Equivariant Attention on Homogeneous Spaces

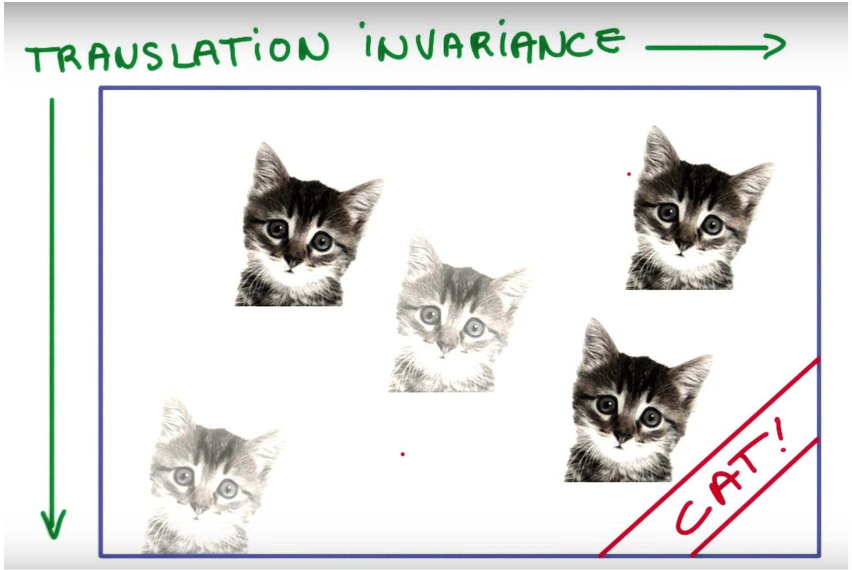
Motivation of this work

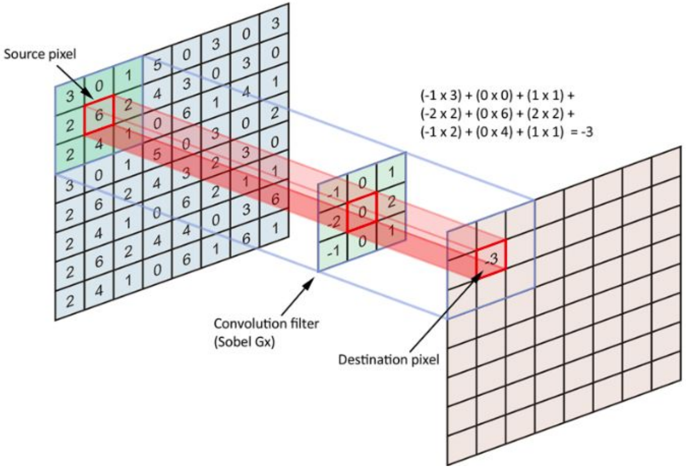
Model Equivariance

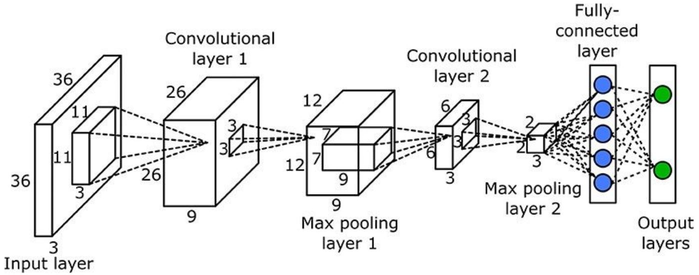
Self Attention Models

Point Cloud Data

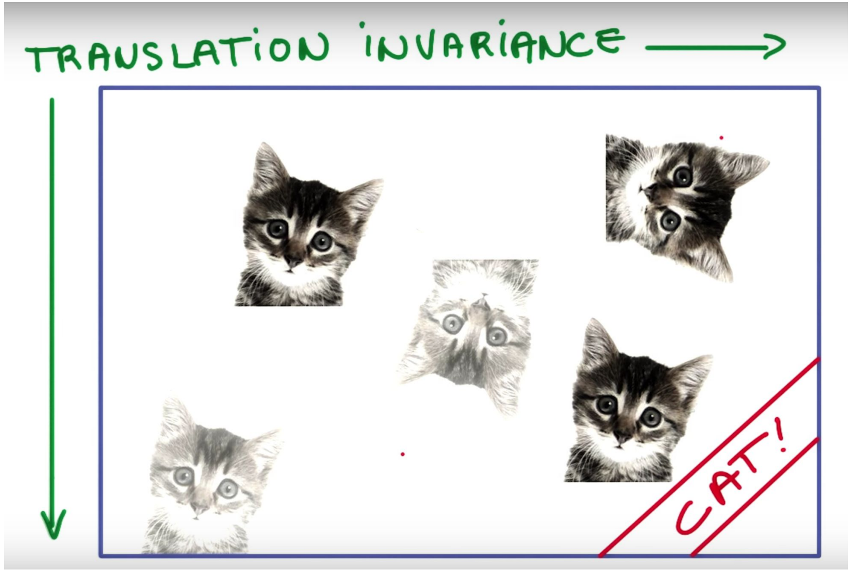
Convolutional Neural Networks

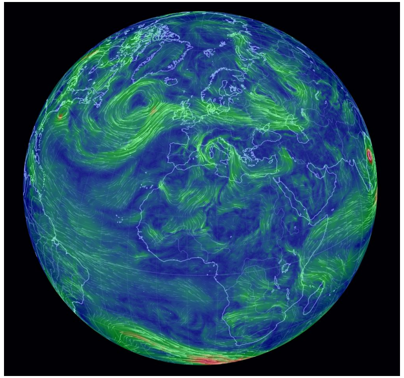


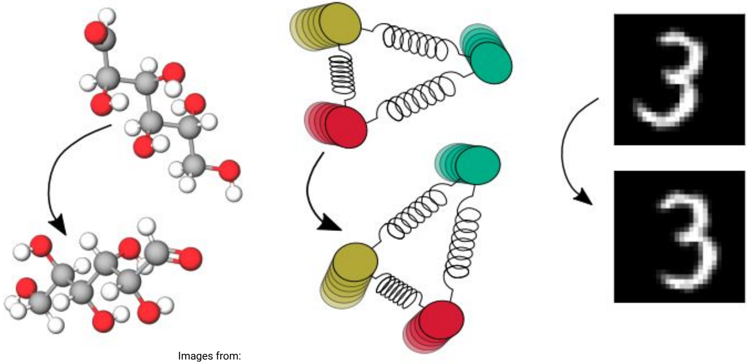




What about other transformations? Or other geometries?







Images from:

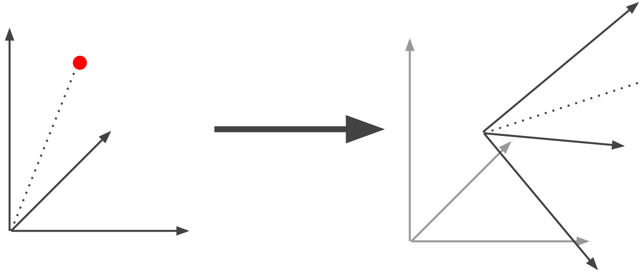
- Generalizing Convolutional Neural Networks for Equivariance to Lie Groups on Arbitrary Continuous Data, *Finzi et al 2020*
- https://earth.nullschool.net/

Some concrete examples of these other groups

$SE(3)$ on $\mathbb{R}^3$

I.e. the 3D rotations and translations of the Euclidean space

action of $g \in SE(3)$

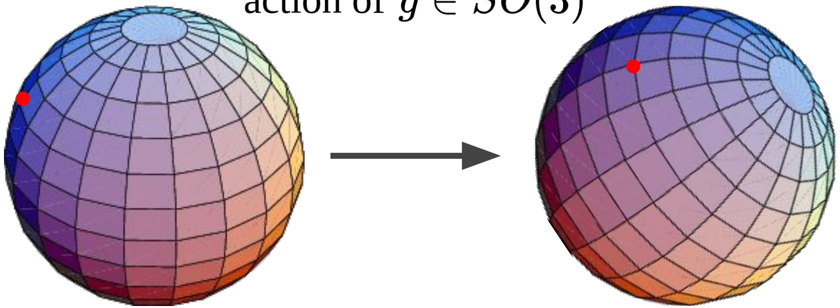


Interesting use case: Data that lives in regular 3D space, such as pose estimation or object classification

$SO(3)$ on S^2

I.e. the 3D rotations of the sphere

action of $g \in SO(3)$



Interesting use case: Data that lives on the surface of the Earth, e.g. climate data, flight path data

What about vector fields?

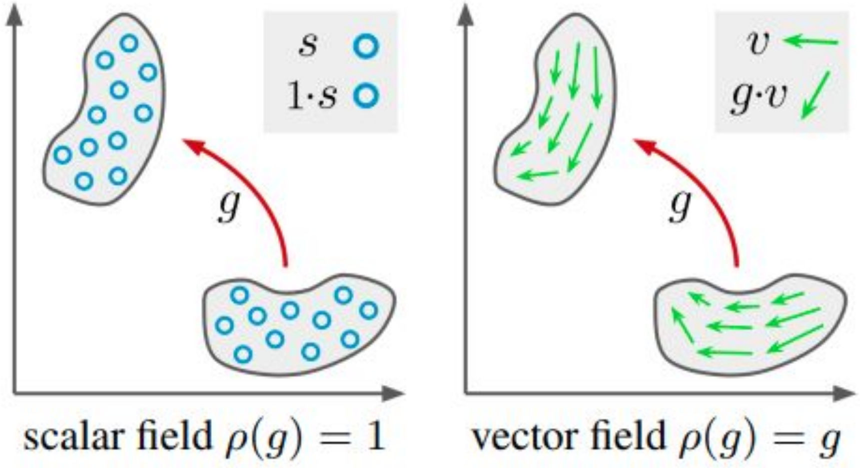


Image from General E(2) - Equivariant Steerable CNNs, *Weiler and Cesa 2019*

Equivariance Summed up

"If a model's inputs transform under a group action, its outputs should transform under a related group action."

G — A group
$\mathcal{X}$ — A space G acts on
$\mathcal{Y}$ — A space G acts on
$\pi_{\mathcal{X}}(g)$ — an action of $g \in G$ on $\mathcal{X}$
$\pi_{\mathcal{Y}}(g)$ — an action of $g \in G$ on $\mathcal{Y}$
$\phi : \mathcal{X} \rightarrow \mathcal{Y}$ — - A map from $\mathcal{X}$ to $\mathcal{Y}$

Equivariance

$$[\pi_{\mathcal{Y}}(g) \circ \phi](x) = [\phi \circ \pi_{\mathcal{X}}(g)](x)$$

Invariance

$$\phi(x) = [\phi \circ \pi_{\mathcal{X}}(g)](x)$$

Recent work on this topic

- A General Theory of Equivariant CNNs on Homogeneous Spaces, *Cohen, Wieler and Kicanaoglu et al 2018*
- Tensor Field Networks, *Thomas and Smidt et al 2018*
- Gauge Equivariant Convolutional Networks and the Icosahedral CNN, *Cohen, Wieler and Kicanaoglu et al 2019*
- General E(2) - Equivariant Steerable CNNs, *Weiler and Cesa 2019*
- Generalizing Convolutional Neural Networks for Equivariance to Lie Groups on Arbitrary Continuous Data, *Finzi et al 2020*

And many more. These however focus mostly on **convolutions**, and as such only **linear** transforms of data (in each layer).

Motivation of this work

Model Equivariance

Motivated by the idea that many datasets display some form of symmetry.

If our data doesn't care about its coordinate system, why should our model?

Self Attention Models

Point Cloud Data

Self-Attention Mechanisms

Each element in a set pays a certain amount of “attention” to each other element in the set

By mixing a location dependant kernel with attention coefficients, you can derive a mixture of attention and convolution

Self-Attention Mechanisms

Each element in a set pays a certain amount of “attention” to each other element in the set

By mixing a location dependant kernel with attention coefficients, you can derive a mixture of attention and convolution

Simplest Form

$$f_i^{l+1} = \sum_j \alpha_{ij} f_j^l$$

$$\bar{\alpha}_{ij} = k(f_i, f_j)$$

$$\sum_j \alpha_{ij} = 1$$

$$\alpha_{ij} = \frac{\exp(\bar{\alpha}_{ij})}{\sum_j \exp(\bar{\alpha}_{ij})}$$

Self-Attention Mechanisms

Each element in a set pays a certain amount of “attention” to each other element in the set

By mixing a location dependant kernel with attention coefficients, you can derive a mixture of attention and convolution

Simplest Form

$$f_i^{l+1} = \sum_j \alpha_{ij} f_j^l \qquad \sum_j \alpha_{ij} = 1$$

$$\bar{\alpha}_{ij} = k(f_i, f_j) \qquad \alpha_{ij} = \frac{\exp(\bar{\alpha}_{ij})}{\sum_j \exp(\bar{\alpha}_{ij})}$$

Common Form (Dot product attention)

$$f_i^{l+1} = \sum_j \alpha_{ij} W_V f_j^l \qquad \sum_j \alpha_{ij} = 1$$

$$\bar{\alpha}_{ij} = (W_K f_i)^T (W_Q f_j) \qquad \alpha_{ij} = \frac{\exp(\bar{\alpha}_{ij})}{\sum_j \exp(\bar{\alpha}_{ij})}$$

Self-Attention Mechanisms

Each element in a set pays a certain amount of “attention” to each other element in the set

By mixing a location dependant kernel with attention coefficients, you can derive a mixture of attention and convolution

Simplest Form

$$f_i^{l+1} = \sum_j \alpha_{ij} f_j^l \qquad \sum_j \alpha_{ij} = 1$$

$$\bar{\alpha}_{ij} = k(f_i, f_j) \qquad \alpha_{ij} = \frac{\exp(\bar{\alpha}_{ij})}{\sum_j \exp(\bar{\alpha}_{ij})}$$

Common Form (Dot product attention)

$$f_i^{l+1} = \sum_j \alpha_{ij} W_V f_j^l \qquad \sum_j \alpha_{ij} = 1$$

$$\bar{\alpha}_{ij} = (W_K f_i)^T (W_Q f_j) \qquad \alpha_{ij} = \frac{\exp(\bar{\alpha}_{ij})}{\sum_j \exp(\bar{\alpha}_{ij})}$$

Attentive Convolution

$$f_i^{l+1} = \sum_j \alpha_{ij} W_V (x_i - x_j) f_j^l \qquad \sum_j \alpha_{ij} = 1$$

$$\bar{\alpha}_{ij} = k(f_i, f_j, x_i - x_j) \qquad \alpha_{ij} = \frac{\exp(\bar{\alpha}_{ij})}{\sum_j \exp(\bar{\alpha}_{ij})}$$

Motivation of this work

Model Equivariance

Motivated by the idea that some things look the same, even when transformed in some way.

If our data doesn't care about its coordinate system, why should our model?

Self Attention Models

Initially developed in NLP and language modelling

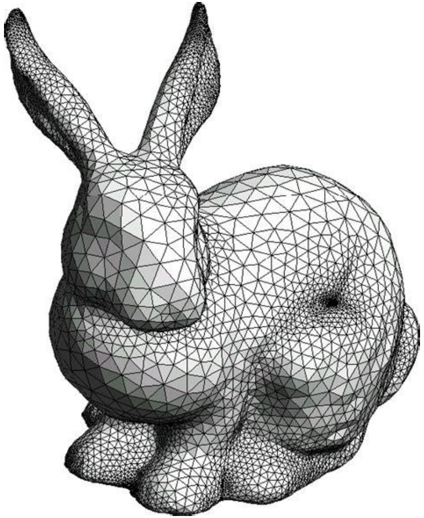
Recent successes in other types of task

Allows for *non-linear* transformations of feature maps

This greater flexibility can be more parameter efficient

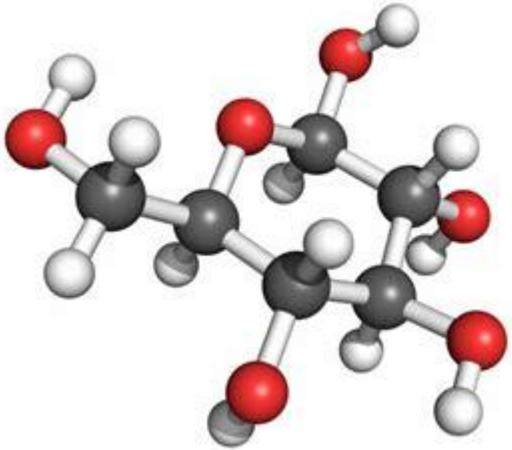
Point Cloud Data

Typical point cloud data tasks



Point cloud object classification
(e.g. 3D scans of objects)

Sensor grid inference
(e.g. weather monitoring)



Molecular property prediction
(e.g. QM9 dataset)

Motivation of this work

Model Equivariance

Motivated by the idea that some things look the same, even when transformed in some way.

If our data doesn't care about its coordinate system, why should our model?

Self Attention Models

Initially developed in NLP and language modelling

Recent successes in other types of task

Allows for *non-linear* transformations of feature maps

This greater flexibility can be more parameter efficient

Point Cloud Data

Typical convolutional architectures rely on grid structured data

Many real world datasets don't have this structure

Developing algorithms that can handle this lack of structure open up many new tasks

Our Approach

Our approach

Our approach centers on projecting input fields over arbitrary representations up to fields over the regular representation.

Our approach

Our approach centers on projecting input fields over arbitrary representations up to fields over the regular representation.

Other models based on this method have appeared before, for the convolution case.

- Spherical CNNs, *Cohen, Geiger and Köhler et al 2018*
- Generalizing Convolutional Neural Networks for Equivariance to Lie Groups on Arbitrary Continuous Data, *Finzi et al 2020*

Our approach

Our approach centers on projecting input fields over arbitrary representations up to fields over the regular representation.

Other models based on this method have appeared before, for the convolution case.

- Spherical CNNs, *Cohen, Geiger and Köhler et al 2018*
- Generalizing Convolutional Neural Networks for Equivariance to Lie Groups on Arbitrary Continuous Data, *Finzi et al 2020*

Our main contributions are:

1. Novel regular representation based self attention and attentive convolution layers, and models built from these

Our approach

Our approach centers on projecting input fields over arbitrary representations up to fields over the regular representation.

Other models based on this method have appeared before, for the convolution case.

- Spherical CNNs, *Cohen, Geiger and Köhler et al 2018*
- Generalizing Convolutional Neural Networks for Equivariance to Lie Groups on Arbitrary Continuous Data, *Finzi et al 2020*

Our main contributions are:

1. Novel regular representation based self attention and attentive convolution layers, and models built from these
2. Application of these models to a number of tasks, including molecular property prediction, system dynamics prediction, and point-cloud classification.

Our approach

Our approach centers on projecting input fields over arbitrary representations up to fields over the regular representation.

Other models based on this method have appeared before, for the convolution case.

- Spherical CNNs, *Cohen, Geiger and Köhler et al 2018*
- Generalizing Convolutional Neural Networks for Equivariance to Lie Groups on Arbitrary Continuous Data, *Finzi et al 2020*

Our main contributions are:

1. Novel regular representation based self attention and attentive convolution layers, and models built from these
2. Application of these models to a number of tasks, including molecular property prediction, system dynamics prediction, and point-cloud classification.
3. A new drop in method to allow any regular representation models to handle general vector fields as inputs, and predict vector field outputs*

Other works in this space

Attentive Group Equivariant Convolutional Networks, *Romero et al 2020*

- Considers attention derived from the visual attention literature, we take a different form of attention from general attention literature
- Experimentally only consider discrete groups and image tasks

SE(3)-Transformers: 3D Roto-Translation Equivariant Attention Networks, *Fuchs and Worrall et al 2020*

- This method is based on an irreducibles approach to equivariance
- General form is limited to semi-direct product groups

Thanks for watching!

Slides available at: MJHutchinson.github.io/mlss