

VoroMesh: Learning Watertight Surface Meshes with Voronoi Diagrams

Nissim Maruani

Inria, Université Côte d’Azur
nissim.maruani@inria.fr

Roman Klokov

LIX, École Polytechnique, IP Paris
klokov@lix.polytechnique.fr

Maks Ovsjanikov

LIX, École Polytechnique, IP Paris
maks@lix.polytechnique.fr

Pierre Alliez

Inria, Université Côte d’Azur
pierre.alliez@inria.fr

Mathieu Desbrun

Inria Saclay - Ecole Polytechnique
mathieu.desbrun@inria.fr

Abstract

In stark contrast to the case of images, finding a concise, learnable discrete representation of 3D surfaces remains a challenge. In particular, while polygon meshes are arguably the most common surface representation used in geometry processing, their irregular and combinatorial structure often make them unsuitable for learning-based applications. In this work, we present VoroMesh, a novel and differentiable Voronoi-based representation of watertight 3D shape surfaces. From a set of 3D points (called generators) and their associated occupancy, we define our boundary representation through the Voronoi diagram of the generators as the subset of Voronoi faces whose two associated (equidistant) generators are of opposite occupancy: the resulting polygon mesh forms a watertight approximation of the target shape’s boundary. To learn the position of the generators, we propose a novel loss function, dubbed VoroLoss, that minimizes the distance from ground truth surface samples to the closest faces of the Voronoi diagram which does not require an explicit construction of the entire Voronoi diagram. A direct optimization of the VoroLoss to obtain generators on the Thingi32 dataset demonstrates the geometric efficiency of our representation compared to axiomatic meshing algorithms and recent learning-based mesh representations. We further use VoroMesh in a learning-based mesh prediction task from input SDF grids on the ABC dataset, and show comparable performance to state-of-the-art methods while guaranteeing closed output surfaces free of self-intersections.

1. Introduction

Geometry processing of three-dimensional shapes typically relies on a discrete representation of its surface boundary. Among various boundary representations, point sets and surface polygon meshes are popular because they are simple to generate, edit, or render. While most tasks can be

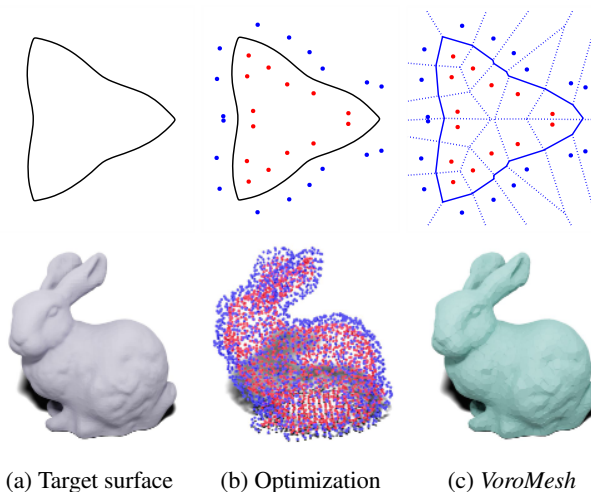


Figure 1: **VoroMesh at a glance:** To fit our representation to (a) a target shape (top: curve; bottom: surface), (b) the positions of generators around the shape boundary are optimized based on our *VoroLoss*, before a binary occupancy is assigned to each generator; (c) the *VoroMesh* is then extracted as a subset of the Voronoi diagram of the generators containing only the Voronoi faces between pairs of generators of opposite occupancy.

efficiently performed with non-learning tools from the standard geometry processing toolbox, some of them (e.g., ill-posed problems such as reconstruction from sparse data, or data-driven tasks) require learning tools. Using neural networks with a differentiable (explicit or implicit) boundary representation can be a powerful way to tackle these problems. Yet, discrete representations such as surface polygon meshes rely on two parts: vertex locations, and vertex connectivity defining the polygonal facets. The first part, continuous in nature, is easy to learn, but the second part, now combinatorial, is substantially more difficult, which has hampered mesh-based learning methods in the past. Indeed, besides the complexity issues arising from modeling vertex connectivity, the discrete nature of topological con-

figurations in meshes poses two accompanying differentiability issues that are particularly relevant in learning applications. First, the way connectivity is induced from corresponding discrete variables generally is not differentiable; even if probabilistic modeling can offer differentiable sampling from discrete variables [17], the combinatorial nature of connectivity forces faces and edges to be considered independently, which often produce polygon soup predictions instead of watertight meshes. Second, vertex position optimization can also be hindered since non-stable mesh prediction during training results in discontinuity in most of the commonly-used loss functions based on sampling from predicted mesh surfaces, such as chamfer distance and point-to-surface distance: any change of topology in the prediction over the course of training can result in discontinuous changes in the coordinates of on-surface samples.

State-of-the-art data-driven methods for mesh prediction can be roughly categorized into two groups: methods producing shapes with neural implicit representations and meshing their zero-level sets [16, 31, 33, 38], and approaches predicting explicit meshes as sets of vertices and faces [6, 9, 13, 32, 40, 47]. While proven to be powerful at capturing surfaces, implicit representations have their drawbacks, such as the need for additional test-time surface discretization to produce meshes (which can be costly for high resolutions), and limited control over the geometry of the produced meshes due to the inability to explicitly define optimization objectives on the output surfaces. While recent works in the second category show promising results in explicit mesh prediction, most of them rely on the use of regular discretization grids, which limit their expressivity and do not guarantee watertightness of the output meshes.

In order to circumvent the issues mentioned thus far, we propose to use a differentiable surface representation based on the *3D Voronoi diagram* [4] of an input point set of generators which canonically derives its combinatorial information from their location. The continuous placement of generators results in an adaptive discretization of 3D surfaces which substantially improves the expressivity of our representation compared to methods relying on regular grids. We leverage the geometric properties of the Voronoi diagram (specifically, the fact that Voronoi cells are defined as intersections of halfspaces) to formulate a novel loss based only on bisectors between generators rather than the full Voronoi diagram: this allows for a differentiable optimization of the 3D Voronoi generators. Since our loss function only requires the position of generators for faithful surface approximation, our Voronoi representation successfully mitigates the differentiability issues associated with explicit connectivity modeling. Compared to previous Voronoi-based representations [48], we do not focus on volume decomposition but on surface approximation, so our loss does not need to be regularized via the addition of the

centroidal Voronoi tessellation energy to favor the spreading of generators.

In summary, the main contributions of this work are:

- We establish a novel loss function, the *VoroLoss*, which directly optimizes 3D Voronoi generators to fit an input target shape (provided as a surface triangle mesh) without requiring an explicit computation of the Voronoi diagram;
- We show that the representation power of our *VoroMesh* outperforms Marching Cubes, Dual Contouring, and two recent learning approaches in terms of geometric fidelity;
- We demonstrate that the proposed *VoroMesh* and *VoroLoss* can be integrated into a learning pipeline to predict 3D Voronoi generators, along with their occupancy, to reconstruct closed and non-self-intersecting meshes from input regular grids of SDF values.

2. Related Work

Numerous prior works have sought 3D shape representations amenable to learning. We review the most relevant approaches while pointing the reader to a recent thorough review on intelligent mesh generation [24] — which, notably, identifies the generation of watertight manifold meshes as one of the most significant open challenges.

Implicit representations. To bypass the issues of shape discretization, pioneering 3D learning methods proposed to use implicit surface representations defining shape boundaries as iso-surfaces of neural distance fields [12, 14, 33] or as decision boundaries of neural occupancy classifiers [8, 29, 30, 50]. Most of the earlier methods, however, rely on global features, trying to encode complete surface information into a single global latent vector. This constraint limits the ability of these models to encode sharp local features and prevents efficient generalization of models across different categories and data domains. Additionally, most of these works resort to Marching Cubes [26] for the final mesh extraction, needing a cubic number of forward passes of the occupancy/distance prediction network. As a consequence, sharp features partially captured by the implicit function are further smoothed out at low resolutions. Single shape optimization [15, 41] is yet another type of implicit representation, for which a shape is encoded by fitting a whole neural network; but learning from previous data is not possible, preventing many data-driven applications. Although our work focuses on the prediction of explicit discrete surface representation, it could potentially be combined similarly to related mesh prediction approaches discussed below, since inferred SDF values could be used as inputs of our model.

Discrete representations. Different shape representations have been proposed to capture geometric information for learning-based applications. Early 3D shape modeling

methods used regular [22, 49] and adaptive [44] grids of binary occupancy, or their decomposition [39] predicted by 3D CNNs to represent reconstructed geometry. A separate line of works proposed to predict shapes as sets of points inferred with MLPs [2, 42], invertible normalizing flows [21, 51], and diffusion-based generative models [28, 52]. Our representation relates to point clouds in two aspects: we predict continuous point coordinates to represent shapes and use distance-based loss functions to provide the training signal; but our representation does not restrict output points to lie on the surface, using them instead as generators of a 3D Voronoi diagram fitting the surface.

Occupancy grids and point clouds can be viewed as approximations of target surfaces suitable for prediction with neural networks and continuous optimization, but containing only partial mesh information. Despite the challenges stated in Section 1, numerous works have proposed models capable of producing meshes. Deformation-based approaches [43] use template meshes and predict vertex offsets to fit target surfaces without any changes in the initial connectivity. The approach of Nash *et al.* [32] treats vertices and connectivity as random variables and uses an autoregressive probabilistic model to derive likelihood-based objectives, suitable for optimization. Several works also proposed to model surfaces with predicted parametric primitives [7, 11], allowing differentiation by continuous relaxation of discrete variables. However, representations from this family do not generalize well to new data.

Most relevant to our work are methods that lift the limitation of global shape encoding and use local features obtained from coarse, noisy, or partial inputs. Deformable volumetric tetrahedral meshes [13, 40] for instance can synthesize shapes from SDF grids. While achieving good performance, these approaches mainly compare themselves to global models, do not provide topological guarantees on the output shapes, and suffer from artifacts due to their use of regular grids. Deep Marching Cubes [25] propose differentiable point-to-mesh distance and curvature losses as expectations over the space of topological configurations to train occupancy and vertex displacements on a grid. Since occupancy is modeled independently, this approach is unable to provide guarantees on topological correctness. A similar approach introduces a new representation inspired by Marching Cubes to capture topological and vertex information [9] with extended sets of discrete and continuous variables for every vertex of a regular grid. Two separate networks predict these variables independently for each grid vertex, so they form a consistent output mesh. Follow-up work [6] substitutes these variables with a reliance on the mesh produced by Dual Contouring [18] and extends the approach to various types of input data. Both of these approaches circumvent the necessity to differentiate through meshing procedures

by the construction of ground truth proxy data structures during pre-processing and training networks to infer them from input data. However, resulting meshes are not always manifold for NDC [6], and both methods often produce self-intersections. Shape-as-points [35] predicts coordinate offsets and normals for noisy point cloud inputs which are used to produce indicator grids with a differentiable version of PSR [19]. These indicator grids are supervised with precomputed grids obtained from ground truth surface samples. Compared to these works, 1) we do not construct output meshes during training, thus avoiding the pitfalls associated with differentiation through explicit meshing; 2) supervision relies on sampled points and does not require any pre-computed data structures; 3) our approach is not limited in expressivity by the use of a fixed grid, since it allows adaptive placement of generators.

Shape representation via computational geometry. A few works have already exploited Voronoi diagrams to represent 3D shapes [1, 3]. However, providing a differentiable loss to compute such a diagram is not straightforward. VoronoiNet [48], for instance, uses a soft definition of the Voronoi diagram to match the occupancy of a given array in 2D applications. However, their formulation offers a differentiable implicit representation of *solid objects*, in the sense that they spread generators across the whole volume (through a CVT-based regularizer) to offer a convex decomposition of the inside and outside of the 2D object, thus extending CvxNet [11]. (A single 3D example, a sphere, was exhibited.) Instead, we focus on *approximating the boundary of a 3D object*: as a consequence, far fewer generators are required and they remain close to the boundary as their role is purely to approximate the surface well, capturing the normal field and sharp features of the surface. Our introduction of a dedicated loss also removes the need to use a soft-argmin version of Voronoi diagrams. Finally, other related works have proposed to use the dual of Voronoi diagrams, i.e., Delaunay triangulations, to offer a differentiable approach to surface triangulations [36, 37]; however, they do not offer topological guarantees due typically to the presence of “non-manifold” triangles in resulting meshes.

3. Voronoi Mesh Representation

We now describe the VoroMesh approach by first explaining our Voronoi-based representation via generators, then the *VoroLoss* which minimizes the distance from an input point cloud sampling a surface to Voronoi facets of the generators, then occupancy determination and mesh extraction. Finally, we discuss the optimization of this representation for a given input shape and its use in learning tasks.

3.1. VoroMesh Overview

Our VoroMesh representation canonically defines a watertight mesh via a Voronoi diagram. Given a set of N dis-

tinct 3D points that we call generators $\mathbf{Q} = \{q_i \in \mathbb{R}^3\}$, the *Voronoi cell* V_i of a generator q_i is the set of points $y \in \mathbb{R}^3$ whose distance to q_i is smaller than their distance to any other site q_j . Each V_i is known to be a convex polytope, whose boundary is composed of flat polygonal *Voronoi faces* lying on bisectors between q_i and some of its closest neighboring generators q_j . The union of these disjoint 3D regions forms a convex decomposition of \mathbb{R}^3 [4]. We assign a binary occupancy $\mathbf{O} = \{o_i\} \in \{0, 1\}^N$ to each generator, such that the union of the cells with non-zero occupancy yields a 3D volume composed of convex polyhedral cells. The boundary of this 3D volume is then a Voronoi-induced surface mesh, that we call *VoroMesh*(\mathbf{Q}, \mathbf{O}). Assuming that all infinite cells are tagged as outside, we obtain a *watertight surface mesh* (i.e., without holes or self-interpenetrations, and with a clearly defined inside). Our approach then consists in optimizing the generators so that induced Voronoi faces best fit the 3D ground truth shape.

VoroLoss loss function. For optimization and learning purposes, we need to be able to minimize the distance between a VoroMesh and a dense sampling X of a target shape. Computing the entire Voronoi diagram induced by the generators at each step of the optimization would be too costly. Instead, we leverage the geometric properties of Voronoi diagrams (and in particular, their locality) to find a loss function that can directly provide the correct distance without invoking a full Voronoi diagram construction. More specifically, we optimize the generators \mathbf{Q} not by minimizing the distance from X to the current Voronoi faces, but by minimizing the distance from X to *all* the faces of the full diagram, i.e., $\sum_{x \in X} \min_i \|x - \partial V_i\|^2$ — which can efficiently be achieved by only considering all close pairs of generators. This property stems from the following theorem (see supplementary material for a proof):

Theorem 1 *The distance from x to its closest face in a Voronoi diagram equals the distance from x to the closest bisector $H_{i_x, j}$ formed between q_{i_x} whose Voronoi cell contains x and another Voronoi site q_j :*

$$\|x - \partial V_i\| = \min_{j \neq i_x} \|x - H_{i_x, j}\|.$$

We thus introduce a loss function, dubbed *VoroLoss*:

$$\text{VoroLoss}(X, \mathbf{Q}) := \sum_{x \in X} \min_{j \neq i_x} \|x - H_{i_x, j}\|^2$$

The simplification implied by evaluating distances to planes instead of distances to the polygonal faces of a computed Voronoi diagram allows for a simple and efficient implementation. In practice, we use a k -nearest neighbor algorithm to select the closest cells i_x and the closest bisector planes $H_{i_x, j}$ to cull a large number of computations. For N generators and a sampling X of size $|X| = D$, the average complexity for computing this loss becomes $\mathcal{O}(ND)$.

Note finally that the *VoroLoss* is computed without any occupancy information: it simply tries to make each sample be on a face of the Voronoi diagram of the generators.

Occupancy and watertight mesh extraction. Once the *VoroLoss* is minimized, we must determine the occupancy of each generator (see colors of the generators in Figure 1). This is quickly achieved by simply using the ground truth occupancy of the barycenter of each Voronoi cell in practice. Once the occupancy of all generators is established, the polygon mesh can be extracted by keeping the Voronoi faces separating generator pairs of complementary occupancy. By construction, this mesh represents the boundary of a volume: it thus has a separate interior and exterior, and it is self-intersection free. In theory, cospherical (resp. cocircular) points of different occupancies may create non-manifold vertices (resp. edges). To fix these topological exceptions one can simply duplicate the corresponding vertices (resp. edges). Note however that in our tests, such a case never happened in practice.

Shape optimization. Based on the building blocks we described thus far, we can now fit a *VoroMesh*(\mathbf{Q}, \mathbf{O}) to a target surface as follows (see Figure 1):

1. Densely sample a set of points $X \in \mathbb{R}^{M \times 3}$ from a ground truth surface mesh (V, F) ;
2. Initialize $\mathbf{Q} \in \mathbb{R}^{N \times 3}$;
3. Minimize *VoroLoss*(X, \mathbf{Q});
4. Compute the Voronoi diagram of \mathbf{Q} ;
5. Determine the ground truth occupancy \mathbf{O} of the barycenter of each Voronoi cell;
6. Compute the final polygonal mesh *VoroMesh*(\mathbf{Q}, \mathbf{O}).

For Step 2, we initialize the generators \mathbf{Q} by selecting, out of a regular voxel grid, every grid node adjacent to one or more voxels containing samples from X .

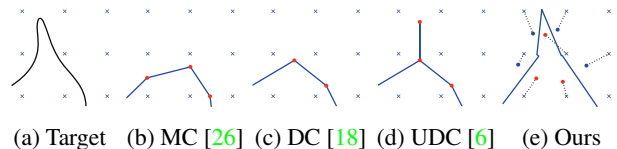


Figure 2: Marching Cubes (b) and Dual Contouring (c) cannot capture details of a target shape (a) smaller than the grid size; UDC (d), based on edge-crossings, can but at the price of a non-manifold elements. *VoroMesh* (e) both captures the details and returns a closed and manifold mesh

3.2. Learning VoroMeshes

To demonstrate the effectiveness of our representation in data-driven settings, we focus on one application: shape reconstruction from a sparse grid sampling of the SDF. Unlike other 3D learning tasks (such as single view reconstruction),

there is no need to encode complete shape information into a single global latent variable, so we can make use of specific features representing local configurations of particular parts of the volume rather than a single global representation: this will allow the network to generalize to new shapes that are strikingly different from the training set.

To generate a *VoroMesh* from a grid $I \in \mathbb{R}^{N \times N \times N}$ of signed distances, we predict the positions of the generators along with their occupancy (see Figure 3). Knowing that most of the information describing the surface is concentrated in the SDF values closest to the surface, we mask the SDF grid through thresholding around the isovalue and convert it into a sparse grid. Selected grid points close to the surface v_i serve as the initial positions of the generators. The sparse grid of SDF values is fed to a generic ResNet-like Sparse 3D CNN (SCNN) [10] with a receptive field of size 19^3 to produce local features. Using SCNNs not only alleviates the computational complexity burden of 3D convolutions, but also forces the CNN filters to focus on near-surface variations in SDF values and to spend the parameter budget efficiently, capturing information that is most relevant for surface reconstruction. We obtain the final position q_i and occupancy o_i of generators with two separate MLPs acting on local features produced by the SCNN, via:

$$\begin{aligned} q_i &= v_i + \text{MLP}_1(\text{SCNN}(\text{SDF}(v_i))), \\ o_i &= \text{MLP}_2(\text{SCNN}(\text{SDF}(v_i))), \end{aligned} \quad (1)$$

where $\text{SDF}(v_i)$ is SDF values at the corresponding point v_i .

We optimize the parameters of our model in two steps. First, we train the SCNN and MLP_1 to predict the position of the generators Q using our *VoroLoss*. Previous offset prediction based networks often introduced hyperbolic tangent non-linearity at the end of the network to bound the values of possible offsets. We observed that at the start of optimization, this may lead to offsets quickly reaching these bounds and never returning to lower values because of the vanishing gradients of the hyperbolic tangent function. Since we do not necessarily want to force generators to stay in their initial cells and wish to obtain more informative gradients during training, we refrained from using hyperbolic tangent and instead, introduced a mild regularization on the maximum values of predicted offsets with:

$$R(\mathbf{Q}) = \max_i \|\Delta v_i\|_2 \quad (2)$$

which we add to *VoroLoss* to form a loss $L(X, \mathbf{Q})$ as:

$$L(X, \mathbf{Q}) = \text{VoroLoss}(X, \mathbf{Q}) + \lambda R(\mathbf{Q}), \quad (3)$$

with λ being a regularization weight. Unlike NDC [6], we do not pre-compute the position of generators, but rely on our loss to move the generators in place without any mesh supervision. By not requiring mesh construction and point sampling to fit surfaces at all, our approach completely evades the difficulties typically arising from the differentiation of these operations.

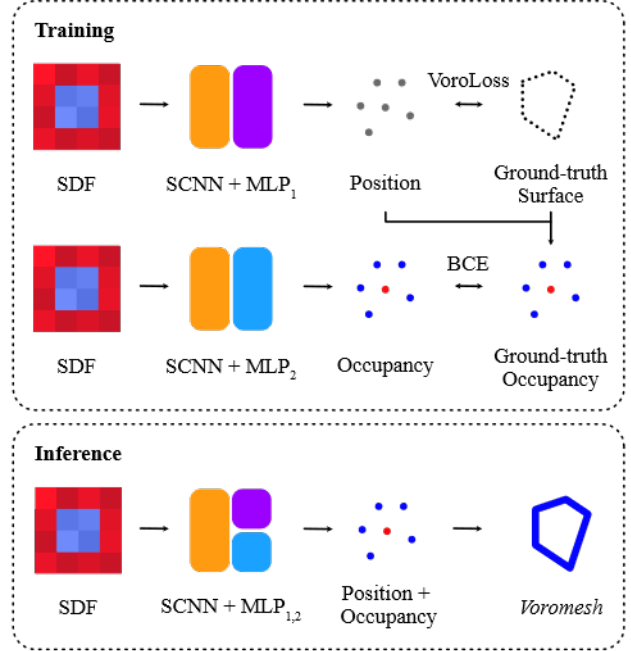


Figure 3: We train SCNN and first MLP to predict the position of generators, supervised by the *VoroLoss* with respect to a sampling of the input shape (top row). We then compute the ground truth occupancy of the predicted generators, and use this information to train a second MLP (middle row). During inference, the whole network predicts the position and occupancy of each generator from which the *VoroMesh* can be extracted (bottom row).

The second training step freezes the convolutional part and optimizes the parameters of MLP_2 with minimization of the binary cross-entropy (BCE) between predicted and ground truth occupancy probabilities, extracted as discussed in the previous section. At inference time, a wrong occupancy prediction for a Voronoi cell could create an undesired surface artifact. Two fallback solutions based on queries to input SDF values can be used to fix incorrect occupancies: (1) increase the thickness of the SDF mask to include additional fixed Voronoi generators with deterministic occupancy, which will then bound the spatial scale of possible artifacts; (2) assign the occupancy of some cells using their vertices’ input SDF values and check that no voxel having a vertex with signed distance larger than the diagonal length of a voxel does not contain the target surface. Since predicted occupancy is only needed for near-surface cells, we found that these two simple actions remove all potential issues in practice.

A last, yet important technical remark is that compared to NMC and NDC, our implementation allows multi-shape batching and parallelized forward and backward computations during training and inference, since MLPs and SCNNs [10] can process several samples simultaneously.

4. Experimental Results

We implemented our VoronoiMesh approach in Python with PyTorch [34], SciPy [46] and CGAL [45]. All timings were computed on a Dell Precision desktop machine with 54 cores, a NVidia A6000 GPU, and 512GB of RAM. Our code can be found online at [our project page](#), along with most of our results.

4.1. Optimization-based 3D Reconstruction

We first compare the ability of different representations — including our *VoroMesh* — to capture ground truth surfaces by fitting them with direct optimization. For all methods, we assume that we have access to the ground truth meshes from Thingi32 dataset, which is a subset of Thingi10K [53], from which we can derive signed distance values and sample dense point clouds. We exclude two non-watertight models (96481 and 58168) to obtain a final dataset composed of 30 valid shapes. We evaluate our method using grid resolutions of sizes g_s^3 with $g_s \in [32, 64, 128]$. We sample the ground truth surface with $150 \times g_s^2$ points, and use grid vertices adjacent to sampled points as initial Voronoi generators. To boost the optimization, we randomly select 20% of the surface points X that we feed in the *VoroLoss* and use 400 optimization steps of the Adam optimizer [20] with a learning rate of 0.005, further divided by two at steps 80, 120, 200 and 250. We timed MC [26] and DC [18] (excluding the SDF extraction step), and measured the optimization time (which is systematically the longest step) for the rest of the methods (see supplementary material for timings).

Baselines. As no public implementation is available, we re-implemented the soft-Voronoi loss from VoronoiNet [48]. We simplified their pipeline by assigning the correct occupancy to each generator prior to optimization in order to focus only on the ability to fit the surface and ignore the two additional losses (*bounds loss* and *signed-distance loss*) designed to relocate the generators to their assigned regions (inside or outside). In addition, we do not utilize centroidal Voronoi loss for two reasons: (1) it requires computing the full Voronoi diagram at each iteration (which we avoid to keep computations low); (2) and because we start with a uniform grid, so the generators are already well distributed. We select only the generators close to the input mesh to be able to use the same number of samples as for our method.

We also compare to Shape-As-Points (SAP) [35] which, similarly to *VoroMesh*, relies on the optimization of point positions, using regular grids in the process. SAP performs an iterative refinement scheme with more iterations at each step, giving it a slim advantage over VoronoiNet [48] and our method, but increasing its computation time. We compare to two axiomatic methods as well, namely Marching Cubes (MC) [26] and Dual Contouring (DC) [18]. These

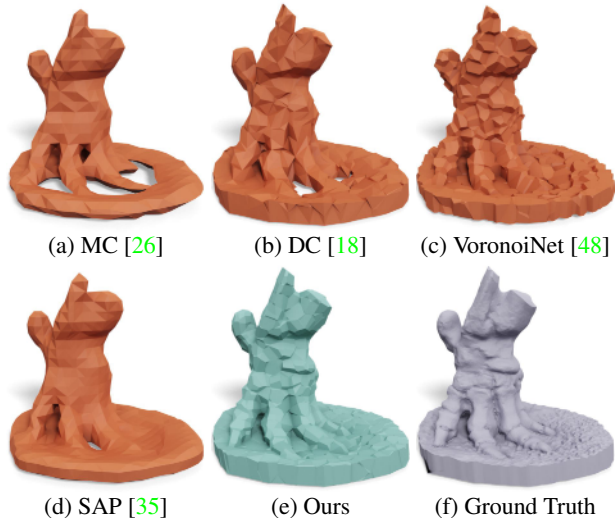


Figure 4: Visual comparison of optimization-based reconstructions. All methods utilize a grid of size 32^3 .

two methods require a sampling of the SDF as input instead of a point cloud. We evaluate their representation power given the same resolution of the input SDF grid that was used to initialize near-surface generators for *VoroMesh*. We also provide additional comparisons with DMTet [40] in the supplementary material, for completeness.

Results. Both MC [26] and DC [18] suffer from discretization artifacts, missing parts as shown in Figure 4 (see supplementary material for examples on more grid sizes), while MC does not reproduce sharp features well.

Originally designed for 2D images, VoronoiNet [48] provides a good fit of the overall volume, but a poor fit of the surface normals, which leads to poor visual appearance and NC distance, see Figure 4. Since all the generators are involved in the loss, it does not scale well with increasing grid resolution either: optimization of a grid larger than 32^3 does not fit into a 50GB GPU. Consequently, our approach is approximately ten times faster, see Table 1.

SAP [35] tends to generate shapes with fewer details and no sharp edges since its reconstruction is based on Marching Cubes. The output mesh, however, is guaranteed to be watertight. SAP requires an optimization of the whole volume, and is thus substantially slower than *VoroMesh*.

Our method performs the best across all considered metrics and resolutions given a similar amount of output information (our choice of initial generators around the surface guarantees that we use approximately the same number of cells as MC [26] and DC [18]). This confirms that our generators can effectively represent surfaces in 3D and that our *VoroLoss* allows for an efficient optimization of their positions to fit target surfaces. This optimized placement of generators (see Figure 2), including the ability to move them

Grid Size	CD ($\times 10^{-5}$)			F1 ($\delta = 0.003$)			NC			Time (s)		
	32 ³	64 ³	128 ³	32 ³	64 ³	128 ³	32 ³	64 ³	128 ³	32 ³	64 ³	128 ³
Marching Cubes [26]	10.653	1.196	0.686	0.585	0.852	0.932	0.900	0.951	0.977	0.001	0.005	0.03
Dual Contouring [18]	5.992	0.814	0.648	0.758	0.911	0.937	0.923	0.961	0.979	0.09	0.7	5.6
VoronoiNet [48]	3.252	-	-	0.543	-	-	0.886	-	-	11.6	-	-
Shape As Points [35]	6.543	1.906	0.669	0.589	0.857	0.934	0.894	0.949	0.978	51.2	106.3	191.3
VoroMesh	0.791	0.645	0.634	0.920	0.938	0.939	0.958	0.975	0.982	2.0	4.1	36.3

Table 1: Quantitative comparisons of Chamfer distance (CD), F1 score, and normal consistency (NC) of *VoroMesh* for an optimization-based 3D reconstruction task on the Thingi32 dataset for three different grid resolutions.

beyond their initial cells, means that *VoroMesh* can capture fine details, yielding faithful reconstructions even for low numbers of generators.

4.2. Robustness to Noise

In theory, a small displacement of the generators or of their occupancy could induce a large deformation of the *VoroMesh*: for instance, swapping the position of two generators while preserving their occupancy creates a 180-degree flip of the associated face. However, in practice, our representation is robust to small perturbations due to the locality of the Voronoi diagram construction. To demonstrate this property, we applied uniform perturbations to the generators with different magnitudes to one of our reconstructions, see Figure 5. This suggests that *VoroMesh* is suitable for learning-based applications.

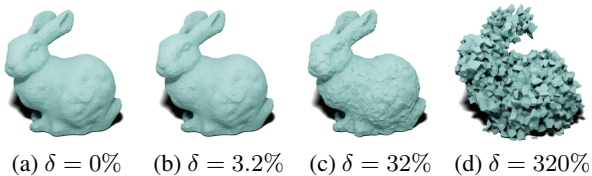


Figure 5: Robustness to noise. We evaluate the impact of noise on the reconstructed *VoroMesh*, where δ is the uniform noise magnitude given as percentage of voxel size.

4.3. Learning-based 3D Reconstruction

We now verify the suitability of *VoroMesh* and *VoroLoss* as target representation and training loss function in an inference-based 3D shape reconstruction from input SDF grids. All methods were trained on the ABC dataset [23], composed of surface triangle meshes of CAD models. We use the train/validation split from [6, 9], with 4,280 models in the training set and 1,071 models in the testing set. We remove approximately 10% of non-watertight models in both sets. We additionally verify the generalization capabilities of our method by reconstructing shapes from the Thingi32 dataset *without any fine-tuning*. For this experiment, we

Method	Grid Size	CD ($\times 10^{-5}$)	F1	NC	Watertight
NDC [6]	32 ³	66.08	0.787	0.941	31%
NMC [9]	32 ³	60.70	0.833	0.954	25%
VoroMesh32	32 ³	13.15	0.819	0.933	100%
VoroMesh32+64	32 ³	2.23	0.835	0.941	100%
NDC [6]	64 ³	2.209	0.882	0.975	20%
NMC [9]	64 ³	2.144	0.891	0.980	18%
VoroMesh64	64 ³	1.317	0.882	0.962	100%
VoroMesh32+64	64 ³	1.199	0.886	0.966	100%

Table 2: Comparison of learning-based 3D reconstruction methods on the ABC dataset.

consider two grid resolutions, 32³ and 64³, for which we trained three models using batches of 32, 16, and 24 shapes (one model for each resolution and one using both) and 10⁵ ground truth sample points per shape with the AdamW optimizer [27] and step-wise scheduling of momentum, learning rate, and regularization weight λ (see [project page](#) for code). We first train the SCNN and MLP predicting generators of *VoroMesh* for 200 epochs, then reuse the convolutional features and train a separate MLP predicting occupancy for each generator for 75 epochs.

Baselines. We compare our method against state-of-the-art NMC [9] and NDC [6] and use their pre-trained models, normalization, and evaluation code to recompute the results on the filtered datasets. Note that NDC proposes an unsigned, augmented version of their algorithm (UNDC); however, the underlying mesh representation does not correspond to the boundary of a volume, as it can collapse thin regions and predict zero thickness surfaces — see Figure 2 — so we do not compare to this variant. The Voronoi diagram could also be potentially leveraged to represent a triangle soup: the occupancy information would then lie on the (dual) Delaunay edges rather than Voronoi cells. Since we focus on generating topologically correct surfaces in this work, we leave this possibility for future work.

Results. *VoroMesh* outperforms both baselines in terms of Chamfer distance (CD), while being comparable in terms of normal consistency (NC) and F1 score, see Table 2. We believe that these results are due to NDC and NMC recon-

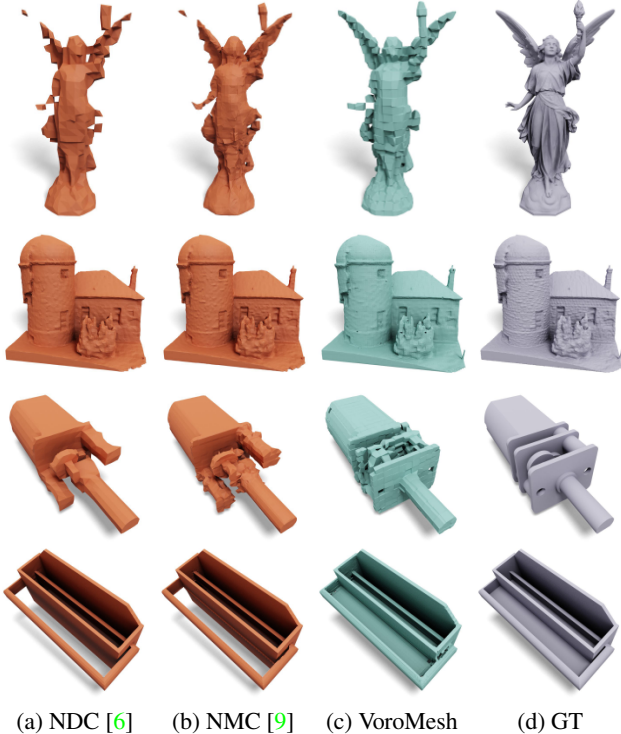


Figure 6: Visual comparison of learning-based reconstruction. Input grids of sizes 32^3 and 128^3 for Thingi32 (top rows), 32^3 and 64^3 for ABC (bottom rows). Our VoroMesh reconstruction, relying on a model trained on 32^3 SDF inputs, is able to reconstruct thin details.

structuring overall smoother surfaces, but failing to capture parts that *VoroMesh* can resolve with proper placement of the generators. NMC, which relies on a custom tessellation, uses more vertices, and thus has a slight advantage in terms of geometric fidelity, as explained by its authors [6]. Note that both NMC and NDC fail to produce *any mesh at all* for certain thin shapes from the test set. We also check for watertightness using CGAL [45] by confirming the absence of self-intersection and verifying that each edge is incident to exactly two faces: as expected, all VoroMeshes are watertight while only a fraction of NDC and NMC meshes are.

Generalization to Thingi32. To test the ability of our model to transfer knowledge across datasets and verify the stability of our *VoroMesh* representation, we reconstruct the same meshes from the Thingi32 dataset used in our direct optimization experiment, with models pretrained on the ABC dataset. As can be seen in Figure 6, our approach systematically produces valid detailed reconstructions of unseen shapes. On a coarse resolution, our approach manages to outperform baselines in terms of CD by a significant margin, meaning *VoroMesh* can capture more details with similar shape parameter budgets. On higher resolu-

Method	Grid Size	CD ($\times 10^{-5}$)	F1	NC	Watertight
NDC [6]	32^3	6.396	0.734	0.918	10%
NMC [9]	32^3	5.212	0.796	0.936	0%
VoroMesh32	32^3	2.146	0.738	0.898	100%
VoroMesh32+64	32^3	2.833	0.758	0.902	100%
NDC [6]	64^3	0.863	0.906	0.960	0%
NMC [9]	64^3	0.779	0.923	0.969	0%
VoroMesh32	64^3	0.868	0.901	0.934	100%
VoroMesh32+64	64^3	1.031	0.906	0.939	100%
NDC [6]	128^3	0.651	0.937	0.980	0%
NMC [9]	128^3	0.642	0.938	0.984	0%
VoroMesh32	128^3	0.659	0.935	0.956	100%
VoroMesh32+64	128^3	0.731	0.932	0.959	100%

Table 3: Quantitative comparisons with *VoroMesh* for inference-based 3D mesh reconstruction on the Thingi32 dataset, using input SDF grids of different resolutions.

tions, NMC slightly pulls ahead, see Table 3; we speculate that it is due to the fact that Thingi32 contains fewer thin structures, so successful reconstructions of such shapes do not demonstrate the ability to capture such details. Note that all three methods outperform classical MC across all resolutions and, in some cases, DC (although it uses richer inputs with normals at edge intersection points): with the same SDF inputs, neural methods are able to recover more information.

5. Conclusion

In this paper, we presented VoroMesh, a differentiable Voronoi-based representation of watertight 3D shape surfaces along with its VoroLoss. As the resulting watertight surface mesh is defined uniquely (but implicitly) by a Voronoi diagram of optimized generators, we showed that our representation provides significant improvement in geometric fidelity compared to previous works, and captures small or sharp features well. We demonstrated how our VoroLoss applies to either direct optimization of generators, or to the training of a neural network for prediction of generators. While we assume for simplicity of exposition that our ground truth inputs for generator optimization are triangle meshes, note that any format supporting inside/outside queries would do as well — even point sets, where, e.g., the generalized fast winding number [5] can be used to extract occupancy information.

Our work offers a number of further research directions. First, the VoroMesh could be post-processed to remove the small faces that can cause surface artifacts in the final mesh, in particular for low-resolution VoroMeshes (see supplementary material for additional details). Second, our initialization of generators has the advantage of being simple, but there may be more efficient ways to proceed, even at the risk of having initially too many generators; if some of them do not affect the VoroMesh, they can simply be dis-

carded later. Third, surface-based regularization may provide a mechanism to further improve preservation of sharp features of the input in the VoroMesh. Finally, our representation can potentially be combined with recent advances in generative modeling of point clouds, in order to produce global mesh prediction models that are suitable for generative applications with guarantees of the topological correctness of output meshes.

Acknowledgments. Nissim Maruani and Pierre Alliez are supported by the French government, through the 3IA Côte d’Azur Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-19-P3IA-0002. Maks Ovsjanikov and Roman Klokov acknowledge ERC Starting Grant No. 758800 (EXPROTEA) and the ANR AI Chair AIGRETTE. Mathieu Desbrun acknowledges the generous support of Ansys and Adobe Research, as well as a Choose France Inria chair.

References

- [1] A. Abdelkader, C. L. Bajaj, M. S. Ebeida, A. H. Mahmoud, S. A. Mitchell, J. D. Owens, and A. A. Rushdi. VoroCrust: Voronoi meshing without clipping. *ACM Transactions on Graphics (TOG)*, 39(3), 2020. [3](#)
- [2] P. Achlioptas, O. Diamanti, I. Mitliagkas, and L. Guibas. Learning representations and generative models for 3D point clouds. In *ICML*, 2018. [3](#)
- [3] N. Amenta, S. Choi, and R. K. Kolluri. The power crust, unions of balls, and the medial axis transform. *Computational Geometry*, 19(2-3), 2001. [3](#)
- [4] F. Aurenhammer. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR)*, 23(3):345–405, 1991. [2](#), [4](#)
- [5] G. Barill, N. G. Dickson, R. Schmidt, D. I. W. Levin, and A. Jacobson. Fast winding numbers for soups and clouds. *ACM Transactions on Graphics (TOG)*, 37(4), 2018. [8](#)
- [6] Z. Chen, A. Tagliasacchi, T. Funkhouser, and H. Zhang. Neural dual contouring. *ACM Transactions on Graphics (TOG)*, 41(4), 2022. [2](#), [3](#), [4](#), [5](#), [7](#), [8](#)
- [7] Z. Chen, A. Tagliasacchi, and H. Zhang. Bsp-net: Generating compact meshes via binary space partitioning. In *CVPR*, 2020. [3](#)
- [8] Z. Chen and H. Zhang. Learning implicit fields for generative shape modeling. In *CVPR*, 2019. [2](#)
- [9] Z. Chen and H. Zhang. Neural marching cubes. *ACM Transactions on Graphics (TOG)*, 40(6), 2021. [2](#), [3](#), [7](#), [8](#)
- [10] C. Choy, J.-Y. Gwak, and S. Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *CVPR*, 2019. [5](#)
- [11] B. Deng, K. Genova, S. Yazdani, S. Bouaziz, G. Hinton, and A. Tagliasacchi. Cvxnet: Learnable convex decomposition. In *CVPR*, 2020. [3](#)
- [12] P. Erler, P. Guerrero, S. Ohrhallinger, N. J. Mitra, and M. Wimmer. Points2surf learning implicit surfaces from point clouds. In *ECCV*, 2020. [2](#)
- [13] J. Gao, W. Chen, T. Xiang, A. Jacobson, M. McGuire, and S. Fidler. Learning deformable tetrahedral meshes for 3d reconstruction. In *NeurIPS*, 2020. [2](#), [3](#)
- [14] K. Genova, F. Cole, A. Sud, A. Sarna, and T. Funkhouser. Local deep implicit functions for 3d shape. In *CVPR*, 2020. [2](#)
- [15] A. Gropp, L. Yariv, N. Haim, M. Atzmon, and Y. Lipman. Implicit geometric regularization for learning shapes. In *ICML*, 2020. [2](#)
- [16] B. Guillard, F. Stella, and P. Fua. MeshUDF: Fast and differentiable meshing of unsigned distance field networks. In *ECCV*, 2022. [2](#)
- [17] E. Jang, S. Gu, and B. Poole. Categorical reparameterization with gumbel-softmax. In *ICLR*, 2017. [2](#)
- [18] T. Ju, F. Losasso, S. Schaefer, and J. Warren. Dual contouring of Hermite data. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, 2002. [3](#), [4](#), [6](#), [7](#)
- [19] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, 2006. [3](#)
- [20] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. [6](#)
- [21] R. Klokov, E. Boyer, and J. Verbeek. Discrete point flow networks for efficient point cloud generation. In *ECCV*, 2020. [3](#)
- [22] R. Klokov, J. Verbeek, and E. Boyer. Probabilistic reconstruction networks for 3D shape inference from a single image. In *BMVC*, 2019. [2](#)
- [23] S. Koch, A. Matveev, Z. Jiang, F. Williams, A. Artemov, E. Burnaev, M. Alexa, D. Zorin, and D. Panozzo. ABC: A big CAD model dataset for geometric deep learning. In *CVPR*, 2019. [7](#)
- [24] Z. Li, Z. Xu, Y. Li, X. Gu, and N. Lei. What’s the situation with intelligent mesh generation: A survey and perspectives. *IEEE transactions on visualization and computer graphics*, 2022. [2](#)
- [25] Y. Liao, S. Donne, and A. Geiger. Deep marching cubes: Learning explicit surface representations. In *CVPR*, 2018. [3](#)
- [26] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. *ACM siggraph computer graphics*, 21(4):163–169, 1987. [2](#), [4](#), [6](#), [7](#)
- [27] I. Loshchilov and F. Hutter. Decoupled weight decay regularization. In *ICLR*, 2019. [7](#)
- [28] S. Luo and W. Hu. Diffusion probabilistic models for 3d point cloud generation. In *CVPR*, 2021. [3](#)
- [29] B. Ma, Y.-S. Liu, M. Zwicker, and Z. Han. Surface reconstruction from point clouds by learning predictive context priors. In *CVPR*, 2022. [2](#)
- [30] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *CVPR*, 2019. [2](#)
- [31] M. Michalkiewicz, J. Pontes, D. Jack, M. Baktashmotlagh, and A. Eriksson. Implicit surface representations as layers in neural networks. In *ICCV*, 2019. [2](#)
- [32] C. Nash, Y. Ganin, S. A. Eslami, and P. Battaglia. Polygen: An autoregressive generative model of 3d meshes. In *ICML*, 2020. [2](#), [3](#)

- [33] J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *CVPR*, 2019. 2
- [34] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In *NeurIPS*, 2019. 6
- [35] S. Peng, C. Jiang, Y. Liao, M. Niemeyer, M. Pollefeys, and A. Geiger. Shape as points: A differentiable Poisson solver. In *NeurIPS*, 2021. 3, 6, 7
- [36] M.-J. Rakotosaona, N. Aigerman, N. J. Mitra, M. Ovsjanikov, and P. Guerrero. Differentiable surface triangulation. *ACM Transactions on Graphics (TOG)*, 40(6), 2021. 3
- [37] M.-J. Rakotosaona, P. Guerrero, N. Aigerman, N. J. Mitra, and M. Ovsjanikov. Learning Delaunay surface elements for mesh reconstruction. In *CVPR*, 2021. 3
- [38] E. Remelli, A. Lukoianov, S. R. Richter, B. Guillard, T. Bagautdinov, P. Baque, and P. Fua. MeshSDF: differentiable iso-surface extraction. In *NeurIPS*, 2020. 2
- [39] S. R. Richter and S. Roth. Matryoshka networks: Predicting 3d geometry via nested shape layers. In *CVPR*, 2018. 2
- [40] T. Shen, J. Gao, K. Yin, M.-Y. Liu, and S. Fidler. Deep marching tetrahedra: a hybrid representation for high-resolution 3d shape synthesis. In *NeurIPS*, 2021. 2, 3, 6
- [41] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein. Implicit neural representations with periodic activation functions. 2020. 2
- [42] H. Su, H. Fan, and L. Guibas. A point set generation network for 3D object reconstruction from a single image. In *CVPR*, 2017. 3
- [43] Q. Tan, L. Gao, Y.-K. Lai, and S. Xia. Variational autoencoders for deforming 3d mesh models. In *CVPR*, 2018. 3
- [44] M. Tatarchenko, A. Dosovitskiy, and T. Brox. Octree generating networks: Efficient convolutional architectures for high-resolution 3D outputs. In *ICCV*, 2017. 2
- [45] The CGAL Project. *CGAL User and Reference Manual*. CGAL Editorial Board, 5.5.2 edition, 2023. 6, 8
- [46] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17, 2020. 6
- [47] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu, and Y. Jiang. Pixel2Mesh: Generating 3D mesh models from single RGB images. In *ECCV*, 2018. 2
- [48] F. Williams, J. Parent-Levesque, D. Nowrouzezahrai, D. Panozzo, K. M. Yi, and A. Tagliasacchi. VoronoiNet: General functional approximators with local support. In *CVPR Workshop*, 2020. 2, 3, 6, 7
- [49] J. Wu, C. Zhang, T. Xue, W. Freeman, and J. Tenenbaum. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. In *NeurIPS*, 2016. 2
- [50] X. Yan, L. Lin, N. J. Mitra, D. Lischinski, D. Cohen-Or, and H. Huang. ShapeFormer: Transformer-based shape completion via sparse representation. In *CVPR*, 2022. 2
- [51] G. Yang, X. Huang, Z. Hao, M.-Y. Liu, S. Belongie, and B. Hariharan. PointFlow: 3D point cloud generation with continuous normalizing flows. In *ICCV*, 2019. 3
- [52] X. Zeng, A. Vahdat, F. Williams, Z. Gojcic, O. Litany, S. Fidler, and K. Kreis. LION: Latent point diffusion models for 3d shape generation. In *NeurIPS*, 2022. 3
- [53] Q. Zhou and A. Jacobson. Thingi10k: A dataset of 10,000 3d-printing models. *arXiv preprint arXiv:1605.04797*, 2016. 6