

Geometric Deep Learning

- From equivariance to weather predictions -

Daniel Persson

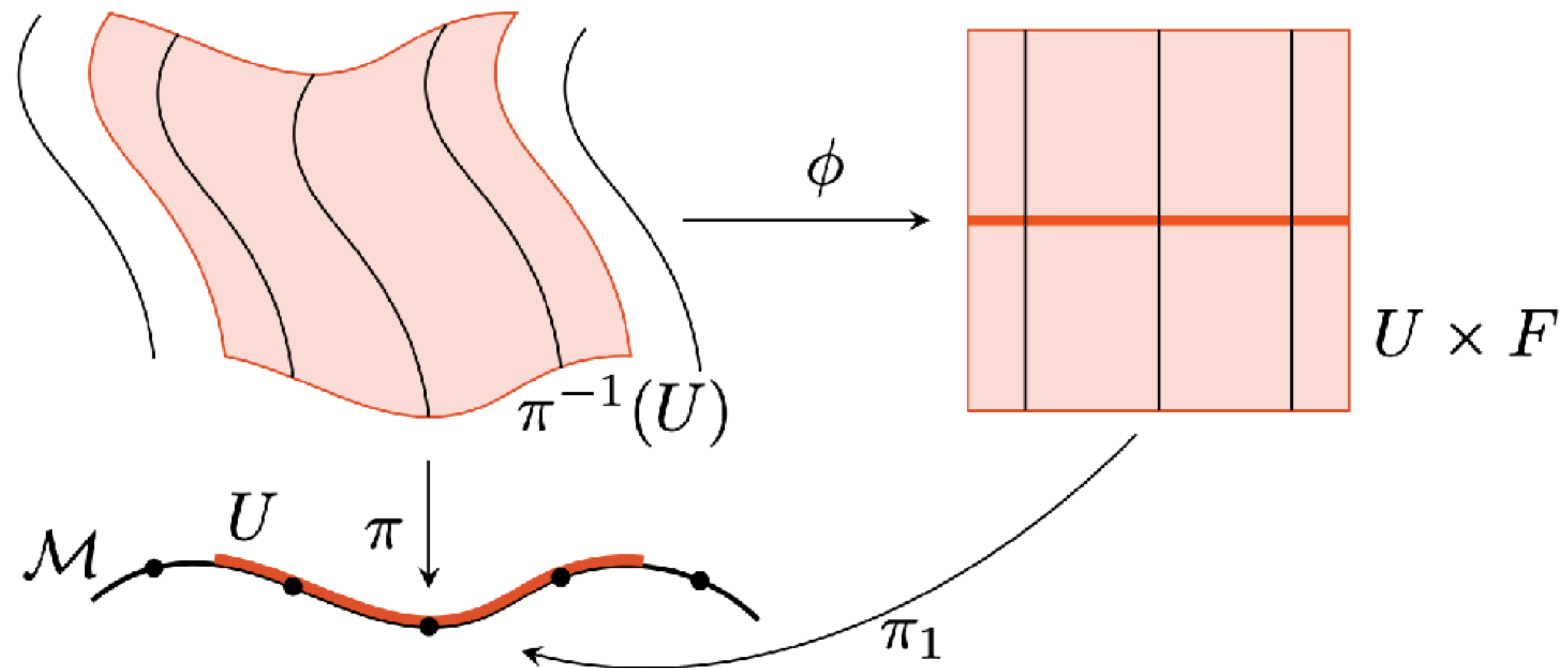
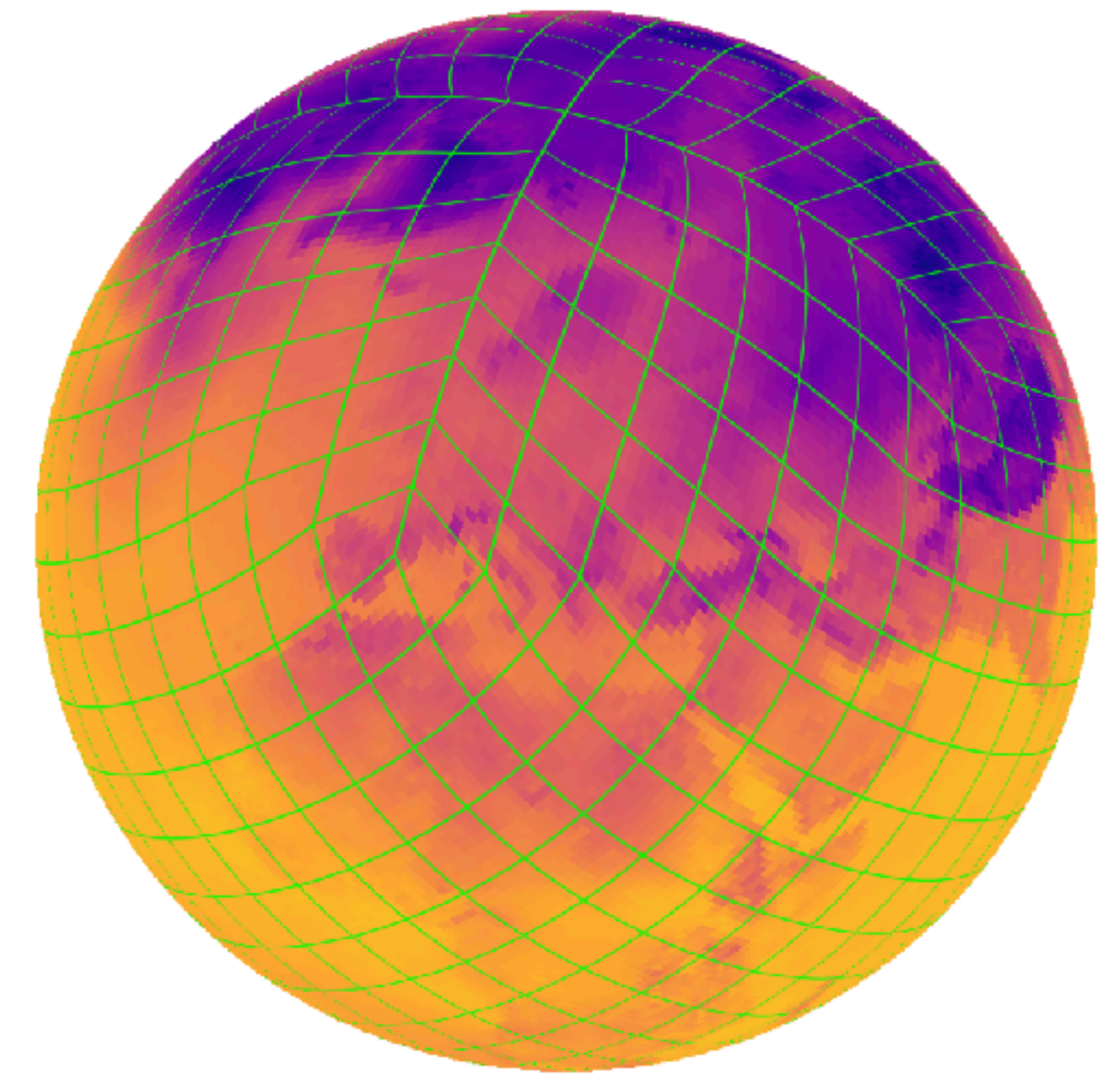
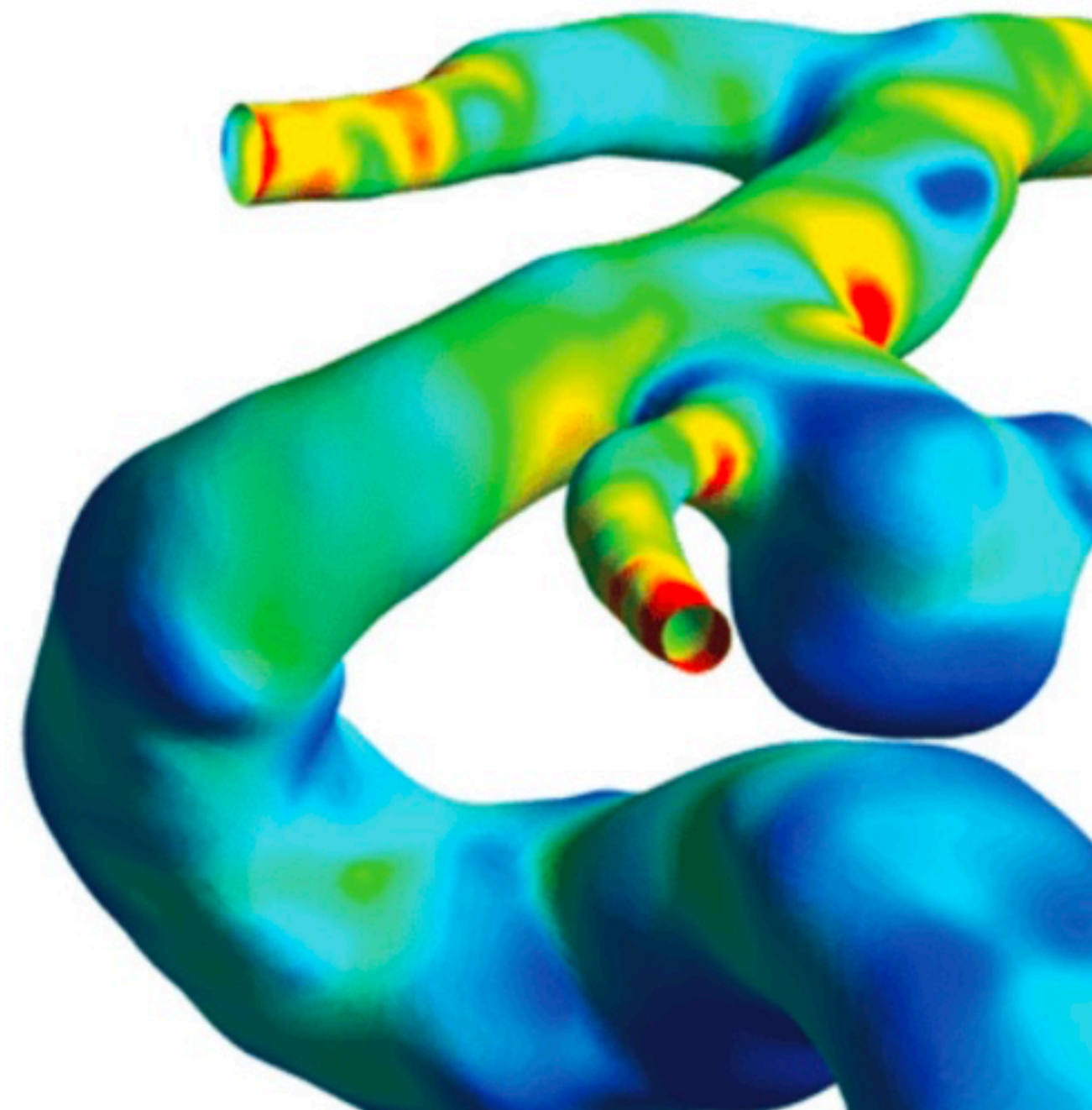
Department of Mathematical Sciences
Chalmers University of Technology
University of Gothenburg

WASP | WALLEMBERG AI
AUTONOMOUS SYSTEMS
AND SOFTWARE PROGRAM

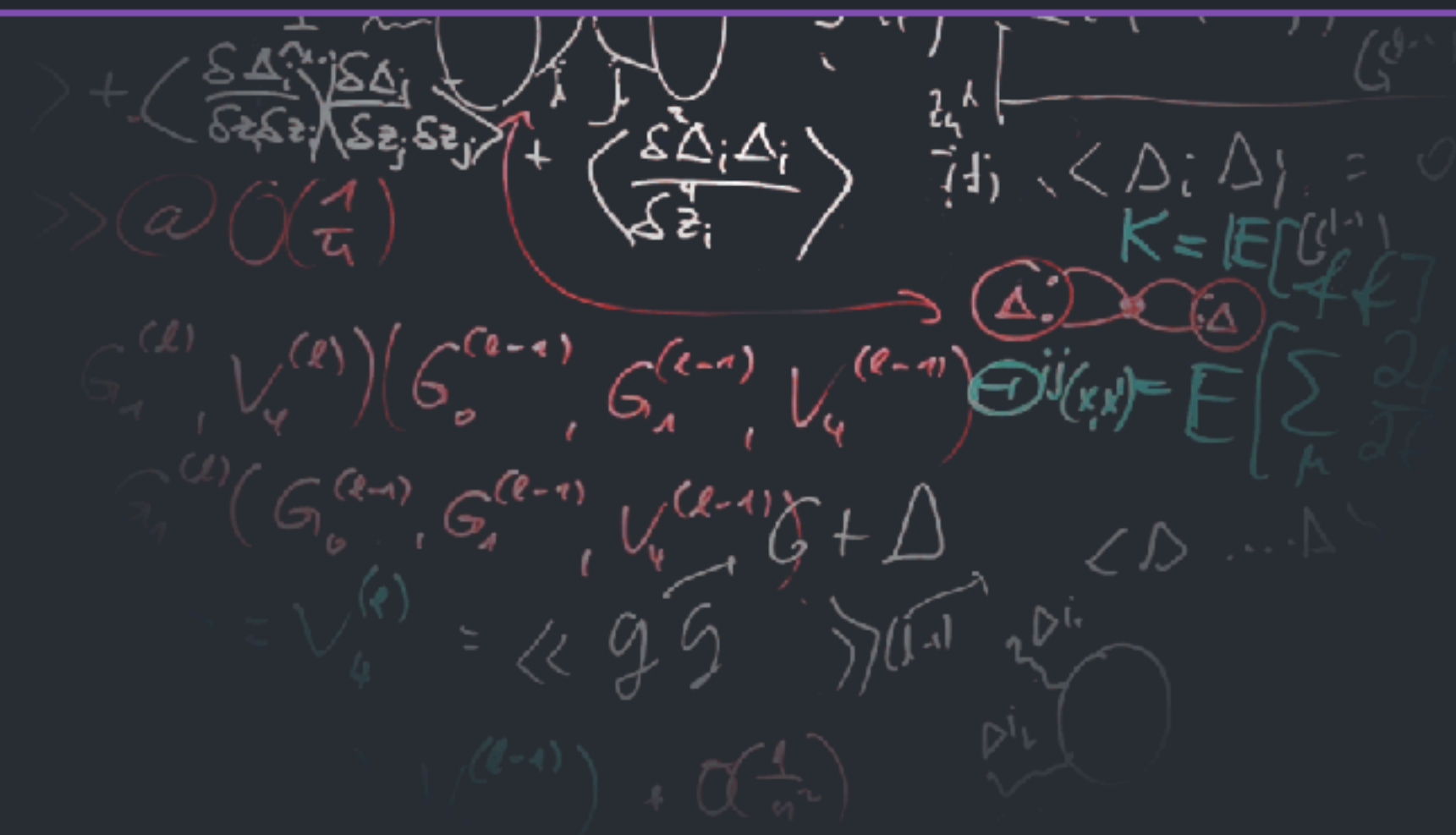
Mathematical Foundations of AI

Chalmers

June 9, 2026



Geometry, Algebra and Physics in Deep Neural Networks



The research group on Geometry, Algebra and Physics in Deep Neural Networks (GAPinDNNs) is based at the Department for Mathematical Sciences at Chalmers University of Technology and the University of Gothenburg. Our vision is to develop a mathematical foundation for deep learning which elevates the field into a theoretically well-grounded science.

News

New Preprint on The Geometry of Polynomial Group Convolutional Neural Networks

31 Mar 2026

[Daniel Persson](#), together with collaborators [Yacoub Hendi](#) and [Magdalena Larfors](#), have published a new preprint on *The Geometry of Polynomial Group Convolutional Neural Networks*.

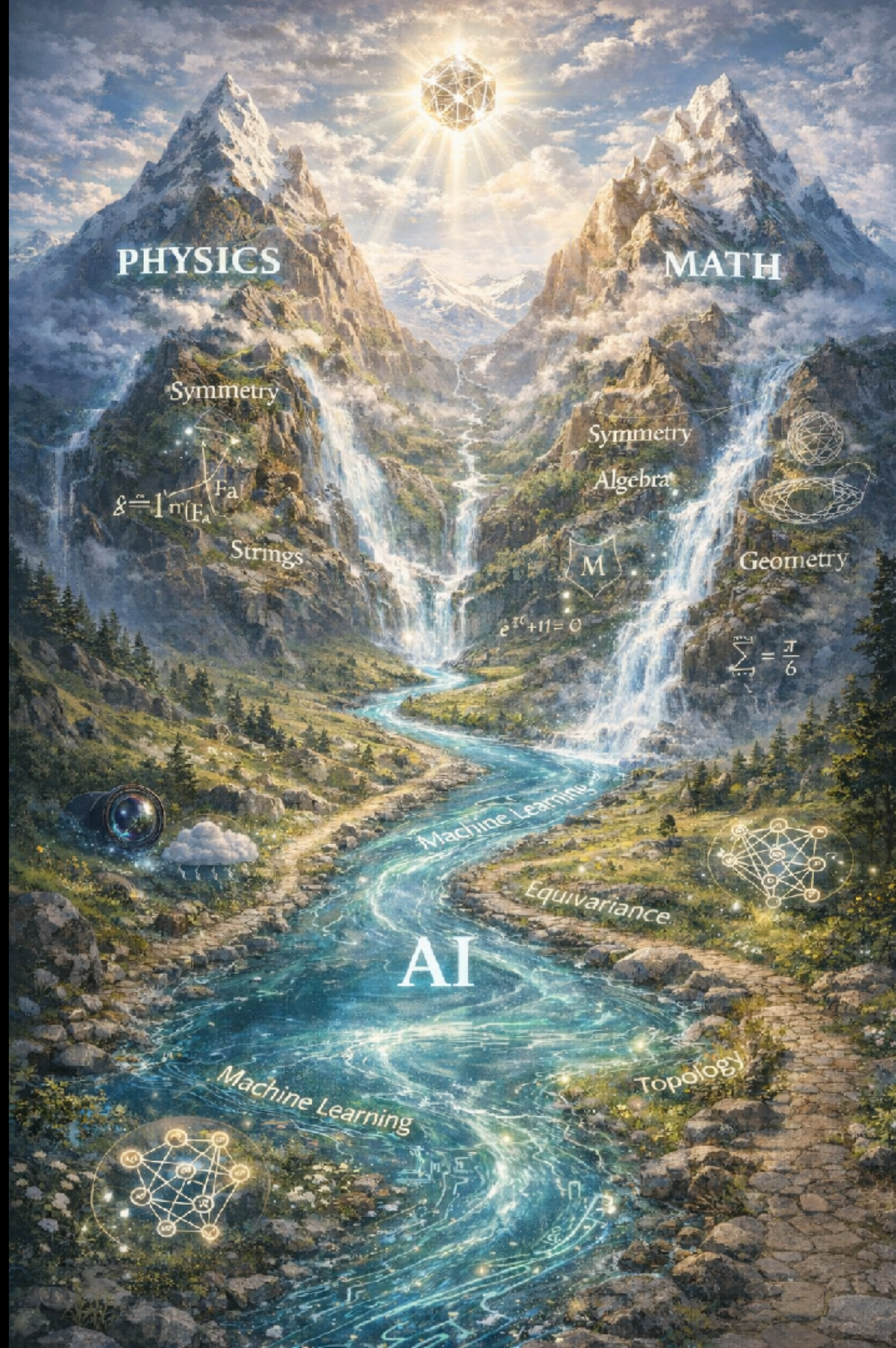
[Read More](#)

Elias in Boston

16 Jan 2026

Our PhD student [Elias](#) will spend the spring in Boston on a WASP-funded research visit, working with [Maurice Weiler](#) at MIT and [Robin Walters](#) at

Math/Physics4AI



Outline

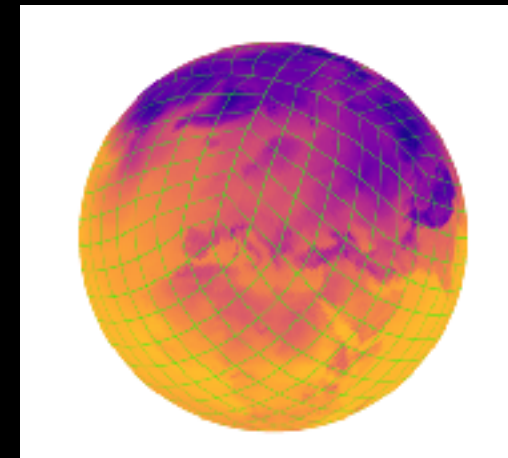
➔ Motivation from physics: Unification



➔ Symmetries in neural networks: Geometric deep learning

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ \downarrow T & & \downarrow T \\ X & \xrightarrow{f} & Y \end{array}$$

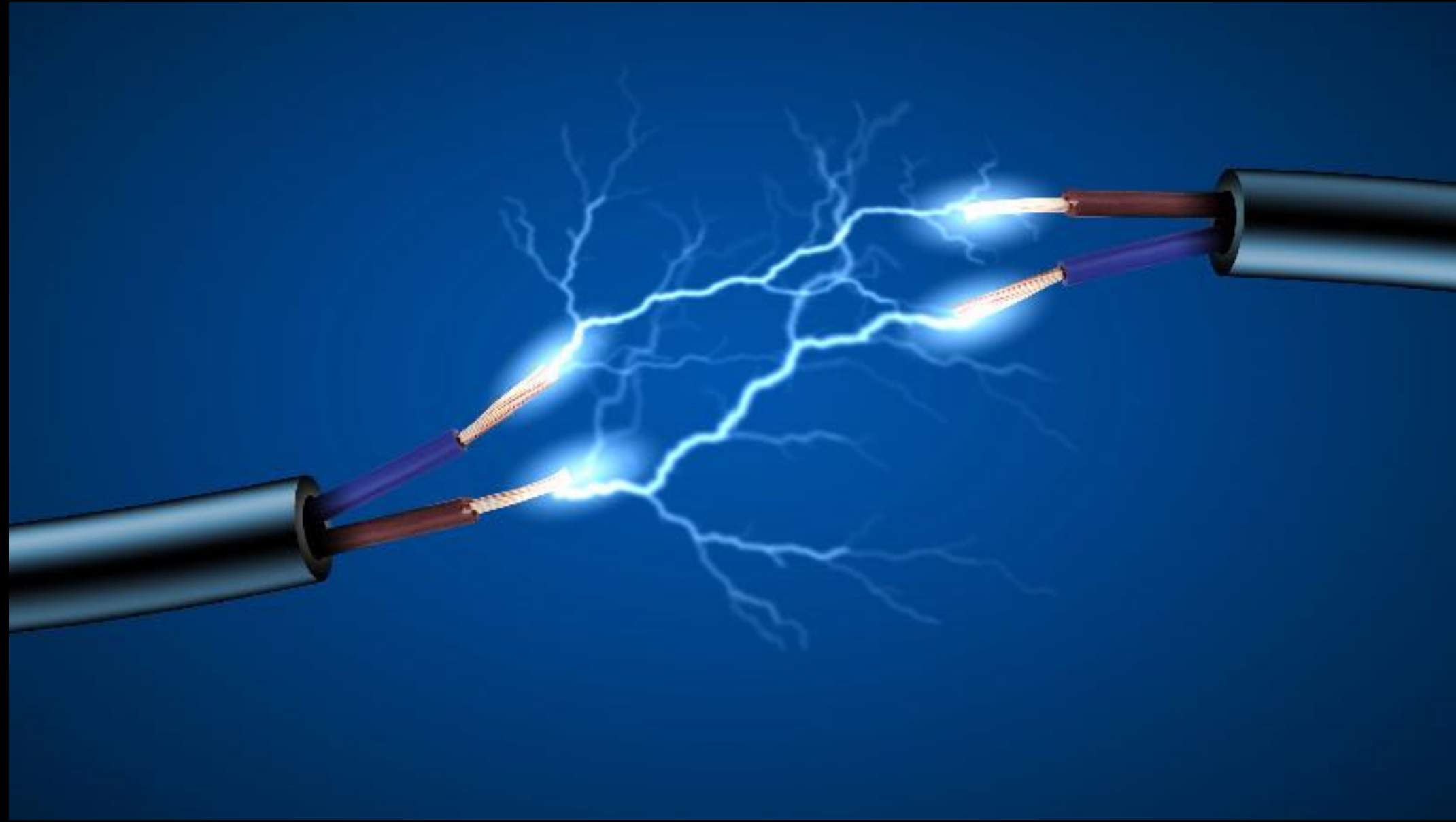
➔ Applications: AI4Science



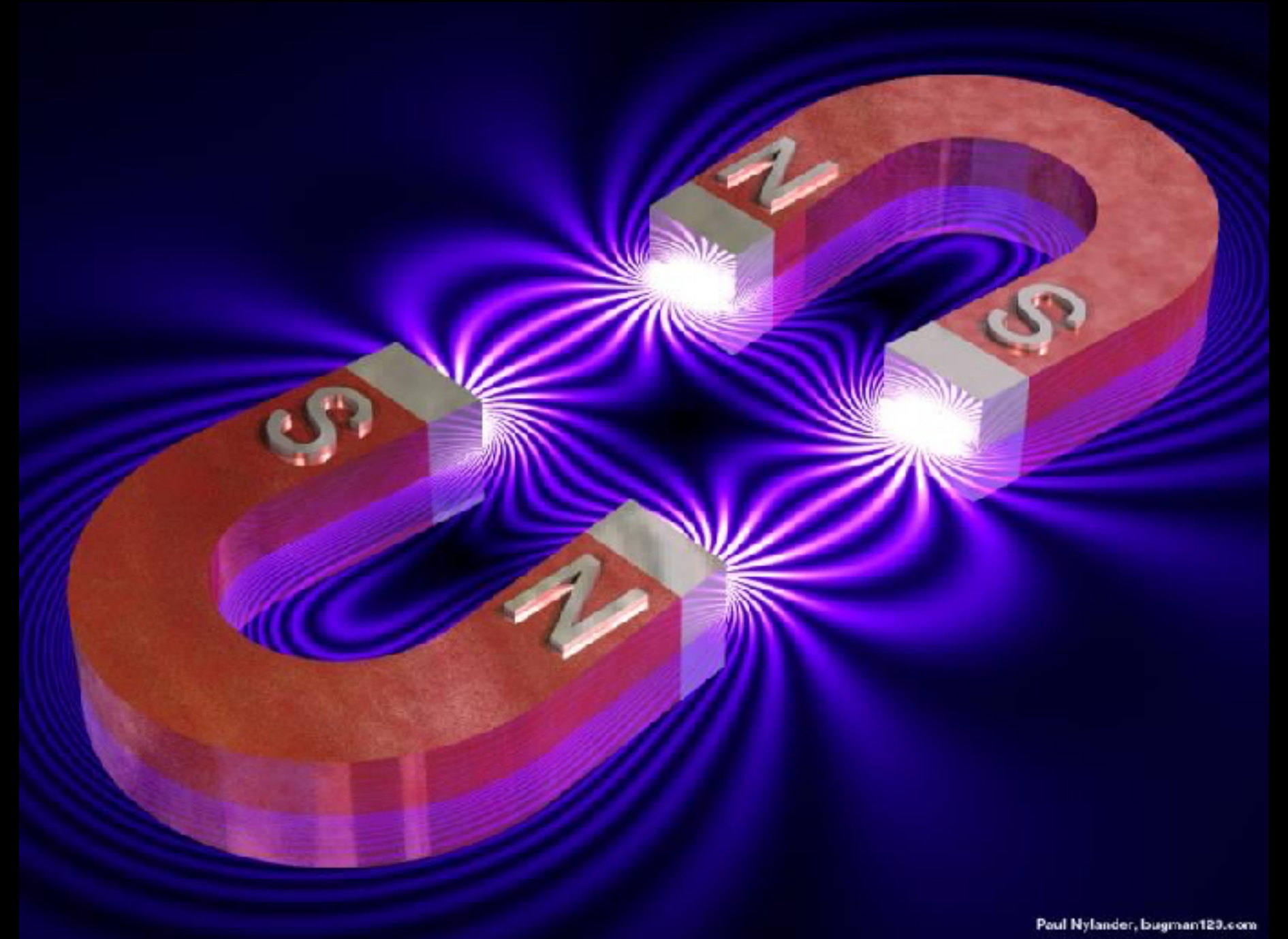
➔ Conclusions

Motivation from physics: Unification

Unification

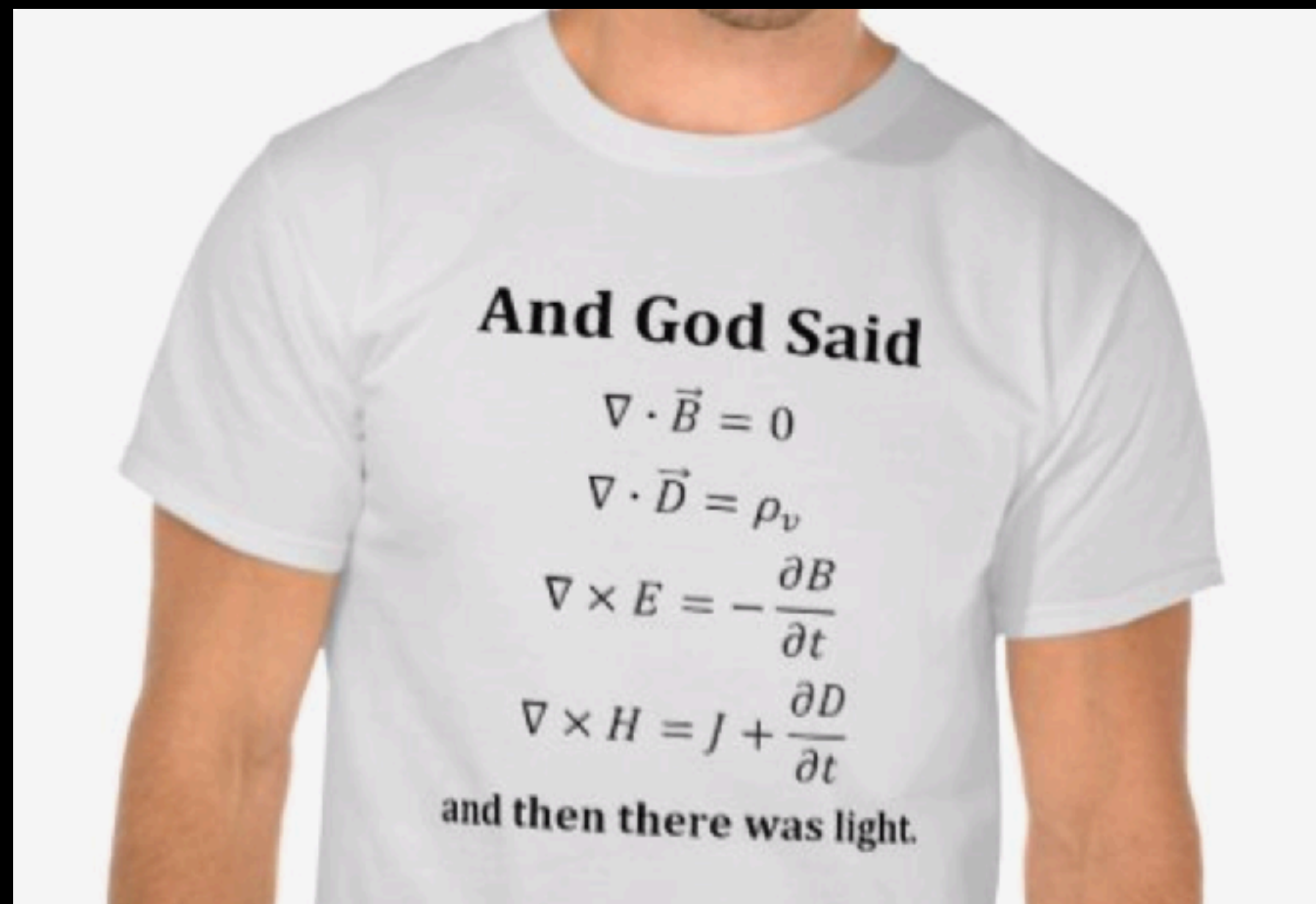
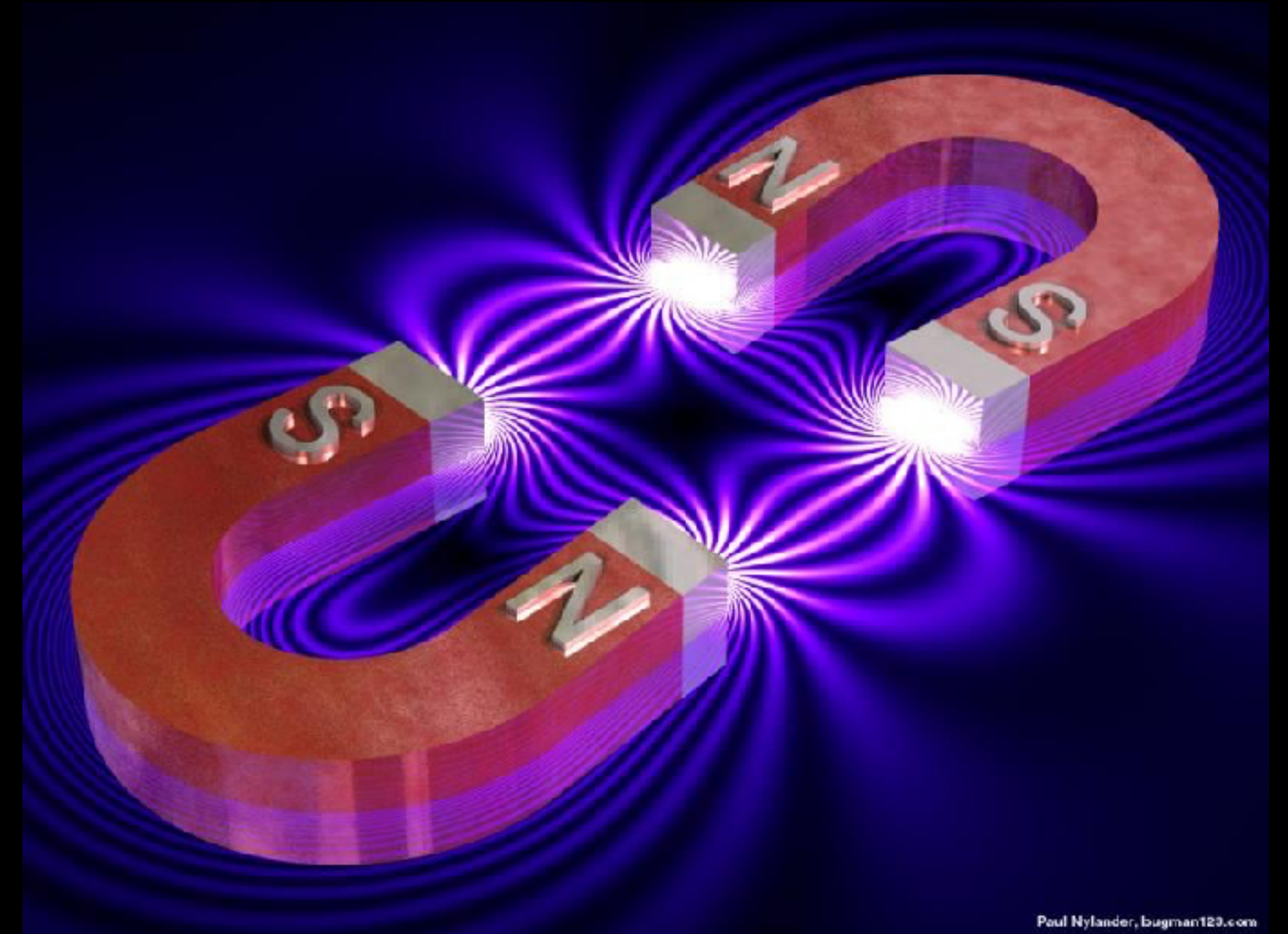


electricity



magnetism

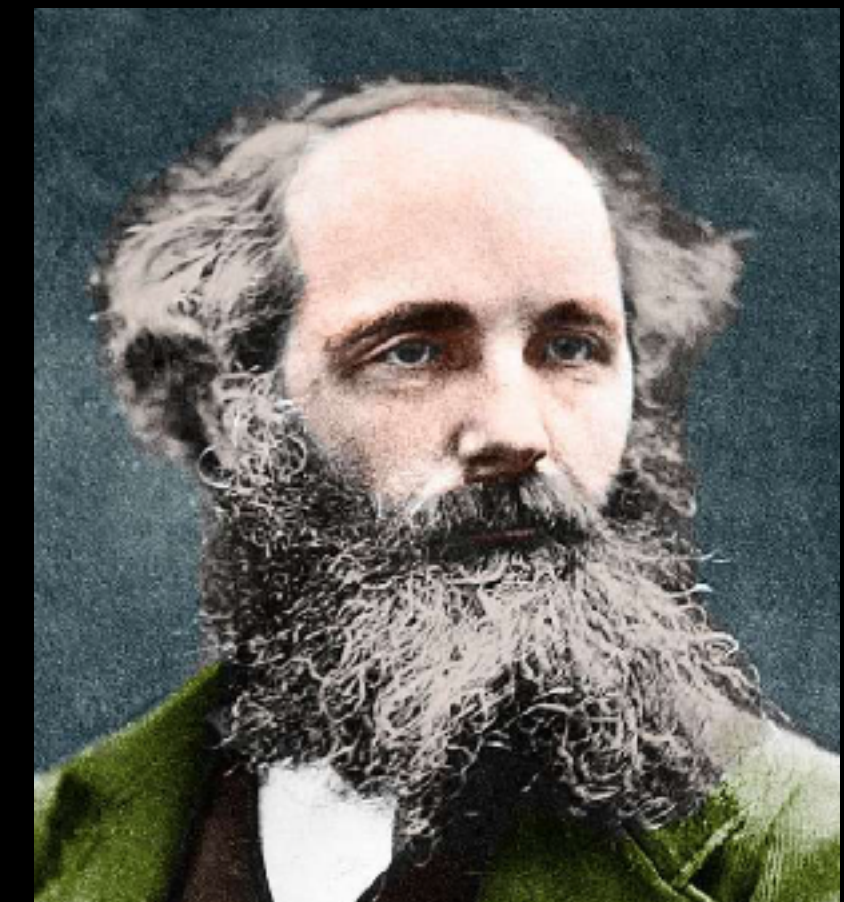
Unification



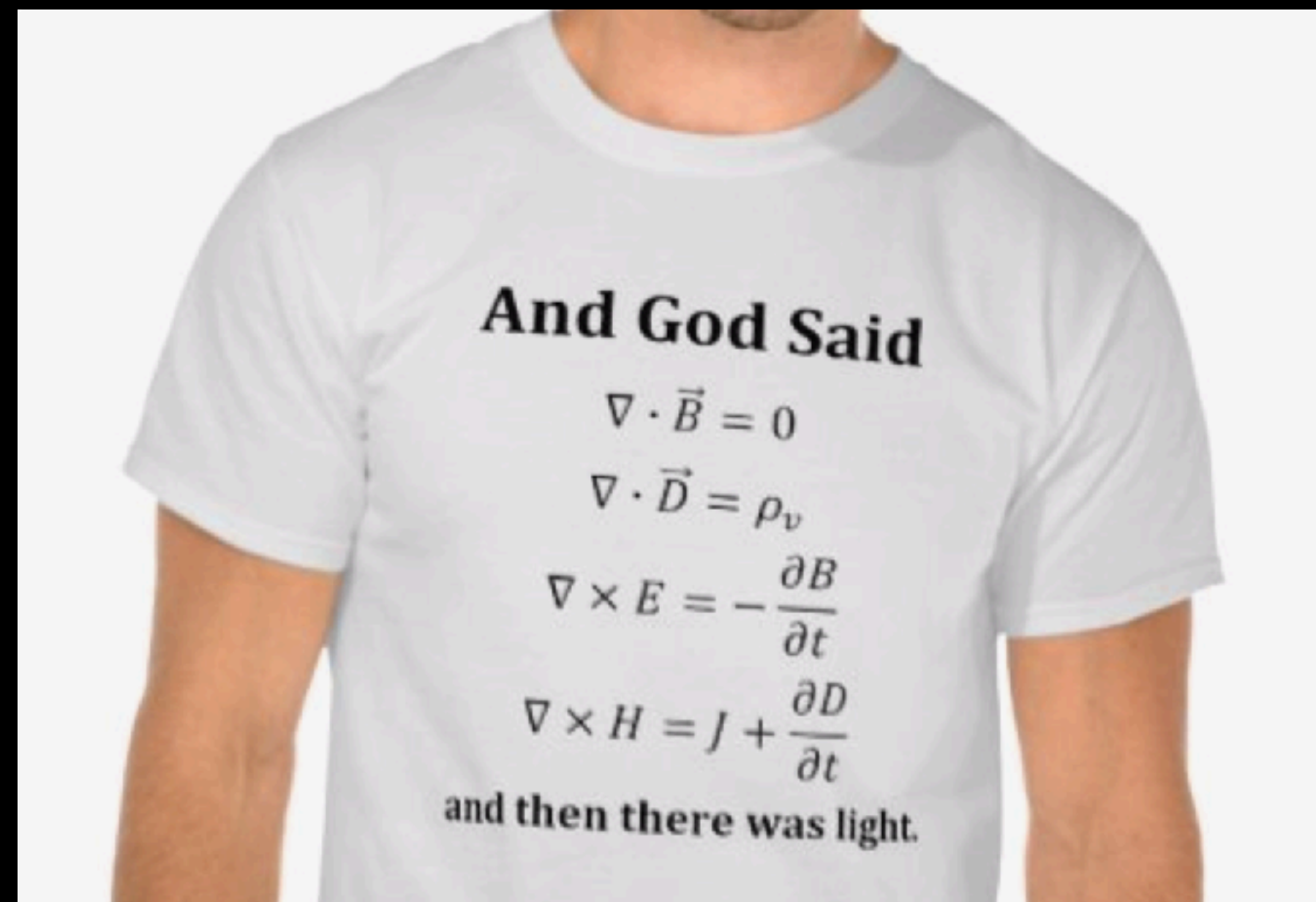
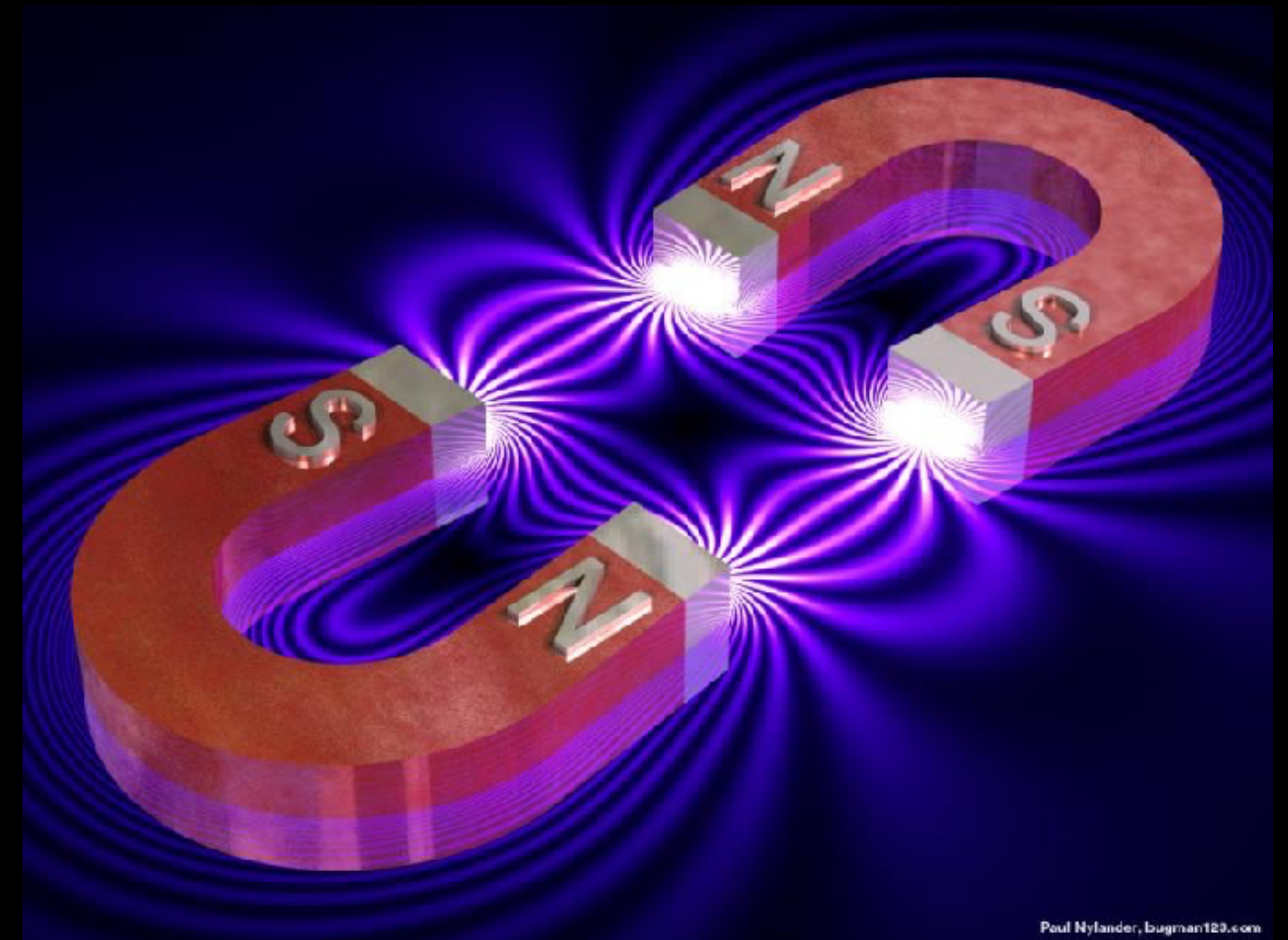
electromagnetism

Maxwell's equations
(1861)

U(1) "gauge theory"
(local symmetry)



Unification

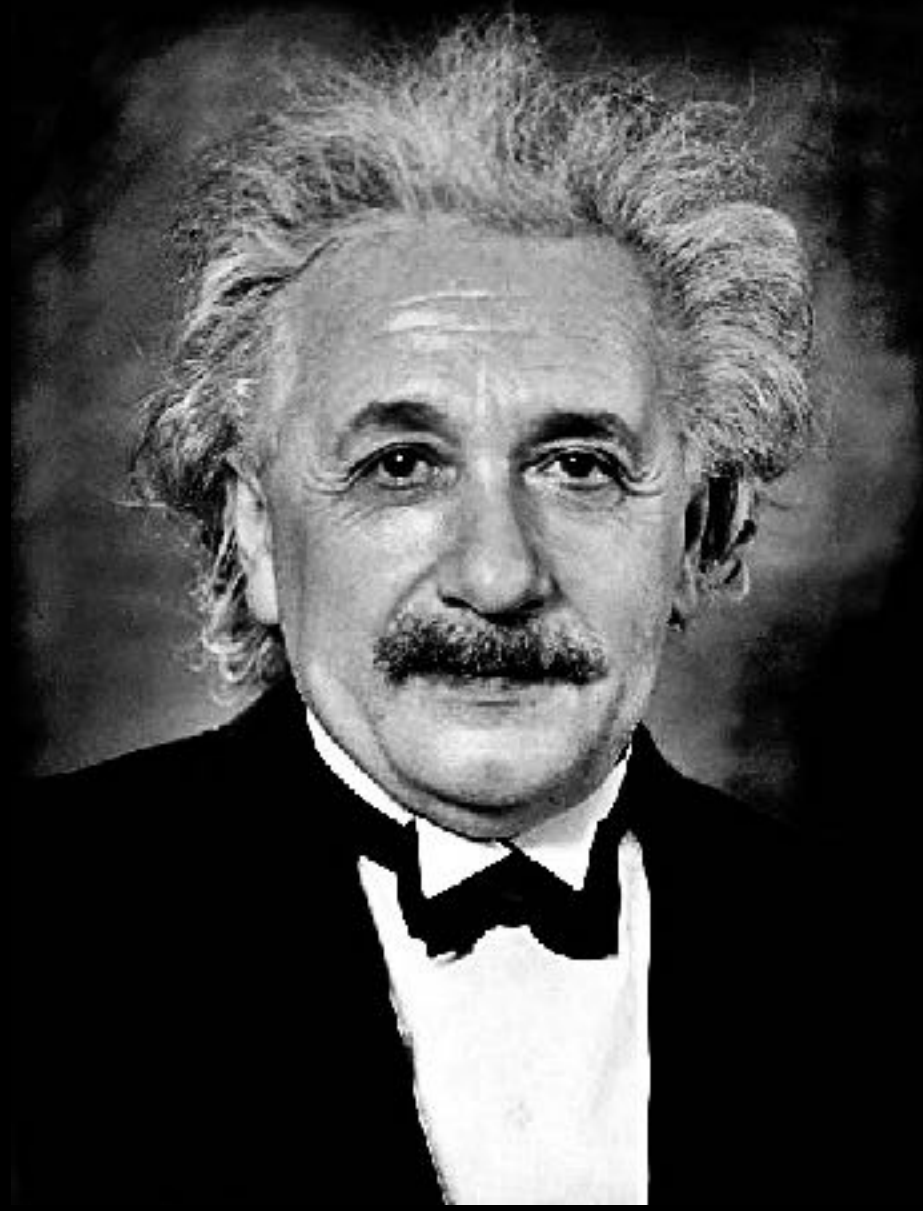


electromagnetism

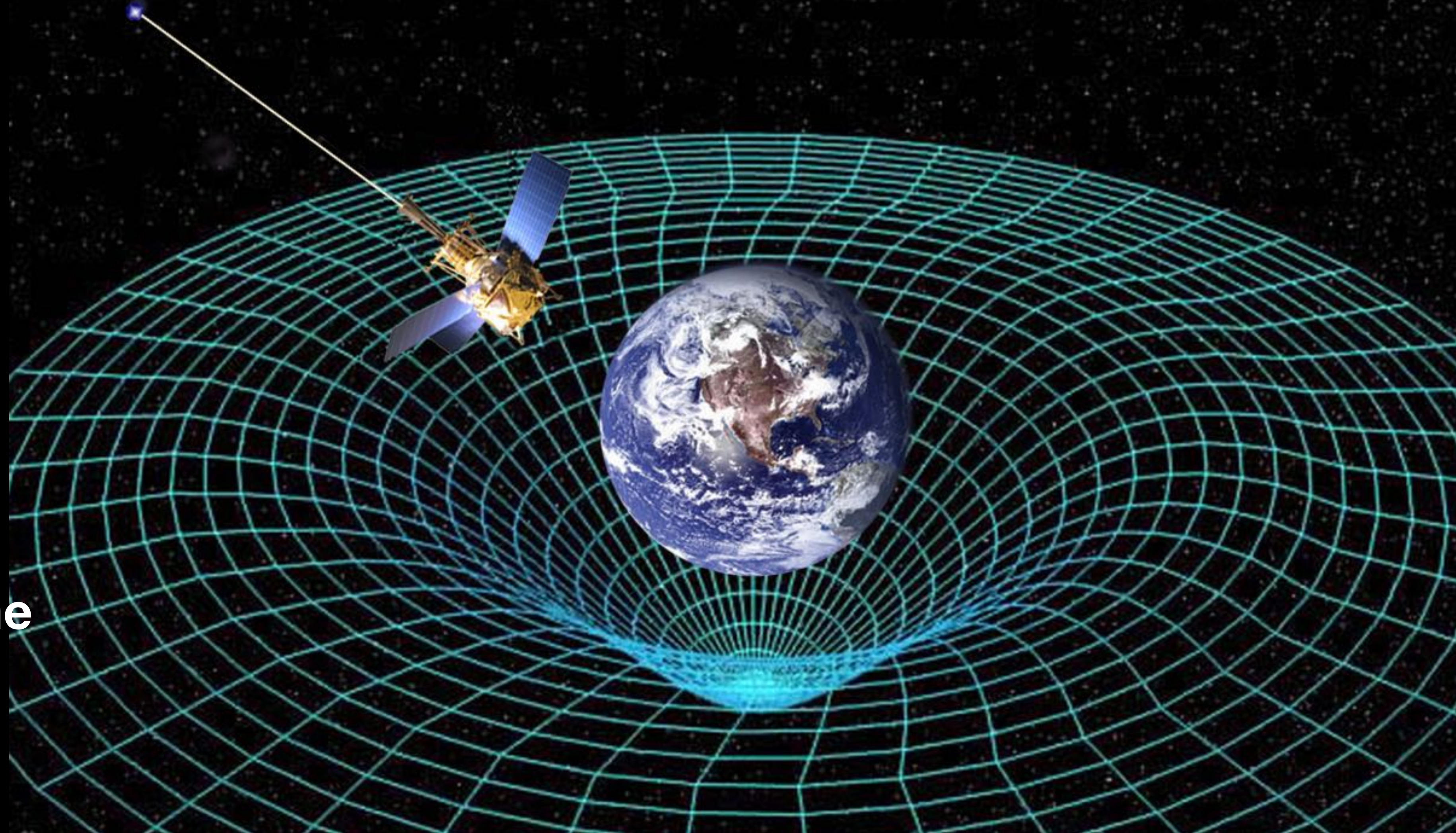
Electric-magnetic duality



(global symmetry)



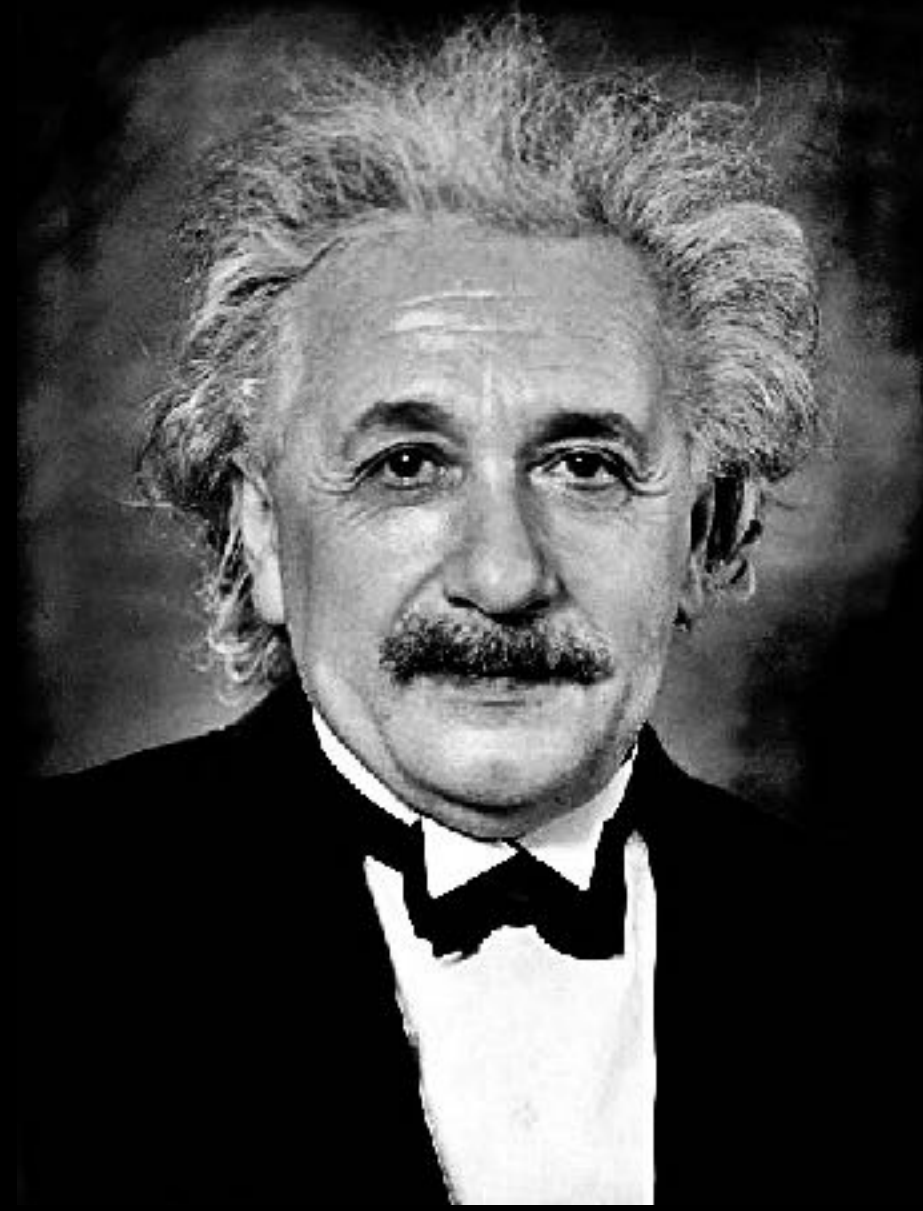
space + time = spacetime



gravity:

mass curves spacetime

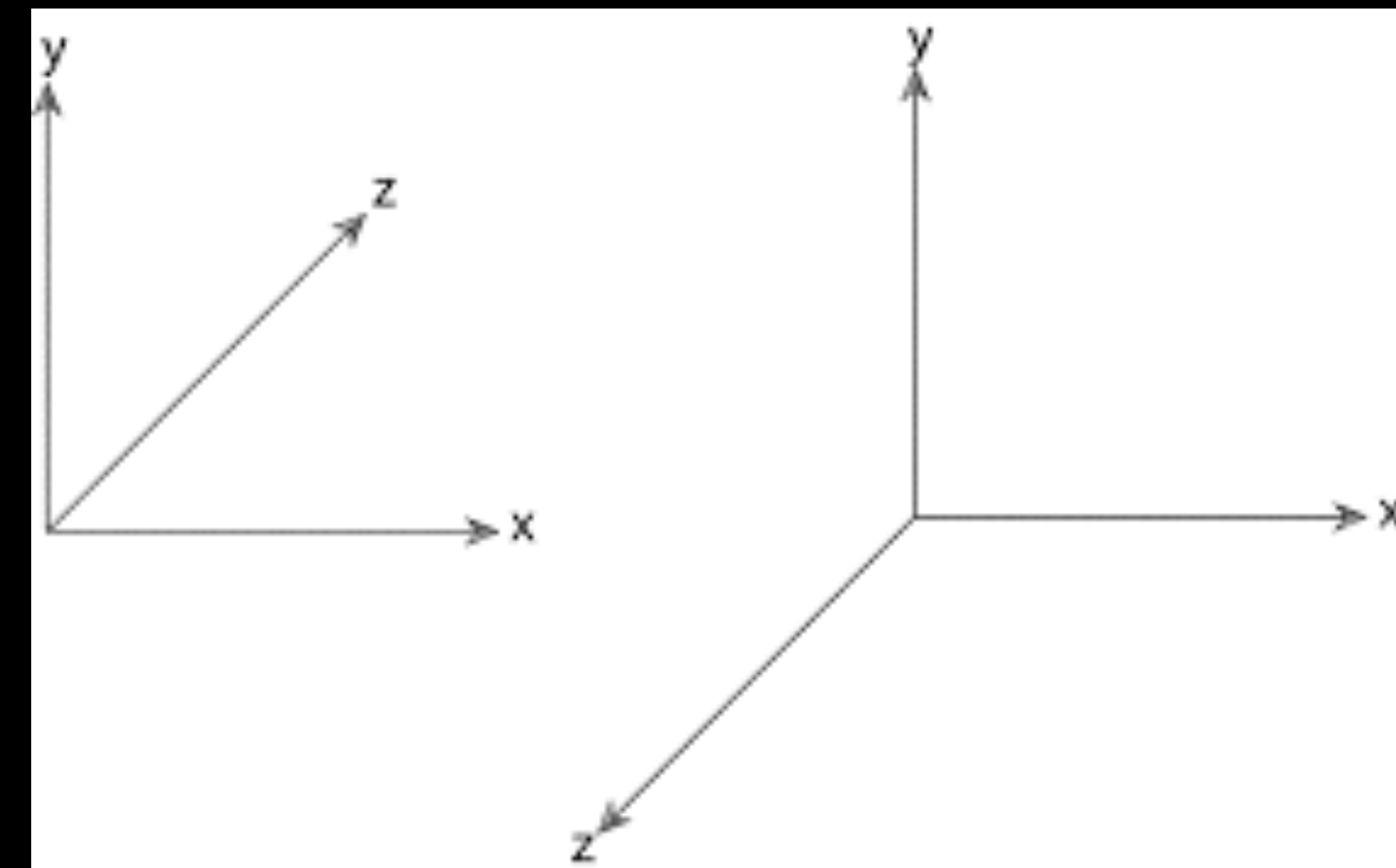
Einstein's general theory of relativity connected differential geometry and physics

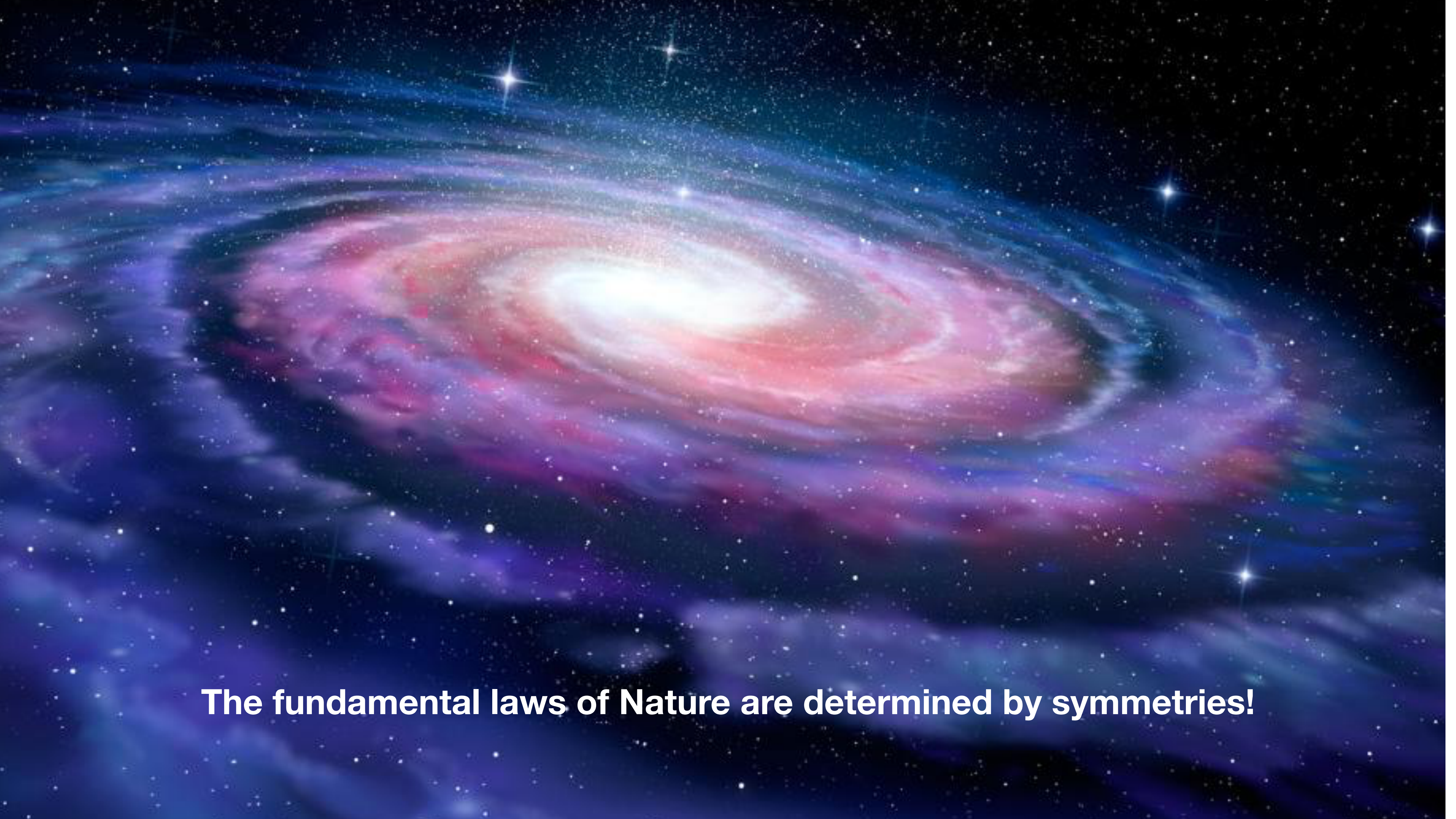


Principle of general covariance:

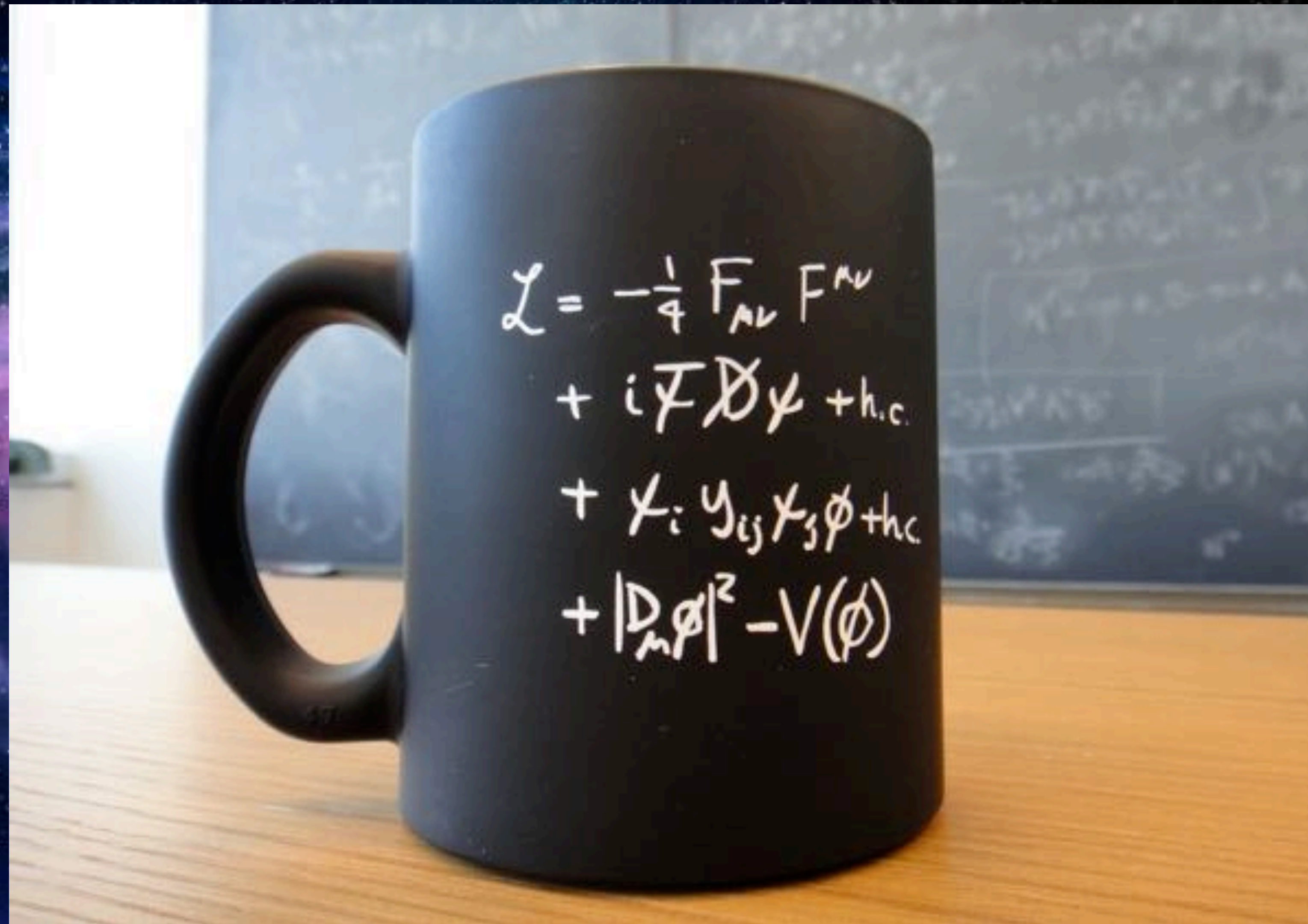
The laws of physics should take the same form independently of which coordinate system we use to represent them

$$R_{\mu\nu}(g) - \frac{1}{2}g_{\mu\nu}R(g) = T_{\mu\nu}(g, F, \dots)$$





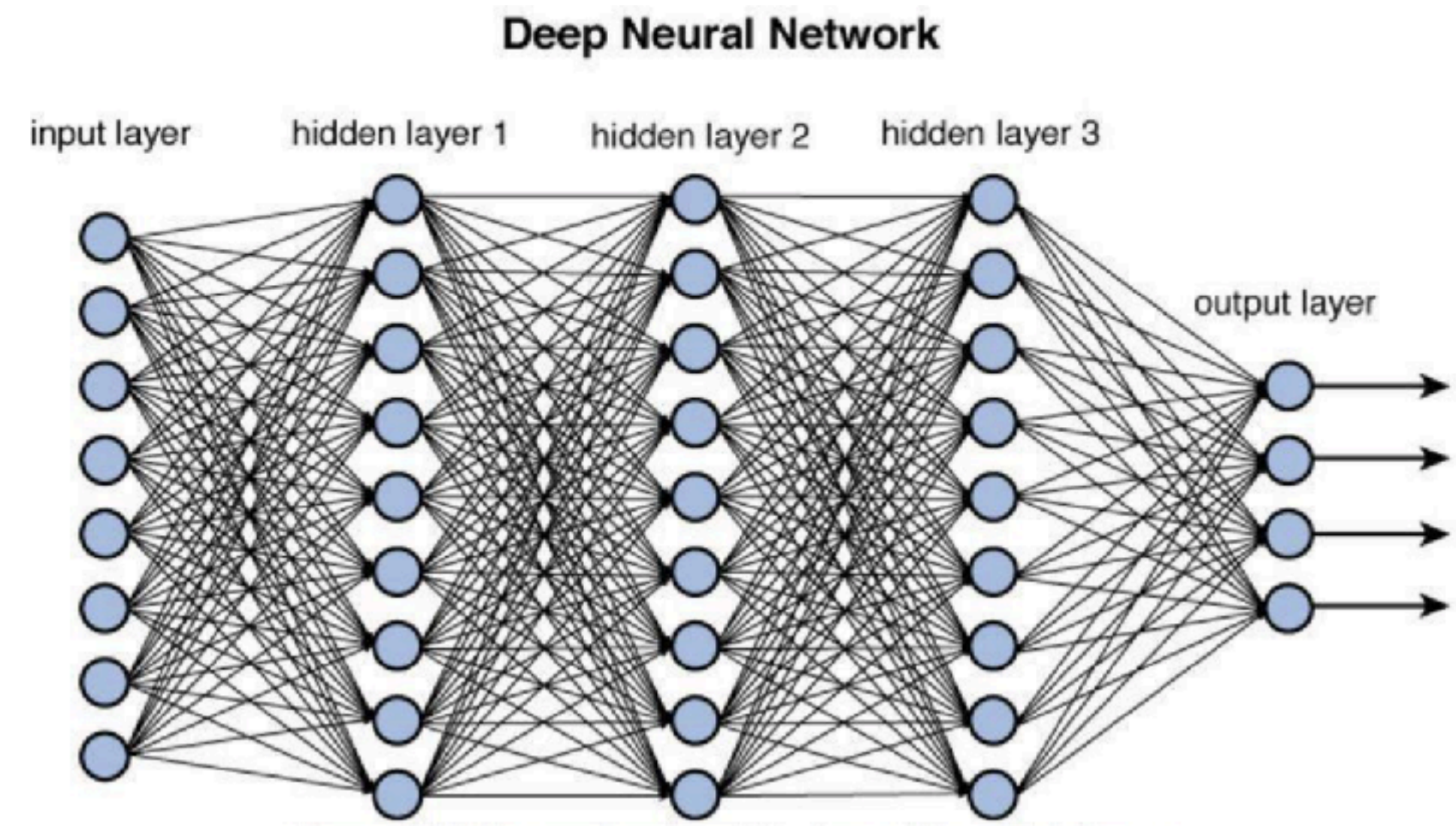
The fundamental laws of Nature are determined by symmetries!



The fundamental laws of Nature are determined by symmetries!

Symmetries in neural networks: Geometric deep learning





Inductive bias

Incorporate known constraints into the architecture/training

symmetries

differential equations

geometry



$$R_{\mu\nu}(g) - \frac{1}{2}g_{\mu\nu}R(g) = T_{\mu\nu}(g, F, \dots)$$



Image classification

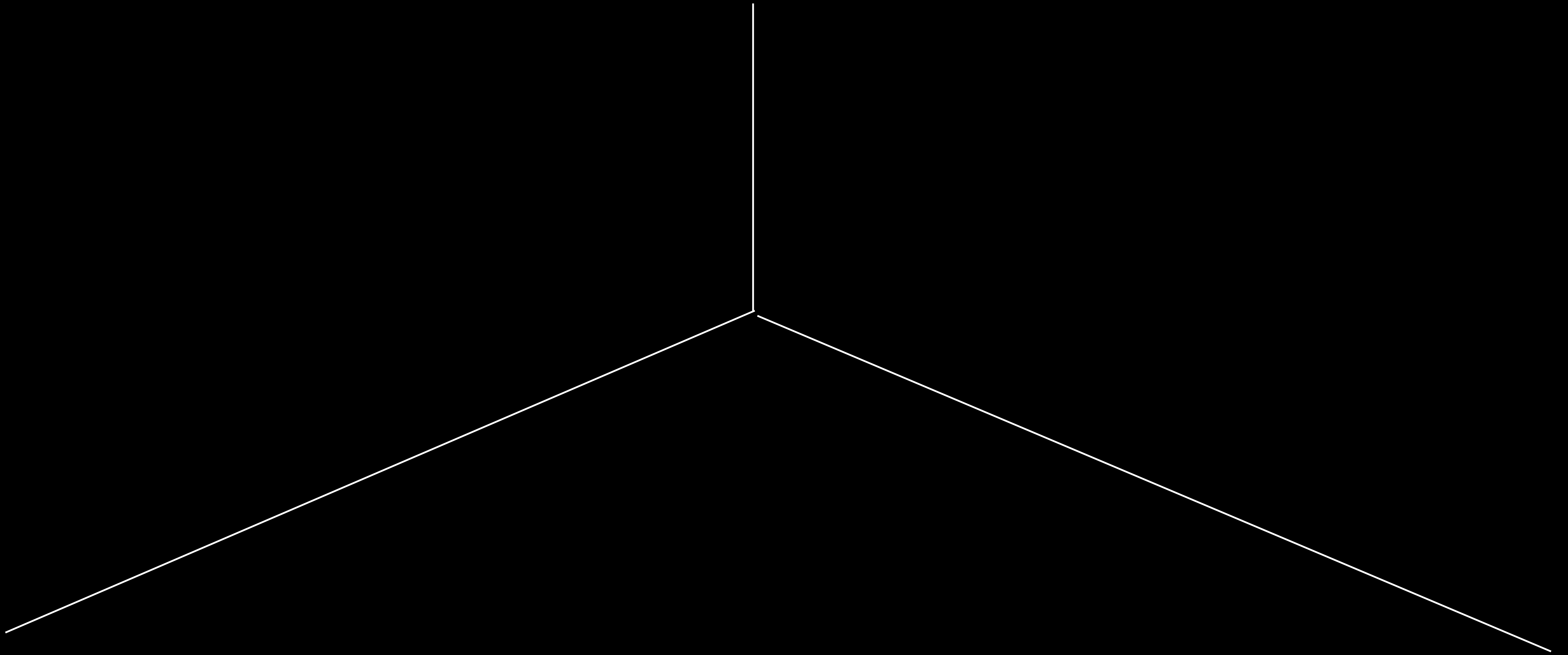


Image classification

“skateboard”



“ribs”



“boxing gloves”



Reflection symmetry



reflect



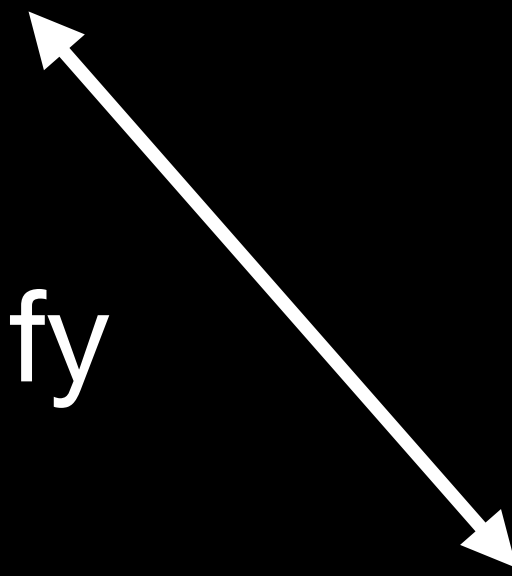
Reflection symmetry



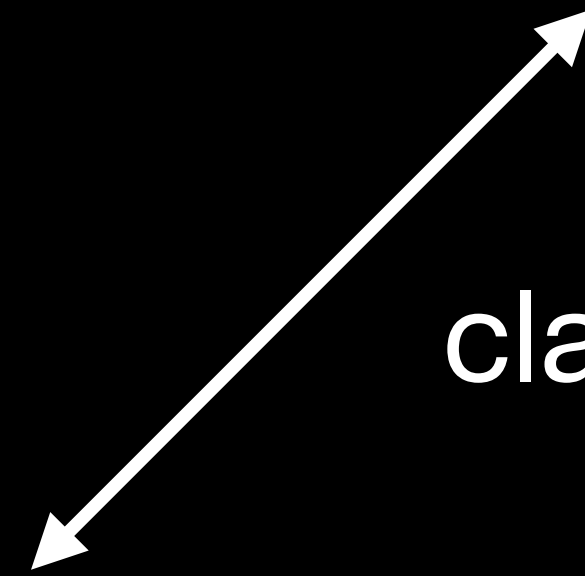
reflect



classify



classify



“boxing glove”

Reflection symmetry



reflect



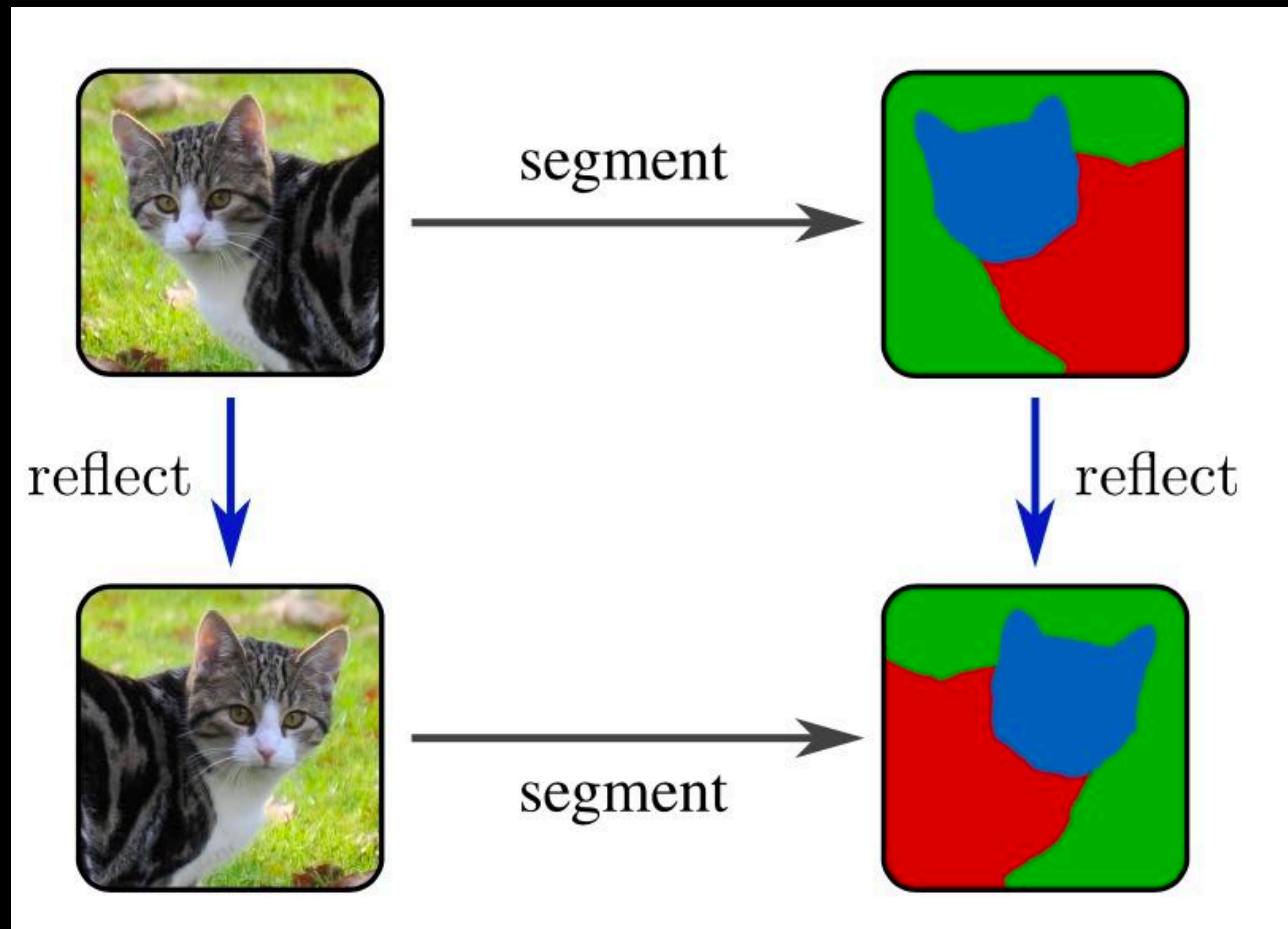
classify

classify

“boxing glove”

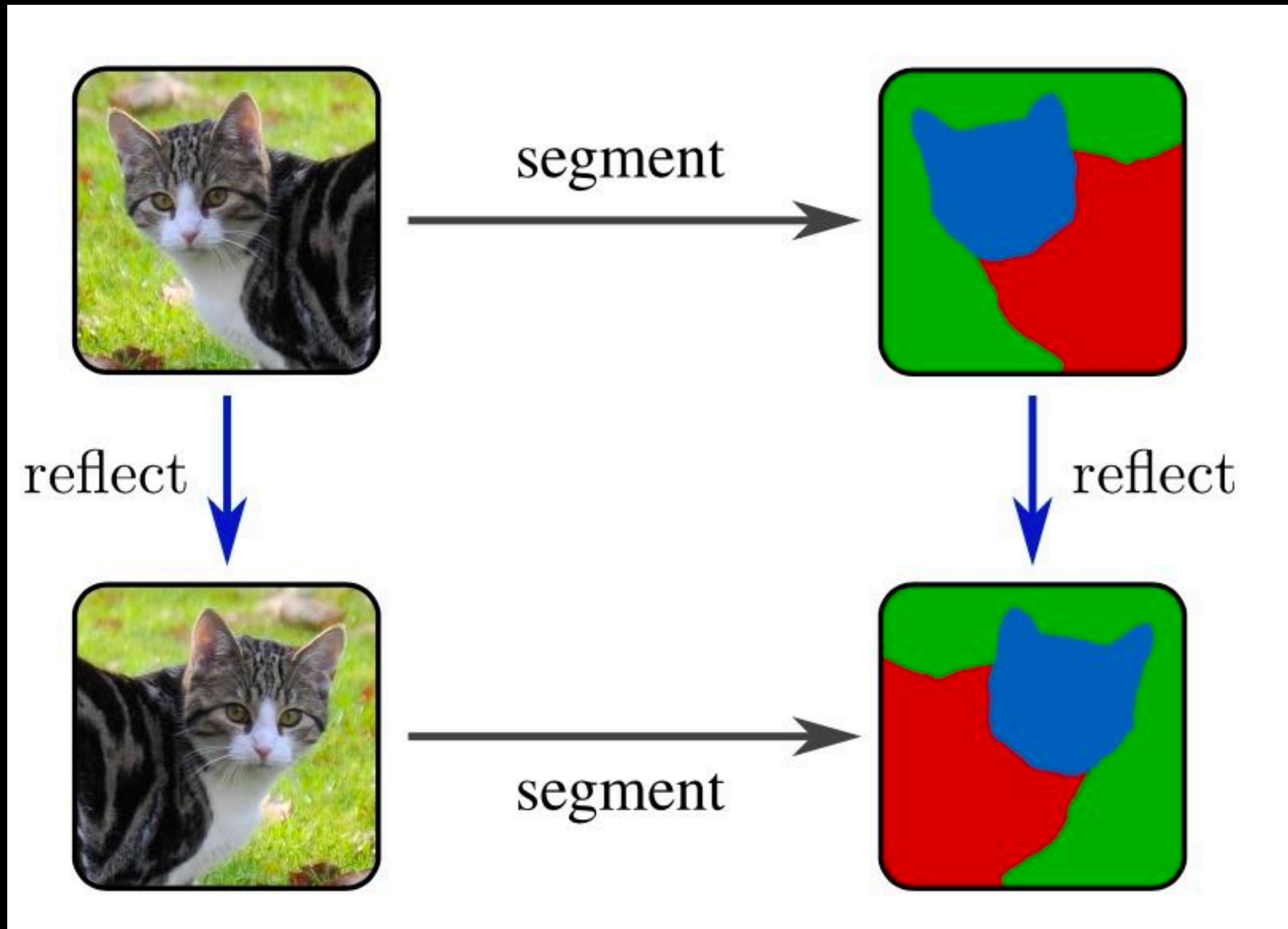
Invariance:
The output is *unchanged*
when we change the input

Segmentation



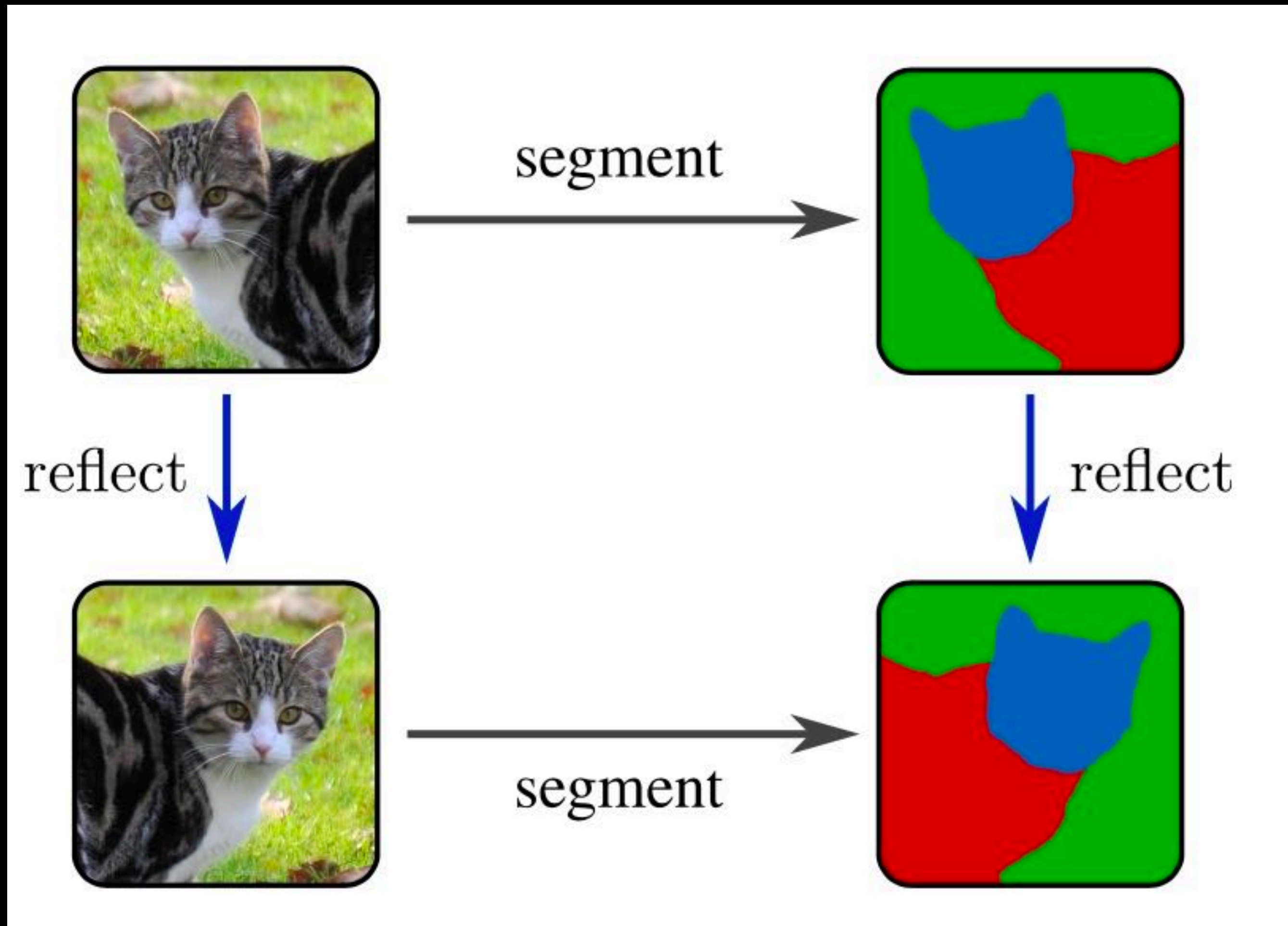
[Image from: Weiler, Forré, Verlinde, Welling (2023)]

Segmentation

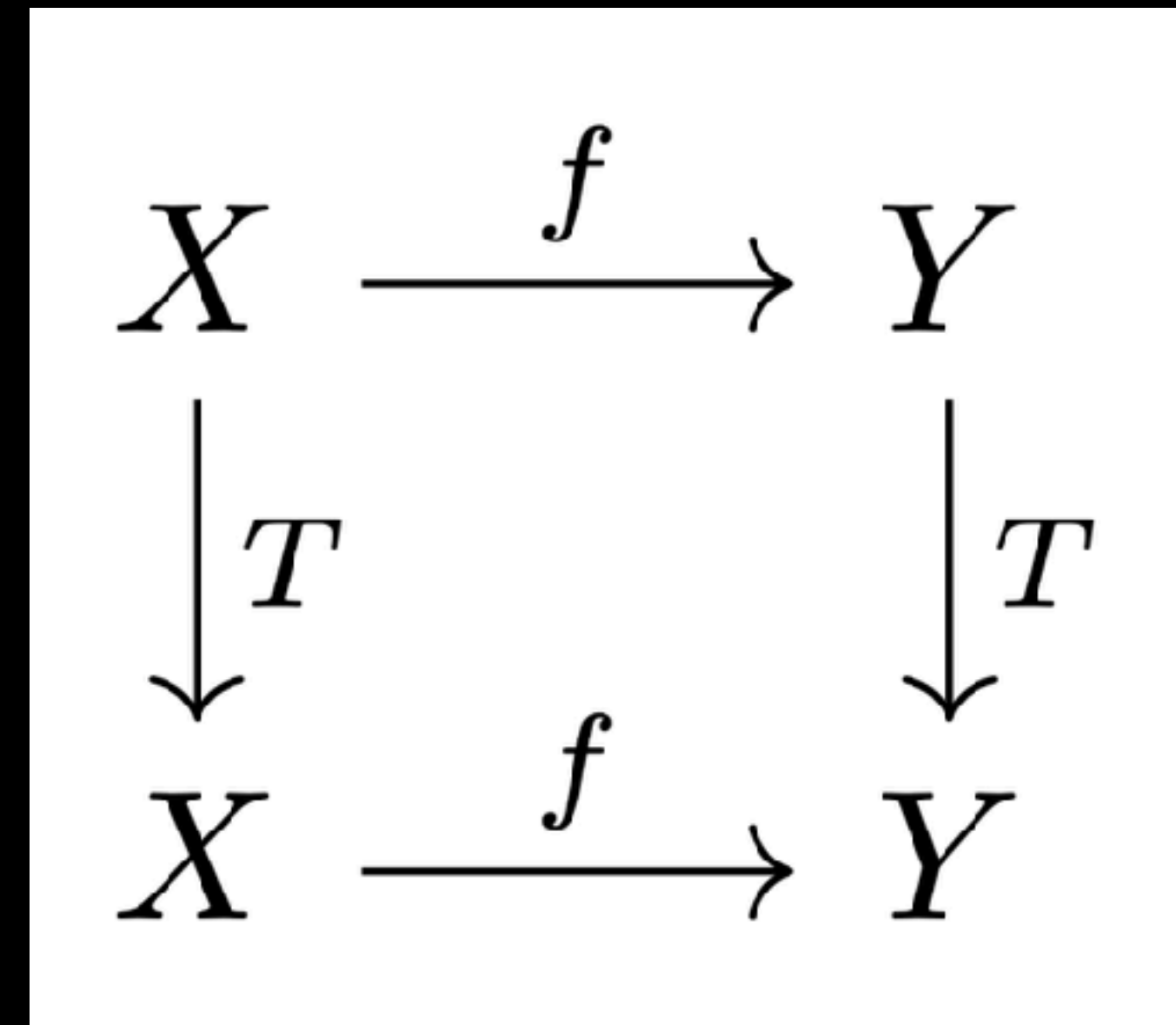


Equivariance:
The output *transforms*
according to the
transformation of the input

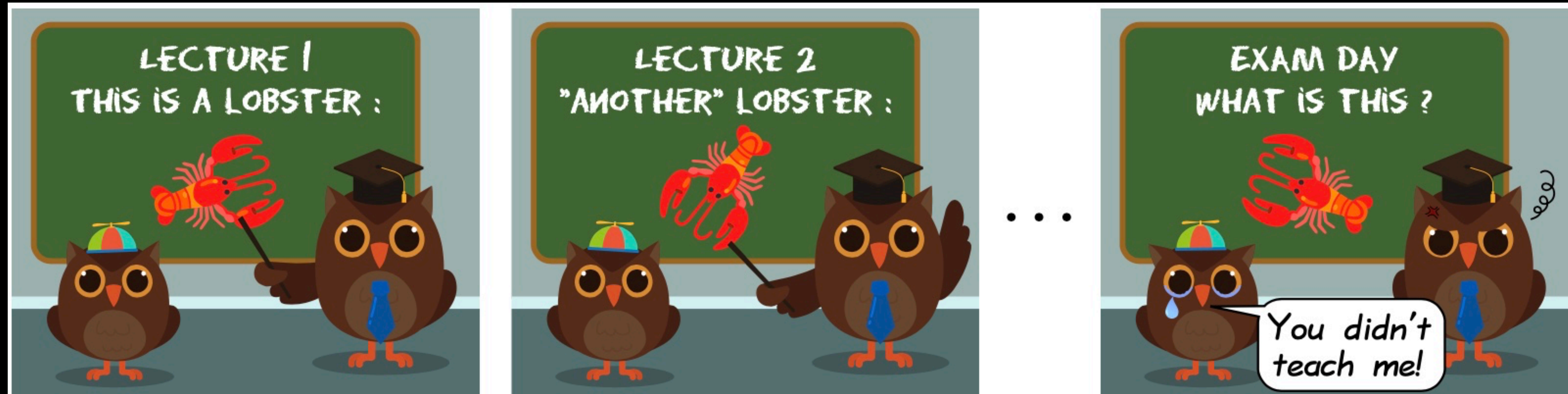
Segmentation



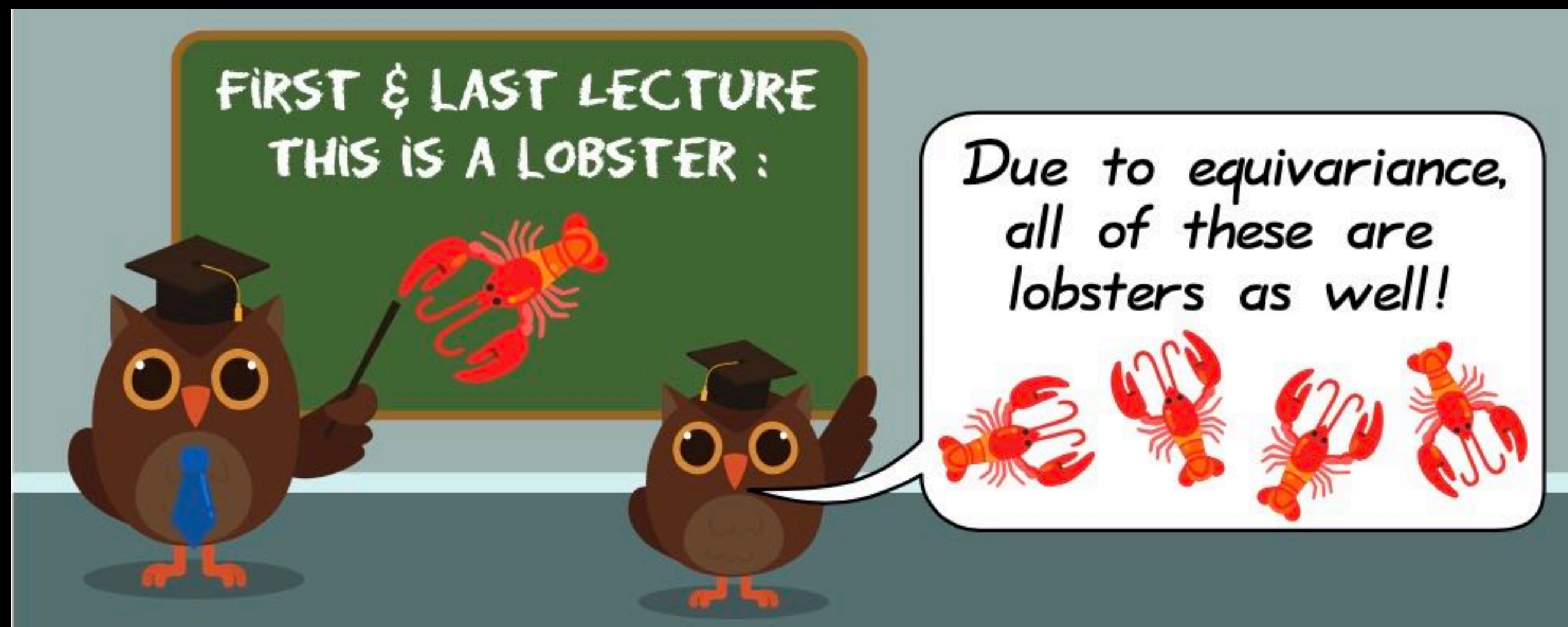
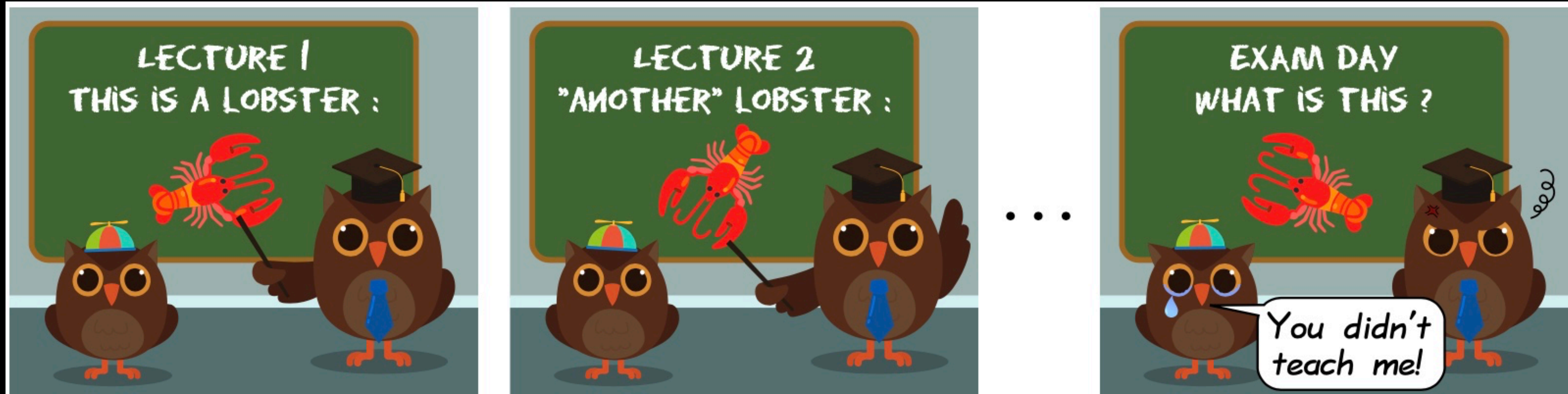
commutative
diagram



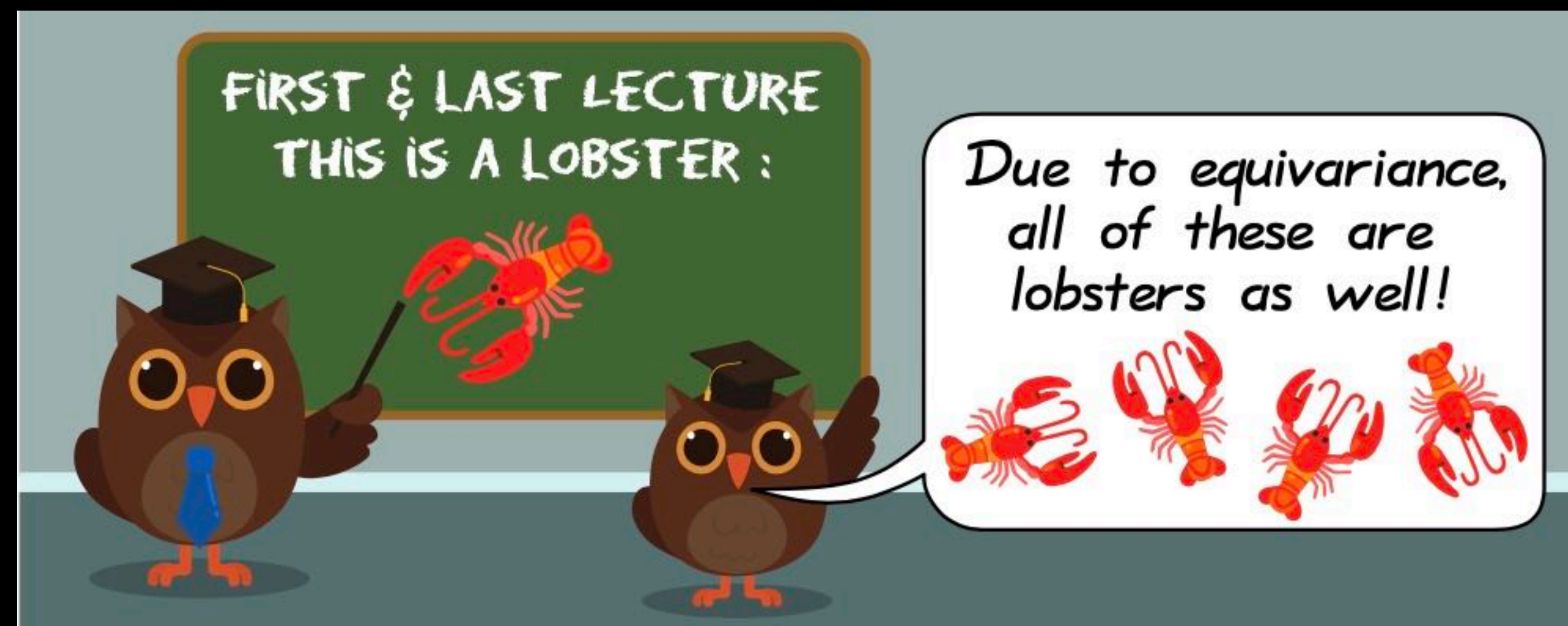
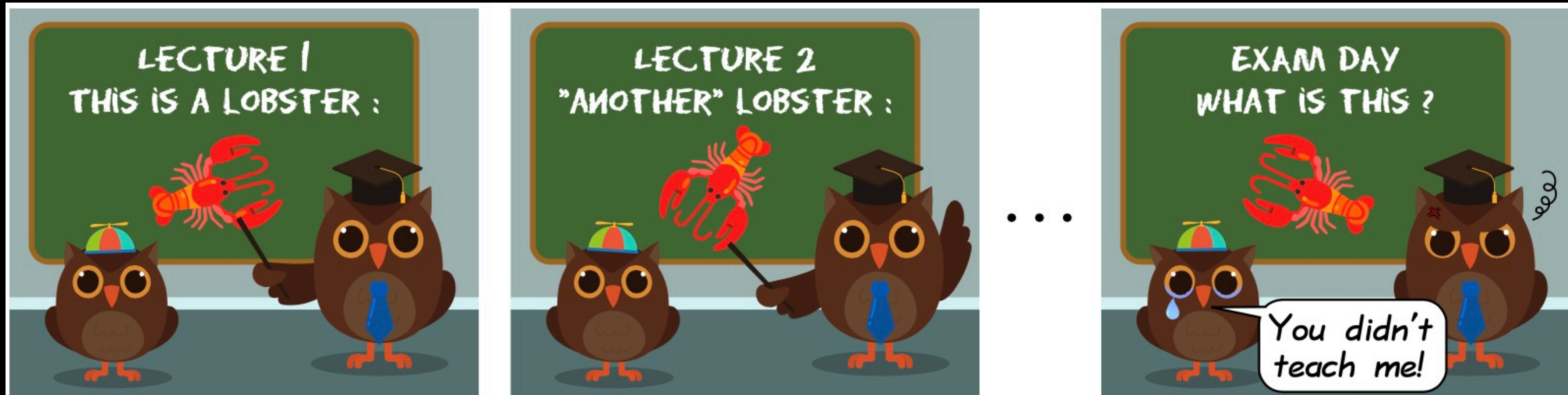
$$f(Tx) = Tf(x)$$



[Image from: Weiler, Forré, Verlinde, Welling (2023)]

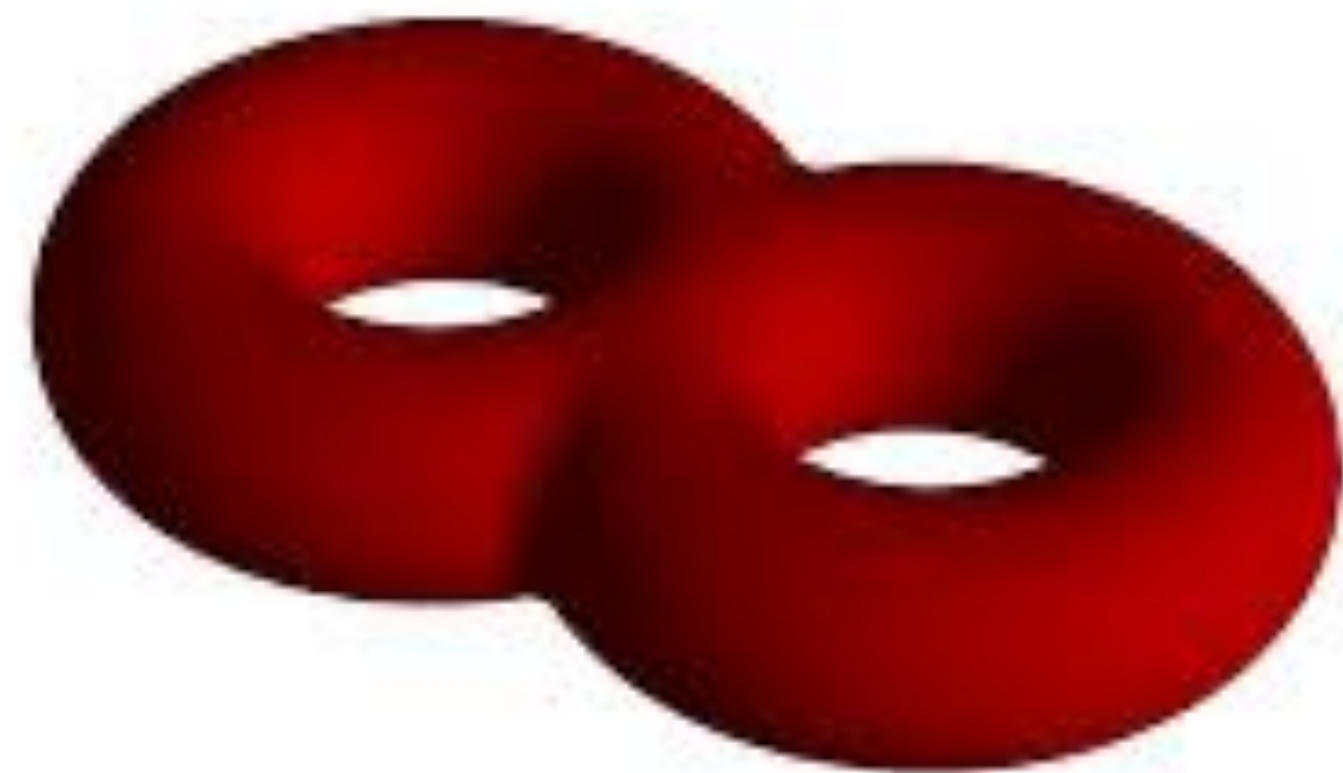
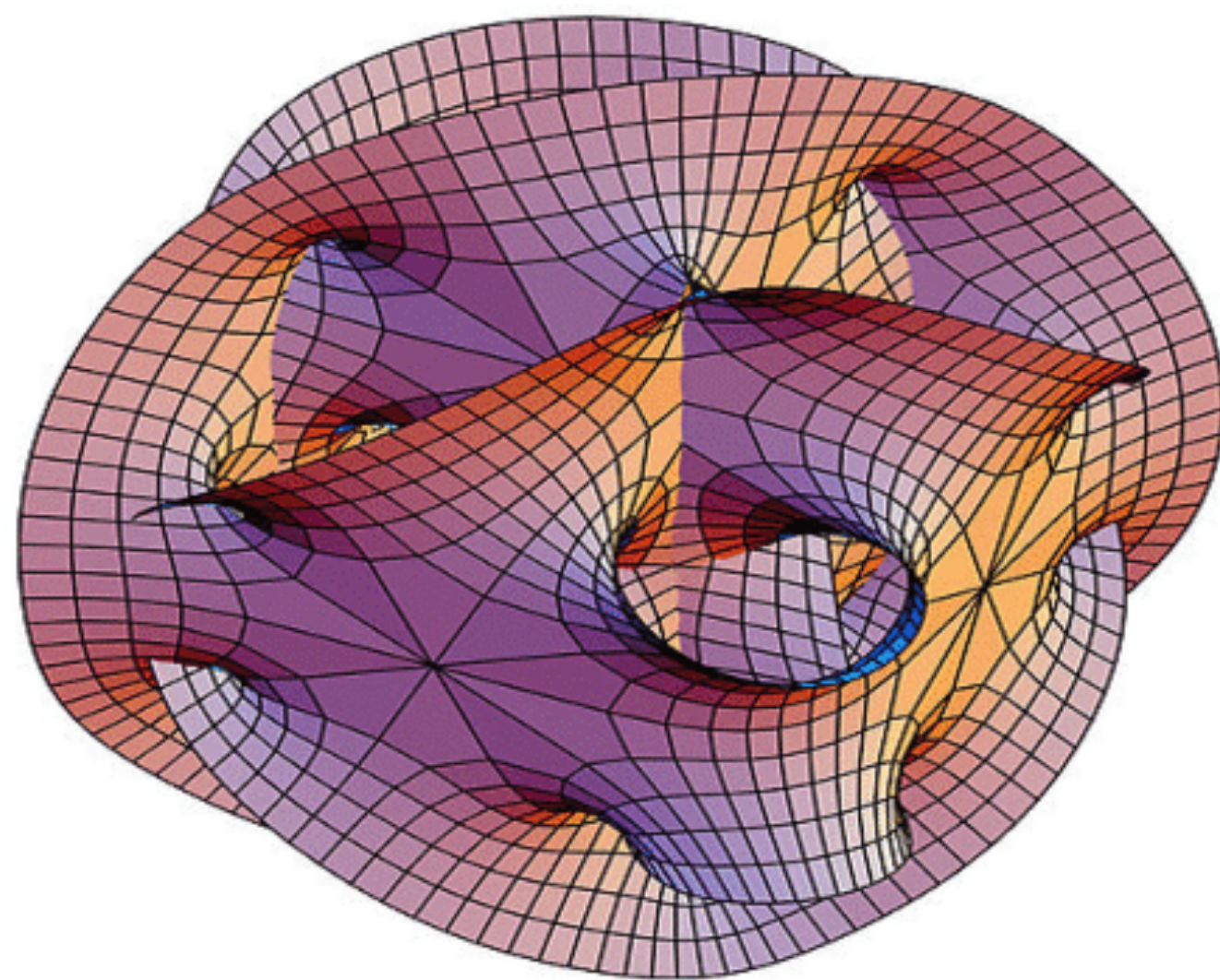
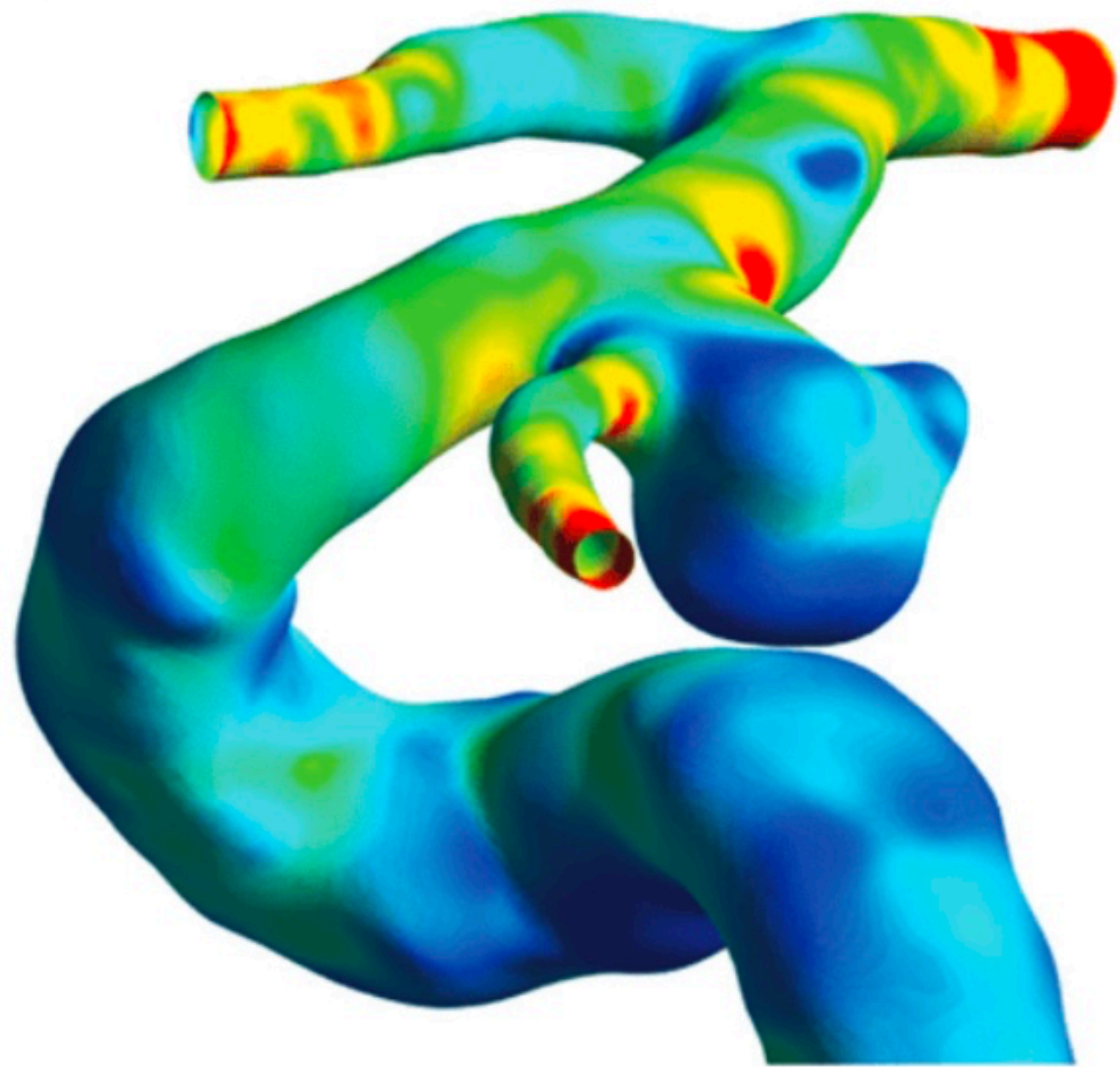


[Image from: Weiler, Forré, Verlinde, Welling (2023)]

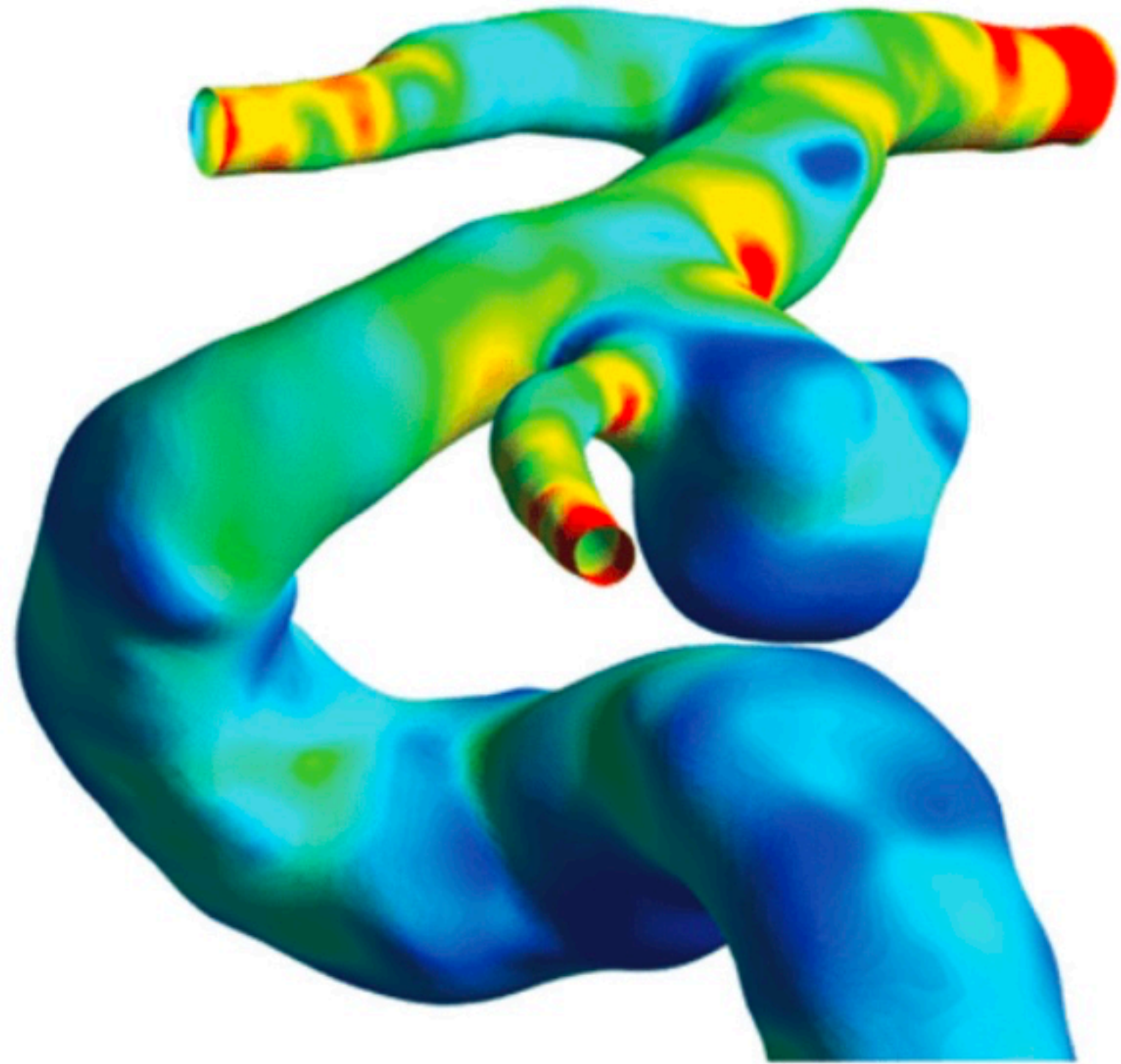


Design neural networks
that have **intrinsic symmetries**
(equivariance)

Geometric deep learning: Deep learning on manifolds

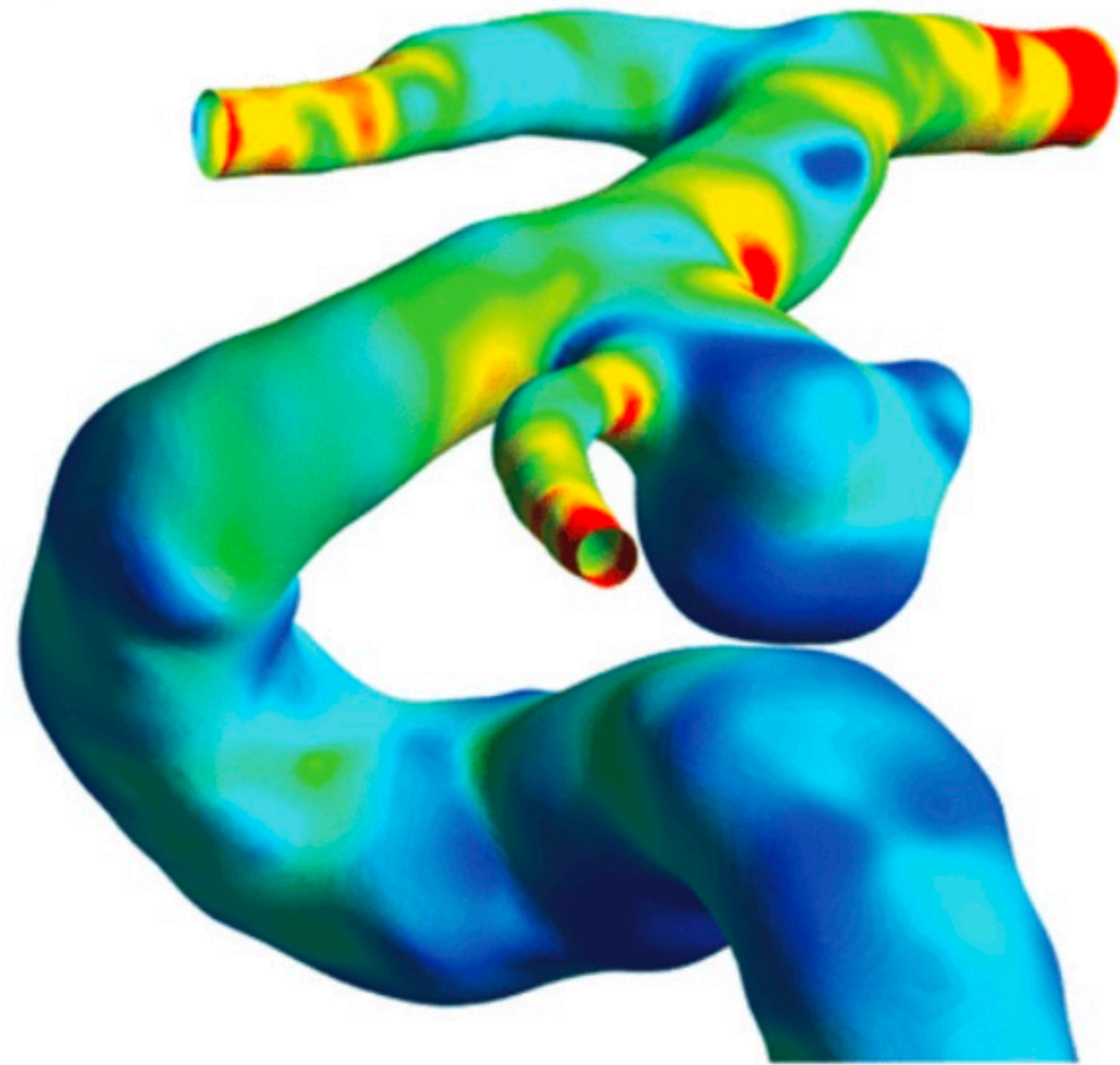


Geometric deep learning - A unified framework for deep learning

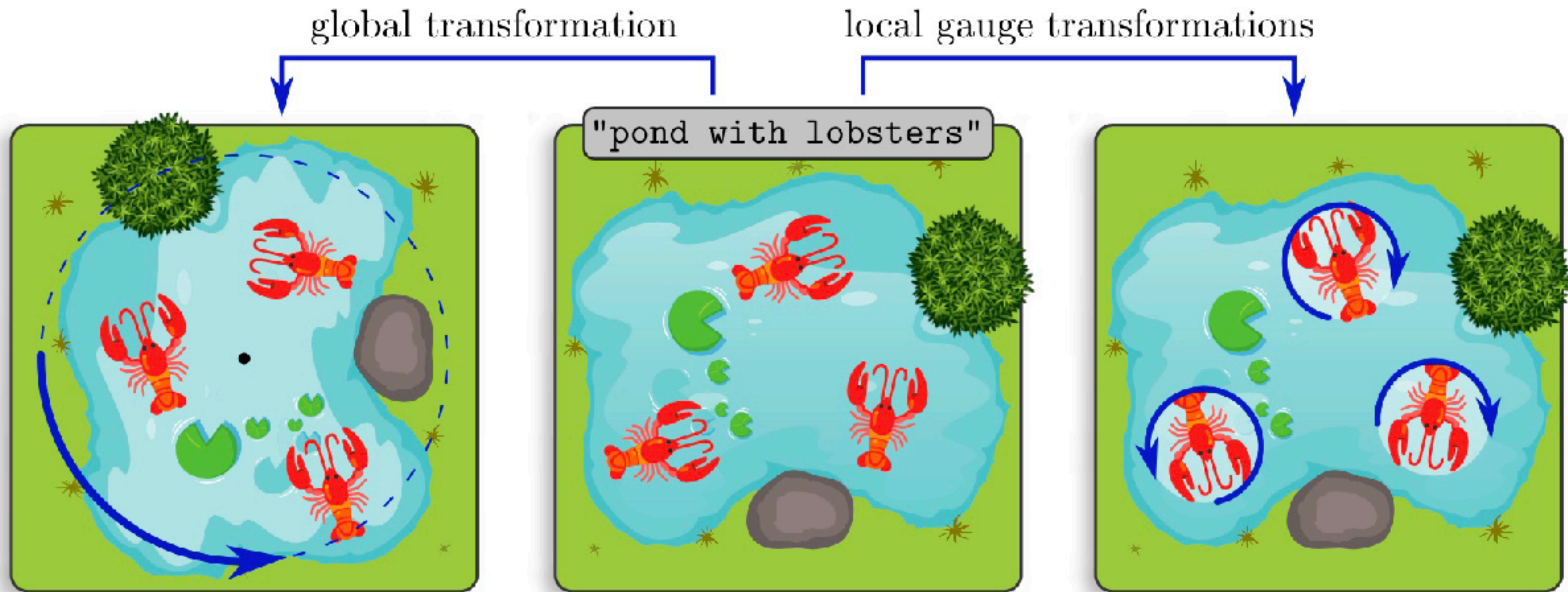


Principle of geometric deep learning:
The equations governing neural networks should be equivariant with respect to all **local** and **global** symmetries of the input data

Geometric deep learning - A unified framework for deep learning

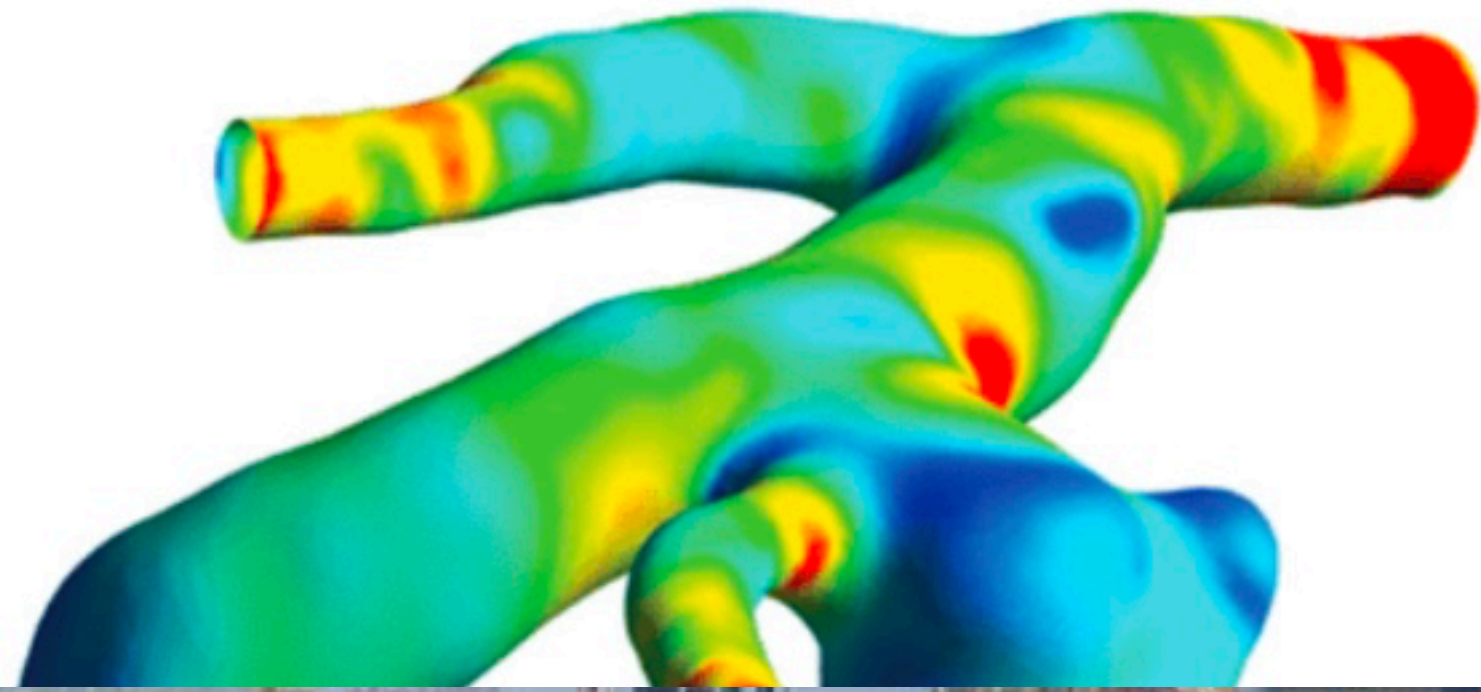


Principle of geometric deep learning:
The equations governing neural networks should be equivariant with respect to all **local** and **global** symmetries of the input data

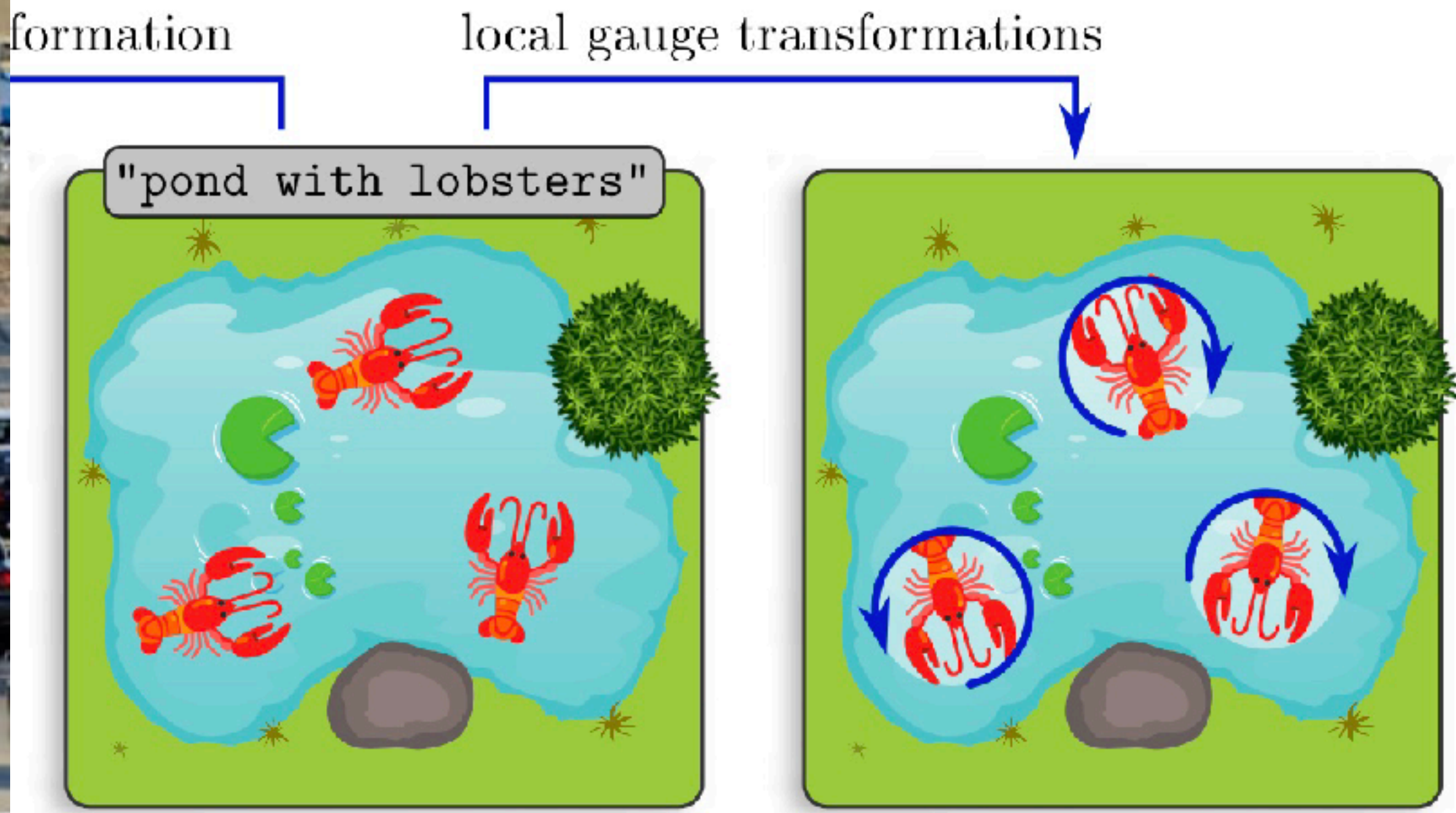


[Images from: Weiler, Forré, Verlinde, Welling (2023)]

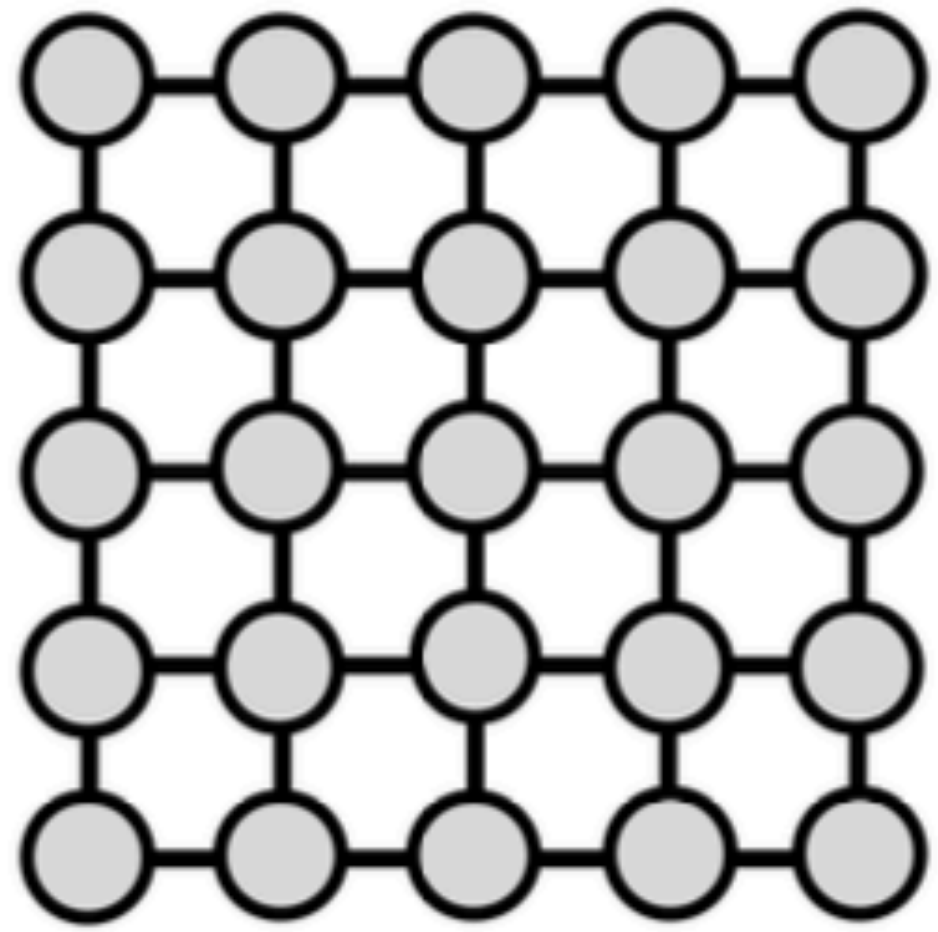
Geometric deep learning - A unified framework for deep learning



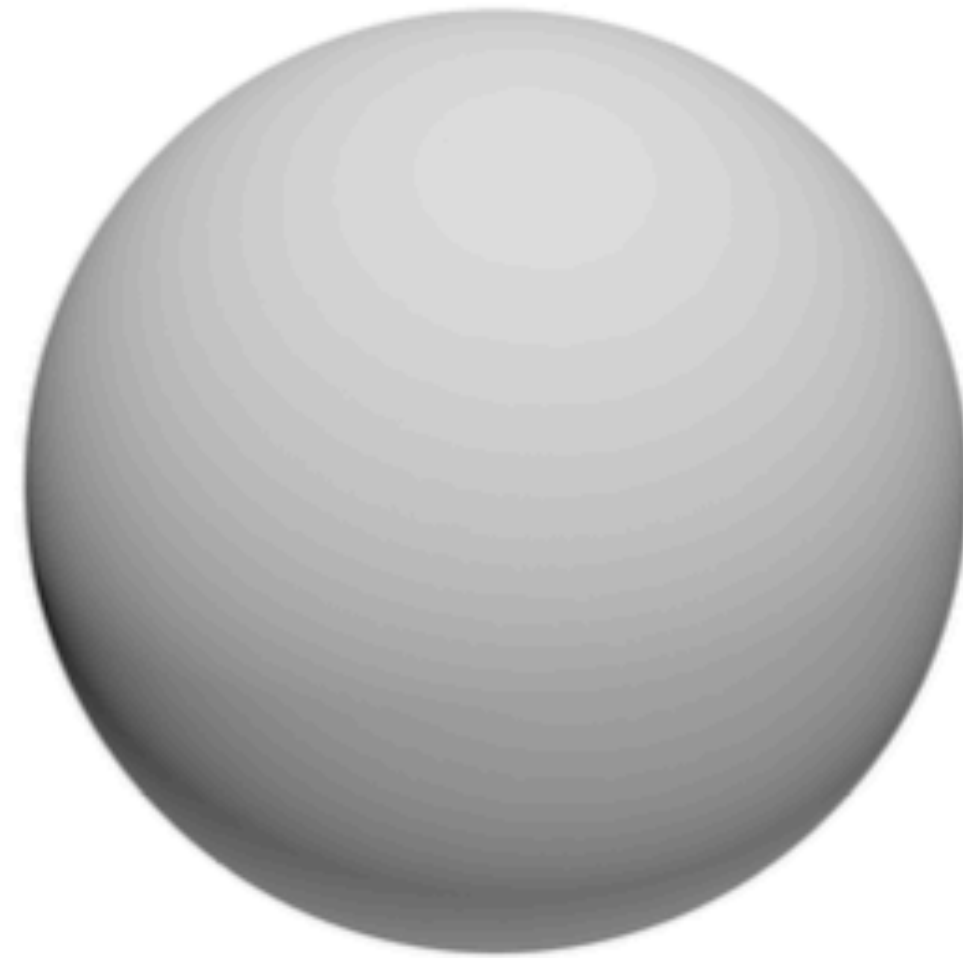
Principle of geometric deep learning:
The equations governing neural networks should be equivariant with respect to all **local** and **global** symmetries of the input data



from: Weiler, Forré, Verlinde, Welling (2023)]



Grids



Groups

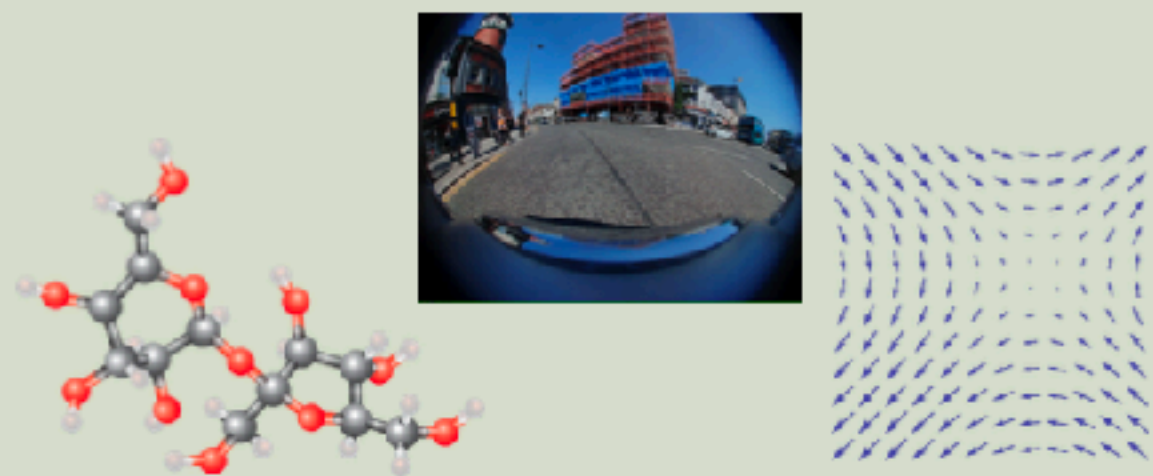


Graphs

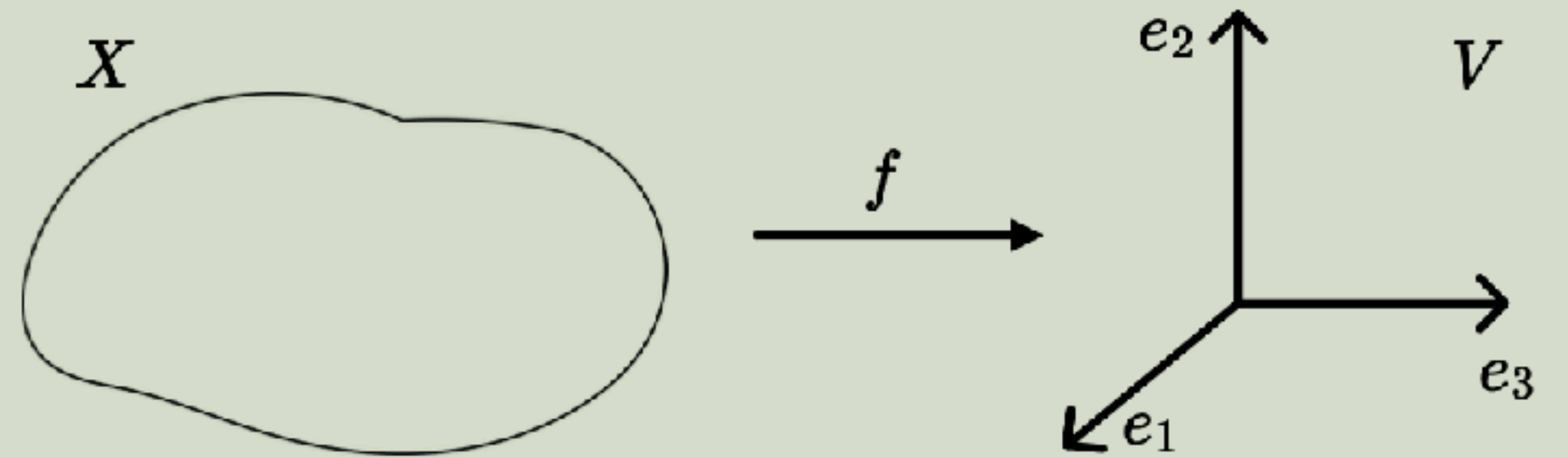


Geodesics & Gauges

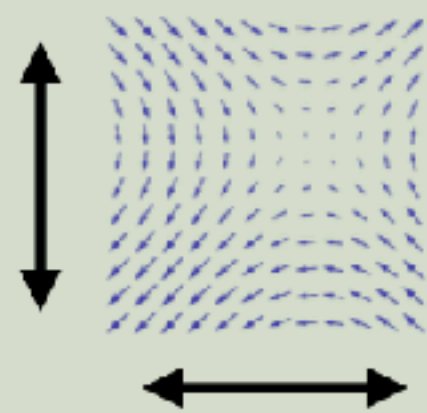
Vector data



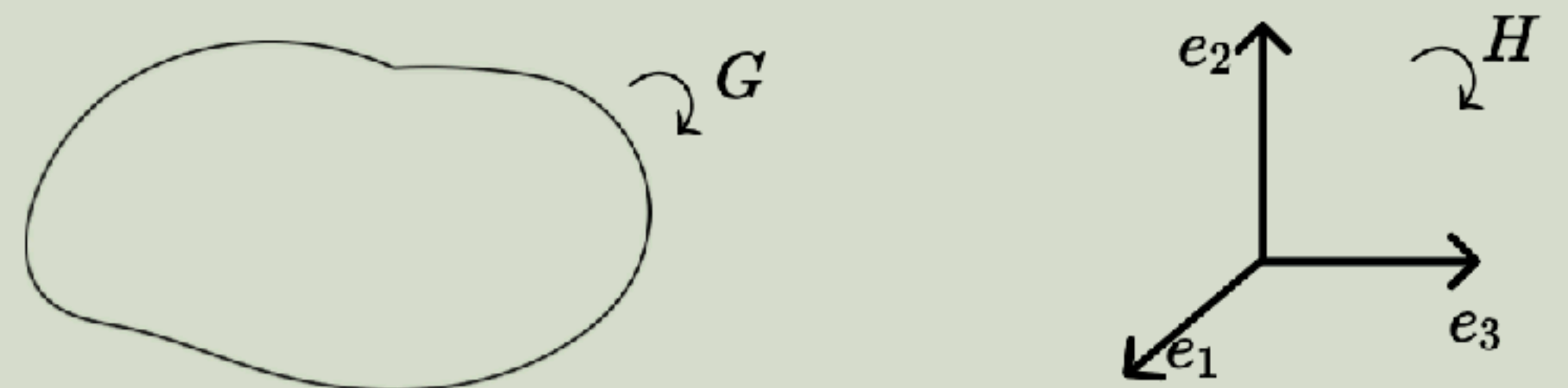
Vector-valued function $f : X \rightarrow V$



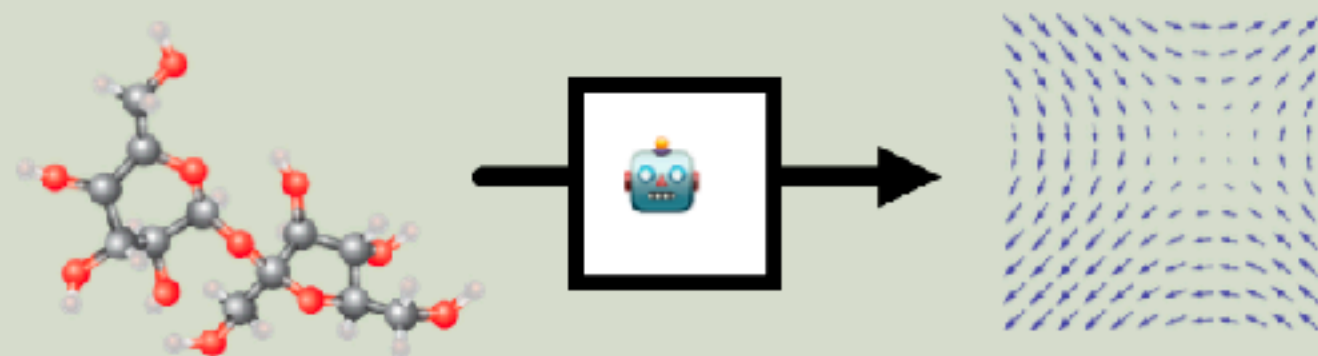
Symmetries in data



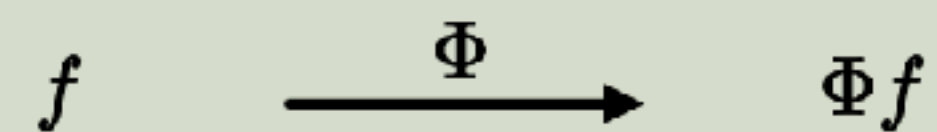
Group action of G on X + representation of H on V



Machine learning model



Operator $\Phi : \text{Map}(X, V) \rightarrow \text{Map}(X, V)$

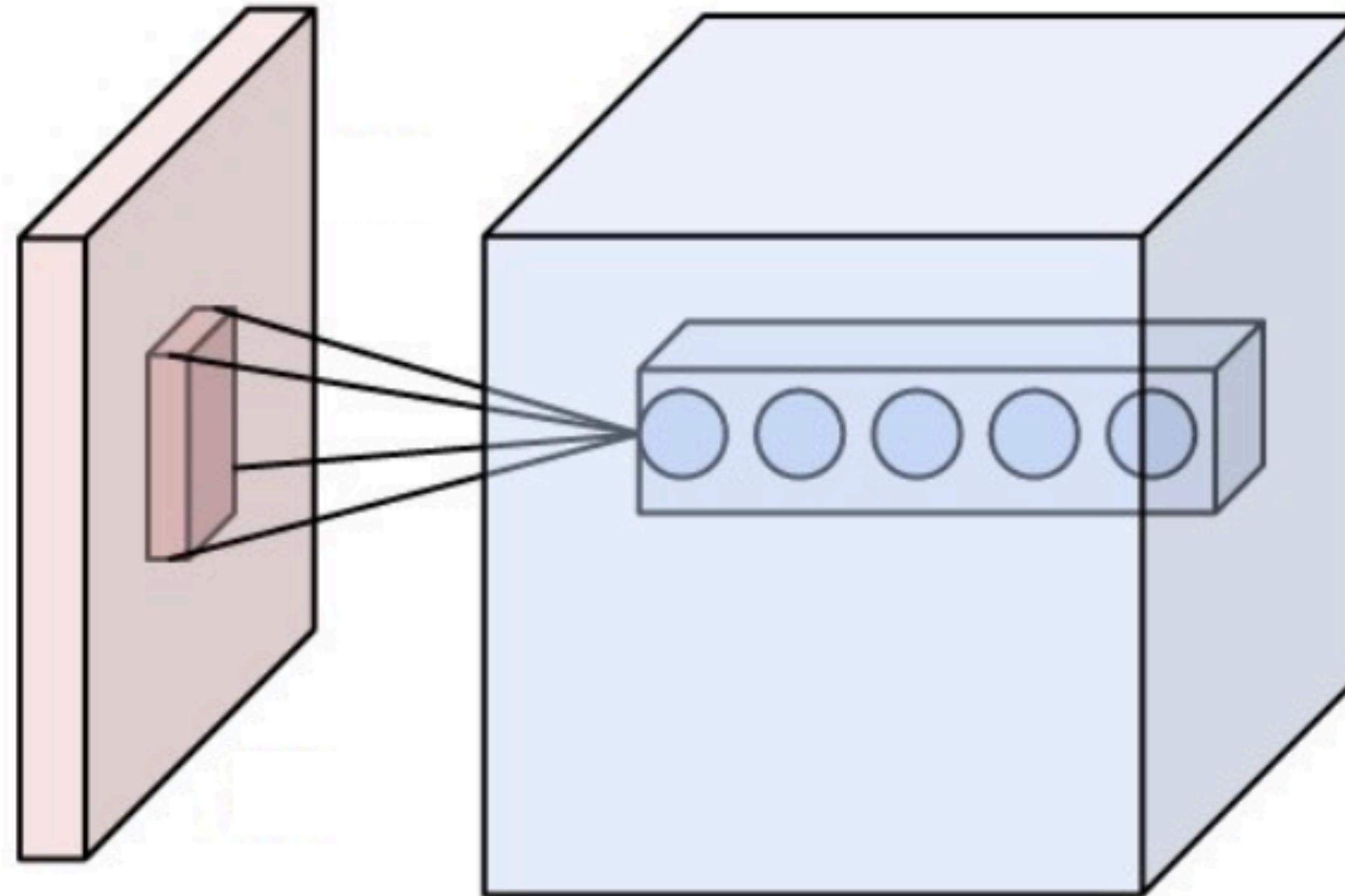


[Slide borrowed from my PhD student Elias!]

Convolutional Neural Networks

“Convolutional networks are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers.”

[Goodfellow, Bengio, Courville]



Mathematical structure

For each layer we have a **feature map**:

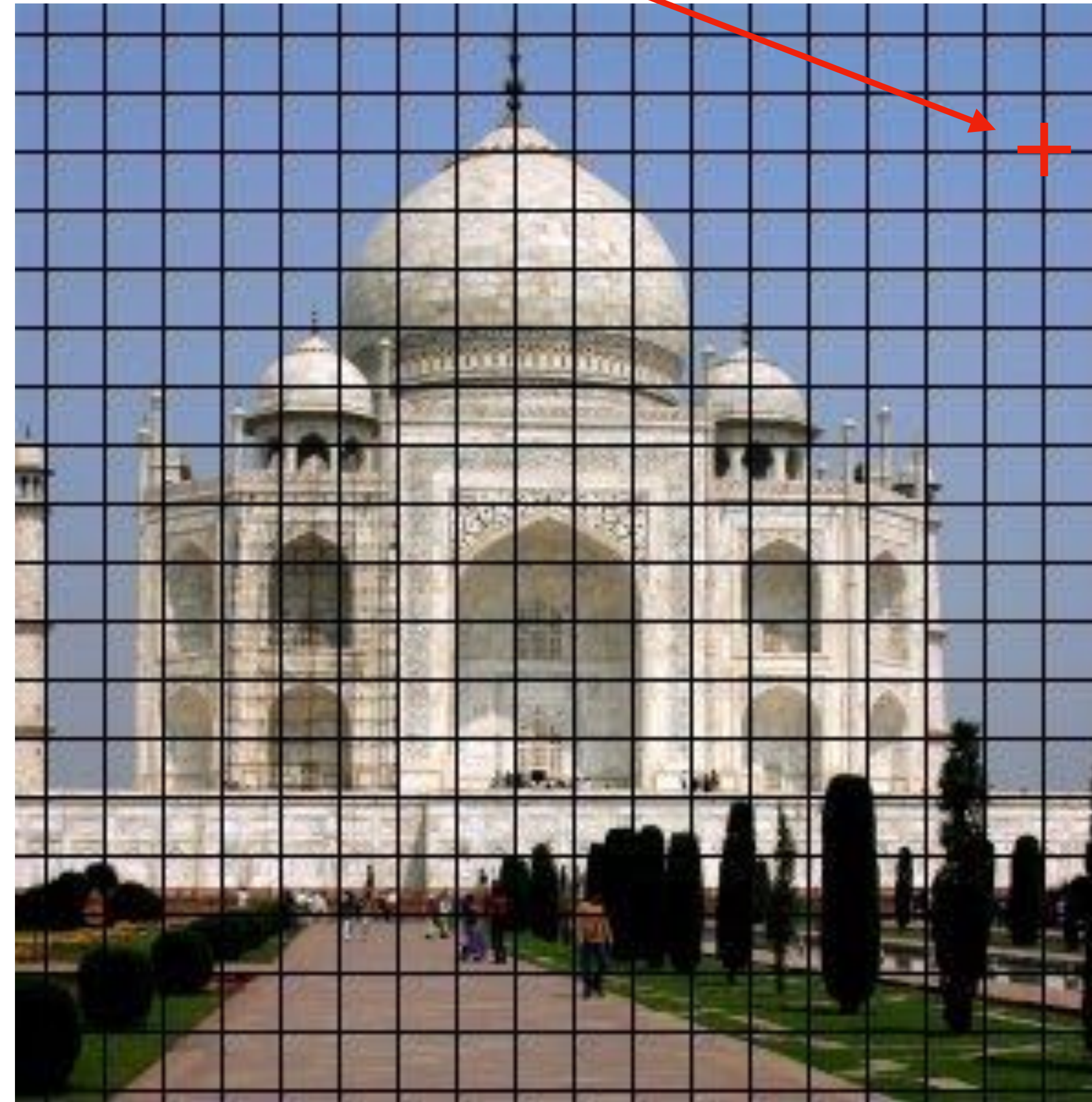
$$f : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$$

Mathematical structure

For each layer we have a **feature map**:

no. of channels

$$f : \mathbb{Z}^2 \rightarrow \mathbb{R}^{\textcircled{K}}$$



(p, q)

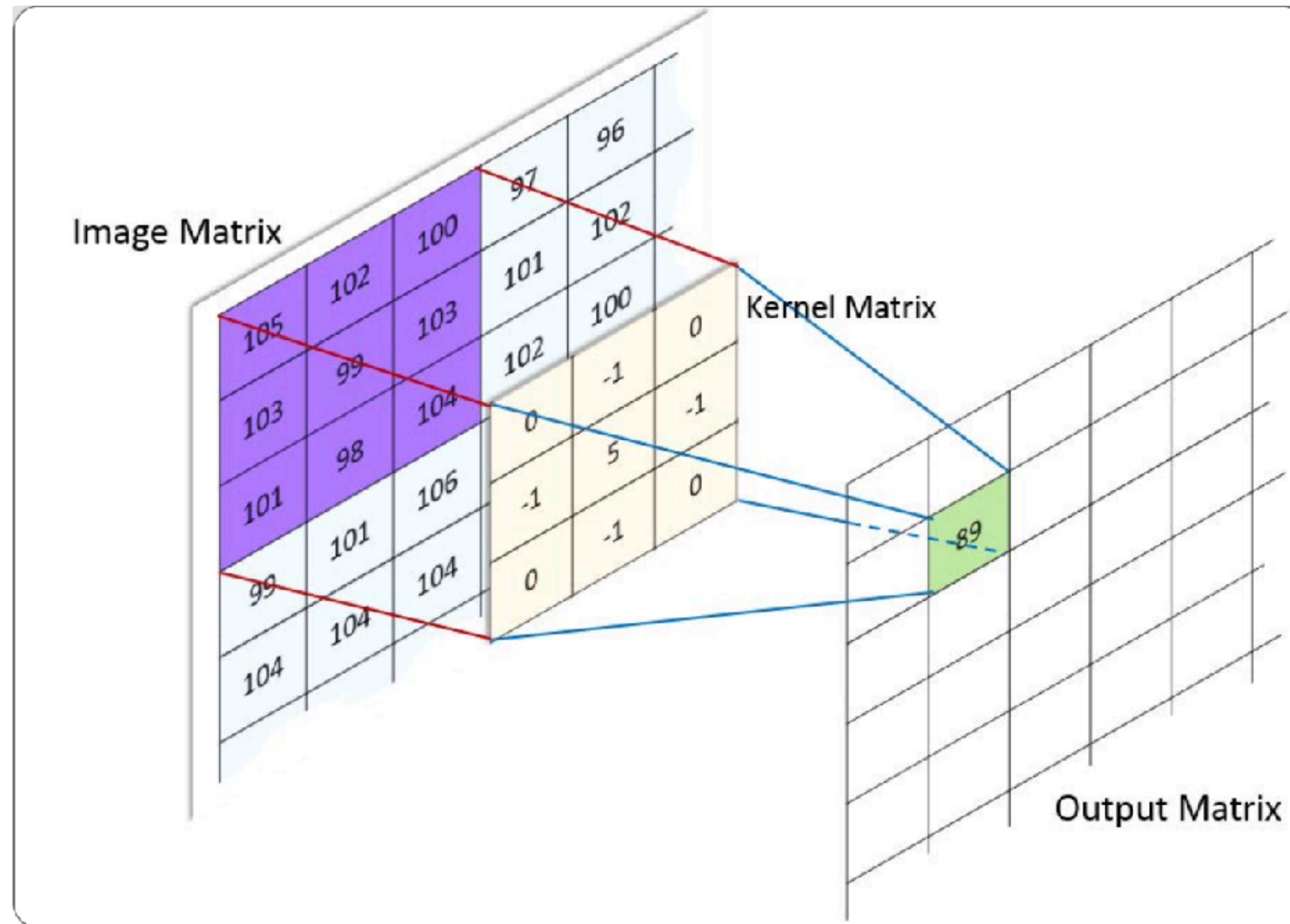
pixel coordinate

Kernel (filter): $\psi : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$

Convolution: $[f * \psi](x) = \sum_{y \in \mathbb{Z}^2} \sum_{k=1}^K f_k(y) \psi_k(x - y)$

Kernel (filter): $\psi : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$

Convolution: $[f * \psi](x) = \sum_{y \in \mathbb{Z}^2} \sum_{k=1}^K f_k(y) \psi_k(x - y)$



[Figure from machinelearninguru.com]

Kernel (filter): $\psi : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$

Convolution: $[f * \psi](x) = \sum_{y \in \mathbb{Z}^2} \sum_{k=1}^K f_k(y) \psi_k(x - y)$

Translation map: $[T(t)f](x) = f(x + t)$

Kernel (filter): $\psi : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$

Convolution: $[f * \psi](x) = \sum_{y \in \mathbb{Z}^2} \sum_{k=1}^K f_k(y) \psi_k(x - y)$

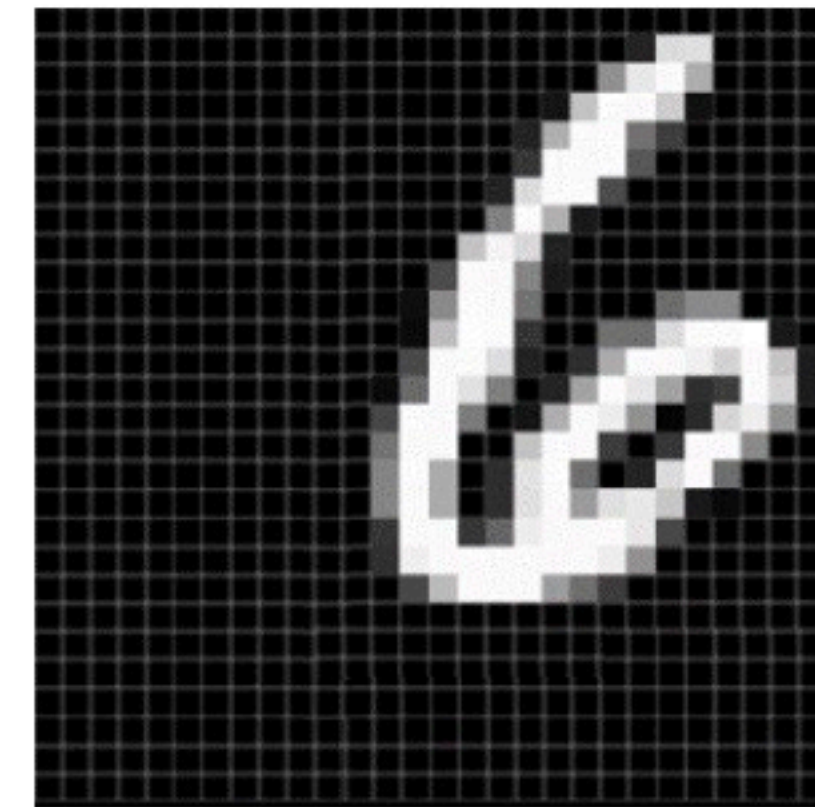
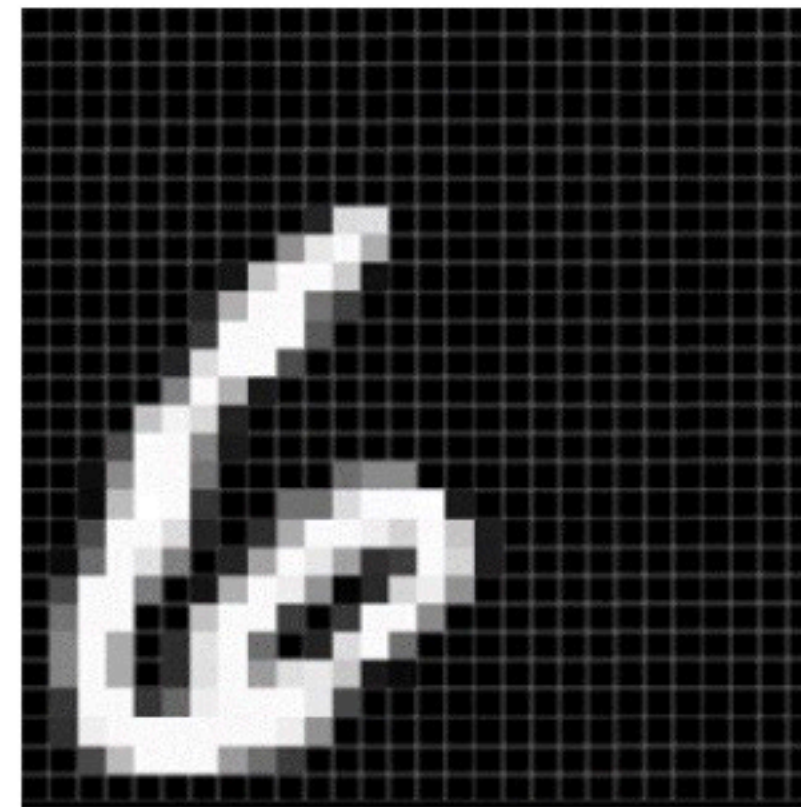
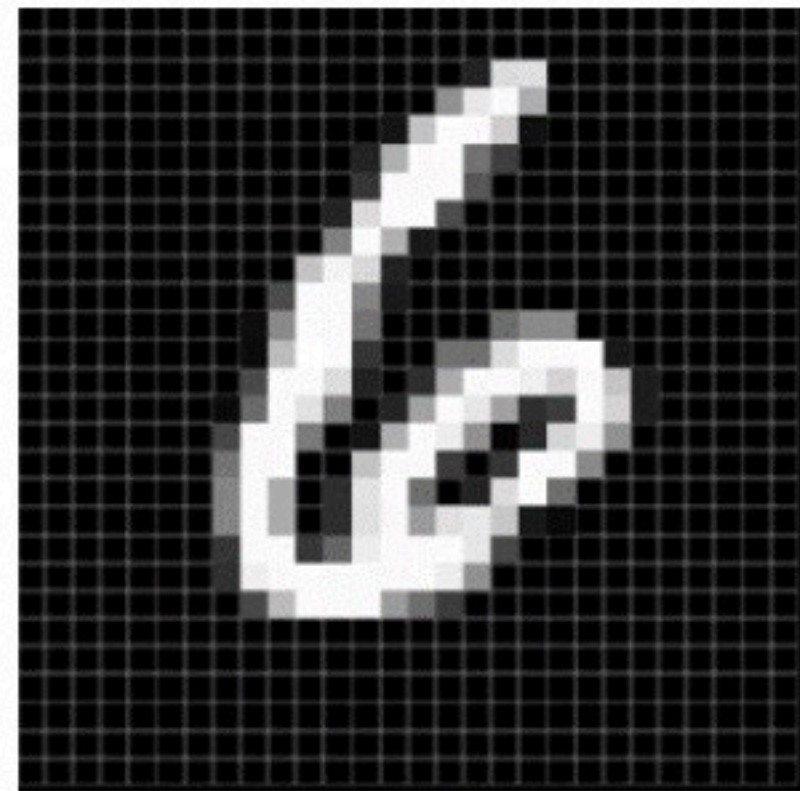
Translation map: $[T(t)f](x) = f(x + t)$

Convolution is equivariant w.r.t. translations

$$[T(t)f] * \psi = T(t)[f * \psi]$$

Convolution is **equivariant w.r.t. translations**

$$[T(t)f] * \psi = T(t)[f * \psi]$$



But what about more general symmetries?

Locally, we can think of feature maps as functions $f : G/H \rightarrow V$

Example: $G = \mathbb{Z}^2$ $H = \{1\}$ $V = \mathbb{R}^K$

Feature maps are sections! $f : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$

Locally, we can think of feature maps as functions $f : G/H \rightarrow V$

Example: $G = \mathbb{Z}^2$ $H = \{1\}$ $V = \mathbb{R}^K$

Feature maps are sections! $f : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$

This is a **vector bundle**:

$$\begin{array}{ccc} V & \longrightarrow & P \\ & & \downarrow p \\ & & G/H \end{array}$$

Locally, we can think of feature maps as functions $f : G/H \rightarrow V$

Example: $G = \mathbb{Z}^2$ $H = \{1\}$ $V = \mathbb{R}^K$

Feature maps are sections! $f : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$

General structure of group equivariant CNNs:

$$\left\{ \begin{array}{l} \text{Feature space of a} \\ G - \text{equivariant CNN} \end{array} \right\} \cong \left\{ \begin{array}{l} \text{Sections of vector} \\ \text{bundles } P \rightarrow G/H \end{array} \right\}$$

Layers defined with group-equivariant convolutions:

$$[f * \psi](g) = \int_G \sum_{k=1}^K f_k(h) \psi_k(gh) dh$$

$$\left\{ \begin{array}{l} \text{Feature space of a} \\ G\text{-equivariant CNN} \end{array} \right\} \cong \left\{ \begin{array}{l} \text{Sections of vector} \\ \text{bundles } P \rightarrow G/H \end{array} \right\}$$

Sections of $P \rightarrow G/H$ belong to the **induced representation**:

$$\mathcal{F} = \text{Ind}_H^G \rho = \{f : G \rightarrow V \mid f(gh) = \rho(h^{-1})f(g)\}$$

$$\left\{ \begin{array}{l} \text{Feature space of a} \\ G - \text{equivariant CNN} \end{array} \right\} \cong \left\{ \begin{array}{l} \text{Sections of vector} \\ \text{bundles } P \rightarrow G/H \end{array} \right\}$$

Sections of $P \rightarrow G/H$ belong to the **induced representation**:

$$\mathcal{F} = \text{Ind}_H^G \rho = \{f : G \rightarrow V \mid f(gh) = \rho(h^{-1})f(g)\}$$

$$\left\{ \begin{array}{l} \text{Maps between layers} \\ \text{in the CNN} \end{array} \right\} \cong \left\{ \begin{array}{l} G - \text{equivariant linear maps} \\ \text{between feature spaces } \mathcal{F} \rightarrow \mathcal{F}' \end{array} \right\}$$

$$\left\{ \begin{array}{l} \text{Feature space of a} \\ G\text{-equivariant CNN} \end{array} \right\} \cong \left\{ \begin{array}{l} \text{Sections of vector} \\ \text{bundles } P \rightarrow G/H \end{array} \right\}$$

Sections of $P \rightarrow G/H$ belong to the **induced representation**:

$$\mathcal{F} = \text{Ind}_H^G \rho = \{f : G \rightarrow V \mid f(gh) = \rho(h^{-1})f(g)\}$$

$$\left\{ \begin{array}{l} \text{Maps between layers} \\ \text{in the CNN} \end{array} \right\} \cong \left\{ \begin{array}{l} G\text{-equivariant linear maps} \\ \text{between feature spaces } \mathcal{F} \rightarrow \mathcal{F}' \end{array} \right\}$$

$$\text{Hom}_G(\mathcal{F}, \mathcal{F}')$$

(intertwining operators)

Equivariant non-linear maps for neural networks on homogeneous spaces

Elias Nyholm, Oscar Carlsson, Maurice Weiler, **D.P.**

[arXiv:2504.20974]

A mathematical formalism for non-linear neural networks!

Equivariant non-linear maps for neural networks on homogeneous spaces

Elias Nyholm, Oscar Carlsson, Maurice Weiler, **D.P.**

[arXiv:2504.20974]

A mathematical formalism for non-linear neural networks!

Linear case:

$$\Phi \xrightarrow{\text{(Homogeneity)}} [\Phi f](g) \xrightarrow{\text{(Linearity)}} [\Phi f](g) = \int_G \kappa(g, g') f(g') dg' \xrightarrow{\text{(Equivariance)}} [\Phi f](g) = \int_G \tilde{\kappa}(g^{-1}g') f(g') dg'$$

Equivariant non-linear maps for neural networks on homogeneous spaces

Elias Nyholm, Oscar Carlsson, Maurice Weiler, **D.P.**

[arXiv:2504.20974]

A mathematical formalism for non-linear neural networks!

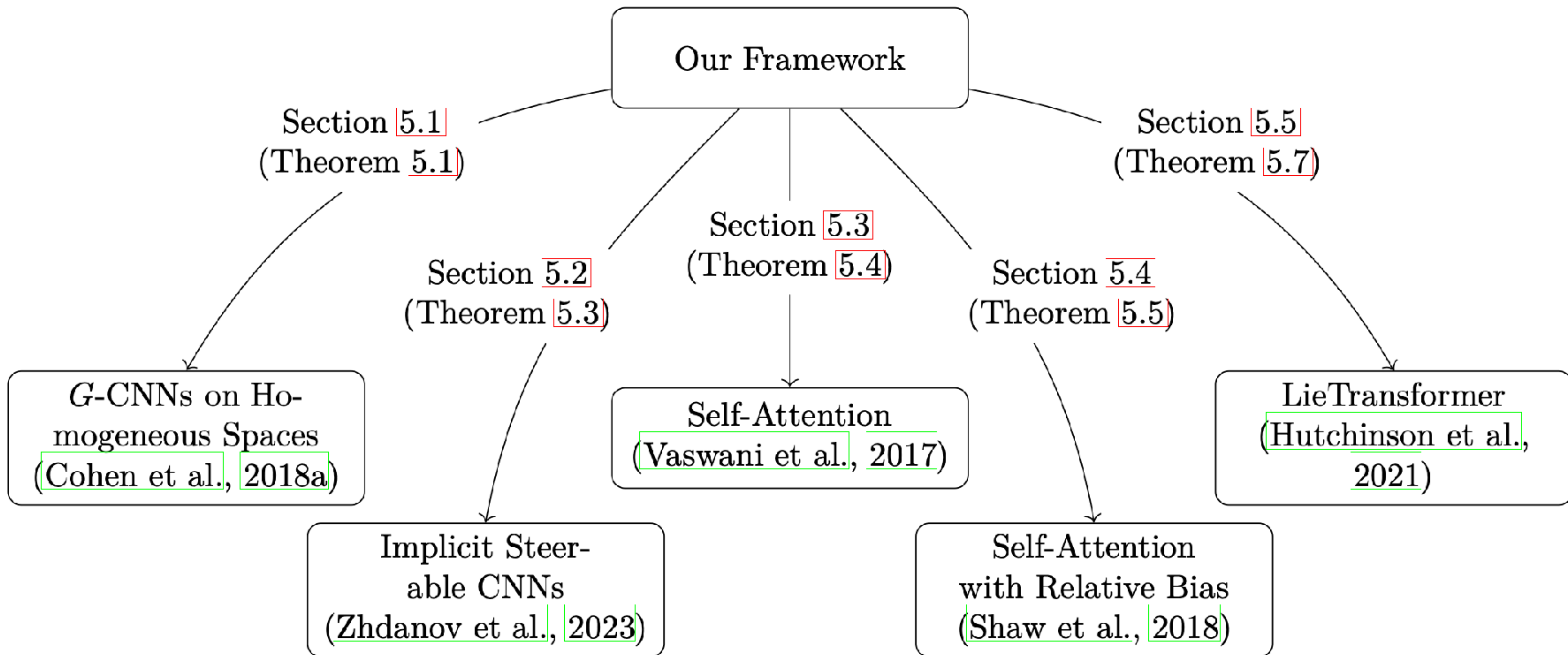
Linear case:

$$\Phi \xrightarrow{\text{(Homogeneity)}} [\Phi f](g) \xrightarrow{\text{(Linearity)}} [\Phi f](g) = \int_G \kappa(g, g') f(g') dg' \xrightarrow{\text{(Equivariance)}} [\Phi f](g) = \int_G \tilde{\kappa}(g^{-1}g') f(g') dg'$$

Non-linear generalization:

$$\Phi \xrightarrow{\text{(Homogeneity)}} [\Phi f](g) \xrightarrow{\text{(Equivariance)}} [\Phi f](g) = \tilde{\Phi}(g^{-1} \triangleright f)$$

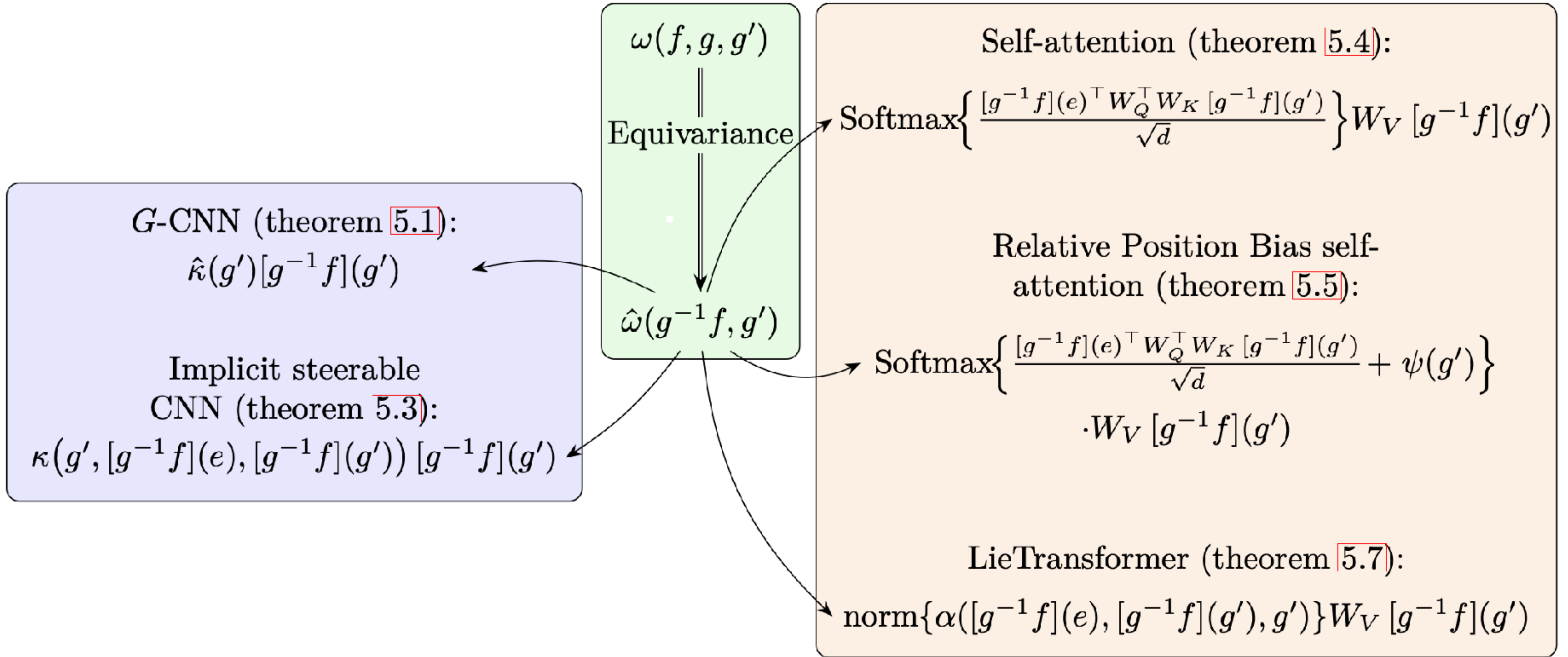
[Slide borrowed from my PhD student Elias!]



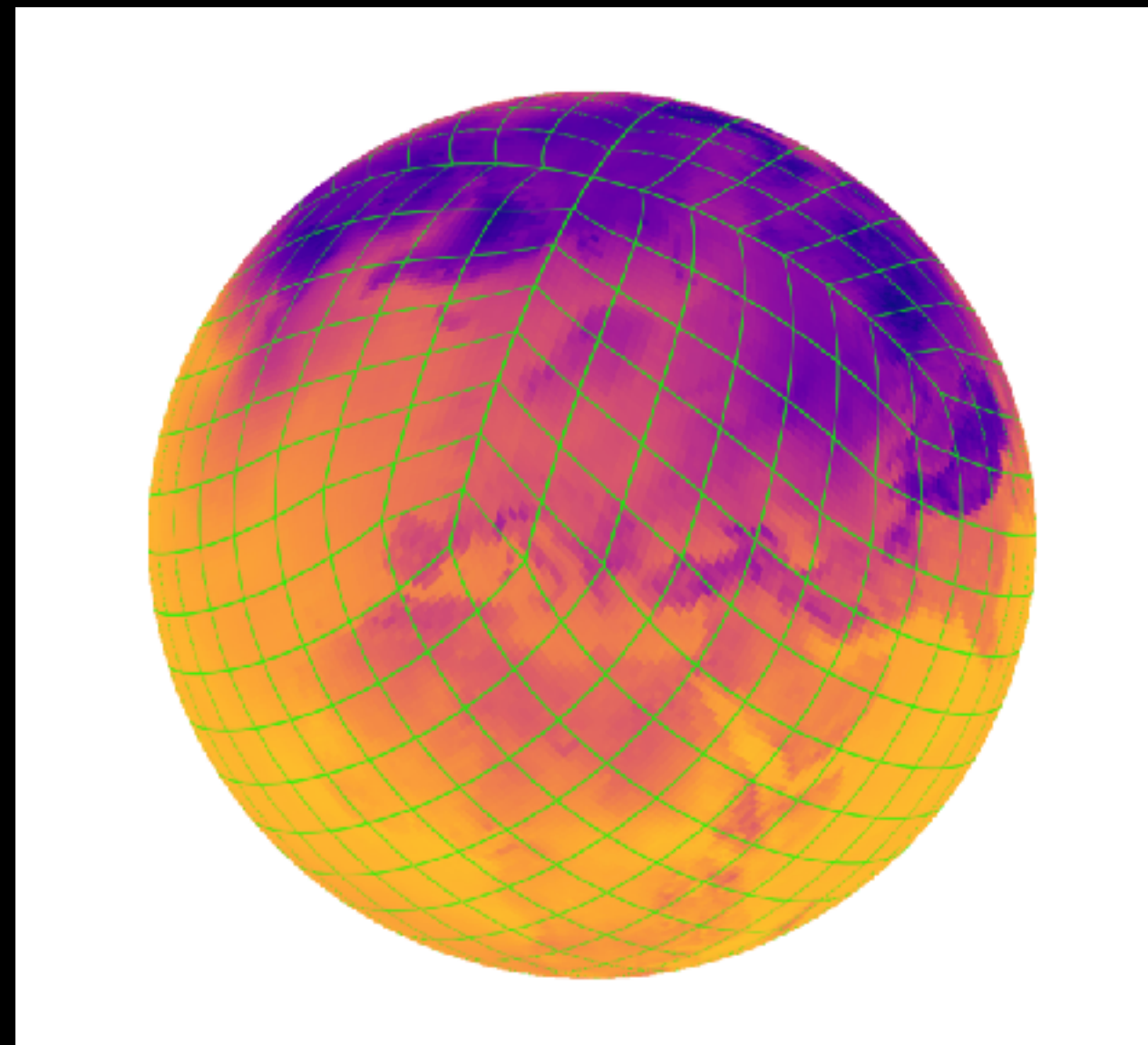
Convolution based

General framework

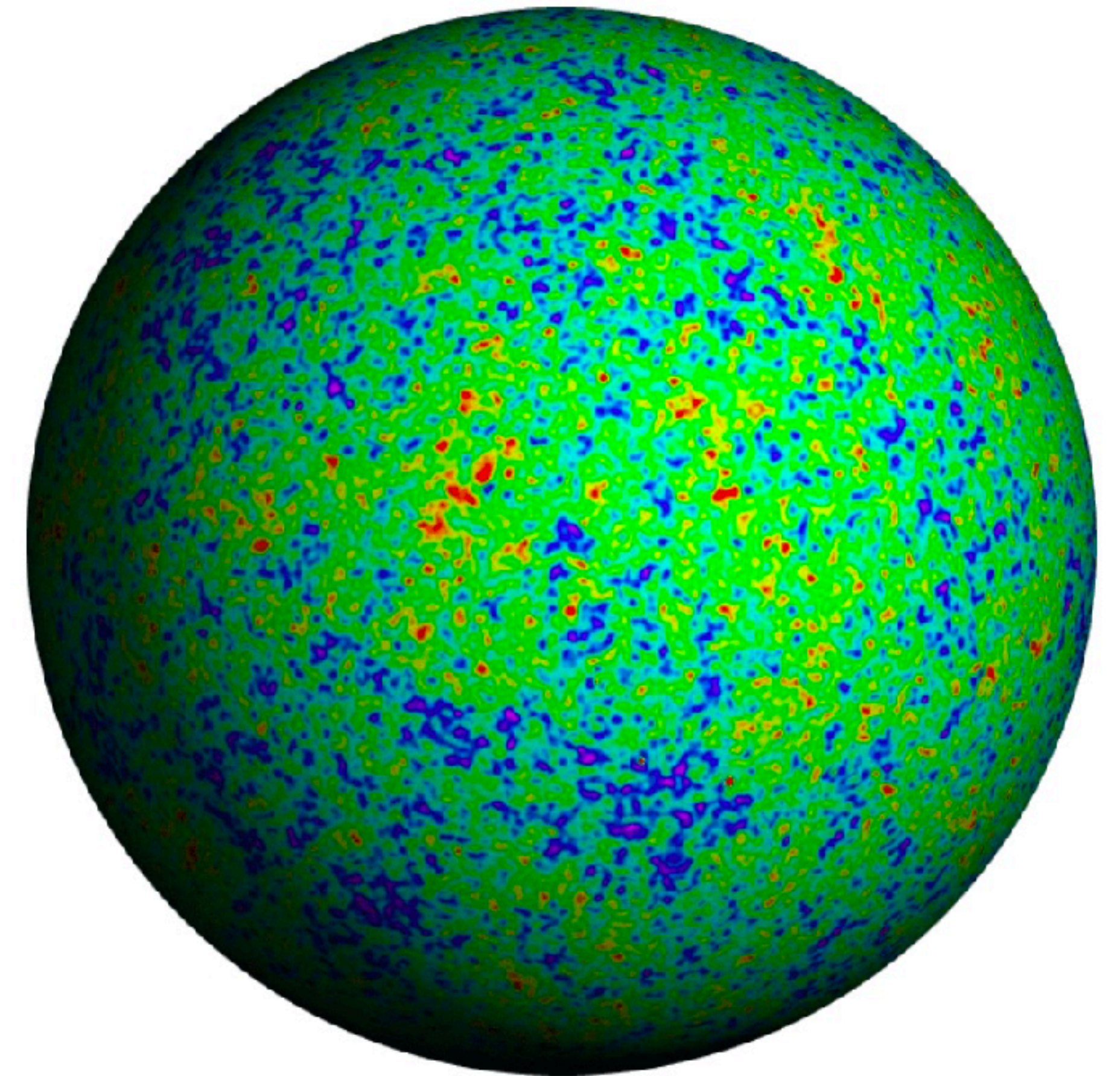
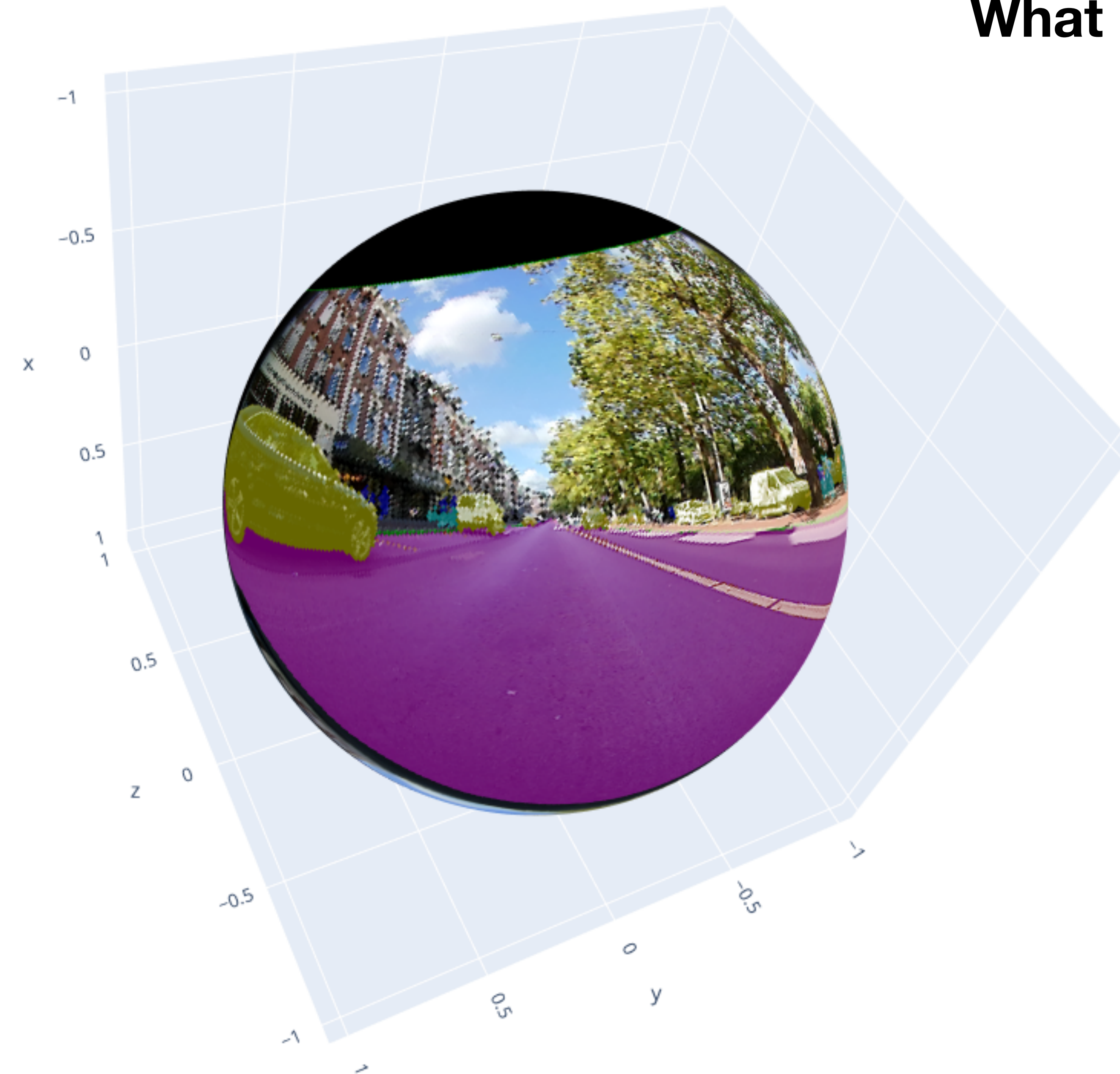
Attention based



Applications: AI4Science



What if the input data is curved?



Cosmic microwave background radiation

[Image from the Woodscape dataset, projected onto a sphere]

Equivariance versus augmentation for spherical signals

By Gerken, Carlsson, Linander, Ohlsson, Petersson, **D.P.**

[ICML 2022, arXiv: 2202.03990]

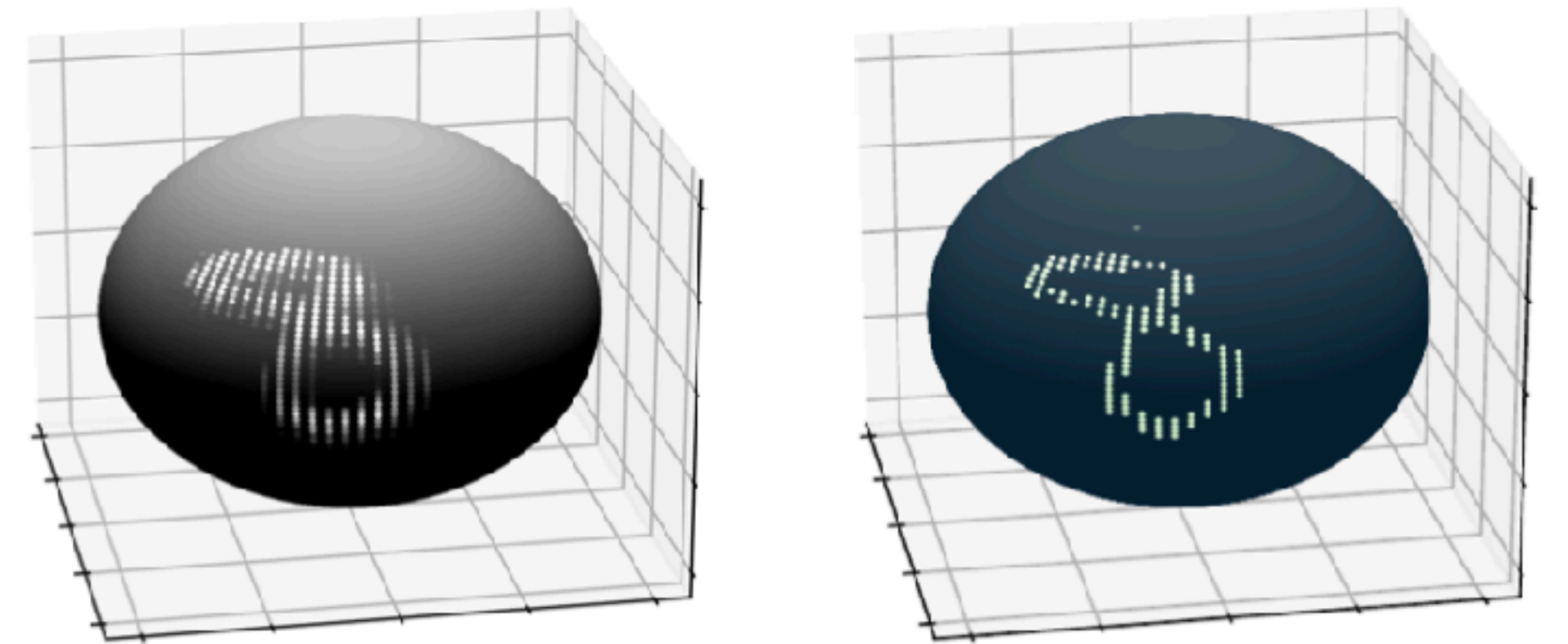
$$G = SO(3)$$

$$H = SO(2)$$

$$G/H \cong S^2$$

**Feature
maps**

$$f : S^2 \rightarrow \mathbb{R}^K$$



- We demonstrate that non-equivariant classification models require considerable data augmentation to reach the performance of smaller equivariant networks
- We show that the performance of non-equivariant semantic segmentation models saturates well below that of equivariant models as the amount of data augmentation is increased

$$(\kappa \star f)(R) = \int_{S^2} \kappa(R^{-1}x) f(x) dx$$

$$(\kappa \star f)(R) = \int_{SO(3)} \kappa(S^{-1}R) f(S) dS$$

Oscar Carlsson^{*ab}

Jan E. Gerken^{*a}

Hampus Linander^a

Heiner Spieß^c

Fredrik Ohlsson^d

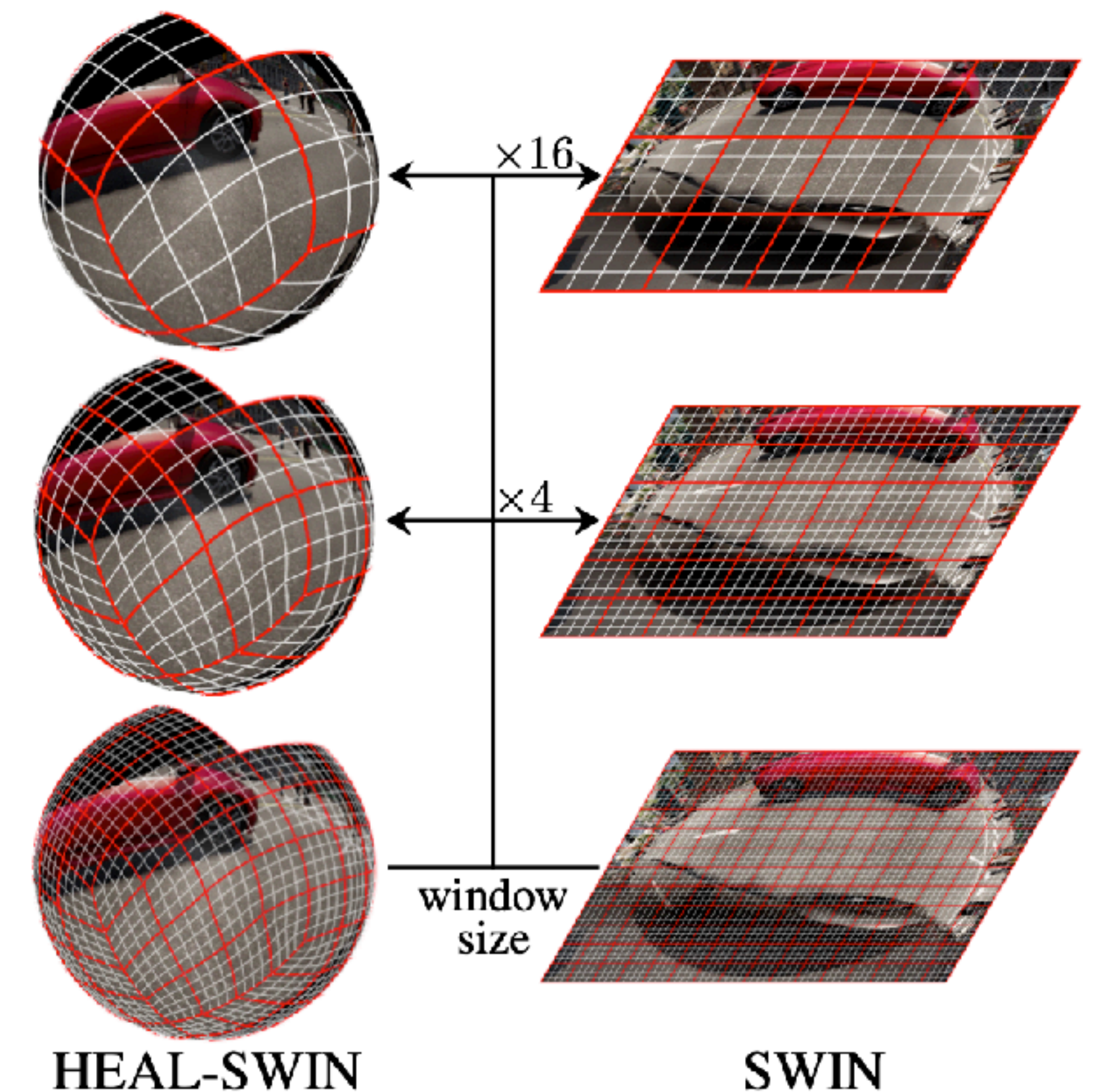
Christoffer Petersson^{ea}

Daniel Persson^a

- We combine an adapted vision transformer with the **Hierarchical Equal Area iso-Latitude Pixelisation** (HEALPix) grid

- HEALPix was developed for capturing the high-resolution measurements of the CMB performed by the WMAP/PLANCK

- Demonstrate superior performance for semantic segmentation of traffic fisheye images, both for synthetic and real automotive datasets

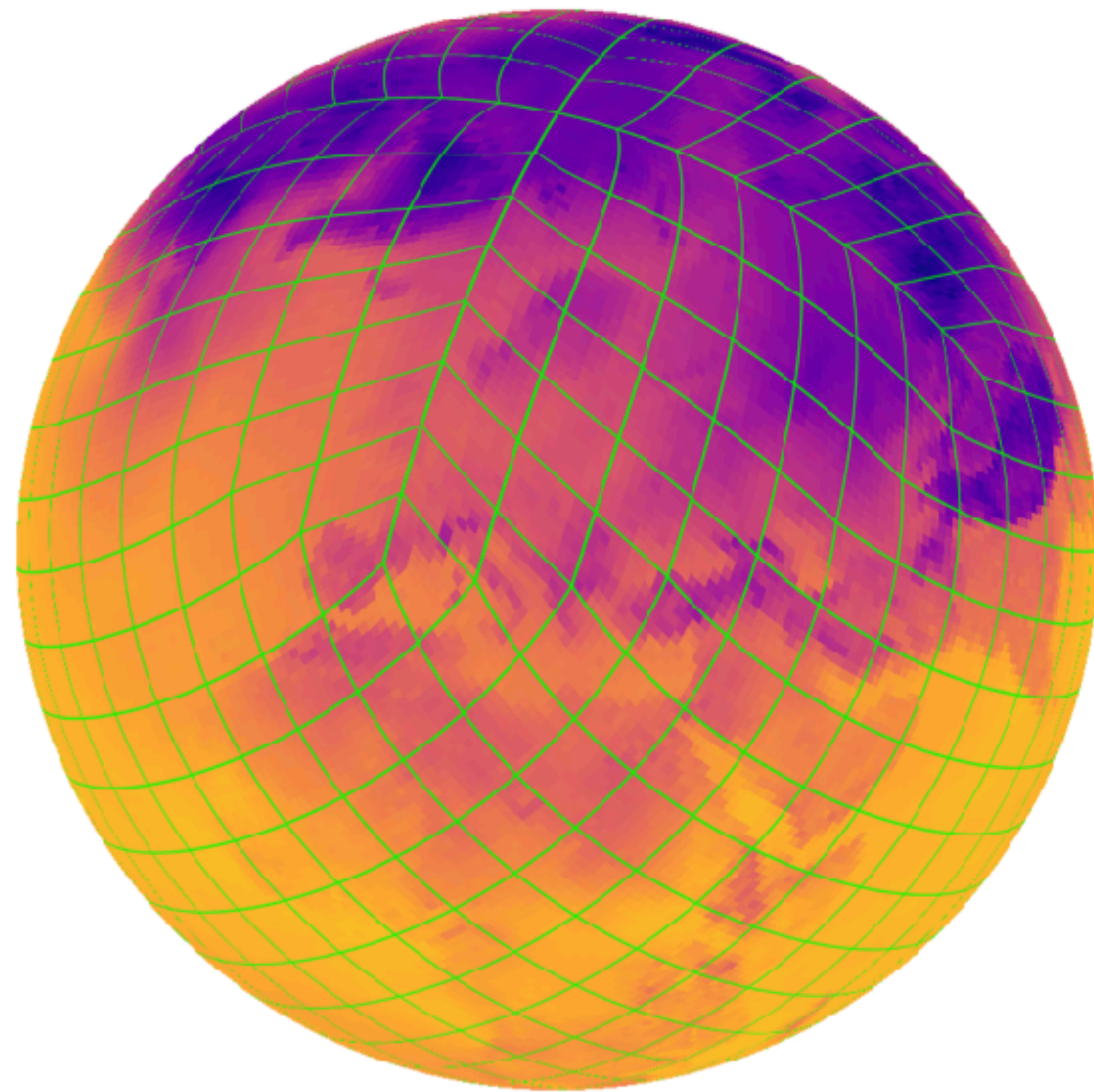


PEAR: Equal Area Weather Forecasting on the Sphere

Hampus Linander^{1,2} Tage Tykesson^{2,4} Pietro Rosso³
Christoffer Petersson^{2,5} Daniel Persson² Jan E. Gerken²

¹ VERSES AI ² Department of Mathematical Sciences
Los Angeles, USA Chalmers University of Technology
University of Gothenburg, Sweden

³ Department of Computer Science and Engineering ⁵ Recohere
⁴ Department of Physics and Astronomy Gothenburg, Sweden
Chalmers University of Technology, Sweden



Predicted surface level temperature

Current AI weather models still work on an **equiangular spherical grid** (Driscoll–Healy), which over-resolves the poles relative to the equator.
Why not forecast directly on an equal-area discretization of the sphere?

PEAR (“Pangu Equal ARea”), is a **transformer-based global weather forecasting model** that operates **natively on HEALPix features**

PEAR also **outperforms Pangu-Large** for forecast horizons of **5 days and beyond**

Learning Chern Numbers of Multiband Topological Insulators with Gauge Equivariant Neural Networks



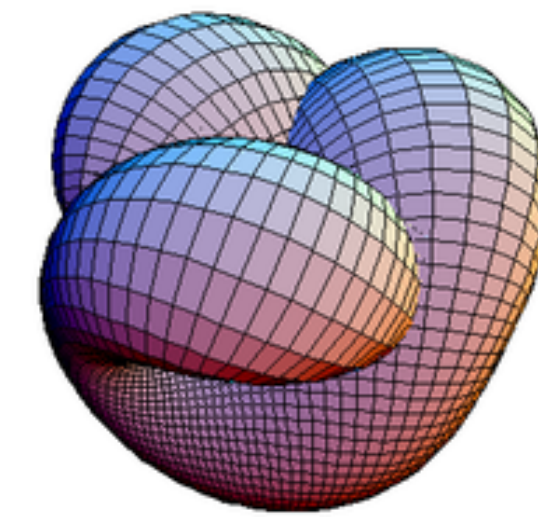
1

Longde Huang, Oleksandr Balabanov, Hampus Linander, Mats Granath
Daniel Persson, Jan E. Gerken

- ➔ **Topological Insulators:** Materials that behave as insulators in their bulk but allow current to flow along their boundaries, due to underlying topological invariants (Chern numbers)
- ➔ **Describe using lattice gauge theory:** Underlying $U(N)$ local symmetry at each site
- ➔ We extend the **gauge equivariant networks** of Favoni et al and apply them to the problem of predicting Chern numbers of multi-band systems

Conclusions and speculations

Gauge equivariant neural networks



[Cheng, Anagiannis, Weiler, de Haan, Cohen, Welling]

[Gerken, Carlsson, Aronsson, Linander, Ohlsson, Petersson, **D.P.**]

Neural networks on arbitrary manifolds
require local equivariance

covariance w. r. t.
gauge transformations
(general coordinate transformations)

gauge equivariant
feature maps

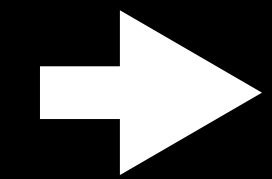
Fields
Sections of vector bundles
(frame bundles)

“elementary feature types”
?

irreducible representations of G
elementary particles
(scalars, vectors, spinors...)

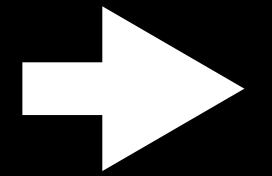
Are these the seeds of a deeper relation between neural networks and gauge theory?

Future work

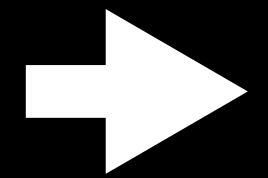


Explore connections with neuroalgebraic geometry

Joint with Magdalena & Yacoub (arXiv:2603.29566)

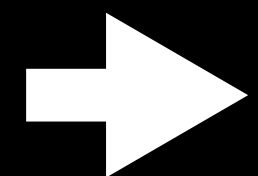


Applications to climate data



Equivariance in weather & climate data?

Joint with Linander, Gerken, Tykesson, Rosso & Petersson
(arXiv:2505.17720)



Geometric deep learning for medical images

Collaboration with Sahlgrenska University Hospital

Joint with Gerken, Bramberger & Hallgren

A vibrant, multi-colored spiral galaxy with a central black hole, surrounded by a complex network of stars and filaments. The galaxy's colors transition from blue and purple in the center to orange and red towards the edges. The surrounding space is filled with a dense network of white and yellow stars connected by thin, glowing lines, creating a web-like structure. The overall scene is set against a dark, starry background.

TACK!

Topological Insulators & the Chern Number

Topological insulators

- ▶ **Bulk insulating**, robust conducting edge states.
- ▶ Topologically protected against deformations/disorder.
- ▶ Applications: spintronics, topological photonics, quantum computing.

The Chern number

- ▶ Integer invariant $C \in \mathbb{Z}$ classifying band topology.
- ▶ Quantised Hall conductance:
 $\sigma_{xy} = C e^2/h$.

Why is prediction hard?

- ▶ C computed from **Berry curvature** — requires a **gauge choice** for Bloch states.
- ▶ For N filled bands: *local* $U(N)$ rotation at every k -point.
- ▶ Total symmetry group: $U(N)^{N_{\text{site}}}$ — *exponentially* large.

The challenge

Standard NNs cannot handle this local symmetry.

Previous ML limited to $N=1$ (abelian $U(1)$).

Learning Chern Numbers of Topological Insulators

The problem

Multiband topological insulators with N filled bands.

Chern number (topological invariant):

$$C = \frac{1}{2\pi i} \int_{\text{BZ}} \text{Tr}[\mathcal{F}(k)] d^2k \in \mathbb{Z}$$

Discretised on a lattice via **Wilson loops** $W_k \in \text{U}(N)$ built from link variables (overlaps of Bloch states).

Gauge symmetry (local, at every site):

$$W_k \longrightarrow \Omega_k^\dagger W_k \Omega_k, \quad \Omega_k \in \text{U}(N)$$

Why gauge equivariance?

Total symmetry group: $\text{U}(N)^{N_{\text{site}}}$

\Rightarrow *exponentially* larger than any global symmetry.

Standard CNNs fail for $N > 1$ band.

Key idea

Adapt **Lattice Gauge Equivariant CNNs** (Favoni et al., from lattice QCD) to condensed matter:

$$f(\Omega^\dagger W \Omega) = f(W) \quad \text{by construction.}$$

Architecture & Key Results

GEBLNet — purely local layers

- ▶ **GEBL** (bilinear):
 $W'^{\gamma} = \alpha_{\mu\nu}^{\gamma} W^{\mu} W^{\nu}$
- ▶ **GEAct** (nonlinearity): $\sigma(\text{Tr } W) \cdot W$
- ▶ **TrNorm** (new): normalise by $|\langle \text{Tr } W \rangle|$
- ▶ **Trace** \rightarrow gauge-invariant scalars $\rightarrow \sum_k$

Universal approximation theorem

Class functions on $U(N)$ are spanned by symmetric polynomials in eigenvalues (Peter–Weyl). GEBLNet generates $\text{Tr } g, \text{Tr } g^2, \dots, \text{Tr } g^M$
 \Rightarrow can approximate *any* gauge invariant.

Key results

- ▶ **Up to $N = 7$ bands** — previous methods: $N = 1$ only. $\sim 96\%$ accuracy ($N=4$, 5×5 grid).
- ▶ **Generalises from $C = 0$** — trained on trivial topology, predicts non-trivial ($\sim 94\%$).
- ▶ **Scales to large lattices** — local architecture; works on unseen grid sizes.
- ▶ **4D Chern numbers** — learns C_2 on 4D Brillouin zones (MAE ≈ 0.25).

