

Supplementary Material for Minimal Neural Atlas: Parameterizing Complex Surfaces with Minimal Charts and Distortion

Weng Fei Low[✉] and Gim Hee Lee[✉]

Institute of Data Science (IDS), National University of Singapore
NUS Graduate School's Integrative Sciences and Engineering Programme (ISEP)
Department of Computer Science, National University of Singapore
{wengfei.low, gimhee.lee}@comp.nus.edu.sg
<https://github.com/low5545/minimal-neural-atlas>

In this supplementary document, we first provide details on the derivation of the *Scaled Symmetric Dirichlet Energy*, and how we improve its numerical stability in practice (Sec. A). Next, we present implementation details of the proposed surface representation and baselines used in the experiments (Sec. B). Lastly, we provide additional quantitative and qualitative results on the surface reconstruction experiment (Sec. C).

A Scaled Symmetric Dirichlet Energy

A.1 Detailed Derivation

When the surface parameterization φ_{θ_k} of each chart preserves the metric of the parametric domain up to a specific common scale of L , the two singular values of its Jacobian J_k , $\sigma_{k,1}$ and $\sigma_{k,2}$ are equal to L at every point \mathbf{u} in the parametric domain, *i.e.* :

$$\sigma_{k,1}(\mathbf{u}) = \sigma_{k,2}(\mathbf{u}) = L . \quad (15)$$

By the definition of the metric tensor g_k , its two singular values and eigenvalues are also equal to L^2 everywhere.

Consequently, it is clear that the *Symmetric Dirichlet Energy* (SDE) [12,13,10] given by:

$$\begin{aligned} & \frac{1}{\sum_{k \in \mathcal{K}} |\mathcal{W}_k|} \sum_{k \in \mathcal{K}} \sum_{\mathbf{u} \in \mathcal{W}_k} \sigma_{k,1}(\mathbf{u})^2 + \sigma_{k,2}(\mathbf{u})^2 + \frac{1}{\sigma_{k,1}(\mathbf{u})^2} + \frac{1}{\sigma_{k,2}(\mathbf{u})^2} \\ &= \frac{1}{\sum_{k \in \mathcal{K}} |\mathcal{W}_k|} \sum_{k \in \mathcal{K}} \sum_{\mathbf{u} \in \mathcal{W}_k} \text{trace}(g_k(\mathbf{u})) + \text{trace}(g_k(\mathbf{u})^{-1}) \\ &= \text{mean}_{\mathcal{W}}(\text{trace} \circ g_k) + \text{mean}_{\mathcal{W}}(\text{trace} \circ g_k^{-1}) \end{aligned} \quad (16)$$

quantifies the isometric distortion, since a global minimum value of 4 is attained if and only if both $\sigma_{k,1}$ and $\sigma_{k,2}$ equal to 1 everywhere. The former and latter terms of the SDE correspond to the *Dirichlet* energy of φ_{θ_k} and $\varphi_{\theta_k}^{-1}$, respectively.

The proposed *Scaled Symmetric Dirichlet Energy* (SSDE) generalizes the SDE to quantify metric distortion up to a specific common scale of L . This is simply done by incorporating a scale factor of L as follows:

$$\begin{aligned}
& \frac{1}{\sum_{k \in \mathcal{K}} |\mathcal{W}_k|} \sum_{k \in \mathcal{K}} \sum_{\mathbf{u} \in \mathcal{W}_k} \frac{\sigma_{k,1}(\mathbf{u})^2}{L^2} + \frac{\sigma_{k,2}(\mathbf{u})^2}{L^2} + \frac{L^2}{\sigma_{k,1}(\mathbf{u})^2} + \frac{L^2}{\sigma_{k,2}(\mathbf{u})^2} \\
&= \frac{1}{\sum_{k \in \mathcal{K}} |\mathcal{W}_k|} \sum_{k \in \mathcal{K}} \sum_{\mathbf{u} \in \mathcal{W}_k} \frac{1}{L^2} \text{trace}(g_k(\mathbf{u})) + L^2 \text{trace}(g_k(\mathbf{u})^{-1}) \quad (17) \\
&= \frac{1}{L^2} \text{mean}_{\mathcal{W}}(\text{trace} \circ g_k) + L^2 \text{mean}_{\mathcal{W}}(\text{trace} \circ g_k^{-1}),
\end{aligned}$$

such that a global minimum value of 4 is attained if and only if both $\sigma_{k,1}$ and $\sigma_{k,2}$ equal to L everywhere.

Furthermore, the SSDE can be employed to quantify metric distortion up to an arbitrary common scale by finding an optimal scale L^* that minimizes it. Since SSDE is a convex function of L , L^* is simply given by the critical point:

$$\left. \frac{\partial \text{SSDE}}{\partial L} \right|_{L=L^*} = 0, \quad (18)$$

which evaluates to:

$$L^{*2} = \sqrt{\frac{\text{mean}_{\mathcal{W}}(\text{trace} \circ g_k)}{\text{mean}_{\mathcal{W}}(\text{trace} \circ g_k^{-1})}}. \quad (19)$$

By substituting Eq. 19 into Eq. 17 with $L = L^*$, we can simplify the SSDE at the optimal scale L^* as:

$$2\sqrt{\text{mean}_{\mathcal{W}}(\text{trace} \circ g_k) \text{mean}_{\mathcal{W}}(\text{trace} \circ g_k^{-1})}. \quad (20)$$

A.2 Improving Numerical Stability

In general, the SSDE at the optimal scale L^* is numerically stable since it is given by the geometric mean of the Dirichlet energies of φ_{θ_k} and $\varphi_{\theta_k}^{-1}$, which are roughly inversely proportional to each other. Although it is rare in practice, the existence of a singular g_k leads to instability in the Dirichlet energy of $\varphi_{\theta_k}^{-1}$, and hence the SSDE at the optimal scale L^* as well as the SSDE in general.

To improve numerical stability of the SSDE under such a scenario, we augment it with a small positive ϵ value as follows:

$$\begin{aligned}
& \frac{1}{\sum_{k \in \mathcal{K}} |\mathcal{W}_k|} \sum_{k \in \mathcal{K}} \sum_{\mathbf{u} \in \mathcal{W}_k} \frac{\sigma_{k,1}(\mathbf{u})^2 + \epsilon}{L^2 + \epsilon} + \frac{\sigma_{k,2}(\mathbf{u})^2 + \epsilon}{L^2 + \epsilon} + \frac{L^2 + \epsilon}{\sigma_{k,1}(\mathbf{u})^2 + \epsilon} + \frac{L^2 + \epsilon}{\sigma_{k,2}(\mathbf{u})^2 + \epsilon} \\
&= \frac{1}{\sum_{k \in \mathcal{K}} |\mathcal{W}_k|} \sum_{k \in \mathcal{K}} \sum_{\mathbf{u} \in \mathcal{W}_k} \frac{1}{L^2 + \epsilon} \text{trace}(g_k(\mathbf{u}) + \epsilon I) + (L^2 + \epsilon) \text{trace}((g_k(\mathbf{u}) + \epsilon I)^{-1}) \\
&= \frac{1}{L^2 + \epsilon} \text{mean}_{\mathcal{W}}(\text{trace} \circ \hat{g}_k) + (L^2 + \epsilon) \text{mean}_{\mathcal{W}}(\text{trace} \circ \hat{g}_k^{-1}), \quad (21)
\end{aligned}$$

where:

$$\hat{g}_k(\mathbf{u}) = g_k(\mathbf{u}) + \epsilon I \quad (22)$$

is the ϵ -conditioned metric tensor with singular values and eigenvalues $\sigma_{k,i}(\mathbf{u})^2 + \epsilon$. This introduces a lower and upper bound on the ϵ -conditioned Dirichlet energy of φ_{θ_k} and $\varphi_{\theta_k}^{-1}$, respectively, given by:

$$\text{mean}_{\mathcal{W}}(\text{trace} \circ \hat{g}_k) \geq 2\epsilon, \quad \text{mean}_{\mathcal{W}}(\text{trace} \circ \hat{g}_k^{-1}) \leq \frac{2}{\epsilon}. \quad (23)$$

The ϵ -conditioned SSDE at L preserves the global minimum value of 4 when both $\sigma_{k,1}$ and $\sigma_{k,2}$ equal to L everywhere. Following the exact same steps in Sec. A.1, it can also be shown that the ϵ -conditioned optimal scale \hat{L}^* is given by:

$$\hat{L}^{*2} = \sqrt{\frac{\text{mean}_{\mathcal{W}}(\text{trace} \circ \hat{g}_k)}{\text{mean}_{\mathcal{W}}(\text{trace} \circ \hat{g}_k^{-1})}} - \epsilon \geq 0 \quad (24)$$

and the ϵ -conditioned SSDE at \hat{L}^* is similarly given by:

$$2\sqrt{\text{mean}_{\mathcal{W}}(\text{trace} \circ \hat{g}_k) \text{mean}_{\mathcal{W}}(\text{trace} \circ \hat{g}_k^{-1})}. \quad (25)$$

In practice, we find $\epsilon = 1 \times 10^{-8}$ to be sufficient for quantifying metric distortion up to an arbitrary common scale with the ϵ -conditioned SSDE at \hat{L}^* .

B Implementation Details

B.1 Minimal Neural Atlas

Architecture. For all experiments, we employ a minimal neural atlas conditioned on an 1024-dimensional latent code $\mathbf{z} \in \mathbb{R}^{1024}$ to reconstruct the family of surfaces described in the CLOTH3D++ and ShapeNet datasets. The encoder architecture adopted for extracting the latent code of a surface depends on the task, or more precisely the form of input. For surface reconstruction where the input is a point cloud, we employ PointNet [2] with all *Batch Normalization* [6] layers removed for better training stability and convergence. On the other hand, we adopt an ImageNet[3]-pretrained ResNet-18 [5] for single-view reconstruction, where the input is an image.

In contrast to the encoder, we employ the same architecture for the conditional minimal neural atlas in all experiments. The architectures adopted for the conditional φ_{θ_k} and \tilde{l}_{θ_k} of each chart k are almost identical and are heavily based on the architecture of the IDR neural scene representation [14], which in turn is based on the DeepSDF [8] implicit neural surface representation.

Particularly, the conditional φ_{θ_k} maps the concatenated inputs of \mathbf{z} and \mathbf{u} to a 3D point on the reconstructed surface via a 4-layer Multi-Layer Perceptron (MLP), where each layer comprises 512 hidden units applied with *Weight Normalization* [11]. Each hidden layer except for the last is followed by a *Soft-Plus* activation with hyperparameter $\beta = 100$. In contrast to *ReLU*, *SoftPlus*

is infinitely-differentiable everywhere. This enables the computation of differentiable geometric properties [1] and facilitates slightly lower distortion, as observed empirically. The inputs are also concatenated with the activations of the second hidden layer to form the next hidden layer inputs. In terms of the model size, this architecture is comparable to that of AtlasNet and DSP, but more lightweight than that of TearingNet.

The conditional \tilde{l}_{θ_k} adopts the same architecture as the conditional φ_{θ_k} , except for some subtle differences. Specifically, a positional encoding with 6 octaves is applied on maximal point coordinates $\tilde{\mathbf{x}}$ before being concatenated with \mathbf{z} to form the input of the network. As shown in the ablation studies, this is important for \tilde{l}_{θ_k} to capture high frequency details for more accurate reconstructions. Moreover, ReLU and sigmoid are also used for the intermediate and output activations respectively.

Training. The training loss weights used in all experiments are given by $\lambda_{rec} = 1.0$, $\lambda_{occ} = 1.0$ and $\lambda_{dist} = 0.00001$. Apart from the encoder, note that training with equal importance on \mathcal{L}_{rec} and \mathcal{L}_{occ} works because they solely supervise φ_{θ_k} and \tilde{l}_{θ_k} , respectively. For surface reconstruction, we adopt the exact same training procedure as AtlasNet. In particular, we adopt the Adam optimizer [7] with a learning rate of 0.001 and *PyTorch*[9]-default hyperparameters. The network is trained for 150 epochs with a learning rate decay of 0.1 at 120, 140 and 145 epochs. The same training procedure is also employed for single-view reconstruction, except that we also adopt a surface reconstruction-pretrained conditional φ_{θ_k} and \tilde{l}_{θ_k} .

Inference. For the evaluation of all experiments, the label frequency c is estimated with a minimum interior rate $\eta = 40\%$. We also adopt the default occupancy probability threshold $\tau = 0.5$ to define the parametric domain of each chart. Furthermore, a reconstructed surface point cloud with approximately 25,000 points is extracted with the proposed two-step batch rejection sampling strategy using an initial UV sample size of 16,667 (*i.e.* $2/3$ of 25,000).

B.2 Baselines

To provide a fair comparison, we train all baselines on the exact same datasets using their official implementations. The AtlasNet training procedure is used for AtlasNet++ and DSP in all experiments. In contrary to all other works, we train DSP without the overlap loss since it requires access to target surface areas. Furthermore, a relatively larger epsilon value of 0.01 is added to the denominator of the deformation loss to significantly improve its numerical stability. Similar to [4], we train AtlasNet++ with Point Cloud CD, Mesh CD and SSDE at the optimal scale weighted by 1.0, 1.0 and 0.00001, respectively.

TearingNet is trained according to its two-step strategy in both experiments, which requires approximately 6 to 7 times the number of epochs compared to other works. Moreover, we omit the optional graph filter as it unnecessarily

constrains the UV sampling density during inference to that of training. This is attributed to its dependence on tearing or graph weight hyperparameters ϵ and r on the UV sampling density. For single-view reconstruction with TearingNet, we adopt the same ResNet-18 encoder and surface reconstruction-pretrained F-Net in the first step of the training. The optimal graph weight hyperparameter ϵ , which defines the parametric domain of the chart and hence the topology of the reconstructed surface, is tuned with respect to the Mesh CD and F-score @ 1% on the validation split of ShapeNet since it contains a wide range of surfaces with complex topologies. Consequently, $\epsilon = 0.025$ is used throughout the experiments for evaluation.

C Additional Results

In this section, we first present a qualitative analysis of distortion (Sec. C.1) and visualizations of SCAR violation artifacts (Sec. C.2) before looking into the occupancy rates of minimal neural atlas (Sec. C.3). Next, we investigate the effect of training with different UV sample sizes (Sec. C.4). Lastly, we examine the sensitivity of our representation on hyperparameters such as the number of positional encoding octaves in \tilde{l}_{θ_k} and minimum interior rate η (Sec. C.5). Unless otherwise stated, results involving only minimal neural atlas are obtained using 3 charts with metric distortion loss from the surface reconstruction experiment on ShapeNet.

C.1 Qualitative Distortion Analysis

Fig. 3 and Fig. 4 illustrate the distortion of the reconstructed chart parameterizations in the surface reconstruction experiment on CLOTH3D++ and ShapeNet, respectively. We employ 2 charts on CLOTH3D++ and 3 charts on ShapeNet for all surface representations with the exception of TearingNet where 1 chart is used. The relative level of distortion observed reflect the quantitative metrics previously reported, where DSP and minimal neural atlas exhibit significantly lower distortion compared to other baselines.

C.2 Artifacts of SCAR Violation

As depicted in Fig. 5, minimal neural atlas suffers from unintended holes on the reconstructed surface. We attribute such artifacts to the severe violation of the SCAR assumption, which is mainly caused by imperfect modeling of the target surface and non-matching sampling distribution between the target and maximal surface.

C.3 Occupancy Rate

Table 5 shows the mean parametric domain occupancy rates of minimal neural atlas in the surface reconstruction experiment. In general, the occupancy rates

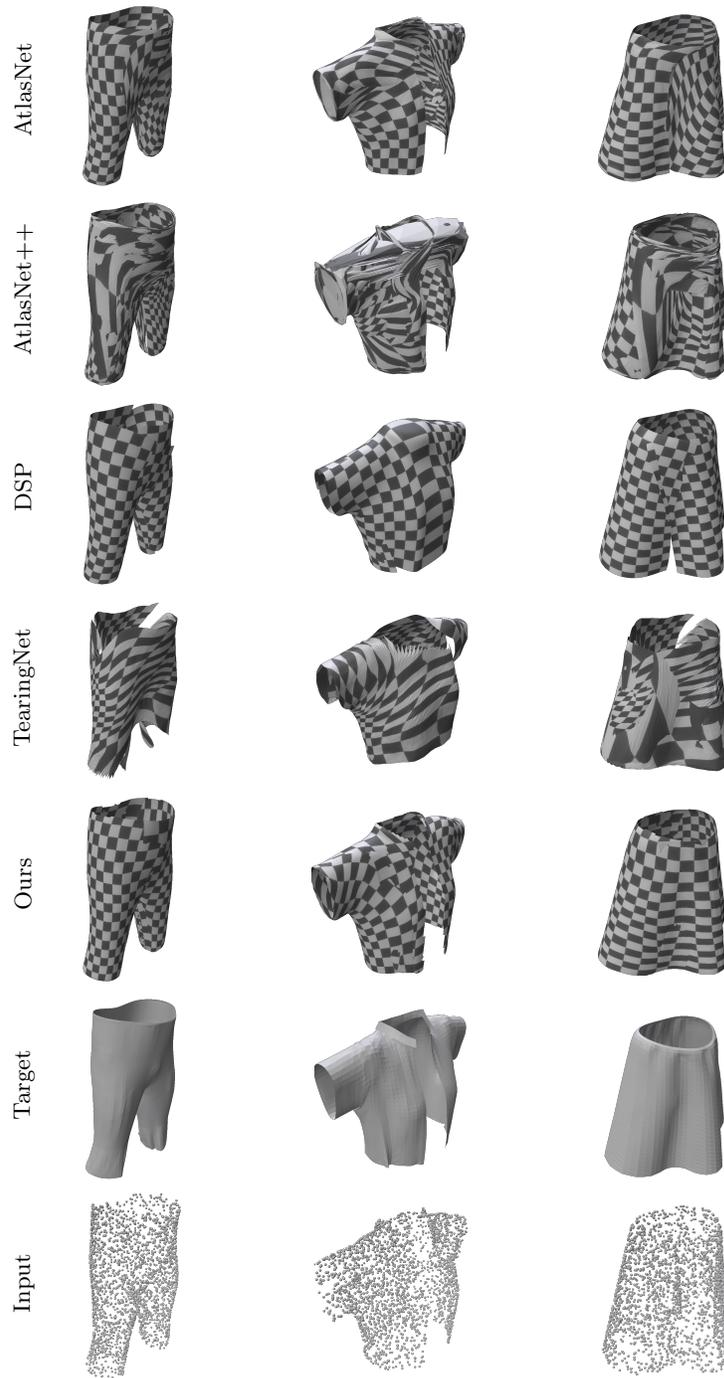


Fig. 3: Distortion of Surface Reconstructions on CLOTH3D++.

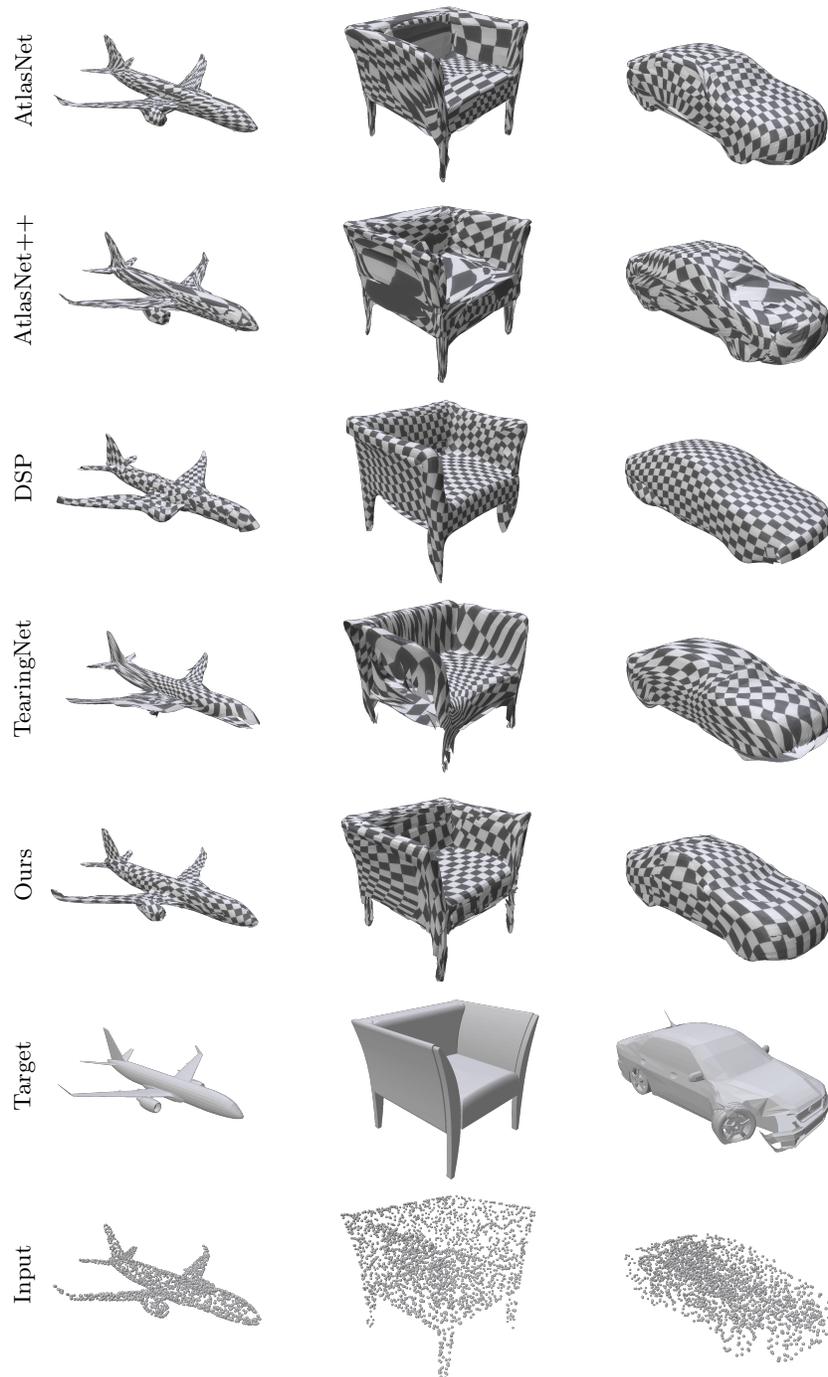


Fig. 4: Distortion of Surface Reconstructions on ShapeNet.



Fig. 5: Artifacts of SCAR Violation.

Table 5: Occupancy Rates in Surface Reconstruction Experiment.

Surface Representation	CLOTH3D++			ShapeNet		
	1 Chart	3 Charts	25 Charts	1 Chart	3 Charts	25 Charts
Ours w/o \mathcal{L}_{dist}	94.51	96.78	94.82	81.84	85.31	84.51
Ours	91.47	96.37	98.20	69.50	77.48	86.50

are fairly high, which allows for the efficient extraction of the reconstructed surface point cloud and mesh. Furthermore, it can also be observed that the occupancy rate generally increases as the number of charts increases, especially when metric distortion is regularized. This may be attributed to the increased flexibility in forming a cover of the target surface as the number of charts increases.

C.4 Ablation on Training UV Sample Size

While the UV sample size used for training is typically chosen to be the same as the target point cloud size (2,500 in our experiments), we show in Table 6 that a larger sample size favors our proposed representation since it leads to better reconstructions and lower distortions when explicitly regularized. It is also important to note that DSP is rather invariant to the increase in training UV sample size. We attribute this to the added influence of occupancy rate to the training, which is absent in other works.

Table 6: Effect of Training UV Sample Size.

Surface Representation	Training UV Sample Size	Point Cloud		Mesh		Metric Distortion ↓	Occupancy Rate ↑
		CD, 10^{-4} ↓	F@1% ↑	CD, 10^{-4} ↓	F@1% ↑		
DSP	2500	10.85	76.29	12.41	74.60	0.3044	-
	3333	10.71	76.50	12.53	74.70	0.3189	-
	5000	10.79	76.39	11.98	74.85	0.4130	-
Ours w/o \mathcal{L}_{dist}	2500	6.603	82.89	7.153	81.07	11.55	91.38
	3333	6.405	83.68	6.996	81.75	9.516	89.22
	5000	6.266	84.04	6.875	82.22	10.23	85.31
Ours	2500	6.866	82.15	7.412	80.39	2.441	87.85
	3333	6.607	82.87	7.131	81.15	2.403	83.43
	5000	6.311	83.63	6.761	82.23	2.189	77.48

Table 7: Sensitivity of the Number of Positional Encoding Octaves.

No. of Octaves	Point Cloud		Mesh		Distortion			Occupancy Rate ↑
	CD, 10^{-4} ↓	F@1% ↑	CD, 10^{-4} ↓	F@1% ↑	Metric ↓	Conformal ↓	Area ↓	
4	6.381	83.34	6.935	81.68	2.100	0.7014	0.2295	79.07
6	6.311	83.63	6.761	82.23	2.189	0.7094	0.2521	77.48
8	6.394	83.46	6.882	81.96	2.154	0.7056	0.2425	77.55

Table 8: Sensitivity of Minimum Interior Rate, η .

η	Point Cloud		Mesh		Distortion			Occupancy Rate ↑
	CD, 10^{-4} ↓	F@1% ↑	CD, 10^{-4} ↓	F@1% ↑	Metric ↓	Conformal ↓	Area ↓	
30	6.318	83.67	6.787	82.26	2.187	0.7807	0.2517	76.94
40	6.311	83.63	6.761	82.23	2.189	0.7094	0.2521	77.48
50	6.326	83.55	6.764	82.17	2.193	0.7099	0.2525	77.96

C.5 Hyperparameter Sensitivity Analysis

Number of Positional Encoding Octaves. While we have demonstrated that it is critical to apply positional encoding on the input maximal point coordinates of l_{θ_k} , Table 7 shows that the specific number of octaves adopted in the encoding does not significantly affect the overall performance of our representation.

Minimum Interior Rate. As observed in Table 8, minimal neural atlas is also not overly sensitive to the specific minimum interior rate η employed for estimating the label frequency c and hence extracting the reconstructed surface.

References

1. Bednarik, J., Parashar, S., Gundogdu, E., Salzmann, M., Fua, P.: Shape Reconstruction by Learning Differentiable Surface Representations. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2020)

2. Charles, R.Q., Su, H., Kaichun, M., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
3. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition (2009)
4. Gupta, K., Chandraker, M.: Neural Mesh Flow: 3D Manifold Mesh Generation via Diffeomorphic Flows. In: Advances in Neural Information Processing Systems (2020)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
6. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning (2015)
7. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015)
8. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
9. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems (2019)
10. Rabinovich, M., Poranne, R., Panozzo, D., Sorkine-Hornung, O.: Scalable locally injective mappings. *ACM Transactions on Graphics* (2017)
11. Salimans, T., Kingma, D.P.: Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In: Proceedings of the 30th International Conference on Neural Information Processing Systems (2016)
12. Schreiner, J., Asirvatham, A., Praun, E., Hoppe, H.: Inter-surface mapping. *ACM Transactions on Graphics* (2004)
13. Smith, J., Schaefer, S.: Bijective parameterization with free boundaries. *ACM Trans. Graph.* (2015)
14. Yariv, L., Kasten, Y., Moran, D., Galun, M., Atzmon, M., Ronen, B., Lipman, Y.: Multiview Neural Surface Reconstruction by Disentangling Geometry and Appearance. In: Advances in Neural Information Processing Systems (2020)