Sign Agnostic Learning with Derivatives (SALD) of 3D Geometry from Raw Data

Debabrata Ghosh

April 17, 2023

Abstract

Lately, several works have addressed the challenge of modeling 3D shapes using deep neural networks to learn implicit surface representations. In this report, we discuss one such method where we learn 3D geometry directly from raw data such as point clouds, or triangle soups which can be used in related tasks like surface reconstruction or generation.

Previous works used training data sampled from a ground truth implicit function like a signed distance function, or occupancy networks to learn the neural network approximating the 3D shape. Such requirement of 3D supervision made it harder to compute. The method discussed in this report solves this issue by making use of only an unsigned distance function from the raw data. We build upon the method used in sign agnostic learning (SAL) by incorporating the derivatives along with the point-wise values of the unsigned distance function in the regression loss. Optimizing such loss with specific network initialization leads to a better quality approximation of the 3D shape which also stems from the fact that including gradients of the unsigned distance function in the loss function results in lower sample complexity and a signed implicit function whose zero-level set is a better fit. We also identify how missing parts are learned through this method as it avails the minimal surface property, consequently learning minimum area solutions.

We also look into the efficacy of the method experimented with two datasets: dynamic Faust or D-Faust [Bog+17] dataset containing about 40,000 raw 3d scans as triangle soups and ShapeNet [Cha+15] dataset consisting of non-manifold and inconsistently oriented synthetic 3D data. Finally, we discuss some further advancements made in terms of learning implicit neural representations of 3D shapes inspired by our approach e.g. LightSAL [BSB21], and a few other recent works such as DiGS [BKG21] which even surpass the effectiveness of our technique.

Keywords *Implicit Neural Representation, Sign Agnostic Learning, Minimal Surface Property*

Contents

1 Introduction

- Recent trend of using Neural networks (NN) to model and reconstruct 3D surfaces. From discrete representations (e.g. voxels, point clouds, triangle soup[1]) to continuous representations.

- NN-based 3D learning techniques can differ in two aspects: the kind of surface representation, and the supervision method of learning the NN.

- Mainly parametric and implicit representations. In the parametric setting NNs used as parameterization mappings, while in implicit representation surfaces as zero level-sets of neural networks:

$$\mathcal{S} = \left\{ \boldsymbol{x} \in \mathbb{R}^3 \mid f(\boldsymbol{x}; \boldsymbol{\theta}) = 0 \right\}$$

  where $f : \mathbb{R}^3 \times \mathbb{R}^m \to \mathbb{R}$ is a neural network.

- The benefit of using neural networks as implicit representations of surfaces stems from their flexibility and approximation power [Atz+19] along with their optimization and generalization properties.

- Different supervision methods include regression of known or approximated volumetric implicit representations [Par+19; Mes+18], direct regression with raw 3D data [AL19; Gro+20].

- Earlier methods to learn neural implicit surface representations using a regression-type loss required data samples from a ground-truth implicit representation of the surface, such as a signed distance function [Par+19] or an occupancy function [Mes+18; CZ18]. No such data are available for the common raw form of acquired 3D data $\mathcal{X} \subset \mathbb{R}^3$ such as a point cloud or a triangle soup, and computing a ground-truth implicit representation for such cases is quite challenging [Ber+17].

- We introduce sign-agnostic learning (SAL) that doesn't need a ground truth signed distance representation of surfaces for training. We then extend the same method by adding derivatives of the unsigned distance function in the regression loss used in SAL and get the sign-agnostic learning with derivatives (SALD) technique.

- We discuss the benefit of using derivatives while training the NN [Cza+17].

- We mention the minimal surface property which holds for SALD.

- We discuss the results of SALD on the two datasets experimented on, i.e., D-Faust and ShapeNet.

- We mention some limitations of SALD and further advancements in the field of learning implicit neural representation (INR).

[1] A collection of triangles in space, not necessarily a manifold or consistently oriented.

2 Background

2.1 Sign Agnostic Learning (SAL)

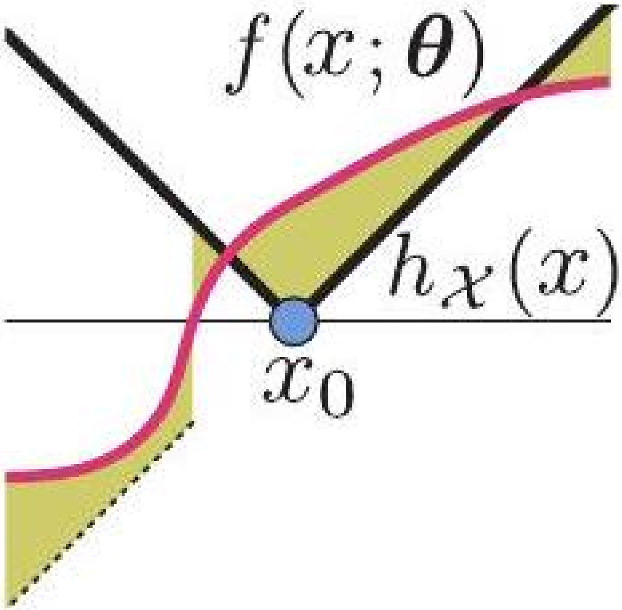


Figure 1: 1-Dimensional Sign Agnostic Learning: Approximating the unsigned distance function with proper initialization results in the unsigned distance function.

- Given raw geometric data as input, $\mathcal{X} \subset \mathbb{R}^3$, we are trying to optimize the weights $\boldsymbol{\theta} \in \mathbb{R}^m$ of a network $f(\boldsymbol{x};\boldsymbol{\theta})$, where $f : \mathbb{R}^3 \times \mathbb{R}^m \to \mathbb{R}$, so that zero level-set of f is the surface approximating $\mathcal{X}$.
- Sign Agnostic Learning (SAL) defined by a loss of the form

$$\text{loss}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{x}\sim D_{\mathcal{X}}} \tau \left(f(\boldsymbol{x};\boldsymbol{\theta}), h_{\mathcal{X}}(\boldsymbol{x}) \right),$$

  where $D_{\mathcal{X}}$ is a probability distribution defined by the input data $\mathcal{X}$; $h_{\mathcal{X}}(\boldsymbol{x})$ is some unsigned distance measure to $\mathcal{X}$; and $\tau : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ is a differentiable unsigned similarity function defined by the following properties:

  (i) Sign agnostic: $\tau(-a, b) = \tau(a, b), \forall a \in \mathbb{R}, b \in \mathbb{R}_+$.

  (ii) Monotonic: $\frac{\partial \tau}{\partial a}(a, b) = g(a - b), \forall a, b \in \mathbb{R}_+$,

  where $g : \mathbb{R} \to \mathbb{R}$ is a monotonically increasing function with $g(0) = 0$ [AL19].
- Unsigned distance functions considered are standard L^2 (Euclidean) distance

$$h_2(\boldsymbol{z}) = \min_{\boldsymbol{x}\in\mathcal{X}} \|\boldsymbol{z} - \boldsymbol{x}\|_2,$$

  and L^0 distance

$$h_0(\boldsymbol{z}) = \begin{cases} 0 & \boldsymbol{z} \in \mathcal{X} \\ 1 & \boldsymbol{z} \notin \mathcal{X} \end{cases}$$

  .

- The unsigned similarity function chosen in SAL is:

$$\tau_\ell(a, b) = ||a| - b|^\ell$$

  where $\ell \geq 1$. The function τ_ℓ satisfies (i) due to the symmetry of $|\cdot|$; and since $\frac{\partial \tau}{\partial a} = \ell ||a| - b|^{\ell-1} \operatorname{sign}(a - b \operatorname{sign}(a))$ it satisfies (ii) as well.

2.2 Sobolev Training

- If we have access to derivatives of the target output concerning the input, Sobolev Training for neural networks incorporates such target derivatives along with the target values while training.
- Optimizing neural networks to approximate the function's derivatives on top of approximating the function's outputs helps to capture additional information within the parameters of the neural network. This leads to better approximation and generalization of our desired function.

2.3 Limitations of SAL

- Rounding of corners in the surface solution.
- Possibly related to the convergence of the learning process but even with tweaked learning rates or a higher number of epochs the learned surface has lack of sharpness in corners.

3 Previous Work

The division is based on the kind of 3D surface representation used.

3.1 Parametric Representation

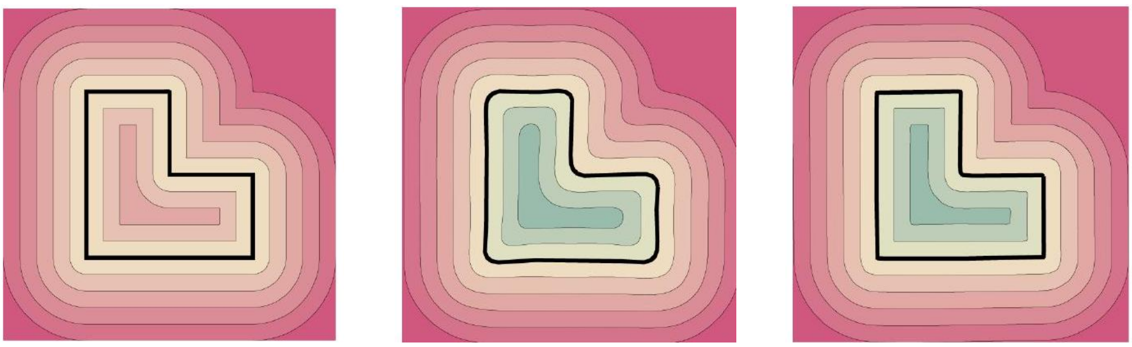
3.2 Implicit Representation

3.3 Template Fitting

3.4 Primitives

4 Method

4.1 Sign Agnostic Learning with Derivatives



(a) Unsigned distance (b) SAL (c) SALD

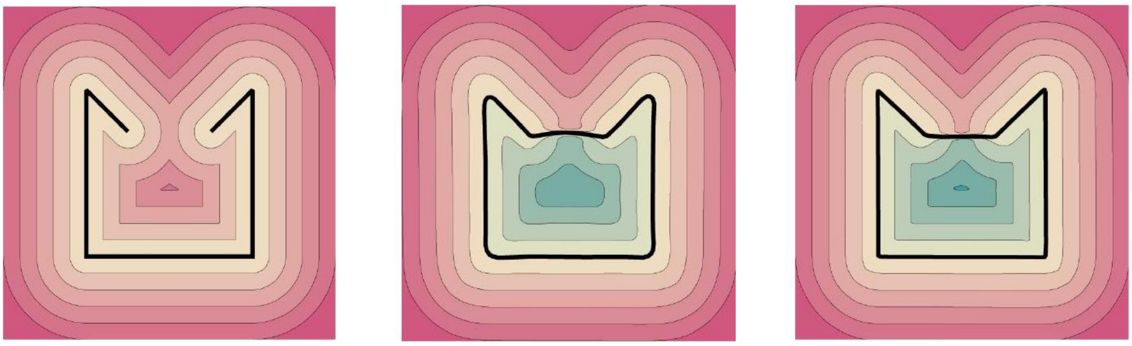
Figure 2: Sign agnostic learning of an unsigned distance function to an L shape (left) where red colors represent positive distance or the outside region, and green colors mean negative distances on the inside region. In the middle, is the result of optimizing the SAL loss, and on the right, is the result of optimizing the SALD loss.

- SALD loss:

$$\text{loss}(\theta) = \mathbb{E}_{\boldsymbol{x}\sim\mathcal{D}}\tau(f(\boldsymbol{x};\theta), h(\boldsymbol{x})) + \lambda\mathbb{E}_{\boldsymbol{x}\sim\mathcal{D}'}\tau\left(\nabla_{\boldsymbol{x}} f(\boldsymbol{x};\theta), \nabla_{\boldsymbol{x}} h(\boldsymbol{x})\right)$$

  where $\lambda > 0$ is a parameter, $\mathcal{D}'$ is a probability distribution, and $\nabla_{\boldsymbol{x}} f(\boldsymbol{x};\theta), \nabla_{\boldsymbol{x}} h(\boldsymbol{x})$ are the gradients f, h (resp.) with respect to their input $\boldsymbol{x}$.

4.2 Minimal Surface Property in 2D



(a) Unsigned distance (b) SAL (c) SALD

Figure 3: Experimental evidence of holding minimal surface property: using SAL (middle) and SALD (right) for the given unsigned distance function of a curve with a missing part (left) leads to a solution (black line) with approximately minimal length in the missing part area.

- Describe minimal surface property [ZOF01].
- Theoretical and experimental justification in case of SAL and SALD.

5 Experiments and Results

5.1 Neural Network Architecture

- Used a modified variational encoder-decoder [KW22], where the encoder $(\boldsymbol{\mu}, \boldsymbol{\eta}) = g(\boldsymbol{X}; \boldsymbol{\theta}_1)$ is taken to be PointNet [Qi+16], $\boldsymbol{X} \in \mathbb{R}^{n \times 3}$ is an input point cloud ($n = 128^2$), $\boldsymbol{\mu} \in \mathbb{R}^{256}$ is the latent vector, and $\boldsymbol{\eta} \in \mathbb{R}^{256}$ represents a diagonal covariance matrix by $\boldsymbol{\Sigma} = \operatorname{diag}(e^{\boldsymbol{\eta}})$.
- The encoder takes in a point cloud $\boldsymbol{X}$ and outputs a probability measure $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The point cloud is drawn uniformly at random from the scans.
- The decoder is the implicit representation $f(\boldsymbol{x}; \boldsymbol{w}, \boldsymbol{\theta}_2)$ with the addition of a latent vector $\boldsymbol{w} \in \mathbb{R}^{256}$.
- The network f is taken to be an 8-layer Multi-layer Perceptron (MLP).

5.2 D-Faust

- Baselines: DeepSDF [Par+19] is chosen as a representative out of the methods that require pre-computed implicit representation for training and for methods that train directly on raw 3D data, SAL [AL19] and IGR [Gro+20] are selected.
- Show quantitative and qualitative results.
- Although SALD does not give the best test quantitative results, it is roughly comparable in every measure to the best among the two baselines.

5.3 ShapeNet

- Show qualitative and quantitative results for the held-out test set.
- Qualitatively the surfaces produced by SALD are smoother, mostly with more accurate sharp features than SAL and DeepSDF generated surfaces.

5.4 Shortcomings

- While inside and outside information is not known and often not even well-defined such as in ShapeNet models, our technique can add surface sheets closing what should be open areas
- Thin structures might be missed while learning.

6 Further Work

6.1 Improvements on Same Idea

- LightSAL, a novel deep convolutional architecture for learning 3D shapes based on the same technique as SAL which concentrates on efficiency both in network training time and resulting model size [BSB21].

6.2 Further Improvements on Related Tasks

- Existing INRs require point coordinates to learn the implicit level sets of the shape and if a normal vector is available for each point, a higher fidelity representation can be learned, however, normal vectors are often not provided as raw data. In DiGS, the authors propose a divergence-guided shape representation learning approach that does not require normal vectors as input [BKG21].
- The authors in the DiGS paper show that incorporating a soft constraint on the divergence of the distance function leads to smooth solutions that reliably orient gradients to match the unknown normal at each point, sometimes even better than approaches that use ground truth normal vectors directly [BKG21].

7 Conclusion

- Talk about the motivation and benefits of the method in short.
- Recapitulate the possibility of further improvements and related current works.
- Mention some possible application areas such as 3D generative modeling.

References

[AL19] Matan Atzmon and Yaron Lipman. "SAL: Sign Agnostic Learning of Shapes from Raw Data". In: *CoRR* abs/1911.10414 (2019). arXiv: `1911.10414`. URL: `http://arxiv.org/abs/1911.10414`.

[Atz+19] Matan Atzmon et al. "Controlling Neural Level Sets". In: *CoRR* abs/1905.11911 (2019). arXiv: `1905.11911`. URL: `http://arxiv.org/abs/1905.11911`.

[BSB21] Abol Basher, Muhammad Sarmad, and Jani Boutellier. "LightSAL: Lightweight Sign Agnostic Learning for Implicit Surface Representation". In: *CoRR* abs/2103.14273 (2021). arXiv: `2103.14273`. URL: `https://arxiv.org/abs/2103.14273`.

[BKG21] Yizhak Ben-Shabat, Chamin Hewa Koneputugodage, and Stephen Gould. "DiGS : Divergence guided shape implicit neural representation for unoriented point clouds". In: *CoRR* abs/2106.10811 (2021). arXiv: `2106.10811`. URL: `https://arxiv.org/abs/2106.10811`.

[Ber+17] Matthew Berger et al. "A Survey of Surface Reconstruction from Point Clouds". In: *Computer Graphics Forum* 36.1 (2017), pp. 301–329. DOI: `https://doi.org/10.1111/cgf.12802`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.12802`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12802`.

[Bog+17] Federica Bogo et al. "Dynamic FAUST: Registering Human Bodies in Motion". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 5573–5582. DOI: `10.1109/CVPR.2017.591`.

[Cha+15] Angel X. Chang et al. *ShapeNet: An Information-Rich 3D Model Repository*. 2015. arXiv: 1512.03012 [cs.GR].

[CZ18] Zhiqin Chen and Hao Zhang. "Learning Implicit Fields for Generative Shape Modeling". In: *CoRR* abs/1812.02822 (2018). arXiv: 1812.02822. URL: http://arxiv.org/abs/1812.02822.

[Cza+17] Wojciech Marian Czarnecki et al. "Sobolev Training for Neural Networks". In: *CoRR* abs/1706.04859 (2017). arXiv: 1706.04859. URL: http://arxiv.org/abs/1706.04859.

[Gro+20] Amos Gropp et al. "Implicit Geometric Regularization for Learning Shapes". In: *CoRR* abs/2002.10099 (2020). arXiv: 2002.10099. URL: https://arxiv.org/abs/2002.10099.

[KW22] Diederik P Kingma and Max Welling. *Auto-Encoding Variational Bayes*. 2022. arXiv: 1312.6114 [stat.ML].

[Mes+18] Lars M. Mescheder et al. "Occupancy Networks: Learning 3D Reconstruction in Function Space". In: *CoRR* abs/1812.03828 (2018). arXiv: 1812.03828. URL: http://arxiv.org/abs/1812.03828.

[Par+19] Jeong Joon Park et al. "DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation". In: *CoRR* abs/1901.05103 (2019). arXiv: 1901.05103. URL: http://arxiv.org/abs/1901.05103.

[Qi+16] Charles Ruizhongtai Qi et al. "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation". In: *CoRR* abs/1612.00593 (2016). arXiv: 1612.00593. URL: http://arxiv.org/abs/1612.00593.

[ZOF01] Hong-Kai Zhao, S. Osher, and R. Fedkiw. "Fast surface reconstruction using the level set method". In: *Proceedings IEEE Workshop on Variational and Level Set Methods in Computer Vision*. 2001, pp. 194–201. DOI: 10.1109/VLSM.2001.938900.