MorphoSkel3D: Morphological Skeletonization of 3D Point Clouds for Informed Sampling in Object Classification and Retrieval

Pierre Onghena Santiago Velasco-Forero Beatriz Marcotegui

Mines Paris, PSL University, Center for Mathematical Morphology (CMM), 77300 Fontainebleau, France
name.surname@minesparis.psl.eu

Abstract

Point clouds are a set of data points in space to represent the 3D geometry of objects. A fundamental step in the processing is to identify a subset of points to represent the shape. While traditional sampling methods often ignore to incorporate geometrical information, recent developments in learning-based sampling models have achieved significant levels of performance. With the integration of geometrical priors, the ability to learn and preserve the underlying structure can be enhanced when sampling. To shed light into the shape, a qualitative skeleton serves as an effective descriptor to guide sampling for both local and global geometries. In this paper, we introduce MorphoSkel3D[1] *as a new technique based on morphology to facilitate an efficient skeletonization of shapes. With its low computational cost, MorphoSkel3D is a unique, rule-based algorithm to benchmark its quality and performance on two large datasets, ModelNet and ShapeNet, under different sampling ratios. The results show that training with MorphoSkel3D leads to an informed and more accurate sampling in the practical application of object classification and point cloud retrieval.*

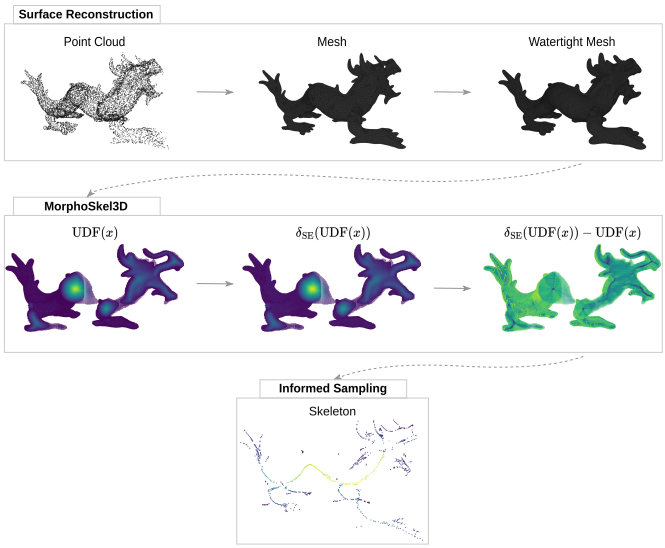


Figure 1. Overview of the proposed shape-agnostic skeletonization pipeline, illustrated with the Stanford dragon [16]. In the MorphoSkel3D module, the local maxima on the unsigned distance function of inner points reveal the set of maximal balls. The local neighbourhood of points is encoded by a structuring element SE.

1. Introduction

A skeleton is a simplified representation of the surface of a 3D shape, which makes it essential in a prior analysis to understand how the surface points are connected at different levels, from coarse- to fine-grained parts. Other common shape descriptors, such as the centroid or volume, are limited to the formulation of global rather than local insights about the geometry. Extracting a skeleton, however, has proven to be a challenging task for 3D shapes due to its extensive complexity in geometric processing. More recently, deep learning-based, skeleton models have been introduced for point clouds in an unsupervised manner to estimate skeletal spheres. While the wide-spread use of deep learning has many benefits, it's also coupled with significant challenges. Whenever skeletons are derived from learnable models, it faces certain limitations despite its advanced capabilities. One drawback is the fixed generalization over shapes, as the network's performance is heavily dependent on the diversity of training data. This dependency can hinder skeleton estimation when encountering data that diverges from the training set. Usually, the model complexity also increases with the prediction of a large number of skeletal spheres such that the network struggles to include finer details. These requirements can be a barrier to widespread adoption and iterative experimentation. Moreover, it remains an open question how to formally define the set of skeletal spheres for 3D shapes. Replacing the skeletal network with the morphological operations depicted in the overview of Fig. 1, can alleviate these

[1]Code: https://github.com/Pierreoo/MorphoSkel3D

issues by providing a more consistent and efficient method. As morphology is not dependent on training data, it can accurately capture features with low computational demands. The contribution of this paper is threefold:

- We propose a lightweight skeletonization method that is able to bring scalability to large-scale datasets.
- We integrate the skeleton as a prior to enable geometry-informed sampling of a representative subset.
- We extensively validate the skeletal quality of our method, demonstrating improved sampling performance in two tasks: object classification and point cloud retrieval.

2. Related Work

Many research efforts in the field of computer vision have focused on the development of skeletonization techniques for 2D images. However, it has been more complex to extend a formal definition of skeletonization to the 3D space for point clouds to assist shape analysis and downstream tasks [41]. This section discusses the concepts of skeletonization and morphology, to support its functionality for guided sampling.

3D shape skeletonization In recent years, several works have been proposed to tackle the sensitivity of the medial axis transform (MAT) [5] to irregular shape surfaces. For instance, Q-MAT [27] / Q-MAT+ [33] or Coverage Axis [18] / Coverage Axis++ [43] developed an algorithm with handcrafted local features to capture the geometric features that approximate a simplified medial axis transform. Whereas Q-MAT relies on watertight surfaces, Coverage Axis uses a winding numbers method [3] to select inner points for point cloud inputs. Learning-based methods learn a geometric transformation to mimic the properties of MAT, pioneered by Point2Skeleton [29], the skeletal points are considered as a local center of surface points. Consequently, rather than directly predicting the skeletal points, a convex combination of input points and weights is learned through multiple loss functions to minimize the reconstruction error. Other methods implement a Laplacian-based contraction process [7, 32], or use the first Laplace-Beltrami eigenfunction to construct Reeb graphs for high level surface information, and the barycenters of these graphs are then connected to find the skeleton [28]. More recently, implicit neural representations have been introduced to provide continuous and detailed modeling of complex shapes [17]. For example, distance functions can be effectively fitted to obtain high resolution outputs of shapes [12, 34]. Neural skeleton [14] leverages the implicit field function for a more accurate distance estimation near the medial axis to extract topological information. In order to provide a more formal definition, our method proposes to transfer the effectiveness of morphology for skeletonization from image analysis to the distance function of 3D point clouds.

3D morphology Research works in image processing have explored techniques that extend traditional 2D morphological operations, such as dilation and erosion, into the 3D image domain. These operations to form a skeleton have shown to be fundamental in preserving the topology of tabular structures in the 3D segmentation of medical images [31, 40]. To avoid the transformation of point clouds to a 2D or 3D image, morphological dilation and erosion operations have been proposed that focus on the addition and removal of points according to a structuring element [2]. Another framework, Point Morphology [6], defined a projection procedure of shifting points to find the operators that sample the medial axis in a meshless context. The extension of morphological operators to large point clouds has also been studied to better define boundaries for segmentation in urban scenes [1]. We suggest an alternative approach to implement morphological operators for efficient 3D point cloud skeletonization.

Point cloud sampling An obstacle in the processing of point clouds is the high density of data, which can hinder the effective use of deep learning models [24]. The aim of sampling is to extract a representative subset from the original point cloud to ensure that the resulting sampled point cloud is both representative of the original data and optimized for further processing and analysis. For instance, to maintain the classification and segmentation performances. This goal aligns with the method of Wen et al. [44] that explores the object skeleton to preserve geometrical information during sampling. In their work, the authors adopted a deep learning approach similar to Point2Skeleton [29] to estimate the skeleton. An ablation study between the learned skeleton and DPC [45] skeleton demonstrated comparable results in their ability to improve the sampling of representative points. Therefore, as a metric to reflect skeletal quality, we incorporate our skeleton to evaluate its ability to inform sampling.

3. Background

To sharpen the intuition before defining our method, this section reviews the background to afterwards introduce morphological skeletonization. The concept of a skeletal structure is first discussed as a geometric approach to introduce the notion of the skeleton of a set of inner points. The goal is to leverage the properties of a watertight mesh to accurately determine an occupancy and distance function of points lying within the object. Based on the distance of inner points to the surface, we propose a highly efficient process that is able to extract the skeletal points from a shape. The interest primarily lies within the inner points, as these serve as the candidate skeletal points. To achieve this, two sequential steps are employed. Firstly, an occupancy function is applied to determine the points within the mesh. Following this determination, an unsigned distance function is computed

on these points to facilitate morphological operators to the extraction of the skeletal structure.

3.1. Skeleton by maximal balls

A rich segment of mathematical literature is devoted to the study of the central part of a set in Euclidean space. While various definitions have been proposed, their central parts are often similar yet not always equivalent. In particular, given a closed surface $\mathcal{S}$, the *medial axis* $\mathcal{MA}(\mathcal{S})$ is identified by the set of centers of maximally inscribed spheres that are tangent to $\mathcal{S}$ at two or more points. Usually referred to as the rowboat analogy, the medial axis has been introduced by Blum [5] in image analysis and by Beucher and Lantuejoul [4, 26] in mathematical morphology. Other studies, including [9] and [10], study the medial axis from a different perspective. For any x in volume $\mathcal{V}$ enclosed by $\mathcal{S}$, let $D(x)$ denote the set of medial balls:

$$D(x) = \{y \in \mathcal{S} : d(x, y) = d(x, \mathcal{S})\}, \tag{1}$$

where $d(x, y)$ denotes the Euclidean distance between x and y. The distance between the point and surface is defined as $d(x, \mathcal{S}) = \inf_{y \in \mathcal{S}} d(x, y)$, where $\inf$ denotes the infimum. Then, the medial axis of a closed surface $\mathcal{S}$ is defined as:

$$\mathcal{MA}(\mathcal{S}) := \{x \in \mathcal{S} : |D(x)| \geq 2\}, \tag{2}$$

where $|D(x)|$ denotes the cardinality of $D(x)$. The medial axis together with the distance from each point to its closest surface point defines the medial axis transform (MAT) [13] as a powerful tool in computer vision and image processing. The related concept of a skeleton is often mistakenly conflated with the medial axis. The *skeleton* of a closed surface $\mathcal{S}$, denoted $\mathcal{SK}(\mathcal{S})$, is namely defined as the set of centers of maximal balls inscribed in $\mathcal{S}$. In specific, a ball $B(x, i)$ defined by its center at point x and radius i is considered maximal within $\mathcal{S}$ if there exists no other ball $B(y, j)$ such that $B(x, i)$ is contained within $B(y, j)$ with a larger radius $j > i$. We emphasize that the set of centers of maximal inscribed disks is generally not connected in discrete distance maps [4, 15, 20]. Consequently, while this set is occasionally referred to as the skeleton, it is primarily useful for compression rather than for shape analysis tasks. A more conventional definition of a skeleton requires that it preserves the same simple connectivity of $\mathcal{S}$, to make it homotopic to the surface. Typically, this involves a linkage of the centers of maximal inscribed balls to produce a connected skeleton [20, 41]. Although things can become more complicated in the 3D space, the proposed approach, that is lightweight and object-agnostic, is introduced in section Sec. 4. The skeletal representation focuses on the maximal inscribed balls to provide a compact and meaningful abstraction of the original set.

3.2. Occupancy Function

Based on a bounding box that encapsulates the object, random points are uniformly distributed in the volume to ensure the coverage of the entire shape. An example shape with bounding box points is illustrated on the left of Fig. 2. Following this, an occupancy function is applied to determine whether each bounding box point lies within the watertight surface mesh. It helps to identify the points inside the shape and excludes those outside its surface. The occupancy function that follows the ray casting algorithm [38], evaluates a given point by analyzing its intersections with the surface triangles. Let $p = (x, y, z) \in \mathbb{R}^3$ be a bounding box point and $\mathcal{M}(T)$ the watertight mesh that consists of a set of triangles T. The number of intersections between the ray cast from p and the triangles in T is denoted by $N_p(T)$. This algorithm is also referred to as the crossing number algorithm or the even–odd rule algorithm [39]. Then, the *occupancy function* $\mathcal{O} : \mathbb{R}^3 \mapsto \{0, 1\}$ for the bounding box point p with respect to the mesh $\mathcal{M}(T)$ can be computed by:

$$\mathcal{O}(p) = \begin{cases} 1, & \text{if } N_p(T) \text{ is odd,} \\ 0, & \text{if } N_p(T) \text{ is even.} \end{cases} \tag{3}$$

If the number of intersections between a ray cast from p and the triangles in T is odd, $\mathcal{O}(p)$ equals to one, indicating that p is inside the mesh. Conversely, if the number of intersections is even, $\mathcal{O}(p)$ equals to zero, indicating that p is outside the mesh. An example illustrating the selection of inner points is provided on the right side of Fig. 2. By $\mathcal{O}$, we consider $\mathcal{M}(T)$ as a closed surface $\mathcal{S}$ to ensure that the mesh forms a completely enclosed surface without gaps.

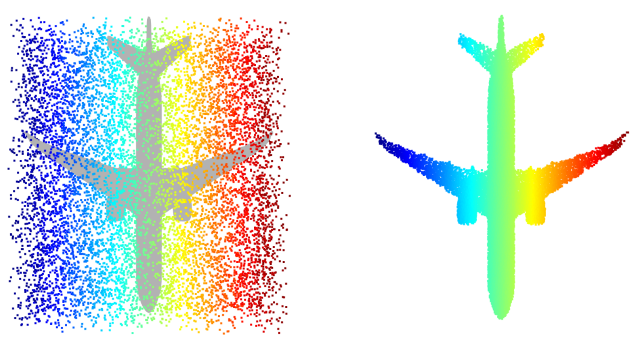


Figure 2. Occupancy function used to retain the inner points from the bounding box points. A case for 10000 points in the bounding box (left). Inner points extracted for the case with 1M points within the box volume (right).

3.3. Signed Distance Function

Given a closed surface $\mathcal{S}$ in metric space $\mathcal{C} \subseteq \mathbb{R}^3$, a *signed distance function* (SDF) is a continuous function that provides the closest distance from a given point $x \in \mathbb{R}^3$ to the

surface $\mathcal{S}$. The sign indicates whether or not the point x lies within the interior volume $\mathcal{V}$ enclosed by $\mathcal{S}$. The signed distance function outputs a negative value for points inside the watertight mesh and increases as it approaches the surface or zero-level set. For points outside the mesh, it takes positive values that become larger as one moves further away from the surface:

$$\mathrm{SDF}(x) = \begin{cases} d(x, \mathcal{S}) & \text{if } x \notin \mathcal{V}, \\ -d(x, \mathcal{S}) & \text{if } x \in \mathcal{V}. \end{cases} \tag{4}$$

The utilization of the occupancy function to discern points within a mesh offers an advantage to circumvent the necessity of performing the distance function for every point within the bounding box of the object. Having already identified the inner points, see Fig. 2 right, the signed distance function could be simplified to the unsigned distance function. This approach eliminates the need to calculate distances for points outside the object's surface, thus optimizing computational efficiency. An example of the distance function is illustrated in Fig. 3b. Additionally, the *unsigned distance function* (UDF) only provides positive distances such that further morphological operations can be simplified to:

$$\mathrm{UDF}(x) = d(x, \mathcal{S}) \quad \text{for } x \in \mathcal{V}. \tag{5}$$

4. Method

The goal of Wen et al. [44] was to identify a subset of points that maintains classification performance using only a few selected points. This process involves the learning of a sampling strategy that is informed by prior knowledge of the skeleton. However, the skeleton estimation network is limited due to its inability to generalize to other shapes than the training set. In this work, we propose a novel skeletonization method, named MorphoSkel3D, that is object-agnostic through a series of morphological operators to give priority to the points belonging to the skeleton $\mathcal{SK}(\mathcal{S})$.

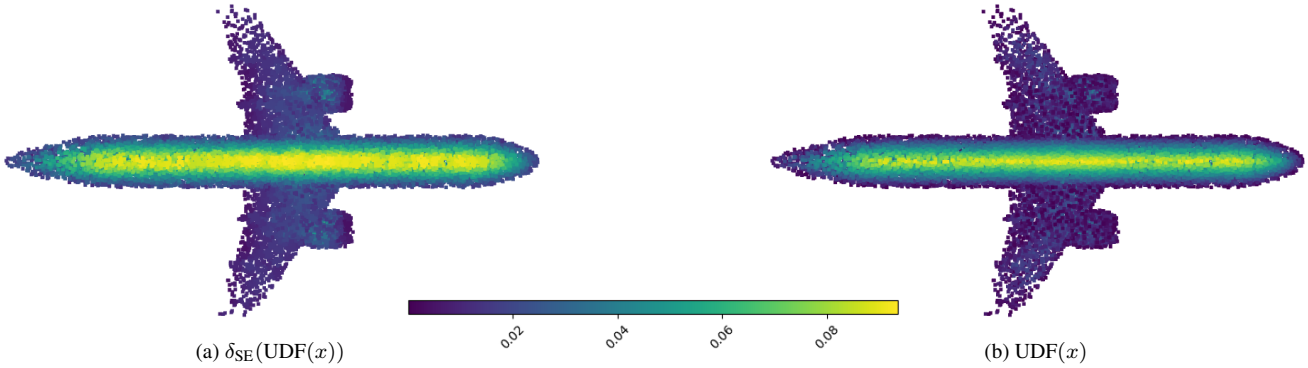
4.1. Morphological Skeletonization

In this method, a sequence of concepts and operators is presented based on the notion of distance. The objective is to outline a technique that characterizes a set of points based on the considerations of metric and geometry. An important concept includes that of a skeleton, which is of fundamental importance in computer vision, and has many practical applications for point clouds. In the analysis of the unsigned distance function, a property becomes transparent that local maxima correspond to maximal balls within the given space. These local maxima represent regions where the distance function reaches its greatest value within a neighborhood. The identification of these maxima becomes essential for understanding the structure of space. An approach to tackle this involves the study of the difference between the original distance and its *dilation*. This analysis serves as a means to identify local maxima and mark the arrangement of a skeleton. Hence, the difference in the distance function among the inner points of $\mathcal{S}$ is described as:

$$\mathrm{MS3D}(x) = \delta_{\mathrm{SE}}(\mathrm{UDF}(x)) - \mathrm{UDF}(x) \quad \text{for } x \in \mathcal{V}, \tag{6}$$

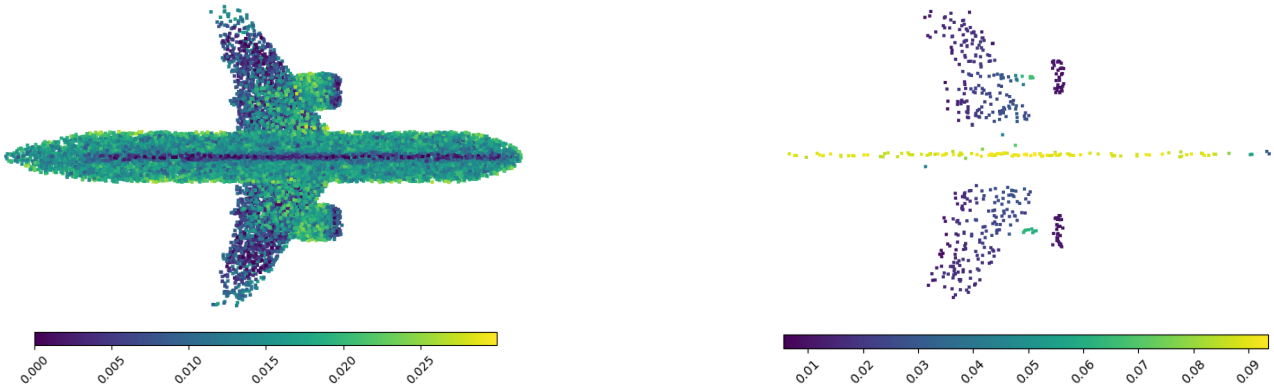
where δ_{SE} denotes a dilation equipped by the structuring element SE. A dilation operator takes the maximum value of UDF within the neighborhood defined by SE. Note that Rosenfeld and Pfaltz [37] have shown for 2D images that, to find the set of maximal balls, it is sufficient to detect the local maxima on the distance function. With a graph to construct local connections, we use the same idea for 3D shapes to highlight the inner points that are candidates to be maximal ball centers. An option to define the structuring element is by using a k-nearest neighbor (k-NN) graph. For a point $x \in \mathcal{V}$, a directed graph $G = (V, E)$ is constructed, representing the local structure of the volume point cloud, where V and E denote the vertices and edges, respectively. The structuring element SE is then defined as:

$$\mathrm{SE}(x) = k\text{-NN}(x) \quad \text{for } x \in \mathcal{V}. \tag{7}$$



(a) $\delta_{\mathrm{SE}}(\mathrm{UDF}(x))$ (b) $\mathrm{UDF}(x)$

Figure 3. Transformed distance function by dilation (left) and its initial unsigned distance function (right). Brighter points indicate greater distances, while darker areas correspond to points closer to the surface.



(a) MS3D(x)

(b) 1024 skeletal points with lowest MS3D(x)

Figure 4. Based on 1M bounding box points, MS3D(x) as the difference between the dilated and original UDF (left). The 1024 inner points with lowest MS3D(x) values are retained to form the skeleton with point colors based on their original UDF to the surface (right).

The local information encapsulated in this graph is leveraged to take the maximum value of the distance function in the neighborhood. We denote this operation of transforming the distance function by dilation as δ_{SE}. An illustration of a dilated distance function is presented in Fig. 3a, next to the initial unsigned distance function on the right in Fig. 3b. The resulting MS3D(x) is illustrated in Fig. 4a to provide a visual understanding of the transformation between the dilation operation and the original distance function.

To make sure that every skeleton used for training has the same number of points, a subset $\mathcal{K}$ from the set of inner points $\mathcal{V}$ is selected with the lowest MS3D(x) values, where MS3D(x) measures how close a point is to being a true skeletal point. We consider that the skeletal points, by definition of maximal balls, have a MS3D(x) value equal to zero. It involves identifying a subset $\mathcal{K}$ of size n that minimizes the sum of MS3D(k) for all elements $k \in \mathcal{K}$:

$$\underset{\mathcal{K}\subset\{1,2,\ldots,S\}:|\mathcal{K}|=n}{\arg\min} \sum_{k\in\mathcal{K}} \text{MS3D}(k). \tag{8}$$

As a result, $\mathcal{K}$ consists of points closest to the center of a maximal ball and defines the skeletal points by minimizing MS3D(x) values to provide a consistent subset for network training. In Fig. 5, the distribution of MS3D(x) values reveals its dependency on the number of bounding box points. As the number of points in the bounding box increases, more accurate skeletal points appear. Although the skeleton becomes better defined, the increase in bounding box lightly prolongs the computational time required for its creation. The Fig. 4b illustrates a subset of skeletal points retained by our proposed method. It shows that the inner points with lowest MS3D(x) values are used to find the maximal balls that form the skeleton. The skeletal points in this representation are color-coded based on their initial UDF values to the surface to represent skeletal spheres.

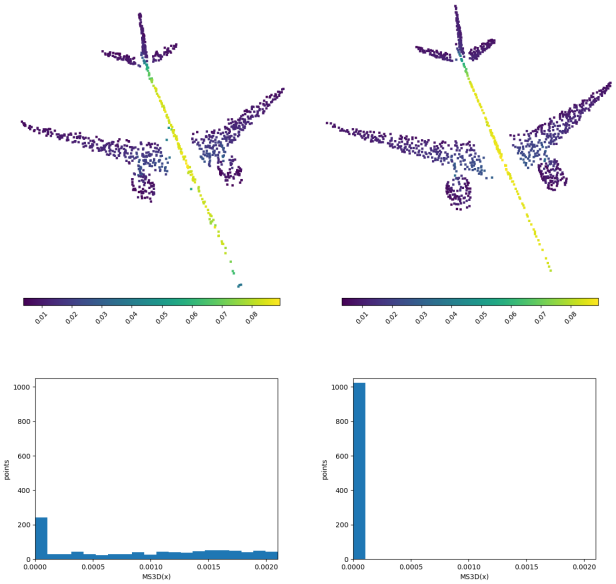


Figure 5. The 1024 inner points selected by MS3D(x), color-coded by their original UDF to the surface (up). Distribution of MS3D(x) for 1024 skeletal points (down), contingent to the density of 1M (left) to 10M (right) bounding box points.

4.2. Prior Sampling Probability

The sampling process can be informed by prior knowledge that sits contained in the object's skeleton. By utilizing the concept of *local feature size* (LFS), which measures the Euclidean distance from a surface point $p \in \mathcal{M}$ to the medial axis or skeleton, it becomes apparent that smaller LFS values denote intricate regions in the object. Hence, the strategy involves an augmentation of sampling probability in small LFS areas to enhance the overall representation. A heatmap where the prior sampling probability is visualized for each

surface point is shown in Fig. 6. In this heatmap, areas with higher weights correspond more to fine-grained regions. In this way, the overall shape will be represented in the subset to improve the task network in its practical application.

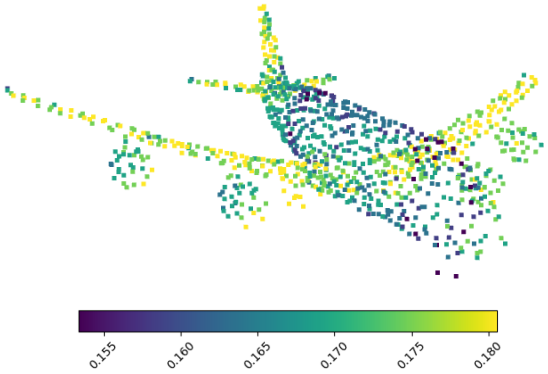


Figure 6. Heatmap of the prior per-point sampling weight, derived from the proposed MS3D skeletal structure.

5. Experiments

To evaluate the quality of the skeleton through morphology, the experimental setup is firstly discussed to assess the skeleton extraction in terms of reconstruction error and processing time. Once these metrics are compared, we apply our skeleton on two downstream tasks: object classification and point cloud retrieval.

5.1. Experimental Setup

Datasets The ShapeNet [8] and ModelNet [46] dataset are employed to establish an evaluation about the skeleton and its functionality as a prior. To compare our morphological skeleton, we use the same subset of the ShapeNet dataset as the Point2Skeleton [29] benchmark that includes 7088 shapes from eight categories. Each object has 2000 points for surface reconstruction to allow the extraction of its mesh and the selection of inner points. For the downstream tasks enhanced by a skeletal prior, the normalized and resampled ModelNet40 dataset of 12311 point clouds spread over 40 categories is used to align the benchmark with existing methods. In this dataset, each shape has 10000 points to reconstruct its surface that is made watertight to determine the inner bounding box points. The skeletons for both datasets are in turn extracted from the inner points by our MorphoSkel3D method. Based on the analysis in Fig. 5, the number of bounding box points is set to 10M to optimize the quality of the skeleton while keeping a low computation time. The resulting dataset consists of 1024 skeletal spheres that are ordered in ascending MS3D(x) value. We notice that this setting includes accurate skeletal points with MS3D(x) values of zero or close to.

Architectures For the first downstream task of object classification, Wen et al. [44] propose to learn a representative subset of the original point cloud with a selection method that is trained with the Gumbel-softmax trick to enable discrete yet differentiable sampling:

$$\mathcal{P}_{\text{sub}} = [Gs_\tau(Lp(\mathcal{W}_{\text{LFS}} \odot \mathcal{W}_\theta)]^\top \mathcal{P}, \tag{9}$$

where $\mathcal{P}_{\text{sub}} \in \mathbb{R}^{M\times 3}$ denotes the extracted subset derived from the input point cloud $\mathcal{P} \in \mathbb{R}^{N\times 3}$. The weight $\mathcal{W}_{\text{LFS}} \in \mathbb{R}^{N\times 1}$ is computed in the prior phase from the skeleton, while $\mathcal{W}_\theta \in \mathbb{R}^{N\times D_{\text{sp}}}$ represents the feature matrix obtained from the sampling network, with D_{sp} denoting the feature dimension. The linear mapping layer $Lp(\cdot) : \mathbb{R}^{N\times D_{\text{sp}}} \rightarrow \mathbb{R}^{N\times M}$ is applied, and the element-wise Hadamard product operation is denoted as $\odot$. The $Gs_\tau(\cdot)$ operation refers to the Gumbel-softmax where the confidence of the softmax is controlled by an annealed temperature τ. For the sampling network to obtain a feature matrix $\mathcal{W}_\theta$, the DGCNN [42] architecture is applied with four edge convolution layers to capture local geometric features between points in the dynamic feature graph. The learned embedding $\mathcal{W}_\theta$ is merged with the skeletal information $\mathcal{W}_{\text{LFS}}$ by Hamard product to facilitate the Gumbel-softmax to learn the optimal subset of points for the task. During the sampling process, a pre-trained PointNet [35] classification model operates in evaluation mode with its loss to guide the sampling network in classifying objects on fewer points. For the second task of point cloud retrieval, the trained sampling model of the first task is in evaluation to sample points for the classification model to form a global descriptor and compare representations for retrieval.

Metrics We follow other skeletonization techniques in the evaluation to ensure a fair comparison. Namely, the reconstruction error is evaluated to quantify how accurately the generated skeletal spheres reconstruct the original point cloud. Another way to evaluate the skeleton is by comparing it to the manually simplified MAT, which serves as the ground truth skeleton in the Point2Skeleton [29] benchmark. This dual evaluation approach uses the Chamfer and Hausdorff distance to provide a reflection of discrepancy between the two sets. However, relying solely on these metrics does not fully capture the method's skeletonization capability, especially when a small reconstruction error is returned for a generated skeleton that closely matches the surface point cloud. Therefore, a benchmark on two downstream applications is also conducted to measure the skeleton's ability to improve sampling. For the classification task, the evaluation metric is based on the overall accuracy (OA) across all categories. The evaluation scheme for point cloud retrieval aligns with the mean Average Precision (mAP) that is derived from averaging the precision scores across point clouds based on L_2 distances between shape descriptors.

	CD-Recon				HD-Recon				CD-MAT				HD-MAT			
	L1	DPC	P2S	MS3D	L1	DPC	P2S	MS3D	L1	DPC	P2S	MS3D	L1	DPC	P2S	MS3D
Airplane	0.0378	0.0348	0.0363	**0.0152**	0.2216	0.1436	0.1266	**0.0933**	0.0793	0.1307	0.0611	**0.0368**	0.2384	0.2580	0.1721	**0.1598**
Chair	0.1126	0.0769	0.0441	**0.0296**	0.4810	0.2478	**0.1618**	0.2347	0.1885	0.2286	0.0974	**0.0659**	0.4991	0.3707	**0.2151**	0.3123
Table	0.1041	0.0853	0.0424	**0.0366**	0.3453	0.2584	**0.1745**	0.2205	0.1541	0.2683	0.0876	**0.0823**	0.3583	0.3690	**0.2085**	0.3107
Lamp	0.1542	0.0712	0.0335	**0.0265**	0.3956	0.1850	**0.1382**	0.1896	0.1870	0.1751	0.0884	**0.0606**	0.4089	0.2627	**0.2003**	0.2595
Guitar	0.0655	0.0212	0.0179	**0.0085**	0.2180	0.0589	0.0625	**0.0553**	0.0817	0.0672	0.0536	**0.0259**	0.2262	0.1226	0.1216	**0.1046**
Earphone	0.0437	0.0573	0.0399	**0.0245**	0.1908	0.2059	**0.1125**	0.1686	0.0607	0.2216	0.1638	**0.0934**	**0.1732**	0.3403	0.2130	0.2815
Mug	0.2864	0.1280	**0.0417**	0.0580	0.9142	0.3510	**0.1419**	0.3201	0.5316	0.4600	0.1179	**0.1126**	0.9057	0.4308	**0.2158**	0.4121
Rifle	0.0260	0.0215	0.0213	**0.0097**	0.1078	0.0702	0.0767	**0.0584**	0.0494	0.0427	0.0356	**0.0250**	0.1234	0.1050	0.0957	**0.0873**
Average	0.1038	0.0668	0.0372	**0.0274**	0.3593	0.2049	**0.1424**	0.1857	0.1665	0.2026	0.0828	**0.0629**	0.3667	0.3047	**0.1898**	0.2601

Table 1. Comparison of reconstruction error between the skeletal spheres of skeletonization methods to the surface point cloud (Recon) and the ground truth skeleton (MAT), Chamfer (CD) and Hausdorff (HD) distances.

5.2. Skeleton Extraction

The initial step before skeletonization is to convert a surface point cloud into its mesh representation with a shape reconstruction method. Among traditional methods, we opt for the Poisson surface reconstruction [23] algorithm to transform the test objects of the ShapeNet subset into meshes. Consequently, we generate watertight meshes [22] in the surface reconstruction module to compute a distance function for the inner points. The k is set to 20 for the k-NN graph to perform a dilation over the inner points in Eq. (6). In the skeletal analysis, we compare against other methods as L_1-medial skeleton [21], deep point consolidation (DPC) [45], and the learning-based Point2Skeleton (P2S) [29]. Whereas the number of estimated skeletal spheres is 100 for P2S, we select 1024 points to form the set of maximal balls for MS3D. For two examples of ShapeNet, a qualitative visualization of our method is given in Fig. 7 to illustrate the skeletal spheres in relation to the original point cloud. The reconstruction results are reported in Tab. 1. We observe that MS3D achieves the lowest Chamfer distance to prove its ability to define maximal balls and effectively extract skeletal spheres. A lower Hausdorff distance of MS3D is observed in three out of eight categories, but the other categories reveal its dependence on a surface reconstruction method to provide reliable meshes. Namely, as reconstruction is not evident with 2000 surface points, certain areas could have incorrect reconstruction to include false inner points for MS3D to skeletonize. Therefore, we implement PoNQ [30] as a state-of-the-art optimization-based reconstruction method to provide meshes for the downstream tasks with the ModelNet dataset. Two examples of ModelNet are presented in Fig. 8 to display the robust, object-agnostic skeletonization without the need to learn a geometric transformation. The corresponding processing times from surface reconstruction to the MS3D module are reported in Tab. 2 for each category to highlight it's efficiency and steadiness on an Intel(R) Core(TM) i7-12800H CPU with 32 GB memory. For reference, the processing time to skeletonize with L_1 or DPC on a 100K point cloud is unsteady but takes approximately 60 seconds next to the high Chamfer and Hausdorff distances.

		Poisson [23]		Watertight [22]		MS3D	
	Num	T_{Total}	$T_{Average}$	T_{Total}	$T_{Average}$	T_{Total}	$T_{Average}$
Airplane	100	34.83	0.35	119.79	1.20	73.99	0.74
Chair	334	98.57	0.30	590.40	1.70	538.77	1.61
Table	334	103.68	0.31	559.71	1.68	661.61	1.98
Lamp	167	53.04	0.32	282.66	1.69	273.76	1.64
Guitar	100	28.66	0.29	190.81	1.91	144.97	1.45
Earphone	13	3.92	0.30	21.47	1.65	17.55	1.35
Mug	36	10.20	0.28	39.01	1.08	105.14	2.92
Rifle	100	33.57	0.34	147.12	1.47	160.25	1.60

Table 2. Comparison of total and average processing times across categories from surface reconstruction to MorphoSkel3D, time (s).

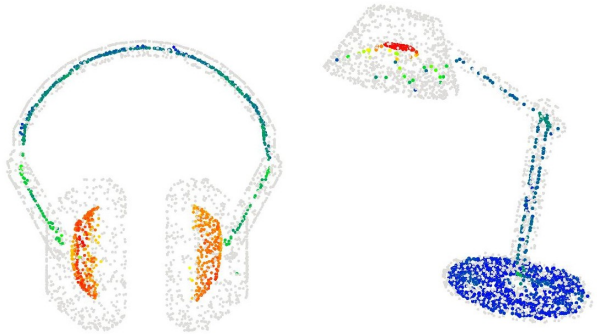


Figure 7. The skeletal spheres of MS3D together with its surface points for an earphone and lamp example of the ShapeNet subset.

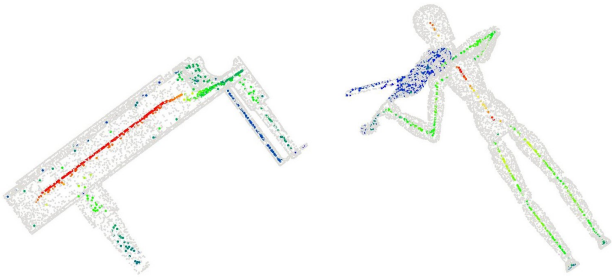


Figure 8. The skeletal spheres of MS3D together with its surface points for a piano and person example of the ModelNet dataset.

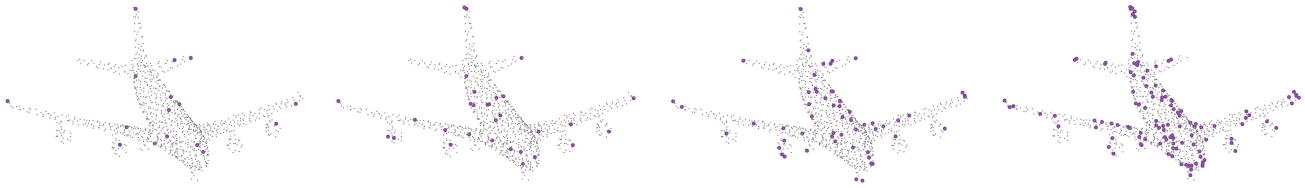


Figure 9. The sampled points through the MorphoSkel3D method, together with the surface points of an airplane example. From left to right the four ratios: 16, 32, 64 and 128 sampled points.

5.3. Object Classification

The pre-trained classification network is optimized for inputs with 1024 surface points. During training, the model underwent data augmentation to improve its generalization and to follow the standard setting for classification. The upper bound classification accuracy is 89.5% when tested on the same number of 1024 points. To further test its robustness against lower point cloud densities, a comparison is conducted between the traditional FPS and learnable methods. Although our MorphoSkel3D method employs the learnable sampling process from the skeleton-aware (SA) method proposed by Wen et al. [44], the objective is to conduct an ablation study that compares the ability of the skeleton as a prior to enhance sampling weights. On the one hand, SA aims to learn the object skeleton in an unsupervised manner similar to P2S [29]. On the other hand, MS3D introduces an efficient, object-agnostic approach to find the maximal balls for skeletal extraction. The methods are evaluated across four sampling ratios: 8, 16, 32, and 64 to reduce the original 1024 points to 128, 64, 32, and 16 points respectively. In Tab. 3, the results confirm that a learnable sampling strategy is significantly more effective than FPS, with the accuracy gap widening as the sampling ratio increases. For a sampling ratio of 8, MS3D maintains the accuracy of the upper bound with a limited decrease at higher ratios. When comparing the influence of the skeleton between SA and MS3D, it can be deducted that the incorporation of a morphological skeleton achieves improvements to better guide the sampling over all four sampling ratios. The advantage is most pronounced at the highest sampling ratio with 85.5%. The outcome of the learned sampling is demonstrated in Fig. 9, where it becomes apparent that the sampled points are spread across the intricate parts of the object.

Ratio	FPS [36]	S-NET [19]	SN [25]	MS [11]	SA [44]	MS3D
8	70.4	77.5	83.7	88.0	89.1	**89.5**
16	46.3	70.4	82.2	85.5	88.8	**88.9**
32	26.3	60.6	80.1	81.5	87.4	**87.8**
64	13.5	36.1	54.1	61.6	82.9	**85.5**

Table 3. Object classification results on ModelNet40, OA (%).

5.4. Point Cloud Retrieval

The integration of morphological skeleton demonstrated an effective selection of points in fine-grained features and, therefore, we transfer the learned sampling networks initially trained for classification to perform shape retrieval. We follow the evaluation scheme of SA [44], where the resulting global feature vector is used to search for similar point clouds based on Euclidean distance. In Tab. 4, the retrieval results are reported to showcase the additional performance boost of MS3D across all four sampling ratios.

Ratio	FPS [36]	S-NET [19]	SN [25]	SA [44]	MS3D
8	58.3	60.4	68.8	72.2	**72.9**
16	49.4	59.0	65.2	70.9	**71.4**
32	37.7	59.0	62.5	67.1	**68.7**
64	27.4	54.5	59.5	62.6	**66.7**

Table 4. Point cloud retrieval results on ModelNet40, mAP (%).

6. Conclusion

We propose a method that studies a distance function of inner points to highlight the maximal balls in the skeleton. In particular, our approach that applies simple morphological operations on a neighborhood graph is shape-agnostic. The defined skeleton is then utilized to formulate a sampling strategy that emphasizes surface points in intricate regions. This technique enhances the sampling network's ability to learn a representative subset, facilitating two downstream tasks in the representation of the original point cloud. Beyond these two applications, we believe our framework can lay the groundwork to scale prior knowledge in geometric processing for deep learning models.

7. Acknowledgement

This work was granted access to the HPC resources of IDRIS under the allocation 2023-AD011014750 made by GENCI.

References

[1] Teo Asplund, Andrés Serna, Beatriz Marcotegui, Robin Strand, and Cris L. Luengo Hendriks. Mathematical mor-

phology on irregularly sampled data applied to segmentation of 3D point clouds of urban scenes. In *Mathematical Morphology and Its Applications to Signal and Image Processing*, pages 375–387, Cham, 2019. Springer International Publishing. 2

[2] Jesús Balado, Peter van Oosterom, Lucía Díaz-Vilariño, and Martijn Meijers. Mathematical morphology directly applied to point cloud data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 168:208–220, 2020. 2

[3] Gavin Barill, Nia Dickson, Ryan Schmidt, David I.W. Levin, and Alec Jacobson. Fast winding numbers for soups and clouds. *ACM Transactions on Graphics*, 2018. 2

[4] Serge Beucher. Digital skeletons in Euclidean and geodesic spaces. *Signal Processing*, 38(1):127–141, 1994. Mathematical Morphology and its Applications to Signal Processing. 3

[5] Harry Blum. A transformation for extracting new descriptions of shape. *Models for the perception of speech and visual form*, pages 362–380, 1967. 2, 3

[6] Stéphane Calderon and Tamy Boubekeur. Point morphology. *ACM Transactions on Graphics (Proc. SIGGRAPH 2014)*, 33 (4):45:1–45:13, 2014. 2

[7] J. Cao, A. Tagliasacchi, M. Olson, H. Zhang, and Z. Su. Point cloud skeletons via Laplacian based contraction. In *Shape Modeling International (SMI 2010)*, pages 187–197, Los Alamitos, CA, USA, 2010. IEEE Computer Society. 2

[8] 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 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago, 2015. 6

[9] John Chaussard, Michel Couprie, and Hugues Talbot. Robust skeletonization using the discrete λ-medial axis. *Pattern Recognition Letters*, 32(9):1384–1394, 2011. 3

[10] Frédéric Chazal and Rémi Soufflet. Stability and finiteness properties of medial axis and skeleton. *Journal of Dynamical and Control Systems*, 10:149–170, 2004. 3

[11] Ta-Ying Cheng, Qingyong Hu, Qian Xie, Niki Trigoni, and Andrew Markham. Meta-sampler: Almost-universal yet task-oriented sampling for point clouds, 2022. 8

[12] Julian Chibane, Aymen Mir, and Gerard Pons-Moll. Neural unsigned distance fields for implicit function learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. 2

[13] Hyeong In Choi, Sung Woo Choi, and Hwan Pyo Moon. Mathematical theory of medial axis transform. *Pacific Journal of Mathematics*, 181(1):57–88, 1997. 3

[14] Mattéo Clémot and Julie Digne. Neural skeleton: Implicit neural representation away from the surface. *Computers & Graphics*, 114:368–378, 2023. 2

[15] Michel Couprie and Hugues Talbot. Distance, granulometry, skeleton. In *Distance, granulometry, skeleton*, pages 291–316. Wiley, 2011. 3

[16] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, page 303–312, New York, NY, USA, 1996. Association for Computing Machinery. 1

[17] Luca De Luigi, Adriano Cardace, Riccardo Spezialetti, Pierluigi Zama Ramirez, Samuele Salti, and Luigi Di Stefano. Deep learning on implicit neural representations of shapes. In *International Conference on Learning Representations (ICLR)*, 2023. 2

[18] Zhiyang Dou, Cheng Lin, Rui Xu, Lei Yang, Shiqing Xin, Taku Komura, and Wenping Wang. Coverage axis: Inner point selection for 3D shape skeletonization. In *Computer Graphics Forum*, pages 419–432. Wiley Online Library, 2022. 2

[19] Oren Dovrat, Itai Lang, and Shai Avidan. Learning to sample. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 8

[20] Yaorong Ge and J Michael Fitzpatrick. On the generation of skeletons from discrete Euclidean distance maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(11):1055–1066, 1996. 3

[21] Hui Huang, Shihao Wu, Daniel Cohen-Or, Minglun Gong, Hao Zhang, Guiqing Li, and Baoquan Chen. L1-medial skeleton of point cloud. *ACM Trans. Graph.*, 32(4), 2013. 7

[22] Jingwei Huang, Hao Su, and Leonidas Guibas. Robust watertight manifold surface generation method for ShapeNet models. *arXiv preprint arXiv:1802.01698*, 2018. 7

[23] Michael M. Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Proceedings of the Fourth Eurographics Symposium on Geometry Processing*, pages 61–70, Aire-la-Ville, Switzerland, Switzerland, 2006. Eurographics Association. 7

[24] Loic Landrieu and Simonovsky Martin. Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs. In *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018)*, Salt Lake City, United States, 2018. 2

[25] Itai Lang, Asaf Manor, and Shai Avidan. SampleNet: Differentiable Point Cloud Sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7578–7588, 2020. 8

[26] Christian Lantuejoul. *La squelettisation et son application aux mesures topologiques des mosaiques polycristallines*. PhD thesis, Ecole des Mines de Paris, 1978. 3

[27] Pan Li, Bin Wang, Feng Sun, Xiaohu Guo, Caiming Zhang, and Wenping Wang. Q-MAT: Computing medial axis transform by quadratic error minimization. 35(1), 2016. 2

[28] Jian Liang, Rongjie Lai, Tsz Wai Wong, and Hongkai Zhao. Geometric understanding of point clouds using Laplace-Beltrami operator. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 214–221, 2012. 2

[29] Cheng Lin, Changjian Li, Yuan Liu, Nenglun Chen, Yi-King Choi, and Wenping Wang. Point2skeleton: Learning skeletal representations from point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4277–4286, 2021. 2, 6, 7, 8

[30] Nissim Maruani, Maks Ovsjanikov, Pierre Alliez, and Mathieu Desbrun. PoNQ: a neural QEM-based mesh representation, 2024. 7

[31] Martin J. Menten, Johannes C. Paetzold, Veronika A. Zimmer, Suprosanna Shit, Ivan Ezhov, Robbie Holland, Monika Probst, Julia A. Schnabel, and Daniel Rueckert. A skeletonization algorithm for gradient-based optimization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. 2

[32] Lukas Meyer, Andreas Gilson, Oliver Scholz, and Marc Stamminger. CherryPicker: Semantic skeletonization and topological reconstruction of cherry trees, 2023. 2

[33] Yiling Pan, Bin Wang, Xiaohu Guo, Hua Zeng, Yuexin Ma, and Wenping Wang. Q-MAT+: An error-controllable and feature-sensitive simplification algorithm for medial axis transform. *Computer Aided Geometric Design*, 71:16–29, 2019. 2

[34] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2

[35] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3D classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 6

[36] Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: deep hierarchical feature learning on point sets in a metric space. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, page 5105–5114, Red Hook, NY, USA, 2017. Curran Associates Inc. 8

[37] Azriel Rosenfeld and John L Pfaltz. Sequential operations in digital picture processing. *Journal of the ACM (JACM)*, 13 (4):471–494, 1966. 4

[38] Scott D Roth. Ray casting for modeling solids. *Computer graphics and image processing*, 18(2):109–144, 1982. 3

[39] Moshe Shimrat. Algorithm 112: position of point relative to polygon. *Communications of the ACM*, 5(8):434, 1962. 3

[40] Suprosanna Shit, Johannes C Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien PW Pluim, Ulrich Bauer, and Bjoern H Menze. clDice-a novel topology-preserving loss function for tubular structure segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16560–16569, 2021. 2

[41] Andrea Tagliasacchi, Thomas Delame, Michela Spagnuolo, Nina Amenta, and Alexandru Telea. 3D skeletons: A state-of-the-art report. In *Computer Graphics Forum*, pages 573–597. Wiley Online Library, 2016. 2, 3

[42] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph CNN for learning on point clouds. *ACM Transactions on Graphics (TOG)*, 2019. 6

[43] Zimeng Wang, Zhiyang Dou, Rui Xu, Cheng Lin, Yuan Liu, Xiaoxiao Long, Shiqing Xin, Lingjie Liu, Taku Komura, Xiaoming Yuan, et al. Coverage Axis++: Efficient inner point selection for 3D shape skeletonization. *arXiv preprint arXiv:2401.12946*, 2024. 2

[44] Cheng Wen, Baosheng Yu, and Dacheng Tao. Learnable skeleton-aware 3D point cloud sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17671–17681, 2023. 2, 4, 6, 8

[45] Shihao Wu, , Hui Huang, Minglun Gong, Matthias Zwicker, and Daniel Cohen-Or. Deep points consolidation. *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia)*, 34(6): 176:1–176:13, 2015. 2, 7

[46] Zhirong Wu, Shuran Song, Aditya Khosla, Xiaoou Tang, and Jianxiong Xiao. 3D ShapeNets for 2.5D object recognition and next-best-view prediction. *CoRR*, abs/1406.5670, 2014. 6