APPROXIMATE EQUIVARIANCE NEEDLET CONVOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper develops a new spherical neural network framework based on needlet convolutions that can capture the multiscale information and rotation invariant feature of spherical data. The spherical needlets are a wavelet tight frame on the sphere $\mathbb{S}^2$ which can transform the data into a frequency domain with different scales: the low-pass and high-passes that extract approximate and detail information from the spherical signal. As the output signal of the spherical needlet convolution lies on the rotation group SO(3), we generalize the needlets to SO(3) and define SO(3) needlet convolution. Wavelet shrinkage is used as a nonlinear activation to reduce the redundancy in the needlet high-pass representation, which enhances the robustness of the neural network. The $\mathbb{S}^2$ needlet convolution can be connected with multiple SO(3) needlet convolution layers to form a Needlet Approximate Equivariance Spherical Neural Network, thus providing a powerful framework to distill the geometric equivariance feature and trainable multiresolution analyzer. Experimental results on quantum chemistry regression and gravitational wave parameter estimation show its great potential for solving scientific challenges.

1 INTRODUCTION

Geometric deep learning (Bronstein et al., 2021; 2017) has gained a great triumph in advancing scientific discovery (Atz et al., 2021; Baek et al., 2021; Davies et al., 2021; Méndez-Lucio et al., 2021; Townshend et al., 2021). It provides a universal blueprint for learning stable representation of high-dimensional data to build equivariant or invariant neural network layers which respect exact or approximate symmetries of input data, such as translation, rotation and permutation (Cohen & Welling, 2016; Qi et al., 2017; Cohen et al., 2018; Fuchs et al., 2020; Bronstein et al., 2021). Many data types in the real world can be modelled as spherical data, such as omnidirectional images, 3D LiDAR scans from autonomous cars, molecules, cosmic microwave background, climate change data on the earth. However, simply mapping spherical signals to $\mathbb{R}^2$ and then applying planar convolutions will result in distorted signals and ineffective shift equivariance. In this paper, we develop a computational strategy to extract and learn multiscale features from spherical data while preserving the rotation equivariance for the network. The multiscale decomposition of the geometric data in the network learning provides a robust **N**eedlet Approximate **E**quivariance **S**pherical Neural Network (**NES**) that can increase the scalability of the geometric deep learning, and reduce the redundancy and noise via wavelet shrinkage in the needlet transforms.

Spherical needlets are a wavelet-like system on the two dimensional sphere $\mathbb{S}^2$ that form a tight frame (Narcowich et al., 2006b;a; Wang et al., 2017). They provide a multiscale representation of signals on the sphere. The spherical needlet convolution turns the spherical data to a signal on the rotation group SO(3). To offer a multiscale representation for the output of the $\mathbb{S}^2$ needlet convolution, we develop a needlet system on SO(3), from which we then design SO(3) needlet convolution. This SO(3) needlet convolution can be repeatedly used in the network, and with $\mathbb{S}^2$ needlet convolution, constitutes a deep neural network that can distill the geometric equivariant feature of spherical data. To further increase the robustness of the needlet convolution network, we apply wavelet shrinkage (Donoho, 1995; Baldi et al., 2009) as a nonlinear activation (Zheng et al., 2021) in the needlet transforms, which filters out the high-pass information in the framelet domain. The needlet convolution with shrinkage activation gains provably approximate equivariance, and the error diminishes as the scale is sufficiently high.

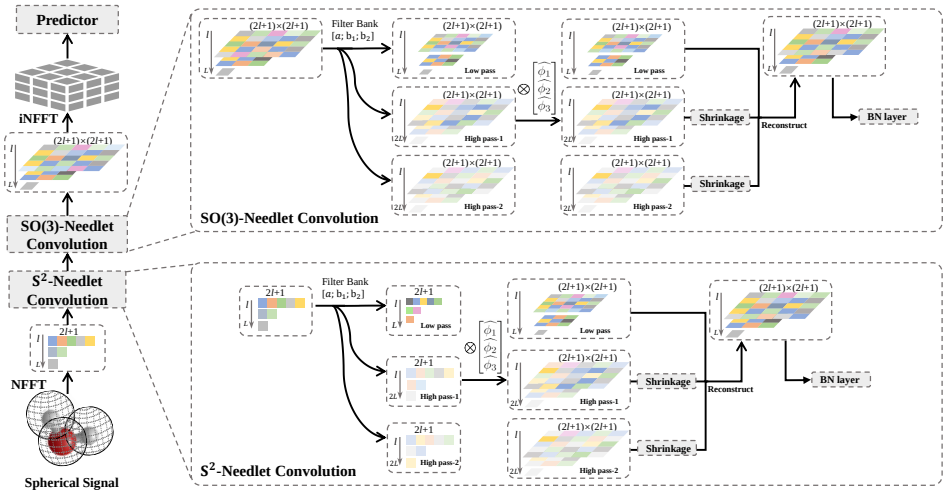

Figure 1: The framework of our needlet spherical neural network. As the left column shows, we first carry on non-equispaced FFT with predefined weights on the spherical signal. The following are an $\mathbb{S}^2$-Needlet Convolution and repeated $\mathrm{SO}(3)$-Needlet Convolutions. Then we use inverse non-equispaced FFT over the output of the $\mathrm{SO}(3)$-Needlet Convolution and feed them into the downstream predictor. The right part shows the intermediate computing process of the convolutions with wavelet shrinkage on high-passes and Batch Normalization layers.

Our proposed needlet network model achieve state-of-the-art performance for real-world scientific problems of regressing quantum chemistry molecules and estimating a cosmological parameter of gravitational wave. Figure 1 shows the framework of our **NES** for the application in bio-molecular prediction, where the input is a set of spherical signals centered at each atom in the molecule. The framework is scalable to any application scenarios that can be represented by spherical signals.

2 Spherical Needlet Framework

2.1 Multiresolution Analysis of Spherical Needlets

Needlets are a form of spherical wavelets which enjoy good localization properties in both spatial and harmonic space and guarantee the uniqueness of the wavelet coefficients. There are two main merits to introduce needlet representations: computationally convenient calculation and multiscale representations of approximation and detail information. Needlets on $\mathcal{M}$ ($\mathcal{M}$ takes $\mathbb{S}^2$ or $\mathrm{SO}(3)$ in our model) is defined by a filter bank $\eta := \{a; b^1, \ldots, b^r\} \subset l_1(\mathbb{Z}) := \{h = \{h_k\}_{k\in\mathbb{Z}} \subset \mathbb{C} : \sum_{k\in\mathbb{Z}} |h_k| < \infty\}$ and a set of associated scaling functions, $\Psi = \{\alpha; \beta^1, \cdots, \beta^r\} \subset L_1(\mathbb{R})$. Filter a is also called low-pass filter which distills approximate information of the signal, and filers $\{b^1, \cdots, b^r\}$ are called high-pass filters which distill more detail and even noise information. The associated generating functions and filter bank satisfy the relations in Eq. (1), where $n = 1, \ldots, r$, and $\xi \in \mathbb{R}$.

$$\widehat{\alpha}(2\xi) = \widehat{a}(\xi)\widehat{\alpha}(\xi), \quad \widehat{\beta^n}(2\xi) = \widehat{b^n}(\xi)\widehat{\alpha}(\xi). \tag{1}$$

To discretize the continuous needlets with zero numerical error, we utilize Polynomial-exact Quadrature Rule that are generated by the tensor product of the Gauss-Legendre nodes on the interval $[-1, 1]$ and equi-spaced nodes in the other dimension with non-equal weights. Let $v_{j,k}$ represent low-pass coefficients, and $w^n_{j,k}$ represent high-pass coefficients of the signal function f. They are defined as the inner products of (low-pass / high-pass) needlets and f. In the implementation of our model, we simply calculate the coefficients in the Fourier space, as shown in Eq. (2), which can be computed fast and conveniently. The $\widehat{f_\ell}$ is the generalized Fourier coefficients of f at degree ℓ. More details of the needlets are given in Appendix A.2.

$$\widehat{v}_{j,\ell} = \widehat{f_\ell}\overline{\widehat{\alpha}\left(\frac{\lambda_\ell}{2^j}\right)}, \quad \widehat{w}^n_{j-1,\ell} = \widehat{f_\ell}\overline{\widehat{\beta^n}\left(\frac{\lambda_\ell}{2^{j-1}}\right)}. \tag{2}$$

2.2 ROTATION-EQUIVARIANT CONVOLUTION

A neural network (or function) Φ is called *rotation-equivariant* if for any rotation operator L_R there exists some operator T_R such that $\Phi \circ L_R = T_R \circ \Phi$. A rotation-equivariant neural network can predict the output of rotated signals with theoretical guarantee more efficiently and precisely than without rotation-equivariance. Formally, the rotation equivariant convolution over $\mathbb{S}^2$ and $\mathrm{SO}(3)$ is defined as $[\phi \star f](R) = \langle L_R\phi, f\rangle = \int_{\mathcal{M}} \phi(R^{-1}x)f(x)dx$, where f is a signal, ϕ is a learnable locally supported filter, $L_R\phi(x) = \phi(R^{-1}x)$, and $\mathcal{M}$ represents $\mathbb{S}^2$ or $\mathrm{SO}(3)$. It is provable that the convolution defined above satisfies a Fourier theorem: $[\widehat{\phi \star f}]_\ell = \widehat{f}_\ell \cdot \widehat{\phi}_\ell^\dagger$, where $\dagger$ denotes the conjugate transpose operation and ℓ is the degree parameter. The operation $\cdot$ denotes matrix multiplication for the case of $\mathcal{M} = \mathrm{SO}(3)$ and outer product for $\mathcal{M} = \mathbb{S}^2$. Spherical CNN (Cohen et al., 2018) uses this formula in its convolution. Different with the existing models, we apply convolution on the needlet coefficients of a spherical signal.

Using the framelet system described in Section 2.1, we take $n = 2$ and get three sets of needlets coefficients $\{\widehat{v}_{1,\ell}\}_{\ell=1}^{\Lambda_{J_0}}$, $\{\widehat{w}^1_{1,\ell}\}_{\ell=1}^{\Lambda_{J_1}}$ and $\{\widehat{w}^2_{1,\ell}\}_{\ell=1}^{\Lambda_{J_1}}$. The Fourier coefficients of signal f of degree ℓ can be reconstructed by the framelet coefficients, and we denote this relation as $\left[\widehat{v}_{1,\ell}, \widehat{w}^1_{1,\ell}, \widehat{w}^2_{1,\ell}\right]^\top \asymp \widehat{f}_\ell$, where $\asymp$ means formal equivalence. Then the following relation holds:

$$\begin{bmatrix}\widehat{\phi_{1\ell}}\\ \widehat{\phi_{2\ell}}\\ \widehat{\phi_{3\ell}}\end{bmatrix} \odot \widehat{f}_\ell \asymp \begin{bmatrix}\widehat{\phi_{1\ell}}\\ \widehat{\phi_{2\ell}}\\ \widehat{\phi_{3\ell}}\end{bmatrix} \odot \begin{bmatrix}\widehat{v}_{1,\ell}\\ \widehat{w}^1_{1,\ell}\\ \widehat{w}^2_{1,\ell}\end{bmatrix} = \begin{bmatrix}\widehat{\phi_{1\ell}} \cdot \widehat{f}_\ell \overline{\widehat{\alpha}\left(\frac{\lambda_\ell}{2^{J_0}}\right)}\\ \widehat{\phi_{2\ell}} \cdot \widehat{f}_\ell \overline{\widehat{\beta^1}\left(\frac{\lambda_\ell}{2^{J_0}}\right)}\\ \widehat{\phi}_{3,\ell} \cdot \widehat{f}_\ell \overline{\widehat{\beta^2}\left(\frac{\lambda_\ell}{2^{J_0}}\right)}\end{bmatrix} = \begin{bmatrix}[\widehat{\phi_1 \star f}]_\ell \overline{\widehat{\alpha}\left(\frac{\lambda_\ell}{2^{J_0}}\right)}\\ [\widehat{\phi_2 \star f}]_\ell \overline{\widehat{\beta^1}\left(\frac{\lambda_\ell}{2^{J_0}}\right)}\\ [\widehat{\phi_3 \star f}]_\ell \overline{\widehat{\beta^2}\left(\frac{\lambda_\ell}{2^{J_0}}\right)}\end{bmatrix} \asymp [\widehat{\phi \star f}]_\ell.$$

Here $\widehat{\phi}_{i\ell}$ $(i = 1, 2, 3)$ are three learnable filters defined in frequency domain, $\odot$ denotes element-wise multiplication. The above relation illustrates that we make a formally equivalent expression of $[\widehat{\phi \star f}]_\ell$ with multiresolution information and rotation equivariance.

2.3 SHRINKAGE AND POOLING

Shrinkage Function One potential drawback of Spherical CNNs is that the non-linear activation in each layer involves redundant Fourier transforms and then a Fourier transform to draw the features into the frequency domain, thus causing heavy computational cost. To preserve rotation-equivariance to the best extent and also reduce computational complexity, we apply non-linear activation on frequency domain directly, with an estimable and tolerable rotation-equivariance error. Similar to the approach of UFGConv (Zheng et al., 2021), the high-pass coefficients in the frequency domain are cut off by shrinkage thresholding

$$\text{Shrinkage}(x) = \text{sgn}(x)(|x| - \lambda)_+ \quad \forall x \in \mathbb{R},$$

where λ is the threshold value, which is taken as $\lambda = \sigma\sqrt{2\log(N)}/\sqrt{N}$ for N coefficients. The hyperparameter σ is an analogue to the noise level of the denoising model. We only apply the shrinkage on high-pass framelet coefficients and maintain rotation-equivariance in low-pass, which mainly distills the approximate information of input data. Therefore, our model is approximately rotation-equivariant. By the theory of needlets, the shrinkage mechanism introduces a controlable rotaton-equivariance error. The proof of Theorem 2.1 is given in Appendix B.1.

Theorem 2.1. *Let $\mathbb{W}_p^s(\mathbb{S}^2)$ with $s \geq 2/p$ and $1 \leq p \leq \infty$ be a Sobelev space embedded in $\mathbb{L}_p(\mathbb{S}^2)$. For $f \in \mathbb{W}_p^s(\mathbb{S}^2)$, ϕ is a fiter, then rotation-equivariance error brought by Shrinkage function is defined as the maximum of the following* Error *over all $R \in \mathrm{SO}(3)$, which has the convergence order $2^{-(J_0+1)s}$:*

$$\text{Error} := \sum_{\ell=0}^{B} \left\| \text{Shr}(\widehat{L_R f \star \phi})_\ell^{(\mathrm{H})} - D^\ell(R)\text{Shr}(\widehat{f \star \phi}_\ell^{(\mathrm{H})}) \right\|^2 \leq C2^{-(J_0+1)s}, \tag{3}$$

where B is the bandwidth, Shr$(\cdot)$ *denotes Shrinkage function, superscript* (H) *denotes the high-pass coefficients. J_0 is the scale of the low pass, and C is a constant depending only on s and ϕ.*

Pooling Operator To avoid repeated Fourier transform, we apply spectral pooling in the frequency domain. Formally, if the spectral feature has the form that $\hat{f} = [\hat{f}_0 \cdots, \hat{f}_{\ell-1}, \hat{f}_\ell]$, spectral pooling removes coefficients with degree larger than $\ell/2$. It is provable that spectral pooling operator is rotation-equivariant (see Appendix B.2).

3 EXPERIMENTS

Local MNIST Classification We use a modified MNIST classification dataset where the images are projected onto a sphere to get the rotated training and testing sets (R/R). In order to evaluate the ability of the needlet convolutions to capture high-frequency information, we downsample the MNIST digits images into different resolutions and then project them onto a scalable area of the sphere. As shown in Table 1, both models are rotation equivariant. However, the performance of Spherical CNN is declining as the ratio increases. Our model consistently performs well as the shrinkage ratio increases.

Table 1: Classification results on spherical MNIST with different scales. Both training and testing are performed on the rotated data (R/R). The values in the first row denote the shrinkage ratio.

Model	10%	30%	50%	70%	90%
SPHERICAL CNN	94.99	92.17	86.92	83.73	78.71
NES	**97.84**	**97.30**	**96.74**	**95.21**	**92.66**

Table 2: RMSE Comparisons on QM7. The standard deviation is computed over 10 trials. † indicates the method is rotation equivariant. Baseline models are MLP (Montavon et al., 2012), GCN (Kipf & Welling, 2017), Spherical CNN (Cohen et al., 2018) and Clebsch-Gordan Net (Kondor et al., 2018).

Method	RMSE	Params
MLP/RANDOM CM	5.96	-
GCN	7.32 ± 0.23	0.8M
SPHERICAL CNN†	8.47	1.4M
CLEBSCH–GORDAN†	7.97	$\geq$1.1M
NES† (Ours)	**7.21** $\pm$ 0.46	0.9M

Table 3: Performance Comparisons in terms MAE of forces in $\mathrm{meV}/\mathring{A}$ on MD17 dataset. The full names of the molecules from top to bottom are: *Aspirin, Ethanol, Malonaldehyde, Naphthalene, Salicylic, Toluene, Uracil*. **NES** achieves better performance on four types of molecules.

Molecule	sGDML	SchNet	DimeNet	SphereNet	**NES**
Asp.	29.5	58.5	21.6	18.6	**15.2**
Eth.	14.3	16.9	10.0	**9.0**	9.2
Mal.	17.8	28.6	16.6	14.7	**13.6**
Nap.	4.8	25.2	9.3	7.7	**3.5**
Sal.	**12.1**	36.9	16.2	15.6	14.2
Tol.	**6.1**	24.7	9.4	6.7	**6.1**
Ura.	**10.4**	24.3	13.1	11.6	10.8

Molecular Property Prediction We test our model's performance on the molecular property prediction task. Experiments are conducted over QM7 and MD17, as shown in Table 2 and Table 3. In QM7 task, our model has about 0.9M parameters and achieves the lowest RMSE of 7.21 ± 0.46 among all rotation equivariant models. In MD17 task, the baseline models are sGDML (Chmiela et al., 2018), SchNet (Schütt et al., 2017), DimeNet (Chmiela et al., 2018) and SphereNet (Liu et al., 2021). Most of the previous models are 3D graph models with geometric information, highly relying on human expertise and extra annotation. Instead, our model utilizes the adaptive learning of input features and incorporates multiscale analysis to improve the representation ability.

Parameter Estimate for Gravitational Wave The existence of a stochastic Primordial Gravitational Wave Background (PGWB), formed when microscopic quantum fluctuations of the metric were stretched up to super-horizon scales by the sudden expansion of space-time that occurred during inflation (Caprini & Figueroa, 2018) is a common prediction in the majority of inflationary models. We utilize our NES to predict the tensor-to-scalar ratio (r) and compare mean square error (MSE) between two kinds of model. The MSE is 0.0173 for ResNet and 0.0099 for NES. As the needlet spherical neural network provides a multiscale learning framework, which captures high-frequency contents in the needlet coefficients, our overall accuracy is considerably higher.

4 CONCLUSION

We develop a Needlet Approximate Equivariance Spherical Neural Network using multiscale representation systems on the sphere and rotation group. The needlet convolution allows the rotation equivariance for network feature extraction with trainable multiscale analyzer. Wavelet shrinkage is used as a network activation to filter out the high-pass redundancy which helps improve the robustness of the network. The shrinkage brings controllable equivariance error, which can be sufficiently small when the scale is high. Empirical study shows the proposed needlet spherical neural network can achieve excellent prediction performance on scientific problems.

REFERENCES

Kenneth Atz, Francesca Grisoni, and Gisbert Schneider. Geometric deep learning on molecular representations. *Nature Machine Intelligence*, pp. 1–10, 2021.

Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N Kinch, R Dustin Schaeffer, et al. Accurate prediction of protein structures and interactions using a 3-track network. *Science*, 373:871–876, 2021.

Paolo Baldi, Gérard Kerkyacharian, Domenico Marinucci, Dominique Picard, et al. Asymptotics for spherical needlets. *The Annals of Statistics*, 37(3):1150–1171, 2009.

Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, Groups, Graphs, Geodesics, and Gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Chiara Caprini and Daniel G Figueroa. Cosmological backgrounds of gravitational waves. *Classical and Quantum Gravity*, 35(16):163001, 2018.

Stefan Chmiela, Huziel E Sauceda, Klaus-Robert Müller, and Alexandre Tkatchenko. Towards exact molecular dynamics simulations with machine-learned force fields. *Nature Communications*, 9(1): 1–10, 2018.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *ICML*, pp. 2990–2999, 2016.

Taco S Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical CNNs. In *ICLR*, 2018.

Alex Davies, Petar Veličković, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev, Richard Tanburn, Peter Battaglia, Charles Blundell, András Juhász, et al. Advancing mathematics by guiding human intuition with AI. *Nature*, 600(7887):70–74, 2021.

David L Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3): 613–627, 1995.

Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. SE (3)-Transformers: 3D roto-translation equivariant attention networks. *NeurIPS*, 33, 2020.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2014.

Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*, 2017.

Risi Kondor, Zhen Lin, and Shubhendu Trivedi. Clebsch–Gordan Nets: a fully Fourier space spherical convolutional neural network. In *NeurIPS*, volume 31, pp. 10117–10126, 2018.

Yi Liu, Limei Wang, Meng Liu, Xuan Zhang, Bora Oztekin, and Shuiwang Ji. Spherical message passing for 3D graph networks. *arXiv preprint arXiv:2102.05013*, 2021.

Oscar Méndez-Lucio, Mazen Ahmad, Ehecatl Antonio del Rio-Chanona, and Jörg Kurt Wegner. A geometric deep learning approach to predict binding conformations of bioactive molecules. *Nature Machine Intelligence*, 3(12):1033–1039, 2021.

Grégoire Montavon, Katja Hansen, Siamac Fazli, Matthias Rupp, Franziska Biegler, Andreas Ziehe, Alexandre Tkatchenko, Anatole Lilienfeld, and Klaus-Robert Müller. Learning invariant representations of molecules for atomization energy prediction. *NeurIPS*, 25:440–448, 2012.

F Narcowich, P Petrushev, and J Ward. Decomposition of Besov and Triebel–Lizorkin spaces on the sphere. *Journal of Functional Analysis*, 238(2):530–564, 2006a.

Francis J Narcowich, Pencho Petrushev, and Joseph D Ward. Localized tight frames on spheres. *SIAM Journal on Mathematical Analysis*, 38(2):574–594, 2006b.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. In *CVPR*, pp. 652–660, 2017.

Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O Anatole Von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical Review Letters*, 108(5):058301, 2012.

Kristof T Schütt, Pieter-Jan Kindermans, Huziel E Sauceda, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *NeurIPS*, 2017.

Raphael JL Townshend, Stephan Eismann, Andrew M Watkins, Ramya Rangan, Maria Karelina, Rhiju Das, and Ron O Dror. Geometric deep learning of RNA structure. *Science*, 373(6558): 1047–1051, 2021.

Yu Guang Wang, Quoc T Le Gia, Ian H Sloan, and Robert S Womersley. Fully discrete needlet approximation on the sphere. *Applied and Computational Harmonic Analysis*, 43(2):292–316, 2017.

Xuebin Zheng, Bingxin Zhou, Junbin Gao, Yu Guang Wang, Pietro Lio, Ming Li, and Guido Montúfar. How framelets enhance graph neural networks. In *ICML*, 2021.

A NEEDLETS

A.1 GENERALIZED FOURIER TRANSFORM

Let $\mathcal{M}$ be the manifold $\mathbb{S}^2$ or $\mathrm{SO}(3)$. The basis functions are spherical harmonics $\{Y_m^\ell(R) : \ell = 0, 1, \ldots, m = -\ell, \ldots, \ell\}$ and Wigner D-functions $\{D_{mn}^\ell(R) : m, n = -\ell, \ldots, \ell, \ell = 0, 1, \ldots\}$ for $\mathbb{S}^2$ and $\mathrm{SO}(3)$ respectively. Denote the basis functions as u_ℓ. We can write the generalized Fourier transform of a function $f : \mathcal{M} \to \mathbb{R}$ with quadrature rule sampling at scale j as

$$\begin{aligned}\widehat{f}_\ell = \langle f, u_\ell \rangle &= \int_{\mathcal{M}} f(\boldsymbol{x})\overline{u_\ell(\boldsymbol{x})}\mathrm{d}\boldsymbol{x} \\ &= \sum_{k=0}^{N_j} f(\boldsymbol{x}_{j,k})\sqrt{\omega_{j,k}}\overline{u_\ell(\boldsymbol{x}_{j,k})}\end{aligned}$$

The inverse Fourier transforms on $\mathbb{S}^2$ and $\mathrm{SO}(3)$ are as follows.

$$f(R) = [\mathcal{F}^{-1}\hat{f}](R) = \sum_{\ell=0}^{\infty}(2\ell+1)\sum_{m=-\ell}^{\ell}\hat{f}_m^\ell Y_m^\ell(R)$$

$$f(R) = [\mathcal{F}^{-1}\hat{f}](R) = \sum_{\ell=0}^{\infty}(2\ell+1)\sum_{m=-\ell}^{\ell}\sum_{n=-\ell}^{\ell}\hat{f}_{mn}^\ell D_{mn}^\ell(R).$$

Let (α, β) with $\alpha \in [0, 2\pi]$ and $\beta \in [0, \pi]$ be the spherical polar coordinates for the point $x \in \mathbb{S}^2$. The spherical harmonics can be explicitly written as

$$Y_{\ell,m}(\alpha, \beta) := \sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}}P_\ell^{(m)}(\cos\beta)e^{im\alpha}$$

where $P_\ell^{(m)}(t)$ is the associated Legendre polynomial of degree ℓ and order m, $m = -\ell, \ldots, \ell, \ell = 0, 1, \ldots$. We use the ZYZ Euler parameterization for $\mathrm{SO}(3)$. An element $R \in \mathrm{SO}(3)$ can be parameterized by $R(\alpha, \beta, \gamma)$ with $\alpha \in [0, 2\pi]$, $\beta \in [0, \pi]$ and $\gamma \in [0, 2\pi]$. Then, there exists a general relationship between Wigner D-functions and spherical harmonics:

$$D_{ms}^\ell(\alpha, \beta, -\gamma) = (-1)^s\sqrt{\frac{4\pi}{2\ell+1}}_s Y_\ell^m(\beta, \alpha)e^{is\gamma}.$$

A.2 NEEDLETS ON $\mathbb{S}^2$ AND $\mathrm{SO}(3)$

Tight framelets on manifold $\mathcal{M}$ is defined by a filter bank (a set of complex-valued filters) $\eta := \left\{a; b^1, \ldots, b^r\right\} \subset l_1(\mathbb{Z}) := \left\{h = \{h_k\}_{k\in\mathbb{Z}} \subset \mathbb{C} : \sum_{k\in\mathbb{Z}} |h_k| < \infty\right\}$ and a set of associated scaling functions, $\Psi = \{\alpha; \beta^1, \cdots, \beta^r\} \subset L_1(\mathbb{R})$, which is a set of complex-valued functions on the real axis satisfying the following equations, for $r = 1, \ldots, n, \xi \in \mathbb{R}$,

$$\widehat{\alpha}(2\xi) = \widehat{a}(\xi)\widehat{\alpha}(\xi), \quad \widehat{\beta^n}(2\xi) = \widehat{b^n}(\xi)\widehat{\alpha}(\xi).$$

Here $a(\cdot)$ is called the *low-pass* filter and $b^{(r)}(\cdot)$ are *high-pass* filters. Let $\{(\boldsymbol{u}_\ell, \lambda_\ell)\}_\ell$ be the eigenvalue and eighenvector pairs for the Laplace-Beltrami operator on $L_2(\mathcal{M})$. The framelets at scale level $j = 1, \cdots, J$ for manifold $\mathcal{M}$ are generated with the above scaling functions and orthonormal eigen-pairs by

$$\boldsymbol{\varphi}_{j,\boldsymbol{y}}(\boldsymbol{x}) = \sum_{\ell=1}^{\Lambda_j} \widehat{\alpha}\left(\frac{\lambda_\ell}{2^j}\right) \overline{\boldsymbol{u}_\ell(\boldsymbol{y})}\boldsymbol{u}_\ell(\boldsymbol{x})$$

$$\boldsymbol{\psi}^n_{j,\boldsymbol{y}}(\boldsymbol{x}) = \sum_{\ell=1}^{\Lambda_j} \widehat{\beta^{(n)}}\left(\frac{\lambda_\ell}{2^j}\right) \overline{\boldsymbol{u}_\ell(\boldsymbol{y})}\boldsymbol{u}_\ell(\boldsymbol{x}).$$

We call $\boldsymbol{\varphi}_{j,\boldsymbol{y}}(\boldsymbol{x})$ and $\{\boldsymbol{\psi}^n_{j,\boldsymbol{y}}(\boldsymbol{x})\}_{n=1}^r$ *low-pass* and *high -pass framelets* at scale j at point $\boldsymbol{y} \in \mathcal{M}$ respectively. The Λ_j is the bandwidth of scale level j and $n = 1, \cdots, r$.

Needlets are a type of framelets on the sphere ($\mathbb{S}^{\mathrm{d}}$) associated with a quadrature rule and a specific filter bank. This type of framelets can also be generalized to rotation group $\mathrm{SO}(3)$ with the same filter bank. For simplicity, we consider the filter bank $\boldsymbol{\eta} = \{a; b^1, b^2\}$ with two high-pass filters. We define the filter bank $\boldsymbol{\eta} = \{a; b^1, b^2\}$ by their Fourier series as follows.

$$\begin{aligned}
\widehat{a}(\xi) &:= \begin{cases} 1, & |\xi| < \frac{1}{8}, \\ \cos\left(\frac{\pi}{2}\nu(8|\xi|-1)\right), & \frac{1}{8} \leqslant |\xi| \leqslant \frac{1}{4}, \\ 0, & \frac{1}{4} < |\xi| \leqslant \frac{1}{2}, \end{cases} \\
\widehat{b^1}(\xi) &:= \begin{cases} 0, & |\xi| < \frac{1}{8}, \\ \sin\left(\frac{\pi}{2}\nu(8|\xi|-1)\right), & \frac{1}{8} \leqslant |\xi| \leqslant \frac{1}{4}, \\ \cos\left(\frac{\pi}{2}\nu(4|\xi|-1)\right), & \frac{1}{4} < |\xi| \leqslant \frac{1}{2}, \end{cases} \\
\widehat{b^2}(\xi) &:= \begin{cases} 0, & |\xi| < \frac{1}{4} \\ \sin\left(\frac{\pi}{2}\nu(4|\xi|-1)\right), & \frac{1}{4} \leqslant |\xi| \leqslant \frac{1}{2}, \end{cases}
\end{aligned} \tag{4}$$

where

$$\nu(t) := \chi_3(t)^2 = t^4\left(35 - 84t + 70t^2 - 20t^3\right), \quad t \in \mathbb{R}.$$

It can be verified that

$$|\widehat{a}(\xi)|^2 + \left|\widehat{b^1}(\xi)\right|^2 + \left|\widehat{b^2}(\xi)\right|^2 = 1 \quad \forall \xi \in [0, 1/2].$$

The associated needlet generators $\Psi = \{\alpha; \beta^1, \beta^2\}$ are explicitly given by

$$\begin{aligned}
\widehat{\alpha}(\xi) &= \begin{cases} 1, & |\xi| < \frac{1}{4}, \\ \cos\left(\frac{\pi}{2}\nu(4|\xi|-1)\right), & \frac{1}{4} \leqslant |\xi| \leqslant \frac{1}{2}, \\ 0, & \text{else}, \end{cases} \\
\widehat{\beta^1}(\xi) &= \begin{cases} \sin\left(\frac{\pi}{2}\nu(4|\xi|-1)\right), & \frac{1}{4} \leqslant |\xi| < \frac{1}{2}, \\ \cos^2\left(\frac{\pi}{2}\nu(2|\xi|-1)\right), & \frac{1}{2} \leqslant |\xi| \leqslant 1, \\ 0, & \text{else}, \end{cases} \\
\widehat{\beta^2}(\xi) &= \begin{cases} 0, & |\xi| < \frac{1}{2} \\ \frac{1}{2}\sin\left(\pi\nu(2|\xi|-2)\right), & \frac{1}{2} \leqslant |\xi| \leqslant 1 \\ 0, & \text{else}. \end{cases}
\end{aligned} \tag{5}$$

The framelet coefficients $v_{j,k}$ represent low-pass coefficients, and $w^n_{j,k}$ represent high-pass coefficients. They are defined as $\langle \boldsymbol{\varphi}_{j,k}, f\rangle$ and $\langle \boldsymbol{\psi}^n_{j,k}, f\rangle$ respectively:

$$v_{j,k} = \sum_{\ell=0}^{\Lambda_j} \widehat{f}_\ell \overline{\widehat{\alpha}\left(\frac{\lambda_\ell}{2^j}\right)} \sqrt{\omega_{j,k}} u_\ell\left(\boldsymbol{x}_{j,k}\right), \quad w^n_{j-1,k} = \sum_{\ell=0}^{\Lambda_j} \widehat{f}_\ell \overline{\widehat{\beta^n}\left(\frac{\lambda_\ell}{2^{j-1}}\right)} \sqrt{\omega_{j,k}} u_\ell\left(\boldsymbol{x}_{j,k}\right). \tag{6}$$

The coefficients of $v_{j,k}$ and $w^n_{j,k}$ in the Fourier space are then given by

$$\widehat{v}_{j,\ell} = \widehat{f}_\ell \overline{\widehat{\alpha}\left(\frac{\lambda_\ell}{2^j}\right)}, \quad \widehat{w}^n_{j-1,\ell} = \widehat{f}_\ell \overline{\widehat{\beta^n}\left(\frac{\lambda_\ell}{2^{j-1}}\right)} \tag{7}$$

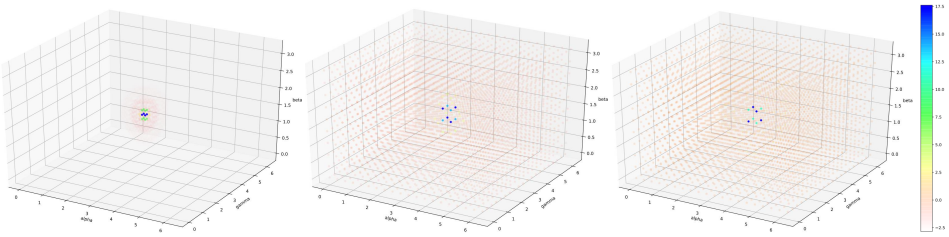

Figure 2: Needlets on $\mathrm{SO}(3)$ centered at $y_0 = (\alpha, \beta, \gamma) = (\pi, \pi/2, \pi)$ at scale $j = 6$. Left: $\boldsymbol{\varphi}_{6,y_0}$; Middle: $\boldsymbol{\psi}^1_{6,y_0}$; Right: $\boldsymbol{\psi}^2_{6,y_0}$

In the implementation of our model, we simply use the Fourier coefficients of $\{v_{j,k}\}$ and $\{w_{j,k}\}$ which can be computed fast and conveniently to represent the framelet coefficients. The tightness of framelets on $\mathcal{M}$ guarantees that the framelet representation of the signal is unique, and the signal can be reconstructed by the coefficients with zero numerical error, no matter in the spatial or frequency domain.

Needlets are a type of framelets on the sphere $\mathbb{S}^2$ associated with a quadrature rule and a specific filter bank. We generalize this type of framelets to rotation group $\mathrm{SO}(3)$ with the same filter bank. In this work, we use the needlet system of two levels ($j = J_0, J_1$) and 2 high-pass filters $\{b^1, b^2\}$ (i.e., $r = 2$). We use the ZYZ Euler parameterization for SO(3) and any element $R \in \mathrm{SO}(3)$ can be written as $R = R(\alpha, \beta, \sigma)$, where $\alpha \in [0, 2\pi]$, $\beta \in [0, \pi]$ and $\sigma \in [0, 2\pi]$. With the needlet filter bank, we can obtain needlets on SO(3) group with the filtered expansion of the Fourier basis (Wigner D-functions). In Figure 2, we show the needlets on SO(3) group at scale $j = 6$ at $y = y(\alpha, \beta, \gamma) \in \mathrm{SO}(3)$, where $\alpha = \pi, \beta = \pi/2, \gamma = \pi$, where the left, and the middle and right illustrate the low-pass needlet and two high-pass needlets respectively. We can observe that the low-pass needlet and high-pass needlets extract information at different scales. They provide the approximate and detailed representation for an $\mathrm{SO}(3)$ signal, as the counterparts of spherical needlets.

A.3 Needlet Decompostion and Reconstruction

Algorithm 1 Decomposition of Multi-Level Needlet Transform

Input: v_J – a (Λ_J, N_J)-sequence
Output: $(\{w^n_{J-1}, w^n_{J-2}, \ldots, w^n_{J_0}\}_{n=1}^r, v_{J_0})$
$v_J \to \hat{v}_J$
for $j \leftarrow J$ **to** $J_0 + 1$ **do**
  $\widehat{v}_{j-1} \longleftarrow \widehat{v}_{j,\cdot}\overline{\widehat{a}}\left(2^{-j}\lambda_\cdot\right)$
  for $n \leftarrow 1$ **to** r **do**
    $\widehat{w}^n_{j-1} \longleftarrow \widehat{v}_{j,\cdot}\overline{\widehat{b^n}}\left(2^{-j}\lambda_\cdot\right)$
    $w^n_{j-1} \leftarrow \hat{w}^n_{j-1}$
$v_{J_0} \leftarrow \hat{v}_{J_0}$

Algorithm 2 Reconstruction of Multi-Level Needlet Transform

Input: $(\{w^n_{J-1}, w^n_{J-2}, \ldots, w^n_{J_0}\}_{n=1}^r, v_{J_0})$
Output: v_J – a (Λ_J, N_J)-sequence
$\hat{v}_{J_0} \leftarrow v_{J_0}$
for $j \leftarrow J_0 + 1$ **to** J **do**
  for $n \leftarrow 1$ **to** r **do**
    $\widehat{w}^n_{j-1} \longleftarrow w^n_{j-1}$
  $\widehat{\mathrm{v}}_j \longleftarrow (\widehat{\mathrm{v}}_{j-1,\cdot})\,\widehat{a}\left(2^{-j}\lambda_\cdot\right) + \sum_{n=1}^r \widehat{\mathrm{w}}^n_{j,\cdot}\widehat{b^n}\left(2^{-j}\lambda_\cdot\right)$
$v_J \leftarrow \hat{v}_J$

As Eq. (6), we have $v_{j,k}$ and $w^n_{j-1,k}$ as (Λ_j, N_j) sequences. We have the following *decomposition relations*:

$$\begin{aligned} v_{j-1,k} &= \sum_{\ell=0}^{\Lambda_{j-1}} \widehat{f}_\ell \overline{\widehat{\alpha}\left(\frac{\lambda_\ell}{2^{j-1}}\right)} \sqrt{\omega_{j-1,k}} u_\ell\left(\boldsymbol{x}_{j-1,k}\right) \\ &= \sum_{\ell=0}^{\Lambda_{j-1}} \widehat{f}_\ell \overline{\widehat{\alpha}\left(\frac{\lambda_\ell}{2^{j}}\right)\widehat{a}\left(\frac{\lambda_\ell}{2^{j}}\right)} \sqrt{\omega_{j-1,k}} u_\ell\left(\boldsymbol{x}_{j-1,k}\right) \\ &= \sum_{\ell=0}^{\Lambda_{j}} \widehat{v_{j,\ell}} \overline{\widehat{\alpha}\left(\frac{\lambda_\ell}{2^{j}}\right)} \sqrt{\omega_{j-1,k}} u_\ell\left(\boldsymbol{x}_{j-1,k}\right) \\ &= \left[\left(v_j *_j a^\star\right) \downarrow_j\right](k), \end{aligned}$$

where $\downarrow$ denotes the down-sampling operator. Similarly, for $k = 0, \ldots, N_{j-1}$ and $n = 1, \ldots, r$,

$$\begin{aligned} w^n_{j-1,k} &= \sum_{\ell=0}^{\Lambda_j} \widehat{f}_\ell \overline{\widehat{\beta^n}\left(\frac{\lambda_\ell}{2^{j-1}}\right)} \sqrt{\omega_{j-1,k}} u_\ell\left(\boldsymbol{x}_{j-1,k}\right) \\ &= \left(v_j *_j \left(b^n\right)^*\right)_k. \end{aligned}$$

Therefore, we have the following identity for reconstruction.

$$\begin{aligned} \widetilde{v} &:= \left(v_{j-1} \uparrow_j\right) *_j a + \sum_{n=1}^{r} w^n_{j-1} *_j b^n \\ &= \left(\left(\left(v_j *_j a^\star\right) \downarrow_j\right) \uparrow_j\right) *_j a + \sum_{n=1}^{r} \left(v_j *_j \left(b^n\right)^\star\right) *_j b^n \end{aligned}$$

The multi-level decomposition and reconstruction algorithms are shown as Algorithms 1 and 2. As plus the Fourier transforms in the algorithms can be implemented by FFT, the fast multi-level needlet transform on $\mathbb{S}^2$ with N the size of the input data has the computational complexity $\mathcal{O}(N\sqrt{\log N})$. With needlets fiters given, we can pre-compute the needlet coefficients and store data as signals in the frequency domain. We can further make decomposition into a finer scale and reconstruction to obtain lower-level approximation information, depending on the specific application scenario.

B Rotation Equivariance Proof

B.1 Error Bound of Rotation Equivariance

In order to reduce the numerical error by the repeated forward and backward FFTs, and also to decrease the model complexity, we apply non-linear shrinkage function on the high-pass coefficients with an controllable parameter σ, which is an analogue to the noise level of the denoising model. Since the low-pass coefficients provide approximate information of the input signal, our model has approximate rotation-equivariance. According to the needlets theory, the rotation-equivariance error in Eq. (3) due to the use of shrinkage on the high-passes has the convergence order $2^{-(J_0+1)s}$.

Proof. Define

$$f_J = f^{(\mathrm{L})}_{J_0} + f^{(\mathrm{H}),J}_{J_0} = f^{(\mathrm{L})}_{J_0} + \sum_{j=J_0}^{J} \langle f, \psi_j \rangle \psi_j$$

as the spherical needlet approximation. By Wang et al. (2017, Theorem 3.12) , for $f \in \mathbb{W}^s_p(\mathbb{S}^2)$ with $s > 0$ and $J \geqslant 0$, we have $\|f - f_J\| \leqslant C_1 2^{-Js}$ and $\left\|f - f^{(\mathrm{L})}_{J_0}\right\| \leqslant C_2 2^{-J_0 s}$, C_1 and C_2 are constants that depends only on d, p, s, h, and fiter smoothness κ. Therefore,

$$\begin{aligned} \left\|f^{(\mathrm{H}),J}_{J_0}\right\|^2 &= \sum_{\ell \leqslant 2^J} \left\|\widehat{f^{(\mathrm{H})}_\ell}\right\|^2 = \left\|f_J - f^{(\mathrm{L})}_{J_0}\right\|^2 \\ &\leqslant \|f - f_J\|^2 + \left\|f - f^{(\mathrm{L})}_{J_0}\right\|^2 \\ &\leqslant C_1 2^{-Js} + C_2 2^{-J_0 s} \leqslant C 2^{-J_0 s}, \end{aligned} \tag{8}$$

where C is a sufficiently large constant depending on d, p, s, h, κ, C_1 and C_2. The Eq. (8) holds for all J. Then, Eq. (2.1) satisfies the following upper bound.

$$\begin{aligned}
\text{Error} &= \sum_{\ell=0}^{B} \left\| \text{Shr}(\widehat{L_R f \star \phi})_\ell^{(\text{H})} - D^\ell(R)\text{Shr}(\widehat{f \star \phi}_\ell^{(\text{H})}) \right\|^2 \\
&= \sum_{\ell=0}^{B} \left\| \text{Shr}(D^\ell(R)\widehat{f}_\ell^{(\text{H})}\widehat{\phi}_\ell) - D^\ell(R)\text{Shr}(\widehat{f}_\ell^{(\text{H})}\widehat{\phi}_\ell) \right\|^2 \\
&\leqslant \sum_{\ell=0}^{B} \left[\left\| \text{Shr}(D^\ell(R)\widehat{f}_\ell^{(\text{H})}\widehat{\phi}_\ell) \right\|^2 + \left\| D^\ell(R)\text{Shr}(\widehat{f}_\ell^{(\text{H})}\widehat{\phi}_\ell) \right\|^2 \right] \\
&\leqslant \left\| D^\ell(R)\widehat{f}_\ell^{(\text{H})}\widehat{\phi}_\ell \right\|^2 + \left\| D^\ell(R) \right\|^2 \left\| \widehat{f}_\ell^{(\text{H})}\widehat{\phi}_\ell \right\|^2 \\
&\leqslant 2\sum_{\ell=0}^{B} \left\| \widehat{f}_\ell^{(\text{H})}\widehat{\phi}_\ell \right\|^2
\end{aligned}$$

If the scale of the low-pass is J_0, we would have

$$\text{Error} \leqslant 2 \sum_{\ell \geqslant 2^{J_0+1}} \left\| \widehat{f}_\ell^{(\text{H})}\widehat{\phi}_\ell \right\|^2 \leqslant C_\phi \sum_{\ell \geqslant 2^{J_0+1}} \left\| \widehat{f}_\ell^{(\text{H})f} \right\|^2 .$$

Here C_ϕ is a constant depending on the fiter ϕ. By Parseval's identity and Eq. (8),

$$\text{Error} \leqslant C_\phi \sum_{\ell \geqslant 2^{J_0+1}} \left\| f_{J_0+1}^{(\text{H})} \right\|^2 \leqslant \tilde{C}_\phi 2^{-(J_0+1)s},$$

thus completes the proof.

□

B.2 Rotation Equivariance of Spectral Pooling

Denote $[D^0(R), \cdots, D^{\ell-1}(R), D^\ell(R)]$ as $D(R)$, then

$$\begin{aligned}
\text{P}(\widehat{L_R f}) &= \text{P}([D^0(R)\hat{f}_0, \cdots, D^\ell(R)\hat{f}_\ell]) \\
&= [D^0(R)\hat{f}_0, \cdots, D^{\ell/2}(R)\hat{f}_{\ell/2}] \\
&= \text{P}(\hat{f}) \odot [D^0(R), \cdots, D^{\ell/2}(R)] \\
&= \text{P}(\hat{f}) \odot \text{P}(D(R))
\end{aligned}$$

where $\odot$ denotes element-wise multiplication, and $\text{P}(\cdot)$ denotes spectral pooling operator. Thus, the spectral pooling operator is equivariant, due to the R-related operator $\text{P}(D(R))$.

C Details of Experiments

C.1 Spherical Signal Construction of Molecular Datasets

The **QM7** dataset contains at total 7165 molecules and each molecule contains at most $N = 23$ atoms of $T = 5$ types (H, C, N, O, S), which is to regress over the atomic energy of molecules given the corresponding position p_i and charge z_i of each atom i. Similar to Spherical CNN Cohen et al. (2018), we generalize the *Coulomb matrix* ($C \in \mathbb{R}^{N \times N}$) proposed by Rupp et al. (2012) and obtain 23 spherical signals for every molecule. We define a sphere S_i centered at p_i for each atom i, and the potential function $U_z(x) = \sum_{j \neq i, z_j = z} \frac{z_i \cdot z}{\|x - p_i\|}$, where z is the charge of the atom, and x is taken from $\mathbb{S}^2$, thus we produce N spherical signals of T channels for every molecule. We use Gauss-Legendre rule to discretize the continuous functions on the sphere with $L = 20$, creating a sparse $N \times T \times (2L + 1) \times (L + 1)$ tensor for each molecule.

To test model's expressive ability for molecular dynamics simulations, we apply our model to the **MD17** dataset, which is to predict the energies and forces at the atomic level for several organic molecules with up to 21 atoms and four chemical elements, using the molecular dynamics trajectories. To incorporate 3D information, we create N spherical signals that are centered at the positions of each atom for every sample, where N is the number of the atoms in the molecule. Considering the atom i, we define a corresponding spherical signal $U_i(x)$, where x is taken from the sphere by Gauss-Legendre sampling method. The MD17 dataset provides the absolute Cartesian coordinates of atoms, thus we can compute the relative position of each atom relative to x. The spherical signal U_i is defined by $U_i(x) = \sum_{j=1}^{N} \mathcal{N}(d_j, \theta_j, \varphi_j)$, where $(d_j, \theta_j, \varphi_j)$ is the relative position of atom j relative to x, where d_j, θ_j, φ_j denote the radial distance, polar angle, and the azimuthal angle respectively, as shown in Figure 3. Different from the QM7 dataset, the MD17 contains no *Coulomb matrix*. To extract effective features with the relative positions, $\mathcal{N}$ can be taken as a neural network function. We choose the approach of neural networks in our implementation, since it provides an adaptive learning of features. We fine-tune each individual hyperparameter for every type of molecules on validation sets with 1000 samplings for both training and validation. The trained models are then applied to the test sets.

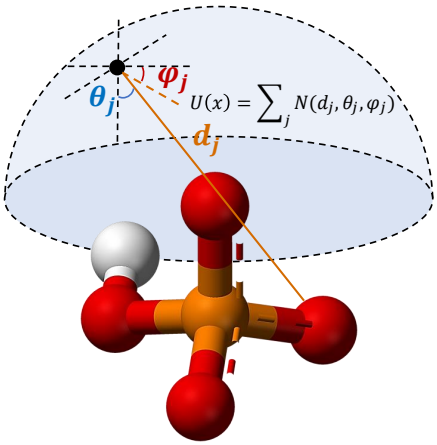

Figure 3: An illustration for computing the spherical signal of a molecule. We aggregate the information of each atom in the molecule with relative distance, polar angle and azimuthal angle.

C.2 Experimental Setting

Our models are trained on 24G NVIDIA GeForce RTX 3090 Ti GPUs. The hyperparemeters are obtained by grid search. Adam (Kingma & Ba, 2014) is used as our optimizer.

For Local MNIST Classification task, we compare our model with Spherical CNNs. These two models have a similar network architecture. The Spherical CNNs use the same architecture in Cohen et al. (2018): $\mathbb{S}^2$conv-ReLU-$\mathrm{SO}(3)$conv-ReLU-FC-softmax, bandwidth L = 30, 10, 6 and k = 20, 40, 10 channels. When it comes to our model, we just replace $\mathbb{S}^2$conv and $\mathrm{SO}(3)$conv with $\mathbb{S}^2$-needlet convolution and $\mathrm{SO}(3)$-needlet convolution, with the same bandwidth L = 30, 10, 6 and k = 20, 40, 10 channels. The implementation uses the batch size of 64 and the learning rate 1e-3.

For QM7 task, we choose the batch size of 32 and learning rate 5e-4. For MD17 tasks, the batch size is 32 and the learning rate is 2e-4. The bandwidth L is from 20, 20, 10, 10, 5, to 5 in the final bock and the feature dimension is from 5, 5, 8, 16, 32, to 64. The hyperparameter σ is taken as 0.001 for shrinkage. We run 10 epochs for QM7 task and 1000 epochs for MD17 task.

For Parameter Estimate for Gravitational Wave task, we transform the original sample rules from HealPix to the Gauss-Legendre tensor product rule with bandwidth $L = 128$ by finding the nearest HEALPix point given the Gauss-Legendre coordinates. The training dataset has 9000 spheres (images), while the test dataset has 1000 spheres. We use 5 blocks containing needlet convolution ($\mathbb{S}^2$ in the first blocks and $\mathrm{SO}(3)$ in the rest), shrinkage activation function and batch normalization layers. The bandwidth is from 128, 64, 32, 32,16 to 8 in the final block and the feature is 16, 32, 64, 128, 256 respectively in each block, with batch size 16 and learning rate 5e-5.

C.3 Ablation Study

Equivariance Error It can be proven that our $\mathbb{S}^2$-needlet convolution, $\mathrm{SO}(3)$-needlet convolution without shrinkage and spectral pooling are equivariant to $\mathrm{SO}(3)$ transforms for the continuous case. In the implementation, we exploit the polynomial-exact quadrature rule to sample the sphere with small numerical integration error. Table 4 shows the rotation-equivariance error of the modules in our framework. Experimental results verify that the errors approximate the machine error of floating

Table 4: Equivariance Error Results. *Single* denotes Single-precision floating-point format. *Double* denotes Double-precision floating-point format. The values are calculated from the average of ten trails. ReLU function in SO(3)+ReLU is applied in the spatial domain, thus involving an FFT and inverse FFT.

Operator	Error (*Single*)	Error (*Double*)
$\mathbb{S}^2$-Conv	2e-7	7e-16
SO(3)-Conv	1e-7	8e-16
SO(3)+ReLU	1e-7	8e-16
SO(3)+Shrinkage	2e-4	5e-7
Pooling	0	0

points, except $\mathrm{SO}(3)$-needlet convolution with the shrinkage mechanism which is governed by approximate equivariance error. We observe that the equivariance errors introduced by the shrinkage mechanism with a small value of σ (e.g., $\sigma = 0.001$) are 2e-4 and 5e-7 with *Single* and *Double* floating-point format respectively, which are tolerable in the neural network learning.

Sensitivity Analysis As σ is a hyperparameter in the shrinkage activation function, it is critical to know how does this value affect our model's equivariance property. We take different values of σ ranging from 1e-7 to 1 to study how the equivariant error changes and how much this signal compresses. We use an $\mathrm{SO}(3)$ signal with the bandwidth $L = 128$ and send it to an $\mathrm{SO}(3)$-needlet convolutional layer with $J = 7$. As shown in Figure 4, when σ is greater than 0.1, the equivariance error is around 0.1, which may affect the accuracy of our equivariant network model. When the σ is smaller than 1e-6, the equivariance error is approaching to single-precision machine error. For the compression rate, the shrinkage mechanism will cut off 20% signal when σ is about 0.1 and approaching to 0 when σ is less than 1e-6.

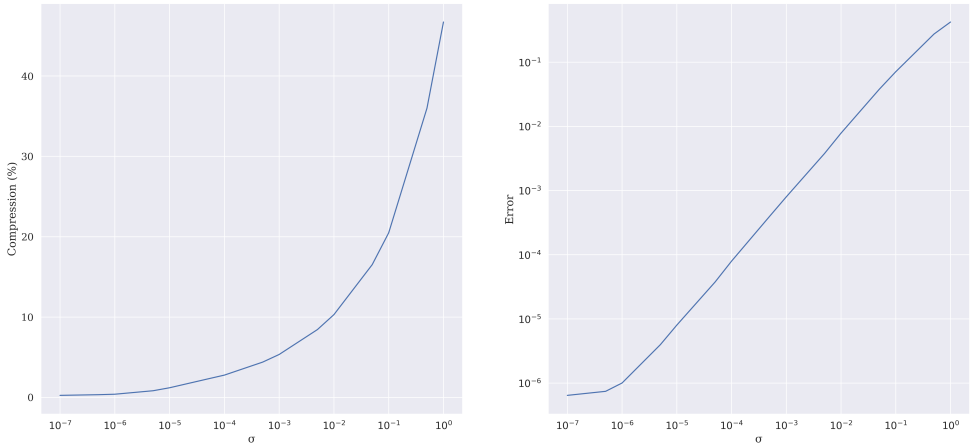

Figure 4: Sensitivity analysis for σ. The equivariance error is near mechanism error when σ is less than 1e-6. The shrinkage activation function will nearly compress 20% signal with $\sigma = 0.1$ and become identity function as σ is close to 0.