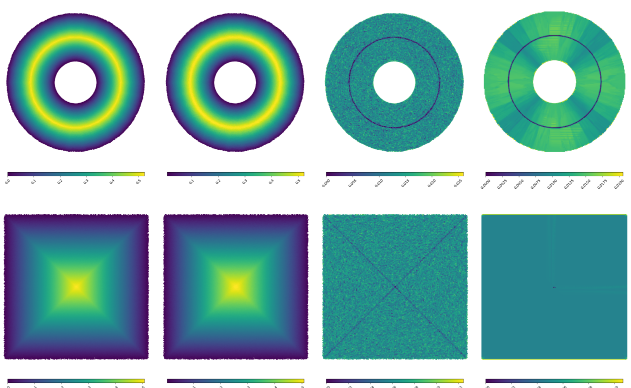
# MorphoSkel3D: Morphological Skeletonization of 3D Point Clouds for Informed Sampling in Object Classification and Retrieval

Supplementary Material

## 1. Rationale of MS3D

The distance function from a set of interior points to the surface is shown in Fig. 1a, to illustrate how the distance increases as it approaches the geometric center. Using a dilation operator that takes the maximum value of the distance function within the neighborhood defined by a structuring element, we demonstrate in Fig. 1b that the distance function is expanded to larger values. Except, for the set of maximal balls, the distance remains unchanged from the dilation in Fig. 1c to reveal the skeleton. An alternative to using random inner points is to form a 3D meshgrid that enforces a stricter selection of skeletal points as illustrated in Fig. 1d.

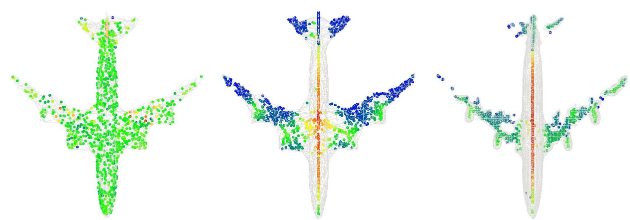


(a) $\mathrm{UDF}(x)$ (b) $\delta_{\mathrm{SE}}(\mathrm{UDF}(x))$ (c) $\mathrm{MS3D}_{\mathrm{rand}}$ (d) $\mathrm{MS3D}_{\mathrm{grid}}$

Figure 1. MorphoSkel3D for torus and box with $k$ set to 20. First three columns represent the process for random points, while last column shows the outcome of inner points formed by a meshgrid.

## 2. Traditional Surface Reconstruction

Given that MorphoSkel3D relies on surface reconstruction within its pipeline, we present a benchmark of different algorithms to evaluate the influence on the skeletal quality. Notably, reconstruction is challenging for the ShapeNet subset with 2000 surface points. In Fig. 2a, the ball pivoting algorithm (BPA) fails to correctly reconstruct the mesh as the inner points remain on the surface. In contrast, Figs. 2b and 2c illustrate how alpha shapes and Poisson both create a closed surface to identify inner points for MS3D to skeletonize. In Tab. 1, the reconstruction results tend to show lower Chamfer distances for Alpha and Poisson, and lower Hausdorff distances for P2S and BPA.



(a) BPA (b) Alpha (c) Poisson

Figure 2. The skeletal spheres that are produced from the mesh by three different traditional surface reconstruction algorithms.

Therefore, given our uncertainty about whether these metrics fully capture skeletal quality, we also assessed in the experiments how well the skeleton can guide sampling. In Tab. 2, the classification performance of the learning-to-sample [**?** ] method is reported for different skeletons.

| Ratio | DPC [**?** ] | SA [**?** ] | MS3D |
|---|---|---|---|
| 8 | 89.1 | 89.1 | **89.5** |
| 16 | 88.8 | 88.8 | **88.9** |
| 32 | 87.5 | 87.4 | **87.8** |
| 64 | 82.8 | 82.9 | **85.5** |

Table 2. Object classification results on ModelNet40, OA (%).

| | CD-Recon | | | | HD-Recon | | | | CD-MAT | | | | HD-MAT | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | P2S | BPA | Alpha | Poisson | P2S | BPA | Alpha | Poisson | P2S | BPA | Alpha | Poisson | P2S | BPA | Alpha | Poisson |
| Airplane | 0.0363 | 0.0187 | 0.0158 | **0.0152** | 0.1266 | 0.0911 | **0.0882** | 0.0993 | 0.0611 | 0.0547 | 0.0431 | **0.0368** | 0.1721 | **0.1310** | 0.1435 | 0.1598 |
| Chair | 0.0441 | 0.0362 | 0.0365 | **0.0296** | **0.1618** | 0.1700 | 0.2415 | 0.2347 | 0.0974 | 0.0906 | 0.0756 | **0.0659** | **0.2151** | 0.2241 | 0.2906 | 0.3123 |
| Table | 0.0424 | 0.0369 | **0.0328** | 0.0366 | 0.1745 | **0.1647** | 0.2208 | 0.2205 | 0.0876 | 0.0745 | **0.0651** | 0.0823 | **0.2085** | 0.2192 | 0.2813 | 0.3107 |
| Lamp | 0.0335 | 0.0233 | **0.0215** | 0.0265 | **0.1382** | **0.1382** | 0.1491 | 0.1896 | 0.0884 | 0.0639 | **0.0575** | 0.0606 | **0.2003** | 0.2095 | 0.2230 | 0.2595 |
| Guitar | 0.0179 | 0.0140 | 0.0089 | **0.0085** | 0.0625 | **0.0486** | 0.0490 | 0.0553 | 0.0536 | 0.0351 | **0.0238** | 0.0259 | 0.1216 | **0.0864** | 0.0992 | 0.1046 |
| Earphone | 0.0399 | 0.0231 | **0.0166** | 0.0245 | **0.1125** | 0.1502 | 0.1859 | 0.1686 | 0.1638 | 0.1015 | 0.1014 | **0.0934** | **0.2130** | 0.2205 | 0.2712 | 0.2815 |
| Mug | 0.0417 | **0.0402** | 0.0434 | 0.0580 | 0.1419 | 0.1439 | **0.1122** | 0.3201 | 0.1179 | 0.1171 | **0.1122** | 0.1126 | **0.2158** | 0.2343 | 0.3354 | 0.4121 |
| Rifle | 0.0213 | 0.0133 | 0.0107 | **0.0097** | 0.0767 | **0.0571** | 0.0602 | 0.0584 | 0.0356 | 0.0353 | 0.0263 | **0.0250** | 0.0957 | **0.0736** | 0.0927 | 0.0873 |
| Average | 0.0372 | 0.0294 | **0.0272** | 0.0274 | 0.1424 | **0.1373** | 0.1796 | 0.1857 | 0.0828 | 0.0714 | **0.0608** | 0.0629 | **0.1898** | 0.1903 | 0.2365 | 0.2601 |

Table 1. Comparison of reconstruction error, for Point2Skeleton as reference and MorphoSkel3D under different surface reconstruction modules, to the surface point cloud (Recon) and the ground truth skeleton (MAT), Chamfer (CD) and Hausdorff (HD) distances.

| Method | Ratio | Mean | <u>Air</u> | Bag | Cap | <u>Car</u> | <u>Cha</u> | Ear | <u>Gui</u> | Kni | <u>Lam</u> | Lap | Mot | Mug | Pis | Roc | Ska | <u>Tab</u> |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 83.7 | 83.4 | 78.7 | 82.5 | 74.9 | 89.6 | 73.0 | 91.5 | 85.9 | 80.8 | 95.3 | 65.2 | 93.0 | 81.2 | 57.9 | 72.8 | 80.6 |
| FPS [? ] | 16 | **80.1** | **80.1** | 67.8 | 66.0 | **65.7** | **88.0** | 64.8 | **88.3** | 79.9 | 76.4 | **95.4** | 51.1 | **89.6** | **78.2** | 45.6 | **74.4** | 77.3 |
| | 32 | **79.6** | **78.9** | **77.5** | **74.3** | 63.1 | **87.8** | 58.5 | 88.5 | **82.7** | 76.3 | **95.3** | **60.6** | **85.4** | **80.6** | **48.3** | **70.6** | 75.8 |
| | 64 | **77.3** | **75.3** | **69.6** | **57.6** | **56.2** | **85.8** | 60.2 | 86.8 | **80.8** | 72.5 | **94.6** | **60.3** | 77.1 | **81.1** | **50.9** | **67.6** | 75.1 |
| | 128 | 70.1 | **73.3** | **74.8** | **50.8** | 46.4 | **77.3** | 46.5 | **89.1** | **82.9** | 72.9 | **94.4** | **47.3** | **86.5** | 71.4 | **48.0** | 57.8 | 61.2 |
| MS3D | 16 | 79.4 | 76.8 | **74.2** | **70.2** | 63.9 | 85.3 | **66.4** | 87.9 | **82.4** | **77.2** | 94.2 | **57.4** | 82.2 | 77.3 | **52.6** | 65.8 | **78.7** |
| | 32 | 77.5 | 73.0 | 74.5 | 47.5 | **64.8** | 82.5 | **59.7** | **88.6** | 80.9 | **77.9** | 95.1 | 52.7 | 74.8 | 73.4 | 45.0 | 63.8 | **76.6** |
| | 64 | 76.9 | 72.3 | 69.5 | 48.8 | 55.8 | 84.7 | **66.2** | **87.8** | 78.2 | **76.3** | 92.4 | 49.0 | **81.3** | 80.4 | 48.1 | 65.5 | **75.5** |
| | 128 | **73.3** | 68.0 | 70.8 | 40.7 | **53.1** | 76.9 | **65.4** | 87.8 | 80.9 | **75.5** | 93.4 | 40.9 | 83.3 | **77.5** | 47.4 | **65.6** | **72.6** |

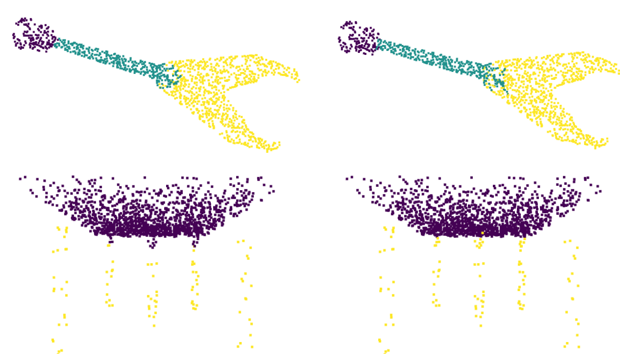
Table 3. Weakly supervised part segmentation results on different sampling ratios, mIoU (%).

## 3. Weakly Supervised Part Segmentation

**Dataset** MorphoSkel3D demonstrated effective point sampling across intricate regions and, therefore, we transfer the learned sampling networks initially trained for classification to perform weakly supervised segmentation on ShapeNet in this section. The ShapeNet dataset includes 16 object categories with a total of 50 classes to train a single model for coarse-level segmentation, six of these categories also represented in ModelNet. In specific, the ShapeNet part segmentation dataset contains 2048 points per object where each is labeled with an annotated ground truth. To enhance segmentation performance with prior knowledge, we leverage the pre-trained sampling network from ModelNet to the downstream task of ShapeNet. The sampling network of ModelNet serves as a backbone to train a segmentation model on ShapeNet. It's important to note that a sampled subset is thus provided to learn segmentation, making the segmentation task weakly supervised since not all points and labels are used. The idea is that a pre-trained sampling network of ModelNet selects a representative subset of points and could be transferred to ShapeNet with the same goal.

**Metric** The evaluation scheme for part segmentation aligns with state-of-the-art, where the intersection over union (IoU) for a shape is derived from the average IoUs of its parts. For each category, the IoU is in turn calculated as the average of all shape IoUs within the category. Finally, the overall instance average mIoU is determined by averaging the IoUs of all instances in the test set. For a fair comparison, we employ no data augmentation techniques to follow the standard setting for part segmentation. Therefore, the fully supervised segmentation results in the upper part of Tab. 3 are identical to those reported in PointNet [? ]. With an instance-averaged mIoU of 83.7%, an upper bound is established to benchmark weakly supervised methods. In our goal to reduce the annotation effort and effectively learn from a limited set of partially labeled points, we compare the sampling of the classic FPS with MS3D. Both methods are evaluated across the four sampling ratios to reduce the original set into subsets of 128, 64, 32, and 16 points.

**Results** In Tab. 3, the performance of FPS is compared against MS3D across 16 categories, with six of these categories underlined as they are also found in ModelNet. To provide more insight in the effectiveness of MS3D, we should focus on, but not limit the analysis to the six categories that occur in ModelNet. At the highest sampling ratio of 128, MS3D indicates an improvement in overall mIoU that surpasses FPS by over 3% with 73.3% compared to 70.1%. For instance, the gap in the table category is evident based on the sampling strategy used by MS3D. It namely targets the corner areas where there's a transition in label between the surface and legs of a table. A segmented example of the table and guitar category is shown in Fig. 3 to compare prediction and ground truth. For the three other sampling ratios, the overall mIoU slightly differs in the favor of FPS. We illustrated that our method focused its sampling in learned regions for classification and, therefore, assume that it fails to cover the entire object as effectively as FPS to learn segmentation. On the other hand, MS3D distributes its few sampled points in different parts. This observation suggests that our proposed method efficiently identifies points in each part to annotate and learn a segmentation. Consequently, the study of objects in a fine-grained segmentation setting arises as an interesting task.



(a) Prediction

(b) Ground truth

Figure 3. Segmentation model trained with 16 labeled points, sampled through MorphoSkel3D. The prediction for a guitar and table is displayed on the left, while the ground truth is on the right.