Fast learning of signed distance functions from noisy point clouds via noise to noise mapping

Nayoung Kwon, Pierrick Bournez

March 18, 2025

1 Introduction

In recent years, 3D scanning and reconstruction technologies have become ubiquitous in fields ranging from autonomous driving to augmented reality. These technologies commonly rely on point cloud data captured by sensors such as Lidar, which, despite their utility, are inherently noisy and unstructured. The noise present in such data can significantly hinder downstream tasks like surface reconstruction, denoising, and upsampling. As a result, developing methods that can robustly extract accurate geometric information from noisy point clouds is a critical challenge in computer vision and graphics.

One promising avenue for addressing this challenge is through the use of implicit representations, specifically Signed Distance Functions (SDFs), which offer a continuous and flexible means to represent surfaces. Traditional approaches to learning SDFs typically require clean point clouds or additional supervision, such as ground truth distances or normals, to ensure accurate reconstructions. However, acquiring such high quality data is often impractical in real world scenarios.

The paper "Fast Learning of Signed Distance Functions from Noisy Point Clouds via Noise to Noise Mapping" introduces a novel framework that circumvents the need for clean data by leveraging a noise-to-noise mapping strategy. Instead of relying on ground truth supervision, the method utilizes multiple noisy observations to statistically infer the underlying geometry. This is achieved by formulating a denoising function that pulls noisy points toward the estimated surface, and by employing Earth Mover's Distance (EMD) as a metric to measure the discrepancy between different noisy observations. Furthermore, the integration of multi-resolution hash encodings allows for a dramatic acceleration in training, achieving significant improvement over conventional methods. While they proposed multiple use-case such as Multi-view reconstruction, Point Clouds Denoising, Upsampling and surface reconstruction.

In our work, we studied the aforementioned paper in detail and sought to replicate its core ideas in a simplified setting. Although the authors propose several applications, including multi-view reconstruction, point cloud denoising, upsampling, and surface reconstruction, we will focus exclusively on surface reconstruction, as it directly aligns with the core principles of the proposed method. Due to resource constraints and for the sake of clarity in visualization and experimentation, our implementation is focused on a two-dimensional (2D) scenario rather than the full three-dimensional (3D) setup presented in the original paper. While the 2D data does not capture the full complexity of 3D point clouds, it retains the essential challenges of noise and the need for robust surface extraction. By implementing the noise-to-noise mapping approach in 2D, we aim to validate the effectiveness of the core concepts and establish a foundation for future extension to 3D data.

This report provides a detailed analysis of the original methodology, a description of our 2D implementation, experimental evaluations, and a critical discussion of the results. Our contributions are as follows:

- Reimplementation of the original paper's multi-resolution hash encoding and proposed loss function (see Section 3.1 and Section 3.2 for the implementation).

- Comparison of the proposed method with a simpler model, specifically a neural network with a sine activation function (see Section 4.2).

- Evaluation of model performance under constrained conditions, including increased noise and reduced input data (see Section 4.4).

- Application of the model to 3D samples (see Section 4.5).

Pierrick Bournez and Nayoung Kwoncontributed to the proof of concept for the different loss functions. Pierrick implemented the multi-resolution hash encoding, conducted 2D experiments (added noise and comparing with the l2 norm). Nayoung adapted the 2D pipeline to 3D by adapting the necessary experiments.

2 State of the art

2.1 Classical surface reconstruction

One of the seminal approaches to surface reconstruction from point clouds is Screened Poisson Surface Reconstruction [3]. It estimates an indicator function by solving a Poisson equation, allowing for a smooth, watertight surface. However, it typically assumes oriented normals and can be sensitive to noise if the point cloud is particularly sparse or lacks accurate normal information.

Methods based on moving least squares [8] attempt to fit local polynomial approximations around each point. While robust in many scenarios, these methods may require careful parameter tuning and can suffer from artifacts in highly noisy data or areas with uneven sampling density.

2.2 Neural implicit representations

Learning continuous functions that map spatial coordinates to occupancy or signed distance values has emerged as a powerful paradigm shown [9]. DeepSDF [9] encodes shapes as SDFs in a latent space, facilitating high quality reconstructions but typically requires clean ground truth data or shape-level supervision. SIREN [10] introduced sine activations to capture high frequency details in implicit neural representations. This approach is well suited for tasks requiring precise boundary localization (e.g., near the zero-level set of an SDF). However, vanilla SIREN alone does not inherently address the challenge of noise in the input data.

Recently, Instant-NGP [6] demonstrated that using a multi-resolution hash encoding dramatically accelerates the training of neural implicit representations. By representing spatial features at multiple scales, it allows for fast convergence and high quality reconstructions. Although highly effective, it was originally designed for radiance fields (NeRF) and requires adaptation for handling noisy point cloud data.

2.3 Learning from noisy point clouds

Some recent works attempt to learn implicit functions without direct supervision of ground truth distances or normals.For instance, SAL [1]propose self-supervised objectives that enforce certain properties of the SDF (e.g., gradient regularity). However, they often assume the input data is relatively clean or come with oriented normals.

Inspired by image-based noise-to-noise methods[4], several studies have explored how statistical reasoning can be applied to unstructured data like point clouds [2]. The core idea is that multiple noisy observations, when compared with each other, can reveal the underlying clean geometry without ever requiring explicitly clean data.

3 Methodology

This section provides a detailed description of the methodology introduced in the original paper. The approach aims to learn a signed distance function (SDF) directly from noisy point clouds, without requiring clean supervision such as ground truth distances or normals. The methodology is outlined in Section 3.1 and consists of three main components: a noise-to-noise mapping presented in Section 3.1.1,

a denoising function with its corresponding loss formulation detailed in Sections 3.1.2 and 3.1.3, and an efficient learning framework utilizing multi-resolution hash encodings, as discussed in Section 3.1.4.

Since the original paper does not provide a fully operational baseline, a complete replicated pipeline is presented in Section 3.2 to facilitate an in-depth evaluation of the method.

3.1 Overview of the original method

3.1.1 Noise-to-noise mapping

Traditional methods for learning SDFs often assume access to clean point clouds or additional supervision to guide the learning process. In contrast, the paper proposes a noise-to-noise mapping strategy. The key idea is that even when only noisy observations are available, the statistical properties of these observations can be exploited to recover the underlying clean geometry.

Instead of training the network to directly output the true signed distances, the method trains the model to map one noisy observation to another noisy observation. By using multiple noisy samples of the same scene or object, the inherent randomness in the noise tends to cancel out when the network is forced to minimize discrepancies between different observations.

This strategy is inspired by the success of Noise2Noise in image denoising [5], where it was shown that learning from pairs of noisy images can still result in a denoised output. Although point clouds lack the spatial correspondence seen in images, the authors demonstrate that a similar statistical reasoning is possible by leveraging a distance metric that establishes correspondences implicitly.

3.1.2 Denoising function and SDF learning

At the heart of the method is a neural network designed to learn an implicit representation of the surface in the form of a signed distance function f_θ. The learned SDF should be such that its zero-level set represents the desired clean surface.

- **SDF Prediction:** The network takes a query point q as input and outputs a scalar $f_\theta(q)$. This scalar is interpreted as the signed distance from q to the surface, with the sign indicating whether q lies inside or outside the object.

- **Denoising function:** To leverage the SDF for denoising, the paper defines a function F that "pulls" noisy points towards the surface. Given a point q and its corresponding SDF value, the denoising function is expressed as: $F(q, f_\theta) = q - \frac{f_\theta(q)}{\|\nabla f_\theta(q)\|} \nabla f_\theta(q)$ This formulation uses both the predicted signed distance and its gradient. The term $\nabla f_\theta(q)$ gives the direction of steepest change of the distance, and scaling it by $\frac{f_\theta(q)}{\|\nabla f_\theta(q)\|}$ computes the minimal displacement required to move the point q onto the estimated surface.

3.1.3 Loss Function

A carefully designed loss function is essential for training the network in the absence of clean data. The paper introduces a composite loss that encourages both the accuracy of the denoised point clouds and the geometric consistency of the learned SDF.

- **Distance Metric : Earth Mover's Distance (EMD)**
  To compare the denoised point cloud $F(S, f_\theta)$ (obtained by applying F to the noisy input S) with another noisy observation S', the method employs Earth Mover's Distance (EMD):

  $$L_{\text{EMD}}(F(S, f_\theta), S')$$

  EMD is chosen because it effectively measures the similarity between two unordered sets of points by finding an optimal one-to-one correspondence. This characteristic is particularly valuable given the unordered nature of point clouds.

- **Geometric consistency regularization**
  While the EMD loss helps to align the denoised points with the noisy observations statistically, it does not ensure that the SDF learned is geometrically meaningful. To address this, the authors

introduce a regularization term that enforces geometric consistency. For an arbitrary query q, the absolute value $|f_\theta(q)|$ should match the minimum distance from q to the denoised surface. Formally, the equation is : We have for any query point q, the absolute value $|f_\theta(q)|$ approximates the minimum distance from q to the denoised surface:

$$L_{\text{geo}} = \sum_{q \in S} \max\left(0, \ |f_\theta(q)| - \min_{q' \in F(S, f_\theta)} \|q - q'\|\right)$$

A penalty is applied when this condition is not met, which in turn encourages the network to produce an SDF that accurately reflects the true geometry of the object.

- **Overall loss**
  The total loss is a weighted sum of the EMD loss and the geometric consistency regularization:

$$L_{\text{total}_1} = \sum_{S \in \mathcal{S}} \sum_{S' \in \mathcal{S}} L_{\text{EMD}}(F(S, f_\theta), S') + \lambda \, L_{geo}(f_\theta) + \lambda' L_{eikonal}(f_\theta)$$

  where λ is a balancing hyperparameter and $L_{eikonal}$ the usual eikonal loss for surface reconstruction.

3.1.4 Fast learning with multi-resolution hash encodings

One of the major contributions of the paper is the introduction of a fast learning framework that significantly reduces training time. Traditional coordinate-based MLPs for implicit representations are known to converge slowly, often requiring extensive training iterations.

- **Multi-resolution hash encoding**
  Inspired by Instant Neural Graphics Primitives [7], the method incorporates multi-resolution hash encodings to represent the input space efficiently. The bounded space containing the shape is partitioned into several voxel grids at different resolutions. For each grid cell, a learnable feature vector is stored in a hash grid When a query point q is processed, its features are extracted from the surrounding grid cells across different resolutions through trilinear (or bilinear in 2D) interpolation. These multi-scale features, concatenated with the original coordinates, serve as input to the MLP that predicts $f_\theta(q)$.

This representation allows the network to capture fine details at multiple scales and drastically accelerates convergence. The paper reports that with the multi-resolution hash encoding, the network can achieve convergence in roughly one minute.

3.2 Our Implementation

The 2D implementation presented in this work is designed to validate the core concepts of the original paper while addressing practical challenges encountered during reproduction. The model selection for evaluation is detailed in Section 3.2.1, while a practical implementation of the loss function is described in Section 3.2.2. Additionally, the dataset used for model evaluation is outlined in Section 3.2.3.

3.2.1 Network architecture

The proposed multi-resolution hash encoding model is compared with a SIREN-style network.

The SIREN-style network takes a 2D coordinate $q \in R^2$ as input and outputs a scalar $f_\theta(q)$, representing the signed distance from q to the surface. The use of sine activations enables the network to capture high-frequency variations, which is essential for accurately modeling fine details of the surface. Unlike the proposed model, the SIREN-style network is a standard MLP that does not perform spatial discretization through voxelization.

In the original paper, the authors employed a multi-resolution hash encoding. In this implementation, despite working in a 2D setting, this approach was also incorporated. However, significant initialization challenges were encountered, which were not discussed in the original work. Specifically, it was observed that the initialization scheme from Instant-NGP, where the hash encoding values were
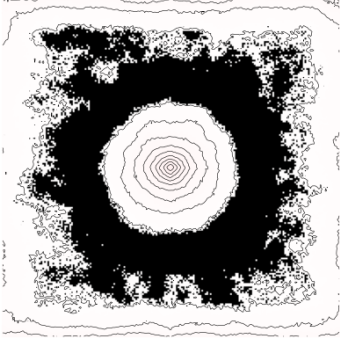
Figure 1: Result with the initial initialization. The output of the model are mostly close to 0.
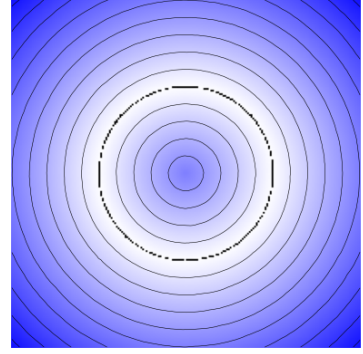

Figure 2: Result of the multi-hash resolution with our initialization. The result are more stable.

sampled from $U(-1e-4, 1e-4)$ (resulting in a variance of approximately 10^{-8}), led to instability in network convergence due to the similarity of initial values.

To address this issue, the shallow MLP was initialized using a normal distribution $N(0,1)$, which increased variance and improved differentiation in network outputs, ultimately leading to more stable training. Figure 1 illustrates the resulting improvements in performance.

3.2.2 Loss function

In addition to the loss described in section 3.1.3, we experimented with two alternative losses to compare with the original paper's proposal:

- **L2 Loss:** We add a basic L2 loss that directly encourages the network output to be zero on the surface. Although this loss performs better on noise-free data, it may struggle under high noise conditions despite potential averaging benefits:

$$L_2 = \sum_{S_i \in S} \|F(S_i, f_\theta)\|^2.$$

- **Alternative to Geometric Consistency Loss:** As the paper notes, the geometric consistency loss can be unstable. Following the approach in Instant-NGP, we precomputed a possible SDF (serving as an approximate ground truth) and enforced that the network's output be as close as possible to this precomputed SDF:

$$\mathcal{L}_{amb} = E_{x \sim X}\left[\|F_\theta(x, 0) - \text{SDF}_{\text{gt}}(x)\|^2\right].$$

The final loss function is defined as a weighted sum of the distance loss and the geometric regularization term (or one of the alternative losses), where the hyperparameter λ controls the relative contributions of each term:

$$L_{\text{total\_2}} = L_{\text{total\_1}} + \lambda_2 L_2 + \lambda_3 L_{\text{amb}}.$$

This formulation provides flexibility in adjusting the influence of different loss components, allowing adaptation to varying noise levels and dataset characteristics. However, our experimental findings indicate that the choice of λ values significantly impacts the quality of the output, and tuning these parameters is not intuitive.

We explored several configurations, including those inspired by the original paper, which were relatively ambiguous. In our experiments, we used $\lambda = 0.01$, $\lambda' = \lambda_1 = 0.1$, and $\lambda_2 = 1$ when disabling L_{EMD}.
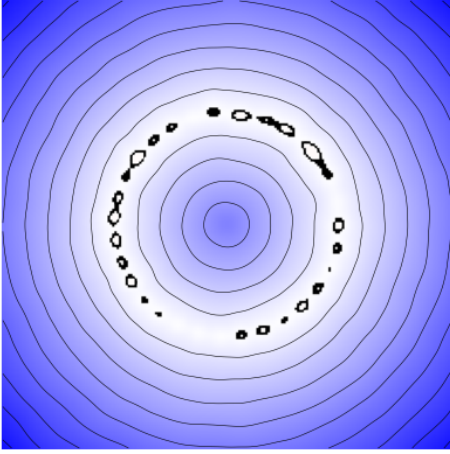
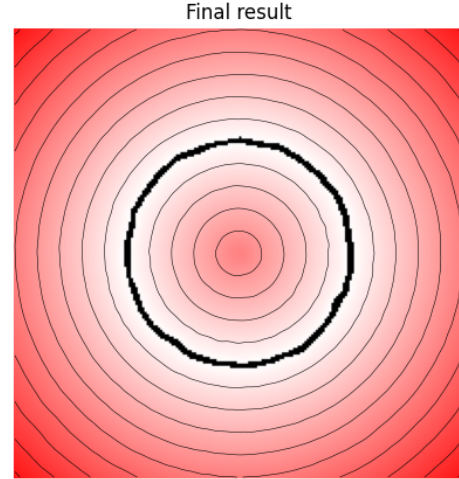Figure 3: SDF obtained with SIREN Net
with the EMD Loss

Figure 4: SDF Obtained with the L2 Loss

3.2.3 Data generation

To evaluate our model, we generate synthetic 2D shapes that serve as the clean underlying surfaces. Specifically, we create simple circular shapes that act as the ground truth boundaries for reconstruction. While our primary focus is on 3D point clouds, using 2D shapes allows for a more comprehensive visualization of the signed distance function, providing deeper insights into the model's output.

To introduce realistic noise, we simulate noisy observations by perturbing points sampled from the clean shapes with Gaussian noise. This approach results in one or multiple noisy datasets that the network utilizes during training, mimicking real-world imperfections in sensor data.

4 Experiments and Results

4.1 Ablation studies

To better understand the contribution of each component of our model, we conducted several ablation experiments:

To better understand the contribution of each component of our model, we conducted several ablation experiments. First, we examined variations in the loss function by comparing the combined loss (distance loss + geometric consistency regularization) with the basic L2 loss and the alternative consistency loss $\mathcal{L}_{amb}$. As discussed in Section 4.1.1, this comparison highlights how different loss formulations affect the quality of the reconstructed surface.

Next, we analyzed the effect of network initialization on training stability and convergence. As detailed in Section 4.2, initializing the MLP with a normal distribution $N(0,1)$ improved variance and led to more robust convergence, whereas low-variance uniform initialization resulted in less stable training dynamics.

Finally, we investigated the impact of noise levels on reconstruction performance. By varying the level of Gaussian noise applied to synthetic shapes (Section 4.2), we assessed the robustness of our approach under different noise conditions.

4.1.1 Comparison between L2 and EMD losses

A comparison between the Earth Mover's Distance (EMD) loss and the L2 loss for surface reconstruction is presented in Figure 2, using InstantNGP and SIRENNet (see Figure 3). When employing the EMD loss, the reconstructed shape resulted in an incomplete circle with a surface exhibiting noticeable thickness rather than a sharp boundary. These results indicate a discrepancy between our observations and the quality reported in the original paper.
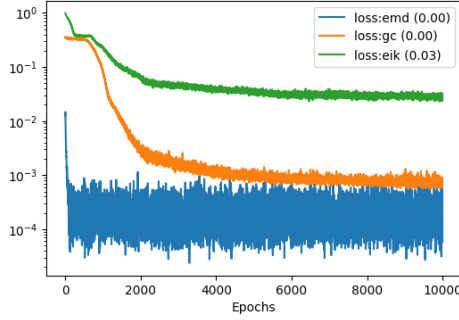
Figure 5: Evolution of the loss with the number of epoch of the SIRENNet

Method	Epochs to Convergence	Training Time (min)	Reconstruction Quality
SIRENNet	10,000	5	Good (L2), Average (EMD)
Instant-NGP	5,000	2	Good (L2), Average (EMD)

Table 1: Comparison between SIRENNet and Instant-NGP in terms of training efficiency and reconstruction quality.

On the contrary, with L2 Loss, the circle is more well-formed (see Figure 4). This suggests that while EMD aligns point clouds statistically, it does not necessarily enforce a sharp boundary, leading to inferior results in certain cases.

4.1.2 Loss evolution

To analyze how the model balances different loss terms during training, we track their evolution over 10,000 epochs (see Figure 5).

The EMD loss decreases rapidly during the first 1,000 epochs, from approximately 10^{-1} to 10^{-3}, suggesting that the model quickly learns to align the denoised points with the noisy observations, effectively mapping them onto a consistent surface. After 2,000 epochs, the loss stabilizes, indicating that the network has reached a stable alignment between the noisy and denoised point clouds.

The geometric consistency loss starts around 10^{-1} and converges to approximately 10^{-2}. This regularization term remains challenging to optimize, as it enforces a stricter constraint: each point's absolute signed distance $|f_\theta(q)|$ should correspond to the minimum distance to the denoised surface.

The Eikonal loss initially has a value around 10^0 (or 1) and gradually decreases to approximately 10^{-1}. This trend indicates that the model progressively enforces the constraint $\|\nabla f_\theta(q)\| \approx 1$, ensuring that the signed distance function (SDF) retains the desired geometric properties.

4.2 Difference between SIRENNet and Instant-NGP

We evaluated the differences in performance between SIRENNet, a fully connected MLP with sine activations, and Instant-NGP, which utilizes multi-resolution hash encoding. The key results are summarized in Table 1.

Both architectures achieve comparable reconstruction accuracy. However, training SIRENNet requires a significantly higher number of iterations to converge, as the network must learn the signed distance function (SDF) over the entire continuous 2D domain without explicit multi-scale guidance. In contrast, Instant-NGP benefits from hierarchical encoding, which accelerates convergence while maintaining similar reconstruction quality.

4.3 Robustness with respect to the number of input points

we assess the model's performance when the number of input points available to define the surface varies. In practical scenarios, an object might not be thoroughly scanned, resulting in fewer points
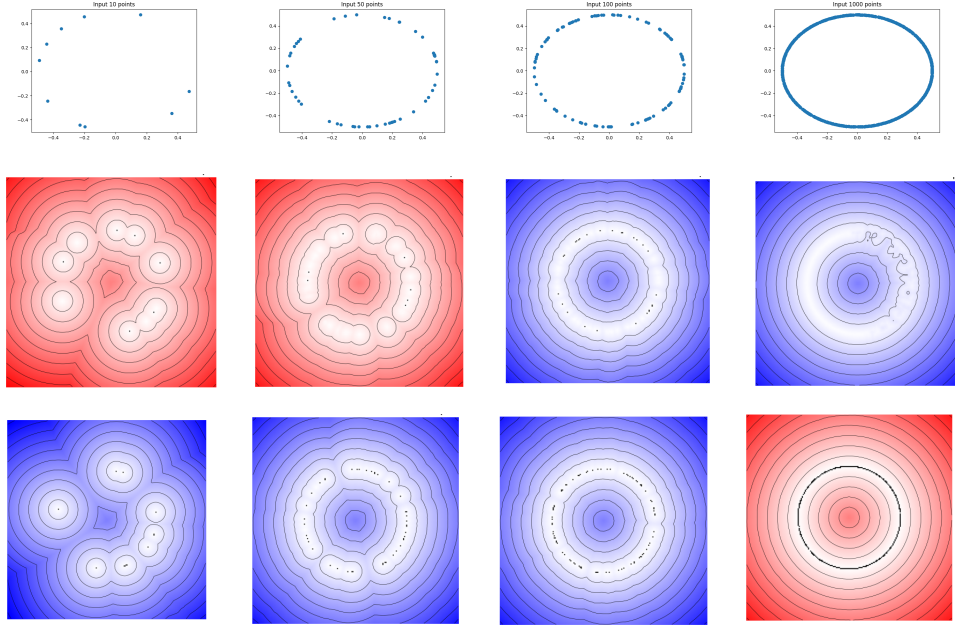
7

Figure 6: Comparison of input data, EMD, and L1 results on the multi-hash resolution networks for different numbers of sampled points (10, 50, 100, and 1000 points). The first row shows the input images, the second row shows the EMD results, and the third row shows the L1 results.

for surface reconstruction. Although this experiment was not explicitly covered in the original paper, we sampled n points from our synthetic circle dataset and compared the reconstruction quality using different loss functions.

Figure 4.3 shows the results. With only 10 or 50 points, both losses struggle to define a precise zero-level set. However, starting from around 100 points, the network begins to form a candidate zero-level set. By 1000 points, the model trained with L2 loss is able to accurately pinpoint the zero-level set, whereas the model trained with EMD loss still exhibits some difficulties.

4.4 Robustness to Noise

To evaluate the robustness of the surface reconstruction approach under noisy conditions, we conducted experiments using two different noise models. First, we introduced additive Gaussian noise with standard deviations $\sigma = 0.1$, 0.05, and 0.01 to the input circle data. These experiments assess the ability of the loss functions to recover a well-defined surface despite both random and systematic noise perturbations. The results are presented in Figure 9.

As noise levels increase, both models exhibit performance degradation. In particular, the zero-level set of the L2 loss expands significantly, causing the model to struggle in accurately capturing the underlying surface. Notably, the geometric consistency term is not strong enough to prevent the model from predicting near-zero values in regions where input points exist, making the L2-based reconstruction highly sensitive to noise.

Conversely, the EMD loss maintains a more consistent surface representation. The results indicate that while both losses are affected by noise, the EMD formulation preserves the integrity of the reconstructed surface more effectively than the L2 loss.

4.5 Scaling to 3D

To extend the implicit SDF approach to three dimensions, the data generation, network architecture, and visualization pipeline were all adapted to accept 3D coordinates. The figure 10 shows the projected 3D point cloud, which forms a nearly perfect sphere of radius 0.5. This visualization confirms that the network successfully learned and reconstructed the geometry. The figure 11 shows 2D slices through the learned 3D SDF volume, where the ring at the center of the middle slice corresponds to the
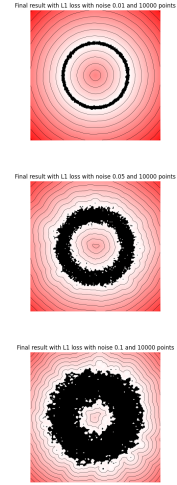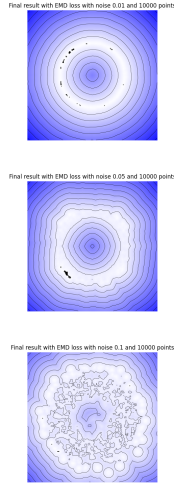
Figure 7: SDF obtained with EMD loss at different noise levels ($\sigma = 0.01$, 0.05, and 0.1).

Figure 8: SDF obtained with L2 loss at different noise levels ($\sigma = 0.01$, 0.05, and 0.1).

Figure 9: Comparison of the SDF reconstructions using L2 and EMD loss under different levels of Gaussian noise.
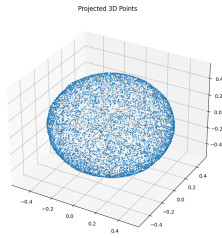

Figure 10: Output of the 3D network on a noise-free 3D sphere

zero-level set of the sphere. Slices on either side of that plane show a continuous transition from inside (negative SDF) to outside (positive SDF). These results indicate that the model can accurately capture the shape's interior and exterior regions in a higher dimensional setting. Subtle differences in how sharply the SDF transitions around the surface can be influenced by hyperparameter choices. We couldn't use a more complex 3D dataset, such as those used in the original paper, because we didn't have sufficient GPU resources. However, even with our adapted pipeline to smaller 3D dataset, we were able to confirm that it functions correctly.

5   Discussion

Our study demonstrates that learning Signed Distance Functions from noisy point clouds via a noise-to-noise mapping strategy is both feasible and effective, as shown by our experiments in 2D and 3D.
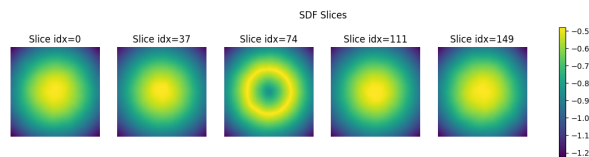

Figure 11:   Slice of the Sign distance function after training.

The model successfully recovers accurate geometrical representations without relying on clean ground truth data.

Despite these promising outcomes, our implementation faces several challenges that warrant further consideration. The performance of the network is noticeably sensitive to hyperparameter choices, such as the weighting of different loss components.

Looking ahead, several improvements could enhance the robustness and efficiency of our approach. For example, exploring hybrid loss functions that combine the benefits of L2 and EMD, or even alternative distance measures, may lead to reconstructions that are both sharp and resilient to noise.

References

[1] Matan Atzmon and Yaron Lipman. Sal: Sign agnostic learning of shapes from raw data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2565–2574, 2020.

[2] Pedro Hermosilla, Tobias Ritschel, and Timo Ropinski. Total denoising: Unsupervised learning of 3d point cloud cleaning. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 52–60, 2019.

[3] Michael Kazhdan and Hugues Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)*, 32(3):1–13, 2013.

[4] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018.

[5] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning image restoration without clean data. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2965–2974. PMLR, 10–15 Jul 2018.

[6] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022.

[7] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):102:1–102:15, July 2022.

[8] Yutaka Ohtake, Alexander Belyaev, Marc Alexa, Greg Turk, and Hans-Peter Seidel. Multi-level partition of unity implicits. In *Acm Siggraph 2005 Courses*, pages 173–es. 2005.

[9] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019.

[10] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.