Revisiting Point Cloud Simplification: Supplementary Material

Rolandos Alexandros Potamias, Giorgos Bouritsas, and Stefanos Zafeiriou

Imperial College London, London, UK
{r.potamias, g.bouritsas, s.zafeiriou}@imperial.ac.uk

1 Implementation Details

We implemented the projector network using three multi-layer perceptrons (MLP) followed by Batch Normalization [1] and ReLU activation functions [4]. The filter sizes were set to 64. The GNN following the stacked MLPs was also ReLU activated with a filter size of 64. Following [5, 6], we use a relatively small neighborhood size of k=7 nearest neighbors as the input graph connectivity. The initial point of FPS is randomly selected since we did not observe any influence on the performance. We selected 15 neighbours for each cluster center selected by FPS, as suggested in . The filter size of the attention-based refinement layer was set to 3, mapping the (64+3) features of the selected points to (x, y, z) coordinates. We trained our model for 150 epochs with learning rate of 0.001 and a weight decay of 0.99 on every epoch using the Adam optimizer [2].

2 Experiments

2.1 Additional Simplification ratios

During this section we report quantitative results for simplification ratios bellow 0.3. Table 1 includes the simplification performance of the proposed and the baseline methods on TOSCA, ModelNet and MeIn3D datasets. Note that the method of [3] run out of time when attempted to simplify meshes for large simplification ratios (over 0.4). Importantly, the proposed method outperforms almost all baselines under the perceptual metrics (NC, RE, SDM) and exhibits comparable CD measures with FPS method.

It is also important to note that the performance of the proposed method degrades linearly as the simplification ratio increases, compared to baseline methods such as TCP and Qi *et al.* [8]. Additionally, the proposed method achieves to be the best or the second best performing method under all measures.

Figure 1 shows simplified point clouds at different scales, comparing QEM and the proposed method, where the point clouds are visualized on top of the mesh surfaces to highlight the salient regions. The proposed method favours point selection at the horse's nape and face in contrast to points at smooth areas, such as the thigh, to preserve salient features of the input point cloud. Given that highly detailed and sharp 3D regions require many planes to be

	TOSCA											
	$N_s/N_{org}=0.8$				$N_s/N_{org}=0.5$				$N_s/N_{org}=0.3$			
Method	CD	NC	RE($\times 10^{-4}$)	SDM($\times 10^{-3}$)	CD	NC	RE($\times 10^{-4}$)	SDM($\times 10^{-3}$)	CD	NC	RE($\times 10^{-4}$)	SDM($\times 10^{-3}$)
Random	0.14	**0.093**	2.87	2.43	0.49	**0.106**	3.04	3.03	1.04	**0.225**	3.65	4.33
TCP	1.11	0.147	2.86	2.61	10.0	0.272	3.36	4.05	30.8	0.357	4.19	6.57
FPS	0.09	0.103	**2.85**	2.32	0.29	0.245	**2.97**	2.92	0.67	0.255	**3.52**	4.22
QEM	0.09	0.103	**2.81**	2.33	0.29	0.214	**2.96**	2.91	0.84	**0.248**	3.54	4.27
Liu *et al.* [3]	-	-	-	-	-	-	-	-	1.56	0.384	3.86	4.52
Qi *et al.* [8]	0.10	0.104	2.87	2.47	0.54	0.209	2.98	3.54	1.71	0.253	3.58	5.32
Yan *et al.* [10]	0.28	0.103	2.58	2.51	0.42	0.208	2.98	2.90	0.71	0.250	3.57	4.20
Proposed-MeIn3D	0.05	0.104	2.88	2.30	0.25	0.244	3.06	2.87	0.65	0.255	3.54	**4.07**
Proposed-ModelNet	**0.03**	0.103	2.87	**2.29**	**0.23**	0.211	3.05	**2.83**	**0.64**	0.259	**3.51**	4.08
Proposed-TOSCA	**0.03**	**0.102**	2.86	**2.21**	**0.23**	**0.193**	3.03	**2.79**	**0.63**	0.254	3.55	**4.04**

	ModelNet											
Method	CD($\times 10^{-4}$)	NC	RE($\times 10^{-4}$)	SDM($\times 10^{-3}$)	CD($\times 10^{-4}$)	NC	RE($\times 10^{-4}$)	SDM($\times 10^{-3}$)	CD($\times 10^{-4}$)	NC	RE($\times 10^{-4}$)	SDM($\times 10^{-3}$)
Random	1.74	**0.181**	4.91	1.14	3.13	**0.201**	5.16	1.53	6.01	**0.333**	5.37	1.99
CP	14.01	0.288	5.01	1.12	55.12	0.371	5.98	1.68	117.11	0.527	6.63	2.71
FPS	**0.89**	**0.195**	**4.71**	**1.01**	**1.93**	**0.213**	4.89	**1.35**	**3.02**	**0.352**	5.57	2.08
QEM	1.35	0.211	4.98	1.14	2.84	0.224	5.12	1.48	3.05	0.382	5.57	2.44
Liu *et al.* [3]	-	-	-	-	-	-	-	-				
Qi *et al.* [8]	1.37	0.210	4.97	1.04	3.31	2.31	5.14	1.54	7.04	0.357	5.31	2.17
Yan *et al.* [10]	2.71	0.209	4.31	1.05	3.37	0.235	5.12	1.43	4.64	0.346	5.28	2.11
Proposed-MeIn3D	2.32	0.353	5.12	1.11	2.81	0.365	5.23	1.50	3.72	0.473	5.53	2.15
Proposed-ModelNet	**0.91**	0.207	**4.61**	**0.99**	**1.12**	0.216	**4.72**	**1.28**	**2.74**	0.371	**5.01**	**1.87**
Proposed TOSCA	2.12	0.270	4.82	1.07	2.98	0.283	**4.86**	1.42	4.11	0.401	**5.26**	**2.03**

	MeIn3D											
Method	CD($\times 10^{-4}$)	NC	RE($\times 10^{-4}$)	SDM($\times 10^{-3}$)	CD($\times 10^{-4}$)	NC	RE($\times 10^{-4}$)	SDM($\times 10^{-3}$)	CD($\times 10^{-4}$)	NC	RE($\times 10^{-4}$)	SDM($\times 10^{-3}$)
Random	0.74	0.108	2.46	0.99	1.12	0.120	2.74	1.24	1.26	0.169	3.43	2.31
TCP	12.35	0.211	2.41	0.97	43.06	0.327	**2.52**	1.54	89.24	0.571	**2.89**	2.91
FPS	**0.59**	**0.103**	**2.32**	**0.86**	**0.97**	**0.105**	2.53	**1.15**	**1.05**	**0.108**	3.21	2.28
QEM	0.94	0.112	2.52	1.06	1.36	0.139	2.76	1.44	1.94	0.150	3.53	2.54
Liu *et al.* [3]	-	-	-	-	-	-	-	-	1.99	0.159	3.91	2.83
Qi *et al.* [8]	1.22	0.110	2.77	4.14	1.97	0.131	2.77	1.40	2.15	0.144	3.55	2.44
Yan *et al.* [10]	1.34	0.178	2.45	1.23	1.84	0.129	2.74	1.31	2.07	0.127	3.21	2.19
Proposed MeIn3D	**0.61**	**0.104**	**2.29**	**0.90**	**0.98**	**0.105**	**2.46**	**1.07**	**1.15**	**0.105**	**2.89**	**1.76**
Proposed ModelNet	1.15	0.111	2.41	1.02	1.28	0.123	2.68	1.43	1.59	0.165	2.99	2.09
Proposed TOSCA	1.04	0.106	2.41	1.08	1.21	0.1171	2.63	1.33	1.41	0.149	**2.96**	**1.85**

Table 1. Simplification performance tested on TOSCA, ModelNet and MeIn3D datasets. Best approaches highlighted are highlighted in **bold** and second best in **red**. We refer to the dataset used for training as "Proposed-*Dataset*"

accurately described, the proposed method can be directly used to simplify large-scale point cloud scans by retaining only the salient points of the input. This enforces applications such as meshing that require many points around highly detailed regions to better describe object characteristics.

2.2 Ablation Studies

Loss function: As mentioned in Section 3.3 of the main manuscript, an important component of the proposed simplification framework is engineered of the curvature guided loss function. In particular, Chamfer Distance (CD) assigns an equal importance weight to each point set, neglecting important points of the point cloud. Thus, semantically meaningful points will be assigned with the same penalty as with points at flat smooth areas. In such way, CD will drive the model to generate smooth results that minimize shape reconstruction without taking into account critical identity details of the object. To break this uniformity, we modified the first term of the CD to assign a different weight to each point according to its curvature. In Table 2, we report the performance of the proposed method trained only with regular CD (Proposed-CD), with adaptive CD (Proposed-ACD), and with both adaptive CD and curvature preservation loss (Proposed-Full). Results reveal that the modified CD exhibits lower percep-

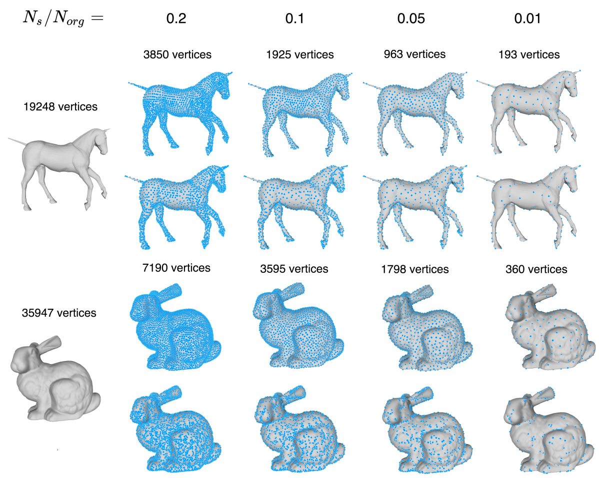

Fig. 1. Qualitative comparison between QEM (top row) and the proposed (bottom row) methods, at different simplification ratios. Point clouds are rendered on top of the original mesh surfaces to better visualize high-curvature areas.

tual error (CE, RE, SDM) compared to simple CD, while adding a curvature preserving loss (Proposed-Full) further boosts the performance of the model.

Model architecture: Additionally, we examined the importance of the GNN-based point projector and the attention refinement network in the proposed architecture. As shown in Table 2, a performance similar to the FPS method is observed when the proposed method is trained without the GNN module (Proposed w/o GNN), with a significant increase taking place in perceptual error measures (RE, SDM). This certifies the importance of the GNN-based point projector that, in contrast to linear layers, enables message passing through neighboring points. Finally, an increase in SMD error is also observed to the model trained without the attention refinement module. As expected the use of attention refinement module further improves perceptual preservation since it weights points within the same cluster and moves the cluster centers closer to salient regions to minimize the curvature error.

2.3 Classification of simplified point clouds

To further assess the simplification quality of the simplified point clouds, we used a pretrained shape classification model and measured its classification accuracy on the simplified point clouds. In such setting we can assess the preservation of high level semantics using an external objective judge such as a neural network. In particular, we trained a PointNet [7] model on the train split of TOSCA dataset. We used the proposed and the baseline methods to simplify

	N_s/N_{org} = 0.2				N_s/N_{org} =0.1				N_s/N_{org} =0.05			
Method	CD	CE	RE($\times 10^{-4}$)	SDM($\times 10^{-4}$)	CD	CE	RE($\times 10^{-4}$)	SDM($\times 10^{-4}$)	CD	CE	RE($\times 10^{-4}$)	SDM($\times 10^{-4}$)
Proposed-CD	1.12	0.40	3.96	5.52	2.41	0.47	4.43	9.64	4.91	0.58	4.99	18.3
Proposed-ACD	1.15	0.39	4.01	5.51	2.54	0.46	4.42	9.61	4.97	0.56	4.96	17.9
Proposed-w/o GNN	**0.86**	0.32	4.67	5.11	**2.15**	0.35	4.76	9.12	**4.15**	0.36	5.32	18.1
Proposed-w/o AttRef	0.95	0.31	4.21	5.05	2.29	0.31	4.52	8.54	4.51	0.34	5.16	17.3
Proposed-Full	1.12	**0.29**	**3.91**	**5.01**	2.45	**0.30**	**4.41**	**7.84**	4.93	**0.33**	**4.93**	**16.5**

Table 2. Ablation study on loss function and model architecture. Proposed-CD denotes the model trained with CD, Proposed-ACD denotes model trained with adaptive CD and Proposed-Full denotes the model trained with the loss functions introduced in Section 3.3. Proposed-w/o GNN refers to the model trained without the GNN layer in point projector module and Proposed-w/o AttRef refers to the model trained without the attention refinement module.

the remaining test split. In Figure 2 we show the classification performance, in terms of accuracy, of the compared methods at different simplification ratios. It is important to note that the scope of this experiment is to demonstrate that the simplified point clouds produced by the proposed method can be better identified, by a pretrained classifier, compared to the ones produced by the baselines. Boosting the point cloud classification performance remains out of the scope of this paper since category indicative points are not always correlated with the visual appearance of the model. We report results of the original test set performance at simplification ratio equal to 1. It can be easily seen that the proposed model degrades with a smaller slope at extreme simplification ratios, compared to the baseline models. This strengthen our argument that the proposed method retains the salient features that characterise each object and indicate its identity. We argue that the performance drop of the baseline models could be attributed to the uniform way of sampling points that may drive to decimation of salient points that characterize the point cloud. In contrast, the perceptually influenced simplification of the proposed method selects points according to their visual importance, ensuring that the salient features will be decimated last.

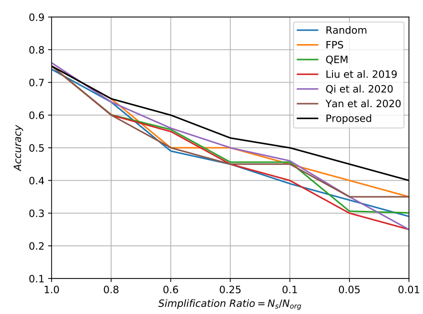

Fig. 2. Classification accuracy of the pretrained PointNet++ on simplified point clouds at different ratios.

2.4 Simplification under noisy conditions

As discussed in Section 5.3 of the main paper, several raw scans contain noise. Thus, a point simplification module should be not be extremely affected by noise. As can be observed in Figure 3, the proposed method preserves most of the structural characteristics of the input, without being affected from the outlier noisy points as much as the baseline methods. In contrast, sampling points directly from the xyz-space using the FPS method produces noisy outputs following the noisy patterns of the inputs. Similarly, QEM selects noisy points in order to minimize the quadric error of the input planes, such as the outliers in cat's foot and human hand (shown in zoomed areas).

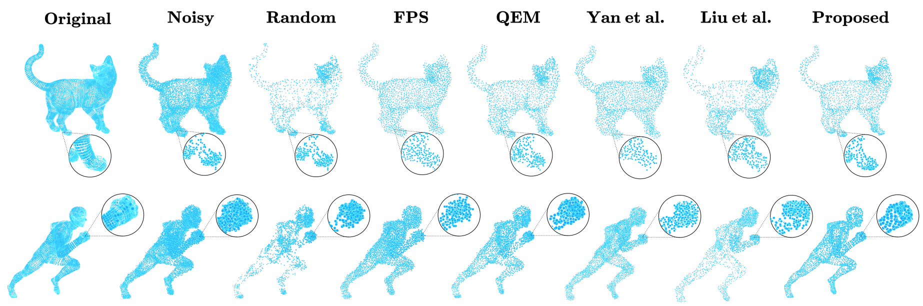

Fig. 3. Qualitative comparison between the baseline and the proposed methods for point clouds with Gaussian noise addition.

2.5 Simplification of Real-World Point Clouds

A significant application of point cloud simplification methods is to sub-sample points of real-world scanners that generate million of points from the representative surface. To test the performance of the proposed method on such scenario, we utilized Torronto3D dataset [9], that contain outdoor point clouds acquired with LIDAR sensors. Again, we utilized the pretrained model on TOSCA dataset, without further training or tuning. Quantitative results summarized in Table 3 demonstrate that the proposed method outperforms baseline methods in perceptual quality measures (CE, RE, SDM).

Torronto3D															
	$N_s/N_{org} = 0.2$					$N_s/N_{org} = 0.1$					$N_s/N_{org} = 0.05$				
Method	CD	NC	CE($\times 10^{-2}$)	RE($\times 10^{-4}$)	SDM($\times 10^{-4}$)	CD	NC	CE($\times 10^{-2}$)	RE($\times 10^{-4}$)	SDM($\times 10^{-4}$)	CD	NC	CE($\times 10^{-2}$)	RE($\times 10^{-4}$)	SDM($\times 10^{-4}$)
Random	0.31	0.577	8.15	11.27	0.47	0.62	0.634	8.89	11.56	0.48	1.27	0.679	9.15	11.91	0.48
TCP	5.90	0.894	15.64	14.47	0.58	7.40	0.912	12.41	12.95	0.53	9.58	0.912	13.02	12.61	0.53
FPS	**0.17**	**0.509**	6.42	11.22	0.46	**0.34**	**0.565**	7.01	11.32	0.46	**0.70**	**0.619**	7.51	11.37	0.47
Proposed	0.18	0.512	**5.67**	**11.02**	**0.34**	0.37	0.595	**6.35**	**11.10**	**0.38**	0.75	0.644	**6.88**	**11.15**	**0.41**

Table 3. Simplification performance tested on outdoor point cloud from Torronto3D dataset. Best approaches highlighted are highlighted in **bold**. The proposed method model is trained with TOSCA dataset.

Although FPS achieves the lower CD and NC errors and produces smooth results that minimize the overall shape loss, it fails to preserve essential details of the object. Figure 4 shows examples of the simplified lidar point clouds at different simplification ratios generated by the proposed method.

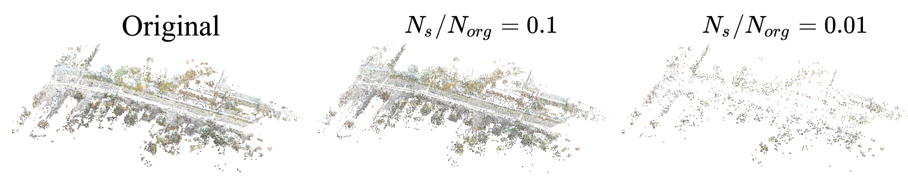

Fig. 4. Simplification of real-world scans using the proposed method. Figure better viewed in zoom.

References

1. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Bach, F., Blei, D. (eds.) Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 37, pp. 448–456. PMLR, Lille, France (07–09 Jul 2015)
2. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
3. Liu, H.T.D., Jacobson, A., Ovsjanikov, M.: Spectral coarsening of geometric operators. ACM Trans. Graph. **38**(4) (Jul 2019)
4. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Icml (2010)
5. Potamias, R.A., Neofytou, A., Bintsi, K.M., Zafeiriou, S.: Graphwalks: Efficient shape agnostic geodesic shortest path estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2968–2977 (2022)
6. Potamias, R.A., Ploumpis, S., Zafeiriou, S.: Neural mesh simplification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18583–18592 (2022)
7. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 652–660 (2017)
8. Qi, J., Hu, W., Guo, Z.: Feature preserving and uniformity-controllable point cloud simplification on graph. In: 2019 IEEE International Conference on Multimedia and Expo (ICME). pp. 284–289. IEEE (2019)
9. Tan, W., Qin, N., Ma, L., Li, Y., Du, J., Cai, G., Yang, K., Li, J.: Toronto-3d: A large-scale mobile lidar dataset for semantic segmentation of urban roadways. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 202–203 (2020)
10. Yan, X., Zheng, C., Li, Z., Wang, S., Cui, S.: Pointasnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5589–5598 (2020)