

# Deep Geometric Prior for Surface Reconstruction

Francis Williams<sup>1</sup>, Teseo Schneider<sup>1</sup>, Claudio Silva<sup>1</sup>, Denis Zorin<sup>1</sup>, Joan Bruna<sup>1</sup>, and Daniele Panozzo<sup>1</sup>

<sup>1</sup>New York University

francis.williams@nyu.edu, teseo.schneider@nyu.edu, csilva@nyu.edu, dzorin@cs.nyu.edu,  
bruna@cims.nyu.edu, panozzo@nyu.edu

## Abstract

*The reconstruction of a discrete surface from a point cloud is a fundamental geometry processing problem that has been studied for decades, with many methods developed. We propose the use of a deep neural network as a geometric prior for surface reconstruction. Specifically, we overfit a neural network representing a local chart parameterization to part of an input point cloud using the Wasserstein distance as a measure of approximation. By jointly fitting many such networks to overlapping parts of the point cloud, while enforcing a consistency condition, we compute a manifold atlas. By sampling this atlas, we can produce a dense reconstruction of the surface approximating the input cloud. The entire procedure does not require any training data or explicit regularization, yet, we show that it is able to perform remarkably well: not introducing typical overfitting artifacts, and approximating sharp features closely at the same time. We experimentally show that this geometric prior produces good results for both man-made objects containing sharp features and smoother organic objects, as well as noisy inputs. We compare our method with a number of well-known reconstruction methods on a standard surface reconstruction benchmark.*

## 1. Introduction

3D geometry is commonly acquired in the form of collections of (possibly incomplete) range images (laser scanning, structured light, etc) or measurements of more complex structure (LIDAR). Unordered set of points (point clouds) is a commonly used representation of combined registered results of scanning objects or scenes. Point clouds can be obtained in other ways, e.g., by sampling an implicit surface using ray casting. Computing a continuous representation of a surface from the discrete point cloud (e.g., a polygonal mesh, an implicit surface, or a set of parameteric patches) in way that is robust to noise, and yet

retains critical surface features and approximates the sampled surface well, is a pervasive and challenging problem.

Different approaches have been proposed, mostly grouped into several categories: (1) using the points to define a volumetric scalar function whose 0 level-set corresponds to the desired surface, (2) attempt to "connect the dots" in a globally consistent way to create a mesh, (3) fit a set of primitive shapes so that the boundary of their union is close to the point cloud, and (4) fit a set of patches to the point cloud approximating the surface.

We propose a novel method, based, on the one hand, on constructing a manifold atlas commonly used in differential geometry to define a surface, and, on the other hand, on observed remarkable properties of deep image priors [37], using an overfitted neural network for interpolation. We define a set of 2D parametrizations, each one mapping a square in parametric space to a region of a surface, ensuring consistency between neighbouring patches. This problem is inherently ambiguous: there are many possible valid parametrizations, and only a small subset will correspond to a faithful representation of the underlying surface. We compute each parametrization by overfitting a network to a part of the point cloud, while enforcing consistency conditions between different patches. We observe that the result is a reconstruction which is superior quantitatively and qualitatively to commonly used surface reconstruction methods.

We use the Wasserstein distance as a training loss, which is naturally robust to outliers, and has the added advantage of providing us explicit correspondences between the parametric domain coordinates and the fitted points, allowing us to explicitly measure, and thus minimize, the disagreement between neighbouring patches covering a surface.

We use a standard shape reconstruction benchmark to compare our method with 12 competing methods, showing that, despite the conceptual simplicity of our algorithm, our reconstructions are superior in terms of quantitative and visual quality.

## 2. Related work

**Geometric Deep Learning** A variety of architectures were proposed for geometric applications. A few work with point clouds as input; in most cases however, these methods are designed for classification or segmentation tasks. One of the first examples are PointNet [33] and PointNet++ [34] originally designed for classification and segmentation, using a set-based neural architecture [39, 36]. PCPNet [14] is version of PointNet architecture, for estimation of local shape properties. A number of learning architectures for 3D geometry work with voxel data converting input point clouds to this form (e.g., [38]). The closest problems these types of networks solve is shape completion and point cloud upsampling.

Shape completion is considered, e.g., in [9], where volumetric CNN is used to predict a very coarse shape completion, which is then refined using data-driven local shape synthesis on small volumetric patches. [16], follows a somewhat similar approach, combining multiview and volumetric low-resolution global data at a first stage, and using a volumetric network to synthesize local patches to obtain higher resolution. Neither of these methods aims to achieve high-quality surface reconstruction.

PU-Net, described in [41], is to the best of our knowledge, the only learning-based work addressing point cloud upsampling directly. The method proceeds by splitting input shapes into patches and learning hierarchical features using PointNet++ type architecture. Then feature aggregation and expansion is used to perform point set expansion in feature space, followed by the (upsampled) point set reconstruction.

In contrast to other methods, the untrained networks in our method take parametric coordinates in square parametric domains as inputs and produce surface points as output. An important exception is the recent work [12] defining an architecture, AtlasNet, in which the decoder part is similar to ours, but with some important distinctions discussed in Section 3. Finally, [4] studied the ability of neural networks to approximate low-dimensional manifolds, showing that even two-layer ReLU networks have remarkable ability to encode smooth structures with near-optimal number of parameters. In our setting, we rely on overparametrisation and leverage the implicit optimization bias of gradient descent.

**Surface Reconstruction** is an established research area dating back at least to the early 1990s (e.g., [17]); [6] is a recent comprehensive survey of the area. We focus our discussion on the techniques that we use for comparison, which are a superset of those included in the surface reconstruction benchmark of Berger et al [5]. Berger tested 10 different techniques in their paper; we will follow their

nomenclature. They separate techniques into four main categories: indicator function, point-set surfaces, multi-level partition of unity, and scattered point meshing.

Indicator function techniques define a scalar function in space that can be used for testing if a given point is inside or outside the surface. There are multiple ways to define such a function from which a surface is generated by isocontouring. Poisson surface reconstruction (**Poisson**) [21] inverts the gradient operator by solving a Poisson equation to define the indicator function. Fourier surface reconstruction (**Fourier**) [20] represents the indicator function in the Fourier basis, while Wavelet surface reconstruction (**Wavelet**) [27] employs a Haar or a Daubechies (4-tap) basis. Screened Poisson surface reconstruction (**Screened**) [22] is an extension of [21] that incorporates point constraints to avoid over smoothing of the earlier technique. This technique is not considered in [5], but is considered by us.

Point-set surfaces [2] define a projection operator that moves points in space to a point on the surface, where the surface is defined to be the collection of stationary points of the projection operator. Providing a definition of the projection operators are beyond the scope of our paper (see [5]). In our experiments, we have used simple point set surfaces (**SPSS**) [1], implicit moving least squares (**IMLS**) [24], and algebraic point set surfaces (**APSS**) [13].

Edge-Aware Point Set Resampling (**EAR**) [18] (also not considered in [5], but considered by us) works by first computing reliable normals away from potential singularities, followed by a resampling step with a novel bilateral filter towards surface singularities. Reconstruction can be done using different techniques on the resulting augmented point set with normals.

Multi-level Partition of Unity defines an implicit function by integrating weight function of a set of input points. The original approach of [29] (**MPU**) uses linear functions as low-order implicits, while [28] (**MPUSm**) defines differential operators directly on the MPU function. The method of [31] (**RBF**) uses compactly-supported radial basis functions.

Scattered Point Meshing [30] (**Scattered**) grows weighted spheres around points in order to determine the connectivity in the output triangle mesh.

The work in [32] uses a manifold-based approach to a direct construction of a global parametrization from a set of range images (a point cloud, or any other surface representation, can be converted to such a set by projection to multiple planes). It uses range images as charts with projections as chart maps; our method computes chart maps by fitting. [40] jointly optimizes for connectivity and geometry to produce a single mesh for an entire input point cloud. In contrast, our method produces a global chart map using only a local optimization procedure.

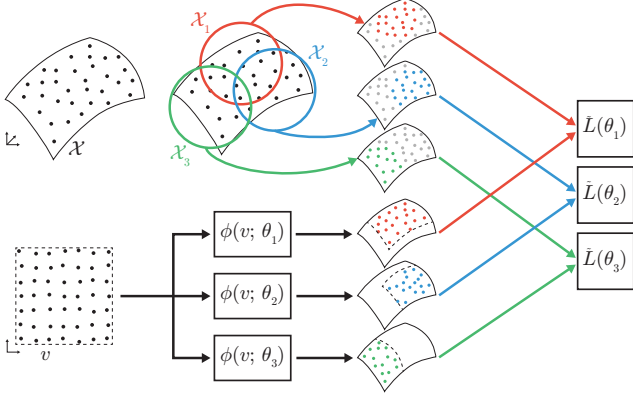


Figure 1. Overview of our proposed surface reconstruction architecture. An input point cloud  $\mathcal{X}$  is split into overlapping patches  $\mathcal{X}_p$ , where a local chart  $\varphi_p(v) = \phi(v, \theta_p)$  is obtained via the minimisation of a Wasserstein loss  $\tilde{L}(\theta)$  (2). The local charts are made globally consistent thanks to the correspondences arising during the Wasserstein loss minimization (Section 3.2)

**Deep Image Prior** Our approach is inspired, in part, by the *deep image prior*. [37] demonstrated that an untrained deep network can be overfitted to input data producing a remarkably high-quality upsampling and even hole filling without training, with the convolutional neural network structure acting as a regularizer. Our approach to surface reconstruction is similar, in that we use untrained networks to represent individual chart embedding functions. However, an important difference is that our loss function measures geometric similarity.

### 3. Method

Our method for surface reconstruction uses a set of deep ReLU networks to obtain local *charts* or parametrizations (Section 3.1). These parameterizations are then made consistent with each other on the parts where they overlap (Section 3.2). The networks are trained using the 2-Wasserstein distance as a loss function. The overall architecture of our technique is illustrated in Figure 1.

In the following, we denote by  $\mathcal{S}$  a smooth surface (possibly with a boundary) in  $\mathbb{R}^3$ . The goal of surface reconstruction is to estimate  $\mathcal{S}$  from a possibly noisy point cloud  $\mathcal{X} = \{x_i = y_i + w_i; y_i \in \mathcal{S}; i \leq N\}$ , where  $w_i$  models the acquisition noise.

#### 3.1. Local Parametrization Model

Let us first consider the reconstruction of a given neighborhood of  $\mathcal{S}$  around a point  $p$ , denoted as  $\mathcal{U}_p$ , from the corresponding point cloud  $\mathcal{X}_p = \mathcal{X} \cap \mathcal{U}_p$ . If  $\mathcal{U}_p$  is sufficiently small, from the implicit function theorem, one can characterize  $\mathcal{U}_p \cap \mathcal{S}$  as the image of the open square  $V = (0, 1)^2$  by a differentiable mapping  $\varphi: V \rightarrow \mathcal{U}_p$ .

We propose to approximate  $\varphi$  using a neural network,  $\phi(\cdot; \theta_p): V \rightarrow \mathbb{R}^3$ , where  $\theta_p$  is a vector of parameters, that we fit so that the image of  $\phi$  approximates  $\mathcal{U}_p \cap \mathcal{S}$ . For that purpose, we first consider a sample  $\{v_1, \dots, v_n\}$  of  $n = |\mathcal{X}_p|$  points in  $V$  using a Poisson disk distribution, and the associated Earth Movers or 2-Wasserstein Distance (EMD) between  $\{\phi(v_i; \theta_p) : i = 1 \dots n\}$  and  $\mathcal{X}_p = \{x_1 \dots, x_n\}$ :

$$L(\theta) = \inf_{\pi \in \Pi_n} \sum_{i \leq n} \|\phi(v_i; \theta) - x_{\pi(i)}\|^2. \quad (1)$$

where  $\Pi_n$  is the set of permutations of  $n$  points. The computation of the EMD in (1) requires solving a linear assignment problem, which can be done in  $O(n^3)$  using, for instance, the Hungarian algorithm [25]. Since this is prohibitive for typical values of  $n$ , we rely instead on the Sinkhorn regularized distance [8]:

$$\tilde{L}(\theta) = \min_{P \in \mathcal{P}_n} \sum_{i,j \leq n} P_{i,j} \|\phi(v_j; \theta) - x_i\|^2 - \lambda^{-1} H(P) \quad (2)$$

where  $\mathcal{P}_n$  is the set of  $n \times n$  bi-stochastic matrices and  $H(P)$  is the entropy,  $H(P) = -\sum_{i,j} P_{i,j} \log P_{i,j}$ . This distance provably approximates the Wasserstein metric as  $\lambda \rightarrow \infty$  and can be computed in near-linear time [3]. Figure 14 in the supplemental material shows the effect of varying the regularization parameter  $\lambda$  on the results.

We choose  $\phi$  to be a MLP with the half-rectified activation function:

$$\phi(v; \theta) = \theta_K \text{ReLU}(\theta_{K-1} \text{ReLU} \dots \text{ReLU}(\theta_1 v)),$$

where  $\theta_i, i = 1 \dots K$ , are per-layer weight matrices. This choice of activation function implies that we are fitting a piece-wise linear approximation to  $\mathcal{X}_p$ . We choose to overparametrize the network such that the total number of trainable parameters  $T = \dim(\theta_1) + \dots + \dim(\theta_K)$ , where  $\dim$  refers to the total number of entries in the matrix, satisfies  $T \gg 3n$ , which is the number of constraints.

Under such overparametrized conditions, one verifies that gradient-descent converges to zero-loss in polynomial time for least-squares regression tasks [10]. By observing that

$$\min_{\theta} \tilde{L}(\theta) = \min_{P \in \mathcal{P}_n} \min_{\theta} \tilde{L}(\theta; P), \text{ with}$$

$$\tilde{L}(\theta; P) = \sum_{i,j \leq n} P_{i,j} \|\phi(v_j; \theta) - x_i\|^2 - \lambda^{-1} H(P),$$

and that  $\tilde{L}(\theta; P)$  is convex with respect to  $P$ , it follows that gradient-descent can find global minimisers of  $\tilde{L}$  in polynomial time.

As  $\lambda \rightarrow \infty$ , the entropic constraint disappears, which implies that by setting  $P$  to any arbitrary permutation matrix  $\Pi$ , we can still obtain zero loss ( $\min_{\theta} \tilde{L}(\theta, \Pi) = 0$ ). In other words, the model has enough capacity to produce any

arbitrary correspondence between the points  $\{v_i\}_i$  and the targets  $\{x_i\}_i$  in the limit  $\lambda \rightarrow \infty$ . A priori, this is an undesirable property of our model, since it would allow highly oscillatory and irregular surface reconstructions. However, our experiments (Section 4) reveal that the gradient-descent optimization of  $\tilde{L}$  remarkably biases the model towards solutions with no apparent oscillation. This implicit regularisation property is reminiscent of similar phenomena enjoyed by gradient-descent in logistic regression [35, 15] or matrix factorization [26]. In our context, gradient-descent appears to be favoring solutions with small complexity, measured for instance in terms of the Lipschitz constant of  $\phi_\theta$ , without the need of explicit regularization. We leave the theoretical analysis of such inductive bias for future work (Section 5). Note that, in practice, we set  $\lambda$  to a large value, which may have an extra regularizing effect.

### 3.2. Building a Global Atlas

Section 3.1 described a procedure to obtain a local chart around a point  $p \in \mathcal{S}$ , with parametric domain  $V_p$  and its associated fitted parametrization  $\varphi_p = \phi(\cdot; \theta_p^*) : V_p \rightarrow \mathbb{R}^3$ . In this section, we describe how to construct an atlas  $\{(V_q, \varphi_q); q \in Q\}$  by appropriately selecting a set of anchor points  $Q$  and by ensuring consistency between charts.

**Consistency.** To define *atlas consistency* more precisely, we need to separate the notions of parametric coordinate-assignment and surface approximation, since the local chart functions  $\varphi_p$  define both. We say that two charts  $p$  and  $q$  overlap, if  $\mathcal{X}_{p,q} = \mathcal{X}_p \cap \mathcal{X}_q \neq \emptyset$ . Each discrete chart  $\varphi_\alpha$  is equipped with a permutation  $\pi_\alpha$ , assigning indices of points in  $\mathcal{X}_p$  to indices of parametric positions in  $V_p$ . Two overlapping charts  $p$  and  $q$  are *consistent* on the surface samples if

$$\varphi_q(v_{\pi_q^{-1}(i)}; \theta_q) = \varphi_p(v_{\pi_p^{-1}(i)}; \theta_p), \forall x_i \in \mathcal{X}_p \cap \mathcal{X}_q$$

i.e., for any point in the patch overlap, the values of the two chart maps at corresponding parametric values coincide. If all chart maps are interpolating, then consistency is guaranteed by construction, but this is in general not the case. We enforce consistency explicitly by minimizing a consistency loss (4).

**Constructing the Atlas.** We construct the set of patch centers  $Q$  using Poisson disk sampling [7] of  $\mathcal{X}$ , with a specified radius,  $r$ . For each  $q \in Q$ , we first extract a neighborhood  $\mathcal{X}_q$  by intersecting a ball of radius  $cr$  centered at  $q$  with  $\mathcal{X}$  ( $\mathcal{X} \cap B(q; cr)$ ), where  $c > 1$  is another hyperparameter. To reduce boundary effects, we consider a larger radius  $\tilde{c} > c$  and use  $\tilde{\mathcal{X}}_q := \mathcal{X} \cap B(q; \tilde{c}r)$  to fit the local chart for  $\mathcal{X}_q$ . In general, the intersection of  $\mathcal{S}$  with the ball

$B(q; \tilde{c}r)$  consists of multiple connected components, possibly of nontrivial genus. We use the heuristic described below to filter out points we expect to be on a different sheet from the ball center  $q$ .

To ensure consistency as defined above, we fit the charts in two phases. In the first phase, we locally fit each chart to its associated points. In the second phase, we compute a joint fitting of all pairs of overlapping charts.

Let  $(\theta_p, \pi_p)$  and  $(\theta_q, \pi_q)$  denote the parameters and permutations of the patches  $p$  and  $q$  respectively at some iteration of the optimization. We compute the first local fitting phase as

$$\min_{\theta_p} \inf_{\pi_p} \sum_{i \leq |\mathcal{X}_p|} \|\phi(v_i; \theta_p) - x_{\pi_p(i)}\|^2. \quad (3)$$

We define the set of indices of parametric points in chart  $p$  of the intersection  $\mathcal{X}_{p,q}$  as

$$\mathcal{T}_{pq} = \{i | x_{\pi_p(i)} \in \mathcal{X}_{p,q}\},$$

where  $\mathcal{T}_{qp}$  is the corresponding set in chart  $q$ . The map between indices of corresponding parametric points in two patches is given by:  $\pi_{p \rightarrow q} := \pi_q \circ \pi_p^{-1} : \mathcal{T}_{pq} \rightarrow \mathcal{T}_{qp}$ .

Equipped with this correspondence, we compute the second joint fitting phase between all patch pairs as:

$$\min_{\theta_p, \theta_q} \inf_{\pi_p, \pi_q} \sum_{i \in \mathcal{T}_{p,q}} \|\phi(v_i; \theta_p) - \phi(v_{\pi_{p \rightarrow q}(i)}; \theta_q)\|^2. \quad (4)$$

Observe that by the triangle inequality,

$$\begin{aligned} \|\phi(v_i; \theta_p) - \phi(v_{\pi_{p \rightarrow q}(i)}; \theta_q)\| &\leq \|\phi(v_i; \theta_p) - x_{\pi_p(i)}\| + \\ &\quad \|\phi(v_{\pi_{p \rightarrow q}(i)}; \theta_q) - x_{\pi_p(i)}\|. \end{aligned}$$

Therefore, the joint fitting term (4) is bounded by the sum of two separate fitting terms (3) for each patch (note that  $x_{\pi_p(i)} = x_{\pi_q(\pi_{p \rightarrow q}(i))}$ ). Consistent transitions are thus enforced by the original Wasserstein loss if the charts are *interpolating*, i.e.  $\phi(v_i; \theta_p) = x_{\pi_p(i)}$  for all  $i, p$ . However, in presence of noisy point clouds, the joint fitting phase enables a smooth transition between local charts without requiring an exact fit through the noisy samples.

If the Sinkhorn distance is used instead of the EMD, then we project the stochastic matrices  $P_q, P_{q'}$  to the nearest permutation matrix by setting to one the maximum entry in each row.

**Filtering Sample Sets  $\mathcal{X}_r$ .** In our experiments, we choose the ball radius to be sufficiently small to avoid most of the artifacts related to fitting patches to separate sheets of the surface-intersection with  $B(q; cr)$ . The radius can be easily chosen adaptively, although at a significant computational cost, by replacing a ball by several smaller balls whenever



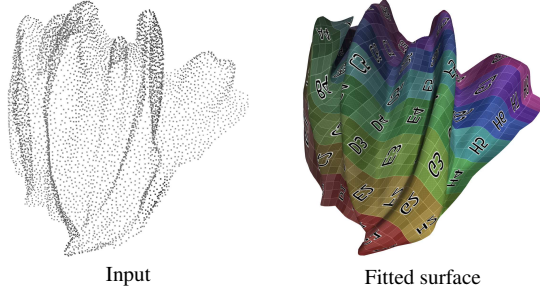


Figure 2. Single patch fitting with uv-mapping illustrated with a checkerboard texture.

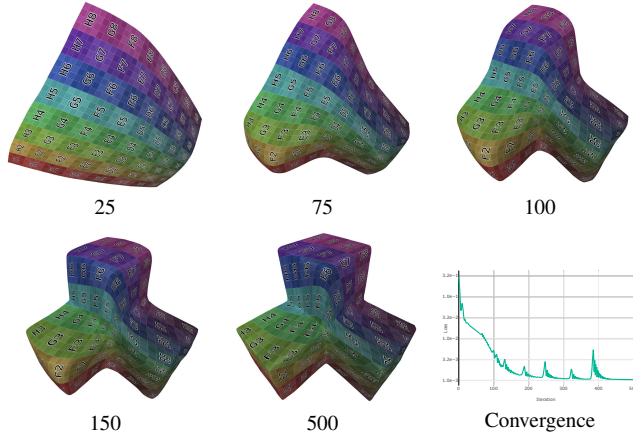


Figure 3. Evolution of single-patch reconstruction as the loss  $\tilde{L}(\theta)$  is minimised. The jumps in the loss are a side-effect of the adaptive gradient descent scheme ADAM.

the quality of the fit is bad. Instead, we use a cheaper heuristic to eliminate points from each set that are likely to be on a different sheet: We assume that the input point cloud  $\mathcal{X}$  also contains normals. If normals are unavailable, they can be easily estimated using standard local fitting techniques. We then discard all vertices in each  $\mathcal{X}_p$  whose normals form an angle greater than a fixed threshold  $\alpha$  with respect to the normal at the center. In all our experiments (Section 4) we used  $r = 2.5\%$  of the bounding box diagonal enclosing the point cloud,  $c = 1.5$   $\tilde{c} = 1.5$ ,  $\alpha = 100$  degrees.

## 4. Experiments

**Experimental Setup.** We run our experiments on a computing node with an Intel(R) Xeon(R) CPU E5-2690 v4, 256 GB of memory, and 4 NVIDIA Tesla P40 GPUs. The runtime of our algorithm are considerably higher than competing methods, requiring around 0.45 minutes per patch, for a total of up to 6.5 hours to optimize an the entire model.

We optimize the losses (3) and (4) using the ADAM [23] implementation in PyTorch with default parameters. Specifically for ADAM, we use a learning rate of  $10^{-3}$ ,  $\beta = (0.9, 0.999)$ ,  $\epsilon = 10^{-8}$ , and no weight decay. For the

Sinkhorn loss, we use a regularization parameter,  $\lambda = 1000$ . For the networks,  $\phi$ , we use an MLP with fully connected layer sizes: (2, 128, 256, 512, 512, 3) and ReLU activations. Our reference implementation is available at <https://github.com/fwilliams/deep-geometric-prior>.

**Single Patch Fitting.** Our first experiment shows the behaviour of a single-patch network overfitted on a complex point cloud (Figure 2 left). Our result is a tight fit to the point cloud. An important side effect of our construction is an explicit local surface parametrization, which can be used to compute surface curvature, normals, or to apply an image onto the surface as a texture map (Figure 2 right).

Figure 3 shows the evolution of the fitting and of the parametrisation  $\varphi_\theta$  as the optimization of  $\tilde{L}(\theta)$  progresses. We observe that the optimization path follows a trajectory where  $\phi_\theta$  does not exhibit distortions, supporting the hypothesis that gradient descent biases towards solutions with low complexity.

**Global Consistency.** As described in Section 3.2, reconstructing an entire surface from local charts requires consistent transitions, leading to the formulation in (3) and (4) which reinforces consistency across overlapping patches. Figure 5 illustrates the effect of adding the extra consistency terms. We verify that these terms significantly improve the consistency.

**Surface Reconstruction Benchmark.** To evaluate quantitatively the performance of our complete reconstruction pipeline, we use the setup proposed by [5], using the first set of range scans for each of the 5 benchmark models. Figure 4 shows the results (and percentage error) of our method on the five models of the benchmark. We compare our results against the baseline methods described in Section 2, and use the following metrics to evaluate the quality of the reconstruction: Let  $\mathcal{X}$  denote the input point cloud. From the ground-truth surface  $\mathcal{S}$  and the reconstruction  $\hat{\mathcal{S}}$ , we obtain two dense point clouds that we denote respectively by  $\mathcal{Y}$  and  $\hat{\mathcal{Y}}$ . We consider

$$\begin{aligned} d_{\text{inp} \rightarrow \text{rec}}(i) &= \min_{\hat{y}_j \in \hat{\mathcal{Y}}} \|x_i - \hat{y}_j\|, i \leq |\mathcal{X}|, \\ d_{\text{rec} \rightarrow \text{GT}}(j) &= \min_{y_k \in \mathcal{Y}} \|\hat{y}_j - y_k\|, j \leq |\hat{\mathcal{Y}}|. \end{aligned} \quad (5)$$

That is,  $d_{\text{rec} \rightarrow \text{GT}}$  measures a notion of precision of the reconstruction, while  $d_{\text{inp} \rightarrow \text{rec}}$  measures a notion of recall. Whereas  $d_{\text{rec} \rightarrow \text{GT}}$  is an indication of overall quality, it does not penalize the methods for not covering undersampled regions of the input. Figure 6 illustrates these one-sided correspondence measures. A naive reconstruction that copies the input satisfies  $d_{\text{inp} \rightarrow \text{rec}} \equiv 0$  but since in general the input point cloud consists of noisy measurements, we will have  $d_{\text{inp} \rightarrow \text{GT}} > 0$ .

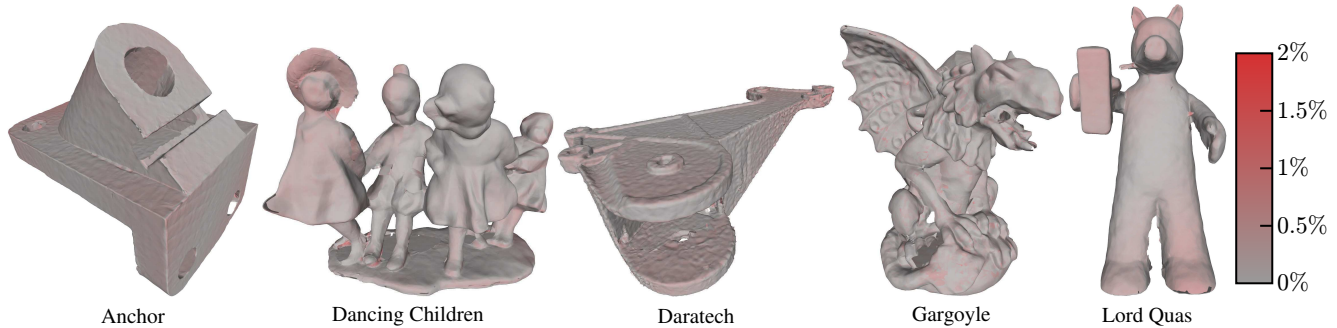


Figure 4. Fitted points on the different models of the benchmark, the color illustrates the error with respect to the ground truth.

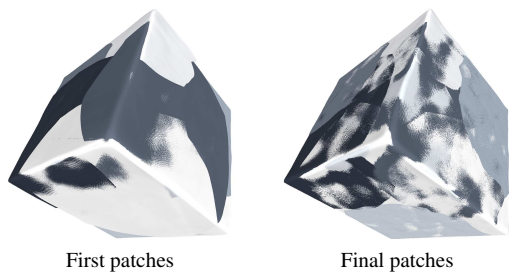


Figure 5. Initial patch fitting (left), we clearly see the “un-mixed” colors due to the disagreeing patches. After optimization (right) all patches are overlapping producing a mix of color.

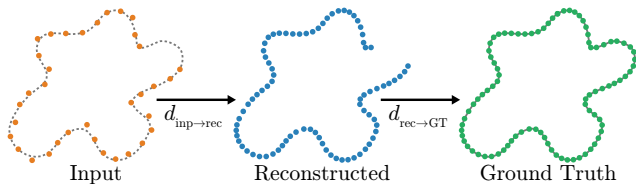


Figure 6. Diagram illustrating the evaluation measures we use to compare different reconstructions.

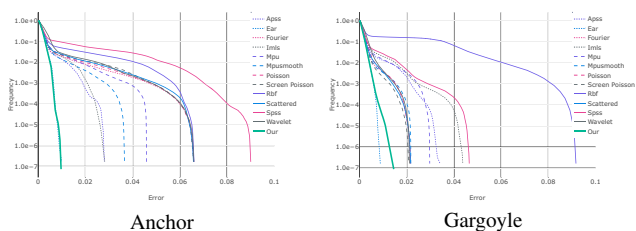


Figure 7. Percentage of fitted vertices ( $y$ -axis, log scale) to reach a given error ( $d_{\text{inp} \rightarrow \text{rec}}$ ,  $x$ -axis) for different methods. The errors are computed from the fitted surface to the ground truth. The plots for the remaining models of the dataset are provided in the supplementary document.

Figures 7 and 8 show respectively the percentage of vertices of  $\hat{\mathcal{Y}}$  and  $\mathcal{X}$  such that  $d_{\text{rec} \rightarrow \text{GT}}$  and  $d_{\text{inp} \rightarrow \text{rec}}$  is below a given error.

Our technique outperforms all the technique we tested,

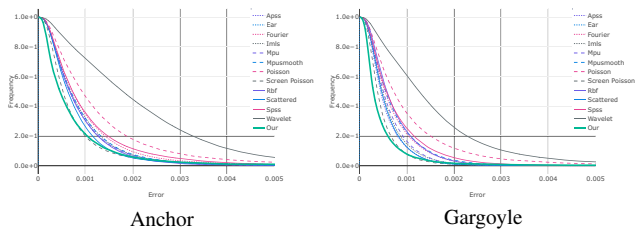


Figure 8. Percentage of fitted vertices ( $y$ -axis) to reach a given error ( $d_{\text{rec} \rightarrow \text{GT}}$ ,  $x$ -axis) for different methods. The errors are measured as distance from the input data to the fitted surface. The plots for the remaining models of the dataset are provided in the supplementary document.

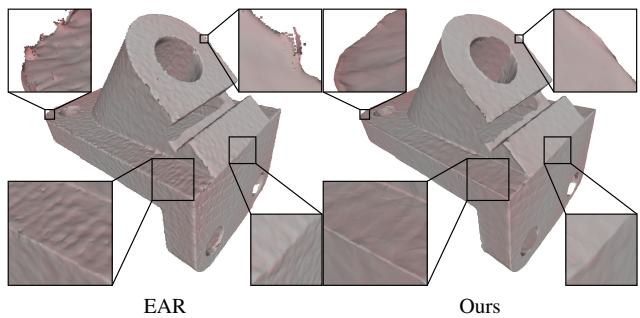


Figure 9. EAR (left) versus Ours (right). The input point cloud contains noise which is smoothed out by our reconstruction (while preserving sharp features quite well) but interpolated by EAR. The result is that EAR produces spurious points and visible artifacts.

and it is on par with the state-of-the art EAR method [19], which achieves a similar score for these 5 models. But, as we will discuss in the next paragraph, EAR is unable to cope with noisy inputs. This is a remarkable result considering that our method, differently from all the others, produces an explicit parametrization of the surface, which can be resampled at arbitrary resolutions, used for texture mapping, or to compute analytic normals and curvature values.

EAR is an interpolative method, which by construction satisfies  $\mathcal{X} \subset \hat{\mathcal{Y}}$ . It follows that the noise in the measurements is necessarily transferred to the surface reconstruction.

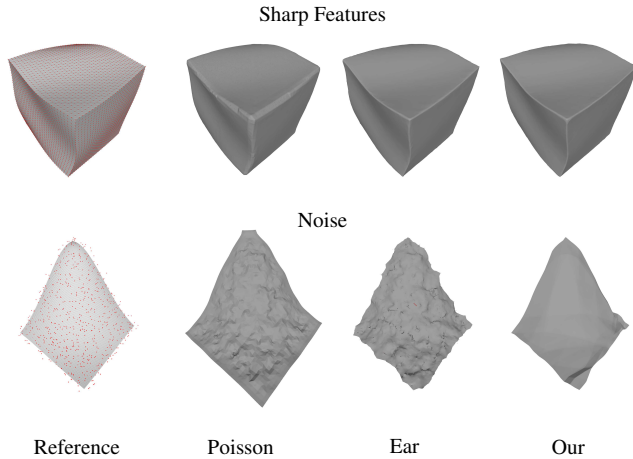


Figure 10. Example of reconstruction on extreme conditions for our method versus the most representative traditional methods. The red dots in the Reference are the input points for the reconstruction.

tion. Figure 9 illustrates that our deep geometric prior preserves the sharp structures while successfully denoising the input point cloud. The mathematical analysis of such implicit regularisation is a fascinating direction of future work.

**Noise and Sharp Features.** As discussed above, the behavior of surface reconstruction methods is particularly challenging in the presence of sharp features and noise in the input measurements. We performed two additional experiments to compare the behaviour of our architecture with the most representative traditional methods on both noisy point clouds and models with sharp features. Screened Poisson Surface Reconstruction [22] (Figure 10, left) is very robust to noise, but fails at capturing sharp features. EAR (Figure 10, middle) is the opposite: it captures sharp features accurately, but being an interpolatory method fails at reconstructing noisy inputs, thus introducing spurious points and regions during the reconstruction. Our method (Figure 10, right) does not suffer from these limitations, robustly fitting noisy inputs and capturing sharp features.

**Generating a Triangle Mesh.** Our method generates a collection of local charts, which can be sampled at an arbitrary resolution. We can generate a triangle mesh by using off-the-shelf-techniques such as Poisson Surface Reconstruction [22] on our dense point clouds. We provide meshes reconstructed in this way for all the benchmark models at <https://github.com/fwilliams/deep-geometric-prior>.

**Comparison with AtlasNet [12].** Our atlas construction is related to the AtlasNet model introduced in [12]. Atlas-

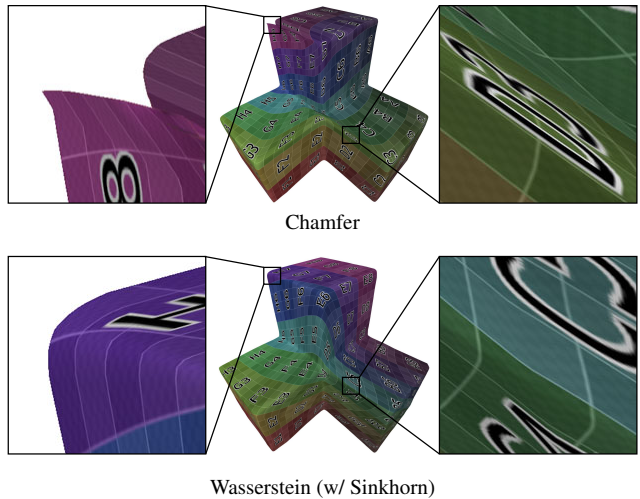


Figure 11. Surface reconstruction of stacked cubes using a single chart, with two different choices of metric. We verify that the Wasserstein metric, even with the Sinkhorn entropy-regularization, provides a more uniform parametrisation, as well as bijective correspondences between overlapping charts.

Net is a data-driven reconstruction method using an autoencoder architecture. While their emphasis was on leveraging semantic similarity of shapes and images on several 3D tasks, we focus on high-fidelity point cloud reconstruction in data-sparse regimes, i.e. in absence of any training data. Our main contribution is to show that in such regimes, an even simpler neural network yields state-of-the-art results on surface reconstruction. We also note the following essential differences between our method and AtlasNet.

- *No Learning:* Our model does not require any training data, and, as a result, we do not need to consider an autoencoder architecture with specific encoders.
- *Transition Functions:* Since we have pointwise correspondences, we can define a transition function between overlapping patches  $V_p$  and  $V_q$  by consistently triangulating corresponding parametric points in  $\mathcal{V}_p$  and  $\mathcal{V}_q$ . In contrast, AtlasNet does not have such correspondences and thus does *not* produce a true manifold atlas.
- *Patch Selection:* We partition the input into point-sets  $\mathcal{X}_p$  that we fit separately. While it is theoretically attractive to attempt to fit each patch to the whole set as it is done in AtlasNet, and let the algorithm figure out the patch partition automatically, in practice the difficulty of the optimization problem leads to unsatisfactory results. In other words, AtlasNet is approximating a *global* matching whereas our model only requires *local* matchings within each patch.



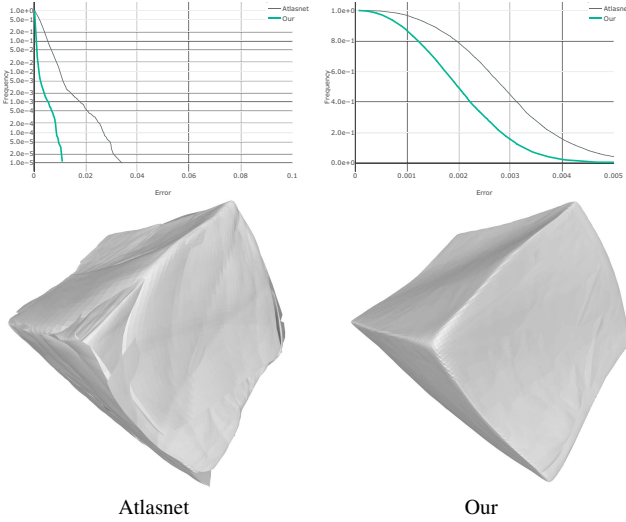


Figure 12. Comparison with Atlasnet. Top:  $d_{\text{rec} \rightarrow \text{GT}}$  and  $d_{\text{inp} \rightarrow \text{rec}}$  cumulative histograms. Bottom: surface reconstruction visualization. Observe that local charts do not have consistent transitions in the Atlasnet output.

- *Wasserstein vs. Chamfer Distance:* As discussed above, the EMD automatically provides transition maps across local charts. AtlasNet considers instead the Chamfer distance between point clouds, which is more efficient to compute but sacrifices the ability to construct bijections in the overlapping regions. Moreover, as illustrated in Figure 11, we observe that Chamfer distances may result in distortion effects even within local charts.
- *Chart Consistency:* We explicitly enforce consistency (4) which has a significant effect on quality, as illustrated in Section 4, whereas AtlasNet does not produce a real manifold atlas, since it has no definition of transition maps.

We provide quantitative and qualitative comparisons to assess the impact of our architecture choices by adapting AtlasNet to a data-free setting. In this setting, we overfit AtlasNet on a single model with the same number of patches used for our method. Figure 12 reports both  $d_{\text{rec} \rightarrow \text{GT}}$  and  $d_{\text{inp} \rightarrow \text{rec}}$  cumulative histograms on a twisted cube surface using 10 local charts. We verify that when the Atlasnet architecture is trained to fit the surface using our experimental setup, it is clearly outperformed both quantitatively and qualitatively by our deep geometric prior. We emphasize however that AtlasNet is designed as a data-driven approach, and as such it can leverage semantic information from large training sets.

## 5. Discussion

Neural networks – particularly in the overparametrised regime – are remarkably efficient at *curve fitting* in high-dimensional spaces. Despite recent progress in understanding the dynamics of gradient descent in such regimes, their ability to learn and generalize by overcoming the curse of dimensionality remains a major mystery. In this paper, we bypass this high-dimensional regime and concentrate on a standard low-dimensional interpolation problem: surface reconstruction. We demonstrate that in this regime neural networks also have an intriguing capacity to reconstruct complex signal structures while providing robustness to noise.

Our model is remarkably simple, combining two key principles: (i) construct local piece-wise linear charts by means of a vanilla ReLU fully-connected network, and (ii) use Wasserstein distances in each neighborhood, enabling consistent transitions across local charts. The resulting architecture, when combined with gradient descent, provides a “deep geometric prior” that is shown to outperform existing surface-reconstruction methods, which rely on domain-specific geometric assumptions. The theoretical analysis of this deep geometric prior is our next focus, which should address questions such as how the geometry of the surface informs the design the neural network architecture, or why is gradient descent biasing towards locally regular reconstructions.

Despite these promising directions, we also note the limitations our approach is facing. In particular, our method is currently substantially more expensive than alternatives. One natural possibility to accelerate it, would be to train a separate neural network model to provide an efficient initialization for the local chart minimisation (2), similarly as in neural sparse coders [11]. Another important question for future research is the adaptive patch selection, which would leverage the benefits of multiscale approximations.

## 6. Acknowledgements

We are grateful to the NYU HPC staff for providing computing cluster service. This project was partially supported by the NSF CAREER award 1652515, the NSF grant IIS-1320635, the NSF grant DMS-1436591, the NSF grant DMS-1835712, the NSF grant RI-IIS 1816753, the SNSF grant P2TIP2\_175859, the Alfred P. Sloan Foundation, the Moore-Sloan Data Science Environment, the DARPA D3M program, NVIDIA, Samsung Electronics, Labex DigiCosme, DOA W911NF-17-1-0438, a gift from Adobe Research, and a gift from nTopology. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA. The authors would also like to thank the anonymous reviewers for their time and effort.



## References

- [1] A. Adamson and M. Alexa. Approximating and intersecting surfaces from points. In *SGP*, page 239, 2003. 2
- [2] M. Alexa, J. Behr, D. Fleishman, D. Levin, and C. Silva. Point Set Surfaces. In *IEEE VIS*, page 21, 2001. 2
- [3] J. Altschuler, J. Weed, and P. Rigollet. Near-linear time approximation algorithms for optimal transport via sinkhorn iteration. In *Advances in Neural Information Processing Systems*, pages 1964–1974, 2017. 3
- [4] R. Basri and D. Jacobs. Efficient representation of low-dimensional manifolds using deep networks. *arXiv preprint arXiv:1602.04723*, 2016. 2
- [5] M. Berger, J. A. Levine, L. G. Nonato, G. Taubin, and C. T. Silva. A benchmark for surface reconstruction. *ACM TOG*, 32(2):20:1–20:17, 2013. 2, 5, 11
- [6] M. Berger, A. Tagliasacchi, L. M. Seversky, P. Alliez, G. Guennebaud, J. A. Levine, A. Sharf, and C. T. Silva. A survey of surface reconstruction from point clouds. *CGF*, 36(1):301–329, 2017. 2
- [7] J. Bowers, R. Wang, L.-Y. Wei, and D. Maletzt. Parallel poisson disk sampling with spectrum analysis on surfaces. In *ACM Transactions on Graphics (TOG)*, volume 29, page 166. ACM, 2010. 4
- [8] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013. 3
- [9] A. Dai, C. R. Qi, and M. Nießner. Shape completion using 3d-encoder-predictor cnns and shape synthesis. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, volume 3, 2017. 2
- [10] S. S. Du, X. Zhai, B. Póczos, and A. Singh. Gradient descent provably optimizes over-parameterized neural networks. *arXiv preprint arXiv:1810.02054*, 2018. 3
- [11] K. Gregor and Y. LeCun. Learning fast approximations of sparse coding. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, pages 399–406. Omnipress, 2010. 8
- [12] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry. Atlasnet: A papier-mâché approach to learning 3d surface generation. *arXiv preprint arXiv:1802.05384*, 2018. 2, 7
- [13] G. Guennebaud and M. Gross. Algebraic Point Set Surfaces. *SIGGRAPH*, 26:23–9, 2007. 2
- [14] P. Guerrero, Y. Kleiman, M. Ovsjanikov, and N. J. Mitra. PCPNet: Learning local shape properties from raw point clouds. *Computer Graphics Forum*, 37(2):75–85, 2018. 2
- [15] S. Gunasekar, J. Lee, D. Soudry, and N. Srebro. Characterizing implicit bias in terms of optimization geometry. *arXiv preprint arXiv:1802.08246*, 2018. 4
- [16] X. Han, Z. Li, H. Huang, E. Kalogerakis, and Y. Yu. High-resolution shape completion using deep neural networks for global structure and local geometry inference. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2017. 2
- [17] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle. Surface reconstruction from unorganized points. In *SIGGRAPH*, pages 71–78, 1992. 2
- [18] H. Huang, S. Wu, M. Gong, D. Cohen-Or, U. M. Ascher, and H. R. Zhang. Edge-aware point set resampling. *ACM TOG*, 32(1):9:1–9:12, 2013. 2
- [19] Q. Huang, F. Wang, and L. Guibas. Functional map networks for analyzing and exploring large shape collections. In *SIGGRAPH*, 2014. 6
- [20] M. Kazhdan. Reconstruction of solid models from oriented point sets. In *SGP*, page 73, 2005. 2
- [21] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *SGP*, page 70, 2006. 2
- [22] M. M. Kazhdan and H. Hoppe. Screened poisson surface reconstruction. *ACM TOG*, 32(3):29:1–29:13, 2013. 2, 7
- [23] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5
- [24] R. Kolluri. Provably Good Moving Least Squares. In *SODA*, pages 1008–1018, 2005. 2
- [25] H. W. Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955. 3
- [26] Y. Li, T. Ma, and H. Zhang. Algorithmic regularization in over-parameterized matrix recovery. *arXiv preprint arXiv:1712.09203*, 2017. 4
- [27] J. Manson, G. Petrova, and S. Schaefer. Streaming surface reconstruction using wavelets. 27(5):1411–1420, 2008. 2
- [28] Y. Nagai, Y. Ohtake, and H. Suzuki. Smoothing of Partition of Unity Implicit Surfaces for Noise Robust Surface Reconstruction. 28(5):1339–1348, 2009. 2
- [29] Y. Ohtake, A. Belyaev, M. Alexa, G. Turk, and H. Seidel. Multi-level partition of unity implicits. *ACM TOG*, 22(3):463–470, 2003. 2
- [30] Y. Ohtake, A. Belyaev, and H. Seidel. An integrating approach to meshing scattered point data. In *Symposium on Solid and Physical Modeling*, page 69, 2005. 2

- [31] Y. Ohtake, A. G. Belyaev, and H.-P. Seidel. 3d scattered data interpolation and approximation with multilevel compactly supported rbfs. *Graphical Models*, 67(3):150–165, 2005. 2
- [32] N. Pietroni, M. Tarini, O. Sorkine, and D. Zorin. Global parametrization of range image sets. In *ACM Transactions on Graphics (TOG)*, volume 30, page 149. ACM, 2011. 2
- [33] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *CVPR*, 2017. 2
- [34] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv preprint arXiv:1706.02413*, 2017. 2
- [35] D. Soudry, E. Hoffer, M. S. Nacson, S. Gunasekar, and N. Srebro. The implicit bias of gradient descent on separable data. *arXiv preprint arXiv:1710.10345*, 2017. 4
- [36] S. Sukhbaatar, R. Fergus, et al. Learning multiagent communication with backpropagation. In *Advances in Neural Information Processing Systems*, pages 2244–2252, 2016. 2
- [37] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. *arXiv preprint arXiv:1711.10925*, 2017. 1, 3
- [38] J. Varley, C. DeChant, A. Richardson, J. Ruales, and P. Allen. Shape completion enabled robotic grasping. In *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*, pages 2442–2447. IEEE, 2017. 2
- [39] O. Vinyals, S. Bengio, and M. Kudlur. Order matters: Sequence to sequence for sets. *arXiv preprint arXiv:1511.06391*, 2015. 2
- [40] S. Xiong, J. Zhang, J. Zheng, J. Cai, and L. Liu. Robust surface reconstruction via dictionary learning. *ACM Transactions on Graphics (TOG)*, 33(6):201, 2014. 2
- [41] L. Yu, X. Li, C.-W. Fu, D. Cohen-Or, and P.-A. Heng. Pu-net: Point cloud upsampling network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2790–2799, 2018. 2