A brief introduction to geometric deep learning

Jason McEwen
www.jasonmcewen.org

Kagenova Limited
Mullard Space Science Laboratory (MSSL), University College London (UCL)

May 2022

Deep learning is hard!

Inductive bias

Universal approximations theorems: Neural networks have the *capacity* to approximate almost arbitrarily complex functions.

Inductive bias

Universal approximations theorems: Neural networks have the *capacity* to approximate almost arbitrarily complex functions.

Does **not** mean we can find good approximators!

Inductive bias

Universal approximations theorems: Neural networks have the *capacity* to approximate almost arbitrarily complex functions.

Does **not** mean we can find good approximators!

Deep learning has given rise to huge variety of **powerful architectures**, *e.g.* CNNs, residual connections, UNet, Inception, depthwise separable convolutions, transformers, ...

Inductive bias

Universal approximations theorems: Neural networks have the *capacity* to approximate almost arbitrarily complex functions.

Does **not** mean we can find good approximators!

Deep learning has given rise to huge variety of **powerful architectures**, *e.g.* CNNs, residual connections, UNet, Inception, depthwise separable convolutions, transformers, ...

Architectures encode inductive bias.

Geometric deep learning

Geometry is a powerful inductive bias.

Term ***geometric deep learning*** first coined by Michael Bronstein (Bronstein *et al.* 2017; Bronstein *et al.* 2022)

Geometric priors

1. Symmetry
2. Stability
3. Multi-scale representation

Symmetry

Equivariance

An operator $\mathcal{A}$ is equivariant to a transformation $\mathcal{T}$ if

$$\mathcal{T}\big(\mathcal{A}(f)\big) = \mathcal{A}\big(\mathcal{T}(f)\big)$$

for all possible signals f.

Transforming the signal after application of the operator, is equivalent to transformation of the signal first, followed by application of the operator.

Symmetry

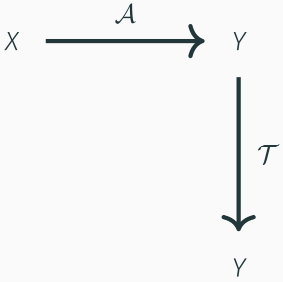
Equivariance

An operator $\mathcal{A}$ is equivariant to a transformation $\mathcal{T}$ if

$$\mathcal{T}\big(\mathcal{A}(f)\big) = \mathcal{A}\big(\mathcal{T}(f)\big)$$

for all possible signals f.

Transforming the signal after application of the operator, is equivalent to transformation of the signal first, followed by application of the operator.



Symmetry

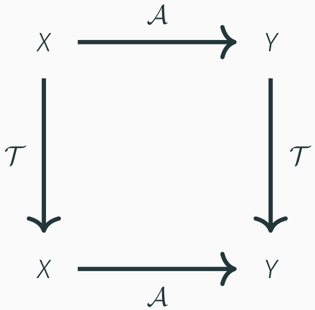
Equivariance

An operator $\mathcal{A}$ is equivariant to a transformation $\mathcal{T}$ if

$$\mathcal{T}(\mathcal{A}(f)) = \mathcal{A}(\mathcal{T}(f))$$

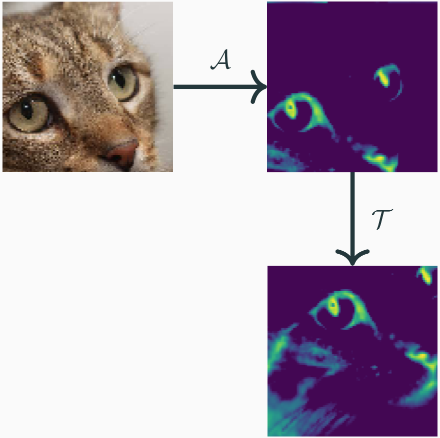
for all possible signals f.

Transforming the signal after application of the operator, is equivalent to transformation of the signal first, followed by application of the operator.



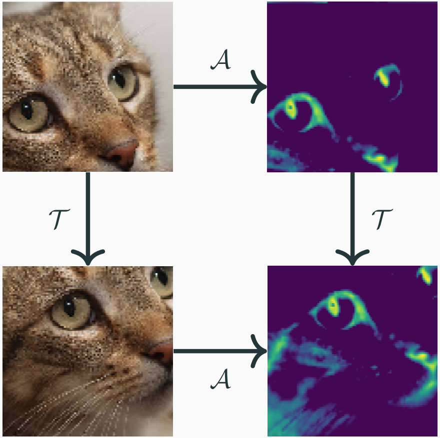
Symmetry

Planar (Euclidean) CNNs exhibit translational equivariance



Symmetry

Planar (Euclidean) CNNs exhibit translational equivariance



Stability

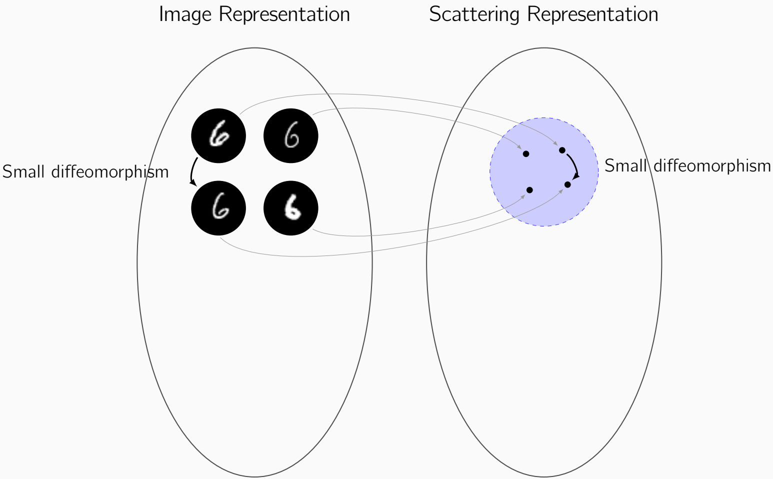


Image Representation
Scattering Representation
Small diffeomorphism
Small diffeomorphism

Stability

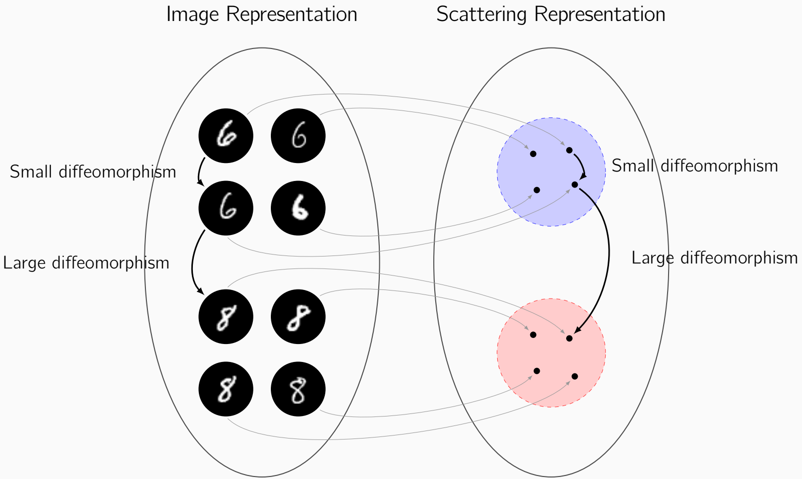
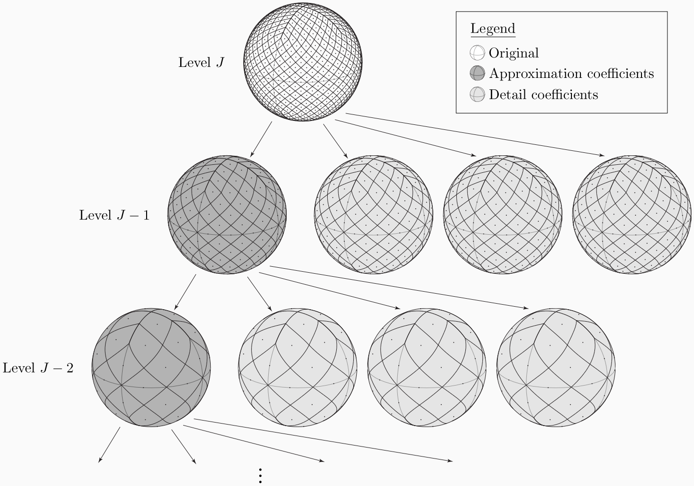
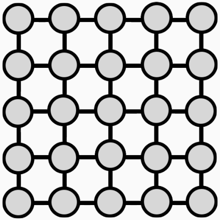
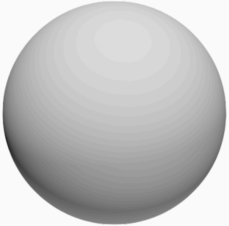
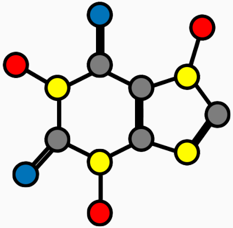


Image Representation
Scattering Representation
Small diffeomorphism
Large diffeomorphism
Small diffeomorphism
Large diffeomorphism

Multi-scale representation



Classes of geometric deep learning

Grids	Groups	Graphs	Geodesics & Gauges
			
Euclidean samples, *e.g. image*	Homogenous spaces with global symmetries, *e.g. sphere*	Nodes and connections, *e.g. social network*	Manifolds, *e.g. 3D mesh*

(Bronstein *et al.* 2022)

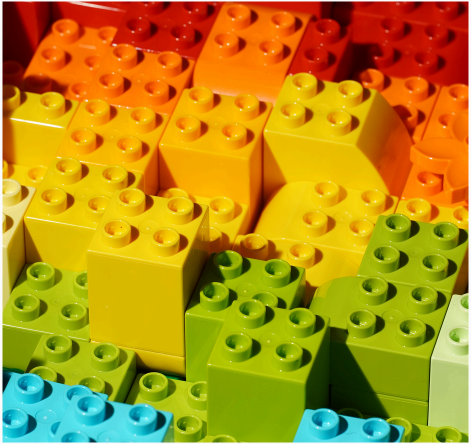
Building blocks

1. Linear equivariant layers,
   e.g. convolutions

2. Non-linear equivariant layers,
   e.g. pointwise activations

3. Local averaging,
   e.g. max pooling

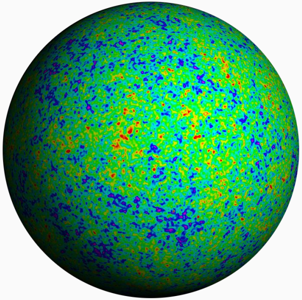
4. Global averaging (invariances),
   e.g. global pooling



Geometric deep learning on the sphere

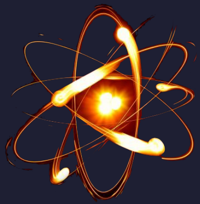
Cosmology and virtual reality

Whenever observe over angles, recover data on 2D sphere (or 3D rotation group).



Cosmic microwave background

360° virtual reality



Since we're concerned with rotational symmetry, leverage the machinery from the study of angular momentum in quantum mechanics.

Generalized spherical CNNs

Consider the s-th layer of a generalized spherical CNN to take the form of a triple (Cobb et al. 2021; arXiv:2010.11661)

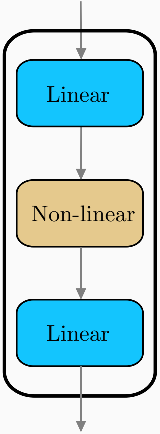
$$\mathcal{A}^{(s)} = (\mathcal{L}_1, \mathcal{N}, \mathcal{L}_2),$$

such that

$$\mathcal{A}^{(s)}(f^{(s-1)}) = \mathcal{L}_2(\mathcal{N}(\mathcal{L}_1(f^{(s-1)}))),$$

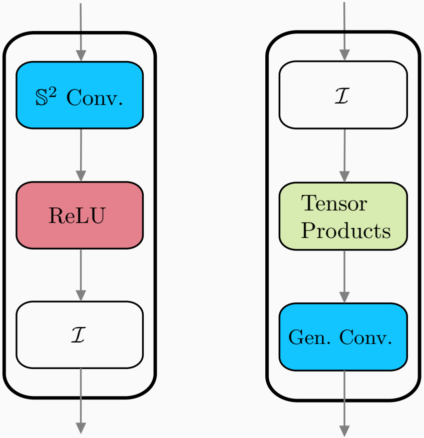
where

- $\mathcal{L}_1, \mathcal{L}_2 : \mathcal{F}_L \rightarrow \mathcal{F}_L$ are spherical convolution operators,
- $\mathcal{N} : \mathcal{F}_L \rightarrow \mathcal{F}_L$ is a non-linear, spherical activation operator.



Generalised spherical CNNs

- Build on other **influential equivariant spherical CNN** constructions:
  - Cohen et al. (2018)
  - Esteves et al. (2018)
  - Kondor et al. (2018)
- Encompass other frameworks as special cases.
- General framework supports hybrids models.



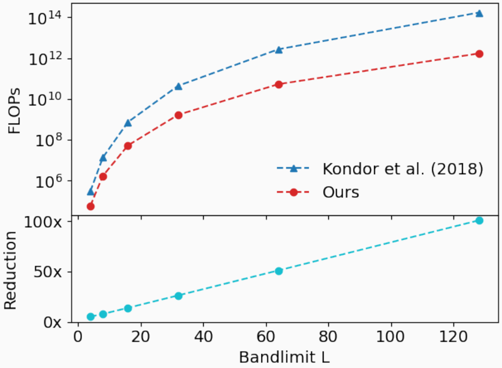
Cohen et al. (2018), Esteves et al. (2018)

Kondor et al. (2018)

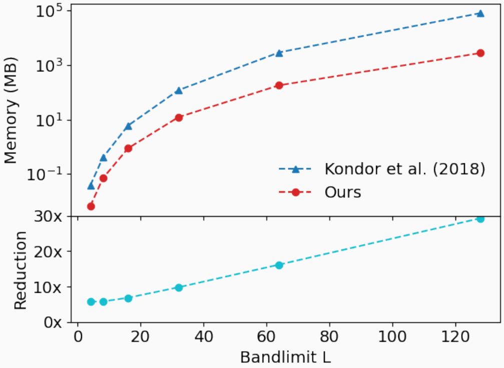
Contributions to improve efficiency

1. Channel-wise structure
2. Constrained generalized convolutions
3. Optimized degree mixing sets
4. Efficient sampling theory on the sphere and rotation group
   (McEwen & Wiaux 2011; McEwen et al. 2015)

Computational cost and memory requirements



Computational cost



Memory requirements

Despite the efficient generalized approach discussed

rotationally equivariant spherical CNNs are not scalable to high-resolution data

Solution: hybrid networks

Efficient generalized spherical CNN framework of Cobb et al. 2021 advocates hybrid networks, with different spherical layers leveraged alongside each other.

(Building on equivariant spherical CNNs of Cohen et al. 2018, Esteves et al. 2018, Kondor et al. 2018.)

Solution: hybrid networks

Efficient generalized spherical CNN framework of Cobb et al. 2021 advocates hybrid networks, with different spherical layers leveraged alongside each other.

(Building on equivariant spherical CNNs of Cohen et al. 2018, Esteves et al. 2018, Kondor et al. 2018.)

Introduce new initial layer, with following properties:

1. Scalable
2. Allow subsequent layers to operate at low-resolution (i.e. mixes information to low frequencies)
3. Rotationally equivariant
4. Stable and locally invariant representation (i.e. effective representation space)

Solution: hybrid networks

Efficient generalized spherical CNN framework of Cobb et al. 2021 advocates hybrid networks, with different spherical layers leveraged alongside each other.

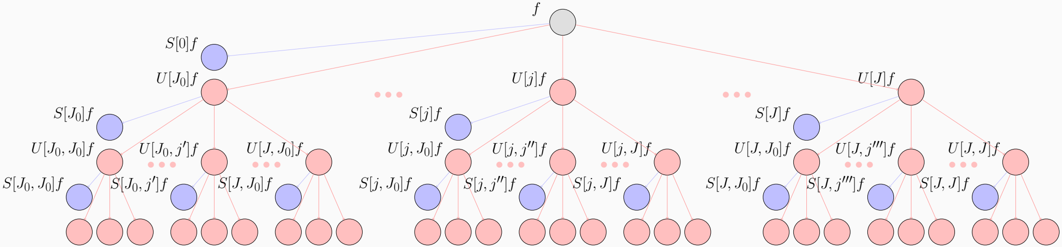
(Building on equivariant spherical CNNs of Cohen et al. 2018, Esteves et al. 2018, Kondor et al. 2018.)

Introduce new initial layer, with following properties:

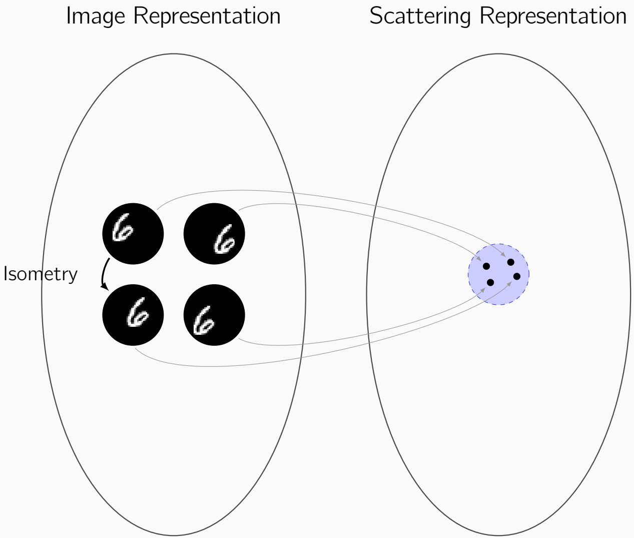
1. Scalable
2. Allow subsequent layers to operate at low-resolution (i.e. mixes information to low frequencies)
3. Rotationally equivariant
4. Stable and locally invariant representation (i.e. effective representation space)

$\Rightarrow$ Scattering networks on the sphere (McEwen et al. 2022; arXiv:2102.02828)

Spherical scattering network is collection of scattering transforms for a number of paths: $\mathcal{S}_{\mathbb{P}} f = \{S[p]f : p \in \mathbb{P}\}$, where the general path set $\mathbb{P}$ denotes the infinite set of all possible paths $\mathbb{P} = \{p = (j_1, j_2, \ldots, j_d) : J_0 \leq j_i \leq J, 1 \leq i \leq d,\ d \in \mathbb{N}_0\}$.



Isometric invariance



Isometric invariance

Theorem (Isometric Invariance)

Let $\zeta \in \mathrm{Isom}(\mathbb{S}^2)$*, then there exists a constant* C *such that for all* $f \in \mathrm{L}^2(\mathbb{S}^2)$*,*

$$\|\mathcal{S}_{\mathbb{P}_D} f - \mathcal{S}_{\mathbb{P}_D} V_\zeta f\|_2 \leq C L^{5/2} (D+1)^{1/2} \, \lambda^{J_0} \, \|\zeta\|_\infty \|f\|_2 .$$

(**Proof:** Follows by straightforward extension of proof of Perlmutter et al. 2020.)

Isometric invariance

Theorem (Isometric Invariance)

Let $\zeta \in \mathrm{Isom}(\mathbb{S}^2)$*, then there exists a constant* C *such that for all* $f \in \mathrm{L}^2(\mathbb{S}^2)$*,*

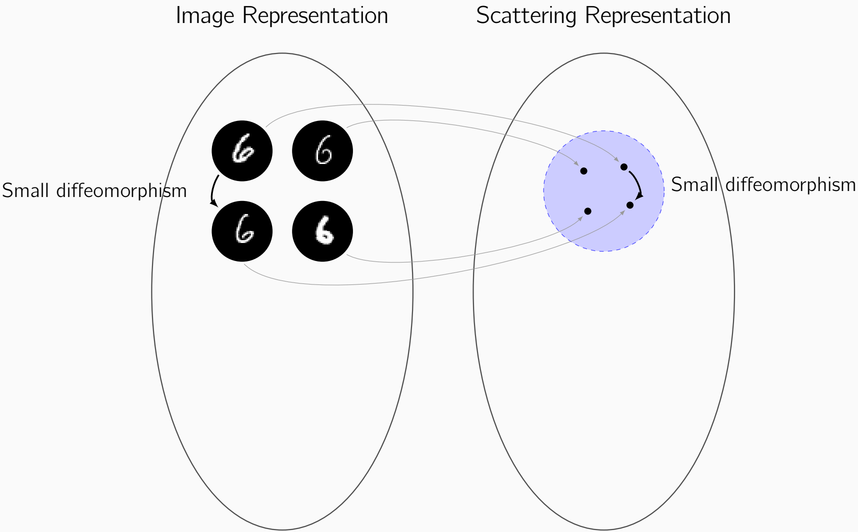
$$\|\mathcal{S}_{\mathbb{P}_D} f - \mathcal{S}_{\mathbb{P}_D} V_\zeta f\|_2 \leq C L^{5/2} (D+1)^{1/2} \lambda^{J_0} \|\zeta\|_\infty \|f\|_2 .$$

Difference in representation

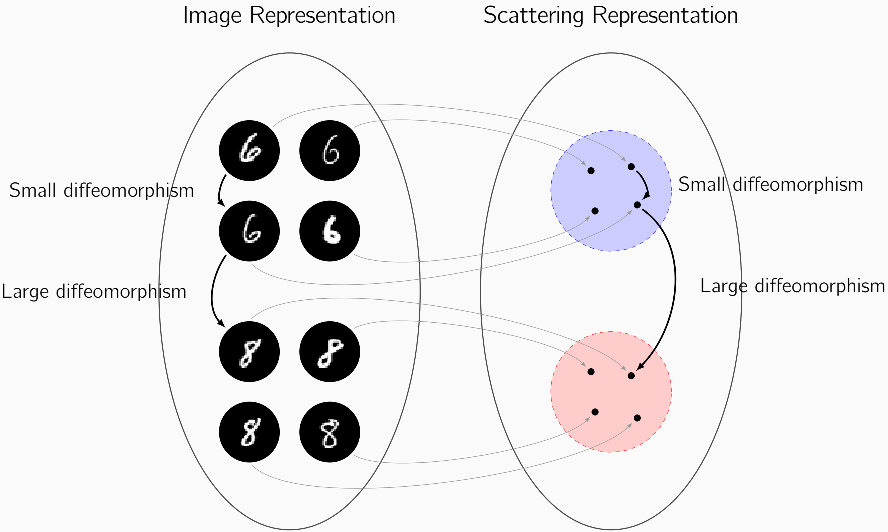
Scattering network representation is invariant to isometries up to a scale.

(**Proof:** Follows by straightforward extension of proof of Perlmutter et al. 2020.)

Stability to diffeomorphisms



Stability to diffeomorphisms



Stability to diffeomorphisms

Theorem (Stability to Diffeomorphisms)

Let $\zeta \in \mathrm{Diff}(\mathbb{S}^2)$. *If* $\zeta = \zeta_1 \circ \zeta_2$ *for some isometry* $\zeta_1 \in \mathrm{Isom}(\mathbb{S}^2)$ *and diffeomorphism* $\zeta_2 \in \mathrm{Diff}(\mathbb{S}^2)$, *then there exists a constant* C *such that for all* $f \in \mathrm{L}^2(\mathbb{S}^2)$,

$$\|\mathcal{S}_{\mathbb{P}_D} f - \mathcal{S}_{\mathbb{P}_D} V_\zeta f\|_2 \le C L^2 \left[L^2 \ \|\zeta_2\|_\infty \ + L^{1/2} (D+1)^{1/2} \lambda^{J_0} \ \|\zeta_1\|_\infty \ \right] \|f\|_2 .$$

(**Proof:** Follows by straightforward extension of proof of Perlmutter et al. 2020.)

Stability to diffeomorphisms

Theorem (Stability to Diffeomorphisms)

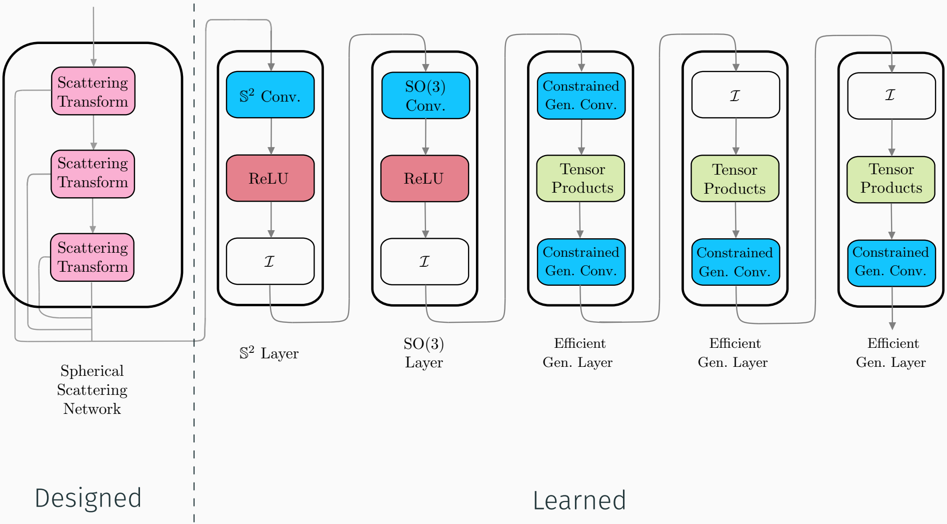
Let $\zeta \in \mathrm{Diff}(\mathbb{S}^2)$. *If* $\zeta = \zeta_1 \circ \zeta_2$ *for some isometry* $\zeta_1 \in \mathrm{Isom}(\mathbb{S}^2)$ *and diffeomorphism* $\zeta_2 \in \mathrm{Diff}(\mathbb{S}^2)$, *then there exists a constant* C *such that for all* $f \in \mathrm{L}^2(\mathbb{S}^2)$,

$$\|\mathcal{S}_{\mathbb{P}_D} f - \mathcal{S}_{\mathbb{P}_D} V_\zeta f\|_2 \leq C L^2 \left[L^2 \|\zeta_2\|_\infty + L^{1/2} (D+1)^{1/2} \lambda^{J_0} \|\zeta_1\|_\infty \right] \|f\|_2 .$$

Difference in representation

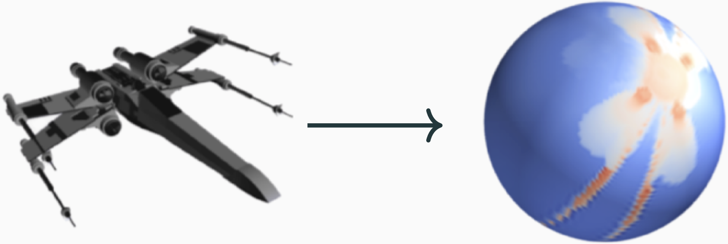
Scattering network representation is stable to small diffeomorphisms about isometry.

(**Proof:** Follows by straightforward extension of proof of Perlmutter et al. 2020.)


Scattering Transform
Scattering Transform
Scattering Transform
Spherical Scattering Network
$\mathbb{S}^2$ Conv.
ReLU
$\mathcal{I}$
$\mathbb{S}^2$ Layer
SO(3) Conv.
ReLU
$\mathcal{I}$
SO(3) Layer
Constrained Gen. Conv.
Tensor Products
Constrained Gen. Conv.
Efficient Gen. Layer
$\mathcal{I}$
Tensor Products
Constrained Gen. Conv.
Efficient Gen. Layer
$\mathcal{I}$
Tensor Products
Constrained Gen. Conv.
Efficient Gen. Layer
Designed
Learned

3D shape classification: problem

Classify 3D meshes and perform shape retrieval.



[Image credit: Esteves et al. 2018]

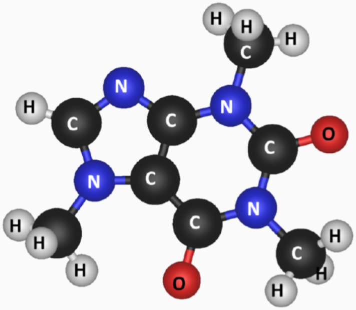
3D shape classification: results

SHREC'17 object retrieval competition metrics (perturbed micro-all)

	P@N	R@N	F1@N	mAP	NDCG	Params
Kondor et al. 2018	0.707	0.722	0.701	0.683	0.756	>1M
Cohen et al. 2018	0.701	0.711	0.699	0.676	0.756	1.4M
Esteves et al. 2018	0.717	**0.737**	-	**0.685**	-	500k
Ours	**0.719**	0.710	**0.708**	0.679	**0.758**	**250k**

Atomization energy prediction: problem

Predict atomization energy of molecule give the atom charges and positions.

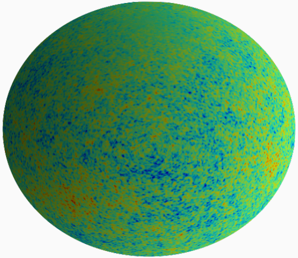


Atomization energy prediction: results

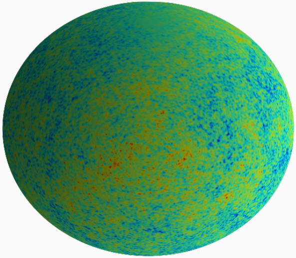
Test root mean squared (RMS) error for QM7 regression problem

	RMS	Params
Montavon et al. 2012	5.96	-
Cohen et al. 2018	8.47	1.4M
Kondor et al. 2018	7.97	>1.1M
Ours (MST)	**3.16**	337k
Ours (RMST)	3.46	**335k**

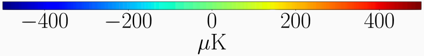
Gaussianity of the cosmic microwave background



Gaussian



Non-Gaussian



At $L = 1024$ (∼2 million pixels), we achieve classification accuracy of:
53% without scattering network versus 95% with scattering network.