896 (2018)
43. Te, G., Hu, W., Zheng, A., Guo, Z.: Rgcnn: Regularized graph cnn for point cloud segmentation. In: *2018 ACM Multimedia Conference on Multimedia Conference*. pp. 746–754. ACM (2018)
44. Wang, Y., Wu, S., Huang, H., Cohen-Or, D., Sorkine-Hornung, O.: Patch-based progressive 3d point set upsampling. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 5958–5967 (2019)
45. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (TOG)* **38**(5), 146 (2019)
46. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3d shapenets: A deep representation for volumetric shapes. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1912–1920 (2015)
47. Xu, Z., Wu, L., Shen, Y., Li, F., Wang, Q., Wang, R.: Tridimensional reconstruction applied to cultural heritage with the use of camera-equipped uav and terrestrial laser scanner. *Remote Sensing* **6**(11), 10413–10434 (2014)
48. Yang, Y., Feng, C., Shen, Y., Tian, D.: Foldingnet: Point cloud auto-encoder via deep grid deformation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 206–215 (2018)
49. Yu, L., Li, X., Fu, C.W., Cohen-Or, D., Heng, P.A.: Ec-net: an edge-aware point set consolidation network. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. pp. 386–402 (2018)
50. 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)
51. 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)
52. Zhang, H., Goodfellow, I., Metaxas, D., Odena, A.: Self-attention generative adversarial networks. In: *International Conference on Machine Learning*. pp. 7354–7363 (2019)
53. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2472–2481 (2018)