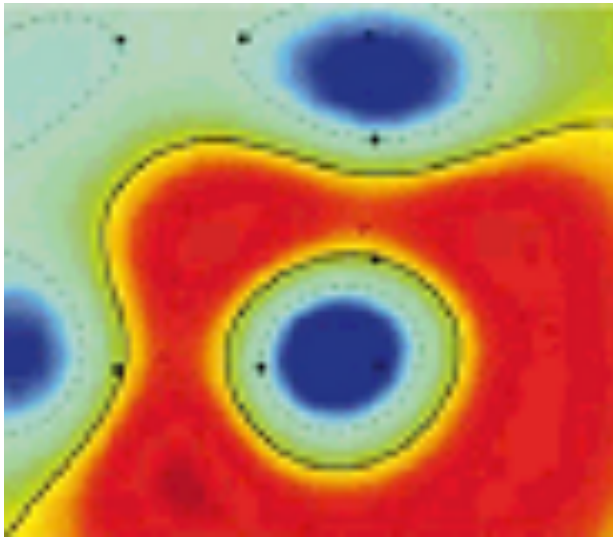
SoSe 2024

Deep Learning 2

Technische Universität Berlin

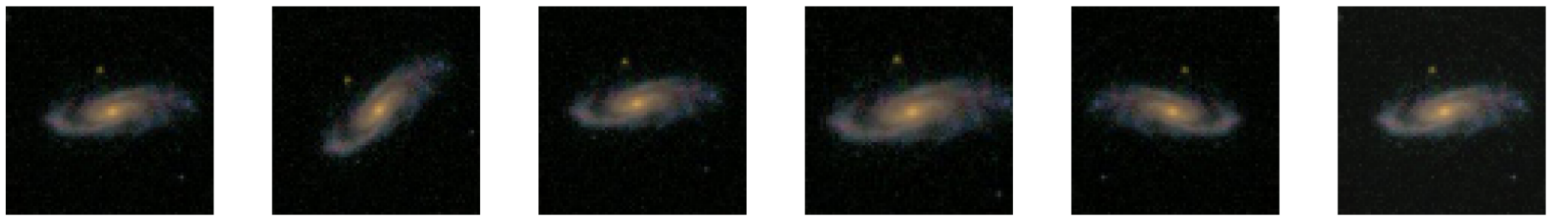


Lecture 6

Equivariant Neural Networks

Symmetry and invariances

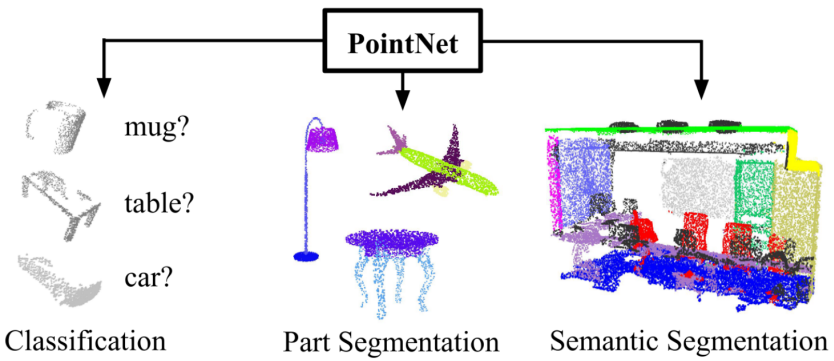
Galaxy classification [Dieleman et al., 2015]



(a) none (b) rotation (c) translation (d) scaling (e) flipping (f) brightness

Figure 6. The five types of random data augmentation used in this model. Note that the effect of translation and brightness adjustment is fairly subtle.

Point clouds [Qi et al., 2017]

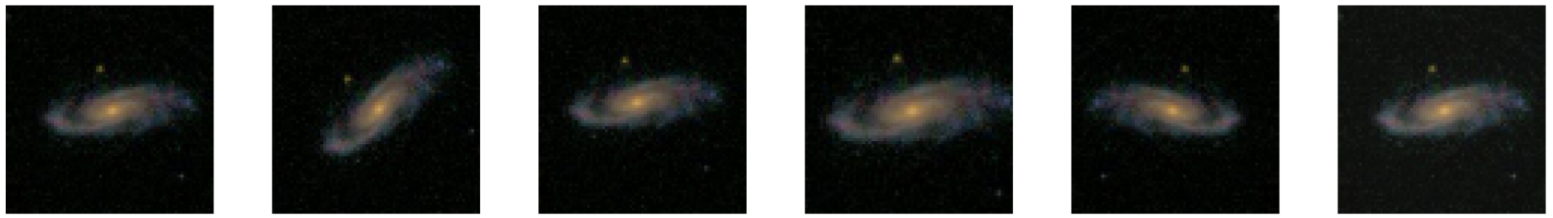


The shape formed by the point cloud is invariant to the order / indexing of points.

Possible solutions

Data augmentation

- One applies the symmetry transformation to the input to learn the invariances and increase robustness.
- It might take longer to train.
- The learned symmetries are not perfect.

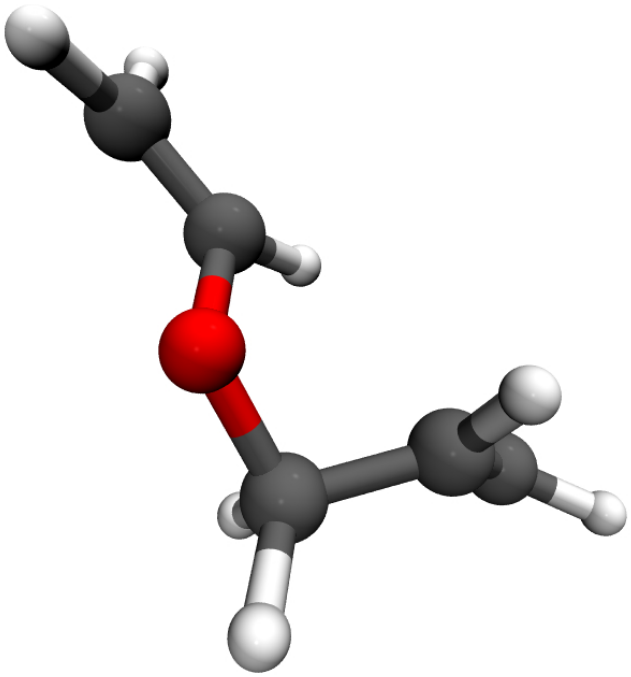


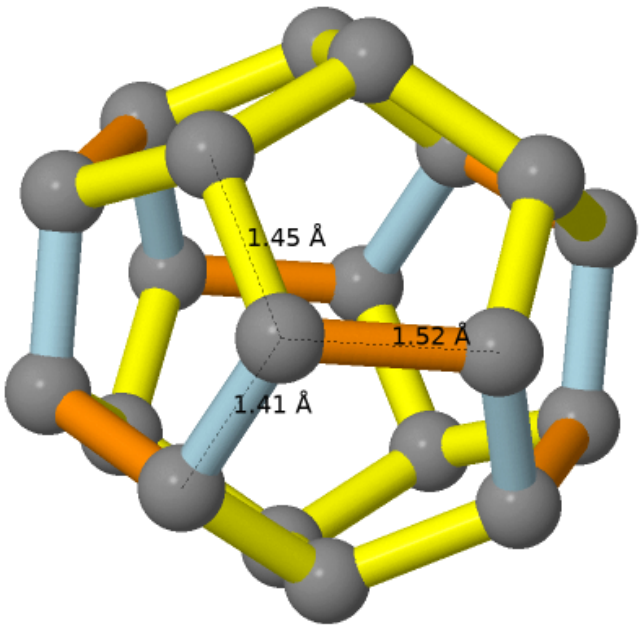
(a) none (b) rotation (c) translation (d) scaling (e) flipping (f) brightness

Figure 6. The five types of random data augmentation used in this model. Note that the effect of translation and brightness adjustment is fairly subtle.

Make the model directly invariant to rotation?

Example: Molecules





The energy of a molecule is invariant to

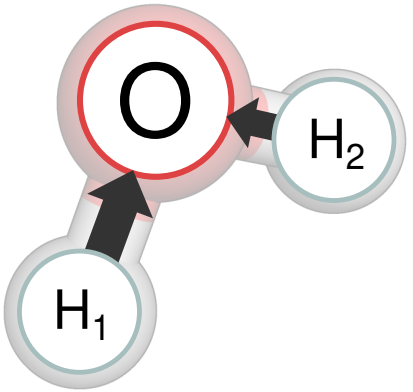
- translation, rotation and reflection
- atom order / indexing

Example: Invariant message-passing neural networks

Message-passing neural network can be applied to 3d molecules with edges being represented by relative positions of atom $\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$:

$$\mathbf{m}_i^{t+1} = \sum_{j \in \mathcal{N}(i)} \mathbf{M}_t(\mathbf{s}_i^t, \mathbf{s}_j^t, \|\vec{r}_{ij}\|)$$

$$\mathbf{s}_i^{t+1} = \mathbf{U}_t(\mathbf{s}_i^t, \mathbf{m}_i^{t+1})$$



In contrast to applications to graph, we can directly define a convolution over 3d space using a continuous filter that can be used as a message function.

The messages are rotational invariant:

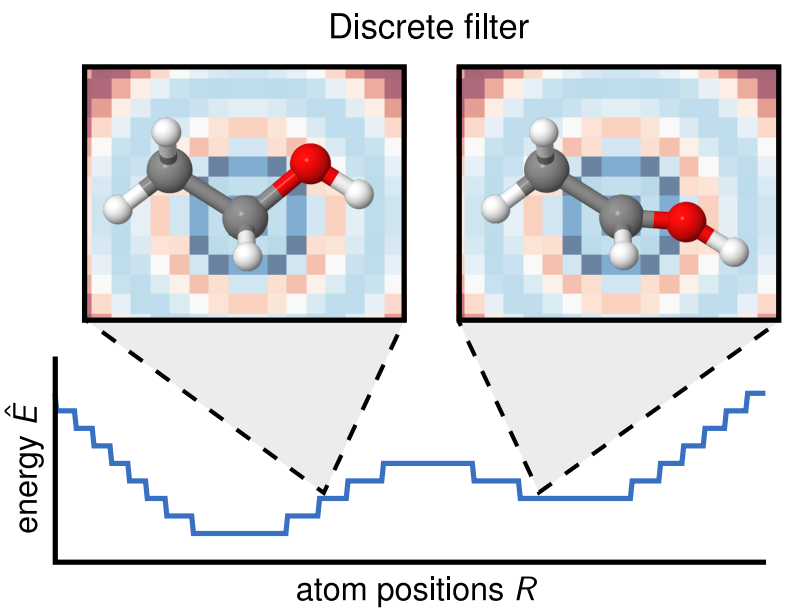
$$\mathbf{M}(\vec{r}_{ij}) = \mathbf{M}(R\,\vec{r}_{ij})$$

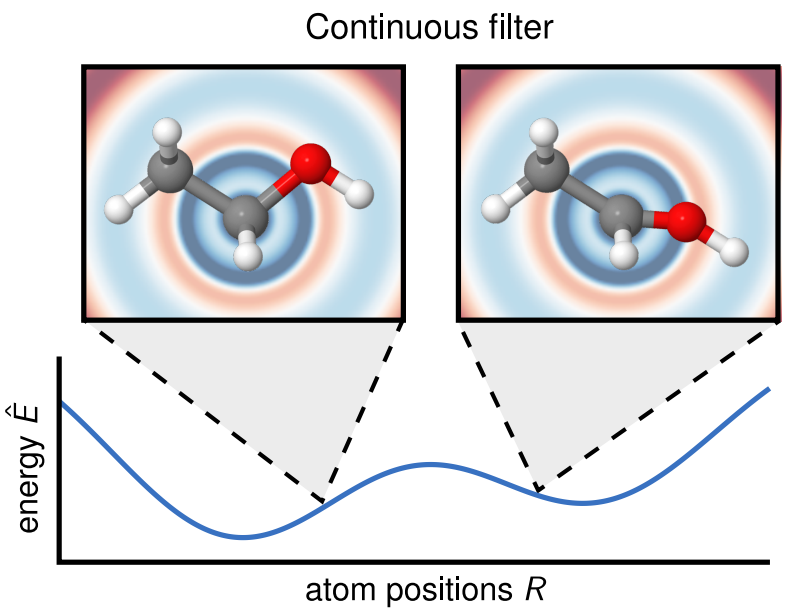
for any rotation R.

The readout function f is invariant to order of indexing:

$$E = \sum_i f(\mathbf{s}_i)$$

Message function: Continuous-filter convolution





$$(\mathbf{x} * W)(\mathbf{r}_i) = \sum_{j=1}^{N_{\text{atom}}} \mathbf{x}_j^{(t)} \circ \underbrace{W^{(t)}_{[\mathbf{r}_i - \mathbf{r}_j]}}_{\text{parameter tensor}}$$

$$(\mathbf{x} * W)(\mathbf{r}_i) = \sum_{j=1}^{N_{\text{atom}}} \mathbf{x}_j^{(t)} \circ \underbrace{W^{(t)}(\|\mathbf{r}_i - \mathbf{r}_j\|)}_{\text{neural network}}$$

[Schütt et al., 2017]

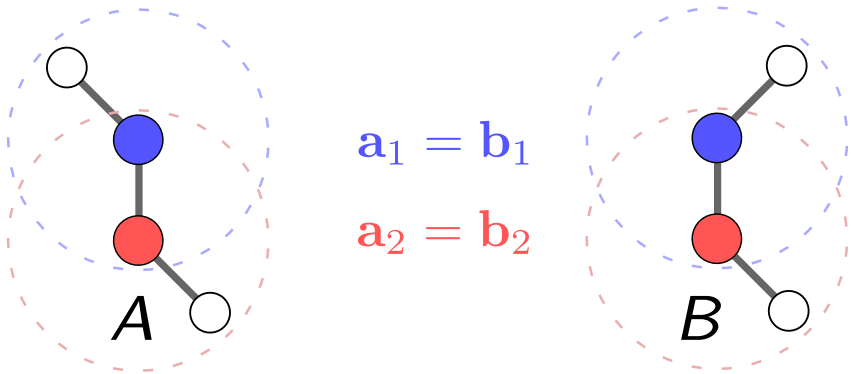
The Picasso problem

The individual parts (eyes, nose, mouth) are there, but the composition is wrong.

When local representations, e.g. of atoms or pixel, are invariant to rotation, translation etc., the neural network might not have enough information to put things back together.

Local symmetry vs global symmetry

Having only nearby particles interact with the message function introduces additional (unwanted) symmetries. For example, distances might not be enough to reconstruct a structure uniquely from local environments:



Representations need to retain some additional directional information!

Rotationally equivariant message passing

Equivariant message passing:

$$\vec{\mathbf{m}}_i^{v,t+1} = \sum_{j \in \mathcal{N}(i)} \vec{\mathbf{M}}_t(\mathbf{s}_i^t, \mathbf{s}_j^t, \vec{\mathbf{v}}_i^t, \vec{\mathbf{v}}_j^t, \vec{r}_{ij})$$

with

$$R\,\vec{\mathbf{M}}(\vec{\mathbf{x}}) = \vec{\mathbf{M}}(R\,\vec{\mathbf{x}})$$

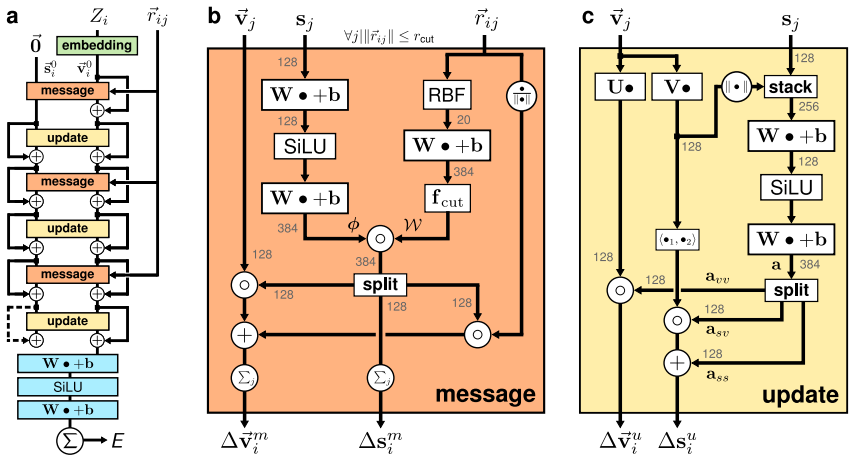
for any rotation R.

Equivariant building blocks:

- Any (nonlinear) function of scalars: $\mathbf{f}(\mathbf{s})$
- Scaling of vectors: $\mathbf{s} \circ \vec{\mathbf{v}}$
- Linear combinations of equivariant vectors: $\mathbf{W}\vec{\mathbf{v}}$
- Vector products: $\vec{\mathbf{v}}_1 \times \vec{\mathbf{v}}_2$
- Scalar products: $\mathbf{s} = \|\vec{\mathbf{v}}\|^2$, $\mathbf{s} = \langle \vec{\mathbf{v}}_1, \vec{\mathbf{v}}_2 \rangle$

[Schütt et al., 2021]

Example: equivariant message passing for molecules



Scalar message

$$\mathbf{m}_i^{t+1} = (\phi_s(\mathbf{s}) * \mathcal{W}_s)_i = \sum_j \underbrace{\phi_s(\mathbf{s}_j)}_{\text{invariant features}} \circ \underbrace{\mathcal{W}_s(\|\vec{r}_{ij}\|)}_{\text{rotationally invariant filter}},$$

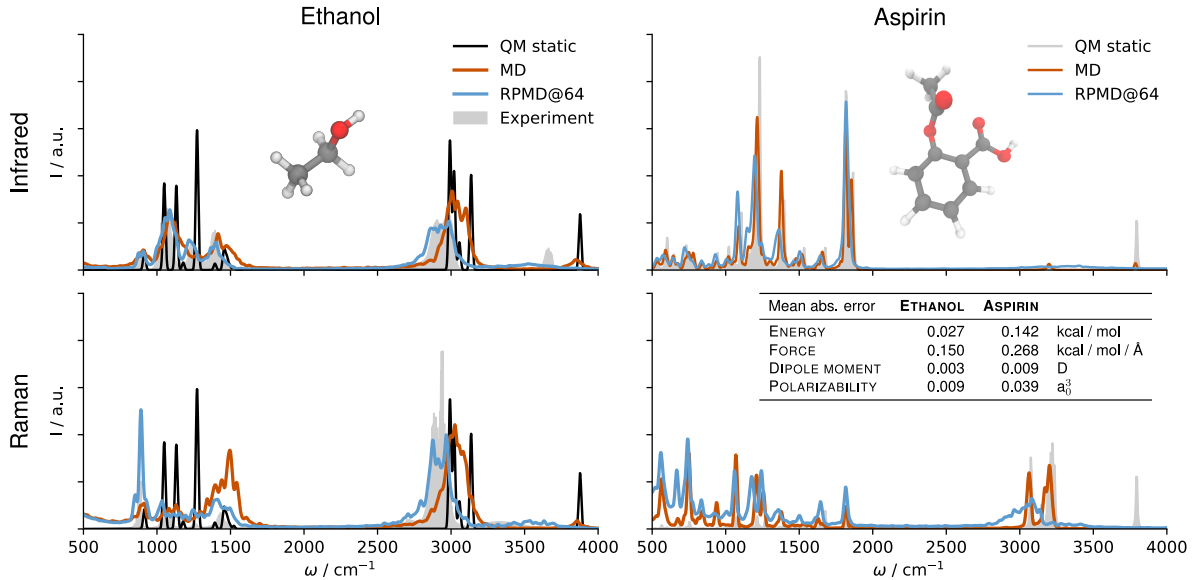
Vectorial message

$$\vec{\mathbf{m}}_i^{v,t+1} = \sum_j \underbrace{\vec{\mathbf{v}}_j \circ \phi_s(\mathbf{s}_j)}_{\text{scaled equivariant features}} \circ \underbrace{\mathcal{W}_{vv}(\|\vec{r}_{ij}\|)}_{\text{rotationally invariant filter}}$$

$$+ \sum_j \underbrace{\phi_{vs}(\mathbf{s}_j)}_{\text{invariant features}} \circ \underbrace{\mathcal{W}'_{vs}(\|\vec{r}_{ij}\|)\frac{\vec{r}_{ij}}{\|\vec{r}_{ij}\|}}_{\text{rotationally equivariant filter}},$$

Predicting equivariant tensorial properties

RPMD @ 64 beads of Aspirin: 25 years $\longrightarrow$ 1 h



Mean abs. error	Ethanol	Aspirin	
Energy	0.027	0.142	kcal / mol
Force	0.150	0.268	kcal / mol / Å
Dipole moment	0.003	0.009	D
Polarizability	0.009	0.039	a_0^3

Tensor rank decomposition

$$T = \sum_{i=1}^{N} \sum_{k=1}^{R} \lambda(\mathbf{s}_i) \vec{\nu}(\vec{\mathbf{v}}_i)_{k,1} \otimes \cdots \otimes \vec{\nu}(\vec{\mathbf{v}}_i)_{k,M}$$

Dipole moment

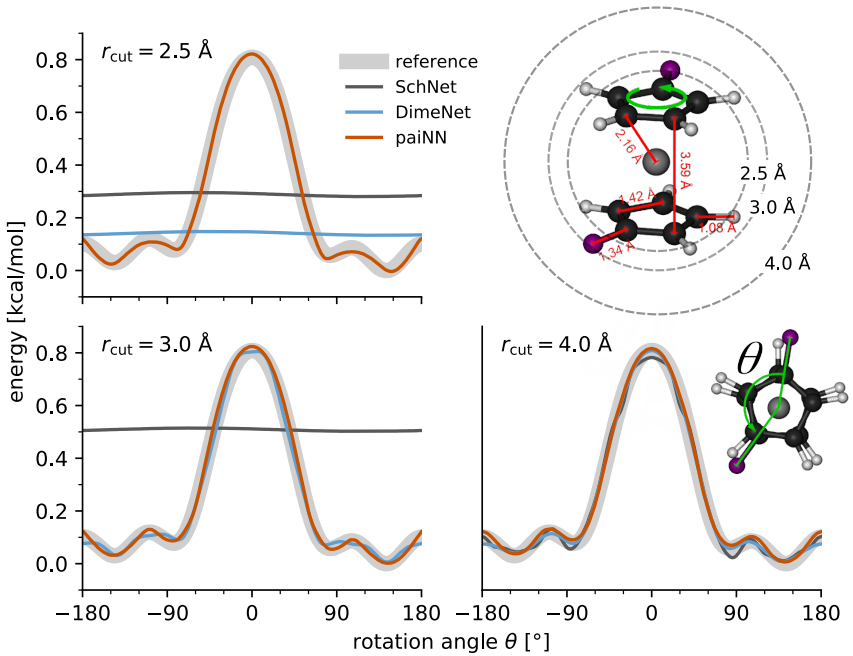
$$\vec{\mu} = \sum_{i=1}^{N} \underbrace{\vec{\mu}_{\text{atom}}(\vec{\mathbf{v}}_i)}_{\text{atomic dipoles}} + \underbrace{q_{\text{atom}}(\mathbf{s}_i)}_{\text{atomic charges}} \vec{r}_i$$

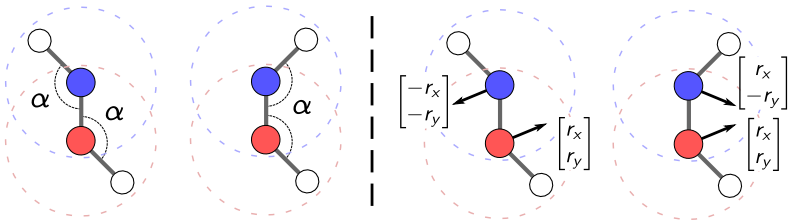
Polarizability tensor

$$\alpha = \sum_{i=1}^{N} \underbrace{\alpha_0(\mathbf{s}_i)\, I_3}_{\text{trace}} + \underbrace{\vec{\nu}(\vec{\mathbf{v}}_i) \otimes \vec{r}_i + \vec{r}_i \otimes \vec{\nu}(\vec{\mathbf{v}}_i)}_{\text{symmetrized outer product}}$$

[Schütt et al., 2021]

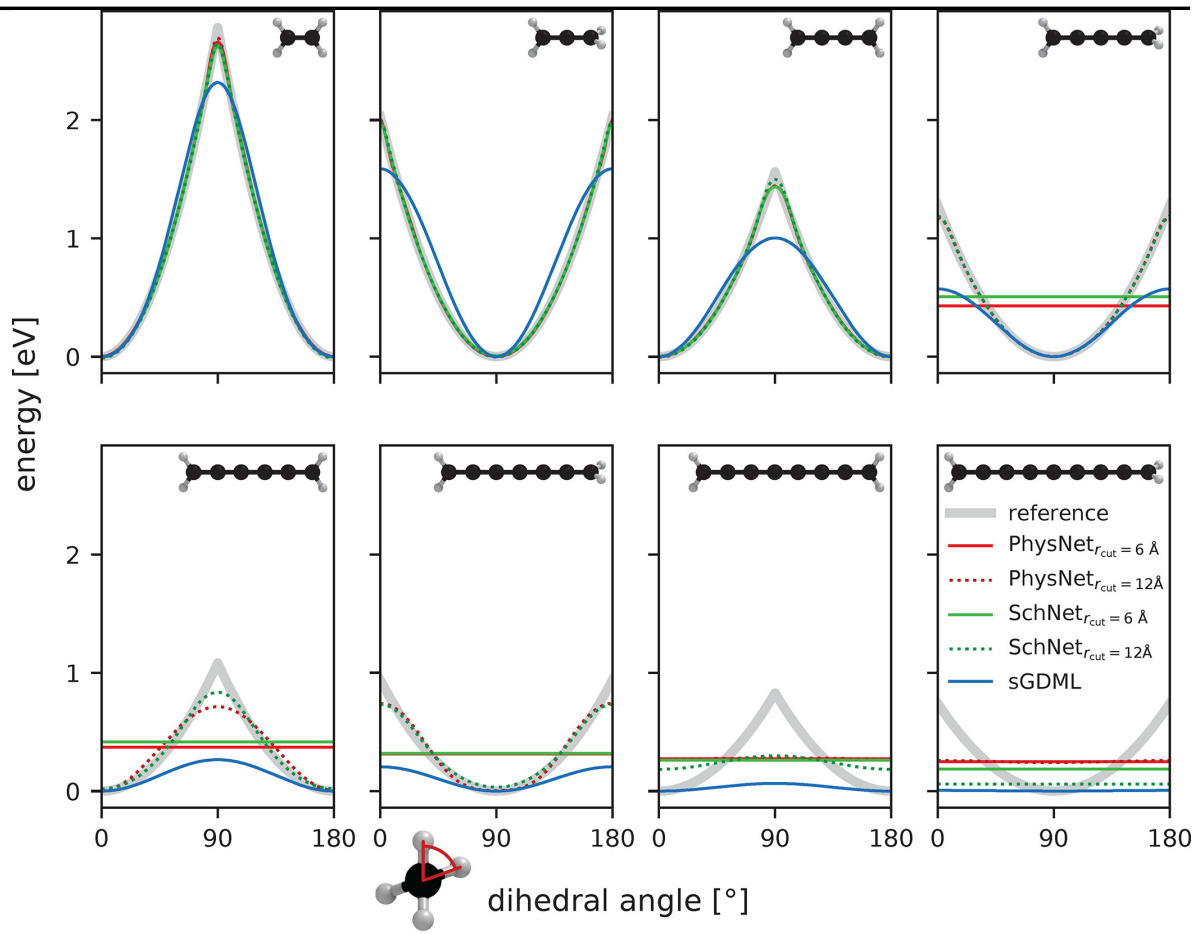
Propagation of directional information





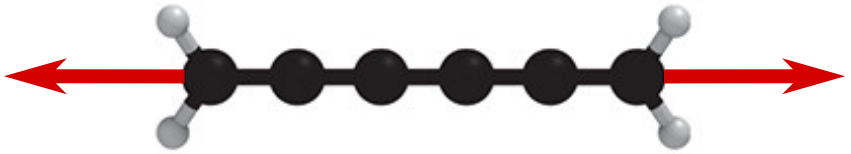
Equivariant message passing is able to propagate directional information beyond the radial cutoff

Vectors might not be enough...



[Unke et al., 2021]

Due to the symmetry of the molecule, all vector features of the carbons will be oriented along the chains, making a propagation of information about the CH2 groups at the end impossible.



Group theory (1)

Group

A group is a pair $(G, \cdot)$ of a set G and a binary operation on G that fulfills:

- Associativity: $\forall a, b, c \in G : (a \cdot b) \cdot c = a \cdot (b \cdot c)$
- Unique identity element: $\exists e : e \cdot a = a, a \cdot e = a$
- Inverse element: $\forall a \in G \exists b \in G : a \cdot b = e, b \cdot a = e$

Examples:

- Integer addition: $(\mathbb{Z}, +)$, $e = 0$, $b = -a$
- Permutation groups, e.g. for 2 elements

$$(\{\underbrace{(1)}_{e}, \underbrace{(1\,2)}_{a=b}\}, \cdot)$$

- *Special orthogonal group* SO(3): group of 3x3 orthogonal matrices with $\det R = 1$, which is isomorphic to rotations in 3D
- Special Euclidean group SE(3): SO(3) + T(3) *(translation)*

Group theory (2)

The binary operation of the group acts on members of the group, for example for combining several permutations. When we perform a permutation in a vector space, this is called a *group action*.

Group action

A **group action** is a map $f(g, x) = gx$ from a group G and a vector space X

$$f : G \times X \to X,$$

satisfying

$$\begin{aligned} ex &= x && \text{(Identity)} \\ g(hx) &= (g \cdot h)x && \text{(Homomorphism)} \end{aligned}$$

Example: Given two permutation $p_1 = p_2 = (1\,2)$, we can apply them to a vector

$$p_2 p_1 \begin{pmatrix} 3 \\ 4 \end{pmatrix} = p_2 \begin{pmatrix} 4 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

Group theory (3)

Group representation

The *general linear group* GL(V) is the group of all automorphism of V. It includes all invertible linear transformations $V \to V$. A group homomorphism from a given group to the general linear group of a given vector space is called a **group representation**.

Example 1: For the previous example, $p_1 = p_2 = (1\,2)$, the corresponding representation in a vector space $\mathbb{R}^2$ is a permutation matrix

$$P = \rho(p_1) = \rho(p_2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

The corresponding group operation is matrix multiplication and the group action on V is the matrix-vector product:

$$\rho(p_2 p_1) \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{p_2 p_1 = e} \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

Group theory (3)

Group representation

The *general linear group* GL(V) is the group of all automorphism of V. It includes all invertible linear transformations $V \to V$. A group homomorphism from a given group to the general linear group of a given vector space is called a **group representation**.

Example 2: Another representation of the permutation group can be obtained from the tensor product $P \otimes P \in \mathbb{R}^{4\times 4}$ acting on a vector space $\mathbb{R}^4$.

$$\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right] \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \\ 2 \\ 1 \end{pmatrix}.$$

Reshaping the vectors to matrices in $\mathbb{R}^{2\times 2}$, this can be rewritten as

$$P \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} P^T = \begin{pmatrix} 4 & 3 \\ 2 & 1 \end{pmatrix}.$$

Following this scheme, one can find representations of a group for arbitrary high tensor orders.

Group Equivariance & Convolution

We generalize our previous notion of equivariance to arbitrary group actions, i.e. a function f is equivariant to group actions

$$f(\rho_1(g)x) = \rho_2(g)f(x) \quad \forall g \in G,$$

with ρ_1, ρ_2 being representations of the group in the respective vector spaces.

The classic convolution layer is an equivariant map to the translation group

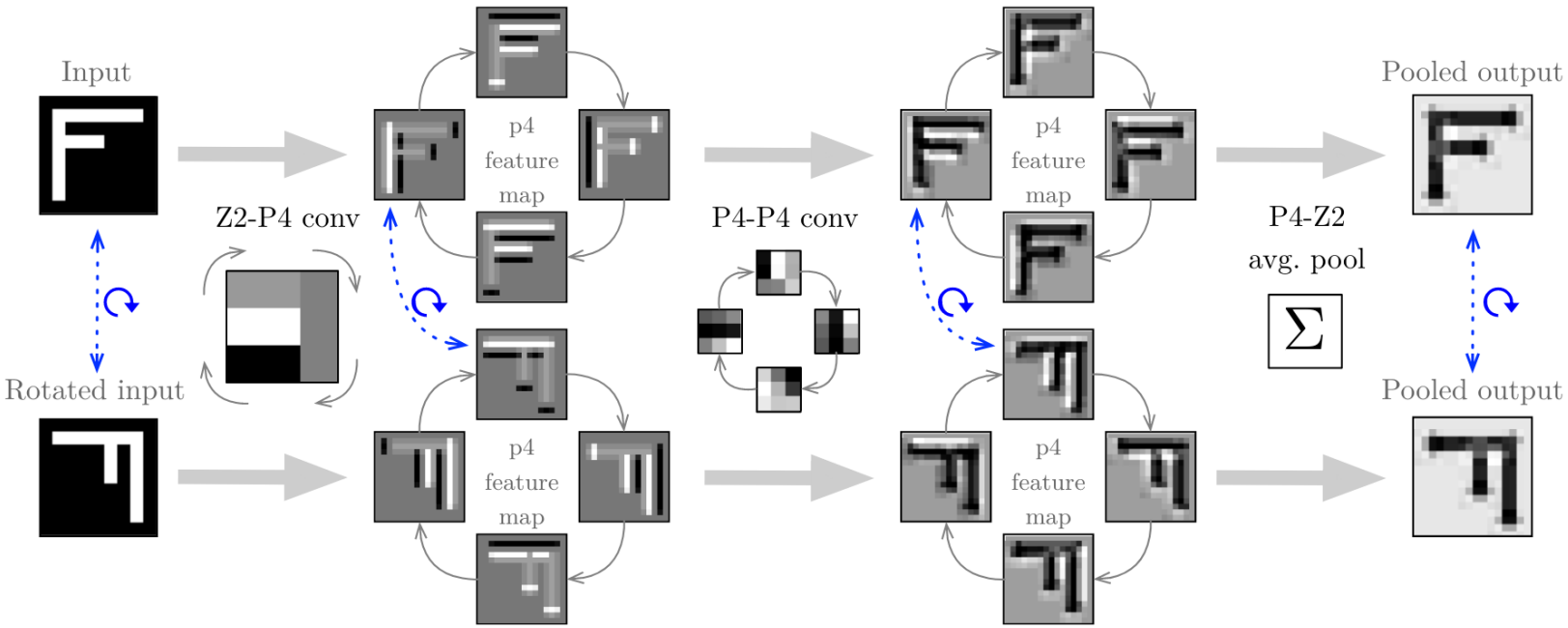
$$(f * W)(x) = \sum_{y \in V} \sum_{k} f_k(y) W_k(y - x),$$

where f is the data (e.g. image), k are the features and x, y positions (e.g. pixel indices).

The **group convolution** is obtained by switching the translation with the desired group operation

$$(f * W)(x) = \sum_{g \in G} \sum_{k} f_k(g) W_k(g^{-1}x)$$

G-CNN for p4 group (90° rotations)



[Veeling et al., 2018]

The image f, written as a function of pixel coordinates $x \in \mathbb{Z}^2$, is transformed to a feature function $h \in p4$ depending on the rotation. While the formula of the convolution looks the same

$$(f * W)(h) = \sum_{g \in G} \sum_{k} f_k(g) W_k(g^{-1}h),$$

the group action g^{-1} here maps to p4 instead of $\mathbb{Z}^2$, changing the operation (see image).

SE(3) steerable CNNs

Group can be decomposed in rotation SO(3) and translation E(3), so standard CNN with *steerable* filters can be used:

$$(f * W)(x) = \sum_y \sum_k f_k(y) W_k(y - x)$$

$$\text{with } W_k(rx) = \rho_2(r) w(x) \rho_1(r)^{-1}$$

The kernel is constrained by the representations of input (ρ_1) and output (ρ_2) space of the SO(3) group for the rotational part r, but free in the radial part x.

To capture the rotational part r of the feature rx accurately, we require tensor-valued neurons, i.e. vectors, matrices and higher-order tensors depending on the required angular resolution.

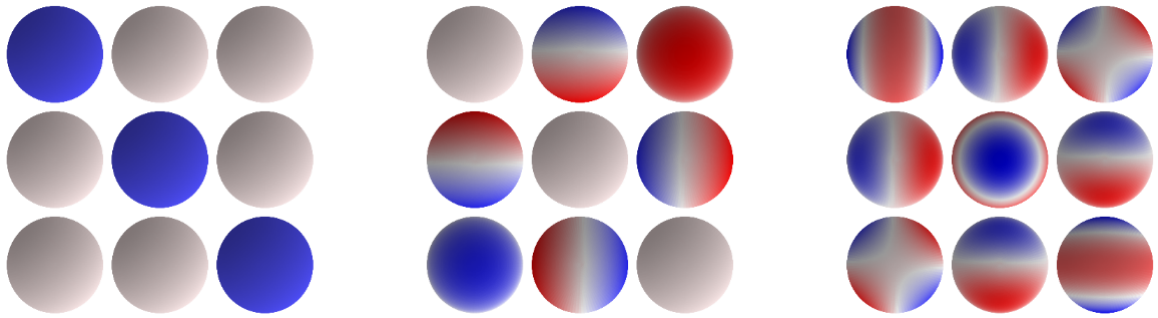
Since we cannot list all rotations (as we did with p4), how can we parametrize the subgroup SO(3) efficiently?

Irreducible SO(3) features

A vector rotation can be represented by a rotation matrix R. As shown, transforms of higher-order tensors can be constructed, e.g. for matrices:

$$R(r)AR(r)^T \rightarrow (R(r) \otimes R(r))(vec)(A) = \rho(r)(vec)(A)$$

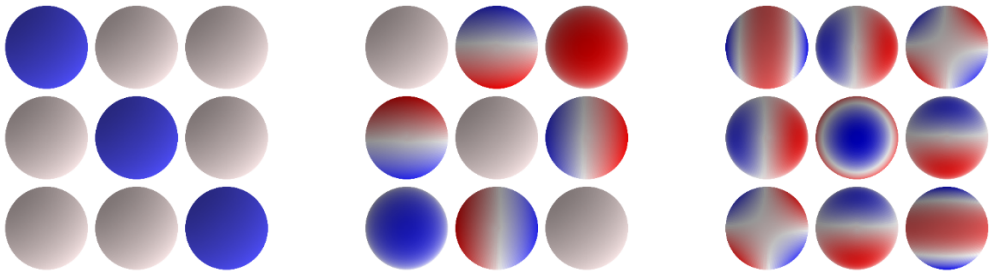
The rotation does not have the full degrees of freedom. It can be decomposed into orthogonal parts that transform independently under the group operation:



trace (1 dimension) | anti-symmetric (3 dimensions) | traceless symmetric (5 dimensions)

This 9-dimensional representation cannot be further decomposed and is called *irreducible*. The higher the order of the tensor, the higher is the representable angular resolution.

SE(3)-equivariant kernel basis



1+3+5=9 spherical harmonics for $l \leq 2$

The SE(3)-filters can be parameterized using an arbitrary radial basis $\phi(x)$ and an angular basis of real spherical harmonics $Y_m^l(r)$ with angular momentum l and $m \in [-l, \ldots, l]$ being the feature indices.

The SE(3)-equivariant convolution can then be written as

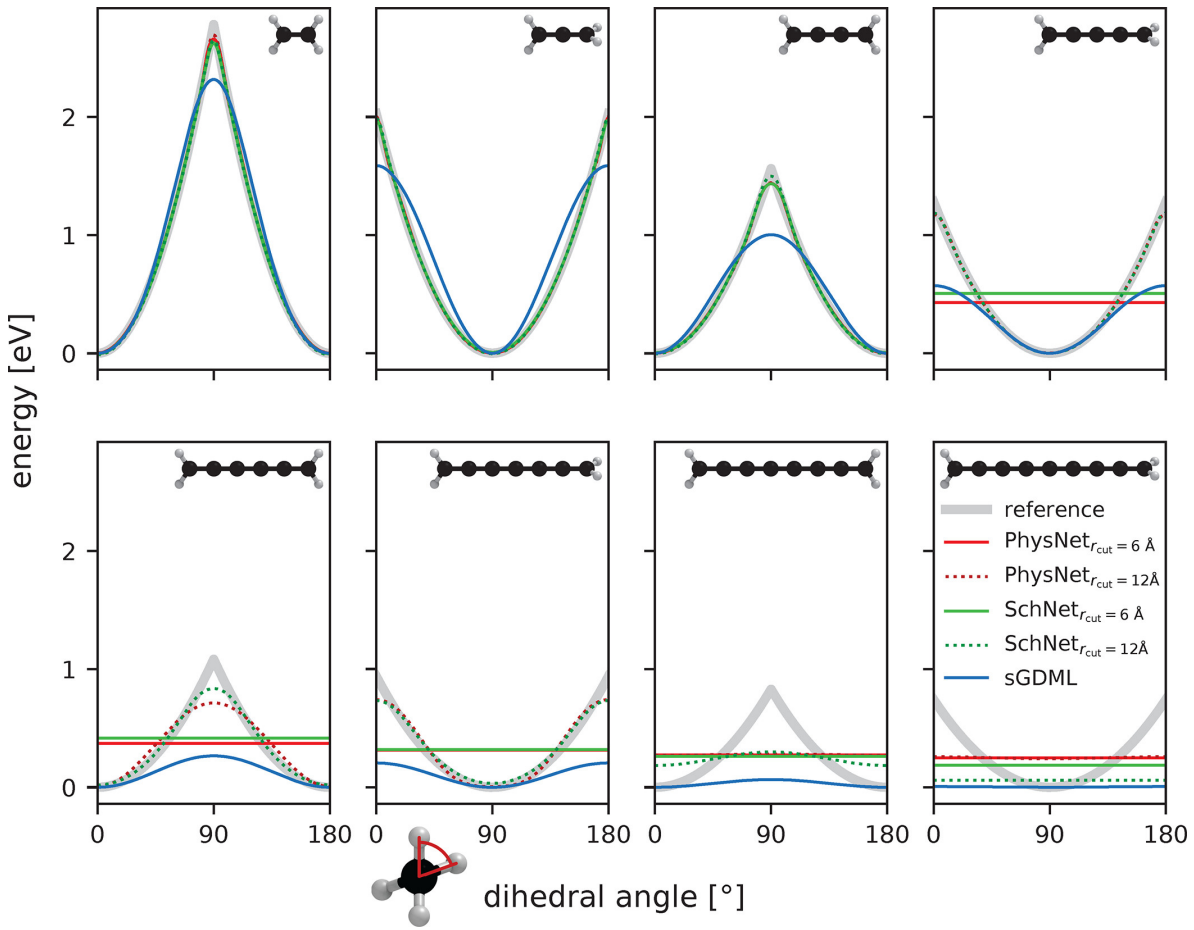
$$(f * W)(x, L, M) = \sum_{y, l_1, m_1, l_2, m_2} \sum_k f_k(y, l_1, m_1) W_k(y - x, l_1, m_1, l_2, m_2)$$

$$W_k(x, l_1, m_1, l_2, m_2) = C^{(L,M)}_{(l_1,m_1),(l_2,m_2)} w_{k,l_1}(\|x\|) Y^{l_2}_{m_2}(x/\|x\|),$$

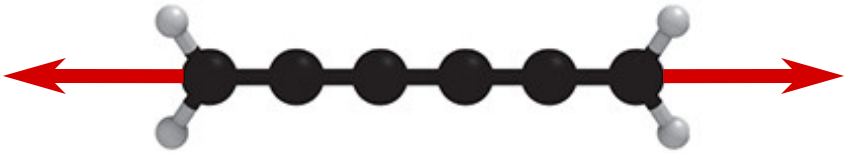
where the Clebsch-Gordon coefficients $C^{LM}_{l_1 m_1 l_2 m_2}$ determine how the irreps of input features and filters couple with each other to produce a given output feature.

Image: Weiler et al. [2018]

Applying what we learned...



Can we combine group equivariance and message passing?



Equivariant message-passing revisited

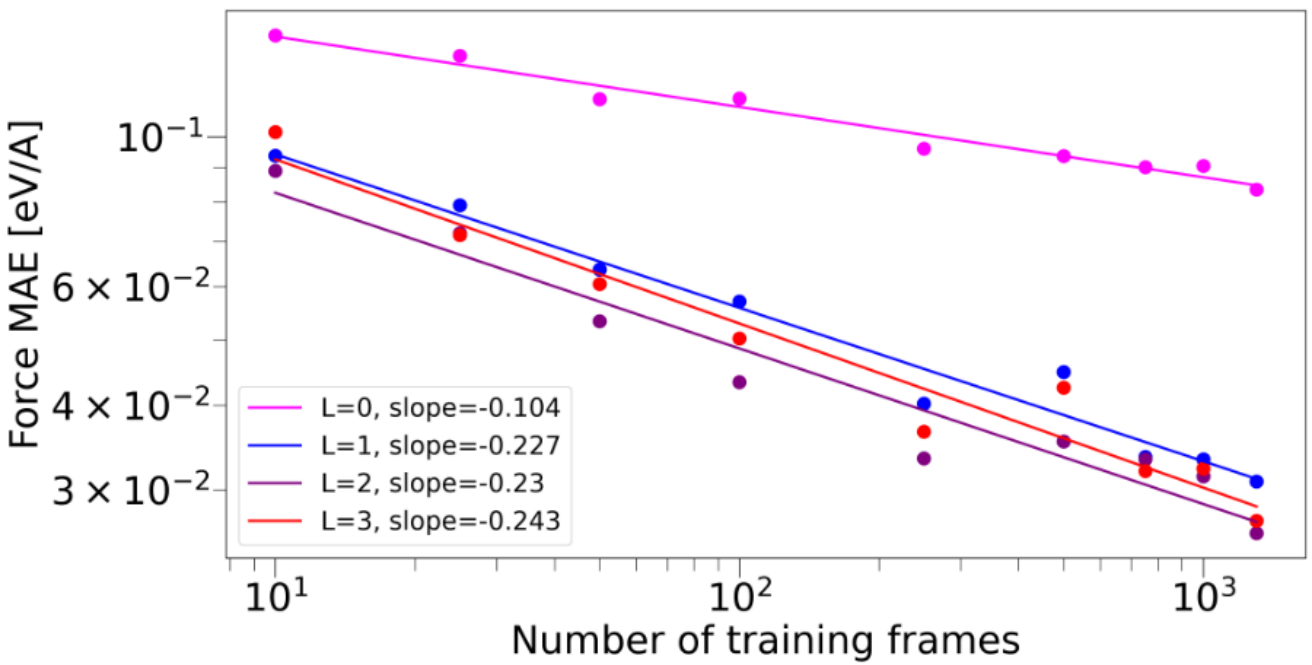
Vectorial convolution in PaiNN:

$$\Delta\vec{\mathbf{v}}_i^m = \sum_j \vec{\mathbf{v}}_j \circ \phi_s(\mathbf{s}_j) \circ \mathcal{W}_{vv}(\|\vec{r}_{ij}\|) + \phi_{vs}(\mathbf{s}_j) \circ \mathcal{W}'_{vs}(\|\vec{r}_{ij}\|) \frac{\vec{r}_{ij}}{\|\vec{r}_{ij}\|},$$

Irreducible representations and Clebsch-Gordan products:

$$(\mathbf{x} * W)_i^{(L,M)} = \sum_j \sum_{l_1,m_1} \sum_{l_2,m_2} \mathbf{x}_{j,}^{(l_1,m_1)} \circ W^{l_2}(\|\vec{r}_{ij}\|) Y_{m_2}^{l_2}(\vec{r}_{ij})\, C_{(l_1,m_1),(l_2,m_2)}^{(L,M)}$$

NequIP force error for water box depending on maximum angular momentum L [Batzner et al., 2022]



Summary

- Equivariance encodes symmetries of the data into neural network architectures
- Group theory provides a general formalism to encode these symmetries
- Local features can be more effective correlated with G-convolutions (Picasso problem)
- Neural networks become more data-efficient and may require less parameters

Bibliography

S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1):1–11, 2022.

S. Dieleman, K. W. Willett, and J. Dambre. Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Monthly notices of the royal astronomical society*, 450(2):1441–1459, 2015.

C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

K. Schütt, P.-J. Kindermans, H. E. Sauceda Felix, S. Chmiela, A. Tkatchenko, and K.-R. Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems*, 30, 2017.

K. Schütt, O. Unke, and M. Gastegger. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pages 9377–9388. PMLR, 2021.

O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schutt, A. Tkatchenko, and K.-R. Muller. Machine learning force fields. *Chemical Reviews*, 121(16):10142–10186, 2021.

B. S. Veeling, J. Linmans, J. Winkens, T. Cohen, and M. Welling. Rotation equivariant cnns for digital pathology. In *International Conference on Medical image computing and computer-assisted intervention*, pages 210–218. Springer, 2018.

M. Weiler, M. Geiger, M. Welling, W. Boomsma, and T. S. Cohen. 3d steerable cnns: Learning rotationally equivariant features in volumetric data. *Advances in Neural Information Processing Systems*, 31, 2018.

Excercise ideas

- proof equivariance for rotation / translation - implement steerable CNN - implement equivariant message function for MPNN from previous homework