FAST 3D ACOUSTIC SCATTERING VIA DISCRETE LAPLACIAN BASED IMPLICIT FUNCTION ENCODERS

Anonymous authors
Paper under double-blind review

ABSTRACT

Acoustic properties of objects corresponding to scattering characteristics are frequently used for 3D audio content creation, environmental acoustic effects, localization and acoustic scene analysis, etc. The numeric solvers used to compute these acoustic properties are too slow for interactive applications. We present a novel geometric deep learning algorithm based on discrete-laplacian and implicit encoders to compute these characteristics for general 3D objects at interactive rates. We use a point cloud approximation of each object, and each point is encoded in a high-dimensional latent space. Our multi-layer network can accurately estimate these acoustic properties for arbitrary topologies and takes less than 1ms per object on a NVIDIA GeForce RTX 2080 Ti GPU. We also prove that our learning method is permutation and rotation invariant and demonstrate high accuracy on objects that are quite different from the training data. We highlight its application to generating environmental acoustic effects in dynamic environments.

1 INTRODUCTION

Acoustic scattering corresponds to the disturbance of a given incident sound field due to an object's shape and surface properties. It can be regarded as one of the fundamental characteristics of an object. The effect of scattering can be expressed in terms of a scattered sound field, which satisfies Sommerfield's radiation condition. There is considerable work on modeling and measuring the acoustic scattering properties in physics and acoustics and these characteristics are widely used for sound rendering in games and virtual reality (Mehra et al., 2015; Rungta et al., 2018), noise analysis in indoor scenes (Morales & Manocha, 2018), acoustic modeling of concert halls (Shtrepi et al., 2015), non-line-of-sight (NLOS) imaging (Lindell et al., 2019), understanding room shapes (Dokmanić et al., 2013), receiver placement (Morales et al., 2019), robot sound source localization (An et al., 2019), 3D mapping (Kim et al., 2020), audio-visual analysis (Sterling et al., 2018), etc.

Acoustic scattering of objects can be modeled accurately using the theory of wave acoustics (Kuttruff, 2016). The scattering characteristics of objects are widely used for sound propagation, which reduces to solving the wave equation in large environments. Given a sound source location and its vibration patterns, acoustic simulation methods are used to predict the perceived sound at another specified location considering the medium it passes through and objects/boundaries it interacts with. While the wave behavior of sound is well understood in physics, it is much more difficult to compute acoustic scattering and sound propagation effects, especially for higher frequencies. Even with state-of-the-art acoustic wave solvers, it can take from hours to days to solve a moderately modeled room environment on a powerful workstation. One of the contributing factors to this difficulty is that wave behaviors are frequency dependent, so that many frequency bands need to be analyzed separately.

Current methods for computing the acoustic scattering characteristics can use numeric solvers like boundary-element methods (BEM). However, their complexity increases as a cubic function of the frequencies and most current implementations are limited to static scenes or environments. No good or practical solutions are known to compute the acoustic scattering properties for dynamic environments or when objects move or undergo deformation.

Main Results: We present novel techniques based on geometric deep learning on differential coordinates to approximate the acoustic scattering properties of arbitrary objects. Our approach is general and makes no assumption about object's shape, genus, or rigidity. We approximate the

objects using point-clouds, and each point in the point cloud representation is encoded in a high-dimensional latent space. Moreover, the local surface shapes in the latent space are encoded using implicit surfaces. This enables us to handle arbitrary topology. Our network takes the point cloud as an input and outputs the spherical harmonic coefficients that represent the acoustic scattering field.

We present techniques to generate the ground truth data using an accurate wave-solver on a large geometric dataset. We have evaluated the performance on thousands of objects that are very different from the training database (with varying convexity, genus, shape, size, orientation) and observe high accuracy. We also perform an ablation study to highlight the benefits of our approach. The additional runtime overhead of estimating the scattering field from neural networks is less than 1ms per object on a NVIDIA GeForce RTX 2080 Ti GPU. We also prove that our learning method is permutation and rotation invariant, which is an important characteristic for accurate computation of acoustic scattering fields.

2 Related Works

2.1 Wave-Acoustic Simulation

Wave-acoustic simulation methods aim to solve the wave equation that governs the propagation and scattering of sound waves. Some conventionally used numeric methods include the finite-element method (Thompson, 2006), the boundary-element method (Wrobel & Kassab, 2003), and the finite-difference time domain (Botteldooren, 1995). The common requirement of these methods is the proper discretization of the problem domain (e.g., spatial resolution, time resolution), which means the time complexity of them will scale drastically with the simulation frequency. Recent works manage to speed up wave simulations by parallelized rectangular decomposition (Morales et al., 2015) or by using pre-computation structures (Mehra et al., 2015). An alternative way to use wave simulation results in real-time applications is to pre-compute a large amount of sound fields in a scene and compressing the results using perceptual metrics (Raghuvanshi & Snyder, 2014). While pre-computation methods can save significant runtime cost, they still require non-trivial pre-computation efforts for unseen scenarios, restricting their use cases.

2.2 Learning-Based Acoustics

Machine learning techniques have been widely applied in many popular computer science research areas. There is considerable work on developing machine learning methods for applications corresponding to audio-visual analysis (Zhang et al., 2017) and acoustic scene classification (Abeßer, 2020) (see Bianco et al. (2019) for a thorough list of work). In comparison, there are much fewer works studying the generation of acoustic data from a physical perspective. Therefore, methodologies from existing popular domains often do not directly apply to our problem of interest. Some more relevant works focus on estimating room acoustics parameters from recorded signals (Eaton et al., 2016; Genovese et al., 2019; Tsokaktsidis et al., 2019; Tang et al., 2020), which can help physically-based simulators to model real-world acoustics more faithfully. Pulkki & Svensson (2019) propose to use a neural network to model the acoustic scattering effect from rectangular plates without running simulation. This is closely related to our goal of bypassing expensive wave simulation while generating plausible sound, although we aim to model objects of more general shapes. Recently, Fan et al. (2020) train convolutional neural networks (CNNs) to learn to map planar sound fields induced by convex scatterers in 2D, which provides promising results. Motivated by the last method, we aim to develop networks that can deal with object geometry in 3D and extend the generality of this learning-based approach.

2.3 Geometric Deep Learning and Shape Representation

There is considerable recent work on generating plausible shape representations for 3D data, including voxel-based (Zhou & Tuzel, 2018; Sindagi et al., 2019; Meng et al., 2019; Wu et al., 2015), point-based (Charles et al., 2017; Qi et al., 2017; Wang et al., 2019; Li et al., 2018a; Monti et al., 2017; Li et al., 2018b; Yi et al., 2017) and mesh-based (Hanocka et al., 2019) shape representations. This includes work on shape representation by learning implicit surfaces on point clouds (Smirnov et al., 2019), designing a mesh Laplacian for convolution (Tan et al., 2018), hierarchical graph

convolution on meshes (Mo et al., 2019), encoding signed distance functions for surface reconstruction (Park et al., 2019), etc. However, previous methods on point cloud shape representations learn by designing loss functions to constrain surface smoothness on global Cartesian coordinates. Such functions only provide spatial information of each point and lack information about local shape of the surface compared to the explicit discretization of the continuous Laplace-Beltrami operator and curvilinear integral (Do Carmo, 2016). Instead, we use point cloud based learning algorithms, which do not require mesh Laplacians for graph neural networks. This makes our approach applicable to all kind of dynamic objects, including changing topologies.

3 Our Approach

3.1 Overview

Our goal is to directly predict the acoustic scattering field (ASF) resulting from an incident wave and a sound scatterer of known shape (described by a triangle mesh). A set of frequency dependent ASFs are visualized in Figure 1 on a disk plane. The two key steps are defining a compact and efficient way to encode the ASF, and designing a neural network architecture that can effectively exploit the 3D geometric information so as to establish the mapping between the object shape and the ASF. In this section, we provide necessary background on wave acoustics and explain the mechanism of our approach. Our notations are summarized in Table 1.

Table 1: Notation and symbols used in the paper.

$\mathbf{x}$	3D Cartesian coordinates.	m, l	Order and degree of basis functions.
v	acoustic particle velocity.	$Y_l^m(\theta, \phi)$	Spherical harmonics basis function.
ω	Frequency of sound.	k	Wavenumber.
(r, θ, ϕ)	Spherical coordinates.	$c_l^m(\omega)$	Spherical harmonics coefficients.
c	Speed of sound in air.	$h_l(kr)$	Hankel function.

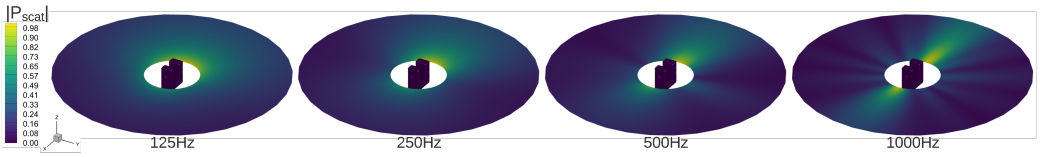

Figure 1: Acoustic scattering of the same object at four different frequencies, assuming a plane wave travelling to the $-x$ direction and a sound-hard boundary condition for the object/scatterer in the center. The scattering patterns are significantly different among frequency bands. Our learning method can compute them with high accuracy for arbitrary objects at interactive rates, including high-genus and deforming objects.

3.2 Acoustic Scattering Fields

We are interested in knowing how the incident sound field will be scattered when it acts on an object surface. Let the sound pressure at location $\mathbf{x}$ and time t be $P(\mathbf{x}, t)$, and the wave equation describes the propagation of sound:

$$(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2})P(\mathbf{x}, t) = 0, \tag{1}$$

where c is the speed of sound. Transforming into the frequency domain, we obtain the homogeneous Helmholtz equation:

$$(\nabla^2 + k^2)p(\mathbf{x}, \omega) = 0, \tag{2}$$

where ω is the wave frequency and $k = \frac{\omega}{c}$ is the wavenumber. When a specific boundary condition is given, its solution in 3D spherical coordinates can be decomposed as

$$p(r, \theta, \phi, \omega) = \sum_{l=0}^{\infty} h_l(kr) \sum_{m=-l}^{+l} c_l^m(\omega) Y_l^m(\theta, \phi), \tag{3}$$

where $h_l(kr)$ is the Hankel function which encodes the radial pressure changes, $Y_l^m(\theta, \phi)$ is the spherical harmonics term which represents the angular part of the solution, and $c_l^m(\omega)$ is the spherical harmonics coefficients for frequency ω. For a maximum order L, there are $(L+1)^2$ coefficients. Since the Hankel and the Spherical Harmonics functions are easy to evaluate, we only need to find the coefficients $c_l^m(\omega)$ to fully describe the acoustic scattering field.

3.3 Network Inference

Our goal is to use an appropriate geometric representation for the underlying objects in the scene so that we can apply geometric deep learning methods to infer the acoustic scattering field from the object shape. It is also important that our approach should be able to handle highly dynamic scenes with arbitrary objects. Moreover, our approach can still produce plausible ASFs results for changing typologies. The input to our system will be the point cloud representation ($N \times 3$ matrix, N being the number of 3D points) of an object. We represent each point and its local surface by a higher dimension implicit surface in the latent space formed by an implicit surface encoder. The desired output is the spherical harmonic coefficients vector up to order L described in § 3.2. In practice, acoustic scattering corresponding to different frequencies can exhibit different distributions. Therefore, we train several networks for different frequency bands and in this paper we present the results for $125Hz \sim 1000Hz$. However, this range is not a limit of our approach, as we can handle a wider frequency range, though the complexity of computing the training data can increase as a cubic function of the frequency. We present a novel network architecture to compute these characteristics for arbitrary objects.

4 Network Design and Properties

In this section we introduce our shape laplacian based point representation using implicit functions. For predicting ASF, the detailed geometric characteristics would play an essential role as the frequency grows and thereby motivate the design of our differential coordinates describing the discretization of the curvilinear integral around a given point. Then we provide the proof of its permutation invariant property via Newton's identities in the appendix. Finally, following de Finetti theorem (Heath & Sudderth, 1976), we justify our design of implicit surface functions and relate it to the acoustic scattering functions.

4.1 Local Surface Shape and Implicit Surface Encoder

Previous works on point cloud learning algorithms mostly focus on designing per-point operations (Charles et al., 2017), encoding per-point features to estimate continuous shape functions (Park et al., 2019; Xu et al., 2019), or minimizing loss between a point normal vector and its connected vertices (Liu et al., 2019).

For each point in the input cloud and its neighborhood in the Euclidean space, we assume that it can form a piecewise smooth surface around the point and each point is encoded by the shared multilayer perceptron (MLP) (Rumelhart et al., 1985) and can be represented by a vector in the higher dimensional latent space (see Figure 2). Thus, a piecewise-linear approximation of the surface around a given point can be used to estimate the local surface shape, where differential coordinates (Sorkine, 2006; Do Carmo, 2016) (i.e., $\delta-$ coordinates) of each vertex v_i can be expressed as:

$$\overrightarrow{\delta_i} = \frac{1}{d_i} \Sigma_{j \in N(i)} (\overrightarrow{v_i} - \overrightarrow{v_j}). \tag{4}$$

Here δ_i encapsulates the local surface shape, $N(i)$ represents the k nearest neighbors of vertex v_i in the Euclidean space, and $d_i = |N(i)|$ is the number of immediate neighbors of v_i. To estimate the mean curvature of the local surfaces formed by each point and its spatial neighbors, we use the radial basis function (RBF) $\varphi(\cdot) = \exp^{-||\cdot||^2}$ to weight each vector, rather than using the uniform weight shown in Equation (4). Since there are $N!$ permutations for a point cloud with N vertices, every operation on point clouds should be permutation invariant (i.e., the permutation of input points should not change the output of our network). Our weight function is designed to be positive definite and symmetric for any choice of data coordinates.

4.2 IMPLICIT SURFACES AND DISCRETE LAPLACIAN

To encapsulate the local surface shape, each point v_i is projected onto higher dimension space using MLPs, and the implicit surface is defined on the latent space z as shown on the top of Figure 2. For one layer MLP, this is $z_i = relu(\overrightarrow{w} \cdot \overrightarrow{v_i} + b)$, where $\overrightarrow{w}$ and b are learnable parameters in our network. To calculate the $\delta-$ coordinates, the closed simple surface curve around v_i and its immediate neighbors in the Euclidean space (illustrated as green circles with blue outline in Figure 2) is used to evaluate the Equation (5). In the latent space, the local surface shape is encoded as an implicit surface. The direction of the differential coordinate vector, as defined in Equation (5), approximates the local normal direction. Following Taubin (1995), the discrete Laplacian of implicit surface signal z given by the weighted average over the neighborhood is represented as:

$$\overrightarrow{\delta_i^{implicit}} = \Sigma_{j \in N_{Euclidean}(i)} \frac{\exp(-||\overrightarrow{v_i} - \overrightarrow{v_j}||)(\overrightarrow{z_i} - \overrightarrow{z_j})}{\Sigma_{j \in N_{Euclidean}(i)} \exp(-||\overrightarrow{v_i} - \overrightarrow{v_j}||)}. \tag{5}$$

To compare the two $\delta-$ coordinate representations, we highlight the error in decibel (dB) between the pressure fields reconstructed from groundtruth spherical harmonics term and the predicted ones using different neural networks in Table 2. We observe that $\delta^{implicit}-$ coordinates result in smaller errors. This indicates that our formulation provides a better approximation of ASFs.

4.3 NEURAL NETWORK DESIGN

Our neural network takes the point cloud as an $N \times 3$ input where N represents the number of points in the point cloud. The output is the spherical harmonic coefficient vector of length 16. For each point x_i, four layers of shared-MLP are applied to encode the implicit surface as demonstrated in the top left of Figure 2. Moreover, its K neighbors in the Euclidean space $N_{Euclidean}(i)$ together with the center point x_i, are fed into the implicit surface encoder and form a higher dimensional representation of the center point in the latent space z_i. Next, the discrete Laplacian defined in Equation (6) is evaluated to estimate the implicit surface in the closed simple surface curve around the given point in the latent space $N_{latent}(i)$. The final representation of the center point, illustrated in the left bottom of Figure 2, for point z_i, is defined as:

$$\begin{aligned} feature(v_i) = \Bigg(& z_i^T, \frac{\varphi(\overrightarrow{v_i} - \overrightarrow{v_0})(\overrightarrow{z_i} - \overrightarrow{z_0})}{\Sigma_{j \in N_{Euclidean}(i)} \exp(-||\overrightarrow{v_i} - \overrightarrow{v_j}||)}^T, \\ & \frac{\varphi(\overrightarrow{v_i} - \overrightarrow{v_1})(\overrightarrow{z_i} - \overrightarrow{z_1})}{\Sigma_{j \in N_{Euclidean}(i)} \varphi(\overrightarrow{v_i} - \overrightarrow{v_j})}^T, \ldots, \frac{\varphi(\overrightarrow{v_i} - \overrightarrow{v_n})(\overrightarrow{z_i} - \overrightarrow{z_n})}{\Sigma_{j \in N_{Euclidean}(i)} \varphi(\overrightarrow{v_i} - \overrightarrow{v_j})}^T \\ & \frac{\varphi(\overrightarrow{z_i} - \overrightarrow{z_0})(\overrightarrow{z_i} - \overrightarrow{z_0})}{\Sigma_{j \in N_{latent}(i)} \varphi(\overrightarrow{z_i} - \overrightarrow{z_j})}^T, \ldots, \frac{\varphi(\overrightarrow{z_i} - \overrightarrow{z_n})(\overrightarrow{z_i} - \overrightarrow{z_n})}{\Sigma_{j \in N_{latent}(i)} \varphi(\overrightarrow{z_i} - \overrightarrow{z_j})}^T \Bigg). \end{aligned} \tag{6}$$

The shape of the final per-point feature representation for point v_i is $(1+2k, 128)$, where $k = 5$ is the number of the nearest neighbors and 128 is the dimension of the latent space. The implicit surface representation, as in Equation (6), is further fed into the MLP, forming the differential coordinates in Equation (5), and global pooling is applied to extract the global features. The global features regress the predicted spherical harmonic term c_l^m using fully connected layers.

4.4 DISCRETE LAPLACIAN AND PERMUTATION INVARIANCE

In this section we provide the proof of permutation invariant property in Equation (6). Without loss of generality, we only discuss the function composed of vectors in the high dimensional latent space projected by the implicit surface encoder shown on the top of Figure 2. Let $f_i(v_1, \ldots, v_N) = \Sigma_{j \in N_{Euclidean(i)}} \varphi(-||MLPs(v_i) - MLPs(v_j)||)$, we would like to show that f_i is **injective** and has **continuous inverse mapping**. The domain of f_i is a compact set, and thereby the image of f_i is a compact set as well. Finally f_i is homeomorphism between its domain and its image.

Without loss of generality, consider $\varphi(v) := [1, v, v^2, \cdots, v^M], \varphi : \mathbb{R} \to \mathbb{R}^{M+1}$, where M is the dimension of our implicit surface encoder space.

$$\begin{aligned} f_i(v_1, \ldots, v_N) &= \big[|N_i|, \sum_{m \in N_i} v_m, \sum_{m \in N_i} v_m^2, \cdots, \sum_{m \in N_i} v_m^M\big] \\ &= \big[|N_i|, p_1(v_1, v_2, \cdots, v_M), p_2(v_1, v_2, \cdots, v_M), \cdots, p_M(v_1, v_2, \cdots, v_M)\big] \end{aligned} \tag{7}$$

,

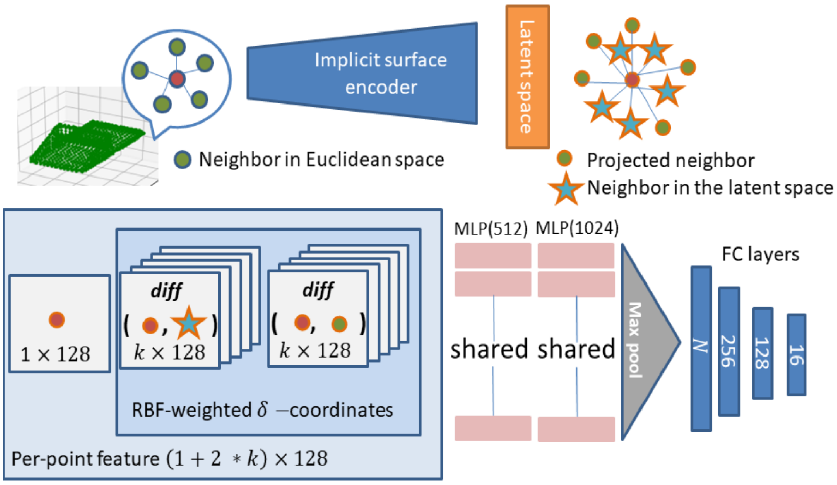

Figure 2: **Network architecture:** The input of our neural network is a $N \times 3$ point cloud ($N = 1024$) and the output of our network is the spherical harmonic coefficients as a vector of length 16 (up to 3rd order spherical harmonics). For each point, marked as red circle with blue outline, we incorporate $\delta-$ coordinates to learn the local shape around the point through a encoder. Next, we continue to find the local shape around the point, marked as red circle with yellow outline. Finally we incorporate the $\delta-$ coordinates both in the latent space and in the Euclidean space to represent the point feature of the given point. By applying 2 MLP layers, we leverage the geometric details and output the predicted spherical harmonic coefficients.

where p is the *power-sums*. According to Lemma 4.1, f_i is injective. Then by Lemma A.2, f_i has inverse continuous mapping.

Lemma 4.1. *Let* $\mathbb{X} = \{(x_1, \cdots, x_M) \in [0,1]^M : x_1 \leq x_2 \leq \cdots \leq x_M\}$. *The sum-of-power mapping* $\mathbb{E} : \mathbb{X} \to \mathbb{R}^{M+1}$ *defined by the coordinate functions* $:E_q(X) := \sum_{m=1}^{M}(x_m)^q, q = 0, \cdots, M$ *is injective.*

4.5 DE FINETTI THEOREM AND IMPLICIT SURFACE FUNCTIONS

The definition of the theorem can be found in Appendix B. Our idea is to approximate the *pressure field* defined by a set of points. Thanks to the continuity nature of neural network, we replace $P(d\theta)$ in a more general form with $P(\theta)d\theta$ since we assume that the underlying latent layer is *absolute continuous* according to Radon–Nikodym theorem (i.e., the distribution on θ has density).

Lemma 4.2. *The acoustic scattering function of a given set of points* $\mathcal{X}$ *and the underlying latent space* z *can be formulated in the form:* $f_{scatter}(\mathcal{X}) = \int dz \int \prod_{i=1}^{n} p(x_i|\theta)P(\theta|z)d\theta$.

5 EVALUATION

5.1 DATA GENERATION AND TRAINING

We need a sufficiently large dataset for training and evaluating our method. We sample 100,000 3D objects from the *ABC Dataset* (Koch et al., 2019). All mesh files are pre-processed to be randomly rotated, scaled between $1 \sim 2m$ and centered. The *FastBEM Acoustics* [1] is used to generate accurate acoustic scattering fields. Specifically, we simulate a plane wave travelling to the $-x$ axis direction, and all objects are assumed to have a zero Neumann boundary condition to solve Equation (2). While other boundary conditions are possible, we do not extensively study this variable in this work. After the acoustic scattering field has been solved at frequencies bands of $\{125, 250, 500, 1000\}Hz$, we use the *pyshtools* [2] package (Wieczorek & Meschede, 2018) to project the directional field to spherical harmonic coefficients c_l^m. By using a maximum order of $L = 3$, we are able to maintain

[1] https://www.fastbem.com/
[2] https://shtools.oca.eu/shtools/public/index.html

a relative projection error below 2%. The original object meshes are also converted to point clouds using furthest sampling for 1024 points. The data generation pipeline takes about 12 days using a desktop with 32 Intel(R) Xeon(R) Gold 5218 CPUs.

We split our dataset into training, validation and test set following a $8 : 1 : 1$ ratio. Our neural network is trained on a NVIDIA GeForce RTX 2080 Ti GPU and takes less than 1ms per object for inference. The Adam algorithm is used in optimizer and it decays exponentially with decay rate and decay step equal to 0.9 and $10\times$ #training examples, respectively. Our code and data will be released upon publication.

5.2 Results and Comparisons

We evaluate our network performance on unseen objects from the *ABC dataset*, perform ablation study on RBF-weighted $\delta-$ coordinates and implicit surface encoders, and show that our network is robust to Gaussian noise ($\sim \mathcal{N}(0, 0.05)$). We compare our results with PointNet (Charles et al., 2017) and DGCNN (Wang et al., 2019) in Table 2 and Table 3. We justify the design of our network with this ablation study, including the use of $\delta-$ coordinates, RBF-weighted function as well as the implicit surface encoder, as highlighted in Figure 2. We use PointNet, where only per-point MLP layers are applied on each point in the point cloud, and DGCNN, which dynamically build graphs according to K nearest neighbors, as the baseline. RBF-weighted $\delta-$ coordinates (as described in Equation (5)), uniformly weighted $\delta-$ coordinates (described in Equation (4)), and implicit surface encoder shown at the top of Figure 2 are considered as subjects of ablation studies. We observe that our fine-grained geometric feature representation in Equation (6) results in larger reduction in dB error. Moreover, our discrete Laplacian based representation is more robust to the Gaussian noise in the test set compared to DGCNN.

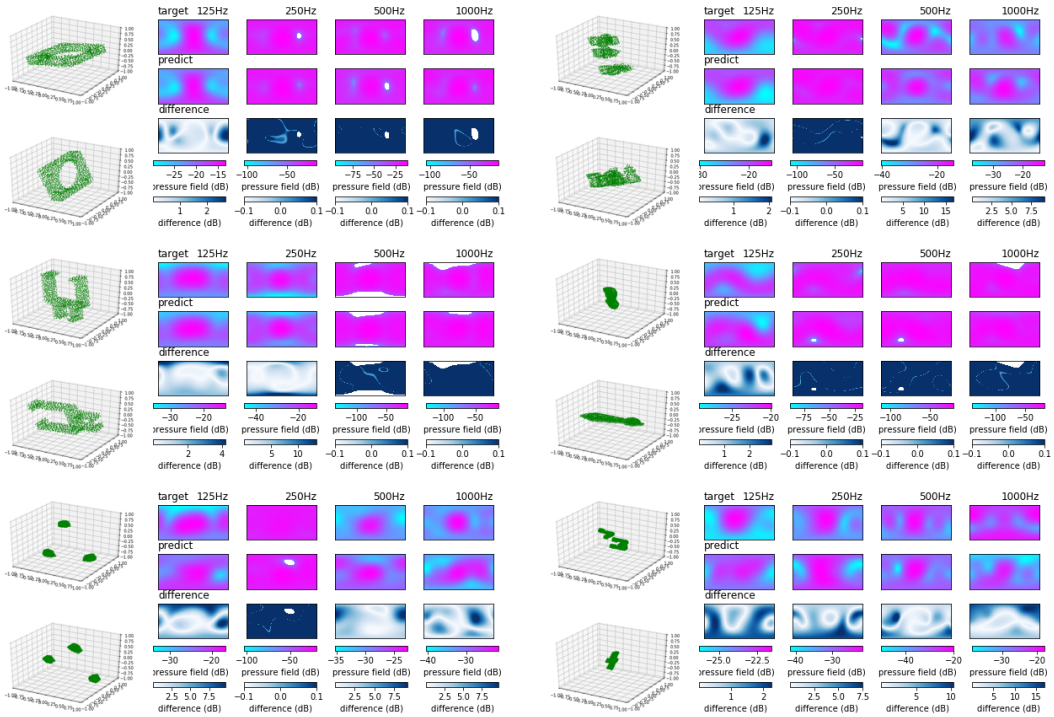

Figure 3: **ASF prediction and groundtruth comparison:** We expand the predicted spherical harmonics coefficients onto a latitude-longitude (lat-long) map, representing the directional distribution of the scattering field. We randomly choose 6 point clouds from the test dataset, which have not been seen during training. Each point cloud is visualized from two viewing angles, followed by their groundtruth/target, predicted, and error/difference lat-long maps in four frequency bands. We see a good match between our predicted fields and the groundtruth fields.

Table 2: **Ablation study:** In this evaluation, we compare the performance on uniform $\delta-$ coordinate in Eq. (4), weighted $\delta-$ coordinates in Eq. (5) and implicit surface estimation in Eq. (6) on our test dataset (including $10k$ objects) at frequency bands {125Hz, 250Hz, 500Hz, 1000Hz}. The best result for each frequency is highlighted in **bold** (lower error is better). We alter between choosing Eq. (4) and Eq. (5) that results in four different combinations: Row 1 and 2, 3 and 4. Next, we experiment the use of implicit surface encoder (Row 3 and 4). Our proposed network design (Row 4) shows superior performance in terms of ASF approximation for all four frequencies. A lower value indicates a better result.

Methods {RBF-weighted δ-coord — implicit surface}	Error in dB				#Params
	125Hz	250Hz	500Hz	1000Hz	
{✗— ✗}	3.49	3.56	3.71	4.23	30k
{✓— ✗}	3.38	3.41	3.57	4.47	30k
{✗— ✓}	3.28	3.38	3.52	3.85	30k
{✓— ✓} (ours)	**2.98**	**3.06**	**3.22**	**3.76**	34k
PointNet (Charles et al., 2017)	3.66	3.44	3.43	4.16	80k
DGCNN (Wang et al., 2019)	3.71	3.56	3.59	4.21	103k

Table 3: **Robustness test**: In this robustness test, we add i.i.d. noise to every point in the point cloud, which follows $\mathcal{N}(0, 0.05)$, to the test set. Note that DGCNN finds $K-$ neighbors at each iteration as we did. However, our algorithm is more robust to the noisy data input thanks to the design of the shape Laplacian based point representation.

Methods	Error in dB			
	125Hz	250Hz	500Hz	1000Hz
PointNet	5.88	5.50	6.02	8.37
DGCNN	7.91	6.65	7.31	11.32
Ours	**5.20**	**5.09**	**4.70**	**5.51**

5.3 Application in Interactive Sound Rendering

One direct application of ASF is sound rendering. Being able to infer ASFs from shapes fast enables us to use wave-acoustic simulation in an interactive context. State-of-the-art interactive sound rendering systems usually employ the geometric ray-tracing algorithm (Schissler & Manocha, 2017), which is a great simplification of wave-acoustics in the high frequency range. We follow a similar hybrid sound rendering approach as in Rungta et al. (2018), where low-frequency scattering is modeled by wave-acoustics and high-frequency scattering is handled by geometric acoustics. This way of coupling allows us to handle even deforming or dynamic objects in real-time due to our efficient computation of ASF. We demonstrate sound rendering applications in our supplemental video.

6 Conclusion and Limitations

We present a novel geometric deep learning approach that infers the 3D acoustic scattering field induced by an object from its shape. Our algorithm uses discrete-laplacian based implicit function encoders to better capture the geometric properties of an object and achieves lower error compared with existing general point-based frameworks. Our shape and sound field representation are well-suited for applications that require real-time wave acoustics characteristics.

Currently we study the case of sound-hard surfaces and a limited range of simulation frequencies, but these are not an inherent limitation of our method. The simulation time needed to create a large training set can grow as a cubic function of the sound frequency. Incorporating more types of sound surfaces can further increase the training overhead of all geometric deep learning methods, including ours. In the future, we propose to extend these methods to real-world objects and use them for localization and non-line-of-sight computations.

REFERENCES

Jakob Abeßer. A review of deep learning based methods for acoustic scene classification. *Applied Sciences*, 10(6), 2020.

Inkyu An, Doheon Lee, Jung-woo Choi, Dinesh Manocha, and Sung-eui Yoon. Diffraction-aware sound localization for a non-line-of-sight source. In *2019 International Conference on Robotics and Automation (ICRA)*, pp. 4061–4067. IEEE, 2019.

Michael J Bianco, Peter Gerstoft, James Traer, Emma Ozanich, Marie A Roch, Sharon Gannot, and Charles-Alban Deledalle. Machine learning in acoustics: Theory and applications. *The Journal of the Acoustical Society of America*, 146(5):3590–3628, 2019.

Dick Botteldooren. Finite-difference time-domain simulation of low-frequency room acoustic problems. *The Journal of the Acoustical Society of America*, 98(6):3302–3308, 1995.

R. Qi Charles, Hao Su, Mo Kaichun, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul 2017. doi: 10.1109/cvpr.2017.16. URL http://dx.doi.org/10.1109/CVPR.2017.16.

Manfredo P Do Carmo. *Differential geometry of curves and surfaces: revised and updated second edition*. Courier Dover Publications, 2016.

Ivan Dokmanić, Reza Parhizkar, Andreas Walther, Yue M Lu, and Martin Vetterli. Acoustic echoes reveal room shape. *Proceedings of the National Academy of Sciences*, 110(30):12186–12191, 2013.

James Eaton, Nikolay D Gaubitch, Alastair H Moore, Patrick A Naylor, James Eaton, Nikolay D Gaubitch, Alastair H Moore, Patrick A Naylor, Nikolay D Gaubitch, James Eaton, et al. Estimation of room acoustic parameters: The ace challenge. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 24(10):1681–1693, 2016.

Ziqi Fan, Vibhav Vineet, Hannes Gamper, and Nikunj Raghuvanshi. Fast acoustic scattering using convolutional neural networks. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 171–175. IEEE, 2020.

Andrea F Genovese, Hannes Gamper, Ville Pulkki, Nikunj Raghuvanshi, and Ivan J Tashev. Blind room volume estimation from single-channel noisy speech. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 231–235. IEEE, 2019.

Rana Hanocka, Amir Hertz, Noa Fish, Raja Giryes, Shachar Fleishman, and Daniel Cohen-Or. Meshcnn: a network with an edge. *ACM Transactions on Graphics (TOG)*, 38(4):1–12, 2019.

David Heath and William Sudderth. De finetti's theorem on exchangeable variables. *The American Statistician*, 30(4):188–189, 1976. doi: 10.1080/00031305.1976.10479175. URL https://www.tandfonline.com/doi/abs/10.1080/00031305.1976.10479175.

Taeyoung Kim, Youngsun Kwon, and Sung-Eui Yoon. Real-time 3-d mapping with estimating acoustic materials. In *2020 IEEE/SICE International Symposium on System Integration (SII)*, pp. 646–651. IEEE, 2020.

Sebastian Koch, Albert Matveev, Zhongshi Jiang, Francis Williams, Alexey Artemov, Evgeny Burnaev, Marc Alexa, Denis Zorin, and Daniele Panozzo. Abc: A big cad model dataset for geometric deep learning. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

Heinrich Kuttruff. *Room Acoustics*. Taylor & Francis Group, London, U. K., 6th edition, 2016.

Jiaxin Li, Ben M. Chen, and Gim Hee Lee. So-net: Self-organizing network for point cloud analysis. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun 2018a. doi: 10.1109/cvpr.2018.00979. URL http://dx.doi.org/10.1109/CVPR.2018.00979.

Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. In *Advances in neural information processing systems*, pp. 820–830, 2018b.

David B Lindell, Gordon Wetzstein, and Vladlen Koltun. Acoustic non-line-of-sight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6780–6789, 2019.

Shichen Liu, Shunsuke Saito, Weikai Chen, and Hao Li. Learning to infer implicit surfaces without 3d supervision. In *Advances in Neural Information Processing Systems*, pp. 8293–8304, 2019.

Ravish Mehra, Atul Rungta, Abhinav Golas, Ming Lin, and Dinesh Manocha. Wave: Interactive wave-based sound propagation for virtual environments. *IEEE transactions on visualization and computer graphics*, 21(4):434–442, 2015.

Hsien-Yu Meng, Lin Gao, Yu-Kun Lai, and Dinesh Manocha. Vv-net: Voxel vae net with group convolutions for point cloud segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 8500–8508, 2019.

Kaichun Mo, Paul Guerrero, Li Yi, Hao Su, Peter Wonka, Niloy J. Mitra, and Leonidas J. Guibas. Structurenet. *ACM Transactions on Graphics (TOG)*, 38(6):1–19, Nov 2019. ISSN 1557-7368. doi: 10.1145/3355089.3356527. URL http://dx.doi.org/10.1145/3355089.3356527.

Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda, and Michael M Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5115–5124, 2017.

Nicolas Morales and Dinesh Manocha. Optimizing source placement for noise minimization using hybrid acoustic simulation. *Computer-Aided Design*, 96:1–12, 2018.

Nicolas Morales, Ravish Mehra, and Dinesh Manocha. A parallel time-domain wave simulator based on rectangular decomposition for distributed memory architectures. *Applied Acoustics*, 97: 104–114, 2015.

Nicolas Morales, Zhenyu Tang, and Dinesh Manocha. Receiver placement for speech enhancement using sound propagation optimization. *Applied Acoustics*, 155:53–62, 2019.

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 Conference on Computer Vision and Pattern Recognition*, pp. 165–174, 2019.

Ville Pulkki and U Peter Svensson. Machine-learning-based estimation and rendering of scattering in virtual reality. *The Journal of the Acoustical Society of America*, 145(4):2664–2676, 2019.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems*, pp. 5099–5108, 2017.

Nikunj Raghuvanshi and John Snyder. Parametric wave field coding for precomputed sound propagation. *ACM Trans. Graph.*, 33(4):38:1–38:11, July 2014. ISSN 0730-0301. doi: 10.1145/2601097.2601184. URL http://doi.acm.org/10.1145/2601097.2601184.

David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

Atul Rungta, Carl Schissler, Nicholas Rewkowski, Ravish Mehra, and Dinesh Manocha. Diffraction kernels for interactive sound propagation in dynamic environments. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1613–1622, 2018.

Carl Schissler and Dinesh Manocha. Interactive sound propagation and rendering for large multi-source scenes. *ACM Transactions on Graphics (TOG)*, 36(1):2, 2017.

Louena Shtrepi, Arianna Astolfi, Sönke Pelzer, Renzo Vitale, and Monika Rychtáriková. Objective and perceptual assessment of the scattered sound field in a simulated concert hall. *The Journal of the Acoustical Society of America*, 138(3):1485–1497, 2015.

Vishwanath A. Sindagi, Yin Zhou, and Oncel Tuzel. Mvx-net: Multimodal voxelnet for 3d object detection. *2019 International Conference on Robotics and Automation (ICRA)*, May 2019. doi: 10.1109/icra.2019.8794195. URL http://dx.doi.org/10.1109/ICRA.2019.8794195.

Dmitriy Smirnov, Matthew Fisher, Vladimir G. Kim, Richard Zhang, and Justin Solomon. Deep parametric shape predictions using distance fields, 2019.

Olga Sorkine. Differential representations for mesh processing. In *Computer Graphics Forum*, volume 25, pp. 789–807. Wiley Online Library, 2006.

Auston Sterling, Justin Wilson, Sam Lowe, and Ming C Lin. Isnn: Impact sound neural network for audio-visual object classification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 555–572, 2018.

Qingyang Tan, Lin Gao, Yu-Kun Lai, Jie Yang, and Shihong Xia. Mesh-based autoencoders for localized deformation component analysis. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

Zhenyu Tang, Nicholas J Bryan, Dingzeyu Li, Timothy R Langlois, and Dinesh Manocha. Scene-aware audio rendering via deep acoustic analysis. *IEEE Transactions on Visualization and Computer Graphics*, 2020.

Gabriel Taubin. A signal processing approach to fair surface design. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 351–358. ACM, 1995.

Lonny L Thompson. A review of finite-element methods for time-harmonic acoustics. *The Journal of the Acoustical Society of America*, 119(3):1315–1330, 2006.

Dimitrios Tsokaktsidis, Timo Von Wysocki, Frank Gauterin, and Steffen Marburg. Artificial neural network predicts noise transfer as a function of excitation and geometry. In *Proc. International Congress on Acoustics (ICA)*, pp. 4392–4396, 2019.

Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics*, 38(5):1–12, Oct 2019. ISSN 0730-0301. doi: 10.1145/3326362. URL http://dx.doi.org/10.1145/3326362.

Mark A Wieczorek and Matthias Meschede. Shtools: Tools for working with spherical harmonics. *Geochemistry, Geophysics, Geosystems*, 19(8):2574–2592, 2018.

Luiz C Wrobel and AJ Kassab. Boundary element method, volume 1: Applications in thermo-fluids and acoustics. *Appl. Mech. Rev.*, 56(2):B17–B17, 2003.

Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1912–1920, 2015.

Qiangeng Xu, Weiyue Wang, Duygu Ceylan, Radomir Mech, and Ulrich Neumann. Disn: Deep implicit surface network for high-quality single-view 3d reconstruction. In *Advances in Neural Information Processing Systems*, pp. 490–500, 2019.

Li Yi, Hao Su, Xingwen Guo, and Leonidas J Guibas. Syncspeccnn: Synchronized spectral cnn for 3d shape segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2282–2290, 2017.

Zhoutong Zhang, Jiajun Wu, Qiujia Li, Zhengjia Huang, James Traer, Josh H McDermott, Joshua B Tenenbaum, and William T Freeman. Generative modeling of audible shapes for object perception. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1251–1260, 2017.

Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun 2018. doi: 10.1109/cvpr.2018.00472. URL http://dx.doi.org/10.1109/CVPR.2018.00472.

Appendices

A PERMUTATION INVARIANCE

Proof for Lemma 4.1: For some $u, v \in \mathbb{X}$, $E(u) = E(v)$, show $u = v$.

Let u, v be the roots of polynomial of degree M, we have:$f_u(x) = \prod_{m=1}^{M}(x - u_m) = x^M + a_1 x^{M-1} + \cdots + a_{M-1}x + a_M$ and $f_v(x) = \prod_{m=1}^{M}(x - v_m) = x^M + b_1 x^{M-1} + \cdots + b_{M-1}x + b_M$.

a, b are elementary symmetric polynomials.. Elementary polynomials(i.e. a, b) can be expressed by power sums(i.e $E(u), E(v)$).

$$
\begin{aligned}
a_M &= \frac{1}{M!} \cdot \begin{vmatrix} E_1(u) & 1 & 0 & 0 & \cdots & 0 \\ E_2(u) & E_1(u) & 2 & 0 & \cdots & 0 \\ E_{M-1}(u) & E_{M-2}(u) & E_{M-3}(u) & E_{M-4}(u) & \cdots & M-1 \\ E_M(u) & E_{M-1}(u) & E_{M-2}(u) & E_{M-3}(u) & \cdots & E_1(u) \end{vmatrix} \\
b_M &= \frac{1}{M!} \cdot \begin{vmatrix} E_1(v) & 1 & 0 & 0 & \cdots & 0 \\ E_2(v) & E_1(v) & 2 & 0 & \cdots & 0 \\ E_{M-1}(v) & E_{M-2}(v) & E_{M-3}(v) & E_{M-4}(v) & \cdots & M-1 \\ E_M(v) & E_{M-1}(v) & E_{M-2}(v) & E_{M-3}(v) & \cdots & E_1(v) \end{vmatrix}.
\end{aligned} \tag{8}
$$

$E(u) = E(v)$ implies that $a = b$, and therefore $u = v$.

Theorem A.1. *The function $f : \mathbb{C}^M \to \mathbb{C}^M$, $f(c) = x^M + c_1 x^{M-1} + \cdots + (-1)^{M-1} c_{M-1} x + (-1)^M c_M, c \in \mathbb{C}^M$, is homeomorphism.*

Lemma A.2. *The sum-of-power mapping defined in Lemma. 4.1 has continuous inverse mapping.*

The domain of E is a compact set, and E is a continuous function. Therefore, the image of E is a compact set.

From Lemma. 4.1, a is a continuous function of the power sums E.

Goal: Show continuity of inverse mapping of E.

Proof: From Theorem. A.1, the continuity of roots u depends on the coefficients a. W.L.G, show continuity from a to u. Due to the nature of homeomorphism, the mapping from u to a as well as its inverse mapping from a to u are both continuous.

B DE FINETTI THEOREM AND IMPLICIT SURFACE FUNCTIONS

Theorem B.1. *a sequence $(x_1, x_2, \ldots, x_n)$ is infinitely exchangeable iff*

$$
p(x_1, \ldots, x_n) = \int \prod_{i=1}^{n} p(x_i|\theta) P(d\theta), \tag{9}
$$

for some measure P on θ