

Meshlet Priors for 3D Mesh Reconstruction

Abhishek Badki^{1,2} Orazio Gallo¹ Jan Kautz¹ Pradeep Sen²

¹NVIDIA ²University of California, Santa Barbara

Abstract

Estimating a mesh from an unordered set of sparse, noisy 3D points is a challenging problem that requires to carefully select priors. Existing hand-crafted priors, such as smoothness regularizers, impose an undesirable trade-off between attenuating noise and preserving local detail. Recent deep-learning approaches produce impressive results by learning priors directly from the data. However, the priors are learned at the object level, which makes these algorithms class-specific, and even sensitive to the pose of the object. We introduce meshlets, small patches of mesh that we use to learn local shape priors. Meshlets act as a dictionary of local features and thus allow to use learned priors to reconstruct object meshes in any pose and from unseen classes, even when the noise is large and the samples sparse.

1. Introduction

The ability to capture, represent, and digitally manipulate objects is crucial for a wide range of important applications, from content creation to animation, robotics, and virtual reality. Among the different representations for 3D objects (which also include depth maps, occupancy grids, and point clouds), meshes are particularly appealing.

Estimating meshes of real-world objects, however, is not straightforward since common capture strategies, such as structured light [30] or multi-view stereo [14, 36], produce point clouds or depth maps instead. These intermediate representations are noisy, sparse, and, when used to estimate a continuous surface, they introduce a trade-off between overfitting to the noise and over-smoothing. Traditional methods require hand-crafted priors (e.g., local smoothness) to balance noise and details, as is the case for Laplacian reconstruction [27]. Figure 1 shows that this balance is difficult to strike when the point cloud is noisy (rows marked as (N)). Recent learning-based methods learn priors directly from a large number of examples [16, 34, 15, 26, 29]. Because

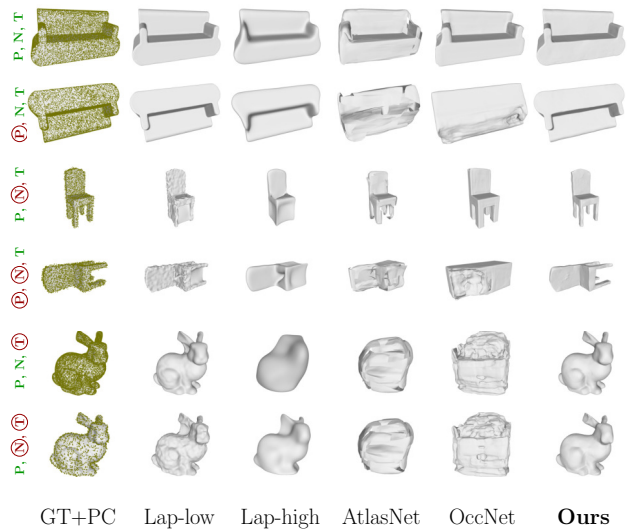


Figure 1: Mesh reconstructions for objects in pose (P) and out of pose ((P)), under low (N) and moderate ((N)) noise, and for objects in (T) and out ((T)) of training. GT+PC ground truth mesh and the point cloud used by the various methods to estimate the mesh. Traditional methods introduce a noise vs. smoothness trade-off (Laplacian low/high [27]). State-of-the-art, deep-learning methods (AtlasNet [11] and OccNet [26]) learn object-level priors, which causes them to fail on objects not seen in training ((T)), or even on objects that are *just rotated* w.r.t. the training set ((P)). Our method learns local priors and forces global consistency with the point cloud.

of their ability to learn priors directly from the data, these approaches can produce impressive results from both point clouds and single images. However, they learn priors *at the object level*, which limits their ability to reconstruct objects from classes not seen during training (rows marked as ((T))). They also struggle to disentangle the shape priors with the object pose: state-of-the-art learning methods can fail completely on a rotated point cloud (rows marked as ((P))) even though they can reconstruct the same point cloud when its pose resembles that of the training set (rows marked as (P)). Tatarchenko *et al.* suggest that many of these methods may

This work was done while A. Badki was interning at NVIDIA.

actually learn a form of classification and nearest-neighbor retrieval from the dataset, rather than a proper 3D reconstruction [33]. In fact, even for rows P and T in Figure 1, a closer look seems to indicate that AtlasNet [11] and OccNet [26] are reconstructing a different couch and chair from the training set.

We present a learning-based method to extract a 3D mesh from a set of sparse, noisy, unordered points that bridges traditional and learning-based approaches. Our key intuition is to learn geometric priors *locally* while enforcing their consistency *globally*. To represent shape priors, we introduce *meshlets*, small patches of mesh that, loosely speaking, serve as a learned dictionary of local features. Specifically, we use a variational auto-encoder (VAE) [20] to learn the latent space of meshlets that can be observed in natural shapes. We call these natural meshlets. Learning these local features offers two key advantages. First, it allows to reconstruct objects from classes never seen in training: at some scale, a couch exhibits similar local features to those of a bunny. Second, it disentangles the global pose of the object and the parametrization learned by the network, which allows our algorithm to be robust to dramatic changes of the object’s pose, see Figure 1.

To fit the meshlets to a point cloud we minimize their distance to the points, while enforcing that they belong to the latent space of natural meshlets. Therefore, the resulting surface will locally satisfy the priors we learned. However, because the meshlets are optimized independently of each other, the mesh extracted from their union will not be watertight. Therefore, we define an auxiliary, watertight mesh and propose to use it in an alternating optimization that ensures that the meshlets are consistent with each other and with the observed point cloud.

We show with extensive comparisons that this iterative method produces results that outperform the state-of-the-art on challenging scenarios such as noisy points clouds in arbitrary poses, as shown in Figure 1. In summary, our contributions are:

- We present meshlets, a new way of representing local shape priors in the latent space of a variational auto-encoder that is trained on local patches from a dataset of real-world objects.
- We propose an alternating optimization which fits meshlets to the measured point samples (enforcing local constraints) while maintaining global consistency for the mesh.
- We demonstrate for the first time, to our knowledge, successful reconstructions of 3D meshes from very sparse, noisy point measurements with a category-agnostic, learning-based method.

2. Related Work

Extracting a mesh from a point cloud is an important problem that has been the focus of much research since the early days of graphics. Traditional methods such as marching cubes [24] or Ball-Pivoting [4] work well for cases where the noise is small as compared to the density of the point cloud.

In general, however, noise does pose issues. One traditional solution, then, is to use the points and their normals to compute a signed-distance function whose zero crossing is the desired surface [8, 13, 2, 17, 18]. An alternative is to use hand-crafted priors, such as smoothness of the vertices and normals of the estimated mesh [27]. However, these priors introduce a trade-off between suppressing the noise and preserving sharp features that becomes increasingly brittle for sparser and noisier point clouds, see Figure 1.

Priors can be more effectively learned from data with neural networks. Deep learning methods, for instance, have shown great success in estimating depth maps from images, whether from multiple views [14, 36], stereo [19], or even single-image [9, 10, 39, 21]. Even meshes can be directly extracted from a single image, provided that the class of the object is known [16, 34, 15].

Rather than requiring to manually tinker with the traditional noise/sharpness trade-off, methods that learn priors to extract meshes from point clouds introduce a new one: generally speaking, the lower the quality of the observations (*e.g.*, strong noise or sparsity of the point cloud), the stronger the priors need to be, thus affecting the algorithm’s ability to generalize to different and unseen classes. For instance, methods that learn local priors are class-agnostic but tend to need dense point clouds with low levels of noise [12, 38, 37, 35]. The recent works of Park *et al.* [29], Groueix *et al.* [11], and Mesheder *et al.* [26] produce impressive results even with sparser and potentially noisier data, but fail to generalize to completely new classes of objects. Often they even struggle when the point cloud is in a pose that differs significantly from the training pose, as show in Figure 1, rows (P). This issue is due, in part, to the fact that these methods lack a mechanism to enforce geometric constraints at inference time. Our method is class-agnostic thanks to its ability to learn and enforce local priors while minimizing the error with respect to the point cloud at inference time.

The idea of learning priors from data and enforcing geometric constraints at inference time was recently explored for depth map [5], point cloud [40] and surface estimation [23, 22, 29]. These approaches use low dimensional representations that allow inference time optimization. However, approaches that learn priors at the object level tend to be category specific. Our meshlet priors directly encode the (local) shape of the surface instead of a viewer centric depth [5]. Meshlets are class agnostic and

can be used to learn and enforce priors at different scales.

Key to solving the mesh estimation problem is how to represent it. Different representations for meshes exist that are amenable to use with neural networks, but they tend to also be class specific [31, 3]. One key ingredient of our method is the use of small mesh patches, called meshlets, which simplify the processing of the mesh, among other things. A related approach is the work of Groueix *et al.* who also represent the mesh as a collection of large parts, which they call charts [11]. However, their method does not offer a mechanism to enforce global consistency and does not leverage local shape priors.

3. Method

Our goal is to estimate a mesh from a set of unordered, non-oriented points. The task is easy when the point cloud is dense and the noise is low. However, when the quality of the observations degrades, *e.g.*, sparser or noisier points, the choice of priors and heuristics becomes central. Hand-crafted priors, such as smoothness, introduce a trade-off between overly smooth and noisy reconstructions, as shown in Figure 1 (Lap-low/high). On the other hand, neural networks can learn priors directly from data, but they introduce other challenges. First, capturing the distribution of generic objects requires training on a large number of examples, possibly larger than what existing datasets can supply. Moreover, generalization can be an issue: the performance of existing learning-based methods quickly degrades when the test objects differ from the ones used in training, as shown in the rows marked as (T) in Figure 1. Finally, it is not straightforward to disentangle object-level priors and the pose of the object. Figure 1 shows that OccNet [26] and AtlasNet [11], both recent state-of-the-art works, fail for classes never seen in training (T), or even when the pose of the object is significantly different from the poses seen in training (P).

To overcome these issues, we propose to learn priors *locally*: even if the Stanford Bunny in Figure 1 was never seen in training, its local features are similar to those found in more common objects from the training set. We introduce *meshlets*, which can be regarded as small patches of a mesh, see Figure 2. Loosely speaking, meshlets act as a dictionary of basic shape features. Meshlets are local and of limited size, and thus offer a simple mechanism to disentangle the (local) priors from the object’s pose. If meshlets are adapted to the point cloud independently of each other, however, they may not result in a watertight surface. Therefore, we explicitly enforce their consistency *globally*. In the following we describe these two stages, and the overall process to extract a mesh.

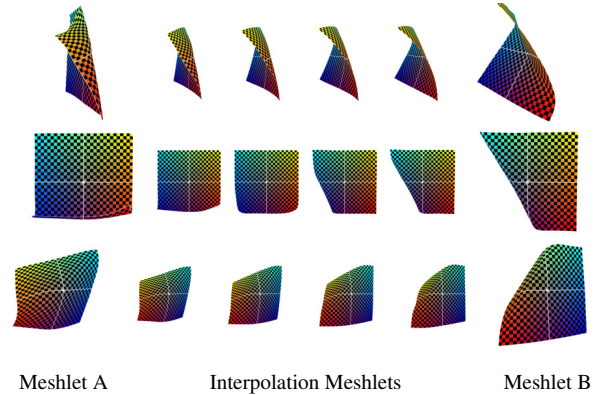


Figure 2: Smoothness of the latent space. We can progressively deform one mesh onto another by interpolating between the corresponding points in latent space (see Section 3.1).

3.1. Local Shape Priors with Meshlets

In this section we introduce meshlets, and describe how we leverage them to enforce local shape priors. Intuitively, a meshlet m is a small patch of mesh deformed to adhere to a region of another, larger mesh, see Figure 2 and 7(a). To extract meshlet m at vertex v of mesh \mathcal{M} , we first compute a local geodesic parametrization [25] that maps the 3D coordinates of the vertices in a neighborhood of v to coordinates on π_T , the plane tangent to \mathcal{M} at v . We then re-sample the geodesic distance function π_T at integer coordinates (μ, ν) . This gives us the correspondence between a vertex on the meshlet at (μ, ν) , and a vertex on the mesh in the neighborhood of v .

Because they only require a local parametrization computed with respect to the center vertex v , meshlets work well even for objects with large, varying curvature. Because they are local, they can learn shape priors that are independent of the pose and class of the object.

Learning local shape priors with meshlets. We want to learn the distribution of “natural meshlets,” *i.e.*, those meshlets that capture the local features of real-world objects. Inspired by recent methods [40, 5], we use a variational auto-encoder (VAE). By training the VAE to reconstruct a large number of meshlets, we force its bottleneck to learn the latent space of natural meshlets. Differently put, vectors sampled on this manifold and fed into the decoder result in natural meshlets. We extract meshlets from objects from the ShapeNet dataset [6] and we feed their 3D coordinates for training. However, we first roto-translate the meshlets to bring them into a canonical pose. This transformation is necessary to make sure that similar meshlets sampled from different 3D locations and orientations map to similar regions in the VAE’s latent space. More specifically, given a

meshlet m_i , we first translate and rotate it so that its center c_i is at the origin, and the normal at c_i is aligned with the z-axis, then we rotate it around the z axis so that the local (μ, ν) coordinates of the meshlet are aligned with the x and y axes. We call this the canonical pose. A meshlet, then, is completely defined by P_i , the transformation from global to canonical pose, and l_i , the latent vector corresponding to the meshlet in canonical pose:

$$m_i = \{l_i, P_i\}. \quad (1)$$

Section 4.2 details the network’s architecture. Since we disentangle pose and shape, smoothly traversing the latent space will smoothly vary the shape of the reconstructed meshlet as shown in Figure 2, where we take the latent vectors l_A and l_B corresponding to meshlets A and B, and we progressively interpolate between them to get vectors l_I ’s. The meshlets reconstructed from the l_I ’s smoothly interpolate between the shape of meshlets A and B.

Fitting a meshlet to 3D points. Assume now that we are given a set of 3D points roughly corresponding to the size of a meshlet (we will generalize this to a complete point cloud in Section 3.2). Deforming a natural meshlet to fit it is now straightforward: we simply traverse the latent space learned by the VAE to minimize the distance between the meshlet and the points. Specifically, we take $m_i(t_0)$, an initialization of the meshlet, and run it through the encoder to find the corresponding latent vector $l_i(t_0)$. This is the starting point of our optimization. We then freeze the weights of the VAE, compute the error between the meshlet and the points, and take a gradient descent step through the decoder. This brings us to a new point in latent space, $l_i(t_1)$, and the corresponding meshlet $m_i(t_1)$. Meshlet $m_i(t_1)$ is a natural meshlet that is closer to the given 3D points. We iterate until convergence, see Figure 3. We note that, although other approaches have also proposed to optimize the latent vector of a VAE to match some measured samples (e.g., [29, 5, 1, 23]), they do so at the object (or scene) level. Because our method learns local surface patches, and therefore reuse surface priors across different object categories, it can better generalize.

3.2. Overall Optimization

Having explained how our meshlets can be used to learn local priors, and can be fit to a set of 3D points, we can describe the overall algorithm, which is fairly straightforward at its core. We start with $\mathcal{M}(t_0)$, an initial, rough approximation of the complete mesh. This could be a sphere, or any other surface that satisfies our meshlet priors, i.e., meshlets extracted from $\mathcal{M}(t_0)$ lie on the manifold learned by the VAE. From $\mathcal{M}(t_0)$ we extract N overlapping meshlets $m_i(t_0)$ ’s and find the corresponding $l_i(t_0)$ ’s and $P_i(t_0)$ ’s,

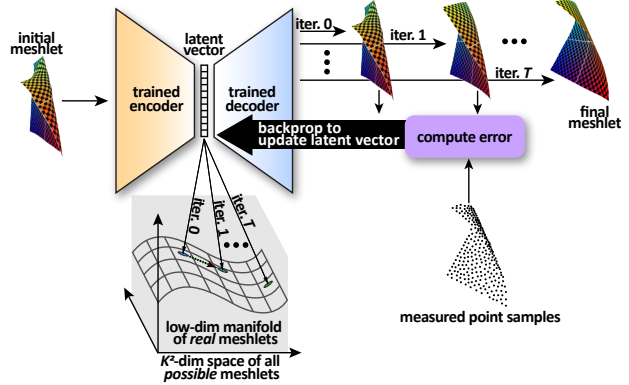


Figure 3: Optimization of our meshlets using the learned latent-space as a prior. By backpropagating the error with respect to the measured points and using it to update the meshlet’s latent vector, we are effectively moving along the low-dimensional manifold of real meshlets while fitting the points.

Figure 5. We select N so that each vertex on $\mathcal{M}(t_0)$ is covered by at least 3 meshlets. Generally, this results in 500 to 1,500 meshlets. We also find the distance between $\mathcal{M}(t_0)$ and the point cloud, whose gradient we can propagate to the meshlets since we have the correspondences between mesh and meshlets by construction. This allows us to update the meshlets to adapt to the points (Section 3.1). However, this optimization is performed on each meshlet independently, so it results in small gaps between the meshlets, see Figure 4(a). Therefore, we enforce and maintain global consistency by adding a step in which we deform \mathcal{M} to match the meshlets and update the meshlets to match \mathcal{M} . Deforming \mathcal{M} brings it closer to the point cloud, deforming the meshlets forces them to be globally consistent. Finally we iterate:

1. Optimize meshlets to fit the point cloud (3.2.1).
2. Optimize meshlets and mesh to match each other (3.2.2).

At convergence, the auxiliary variable \mathcal{M} , watertight by construction, is our estimation of the mesh. We now explain the two steps in detail.

3.2.1 Enforcing Local Shape Priors

To optimize the N meshlets $\{m_i(t_0)\}_{i=1:N}$ with respect to the point cloud we need to define an error. Unfortunately, the correspondences between the point cloud and the vertices of the meshlets are not readily available. A Chamfer distance, then, is not straightforward to use because without correspondences all the points in the point cloud would contribute to the error of all the meshlets—even if they are on opposite sides of the object. However, we do have the correspondences between the vertices of $\mathcal{M}(t_0)$ and the mesh-

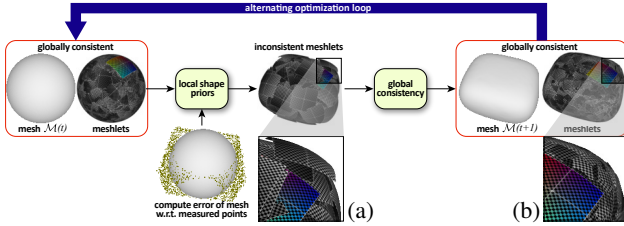


Figure 4: Alternating optimization used in our algorithm. The first stage updates the meshlets based on the errors between the underlying mesh and the measured point cloud. However, because meshlets are localized representations, optimizing them individually causes inconsistencies across the object. Hence, in the second stage we enforce global consistency across all meshlets to reconstruct an updated version of the mesh which is used in the next iteration of the algorithm.

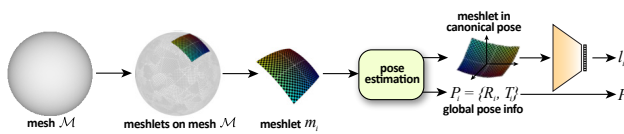


Figure 5: Encoding of meshlets on a given mesh, \mathcal{M} . Each meshlet is represented by a latent vector l_i in the low-dimensional manifold as well as its global pose P_i (composed of rotation R_i and translation T_i between the canonical space and global coordinates).

lets. Therefore, we compute the Chamfer distance between the point cloud and the mesh instead:

$$C^{\text{PC}} = \sum_{v_j \in \mathcal{M}(t_0)} \min_{p \in \text{PC}} \|v_j - p\|_2^2 + \sum_{p \in \text{PC}} \min_{v_j \in \mathcal{M}(t_0)} \|v_j - p\|_2^2, \quad (2)$$

where p is a 3D point in the input point cloud, PC. Equation 2 gives us per-vertex error on the mesh, which we can propagate to the corresponding meshlets. We then update the meshlets to minimize C^{PC} as explained in Section 3.1, and get a new set of natural meshlets $\{m_i(t_1)\}_{i=1:N}$.

3.2.2 Enforcing Global Consistency

To enforce that meshlets $\{m_i(t_1)\}_{i=1:N}$ are globally consistent, *i.e.*, that their union is a watertight mesh, we use, once again, \mathcal{M} . Specifically, we compute the Chamfer distance between the vertices of \mathcal{M} and the vertices of *all* the meshlets as

$$C^m = \sum_{v_j \in \mathcal{M}} \min_{v_k \in \{m_i\}_{i=1:N}} \|v_j - v_k\|_2^2 + \sum_{v_k \in \{m_i\}_{i=1:N}} \min_{v_j \in \mathcal{M}} \|v_j - v_k\|_2^2. \quad (3)$$

First we keep the meshlets fixed and deform $\mathcal{M}(t_0)$ to minimize C^m . Then we fix the resulting mesh $\mathcal{M}(t_1)$ and adjust

the meshlets with the algorithm described in Section 3.1, but this time to minimize C^m . We iterate until Equation 3 is minimized. At this point the meshlets will be consistent with the mesh and, in turn, *globally*. This process corresponds to the block “global consistency” in Figure 4.

4. Implementation Details

4.1. Optimization

We start by describing a few details that improve the efficiency of the optimization procedure or the quality of the resulting meshes.

Mesh initialization. The auxiliary mesh $\mathcal{M}(t_0)$ can be any genus-zero mesh that satisfies the meshlets’ priors (see Section 3.2). The actual choice, however, does have a bearing on the number of iterations required to converge. We initialize our approach with an overly-smoothed Laplacian reconstruction. Empirically, we have observed that the results of our algorithm initialized in this way are effectively indistinguishable from the results obtained by using a sphere as an initialization; convergence, however, does take a fraction of the time. For reference, we show a few examples of $\mathcal{M}(t_0)$ in the Supplementary.

Meshlets re-sampling. As the optimization progresses, the shape and the size of the auxiliary mesh \mathcal{M} may change significantly. On the one hand, this is a desirable behavior: if the mesh can scale to arbitrary sizes, it can properly match the size of the underlying mesh, even when the initialization is far from it. On the other, it results in a sparser meshlet coverage and, potentially, no coverage in some areas. Moreover, it could cause meshlets to be overly-stretched. Therefore, every 20 iterations of enforcing local shape priors and global consistency (blue arrow in Figure 4), we re-sample the meshlets on the current mesh.

Re-meshing. Large changes from the initialization may also cause issues to the mesh itself, which may stretch in some regions or become otherwise irregular. One way to prevent this is to use strong smoothness priors when enforcing global consistency, but that would hinder our ability to reconstruct sharp features. At the end of every iteration, we re-mesh \mathcal{M} using Screened Poisson Reconstruction [18] to encourage smoothness while respecting the priors enforced by our approach, *i.e.*, preserving the sharpness of local features. We provide more details in the Supplementary.

4.2. Meshlet training

To train the meshlets network, we sample 2.2×10^6 meshlets from the ShapeNet dataset [6]. We extract meshlets by randomly selecting objects across several classes.

We then apply three different scales to each object and extract 256 meshlets for each scale, so that our meshlet dataset captures both fine and coarse details. Note that we disregard meshlets that are problematic. Specifically, we use the geodesic distance algorithm by Melv er *et al.* [25] and reject those meshlets for which the geodesic distance calculation results in a large anisotropic stretch, or fails altogether. The network, then, is trained to reconstruct these meshlets using ℓ_2 as a loss. In all of our experiments we use meshlets of size $31 \times 31 \times 3$. To exploit the latent space of natural meshlets, we use a fully-connected encoder decoder network that takes as input a $(31 \cdot 31) \times 3$ vector (*i.e.*, a vectorized version of the meshlet). The encoder and the decoder are symmetric with 6 layers each, and the latent code vector is one third of the input dimension.

5. Experiments

Comparisons. In this section we evaluate our method against state-of-the-art approaches. Then, we compare our meshlets to other local shape priors to validate their importance.

We compare our method with several state-of-the-art mesh reconstruction approaches. The first is Screened Poisson [18], a widely used, traditional technique that creates watertight surfaces from oriented point clouds. Because our input is a raw point cloud, then, we need to estimate normals. We use two methods to estimate normals. One is MeshLab’s normal estimation, which fits local planes and uses them to estimate normals [7]. The other is a recently published, learning-based method called PCP-Net [12]. The second mesh reconstruction approach is the method by  ztireli *et al.* [28], which also requires oriented points. They propose a recent variant of marching cubes that preserves sharp features using non-linear regression. In addition, we compare against Laplacian mesh optimization [27]. Leveraging the fact that the norm of the mesh’s Laplacian captures the local mean curvature, this mesh optimization algorithm optimizes the Laplacian at the vertices in a weighted least-square sense. The algorithm has a free parameter that regulates the smoothness of the resulting surface. After a parameter sweep we found that no single parameter would yield the best results over the whole dataset. Therefore we settled for two values, each offering a different compromise between denoising and over-smoothing. We also compare with Deep Geometric Prior (DGP) [35], OccNet [26], and AtlasNet [11], all of which are deep learning methods. The last two approaches learn priors at the object level.

Data. We test all the methods on 20 objects. To validate that our method generalizes well, we also include four objects that are commonly used by the graphics community (Suzanne, the Stanford Bunny, Armadillo, and the Utah

		Chamfer- ℓ_1	Hausdorff
Meshlab [7] +	Scr.Pois. [18]	0.0285 / 0.0112	0.339 / 0.102
	RILMS [28]	0.0177 / 0.0166	0.149 / 0.148
PCPNet [12] +	Scr.Pois. [18]	0.0122 / 0.0109	0.147 / 0.140
	RILMS [28]	0.0181 / 0.0176	0.151 / 0.153
Laplacian [27]	Low	0.0104 / 0.0103	0.100 / 0.065
	High	0.0096 / 0.0094	0.103 / 0.069
Deep Geometric Prior [35]		0.0128 / 0.0130	0.147 / 0.148
AtlasNet [11]		0.0415 / 0.0377	0.293 / 0.263
OccNet [26]		0.0630 / 0.0627	0.304 / 0.285
Ours		0.0090 / 0.0092	0.054 / 0.047

Table 1: We compare our method with state-of-the-art approaches, both traditional and learning-based, using two metrics. For each metric we report mean/median values over all of the objects reconstructed, and across multiple levels of noise. **Green** and **Red** indicate the best and second best method respectively.

Teapot). We select the rest of the meshes from the test set of ShapeNet dataset [6]. We show all the objects in the Supplementary. However, because the ShapeNet meshes are not always watertight, we pre-process them with the algorithm proposed by Stutz and Geiger [32]. Given the watertight meshes we randomly decimate the number of vertices by different factors, obtaining three different sparsity levels. For each sparsity level we also add an increasingly large amount of Gaussian noise. We describe the parameters we use, and offer visualization of the different levels of noise in the Supplementary.

Numerical evaluation. For our numerical evaluation we use the symmetric Hausdorff distance, which reports the largest vertex reconstruction error for each mesh, and the Chamfer- ℓ_1 distance, which computes the distance between two meshes after assigning correspondences based on closest vertices. Table 1 shows that our method performs consistently better than all the competitors. The gap is most apparent when comparing with deep-learning methods that learn priors at the object level, further suggesting that our strategy to learn local priors is a promising direction. We list the numbers for each object across the different noise settings in the Supplementary.

Qualitative evaluation. In Figure 7 we show both the meshlets at the end of our optimization (a) and the quality of the final mesh reconstructed by our algorithm (b). Note that, thanks to the re-meshing steps during our optimization procedure (Section 4.1), our output is a high-quality, regular mesh. We also show a subset of the objects used in the numerical evaluation in Figures 1 and 6 for different levels of sparsity and noise. Additional results are in the Supplementary. Competing methods are significantly impacted by noise and produce overly smooth results to attenuate its effect. For example, the Laplacian reconstructions obtained with low regularization (Lap-low) are still

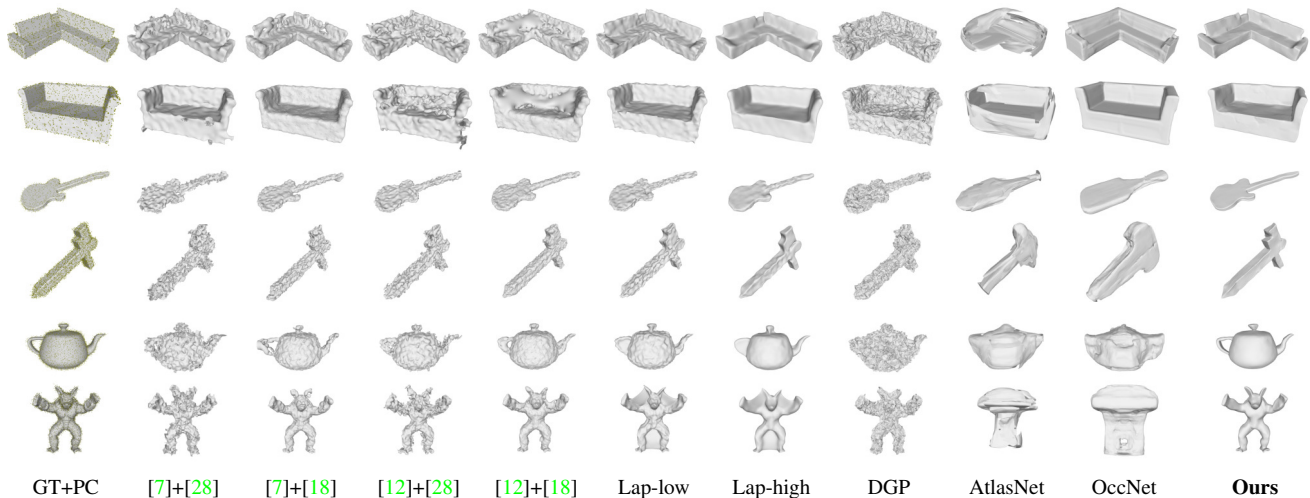


Figure 6: Qualitative comparison of several reconstruction methods and our approach. On the left is the input, the ground-truth (GT) mesh overlaid with the sparse, noisy point cloud (PC). We show results with normals estimated with Meshlab [7] and PCPNet [12]. We reconstruct the resulting point clouds with both RILMS [28] and Screened Poisson [18]. Laplacian regularizer [27] is shown for two levels of smoothing. We also show three recent deep learning approaches: Deep Geometric Prior (DGP) [35], AtlasNet [11] and OccNet [26]. All of these methods struggle to cope with noise, classes not seen in training, or both.

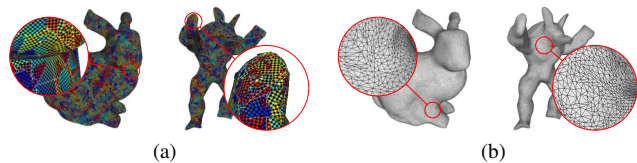


Figure 7: Our final meshlets are globally consistent and capture the local shape of the mesh (a). The resulting mesh is regular over the whole reconstructed object (b).

noisy, while those for which we used high regularization (Lap-high) are over-smoothed. Even Screened Poisson reconstruction [18], the de facto standard among traditional methods, used in conjunction with PCPNet [12] to estimate the normals, produces visibly noisy results. Finally, as also shown in Figure 1, state-of-the-art deep learning methods only work on objects seen in training and for low levels of noise. Our results, on the other hand, offer the best trade-off between detail and noise by recovering locally sharp features and small details, despite the sparsity and noise of the point clouds.

On the importance of natural meshlets. Our meshlet priors are the core of our method. Here we compare our natural meshlets with other shape priors to isolate their contribution to the overall quality of the result. The first is a Laplacian regularizer, which is a standard smoothness prior [27]. The second, is the recent work by Williams *et al.*, which suggests that a neural network is also, in itself, a prior for local geometry [35]. We use these priors and

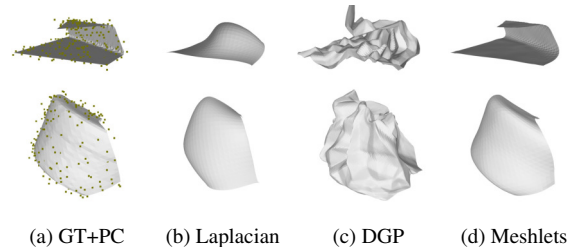


Figure 8: Meshlets reconstruct local features more accurately than other priors. Input point clouds shown on the GT mesh in (a).

our natural meshlet prior to optimize small patches of mesh to small point clouds extracted from real objects. Figure 8 shows two representative examples. Despite the complexity of the local shape, and the level of noise, the optimization that uses our strategy (Section 3.1) is able to correctly estimate the underlying meshlet. On the contrary, Deep Geometric Prior over-fits to the noise, and the Laplacian regularizer over-smooths the surface. On 10^5 meshlets, the average symmetric Hausdorff distance is 0.027 for DGP, 0.029 for Laplacian, and 0.024 for our method.

To further validate the importance of our natural meshlet priors towards our overall optimization procedure, we swap our latent space search with a Laplacian prior, and leave the rest of the algorithm untouched. As shown in Figure 9, using meshlet priors allows to better smooth noise, while still reconstructing sharp features, such as edges and corners. However, we also note that enforcing global consistency (GC) is an important step of our method, without which the quality of the final reconstruction degrades sig-

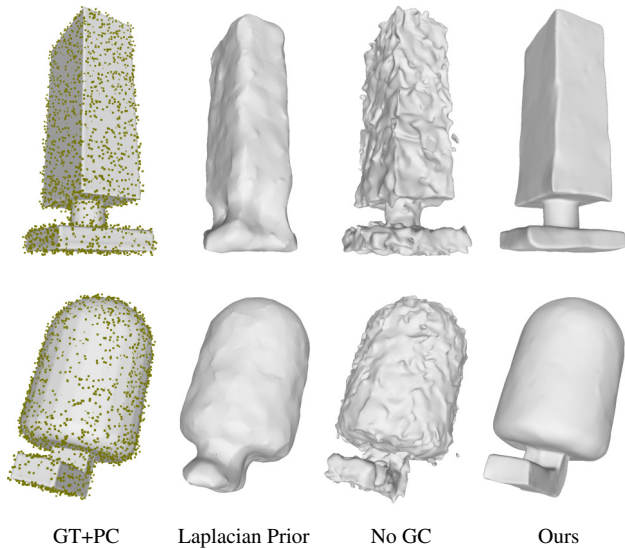


Figure 9: Replacing our meshlet priors with Laplacian priors leads to overly-smoothed results. However, using our meshlet priors without using global consistency (GC) produces irregular meshes.

nificantly, as shown in Figure 9.

6. Discussion and Limitations

Our approach optimizes a mesh and a number of meshlets based on the gradients available at mesh vertexes, while enforcing the meshlets priors. In this paper, the gradients for the mesh were obtained by computing the distance of the mesh to the point cloud. However, our method can take gradients from any source, including a differentiable renderer [16]. This adds to the flexibility of our approach.

In our work we learn and enforce priors for mesh estimation using meshlets, which have an intrinsic scale and resolution. Our current approach uses a single fixed scale of the meshlet for all the object reconstructions, although we effectively learn meshlets at multiple scales (see Section 4.2). This poses limitations on the level of details we can reconstruct: they cannot be smaller than the resolution of the meshlet. When this happens, the fine details may not be reconstructed, as shown in Figure 10(a). Using a meshlet at a single scale throughout the mesh deformation process may also lead to local minima, a particularly pressing problem if the initial mesh is significantly far from the target. Figure 10(b) shows a reconstruction that fell in a local minimum. A natural extension, then, would be to use a coarse-to-fine approach. Finally, our approach may fail for objects that present very thin structures, such as those shown in Figure 10(c).

To ensure that our auxiliary mesh stays regular throughout the optimization process, we perform Poisson reconstruction (see Section 4.1). As a by-product, we inherit its ability to deal with different topologies as can be seen for

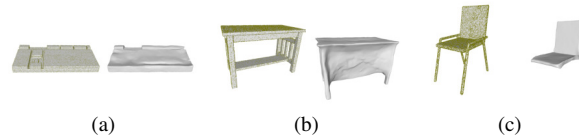


Figure 10: Our reconstruction method may yield poor reconstructions if when the object features are smaller than the resolution of the meshlets (a), when the optimization falls in a local minimum (b), or for thin structures (c).

the Stanford Teacup in Figure 6, which is genus-one.

Our current approach is computationally expensive and not optimized for speed. Hence, it can take from hours to dozens of hours, depending on the initialization to run the full optimization. The Chamfer distance computation in Equations 2 and 3, which has quadratic complexity in the number of points, is the bottleneck. In the current implementation, we consider all the points in the point cloud every time we compute the distance. Limiting the search to a local neighborhood would help. Improving the efficiency of the meshlets extraction is another obvious venue to speed up the algorithm.

7. Conclusions

We present meshlets, a novel local shape representation that allows to reconstruct 3D meshes from sparse, noisy point clouds. To do this, we train a variational autoencoder to learn the manifold of naturally-occurring meshlets. Meshlets on this manifold act as priors for local, class-agnostic, natural features. Therefore, meshlets allow us to disentangle the overall pose of the object from its shape, and can be used to reconstruct objects from classes not seen in training. To reconstruct a full mesh we use a number of meshlets. Then, we propose an alternating optimization procedure that first optimizes the meshlets to match the points (locally) and then enforces their consistency (globally). Our algorithm reconstructs objects from classes not seen in training, in arbitrary pose, and under significant noise and sparsity of the input points, even when existing state-of-the-art methods fail.

Acknowledgments

We would like to thank Kihwan Kim, Aleandro Troccoli, and Ben Eckart for the discussions on comparisons and evaluation. Arash Vahdat for the discussions and feedback on latent space regularization. UCSB acknowledges partial support from NSF grant IIS 16-19376 and an NVIDIA fellowship for A. Badki.

References

- [1] T. Bagautdinov, C. Wu, J. Saragih, P. Fua, and Y. Sheikh. Modeling facial geometry using compositional VAEs. In

- Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 4
- [2] Chandrajit L Bajaj, Fausto Bernardini, and Guoliang Xu. Automatic reconstruction of surfaces and scalar fields from 3D scans. In *Proceedings of SIGGRAPH*, 1995. 2
- [3] Heli Ben-Hamu, Haggai Maron, Itay Kezurer, Gal Avineri, and Yaron Lipman. Multi-chart generative surface modeling. *ACM Transactions on Graphics*, 2018. 3
- [4] Fausto Bernardini, Joshua Mittleman, Holly Rushmeier, Cláudio Silva, and Gabriel Taubin. The ball-pivoting algorithm for surface reconstruction. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 1999. 2
- [5] Michael Bloesch, Jan Czarnowski, Ronald Clark, Stefan Leutenegger, and Andrew J. Davison. CodeSLAM — learning a compact, optimisable representation for dense visual SLAM. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2, 3, 4
- [6] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. ShapeNet: An information-rich 3D model repository. Technical Report arXiv:1512.03012, 2015. 3, 5, 6
- [7] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. MeshLab: An open-source mesh processing tool. In *Eurographics Italian Chapter Conference*, 2008. 6, 7
- [8] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *ACM Transactions on Graphics (SIGGRAPH)*, 1996. 2
- [9] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. In *Advances in Neural Information Processing Systems (NIPS)*, 2014. 2
- [10] Clément Godard, Oisín Mac Aodha, and Gabriel J. Brostow. Unsupervised monocular depth estimation with left-right consistency. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2
- [11] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan C. Russell, and Mathieu Aubry. A papier-mâché approach to learning 3D surface generation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1, 2, 3, 6, 7
- [12] Paul Guerrero, Yanir Kleiman, Maks Ovsjanikov, and Niloy J Mitra. PCPNet: Learning local shape properties from raw point clouds. In *Computer Graphics Forum*, 2018. 2, 6, 7
- [13] Hugues Hoppe, Tony DeRose, Tom Duchamp, John McDonald, and Werner Stuetzle. Surface reconstruction from unorganized points. In *Proceedings of SIGGRAPH*. 1992. 2
- [14] Po-Han Huang, Kevin Matzen, Johannes Kopf, Narendra Ahuja, and Jia-Bin Huang. DeepMVS: Learning multi-view stereopsis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1, 2
- [15] Angjoo Kanazawa, Shubham Tulsiani, Alexei A. Efros, and Jitendra Malik. Learning category-specific mesh reconstruction from image collections. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. 1, 2
- [16] Hiroharu Kato, Yoshitaka Ushiku, and Tatsuya Harada. Neural 3D mesh renderer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1, 2, 8
- [17] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Eurographics Symposium on Geometry Processing*, 2006. 2
- [18] Michael Kazhdan and Hugues Hoppe. Screened Poisson surface reconstruction. *ACM Transactions on Graphics*, 2013. 2, 5, 6, 7
- [19] Alex Kendall, Hayk Martirosyan, Saumitro Dasgupta, Peter Henry, Ryan Kennedy, Abraham Bachrach, and Adam Bry. End-to-end learning of geometry and context for deep stereo regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2
- [20] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 2
- [21] Katrin Lasinger, René Ranftl, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *arXiv preprint arXiv:1907.01341*, 2019. 2
- [22] Chen-Hsuan Lin, Oliver Wang, Bryan C Russell, Eli Shechtman, Vladimir G Kim, Matthew Fisher, and Simon Lucey. Photometric mesh optimization for video-aligned 3D object reconstruction, 2019. 2
- [23] O. Litany, A. Bronstein, M. Bronstein, and A. Makadia. Deformable shape completion with graph convolutional autoencoders. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2, 4
- [24] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3D surface construction algorithm. In *Proceedings of SIGGRAPH*, 1987. 2
- [25] Eivind Lyche Melvæ and Martin Reimers. Geodesic polar coordinates on polygonal meshes. In *Computer Graphics Forum*, 2012. 3, 6
- [26] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3D reconstruction in function space. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 2, 3, 6, 7
- [27] Andrew Nealen, Takeo Igarashi, Olga Sorkine, and Marc Alexa. Laplacian mesh optimization. In *ACM International Conference on Computer Graphics and Interactive Techniques (GRAPHITE)*, 2006. 1, 2, 6, 7
- [28] A. Cengiz Öztireli, Gael Guennebaud, and Markus Gross. Feature preserving point set surfaces based on non-linear kernel regression. In *Computer Graphics Forum*, 2009. 6, 7
- [29] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 1, 2, 4
- [30] Daniel Scharstein and Richard Szeliski. High-accuracy stereo depth maps using structured light. In *Proceedings*

of the *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003. 1

- [31] Ayan Sinha, Asim Unmesh, Qixing Huang, and Karthik Ramani. SurfNet: Generating 3D shape surfaces using deep residual networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 3
- [32] David Stutz and Andreas Geiger. Learning 3D shape completion under weak supervision. *CoRR*, 2018. 6
- [33] Maxim Tatarchenko, Stephan R Richter, René Ranftl, Zhuwen Li, Vladlen Koltun, and Thomas Brox. What do single-view 3D reconstruction networks learn? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2
- [34] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3D mesh models from single RGB images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. 1, 2
- [35] Francis Williams, Teseo Schneider, Claudio Silva, Denis Zorin, Joan Bruna, and Daniele Panozzo. Deep geometric prior for surface reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 2, 6, 7
- [36] Yao Yao, Zixin Luo, Shiwei Li, Tian Fang, and Long Quan. MVSNet: Depth inference for unstructured multi-view stereo. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. 1, 2
- [37] Lequan Yu, Xianzhi Li, Chi-Wing Fu, Daniel Cohen-Or, and Pheng-Ann Heng. EC-Net: An edge-aware point set consolidation network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. 2
- [38] Lequan Yu, Xianzhi Li, Chi-Wing Fu, Daniel Cohen-Or, and Pheng-Ann Heng. PU-Net: Point cloud upsampling network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2
- [39] Tinghui Zhou, Matthew Brown, Noah Snavely, and David G. Lowe. Unsupervised learning of depth and ego-motion from video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2
- [40] Rui Zhu, Chaoyang Wang, Chen-Hsuan Lin, Ziyang Wang, and Simon Lucey. Object-centric photometric bundle adjustment with deep shape prior. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018. 2, 3