Towards Geometric Normalization Techniques in SE(3) Equivariant Graph Neural Networks for Physical Dynamics Simulations

Ziqiao Meng [1*], **Liang Zeng** [3], **Zixing Song** [1], **Tingyang Xu** [2], **Peilin Zhao** [2] and **Irwin King** [1]
[1]The Chinese University of Hong Kong
[2]Tencent AI Lab
[3]Tsinghua University
king@cse.cuhk.edu.hk, masonzhao@tencent.com,

Abstract

SE(3) equivariance is a fundamental property that is highly desirable to maintain in physical dynamics modeling. This property ensures neural outputs to remain robust when the inputs are translated or rotated. Recently, there have been several proposals for SE(3) equivariant graph neural networks (GNNs) that have shown promising results in simulating particle dynamics. However, existing works have neglected an important issue that current SE(3) equivariant GNNs cannot scale to large particle systems. Although some simple normalization techniques are already in use to stabilize the training dynamics of equivariant graph networks, they actually break the SE(3) equivariance of the architectures. In this work, we first show the numerical instability of training equivariant GNNs on large particle systems and then analyze some existing normalization strategies adopted in modern works. We propose a new normalization layer called GeoNorm, which can satisfy the SE(3) equivariance and simultaneously stabilize the training process. We conduct comprehensive experiments on N-body system simulation tasks with larger particle system sizes. The experimental results demonstrate that GeoNorm successfully preserves the SE(3) equivariance compared to baseline techniques and stabilizes the training dynamics of SE(3) equivariant GNNs on large systems.

1 Introduction

Particle dynamics simulations have recently been successfully formulated as geometric graph learning problems. The message passing mechanism of graph neural networks (GNNs) [Kipf and Welling, 2017; Hamilton *et al.*, 2017; Veličković *et al.*, 2018] can effectively capture the interactions between each pair of particles. However, applying GNNs to particle dynamics modeling poses a special challenge in preserving physical symmetry biases, particularly

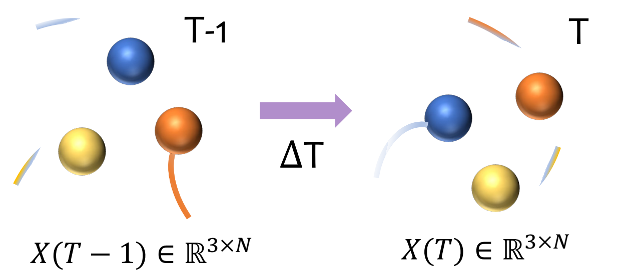

Figure 1: The illustration of particle dynamics simulations. At time step $T-1$, input particle locations $\mathbf{X}(T-1) \in \mathbb{R}^{3\times N}$, where each column $\mathbf{x}_i$ denotes the 3D coordinate of particle i. Then we predict next location $\mathbf{X}(T) \in \mathbb{R}^{3\times N}$ at time step T. Note that ΔT is a pre-defined delta time frame, which determines the time slice to discretize an inherent continuous trajectory.

the SE(3) transformation (3D rotation and translation) equivariance. Since it is not feasible to consider all possible 3D rotations and translations of input systems through data augmentations in continuous space, a promising approach is to directly incorporate symmetry inductive biases into the design of the GNN architecture. To achieve SE(3) equivariance in GNN learning mechanisms, several SE(3) equivariant GNNs [Satorras *et al.*, 2021b; Huang *et al.*, 2022; Du *et al.*, 2022; Thomas *et al.*, 2018; Fuchs *et al.*, 2020; Köhler *et al.*, 2020] have been proposed and have shown remarkable empirical performance. Among these models, EGNN [Satorras *et al.*, 2021b] is considered the most important and fundamental equivariant GNN model due to its simplicity and efficiency. However, the current EGNN is typically applied to small particle systems (3-5 particles), and it faces numerical instability issues when trained on larger systems (50-100 particles), making it impractical for large particle systems. Specifically, the position update step of EGNN shown in Eq. 2 can be highly unstable, and this instability becomes more severe with deeper models and larger systems. We illustrate this issue in Figure 2 through n-body system simulation experiments involving charged particles following simple physics rules. To address this issue, some recent works have employed simple normalization techniques. One popular strategy is scaling geometric vectors by their norm to make them unit-length vectors. However, this approach vio-

*This work is done when Ziqiao Meng worked as an intern in Tencent AI Lab.

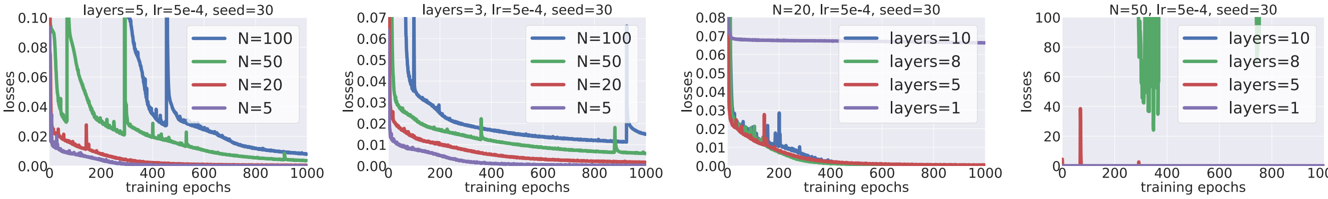

Figure 2: The visualization of the numerical instability issue of EGNN. The upper panel shows the training curve of EGNN on n-body system with $N = 20, 50$. When the layer depth of EGNN increases from 1 to 10, the numerical explosion occurs (training dynamics becomes extremely unstable). When the $N = 50$, the training dynamics becomes unstable even when layer depth is only 5. Note that we do not draw the training curve of layer depth 10 when $N = 50$ since the loss value is too large to plot in the figure; The lower panel shows that the training dynamics becomes increasingly unstable when the size of system becomes increasingly large. Even when the layer depth is only 3, the numerical explosion will easily occur when the system size goes to $N = 100$.

lates SE(3) equivariance, although it can stabilize the training dynamics. In the following sections, we will systematically analyze the weaknesses of existing approaches. In general, designing novel SE(3) equivariant normalization layers that can be easily integrated into current equivariant GNN architectures is challenging yet necessary. This will pave the way for training SE(3) equivariant GNNs on large particle system dynamics.

In this study, we begin by examining various normalization strategies. We analyze their properties and shortcomings through theoretical analysis. Following this, we introduce a new normalization layer called GEONORM. This ensures stable training dynamics and upholds the SE(3) equivariance constraints with a theoretical guarantee. We perform extensive experiments on n-body system simulation tasks with large system sizes, specifically $N = 50, 100$. It is observed that GEONORM effectively stabilizes training, irrespective of different weight initialization and learning rate choices. We further evaluate the SE(3) equivariance preservation of GEONORM by applying random rotations and translations to test inputs. Empirical results show that GEONORM achieves the lowest rotational and translational errors (RTE) when compared to baseline methods. In conclusion, our contributions can be summarized as follows:

- We systematically analyze the equivariance properties of several simple and existing normalization strategies. To the best of our knowledge, this is the first attempt to investigate geometric normalization techniques with equivariance analysis;
- We propose a SE(3) equivariant normalization layer, called GEONORM, which can stabilize the training dynamics of EGNN on much larger systems;
- We conduct extensive experiments to verify that our proposed GEONORM method can effectively address the numerical explosion issue, while maintaining the SE(3) group equivariance of the original EGNN architecture.

2 Related Works

Normalization. Normalization layers are essential components of modern large models. In language and vision modeling, various classic normalization techniques have been introduced to improve the stability of large model training. These normalization layers aim to normalize the output of neurons by removing the mean and dividing by the standard deviation along specific dimensions, ensuring that the values fall within a reasonable range. For example, BatchNorm [Ioffe and Szegedy, 2015] standardizes the output across the entire batch of samples to address internal covariate shift. LayerNorm [Ba *et al.*, 2016], on the other hand, normalizes the output across the feature dimensions of each individual sample, resulting in improved training stability for transformer-based models. Other variations follow a similar normalization approach but with different dimensions [Ulyanov *et al.*, 2016; Wu and He, 2018; Brock *et al.*, 2021], or use simplified versions by scaling with different norms [Singh and Krishnan, 2020; Daneshmand *et al.*, 2020; Salimans and Kingma, 2016; Miyato *et al.*, 2018]. Recent research has explored incorporating symmetry priors into normalization layers. In the context of graph structures, normalization techniques have been developed to stabilize deep GNN training by leveraging graph-specific information [Cai *et al.*, 2021; Zhao and Akoglu, 2020; Yang *et al.*, 2020; Li *et al.*, 2022; Zhou *et al.*, 2020; Dwivedi *et al.*, 2023]. Similarly, there are SO(3) equivariant normalization layers designed for point cloud modeling [Deng *et al.*, 2021; Shen *et al.*, 2020]. Additionally, different normalization techniques have been proposed for chemical modeling [Meng *et al.*, 2023] and protein modeling, taking into account corresponding geometric considerations [Schütt *et al.*, 2017; Batatia *et al.*, 2022; Jing *et al.*, 2021].

Equivariant Graph Neural Networks. There are three main types of equivariant GNNs. Irreducible representation methods [Fuchs *et al.*, 2020; Anderson *et al.*, 2019; Thomas *et al.*, 2018; Batzner *et al.*, 2022; Zitnick *et al.*, 2022; Brandstetter *et al.*, 2022; Frank *et al.*, 2022] leverage specific equivariant basis functions to process relative position signals. However, these models suffer from excessive computational overhead and produce higher-order outputs. Regular representations [Finzi *et al.*, 2020; Hutchinson *et al.*, 2021] map each geometric vector to a group element by designing equivariant group convolutions. However, these models are usually SE(3)-invariant. Scalarization methods [Satorras *et al.*, 2021b; Schütt *et al.*, 2017; Gasteiger *et al.*, 2020; Liu *et al.*, 2022; Köhler *et al.*, 2020; Jing *et al.*, 2021; Huang *et al.*, 2022; Schütt *et al.*, 2021; Thölke and Fabritiis, 2022; Klicpera *et al.*, 2021] transform relative position information into invariant scalars and use

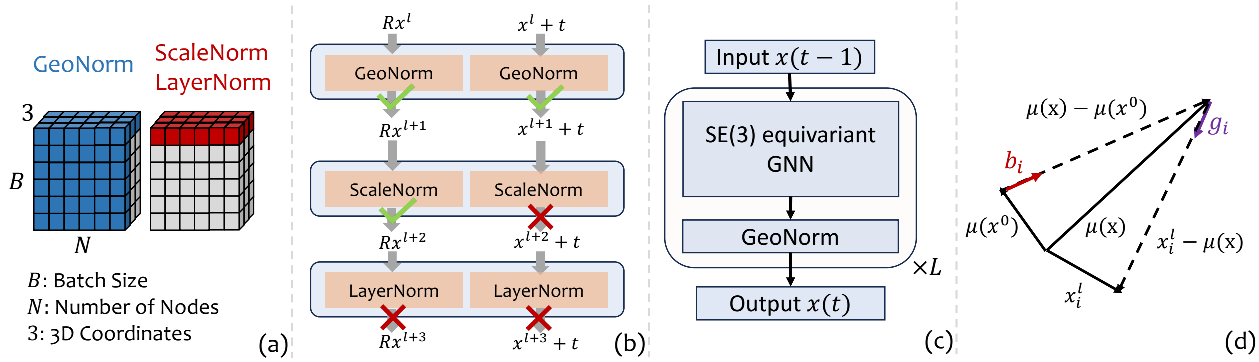

Figure 3: The proposed GeoNorm method. Figure (a) illustrates the comparisons over three different norms. GeoNorm summarizes statistics over dimension N and coordinate dimension 3 while the feature scaling norm and the vanilla LN computes summary statistics only over the coordinate dimension; Figure (b) illustrates the comparison over three different norms in terms of SE(3) equivariance constraints fulfillment; Figure (c) shows how to concatenate GeoNorm with SE(3) equivariant graph neural networks in the general architectures; Figure (d) illustrates the comparison between b_i and g_i learning mechanism.

this scalar information to update coordinates. This class of

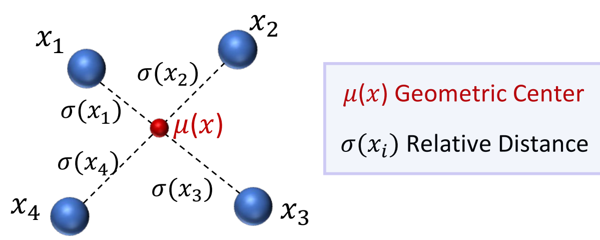

Figure 4: Two important geometric notions in a particle system.

models dominates current research and benefits many downstream applications like molecules [Hoogeboom *et al.*, 2022], proteins [Ganea *et al.*, 2022] and materials [Jiao *et al.*, 2023; Song *et al.*, 2024].

3 Preliminary

We first define the problem formulation of particle dynamics modeling. Given a set of N particles $P = \{p_i\}_{i=1}^N$ with an input 3D position vector matrix $\mathbf{X}(T) \in \mathbb{R}^{3\times N}$ at time T, each particle p_i is associated with an input geometric vector $\mathbf{x}_i(T) \in \mathbb{R}^3$ (i-th column of $\mathbf{X}(0)$) and an input scalar feature vector $\mathbf{h}_i(T) \in \mathbb{R}^H$ with hidden dimension H. A geometric graph $G = (P, E)$ can be constructed by building local edges between each pair of particles. Usually, each particle connects with its neighbors within a pre-defined cutoff distance threshold γ such that $E = \{e_{ij} | \|\mathbf{x}_i(T) - \mathbf{x}_j(T)\|_2 < \gamma\}$. In this work, we only consider fully connected geometric graphs. Each edge e_{ij} can be associated with an edge feature $\mathbf{e}_{ij}$, such as bond types in chemical molecules. The target of dynamics modeling is predicting the system state $\mathbf{X}(T+\Delta T)$ after ΔT steps given the initial state $\mathbf{X}(T)$ as input.

SE(3) Equivariance. Let $T_g : \mathcal{X} \to \mathcal{X}$ be a set of transformations on $\mathcal{X}$ for the abstract group $g \in G$. We say a function $f : \mathcal{X} \to \mathcal{Y}$ is equivariant to g if there exists an equivalent transformation on its output space $S_g : \mathcal{Y} \to \mathcal{Y}$ such that:

$$f(T_g(\mathbf{X})) = S_g(f(\mathbf{X})). \tag{1}$$

If S_g is an identity mapping function, then the function $f(\cdot)$ is invariant to group action g. Function $f(\cdot)$ is considered to be SE(3) equivariant if it is equivariant to both 3D rotations $\mathbf{R} \in \mathbb{R}^{3\times 3}$ (SO(3) group elements) and 3D translations $\mathbf{t} \in \mathbb{R}^3$. Specifically, function $f(\cdot)$ is SE(3)-equivariant if $f(\mathbf{RX} + \mathbf{t}) = \mathbf{R}f(\mathbf{X}) + \mathbf{t}$.

Equivariant Graph Neural Networks (EGNN). In this work, we mainly focus on the EGNN architecture since its output is not higher-order outputs like irreducible representation-based methods and it is the fundamental prototype of various scalarization-based methods. Currently, EGNN is still regarded as the most efficient equivariant GNN models. The following is the general message passing scheme proposed by EGNN [Satorras *et al.*, 2021b]:

$$\begin{aligned}
\mathbf{m}_{ij}^l &= \phi_{\mathbf{m}}(\mathbf{h}_i^l, \mathbf{h}_j^l, \|\mathbf{x}_i^l - \mathbf{x}_j^l\|_2^2, \mathbf{e}_{ij}), \\
\mathbf{x}_i^{l+1} &= \mathbf{x}_i^l + \sum_{j\neq i} (\mathbf{x}_i^l - \mathbf{x}_j^l)\phi_{\mathbf{x}}(\mathbf{m}_{ij}^l), \\
\mathbf{h}_i^{l+1} &= \phi_{\mathbf{h}}(\mathbf{h}_i^l, \sum_{j\in N(i)} \mathbf{m}_{ij}^l),
\end{aligned} \tag{2}$$

where $\mathbf{x}_i^l$ denotes the coordinate of particle i at layer l, $\mathbf{h}_i^l$ denotes the invariant embedding of particle i at layer l, $\phi_{\mathbf{m}}$ and $\phi_{\mathbf{h}}$ denote two separate multi-layer perceptrons (MLPs). For dynamics modeling, the EGNN is trained in an Markov way with input $\mathbf{x}_i(T)$ and output $\mathbf{x}_i(T + \Delta T)$. The layer input $\mathbf{x}_i^0$ is the input system state $\mathbf{x}_i(T)$ at time T and the final layer output $\mathbf{x}_i^L$ of the above learning mechanism (stacking L layers) is the output predicted $\mathbf{x}_i(T + \Delta T)$.

4 Proposed Methods

In this section, we first demonstrate the issue of numerical instability in SE(3) equivariant GNNs and identify the specific challenges in designing geometric normalization. Next, we summarize the temporary normalization techniques used in previous works and analyze their limitations. To address this challenge, we introduce a new normalization layer called GeoNorm, which provides SE(3) equivariance with theoretical guarantee. Finally, we discuss some trade-off issues recently discovered in GeoNorm.

4.1 Numerical Instability & Particular Challenges

From Eq. 2, we can see that there are two separate updating steps which are invariant embedding updates and geometric vector updates. Indicated by equivariant normalizing flow [Satorras *et al.*, 2021a], the coordinate update step easily explodes when utilized in an ODE. Based on our experimental observations shown in Figure 2, the coordinate update step also explodes when training EGNN on large particle systems. In fact, both invariant embedding updates and geometric vector updates can lead to numerical explosions. The numerical instability of the invariant embedding update part can be alleviated through a vanilla layer normalization LayerNorm [Ba *et al.*, 2016]. If we apply LayerNorm on geometric vectors, the normalization layer would be as follows:

$$\mu(\mathbf{x}_i^l) = \frac{\sum_{d=1}^{3} \mathbf{x}_{id}^l}{3}, \sigma(\mathbf{x}_i^l) = \sqrt{\frac{\sum_{d=1}^{3}(\mathbf{x}_{id}^l - \mu(\mathbf{x}_i^l))^2}{3}},$$
$$\mathbf{x}_i^{l+1} = F(\mathbf{x}_i^l) = g_i \cdot (\frac{\mathbf{x}_i^l - \mu(\mathbf{x}_i^l)}{\sigma(\mathbf{x}_i^l)}) + \mathbf{b}_i, g_i \in \mathbb{R}, \mathbf{b}_i \in \mathbb{R}^3 \tag{3}$$

where $\mathbf{x}_{id}^l \in \mathbb{R}$ denotes the dth dimension of the position coordinate of particle i in layer l, $\mu(\mathbf{x}_i^l) \in \mathbb{R}$ and $\sigma(\mathbf{x}_i^l)$ denote the mean and standard deviation across coordinate dimensions respectively, $\mathbf{x}_i^{l+1}$ denotes the output position vector of particle i after the lth layer block, $g_i \in \mathbb{R}$ denotes a learnable re-scaling scalar parameter, $\mathbf{b}_i \in \mathbb{R}^3$ denotes a learnable bias parameter. Note that unlike the vanilla LayerNorm on the invariant embedding update, the standard deviation $\sigma(\mathbf{x}_i^l)$ in Eq. 3 is a scalar value since each dimension of the particle should be scaled by the same magnitude. However, LayerNorm has two important deflects. First, the $\mu(\mathbf{x}_i^l)$ and $\sigma(\mathbf{x}_i^l)$ computation are measuring the mean and variance across coordinate dimension values, which does not summarize useful statistics that can reflect geometric information. Second, the above transformation is not SE(3) equivariant and thereby we present the following lemma:

Lemma 1. *The vanilla* LayerNorm *on geometric vectors are neither rotation equivariant nor translation equivariant.*

The proof of lemma 1 is straightforward since $\mu(\mathbf{x}_i^l)$ is clearly not SE(3) equivariant and thus we show the detailed procedures in Appendix. To sum up, the geometric vector update has instability issue and it cannot be easily solved by simple modifications of the vanilla LayerNorm. Hence, designing SE(3) equivariant normalization layers for geometric vectors is a non-trivial challenge.

4.2 Existing Solutions

There are some straightforward solutions tackling the instability issue. One of the popular strategies adopted in many recent works [Liao and Smidt, 2023; Fuchs *et al.*, 2020; Deng *et al.*, 2021; Jing *et al.*, 2021; Batatia *et al.*, 2022] is scaling the geometric vector by its norm. Specifically, most of the variants can be concluded in the following form:

$$\mathbf{x}_i^{l+1} = F(\mathbf{x}_i^l) = g_i \cdot \frac{\mathbf{x}_i^l}{\mathrm{norm}(\mathbf{x}_i^l)}, \quad g_i \in \mathbb{R}, \tag{4}$$

where g_i is still a learnable re-scaling scalar parameter, $\mathrm{norm}(\cdot)$ denotes a vector norm computation on $\mathbf{x}_i$. Some adopt the L2-norm and some adopt the RMS-norm (Root-Mean-Square norm). To summarize, the temporary solution presented in Eq. 4 involves using various norms to scale geometric vectors into unit-length vectors, depending on the selected norm. Experimental evidence from related works suggests that this method of feature scaling can significantly facilitate the training stability of EGNN. However, it also has issues with SE(3) equivariance constraints and we present the following lemma:

Lemma 2. *The feature scaling norm* ScaleNorm *is rotation equivariant but not translation equivariant.*

We use L2-norm as an example to prove the Lemma 2:

Proof. $F(\mathbf{R}\mathbf{x}_i) = g_i \cdot \frac{\mathbf{R}\mathbf{x}_i}{\|\mathbf{R}\mathbf{x}\|_2} = g_i \cdot \frac{\mathbf{R}\mathbf{x}_i}{\sqrt{(\mathbf{R}\mathbf{x}_i)^T \mathbf{R}\mathbf{x}_i}}$
$= g_i \cdot \frac{\mathbf{R}\mathbf{x}_i}{\sqrt{(\mathbf{x}_i)^T \mathbf{R}^T \mathbf{R}\mathbf{x}_i}} = \mathbf{R}(g_i \cdot \frac{\mathbf{x}_i}{\sqrt{(\mathbf{x}_i)^T \mathbf{x}_i}}) = \mathbf{R}F(\mathbf{x}_i)$
$F(\mathbf{x}_i + \mathbf{t}) = g_i \frac{\mathbf{x}_i + \mathbf{t}}{\sqrt{(\mathbf{x}_i+\mathbf{t})^T(\mathbf{x}_i+\mathbf{t})}}$
$\neq g_i \frac{\mathbf{x}_i}{\sqrt{(\mathbf{x}_i)^T(\mathbf{x}_i)}} + \mathbf{t} = F(\mathbf{x}_i) + \mathbf{t}$ $\square$

Therefore, the ScaleNorm is rotation equivariant but not translation equivariant. Also, this feature scaling normalization does not collect the mean statistics and consequently it does not contain learnable bias parameters, which impairs the expressiveness of the normalization layer.

Another temporary solution proposed in E-NF [Satorras *et al.*, 2021a] is turning the relative coordinate difference $(\mathbf{x}_i^l - \mathbf{x}_j^l)$ in the second line of Eq. 2 into the scaled relative coordinate difference $\frac{(\mathbf{x}_i^l - \mathbf{x}_j^l)}{\|\mathbf{x}_i^l - \mathbf{x}_j^l\|_2}$. This solution is not a real normalization layer since it does not contain the re-scaling operation and $(\mathbf{x}_i^l - \mathbf{x}_j^l)$ is not a direct neural network output in EGNN. Consequently, no recovery of the coordinate difference norm would hurt the expressiveness of the networks as it results in some geometry and force information loss. In addition, this method still cannot stabilize the training dynamics of EGNN on real large particle systems.

4.3 New Geometric Normalization Layer GeoNorm

The very first challenge we need to solve is finding summary statistics $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ in particle systems that make geometric sense. Inspired by point cloud geometry observations, two important geometric notions are emerged as strong candidates for $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$. First, we adopt the center of mass of the particle system as $\mu(\mathbf{x})$. Since our problem setting assumes that each particle has exactly the same mass to simplify the problem setting, the center of mass is equal to the geometric center of the particle system such that $\mu(\mathbf{x}) = \frac{\sum_{i=1}^{N} \mathbf{x}_i}{N}$.

For the standard deviation $\sigma(\mathbf{x})$, we adopt the euclidean distance between each particle and the geometric center such that $\sigma(\mathbf{x}_i) = \|\mathbf{x}_i - \mu(\mathbf{x})\|_2 = \sqrt{(\mathbf{x}_i - \mu(\mathbf{x}))^T(\mathbf{x}_i - \mu(\mathbf{x}))}$. Here we abuse the notation σ a bit for alignment with the terminologies used in the vanilla LN setting although the selected $\sigma(\mathbf{x}_i)$ is not a typical standard deviation. We illustrate these two geometric notions in Figure 4. Importantly, these two statistics include fundamental geometric information. The geometric center $\mu(\mathbf{x})$ reflects the global state of the whole particle system and the $\sigma(\mathbf{x}_i)$ describes how distant from each particle to the geometric center. Furthermore, these two geometric notions enjoy great SE(3) equivariance properties, which brings extra convenience for designing SE(3) equivariant normalization layers. Therefore, we present the following two lemmas:

Lemma 3. *The geometric center $\mu(\mathbf{x})$ is both rotation equivariant and translation equivariant.*

Lemma 4. *The standard deviation $\sigma(\mathbf{x}_i)$ is both rotation invariant and translation invariant.*

The proof procedures of the above two lemmas are shown in the Appendix.

To consider the SE(3) equivariance guarantee of GeoNorm design, our strategy is splitting the normalization layer into two independent operations. The first operation is the normalization that projects geometric vectors to the standard normal distribution. The second operation is the re-scaling operation that learns to recover the normalized vectors back to the original. The general idea is turning the first operation into a rotation equivariant and translation invariant operation and then adding the translation back with a rotation-equivariant re-scaling operation. From the Lemma 2, we know that the feature scaling norm is rotation equivariant but not translation equivariant. Inspired by [Satorras *et al.*, 2021a] and the two previously selected geometric summary statistics, we modify the feature scaling operation in Eq. 4 to be a rotation equivariant and translation invariant operation $\frac{\mathbf{x}_i^l - \mu(\mathbf{x}^l)}{\|\mathbf{x}_i^l - \mu(\mathbf{x}^l)\|_2} = \frac{\mathbf{x}_i^l - \mu(\mathbf{x}^l)}{\sigma(\mathbf{x}_i^l)}$. We present the following lemma:

Lemma 5. *The feature scaling of coordinate difference between each particle coordinate and the geometric center is rotation equivariant and translation invariant.*

The proof of this lemma is shown in the Appendix. At this time, the feature scaling norm becomes $g_i \cdot \frac{\mathbf{x}_i^l - \mu(\mathbf{x}^l)}{\sigma(\mathbf{x}_i^l)}$.

The remaining problem is how to add the missing translation back in the re-scaling operation while still maintaining the rotation equivariance. Note that since the geometric center has been removed from position coordinates, we need a learnable bias to recover the geometric center such that $g_i \cdot \frac{\mathbf{x}_i^l - \mu(\mathbf{x}^l)}{\sigma(\mathbf{x}_i^l)} + \mathbf{b}_i$. However, adding the additional bias vector $\mathbf{b}_i$ would break the rotation equivariance. A straightforward solution to solve this tricky issue is multiplying a rotation equivariant and translation invariant vector on $\mathbf{b}$ and replace the bias vector $\mathbf{b}_i$ with a scalar parameter b_i. Therefore, we multiply a direction vector $\frac{\mu(\mathbf{x}^l) - \mu(\mathbf{x}^0)}{\|\mu(\mathbf{x}^l) - \mu(\mathbf{x}^0)\|_2}$ to b_i such that the norm becomes $g_i \cdot \frac{\mathbf{x}_i^l - \mu(\mathbf{x}^l)}{\sigma(\mathbf{x}_i^l)} + \frac{\mu(\mathbf{x}^l) - \mu(\mathbf{x}^0)}{\|\mu(\mathbf{x}^l) - \mu(\mathbf{x}^0)\|_2} \cdot b_i$ where $\mu(\mathbf{x}^0)$ denotes the computed geometric center of the input state of particle system (not the intermediate layer l state).

Note that we apply $\frac{\mu(\mathbf{x}^l) - \mu(\mathbf{x}^0)}{\|\mu(\mathbf{x}^l) - \mu(\mathbf{x}^0)\|_2}$ as the bias direction instead of using $\frac{\mathbf{x}_i^l - \mu(\mathbf{x}^l)}{\|\mathbf{x}_i^l - \mu(\mathbf{x}^l)\|_2}$ and $\frac{\mathbf{x}_i^l - \mu(\mathbf{x}^0)}{\|\mathbf{x}_i^l - \mu(\mathbf{x}^0)\|_2}$. If we use $\frac{\mathbf{x}_i^l - \mu(\mathbf{x}^l)}{\|\mathbf{x}_i^l - \mu(\mathbf{x}^l)\|_2}$, then the bias b_i learning would degenerate into the g_i learning since the mean removal direction and given bias direction are exactly the same, which makes the bias learning very trivial. If we use $\frac{\mathbf{x}_i^l - \mu(\mathbf{x}^0)}{\|\mathbf{x}_i^l - \mu(\mathbf{x}^0)\|_2}$, then GeoNorm would be unstable particularly when a single particle position vector $\mathbf{x}_i^l$ has extreme values. Therefore, using $\frac{\mu(\mathbf{x}^l) - \mu(\mathbf{x}^0)}{\|\mu(\mathbf{x}^l) - \mu(\mathbf{x}^0)\|_2}$ would avoid both problems. However, this leads to a direction inconsistency issue between the mean removal direction $\frac{\mathbf{x}_i^l - \mu(\mathbf{x}^l)}{\|\mathbf{x}_i^l - \mu(\mathbf{x}^l)\|_2}$ and the bias learning direction $\frac{\mu(\mathbf{x}^l) - \mu(\mathbf{x}^0)}{\|\mu(\mathbf{x}^l) - \mu(\mathbf{x}^0)\|_2}$. The scalar bias b_i is not able to calibrate the bias direction to align with the mean removal direction. Fortunately, this problem can be perfectly solved by adding an additional $\mu(\mathbf{x}^0)$ since $\frac{\mu(\mathbf{x}^l) - \mu(\mathbf{x}^0)}{\|\mu(\mathbf{x}^l) - \mu(\mathbf{x}^0)\|_2} b_i + \mu(\mathbf{x}^0)$ can perfectly recover the removed $\mu(\mathbf{x}^l)$ with b_i learns different information from g_i (No degeneracy). We illustrate this mechanism in Figure 3 (d).

To conclude, the complete solution of our GeoNorm normalization layer is as follows:

$$
\begin{aligned}
&\mu(\mathbf{x}^l) = \frac{\sum_{i=1}^{N} \mathbf{x}_i^l}{N}, \sigma(\mathbf{x}_i^l) = \|\mathbf{x}_i^l - \mu(\mathbf{x}^l)\|_2, \\
&F(\mathbf{x}_i^l) = g_i \cdot (\frac{\mathbf{x}_i^l - \mu(\mathbf{x}^l)}{\sigma(\mathbf{x}_i^l)}) + \frac{\mu(\mathbf{x}^l) - \mu_0}{\|\mu(\mathbf{x}^l) - \mu_0\|_2} \cdot b_i + \mu_0. \\
&\mathbf{x}_i^{l+1} = F(\mathbf{x}_i^l), \mu_0 = \mu(\mathbf{x}^0), g_i \in \mathbb{R}, b_i \in \mathbb{R}
\end{aligned}
\tag{5}
$$

Then we prove the following theorem:

Theorem 1. GeoNorm *is an SE(3) equivariant normalization layer, which is both rotation-equivariant and translation-equivariant.*

We show the proof of the above theorem as follows:

Proof.
$$
\begin{aligned}
F(\mathbf{R}\mathbf{x}_i + \mathbf{t}) &= g_i \cdot (\tfrac{\mathbf{R}\mathbf{x}_i + \mathbf{t} - \mathbf{R}\mu(\mathbf{x}) - \mathbf{t}}{\sigma(\mathbf{x}_i)}) \\
&+ \tfrac{\mathbf{R}\mu(\mathbf{x}) + \mathbf{t} - \mathbf{R}\mu_0 - \mathbf{t}}{\|\mu(\mathbf{x}) - \mu_0\|_2} \cdot b_i + \mathbf{R}\mu_0 + \mathbf{t} \\
&= g_i \cdot \mathbf{R}(\tfrac{\mathbf{x}_i - \mu(\mathbf{x})}{\sigma(\mathbf{x}_i)}) + \mathbf{R}\tfrac{\mu(\mathbf{x}) - \mu_0}{\|\mu(\mathbf{x}) - \mu_0\|_2} \cdot b_i + \mathbf{R}\mu_0 + \mathbf{t} \\
&= \mathbf{R}(g_i \cdot (\tfrac{\mathbf{x}_i - \mu(\mathbf{x})}{\sigma(\mathbf{x}_i)}) + \tfrac{\mu(\mathbf{x}) - \mu_0}{\|\mu(\mathbf{x}) - \mu_0\|_2} \cdot b_i + \mu_0) + \mathbf{t} \\
&= \mathbf{R}F(\mathbf{x}_i) + \mathbf{t}
\end{aligned}
$$
□

Therefore, GeoNorm is an SE(3) equivariant normalization layer which would not break the whole SE(3) equivariance of the particle system.

4.4 Further Discussions

Although we obtain an SE(3) equivariant normalization layer, a new challenge arises in the bias learning design. Note that in Eq. 5, the bias b_i is a scalar parameter, which is only able to learn scalar information such as distance particularly the

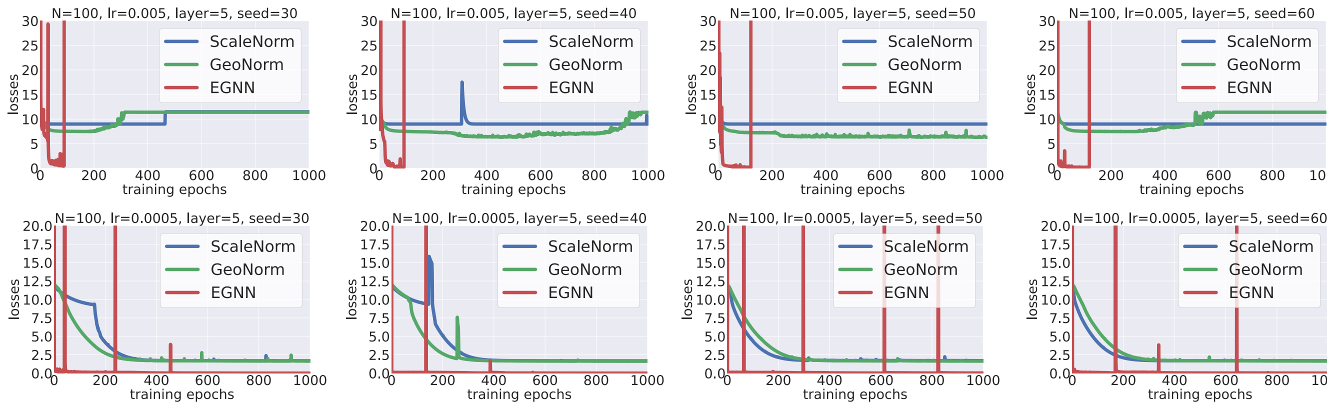

Figure 5: Comparisons of training dynamics of EGNN equipped with GEONORM and other mentioned baseline methods over different random seeds and learning rates. The training dynamics of GEONORM is highlighted by red.

direction information is determined by both $\frac{\mu(\mathbf{x})-\mu_0}{\|\mu(\mathbf{x})-\mu_0\|_2}$ and μ_0. This design apparently downgrades the expressiveness of GEONORM although we avoid the bias b_i to be degenerated into the re-scaling parameter g_i. In contrast, if we want to transform the bias b_i into a vector parameter $\mathbf{b}_i \in \mathbb{R}^3$, the rotation matrix $\mathbf{R}$ cannot be extracted to the front to ensure the rotation equivariance of GEONORM. To sum up, there exists an temporary trade-off issue between expressiveness and SE(3) equivariance guarantee for the current GEONORM design. To enable the bias parameter to learn both direction and distance information, we need to come up with a new solution that set biases to be learnable vectors instead of scalars meanwhile not breaking the rotation equivariance of GEONORM. Since this is a tough challenge that requires tremendous additional research efforts, it is hard to cover the corresponding solutions in this work. Therefore, we leave this problem as an important future work to explore.

5 Experiments

Dataset & Experimental Setting. We also adopt the 3D extension [Fuchs *et al.*, 2020] of the N-body system simulation dataset provided by Kipf [Kipf *et al.*, 2018]. It provides simulation trajectories of charged particle systems following simple physical force rules. Specifically, we train EGNNs on thousands of provided training trajectories that can generalize to the validation and test sets containing 2000 trajectories. We utilize the MSE (Mean-Squared-Error) loss as the evaluation metric to measure the state prediction accuracy between predicted particle positions and ground-truth particle positions after 1000 steps. Unlike the previous works adopting small particle systems ($N = 3, 5$), we set large system simulations with $N = 20, 50, 100$ particles to test the model training stability. The two commonly used datasets QM9 and MD-17 are not used in this work since the involved particle system size is around 5-10 (small chemical molecular conformers), which is not large enough to test the stability of large EGNN training. In addition to stability testing, we also evaluate whether GEONORM satisfy the SE(3) equivariance constraints. During testing, we apply random 3D rotations and translations to mutate the inputs. Then we report MSE scores of each selected baseline after SE(3) transformations mutations. We mainly report mean, median and incremental rotational and translational errors (RTE).

Implementation Details, Hyperparameters and Configurations. In preliminary sections, we explore the training instability issue by setting the number of particles $n = 20, 50, 100$, the number of layers $l = 3, 5, 8, 10$. For the equivariant GNN model selection, we adopt the alternative velocity version of EGNN as our backbone model (Note that the velocity vector update is omitted from Eq. 2, please refer details in [Satorras *et al.*, 2021b]). In main experiments for normalization comparisons, there are mainly two important configurations which are random seed and learning rate. We select random seed$= 30, 40, 50, 60$ to test different weight initialization. For learning rate, we select $lr = 5e-3, 5e-4$ to test different learning rates. The optimization is conducted using Adam [Kingma and Ba, 2015] and the weight decay of 1×10^{-12}. The feature dimensions in EGNN is 64 as its original paper reported. All algorithms have been trained under the same conditions with batch size of 50. All experiments on N-body particle system use the default trajectory sampling method in EGNN [Satorras *et al.*, 2021b]. We train all models with 1000 total number of epochs with batch size 50. The target we optimize is the averaged Mean Squared Error (MSE) between predicted positions and corresponding ground-truth positions. A single Tesla V100 GPU is needed to reproduce each hyperparameter's experimental results. Running each configuration may take up to 6 hours.

Experimental Results. We mainly compare the training stability of vanilla EGNN, the feature scaling norm SCALENORM and our GEONORM. From the training dynamics shown in Figure 5, we can see that both GEONORM and SCALENORM can stabilize the training dynamics over all system configurations while the vanilla EGNN easily encounters numerical explosions. Furthermore, compared to SCALENORM, our GEONORM perform better convergence properties in optimizations over different settings. We clearly see that in most cases GEONORM (highlighted by red curves) can reach lower training errors compared to SCALENORM. Even in some cases, SCALENORM may converge faster than

	$N=100, l=5, seed=30$							
Model	Mean $R_{err}(^\circ)$	Median $R_{err}(^\circ)$	Max $R_{err}(^\circ)$	ΔR_{err}	Mean T_{err}	Median T_{err}	Max T_{err}	ΔT_{err}
EGNN-LN	77.6337	76.2877	120.6114	70.2623	76.4022	76.5329	108.3657	69.0064
EGNN-SN	**11.6812**	**11.6423**	**12.8005**	**0.00**	79.9684	79.0553	113.0788	68.3327
EGNN-GN	**12.0141**	**11.9806**	**13.1563**	**0.00**	**12.0149**	**11.9934**	**13.1494**	**0.00**
	$N=100, l=3, seed=30$							
Model	Mean $R_{err}(^\circ)$	Median $R_{err}(^\circ)$	Max $R_{err}(^\circ)$	ΔR_{err}	Mean T_{err}	Median T_{err}	Max T_{err}	ΔT_{err}
EGNN-LN	75.7769	77.0998	114.7744	68.4406	74.4919	73.3482	120.7420	67.1195
EGNN-SN	**11.6409**	**11.5987**	**12.6707**	**0.00**	81.2245	81.2323	108.4697	69.7046
EGNN-GN	**12.0191**	**11.997**	**13.0744**	**0.00**	**12.01149**	**12.00350**	**13.1469**	**0.00**
	$N=50, l=5, seed=30$							
Model	Mean $R_{err}(^\circ)$	Median $R_{err}(^\circ)$	Max $R_{err}(^\circ)$	ΔR_{err}	Mean T_{err}	Median T_{err}	Max T_{err}	ΔT_{err}
EGNN-LN	43.4534	43.2547	74.5962	38.1556	42.3317	41.4941	66.3715	37.0249
EGNN-SN	**8.5636**	**8.5396**	**9.5382**	**0.00**	51.8175	50.1510	68.9853	43.2368
EGNN-GN	**9.0896**	**9.0807**	**9.9733**	**0.00**	**9.1155**	**9.1257**	**10.0839**	**0.00**

	$N=100, l=5, seed=40$							
Model	Mean $R_{err}(^\circ)$	Median $R_{err}(^\circ)$	Max $R_{err}(^\circ)$	ΔR_{err}	Mean T_{err}	Median T_{err}	Max T_{err}	ΔT_{err}
EGNN-LN	79.7006	78.9167	119.5280	72.3209	80.5087	80.8898	117.7024	73.1220
EGNN-SN	**11.7812**	**11.7750**	**12.9575**	**0.00**	84.7791	82.4171	139.6815	73.0484
EGNN-GN	**11.9976**	**11.9577**	**13.1755**	**0.00**	**12.0059**	**11.9898**	**13.0797**	**0.00**
	$N=100, l=3, seed=40$							
Model	Mean $R_{err}(^\circ)$	Median $R_{err}(^\circ)$	Max $R_{err}(^\circ)$	ΔR_{err}	Mean T_{err}	Median T_{err}	Max T_{err}	ΔT_{err}
EGNN-LN	76.2032	73.3847	116.7627	68.8349	73.0035	72.6746	104.7594	65.6595
EGNN-SN	**11.4117**	**11.3684**	**12.5119**	**0.00**	81.2904	79.4168	120.2692	69.9343
EGNN-GN	**11.9349**	**11.9154**	**13.1020**	**0.00**	**11.8147**	**11.8018**	**12.8062**	**0.00**
	$N=50, l=5, seed=40$							
Model	Mean $R_{err}(^\circ)$	Median $R_{err}(^\circ)$	Max $R_{err}(^\circ)$	ΔR_{err}	Mean T_{err}	Median T_{err}	Max T_{err}	ΔT_{err}
EGNN-LN	42.7479	41.4761	75.6686	37.4484	43.5382	41.7819	71.8296	38.2338
EGNN-SN	**8.6886**	**8.6783**	**9.6922**	**0.00**	52.4891	50.8263	84.0599	43.7528
EGNN-GN	**9.1182**	**9.0935**	**10.1146**	**0.00**	**9.0662**	**9.0217**	**10.0238**	**0.00**

Table 1: In this table, all reported results are MSE scores between predicted positions and ground-truth positions. Therefore, each score in the table lower is better. EGNN-LN, EGNN-SN, EGNN-GN denote the LAYERNORM, SCALENORM and GEONORM concatenated with EGNN training. We mainly report 8 metrics: Mean $R_{err}(^\circ)$, Median $R_{err}(^\circ)$, Max $R_{err}(^\circ)$, ΔR_{err}, Mean T_{err}, Median T_{err}, Max T_{err}, ΔT_{err}, which are reporting the mean, median, max and incremental RTE errors over all testing samples after rotation and translation mutations. Note that the most important metrics are ΔR_{err} and ΔT_{err} which reflect additional MSE errors caused by SE(3) transformations to the original testing input. System configurations include the number of particles $N = 50, 100$, the EGNN layers $l = 3, 5$ and the random seed $30, 40$.

GEONORM, GEONORM can still catch up with the performance of SCALENORM in few more epochs. This demonstrates the effectiveness of GEONORM compared to baseline methods. However, we can also observe that GEONORM is still not good enough since the training errors cannot be further minimized at some training stage. Although the vanilla EGNN is very unstable, its training errors can be minimized to very decent accuracy at some certain epochs. This may suggest that GEONORM is still not expressive enough and some operations may hurt the expressiveness of the original EGNN.

From the table 5, we can see that using the LAYERNORM cannot preserve both rotation equivariance and translation equivariance in all experimental configurations. And the SCALENORM successfully preserves the rotation equivariance but fail to preserve the translation equivariance in all settings. Only our GEONORM preserves the rotation equivariance and the translation equivariance at the same time. Although the SCALENORM enjoys a bit lower rotational errors in some system configurations, it also suffers from large translational errors. And we surprisingly find that the rotational errors of GEONORM become relatively lower when the system becomes larger (from the bottom to top). This might indicate the superiority of GEONORM on large particle system compared to baseline methods. To sum up, GEONORM is the only SE(3) equivariant normalization layer among these methods. Equipping EGNN models with GEONORM would not hurt its generalization to testing trajectories with respect to rotation and translation input mutations. Note that we only train few epochs to produce these experimental results since we focus on the generalization error instead of deriving the final optimal performance. In conclusions, we show that our GEONORM can stabilize the training dynamics of EGNN on larger systems while additionally satisfying SE(3) group equivariance properties. More experiments on different variants are shown in Appendix since the experimental settings are hard to be exhausted.

6 Limitations and Future Works

The most critical limitation is that GEONORM is only SE(3) equivariant but not permutation equivariant, which is another fundamental equivariance property that GNN models should ensure. This may hurt the model generalization when permuting the input order of particles. Secondly, to stabilize the training of large EGNNs, we should also incorporate other important techniques to further enhance the neural architectures from different perspectives. However, this necessitates additional research efforts and can result in a series of works. Additionally, it would be beneficial to conduct more experiments on real-world large particle systems. Unfortunately, most popular particle system simulation datasets only include small systems, making large real-world particle system simulation datasets extremely limited. Moreover, GEONORM needs to store the initial state or the initial geometric center, which occupies additional storage space. Hence, more advanced techniques should be explored to avoid this issue in the following works. Last but not the least, more diverse equivariant models should be involved in the discussions. For example, some of the models might be E(3) equivariant, which further include reflection equivariance. And some models might follow a special case of SE(3) equivariance (e.g. translation invariant). Therefore, more comprehensive investigations should be included in the future works.

7 Conclusion

In this work, we investigate geometric normalization techniques for deep SE(3) equivariant graph neural networks on larger particle system simulation tasks. We demonstrate that both naive and existing solutions cannot strictly satisfy the SE(3) equivariance geometric constraints. To address this issue and maintain both stable training dynamics and SE(3) equivariance, we propose a new technique called GEONORM as a replacement for previous normalization strategies. The effectiveness of our approach is supported by theoretical analysis and experimental results.

Acknowledgments

The work described in this paper was partially supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (RGC GRF 2151185). Ziqiao Meng and Liang Zeng contributed equally to this work.

References

[Anderson *et al.*, 2019] Brandon Anderson, Truong Son Hy, and Risi Kondor. Cormorant: Covariant molecular neural networks. In *NeurIPS*, volume 32. Curran Associates, Inc., 2019.

[Ba *et al.*, 2016] Lei Jimmy Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization. *CoRR*, abs/1607.06450, 2016.

[Batatia *et al.*, 2022] Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gabor Csanyi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. In *NeurIPS*, volume 35, pages 11423–11436. Curran Associates, Inc., 2022.

[Batzner *et al.*, 2022] Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1), 2022.

[Brandstetter *et al.*, 2022] Johannes Brandstetter, Rob Hesselink, Elise van der Pol, Erik J Bekkers, and Max Welling. Geometric and physical quantities improve e(3) equivariant message passing. In *ICLR*, 2022.

[Brock *et al.*, 2021] Andrew Brock, Soham De, and Samuel L Smith. Characterizing signal propagation to close the performance gap in unnormalized resnets. In *ICLR*, 2021.

[Cai *et al.*, 2021] Tianle Cai, Shengjie Luo, Keyulu Xu, Di He, Tie-Yan Liu, and Liwei Wang. Graphnorm: A principled approach to accelerating graph neural network training. In *ICML*, volume 139 of *Proceedings of Machine Learning Research*, pages 1204–1215. PMLR, 18–24 Jul 2021.

[Daneshmand *et al.*, 2020] Hadi Daneshmand, Jonas Kohler, Francis Bach, Thomas Hofmann, and Aurelien Lucchi. Batch normalization provably avoids ranks collapse for randomly initialised deep networks. In *NeurIPS*, volume 33, pages 18387–18398. Curran Associates, Inc., 2020.

[Deng *et al.*, 2021] Congyue Deng, Or Litany, Yueqi Duan, Adrien Poulenard, Andrea Tagliasacchi, and Leonidas J. Guibas. Vector neurons: A general framework for so(3)-equivariant networks. In *ICCV*, pages 12180–12189. IEEE, 2021.

[Du *et al.*, 2022] Weitao Du, He Zhang, Yuanqi Du, Qi Meng, Wei Chen, Nanning Zheng, Bin Shao, and Tie-Yan Liu. SE(3) equivariant graph neural networks with complete local frames. In *ICML*, volume 162 of *Proceedings of Machine Learning Research*, pages 5583–5608. PMLR, 17–23 Jul 2022.

[Dwivedi *et al.*, 2023] Vijay Prakash Dwivedi, Chaitanya K. Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. *J. Mach. Learn. Res.*, 24:43:1–43:48, 2023.

[Finzi *et al.*, 2020] Marc Finzi, Samuel Stanton, Pavel Izmailov, and Andrew Gordon Wilson. Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data. In *ICML*, volume 119 of *Proceedings of Machine Learning Research*, pages 3165–3176. PMLR, 2020.

[Frank *et al.*, 2022] Thorben Frank, Oliver Thorsten Unke, and Klaus Robert Muller. So3krates: Equivariant attention for interactions on arbitrary length-scales in molecular systems. In *NeurIPS*, 2022.

[Fuchs *et al.*, 2020] Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. Se(3)-transformers: 3d roto-translation equivariant attention networks. In *NeurIPS*, volume 33, pages 1970–1981. Curran Associates, Inc., 2020.

[Ganea *et al.*, 2022] Octavian-Eugen Ganea, Xinyuan Huang, Charlotte Bunne, Yatao Bian, Regina Barzilay, Tommi S. Jaakkola, and Andreas Krause. Independent SE(3)-equivariant models for end-to-end rigid protein docking. In *ICLR*, 2022.

[Gasteiger *et al.*, 2020] Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. In *ICLR*, 2020.

[Hamilton *et al.*, 2017] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *NIPS*, volume 30. Curran Associates, Inc., 2017.

[Hoogeboom *et al.*, 2022] Emiel Hoogeboom, Victor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d. In *ICML 2022*, volume 162 of *Proceedings of Machine Learning Research*, pages 8867–8887. PMLR, 2022.

[Huang *et al.*, 2022] Wenbing Huang, Jiaqi Han, Yu Rong, Tingyang Xu, Fuchun Sun, and Junzhou Huang. Equivariant graph mechanics networks with constraints. In *ICLR*, 2022.

[Hutchinson *et al.*, 2021] Michael J Hutchinson, Charline Le Lan, Sheheryar Zaidi, Emilien Dupont, Yee Whye Teh, and Hyunjik Kim. Lietransformer: Equivariant self-attention for lie groups. In *ICML*, volume 139 of *Proceedings of Machine Learning Research*, pages 4533–4543. PMLR, 18–24 Jul 2021.

[Ioffe and Szegedy, 2015] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France, 07–09 Jul 2015. PMLR.

[Jiao *et al.*, 2023] Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, and Yang Liu. Crystal structure prediction by joint equivariant diffusion. In *NeurIPS 2023*, 2023.

[Jing *et al.*, 2021] Bowen Jing, Stephan Eismann, Patricia Suriana, Raphael John Lamarre Townshend, and Ron Dror. Learning from protein structure with geometric vector perceptrons. In *ICLR*, 2021.

[Kingma and Ba, 2015] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.

[Kipf *et al.*, 2018] Thomas Kipf, Ethan Fetaya, Kuan-Chieh Wang, Max Welling, and Richard Zemel. Neural relational inference for interacting systems. In *ICML*, pages 2688–2697. PMLR, 2018.

[Klicpera *et al.*, 2021] Johannes Klicpera, Florian Becker, and Stephan Günnemann. Gemnet: Universal directional graph neural networks for molecules. In *NeurIPS*, 2021.

[Köhler *et al.*, 2020] Jonas Köhler, Leon Klein, and Frank Noe. Equivariant flows: Exact likelihood generative learning for symmetric densities. In *ICML*, pages 5361–5370, 2020.

[Li *et al.*, 2022] Guohao Li, Chenxin Xiong, Guocheng Qian, Ali Thabet, and Bernard Ghanem. DeeperGCN: All you need to train deeper GCNs, 2022.

[Liao and Smidt, 2023] Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. In *ICLR*, 2023.

[Liu *et al.*, 2022] Yi Liu, Limei Wang, Meng Liu, Yuchao Lin, Xuan Zhang, Bora Oztekin, and Shuiwang Ji. Spherical message passing for 3d molecular graphs. In *ICLR*, 2022.

[Meng *et al.*, 2023] Ziqiao Meng, Peilin Zhao, Yang Yu, and Irwin King. A unified view of deep learning for reaction and retrosynthesis prediction: Current status and future challenges. In *IJCAI 2023*, pages 6723–6731. ijcai.org, 2023.

[Miyato *et al.*, 2018] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *ICLR*, 2018.

[Salimans and Kingma, 2016] Tim Salimans and Durk P Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In *NIPS*, volume 29. Curran Associates, Inc., 2016.

[Satorras *et al.*, 2021a] Victor Garcia Satorras, Emiel Hoogeboom, Fabian Bernd Fuchs, Ingmar Posner, and Max Welling. E(n) equivariant normalizing flows. In *NeurIPS*, 2021.

[Satorras *et al.*, 2021b] Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In *ICML*, volume 139 of *Proceedings of Machine Learning Research*, pages 9323–9332. PMLR, 18–24 Jul 2021.

[Schütt *et al.*, 2017] Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In *NIPS*, volume 30. Curran Associates, Inc., 2017.

[Schütt *et al.*, 2021] Kristof Schütt, Oliver Unke, and Michael Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *ICML*, volume 139 of *Proceedings of Machine Learning Research*, pages 9377–9388. PMLR, 18–24 Jul 2021.

[Shen *et al.*, 2020] Wen Shen, Binbin Zhang, Shikun Huang, Zhihua Wei, and Quanshi Zhang. 3d-rotation-equivariant quaternion neural networks. In *ECCV*, volume 12365 of *Lecture Notes in Computer Science*, pages 531–547. Springer, 2020.

[Singh and Krishnan, 2020] Saurabh Singh and Shankar Krishnan. Filter response normalization layer: Eliminating batch dependence in the training of deep neural networks. In *CVPR*, pages 11234–11243. Computer Vision Foundation / IEEE, 2020.

[Song *et al.*, 2024] Zixing Song, Ziqiao Meng, and Irwin King. A diffusion-based pre-training framework for crystal property prediction. In *AAAI 2024*, pages 8993–9001. AAAI Press, 2024.

[Thölke and Fabritiis, 2022] Philipp Thölke and Gianni De Fabritiis. Equivariant transformers for neural network based molecular potentials. In *ICLR*, 2022.

[Thomas *et al.*, 2018] Nathaniel Thomas, Tess E. Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. *CoRR*, abs/1802.08219, 2018.

[Ulyanov *et al.*, 2016] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Instance normalization: The missing ingredient for fast stylization. *CoRR*, abs/1607.08022, 2016.

[Veličković *et al.*, 2018] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *ICLR*, 2018.

[Wu and He, 2018] Yuxin Wu and Kaiming He. Group normalization. In *ECCV*, volume 11217 of *Lecture Notes in Computer Science*, pages 3–19. Springer, 2018.

[Yang *et al.*, 2020] Chaoqi Yang, Ruijie Wang, Shuochao Yao, Shengzhong Liu, and Tarek F. Abdelzaher. Revisiting "over-smoothing" in deep gcns. *CoRR*, abs/2003.13663, 2020.

[Zhao and Akoglu, 2020] Lingxiao Zhao and Leman Akoglu. Pairnorm: Tackling oversmoothing in gnns. In *ICLR*, 2020.

[Zhou *et al.*, 2020] Kaixiong Zhou, Xiao Huang, Yuening Li, Daochen Zha, Rui Chen, and Xia Hu. Towards deeper graph neural networks with differentiable group normalization. In *NeurIPS*, volume 33, pages 4917–4928. Curran Associates, Inc., 2020.

[Zitnick *et al.*, 2022] C. Lawrence Zitnick, Abhishek Das, Adeesh Kolluru, Janice Lan, Muhammed Shuaibi, Anuroop Sriram, Zachary Ward Ulissi, and Brandon M Wood. Spherical channels for modeling atomic interactions. In *NeurIPS*, 2022.