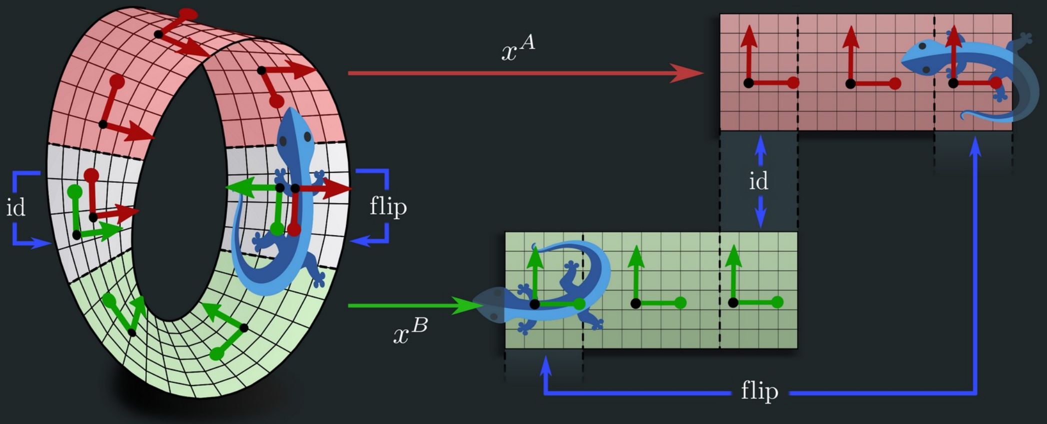
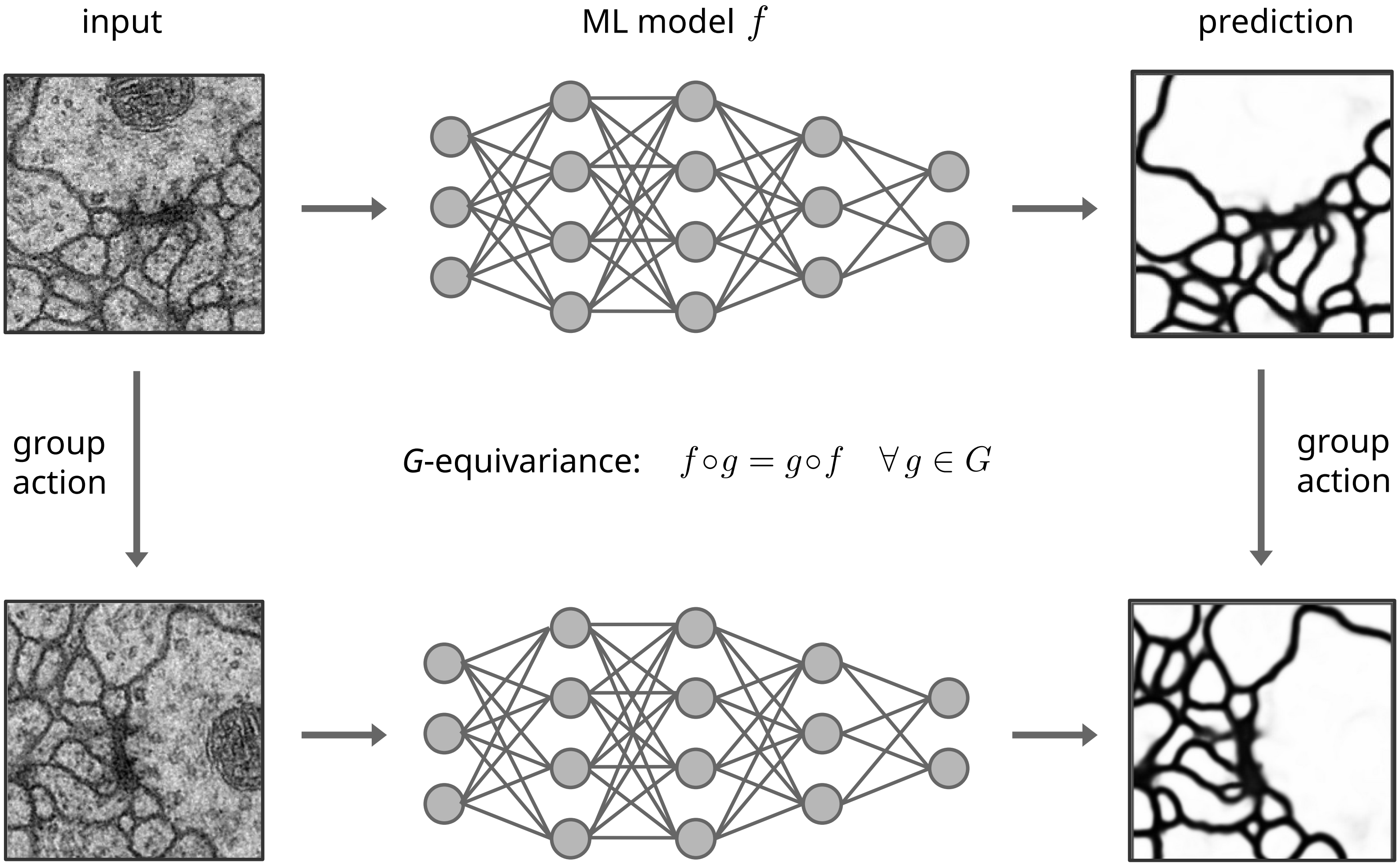
Equivariant & Coordinate Independent Convolutional Neural Networks

Maurice Weiler
AMLab, QUVA Lab
University of Amsterdam
@maurice_weiler



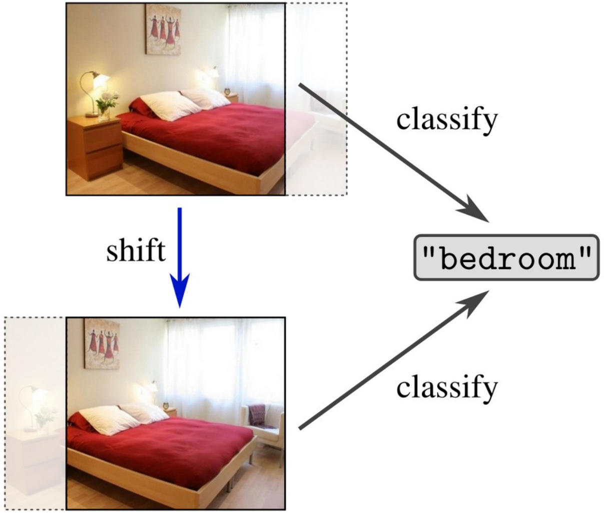
QUVA
Deep Vision Lab



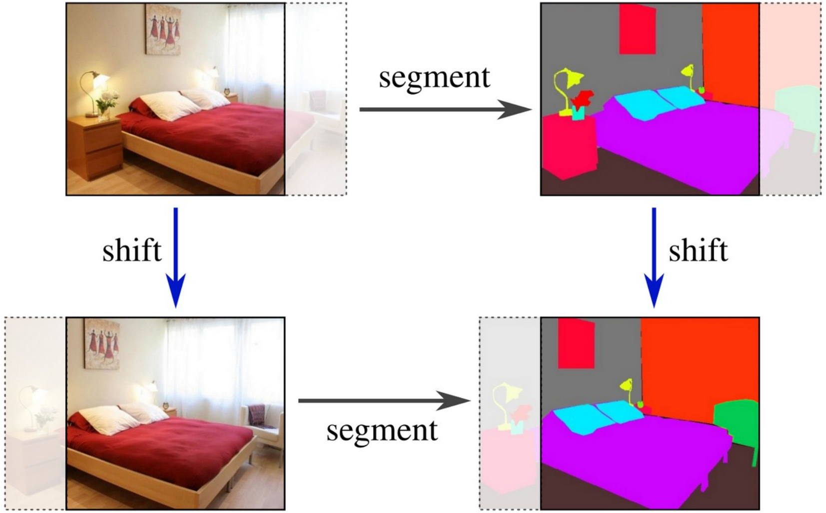
G-equivariance: $f \circ g = g \circ f \quad \forall\, g \in G$

convolutional neural networks are **translation invariant / equivariant**

invariant image classification



equivariant image segmentation



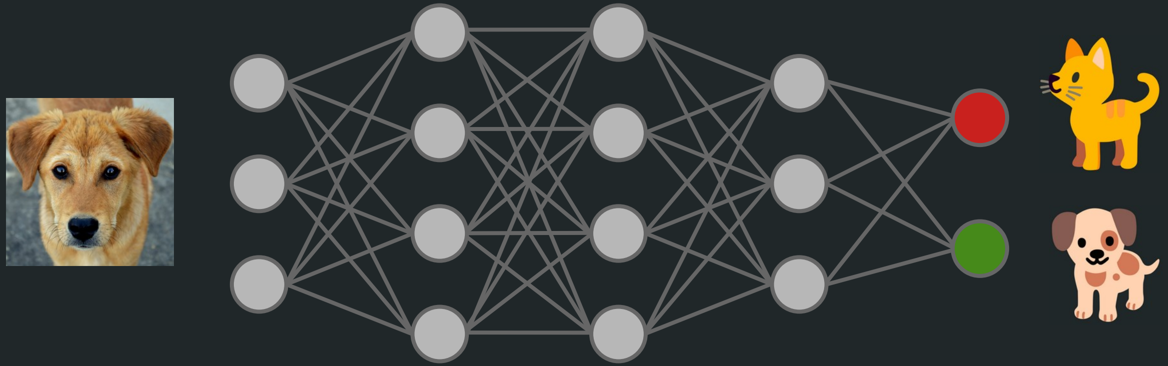
research goals: generalize equivariant convolutions to… …larger symmetry groups (of Euclidean spaces)

…more general manifolds

Outline

MLPs for image processing?	
Translation equivariant CNNs	(Euclidean spaces)
Affine equivariant CNNs	(Euclidean spaces)
Coordinate independent CNNs	(Riemannian manifolds)

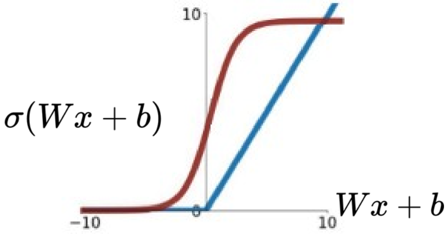
MLPs for image processing ?



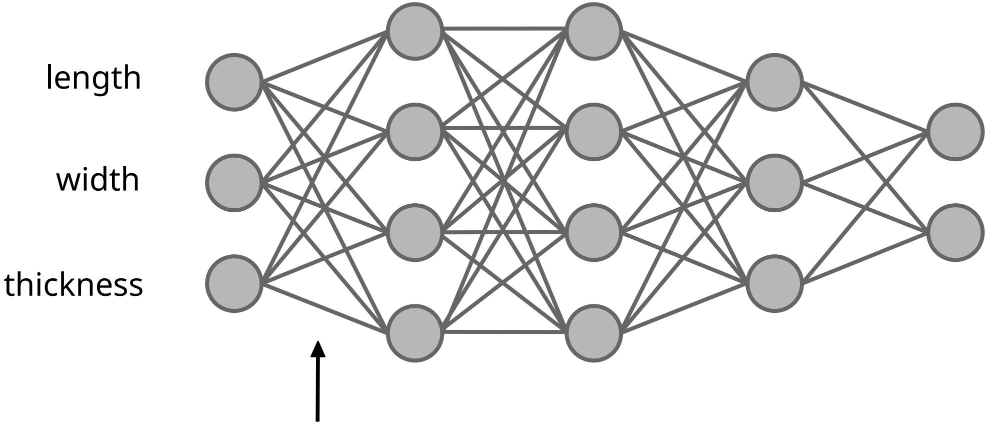
Multilayer Perceptrons (MLPs)

universal function approximators $f : \mathbb{R}^N \to \mathbb{R}^M$

composed of affine maps + nonlinearities: $x_{i+1} = \sigma(Wx_i + b)$



$$\mathbb{R}^3 \xrightarrow{f_1} \mathbb{R}^4 \xrightarrow{f_2} \mathbb{R}^4 \xrightarrow{f_3} \mathbb{R}^3 \xrightarrow{f_4} \mathbb{R}^2$$



$W \in \mathbb{R}^{4\times 3}$

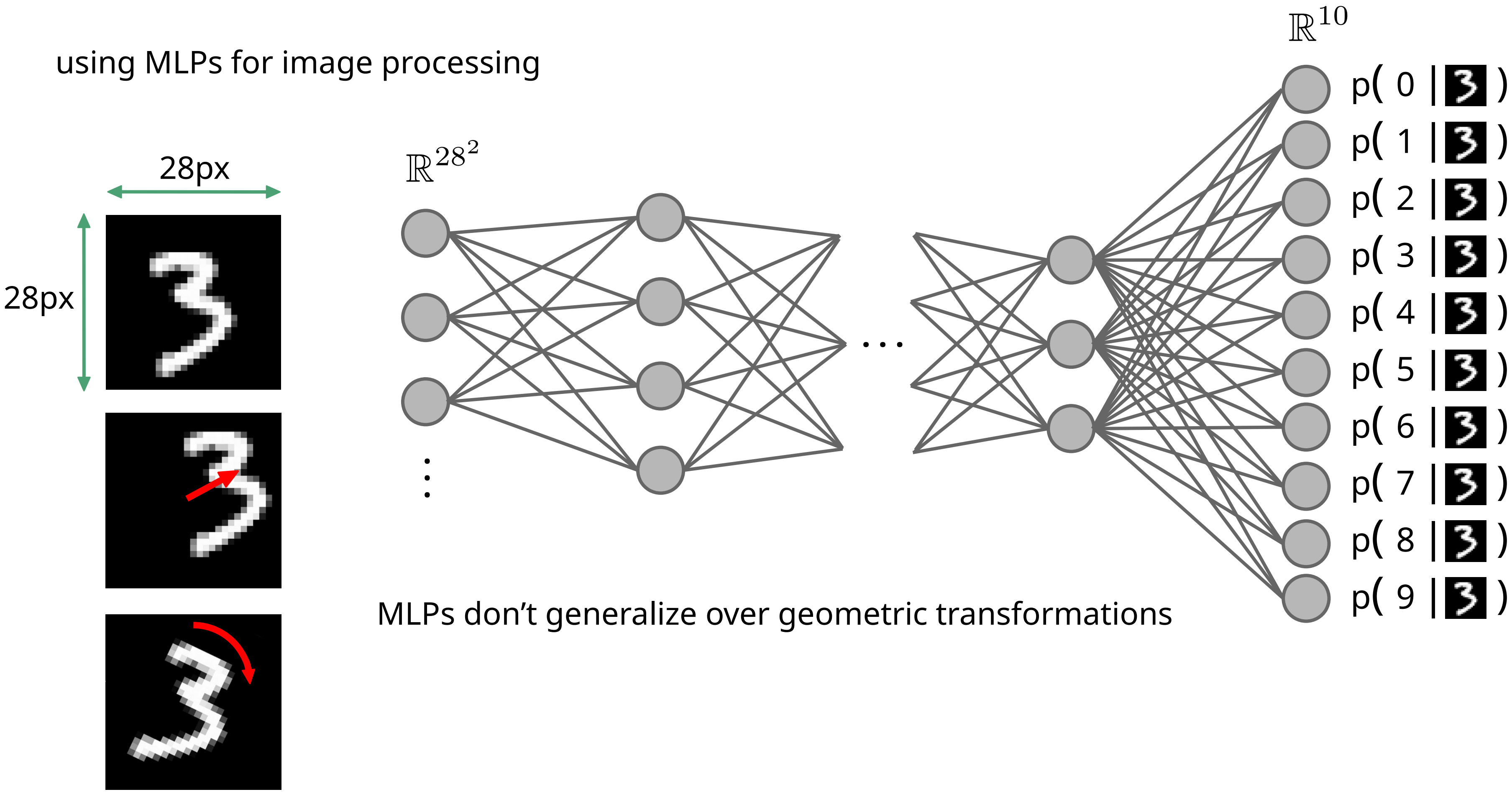
$b \in \mathbb{R}^4$

p(species A | )

p(species B | )

Multilayer Perceptrons (MLPs)

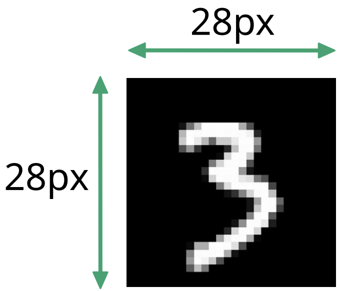
using MLPs for image processing

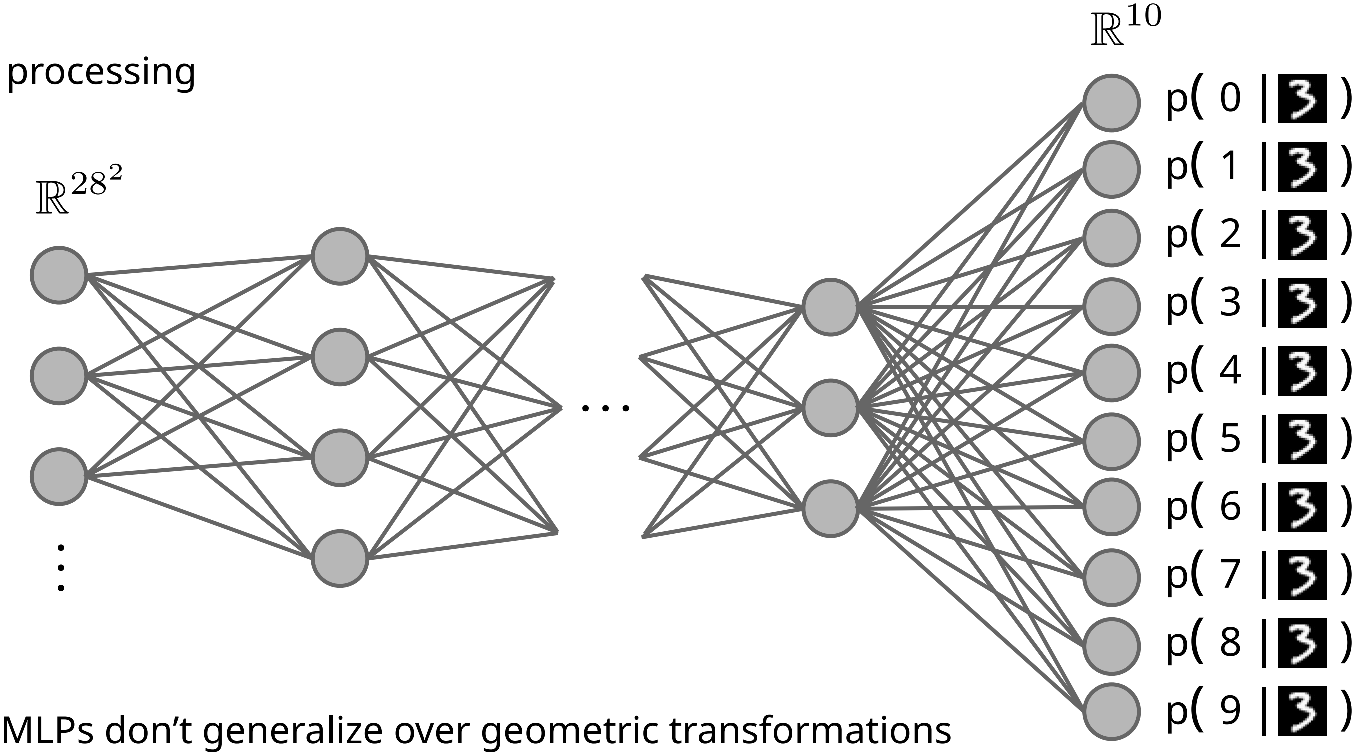


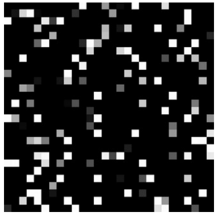
MLPs don’t generalize over geometric transformations

Multilayer Perceptrons (MLPs)

using MLPs for image processing



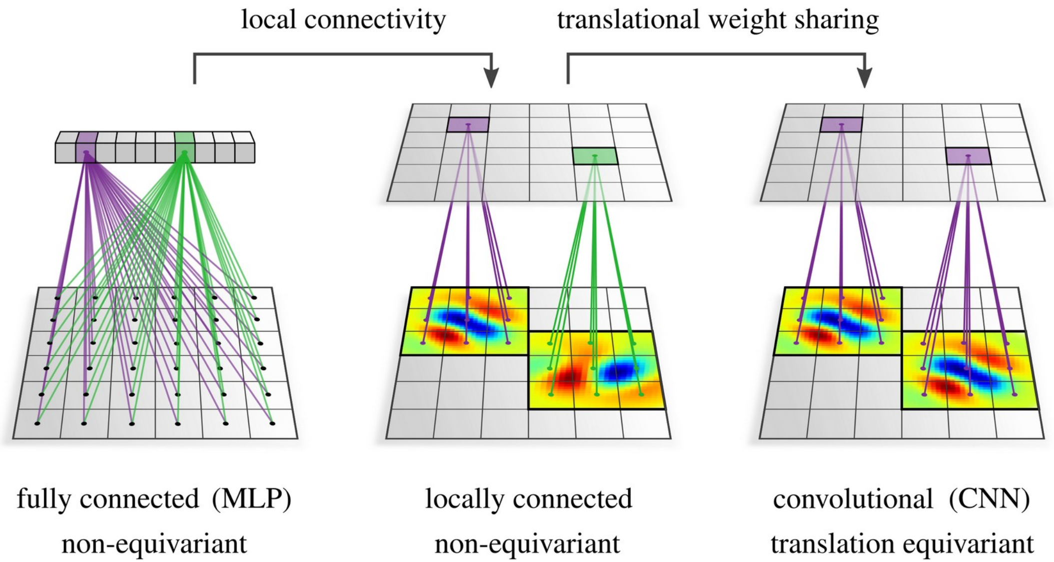




MLPs don’t generalize over geometric transformations

MLPs are ignorant of the geometric arrangement of pixels
(any permutation of pixels would be equivalent)

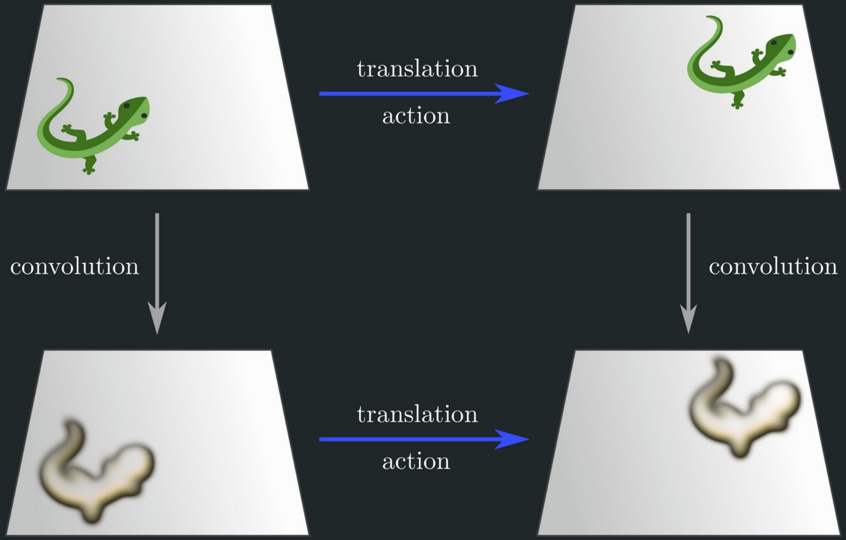
convolutional networks == MLPs + geometric inductive biases



usually: weight sharing $\Longrightarrow$ equivariance (sufficiency)

our approach: weight sharing $\Longleftarrow$ equivariance (necessity)

Translation equivariant CNNs on Euclidean spaces



Equivariant Neural Networks

(feed forward) neural networks are sequences of layers:

$$\begin{tikzcd}
\mathcal{F}_0 \arrow[r, "L_1"] & \mathcal{F}_1 \arrow[r, "L_2"] & \mathcal{F}_2 \arrow[r, "L_3"] & \dots \arrow[r, "L_{N-1}"] & \mathcal{F}_{N-1} \arrow[r, "L_N"] & \mathcal{F}_N
\end{tikzcd}$$

equivariant NNs are sequences of equivariant layers:

$$\begin{tikzcd}
\mathcal{F}_0 \arrow[r, "L_1"] \arrow[d, "g \triangleright_0"'] & \mathcal{F}_1 \arrow[r, "L_2"] \arrow[d, "g \triangleright_1"'] & \mathcal{F}_2 \arrow[r, "L_3"] \arrow[d, "g \triangleright_2"'] & \dots \arrow[r, "L_{N-1}"] & \mathcal{F}_{N-1} \arrow[r, "L_N"] \arrow[d, "g \triangleright_{N-1}"'] & \mathcal{F}_N \arrow[d, "g \triangleright_N"] \\
\mathcal{F}_0 \arrow[r, "L_1"'] & \mathcal{F}_1 \arrow[r, "L_2"'] & \mathcal{F}_2 \arrow[r, "L_3"'] & \dots \arrow[r, "L_{N-1}"'] & \mathcal{F}_{N-1} \arrow[r, "L_N"'] & \mathcal{F}_N
\end{tikzcd}$$

to design an equivariant network, we need to ...

... specify the *feature spaces* and *group actions* on them ⟶ feature maps with translation action

... design *equivariant layers,* which commute with the group actions ⟶ convolutions, bias summation, nonlinearities, etc.

Feature maps

discretized feature maps on $\mathbb{R}^d$ are implemented as “tensors” of shape $(\overbrace{X_1, \ldots, X_d}^{\text{spatial / pixel dimensions}}, \overset{\text{feature channels}}{C})$

continuous feature maps are functions $f : \mathbb{R}^d \to \mathbb{R}^c$ that assign feature vectors $f(x) \in \mathbb{R}^c$ to points $x \in \mathbb{R}^d$

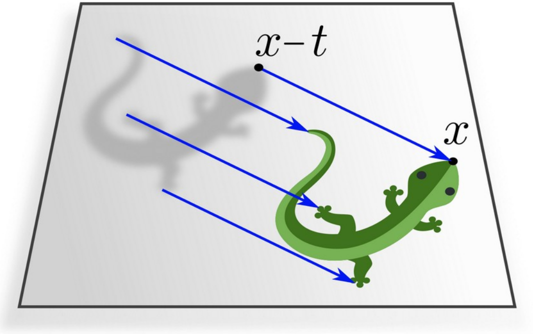
$\pitchfork$

$L^2(\mathbb{R}^d, \mathbb{R}^c)$ = **feature vector space**

linear

feature maps carry a translation **group action** $[t \triangleright f](x) := f(x - t)$

feature maps form the *regular* $(\mathbb{R}^d, +)$-representation



Translation equivariant NNs

translation equivariant networks consist of layers $\mathcal{L}: L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}) \to L^2(\mathbb{R}^d, \mathbb{R}^{c_{\text{out}}})$ that ...

... map between c_{in} and c_{out}-dimensional input and output feature maps

... commute with the group action:

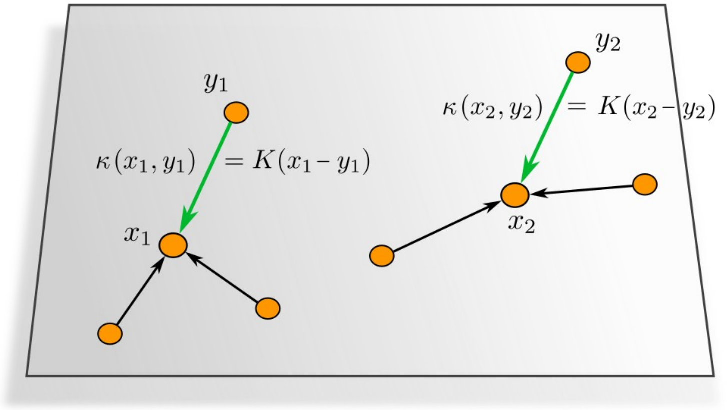
$$\begin{tikzcd}
L^2\left(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}\right) \arrow[r, "\mathcal{L}"] \arrow[d, "t \triangleright_{\text{in}}"'] & L^2\left(\mathbb{R}^d, \mathbb{R}^{c_{\text{out}}}\right) \arrow[d, "t \triangleright_{\text{out}}"] \\
L^2\left(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}\right) \arrow[r, "\mathcal{L}"'] & L^2\left(\mathbb{R}^d, \mathbb{R}^{c_{\text{out}}}\right)
\end{tikzcd}$$

Linear equivariant maps $\Leftrightarrow$ convolutions

ansatz for linear map:

generic integral transform $$\mathtt{I}_\kappa\big[f\big](x) := \int_{\mathbb{R}^d} dy\ \kappa(x,y)\, f(y)$$

parameterized by 2-point correlator $$\kappa : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{c_\text{out} \times c_\text{in}}$$



Theorem (linearity + translation equivariance $\Rightarrow$ convolution)

The integral transform $\mathtt{I}_\kappa$ is translation equivariant iff the 2-point correlator is *invariant*:

$$\kappa(x+t,\ y+t) \ = \ \kappa(x,y) \qquad \forall\, x,y,t \in \mathbb{R}^d$$

It depends only on the *relative* distance $x - y$, that is,

$$\kappa(x,\ y) \ = \ K(x-y) \qquad \text{for some} \quad \boxed{K : \mathbb{R}^d \to \mathbb{R}^{c_\text{out} \times c_\text{in}}}\,.$$

The integral transform is therefore given by a *convolution integral*:

$$\mathtt{I}_\kappa\big[f\big](x) \ = \ \big[K * f\big](x) \ = \ \int_{\mathbb{R}^d} dy\ K(x-y)\, f(y)$$

on pixel grids: tensors of shape

$$\big(\underbrace{X_1, \ldots, X_d}_{\mathbb{R}^d \qquad\longrightarrow},\ \underbrace{C_\text{out}, C_\text{in}}_{\mathbb{R}^{c_\text{out} \times c_\text{in}}}\big)$$

Translation equivariant bias summation

consider a general bias summation operation $f \mapsto f + \mathfrak{b}$

parameterized by a **bias field** $\mathfrak{b} : \mathbb{R}^d \to \mathbb{R}^c$ $\implies$ allows to sum a *different bias* $\mathfrak{b}(x) \in \mathbb{R}^c$ at each $x \in \mathbb{R}^d$

Theorem (translation equivariant bias summation)

Bias summation is translation equivariant iff the bias field is *invariant*:

$$\mathfrak{b}(x) = b \qquad \text{for some } b \in \mathbb{R}^c \text{ and any } x \in \mathbb{R}^d$$

similar *spatial invariance* results hold for other operations like nonlinearities, pooling, ...

Translation equivariant CNNs - Summary

we defined **feature vector spaces** as spaces of feature maps

we defined a (linear) **translation group action** on feature maps

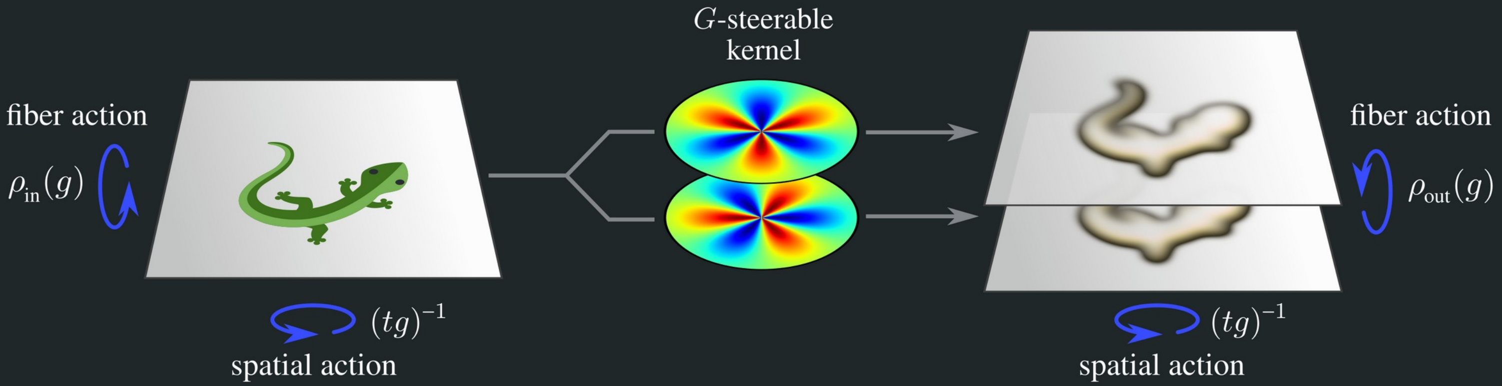
} (regular) **translation group representation**

we derived **CNN operations** like convolutions / bias summation / etc by:

1) asuming a flexible **ansatz** (linear map, bias field summation)
2) demanding **translation equivariance** → resulting in **spatial invariance / relativity / weight sharing**

next we do the same with more general symmetries of Euclidean space

Steerable CNNs on Euclidean spaces



Affine group equivariant CNNs

translations

stabilizer / local symmetries (rotations / reflections / scaling / shearing / ...)

affine groups: $\mathrm{Aff}(G) \;:=\; (\mathbb{R}^d, +) \rtimes G \qquad G \leq \mathrm{GL}(d)$

action on $\mathbb{R}^d$: $(tg)x \;:=\; gx + t$

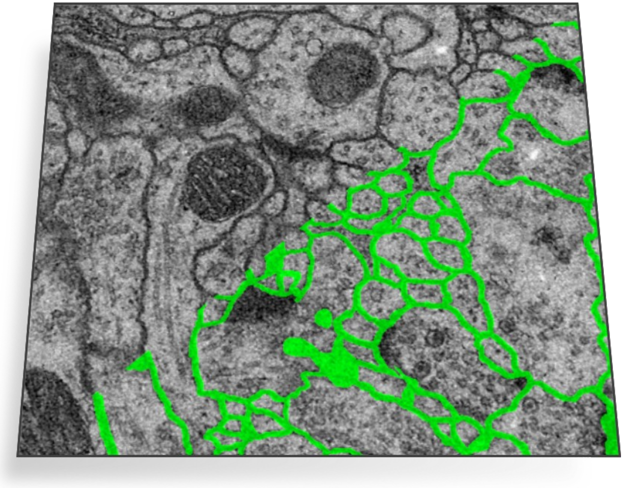
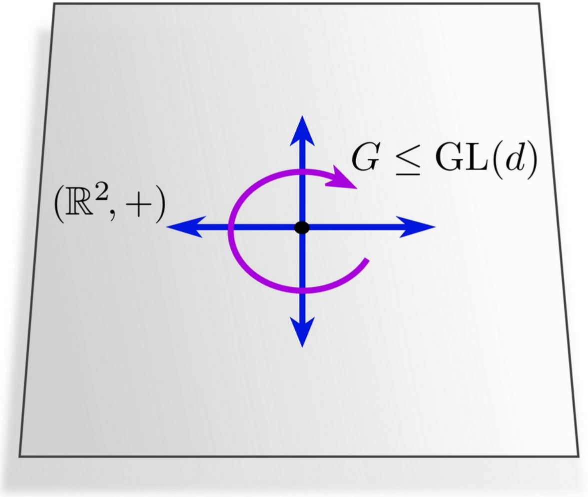


Image from ISBI 2012 EM segmentation challenge



Affine group equivariant CNNs

translations

stabilizer / local symmetries (rotations / reflections / scaling / shearing / ...)

affine groups: $\mathrm{Aff}(G) \;:=\; (\mathbb{R}^d, +) \rtimes G \qquad G \leq \mathrm{GL}(d)$

action on $\mathbb{R}^d$: $(tg)x \;:=\; gx + t$

action on feature spaces ?

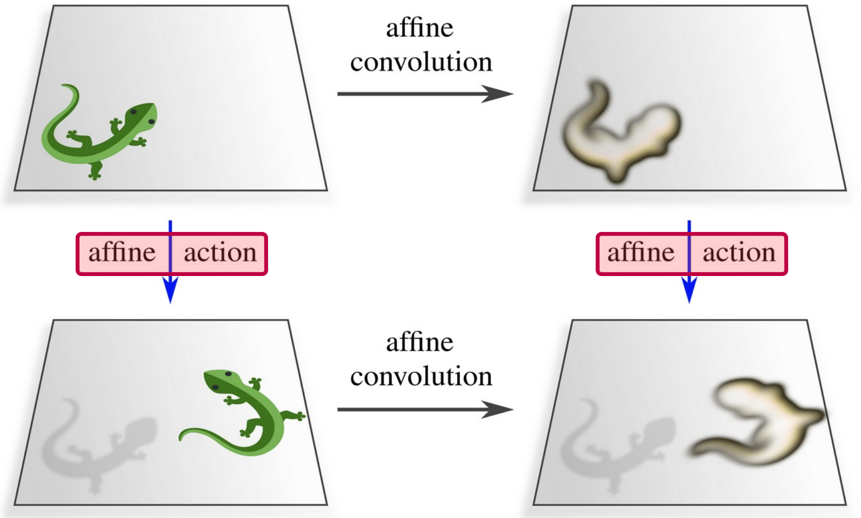


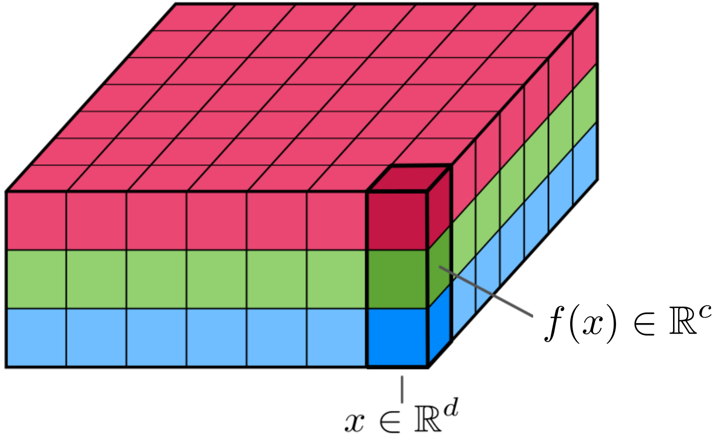
Image from ISBI 2012 EM segmentation challenge

Feature vector fields

feature vector fields on Euclidean spaces ...

... are functions $f : \mathbb{R}^d \to \mathbb{R}^c$ that assign feature vectors $f(x) \in \mathbb{R}^c$ to points $x \in \mathbb{R}^d$ (like feature maps)

... carry an $\mathrm{Aff}(G)$-action (the details depend on their *field type* ρ)



Feature vector fields

feature vector fields on Euclidean spaces ...

... are functions $f : \mathbb{R}^d \to \mathbb{R}^c$ that assign feature vectors $f(x) \in \mathbb{R}^c$ to points $x \in \mathbb{R}^d$ (like feature maps)

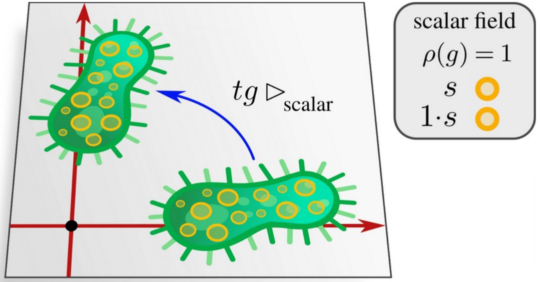
... carry an $\mathrm{Aff}(G)$-action (the details depend on their *field type* ρ)

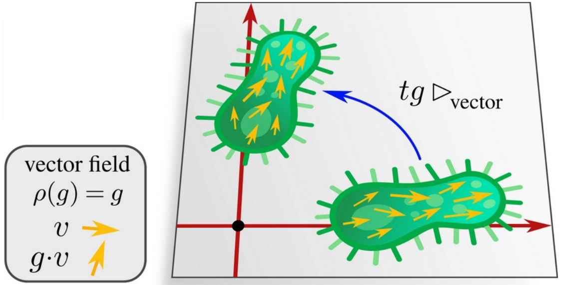
examples: scalar fields $s : \mathbb{R}^d \to \mathbb{R}^1$ transform like: $\big[(tg) \triangleright s\big](x) = \boxed{1 \cdot} \, s \boxed{\big((tg)^{-1}x\big)}$

tangent vector fields $v : \mathbb{R}^d \to \mathbb{R}^d$ transform like: $\big[(tg) \triangleright v\big](x) = \boxed{g \cdot} \, v \boxed{\big((tg)^{-1}x\big)}$

$\mathrm{Aff}(G)$ acts here by... 1) moving feature vectors on $\mathbb{R}^d$

2) G-transforming feature vectors in $\mathbb{R}^c$



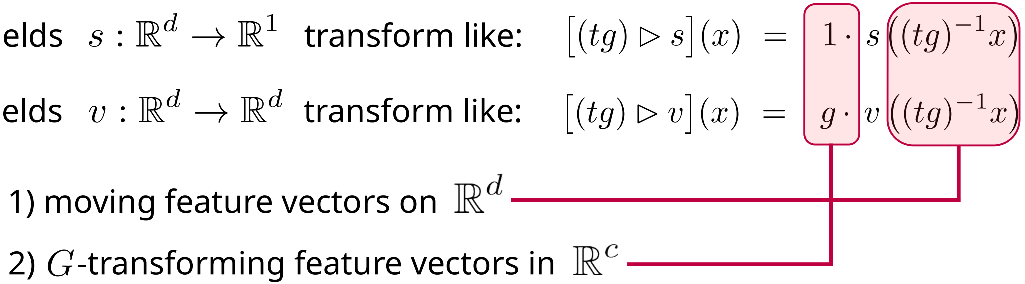


Feature vector fields

feature vector fields on Euclidean spaces ...

... are functions $f : \mathbb{R}^d \to \mathbb{R}^c$ that assign feature vectors $f(x) \in \mathbb{R}^c$ to points $x \in \mathbb{R}^d$ (like feature maps)

... carry an $\mathrm{Aff}(G)$-action (the details depend on their *field type* ρ)



examples:

scalar fields $s : \mathbb{R}^d \to \mathbb{R}^1$ transform like: $\big[(tg) \triangleright s\big](x) \;=\; \boxed{1 \cdot}\, s\, \boxed{\big((tg)^{-1}x\big)}$

tangent vector fields $v : \mathbb{R}^d \to \mathbb{R}^d$ transform like: $\big[(tg) \triangleright v\big](x) \;=\; \boxed{g \cdot}\, v\, \boxed{\big((tg)^{-1}x\big)}$

$\mathrm{Aff}(G)$ acts here by...

1) moving feature vectors on $\mathbb{R}^d$

2) G-transforming feature vectors in $\mathbb{R}^c$

ρ-feature fields $f : \mathbb{R}^d \to \mathbb{R}^c$ transform like: $\big[(tg) \triangleright f\big](x) \;=\rho(g)\, f\,\big((tg)^{-1}x\big)$

where $\rho : G \to \mathrm{GL}(c)$ is a G-representation acting on individual feature vectors in $\mathbb{R}^c$

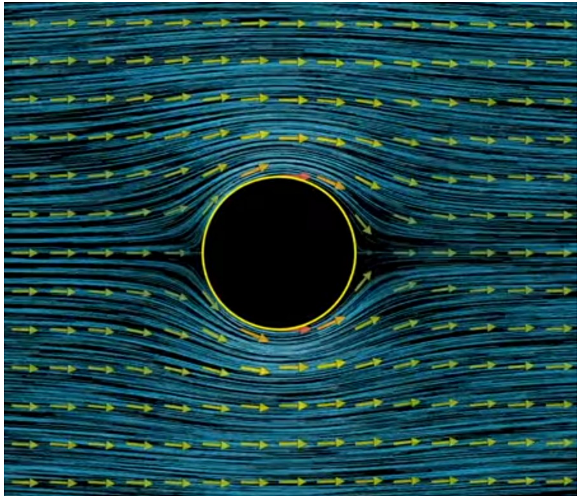
ρ-feature fields form an $\mathrm{Aff}(G)$-representation, denoted as **induced representation** $\mathrm{Ind}_G^{\mathrm{Aff}(G)}\,\rho$

Feature vector fields - examples

fluid flow

(vector)

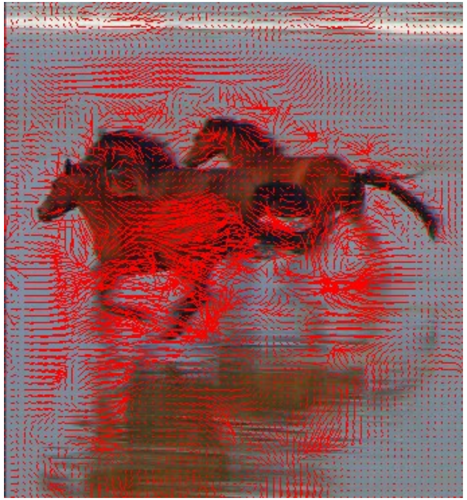
$\rho(g) = g$



optical flow

(vector)

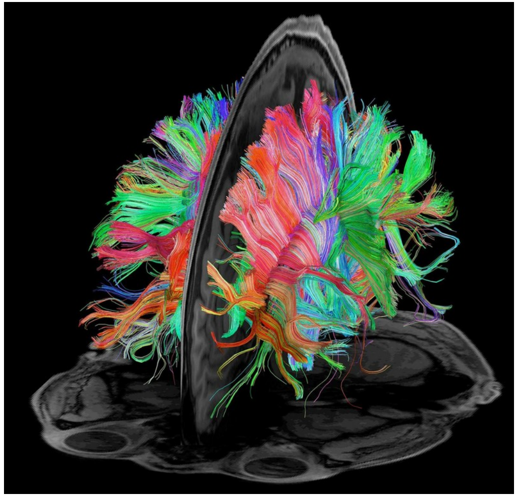
$\rho(g) = g$



diffusion tensor image

(symmetric pos. def. (1,1)-tensor)

(subspace of) $\rho(g) = g \otimes g^{-\top}$



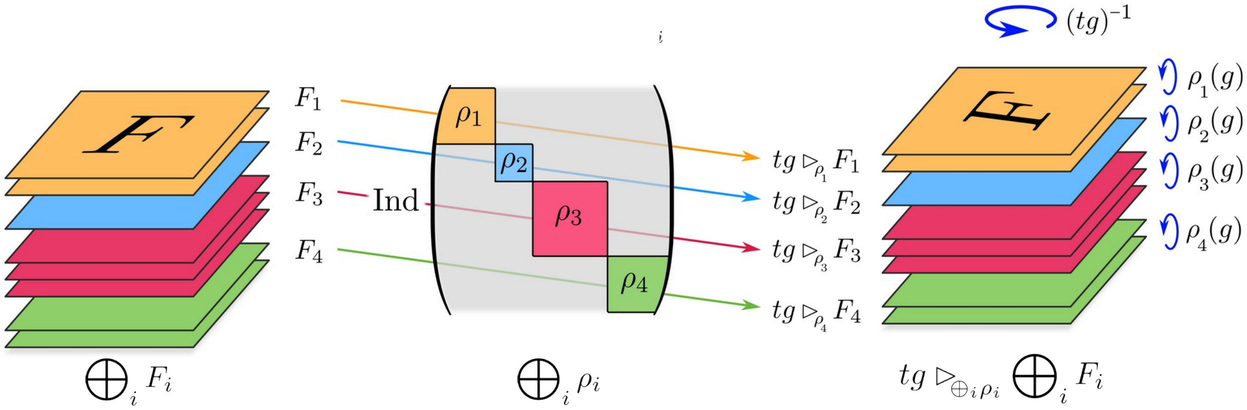
Feature vector fields – direct sum

conventional CNNs operate on a “stack” of multiple independent feature map channels

$\Rightarrow$ *#channels* as hyperparameter

steerable CNNs operate on ”stacks” $\bigoplus_i f_i$ of multiple independent feature fields

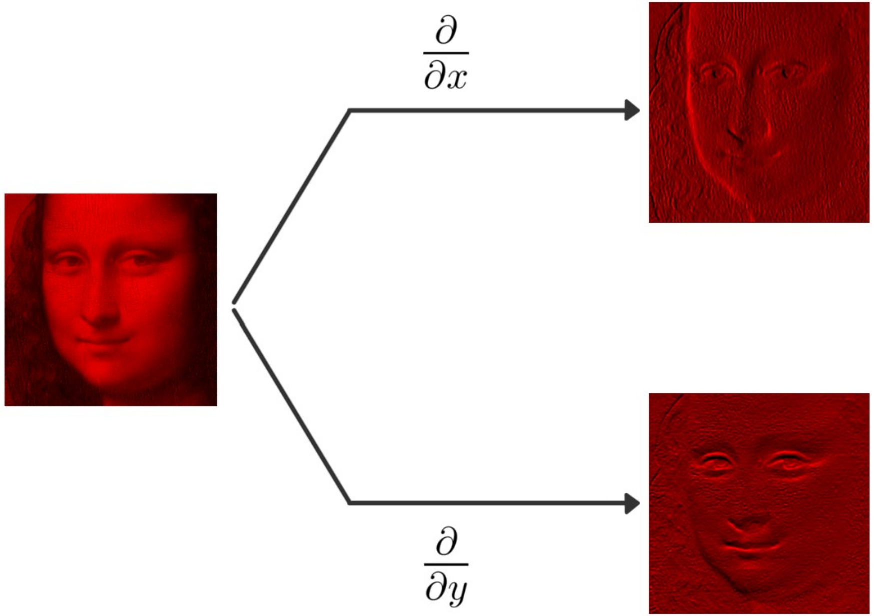
$\Rightarrow$ field *types* ρ_i and *multiplicities* as hyperparameters



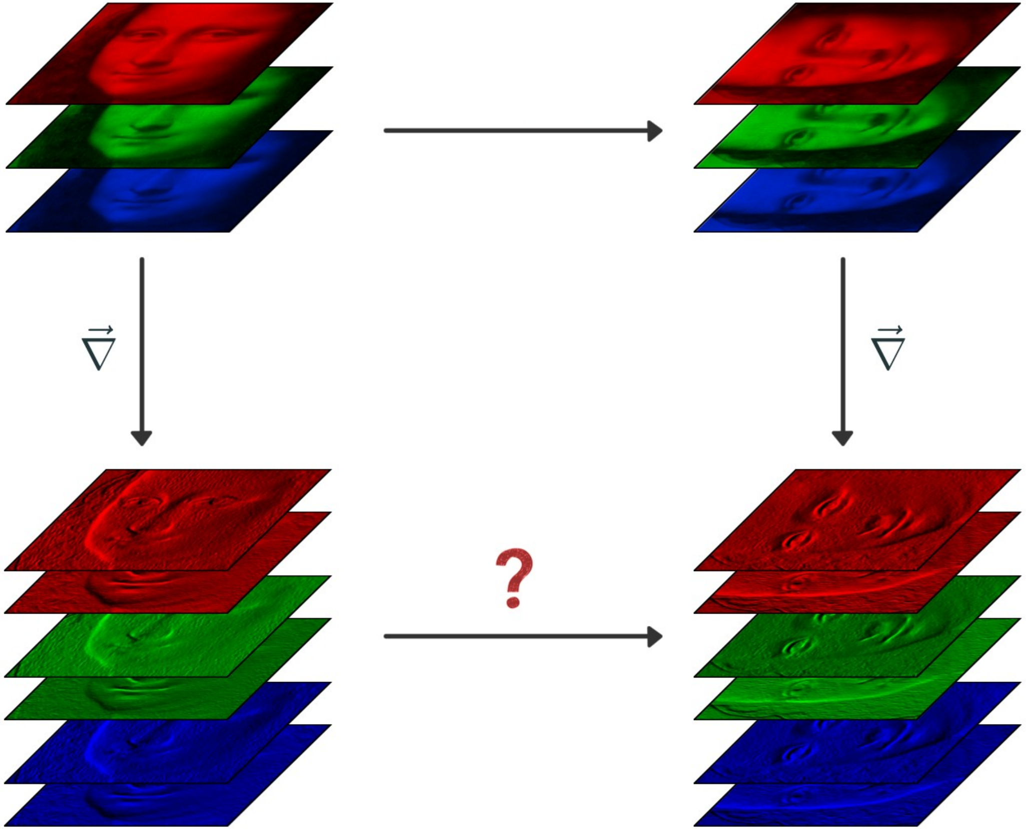
Feature vector fields - examples

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} \longrightarrow \underbrace{\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}}_{\rho(g) = \bigoplus_{i=1}^{3} (1)} \cdot \begin{pmatrix} r \\ g \\ b \end{pmatrix}$$

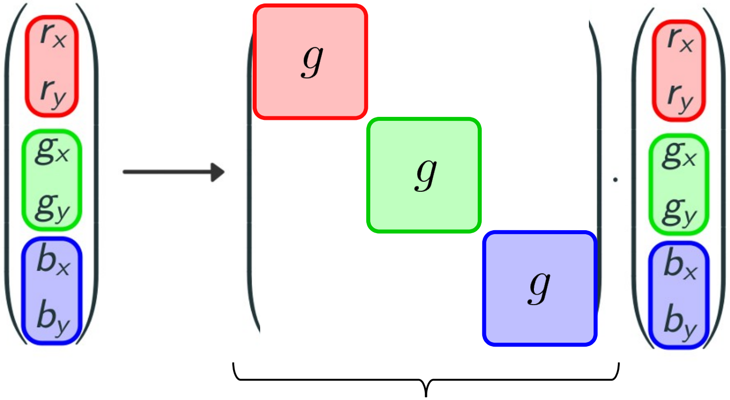
Feature vector fields - examples



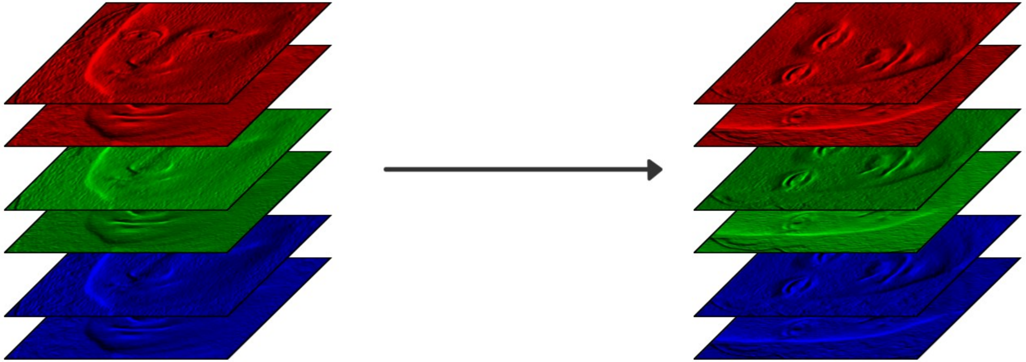
Feature vector fields - examples



Feature vector fields - examples



$$\rho(g) = \bigoplus_{i=1}^{3} g$$



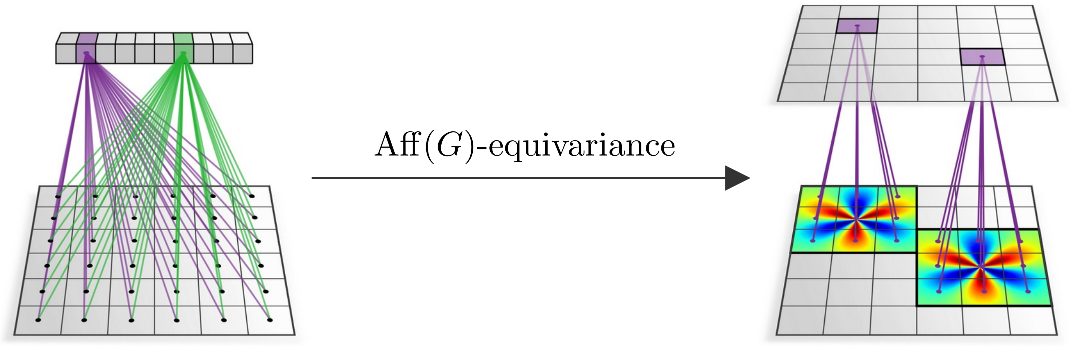
Steerable CNNs

Steerable CNN layers map between feature fields of types ρ_{in} and ρ_{out}

$$\begin{tikzcd}
\mathcal{L}^2\left(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}\right) \arrow[r, "L"] \arrow[d, "\operatorname{Ind}\rho_{\text{in}}(tg)"'] & \mathcal{L}^2\left(\mathbb{R}^d, \mathbb{R}^{c_{\text{out}}}\right) \arrow[d, "\operatorname{Ind}\rho_{\text{out}}(tg)"] \\
\mathcal{L}^2\left(\mathbb{R}^d, \mathbb{R}^{c_{\text{in}}}\right) \arrow[r, "L"'] & \mathcal{L}^2\left(\mathbb{R}^d, \mathbb{R}^{c_{\text{out}}}\right)
\end{tikzcd}$$

Steerable CNNs

Steerable CNN layers map between feature fields of types ρ_{in} and ρ_{out}



flexible ansatz

weight sharing + G-steerability

approach: - start with flexible ansatz for layers

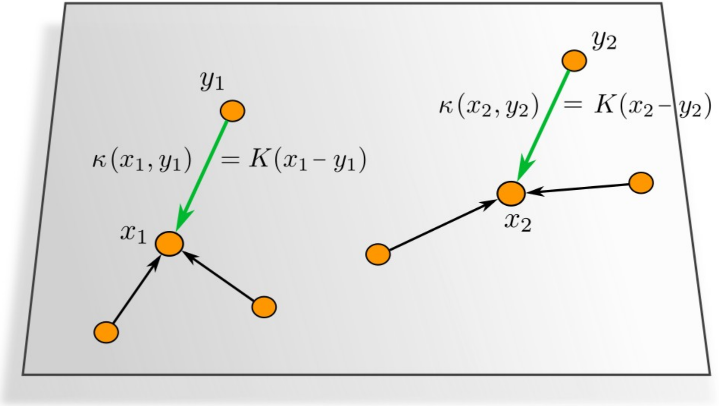
- demand $\mathrm{Aff}(G)$-equivariance, resulting in...
  1) spatial weight sharing —— $(\mathbb{R}^d, +) \rtimes G =: \mathrm{Aff}(G)$
  2) *G*-steerability —— G

Linear equivariant maps $\Leftrightarrow$ G-*steerable* convolutions

ansatz for linear map:

generic integral transform $\mathtt{I}_\kappa[f](x) := \int_{\mathbb{R}^d} dy\ \kappa(x,y)\, f(y)$

parameterized by 2-point correlator $\kappa : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$



demanding Aff(*G*)-equivariance:

Theorem. *The integral transform* $\mathtt{I}_\kappa$ *is* $\operatorname{Aff}(G)$ *equivariant iff:*

1) *it is a* convolution integral

$$\mathtt{I}_\kappa[f](x) \ = \ [K * f](x) \ = \ \int_{\mathbb{R}^d} dy\ K(x-y)\, f(y)\,.$$

with a matrix valued kernel $K : \mathbb{R}^d \to \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ *defined by translation relativity* $\kappa(x,y) = K(x-y)$

2) *the kernel is* G-steerable: $K(gx) \ = \ \frac{1}{|\det g|} \rho_{\text{out}}(g)\, K(x)\, \rho_{\text{in}}(g)^{-1} \quad \forall\, g \in G,\ x \in \mathbb{R}^d$

Linear equivariant maps $\Leftrightarrow$ *G-steerable* convolutions

ansatz for linear map:

generic integral transform $\mathtt{I}_\kappa\big[f\big](x) := \int_{\mathbb{R}^d} dy\ \kappa(x,y)\, f(y)$

parameterized by 2-point correlator $\kappa : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{c_\text{out}\times c_\text{in}}$

$$\begin{tikzcd}
\mathbb{R}^d \arrow[r, "K"] \arrow[d, "g\,\cdot"'] & \mathbb{R}^{c_\text{out}\times c_\text{in}} \arrow[d, "\frac{\rho_\text{in}^* \otimes \rho_\text{out}}{|\det|}(g)"] \\
\mathbb{R}^d \arrow[r, "K"'] & \mathbb{R}^{c_\text{out}\times c_\text{in}}
\end{tikzcd}$$

demanding Aff(*G*)-equivariance:

Theorem. *The integral transform* $\mathtt{I}_\kappa$ *is* $\mathrm{Aff}(G)$ *equivariant iff:*

1) *it is a* convolution integral

$$\mathtt{I}_\kappa\big[f\big](x) \;=\; \big[K * f\big](x) \;=\; \int_{\mathbb{R}^d} dy\, K(x-y)\, f(y)\,.$$

with a matrix valued kernel $K : \mathbb{R}^d \to \mathbb{R}^{c_\text{out}\times c_\text{in}}$ *defined by translation relativity* $\kappa(x,y) = K(x-y)$

2) *the kernel is G-*steerable*:* $\quad K(gx) \;=\; \frac{1}{|\det g|}\rho_\text{out}(g)\, K(x)\, \rho_\text{in}(g)^{-1} \quad \forall\, g \in G,\ x \in \mathbb{R}^d$

Linear equivariant maps $\Leftrightarrow$ *G-steerable* convolutions

ansatz for linear map:

gen

$$\mathbb{R}^d \xrightarrow{K} \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$$

parameterize

$$\frac{\rho_{\text{in}}^* \otimes \rho_{\text{out}}}{|\det|}(g)$$

c_{in}

takeaway:

linearity + translation equivariance $\Leftrightarrow$ convolution

linearity + Aff(G) equivariance $\Leftrightarrow$ convolution with G-steerable kernel

demanding

Theorem. *The integral transform* $\mathtt{I}_\kappa$ *is* $\mathrm{Aff}(G)$ *equivariant iff:*

1) *it is a* convolution integral

$$\mathtt{I}_\kappa[f](x) = [K * f](x) = \int_{\mathbb{R}^d} dy\, K(x-y)\, f(y)\,.$$

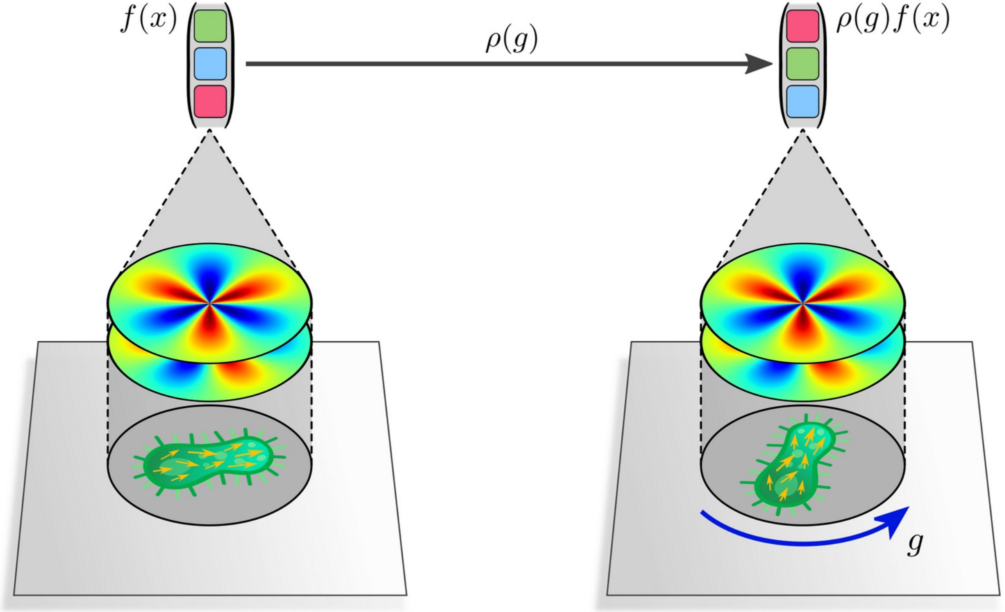
with a matrix valued kernel $K : \mathbb{R}^d \to \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ *defined by translation relativity* $\kappa(x,y) = K(x-y)$

2) *the kernel is G-*steerable*.* $\quad K(gx) = \frac{1}{|\det g|}\rho_{\text{out}}(g)\, K(x)\, \rho_{\text{in}}(g)^{-1} \quad \forall\, g \in G,\ x \in \mathbb{R}^d$

G-*steerable* kernels

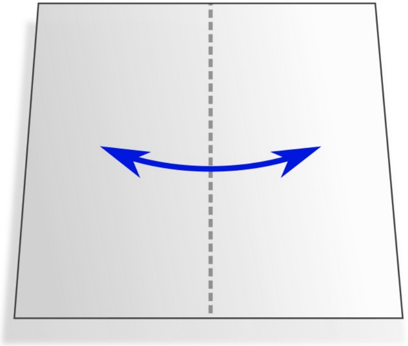
convolution kernels summarize their field of view around $x \in \mathbb{R}^d$ into a feature vector $f(x) \in \mathbb{R}^{c_\text{out}}$

G-steerable kernels guarantee: G-trafo of their input field of view $\Rightarrow$ G-trafo of the output feature vector



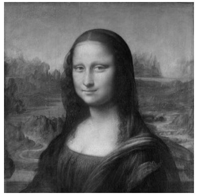
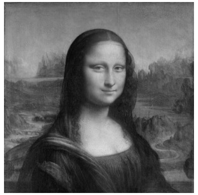
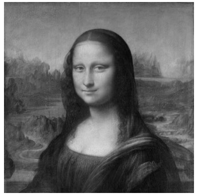
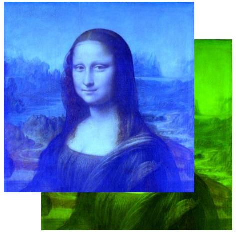
G-*steerable* kernels – reflection group example

example: *reflection* steerable kernels $G = \{e, s\}, \quad s^2 = e$



G-*steerable* kernels – reflection group example

example: *reflection* steerable kernels $G = \{e, s\}, \quad s^2 = e$

field type ρ	$\rho(e)$	$\rho(s)$	original field	transformed field
trivial / scalar	(1)	(1)		
sign-flip / pseudo-scalar	(1)	(-1)		
regular	$\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$	$\begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}$		

G-*steerable* kernels – reflection group example

example: *reflection* steerable kernels $G = \{e, s\}, \quad s^2 = e$

general steerability constraint:

$$K(gx) = \frac{1}{\boxed{\xcancel{|\det g|}}} \rho_{\mathrm{out}}(g) K(x) \rho_{\mathrm{in}}(g)^{\boxed{\xcancel{-1}}} \qquad \forall\, \boxed{g \in G,}\ x \in \mathbb{R}^2$$

$|\det g| = 1$

$g = g^{-1}$

constraint for $g = e$ trivial,
only $g = s$ left

⇒ specific reflection steerably constraint:

$$K(sx) = \rho_{\mathrm{out}}(s) K(x) \rho_{\mathrm{in}}(s) \qquad \forall x \in \mathbb{R}^2$$

spatially
reflected
kernel

channel-
transformed
kernel

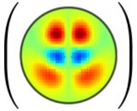
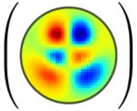
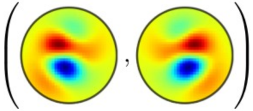
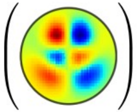
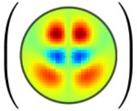
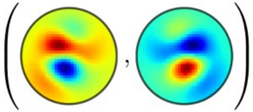
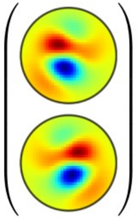
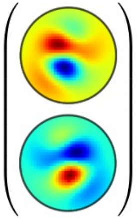
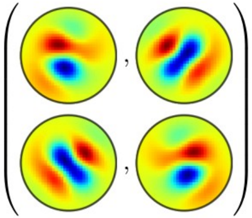
G-*steerable* kernels – reflection group examples

example: *reflection* steerable kernels

- **regular → regular:** $K:\mathbb{R}^2\to\mathbb{R}^{2\times 2}$

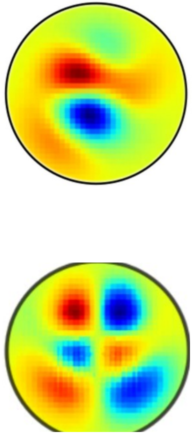
$$\begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}(s.x) \;=\; \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\cdot\begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}(x)\cdot\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \;=\; \begin{bmatrix} K_{22} & K_{21} \\ K_{12} & K_{11} \end{bmatrix}(x)$$

representation	group elements	
ρ	identity e	reflection s
trivial / scalar	(1)	(1)
sign-flip / pseudo-scalar	(1)	(-1)
regular	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

ρ_{in} ρ_{out}	trivial	sign-flip	regular
trivial	$K_{11}(sx) = K_{11}(x)$ 	$K_{11}(sx) = -K_{11}(x)$ 	$K_{11}(sx) = K_{12}(x)$ 
sign-flip	$K_{11}(sx) = -K_{11}(x)$ 	$K_{11}(sx) = K_{11}(x)$ 	$K_{11}(sx) = -K_{12}(x)$ 
regular	$K_{11}(sx) = K_{21}(x)$ 	$K_{11}(sx) = -K_{21}(x)$ 	$K_{11}(sx) = K_{22}(x)$ $K_{12}(sx) = K_{21}(x)$ 

full derivation of these examples @ Weiler et al. 2021, *Coordinate Independent Convolutional Networks*, Section 5.3.3

G-*steerable* kernels – expansion in linear basis



to solve the G-steerability kernel constraint in general, observe that:

- the set $\left\{K : \mathbb{R}^d \to \mathbb{R}^{c_\text{out} \times c_\text{in}}\right\}$ of *unconstrained* convolution kernels forms a *vector space*

- the constraint $K(gx) \;=\; \frac{1}{|\det g|} \rho_\text{out}(g)\, K(x)\, \rho_\text{in}(g)^{-1} \quad \forall\, g \in G,\; x \in \mathbb{R}^d$ is *linear*

$\Longrightarrow$ G-steerable kernels form a *linear (vector) subspace* !

to parameterize steerable convolutions:

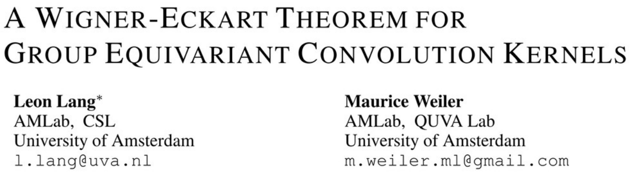
1) solve for a *basis* $\{K_1, \dots, K_N\}$ of G-steerable kernels (precomputation step)

2) expand kernel in this basis with trainable weights: $K = \sum_{i=1}^{N} w_i K_i$ (during forward pass)

G-*steerable* kernels – Wigner-Eckart theorem

analytical solution for compact G (including in particular any $G \leq \mathrm{O}(d)$)

based on an analogy: G-steerable kernels $\Leftrightarrow$ tensor operators in QM



A WIGNER-ECKART THEOREM FOR
GROUP EQUIVARIANT CONVOLUTION KERNELS

Leon Lang*
AMLab, CSL
University of Amsterdam
l.lang@uva.nl

Maurice Weiler
AMLab, QUVA Lab
University of Amsterdam
m.weiler.ml@gmail.com

A PROGRAM TO BUILD
E(n)-EQUIVARIANT STEERABLE CNNS

Gabriele Cesa
Qualcomm AI Research*
University of Amsterdam
gcesa@qti.qualcomm.com

Leon Lang
University of Amsterdam
l.lang@uva.nl

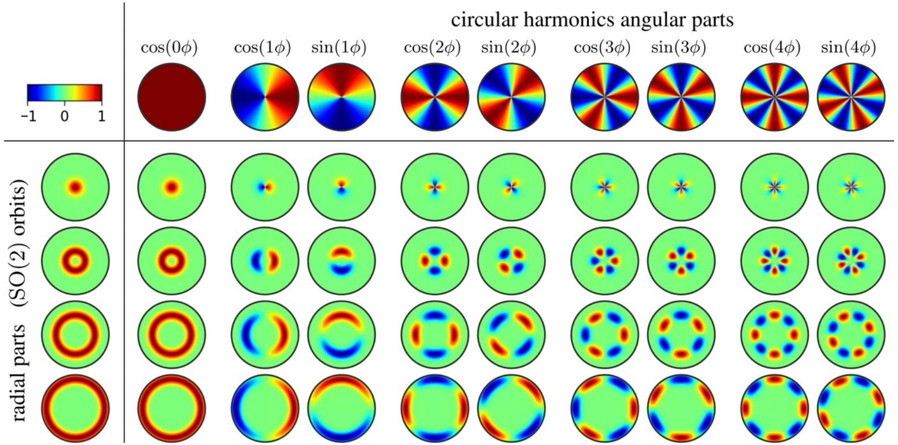
Maurice Weiler
University of Amsterdam
m.weiler.ml@gmail.com

Leon Lang

Gabriele Cesa

the solution decomposes steerable kernels into:

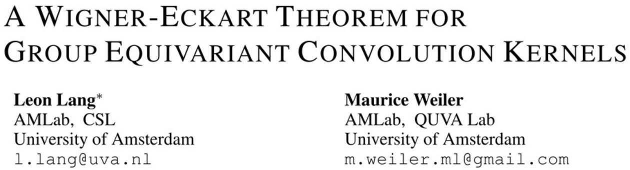
- harmonics on *G*-orbits (Peter-Weyl)
- *Clebsch-Gordan coefficients*
- *irrep endomorphisms* (reduced matrix elements)



G-*steerable* kernels – Wigner-Eckart theorem

analytical solution for compact G (including in particular any $G \leq \mathrm{O}(d)$)

based on an analogy: G-steerable kernels $\Leftrightarrow$ tensor operators in QM



A WIGNER-ECKART THEOREM FOR GROUP EQUIVARIANT CONVOLUTION KERNELS

Leon Lang[*]
AMLab, CSL
University of Amsterdam
`l.lang@uva.nl`

Maurice Weiler
AMLab, QUVA Lab
University of Amsterdam
`m.weiler.ml@gmail.com`

A PROGRAM TO BUILD E(n)-EQUIVARIANT STEERABLE CNNS

Gabriele Cesa
Qualcomm AI Research[*]
University of Amsterdam
`gcesa@qti.qualcomm.com`

Leon Lang
University of Amsterdam
`l.lang@uva.nl`

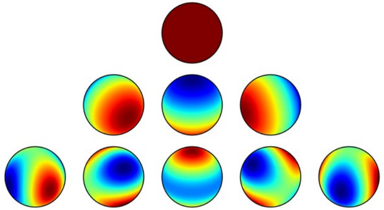
Maurice Weiler
University of Amsterdam
`m.weiler.ml@gmail.com`

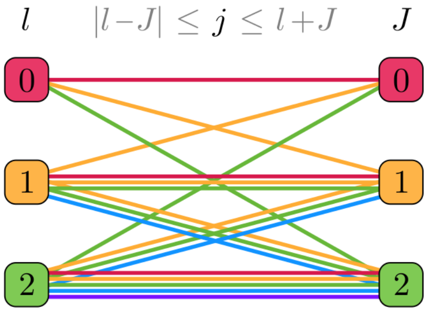
Leon Lang

Gabriele Cesa

the solution decomposes steerable kernels into:

- harmonics on *G*-orbits (Peter-Weyl)
- *Clebsch-Gordan coefficients*
- *irrep endomorphisms* (reduced matrix elements)

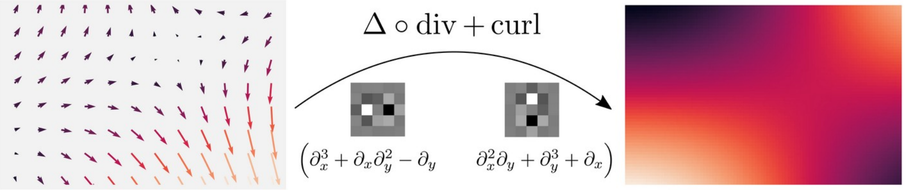


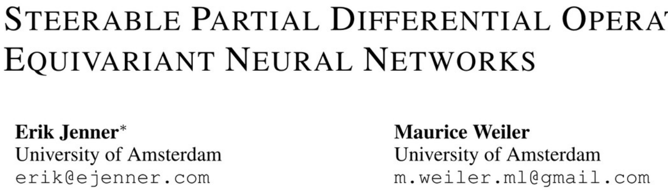


transition rules for $\mathrm{SO}(3)$

we get *transition rules* between irrep-fields (as in quantum mechanics)

Linear equivariant maps $\Leftrightarrow$ *G-steerable* convolutions




STEERABLE PARTIAL DIFFERENTIAL OPERATORS FOR EQUIVARIANT NEURAL NETWORKS

Erik Jenner*
University of Amsterdam
`erik@ejenner.com`

Maurice Weiler
University of Amsterdam
`m.weiler.ml@gmail.com`


Erik Jenner

linear maps revisited:

our integral transform ansatz $\mathtt{I}_\kappa\big[f\big](x) := \int_{\mathbb{R}^d} dy\ \kappa(x,y)\, f(y)$ does not cover all possible linear maps

a stronger version of the theorem proves:

continuous, $\mathrm{Aff}(G)$-equivariant linear maps $\Leftrightarrow$ convolutions with ***G*-steerable Schwartz distributions**

the distributional setting covers in particular **equivariant partial differential operators**

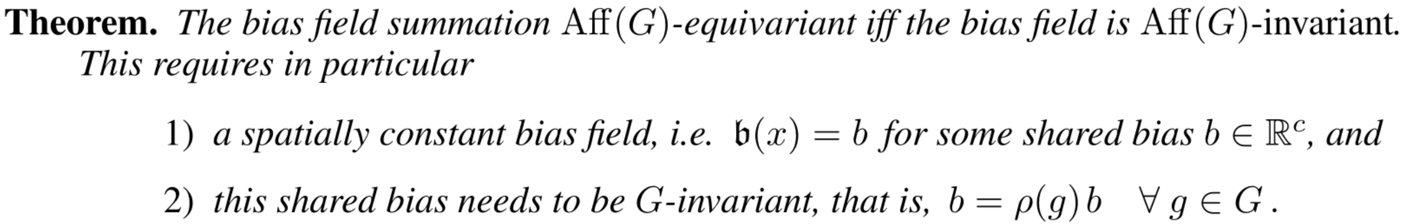
$\mathrm{Aff}(G)$-equivariant bias summation

flexible ansatz:

consider a general bias summation operation $f \mapsto f + \mathfrak{b}$

parameterized by a **bias field** $\mathfrak{b} : \mathbb{R}^d \to \mathbb{R}^c \implies$ allows to sum a *different bias* $\mathfrak{b}(x) \in \mathbb{R}^c$ at each $x \in \mathbb{R}^d$

demanding equivariance, we get:



Theorem. *The bias field summation* $\mathrm{Aff}(G)$*-equivariant iff the bias field is* $\mathrm{Aff}(G)$-invariant.
This requires in particular

1) *a spatially constant bias field, i.e.* $\mathfrak{b}(x) = b$ *for some shared bias* $b \in \mathbb{R}^c$*, and*

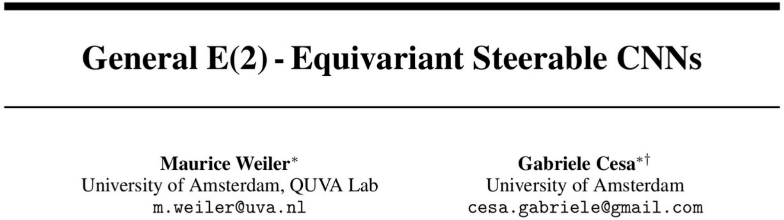
2) *this shared bias needs to be G-invariant, that is,* $b = \rho(g)\, b \quad \forall\, g \in G\,.$

similar results for nonlinearities, pooling operations, etc.

e2cnn / escnn library

github: https://github.com/QUVA-Lab/e2cnn
https://github.com/QUVA-Lab/escnn

PyTorch extension for $\mathrm{Aff}(G)$-steerable CNNs (for compact *G*)



General E(2) - Equivariant Steerable CNNs

Maurice Weiler*
University of Amsterdam, QUVA Lab
m.weiler@uva.nl

Gabriele Cesa*†
University of Amsterdam
cesa.gabriele@gmail.com

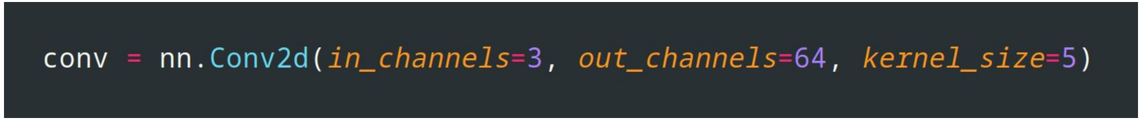
A PROGRAM TO BUILD E(n)-EQUIVARIANT STEERABLE CNNS

Gabriele Cesa
Qualcomm AI Research*
University of Amsterdam
gcesa@qti.qualcomm.com

Leon Lang
University of Amsterdam
l.lang@uva.nl

Maurice Weiler
University of Amsterdam
m.weiler.ml@gmail.com

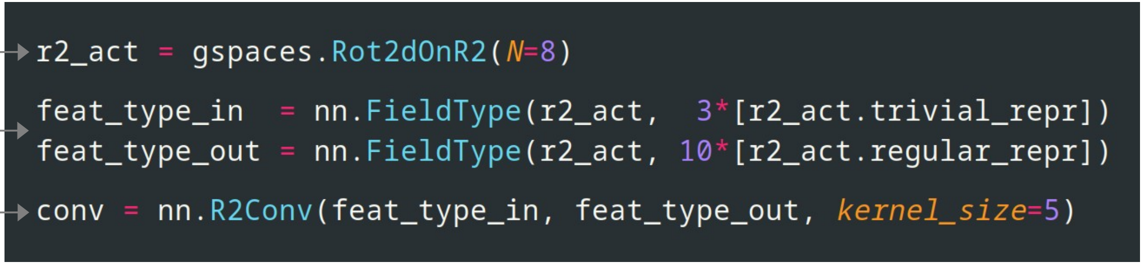
convolution in native PyTorch:



```
conv = nn.Conv2d(in_channels=3, out_channels=64, kernel_size=5)
```

convolution in e2cnn / escnn:

- fix symmetry group $G = \mathbb{Z}_8$ + action on $\mathbb{R}^2$ →
- fix types + multiplicities of feature fields →
- construct Aff(*G*)-equivariant convolution →

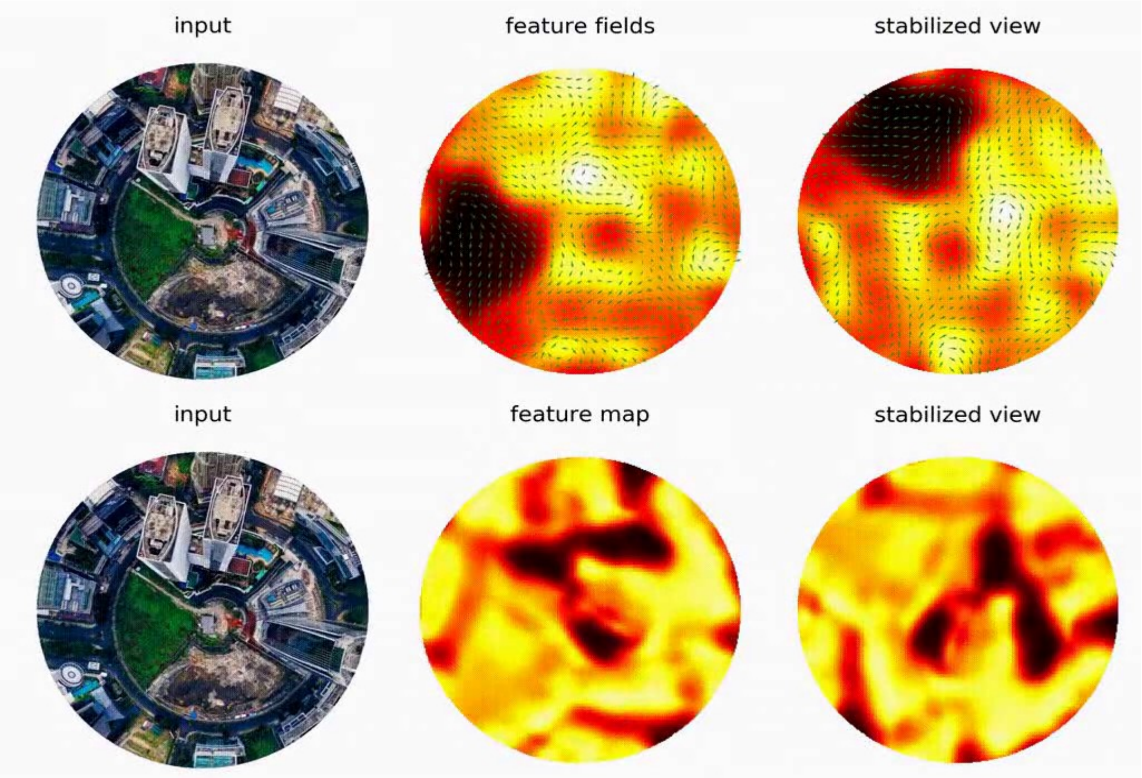


```
r2_act = gspaces.Rot2dOnR2(N=8)

feat_type_in  = nn.FieldType(r2_act,  3*[r2_act.trivial_repr])
feat_type_out = nn.FieldType(r2_act, 10*[r2_act.regular_repr])

conv = nn.R2Conv(feat_type_in, feat_type_out, kernel_size=5)
```

Equivariance demonstration

SE(2)-steerable CNN:

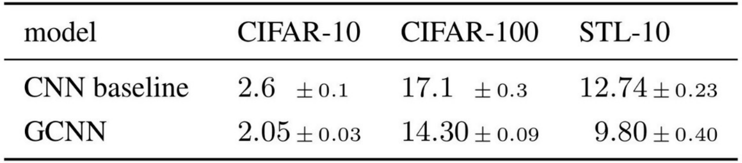
conventional CNN:



Emperical results – natural images

group convolutions as drop in replacement

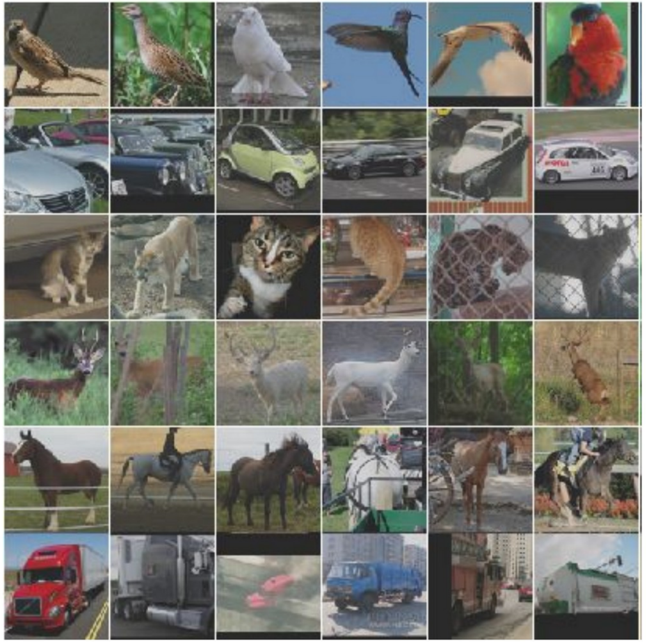
- same number of parameters
- same training setup
- no hyperparameter tuning



model	CIFAR-10	CIFAR-100	STL-10
CNN baseline	2.6 ± 0.1	17.1 ± 0.3	12.74 ± 0.23
GCNN	2.05 ± 0.03	14.30 ± 0.09	9.80 ± 0.40

Test errors on natural image datasets

[70]



Emperical results - benchmarking

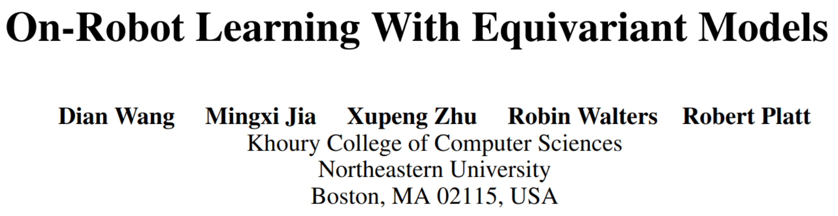
extensive benchmark of:

- groups $G \leq \mathrm{O}(2)$
- *G*-representations / field types
- *G*-equivariant nonlinearities
- invariant maps

covering a wide range of related work and new models

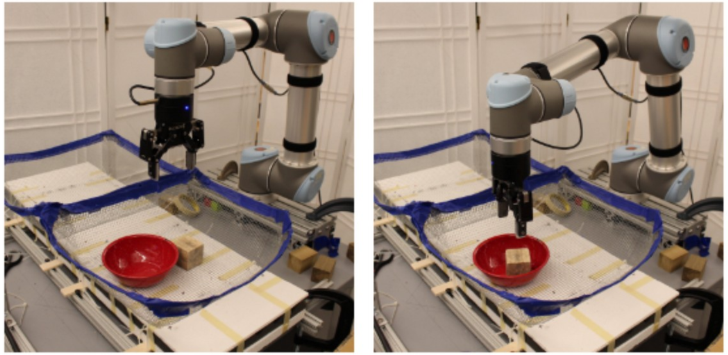
	group	representation		nonlinearity	invariant map	citation	MNIST O(2)	MNIST rot	MNIST 12k
1	$\{e\}$	(conventional CNN)		ELU	-	-	5.53 ± 0.20	2.87 ± 0.09	0.91 ± 0.06
2	C_1	regular	ρ_{reg}	ELU	G-pooling	[7, 9]	5.19 ± 0.08	2.48 ± 0.13	0.82 ± 0.01
3	C_2					[7, 9]	3.29 ± 0.07	1.32 ± 0.02	0.87 ± 0.04
4	C_3					-	2.87 ± 0.04	1.19 ± 0.06	0.80 ± 0.03
5	C_4					[6, 1, 7, 9, 10]	2.40 ± 0.05	1.02 ± 0.03	0.99 ± 0.03
6	C_6					[8]	2.08 ± 0.03	0.89 ± 0.03	0.84 ± 0.02
7	C_8					[7, 9]	1.96 ± 0.04	0.84 ± 0.02	0.89 ± 0.03
8	C_{12}					[7]	1.95 ± 0.07	0.80 ± 0.03	0.89 ± 0.03
9	C_{16}					[7, 9]	1.93 ± 0.04	0.82 ± 0.02	0.95 ± 0.04
10	C_{20}					[7]	1.95 ± 0.05	0.83 ± 0.05	0.94 ± 0.06
11	C_4	quotient	$5\rho_{\text{reg}}\oplus2\rho_{\text{quot}}^{C_4/C_2}\oplus2\psi_0$			[1]	2.43 ± 0.05	1.03 ± 0.05	1.01 ± 0.03
12	C_8		$5\rho_{\text{reg}}\oplus2\rho_{\text{quot}}^{C_8/C_2}\oplus2\rho_{\text{quot}}^{C_8/C_4}\oplus2\psi_0$			-	2.03 ± 0.05	0.84 ± 0.05	0.91 ± 0.02
13	C_{12}		$5\rho_{\text{reg}}\oplus2\rho_{\text{quot}}^{C_{12}/C_2}\oplus2\rho_{\text{quot}}^{C_{12}/C_4}\oplus3\psi_0$			-	2.04 ± 0.04	0.81 ± 0.02	0.95 ± 0.02
14	C_{16}		$5\rho_{\text{reg}}\oplus2\rho_{\text{quot}}^{C_{16}/C_2}\oplus2\rho_{\text{quot}}^{C_{16}/C_4}\oplus4\psi_0$			-	2.00 ± 0.01	0.86 ± 0.04	0.98 ± 0.04
15	C_{20}		$5\rho_{\text{reg}}\oplus2\rho_{\text{quot}}^{C_{20}/C_2}\oplus2\rho_{\text{quot}}^{C_{20}/C_4}\oplus5\psi_0$			-	2.01 ± 0.05	0.83 ± 0.03	0.96 ± 0.04
16	C_{16}	regular/scalar	$\psi_0 \xrightarrow{\text{conv}} \rho_{\text{reg}} \xrightarrow{G\text{-pool}} \psi_0$	ELU, G-pooling		[6, 36]	2.02 ± 0.02	0.90 ± 0.03	0.93 ± 0.04
17		regular/vector	$\psi_1 \xrightarrow{\text{conv}} \rho_{\text{reg}} \xrightarrow{\text{vector pool}} \psi_1$	vector field		[13, 37]	2.12 ± 0.02	1.07 ± 0.03	0.78 ± 0.03
18		mixed vector	$\rho_{\text{reg}}\oplus\psi_1 \xrightarrow{\text{conv}} 2\rho_{\text{reg}} \xrightarrow[\text{pool}]{\text{vector}} \rho_{\text{reg}}\oplus\psi_1$	ELU, vector field		-	1.87 ± 0.03	0.83 ± 0.02	0.63 ± 0.02
19	D_1	regular	ρ_{reg}	ELU	G-pooling	-	3.40 ± 0.07	3.44 ± 0.10	0.98 ± 0.03
20	D_2					-	2.42 ± 0.07	2.39 ± 0.04	1.05 ± 0.03
21	D_3					-	2.17 ± 0.06	2.15 ± 0.05	0.94 ± 0.02
22	D_4					[6, 1, 38]	1.88 ± 0.04	1.87 ± 0.04	1.69 ± 0.03
23	D_6					[8]	1.77 ± 0.06	1.77 ± 0.04	1.00 ± 0.03
24	D_8					-	1.68 ± 0.06	1.73 ± 0.03	1.64 ± 0.02
25	D_{12}					-	1.66 ± 0.05	1.65 ± 0.05	1.67 ± 0.01
26	D_{16}					-	1.62 ± 0.04	1.65 ± 0.02	1.68 ± 0.04
27	D_{20}					-	1.64 ± 0.06	1.62 ± 0.05	1.69 ± 0.03
28	D_{16}	regular/scalar	$\psi_{0,0} \xrightarrow{\text{conv}} \rho_{\text{reg}} \xrightarrow{G\text{-pool}} \psi_{0,0}$	ELU, G-pooling		-	1.92 ± 0.03	1.88 ± 0.07	1.74 ± 0.04
29	$\mathrm{SO}(2)$	irreps ≤ 1	$\bigoplus_{i=0}^{1} \psi_i$	ELU, norm-ReLU	conv2triv	-	2.98 ± 0.04	1.38 ± 0.09	1.29 ± 0.05
30		irreps ≤ 3	$\bigoplus_{i=0}^{3} \psi_i$			-	3.02 ± 0.18	1.38 ± 0.09	1.27 ± 0.03
31		irreps ≤ 5	$\bigoplus_{i=0}^{5} \psi_i$			-	3.24 ± 0.05	1.44 ± 0.10	1.36 ± 0.04
32		irreps ≤ 7	$\bigoplus_{i=0}^{7} \psi_i$			-	3.30 ± 0.11	1.51 ± 0.10	1.40 ± 0.07
33		$\mathbb{C}$-irreps ≤ 1	$\bigoplus_{i=0}^{1} \psi_i^{\mathbb{C}}$			[12]	3.39 ± 0.10	1.47 ± 0.06	1.42 ± 0.04
34		$\mathbb{C}$-irreps ≤ 3	$\bigoplus_{i=0}^{3} \psi_i^{\mathbb{C}}$			[12]	3.48 ± 0.16	1.51 ± 0.05	1.53 ± 0.07
35		$\mathbb{C}$-irreps ≤ 5	$\bigoplus_{i=0}^{5} \psi_i^{\mathbb{C}}$			-	3.59 ± 0.08	1.59 ± 0.05	1.55 ± 0.06
36		$\mathbb{C}$-irreps ≤ 7	$\bigoplus_{i=0}^{7} \psi_i^{\mathbb{C}}$			-	3.64 ± 0.12	1.61 ± 0.06	1.62 ± 0.03
37		irreps ≤ 3	$\bigoplus_{i=0}^{3} \psi_i$	ELU, squash		-	3.10 ± 0.09	1.41 ± 0.04	1.46 ± 0.05
38				ELU, norm-ReLU	norm	-	3.23 ± 0.08	1.38 ± 0.08	1.33 ± 0.03
39				ELU, shared norm-ReLU		-	2.88 ± 0.11	1.15 ± 0.06	1.18 ± 0.03
40				shared norm-ReLU		-	3.61 ± 0.09	1.57 ± 0.05	1.88 ± 0.05
41				ELU, gate	conv2triv	-	2.37 ± 0.06	1.09 ± 0.03	1.10 ± 0.02
42				ELU, shared gate		-	2.33 ± 0.06	1.11 ± 0.03	1.12 ± 0.04
43				ELU, gate	norm	-	2.23 ± 0.09	1.04 ± 0.04	1.05 ± 0.06
44				ELU, shared gate		-	2.20 ± 0.06	1.01 ± 0.03	1.03 ± 0.03
45	$\mathrm{O}(2)$	irreps $= 0$	$\psi_{0,0}$	ELU	-	-	5.46 ± 0.46	5.21 ± 0.29	3.98 ± 0.04
46		irreps ≤ 1	$\psi_{0,0} \oplus \psi_{1,0} \oplus 2\psi_{1,1}$	ELU, norm-ReLU	$\mathrm{O}(2)$-conv2triv	-	3.31 ± 0.17	3.37 ± 0.18	3.05 ± 0.09
47		irreps ≤ 3	$\psi_{0,0} \oplus \psi_{1,0} \bigoplus_{i=1}^{3} 2\psi_{1,i}$			-	3.42 ± 0.03	3.41 ± 0.10	3.86 ± 0.09
48		irreps ≤ 5	$\psi_{0,0} \oplus \psi_{1,0} \bigoplus_{i=1}^{5} 2\psi_{1,i}$			-	3.59 ± 0.13	3.78 ± 0.31	4.17 ± 0.15
49		irreps ≤ 7	$\psi_{0,0} \oplus \psi_{1,0} \bigoplus_{i=1}^{7} 2\psi_{1,i}$			-	3.84 ± 0.25	3.90 ± 0.18	4.57 ± 0.27
50		Ind-irreps ≤ 1	$\mathrm{Ind}\,\psi_0^{\mathrm{SO}(2)} \oplus \mathrm{Ind}\,\psi_1^{\mathrm{SO}(2)}$	ELU, Ind norm-ReLU	Ind-conv2triv	-	2.72 ± 0.05	2.70 ± 0.11	2.39 ± 0.07
51		Ind-irreps ≤ 3	$\mathrm{Ind}\,\psi_0^{\mathrm{SO}(2)} \bigoplus_{i=1}^{3} \mathrm{Ind}\,\psi_i^{\mathrm{SO}(2)}$			-	2.66 ± 0.07	2.65 ± 0.12	2.25 ± 0.06
52		Ind-irreps ≤ 5	$\mathrm{Ind}\,\psi_0^{\mathrm{SO}(2)} \bigoplus_{i=1}^{5} \mathrm{Ind}\,\psi_i^{\mathrm{SO}(2)}$			-	2.71 ± 0.11	2.84 ± 0.10	2.39 ± 0.09
53		Ind-irreps ≤ 7	$\mathrm{Ind}\,\psi_0^{\mathrm{SO}(2)} \bigoplus_{i=1}^{7} \mathrm{Ind}\,\psi_i^{\mathrm{SO}(2)}$			-	2.80 ± 0.12	2.85 ± 0.06	2.25 ± 0.08
54		irreps ≤ 3	$\psi_{0,0} \oplus \psi_{1,0} \bigoplus_{i=1}^{3} 2\psi_{1,i}$	ELU, gate	$\mathrm{O}(2)$-conv2triv	-	2.39 ± 0.05	2.38 ± 0.07	2.28 ± 0.07
55					norm	-	2.21 ± 0.09	2.24 ± 0.06	2.15 ± 0.03
56		Ind-irreps ≤ 3	$\mathrm{Ind}\,\psi_0^{\mathrm{SO}(2)} \bigoplus_{i=1}^{3} \mathrm{Ind}\,\psi_i^{\mathrm{SO}(2)}$	ELU, Ind gate	Ind-conv2triv	-	2.13 ± 0.04	2.09 ± 0.05	2.05 ± 0.05
57					Ind-norm	-	1.96 ± 0.06	1.95 ± 0.05	1.85 ± 0.07

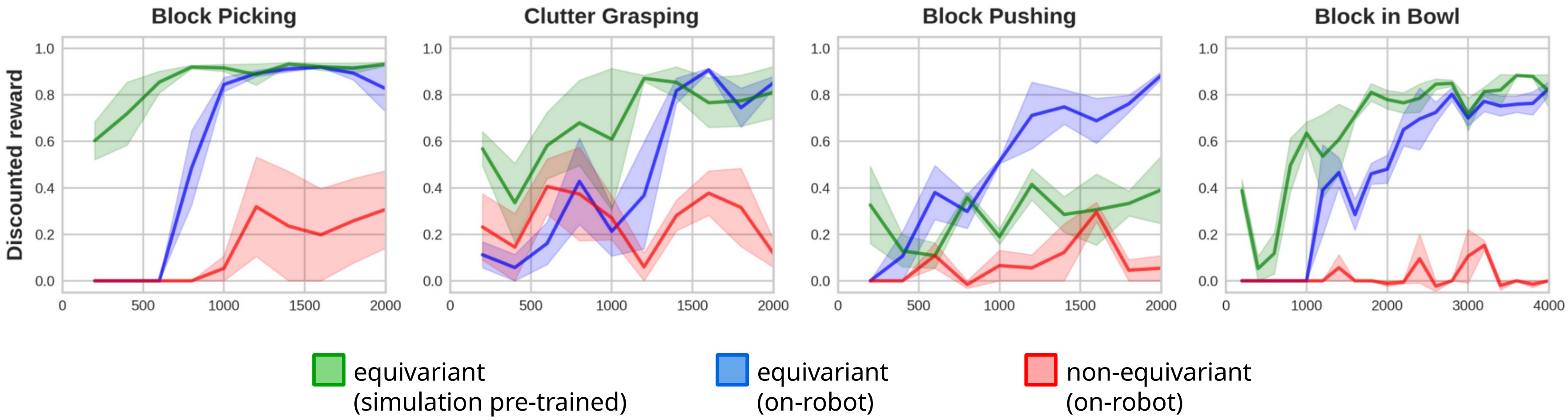
Emperical results – reinforcement learning



On-Robot Learning With Equivariant Models

Dian Wang **Mingxi Jia** **Xupeng Zhu** **Robin Walters** **Robert Platt**
Khoury College of Computer Sciences
Northeastern University
Boston, MA 02115, USA

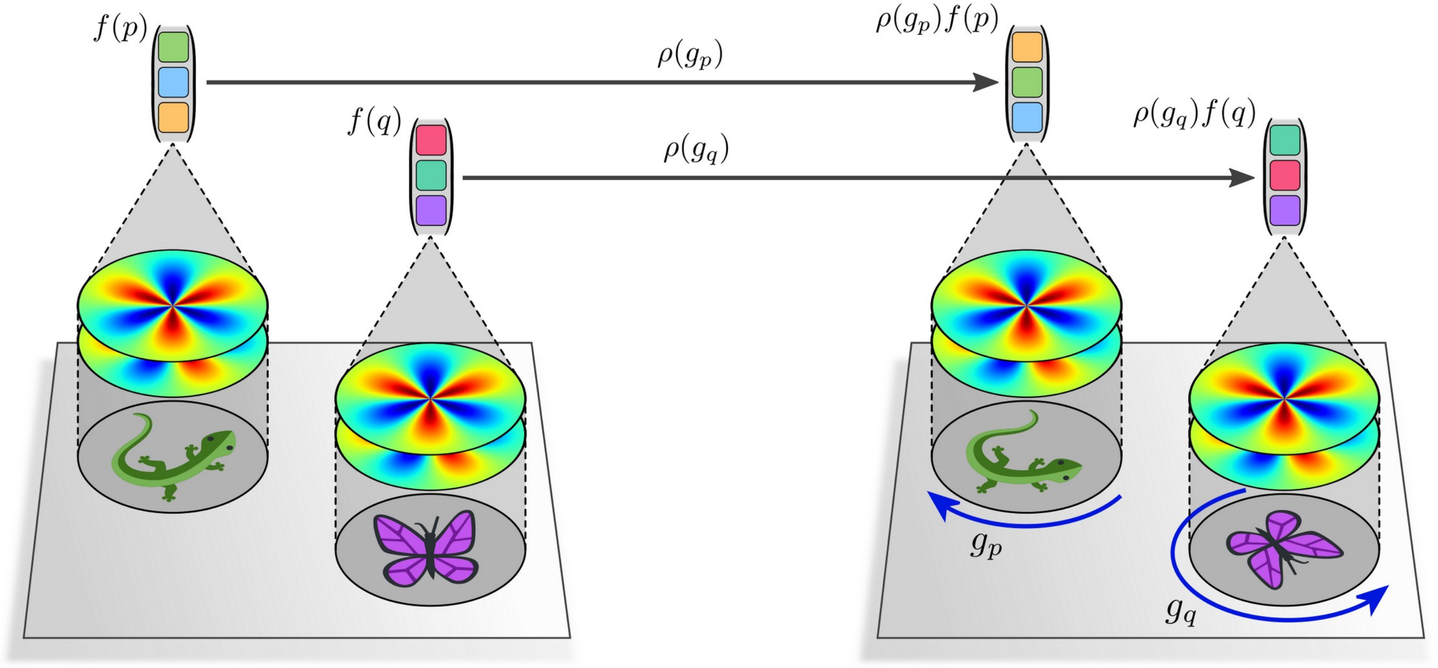




Local gauge equivariance

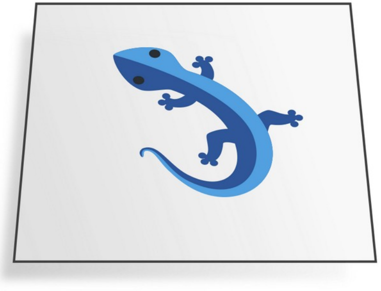
steerable CNNs are not only *globally* Aff(*G*)-equivariant, but *locally* *G*-equivariant (gauge equivariant)

formalized as *coordinate independent CNN*



Active & passive transformations

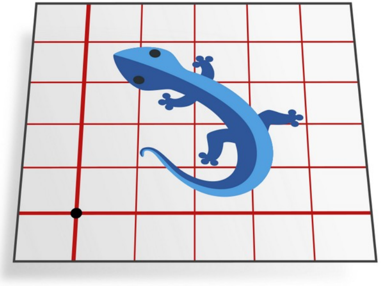
active transformations - acting on the data itself:



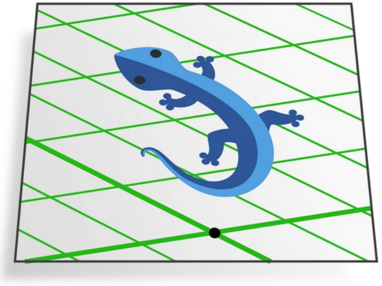
vs.

⟵ **global** transformations

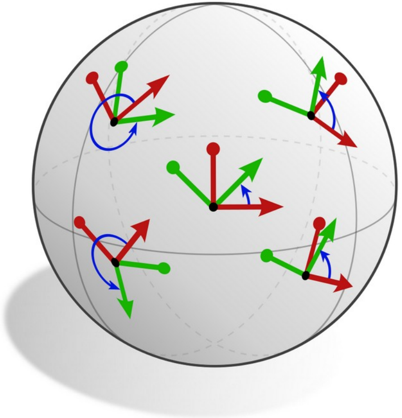
passive transformations - acting on coordinatization of data:



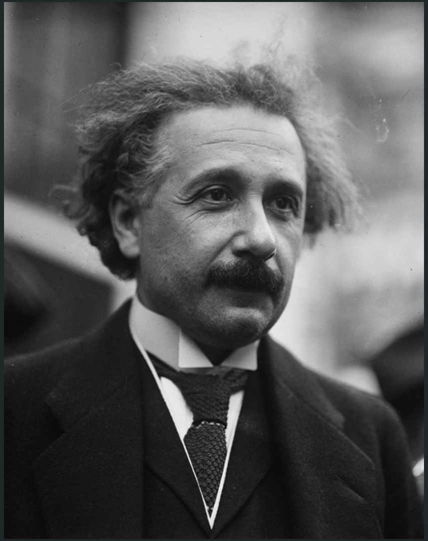
vs.



local *gauge* trafos

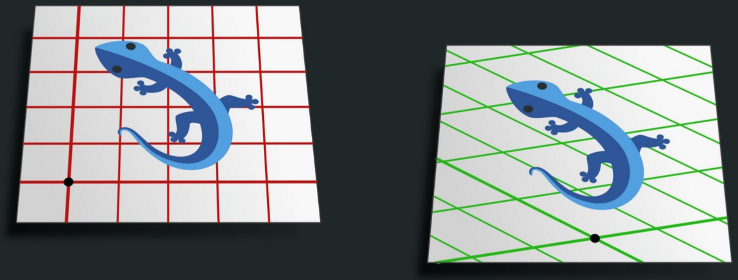


Coordinate independent CNNs on Riemannian manifolds



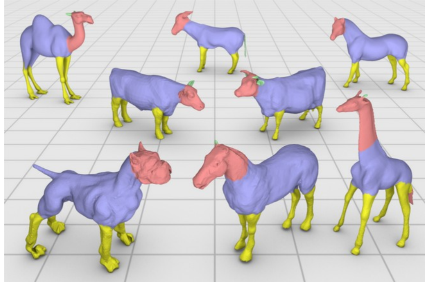
Principle of Covariance (Einstein, 1916)

“Universal laws of AI *are to be expressed by equations which hold good for all systems of coordinates.”*

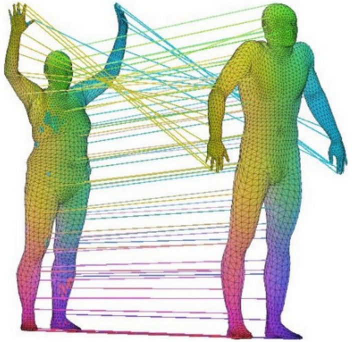


Convolutions on Riemannian manifolds

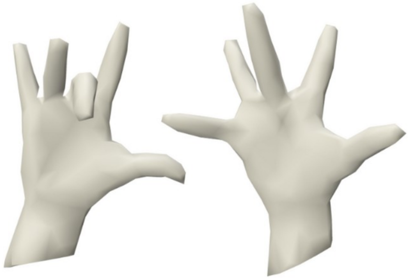
mesh segmentation



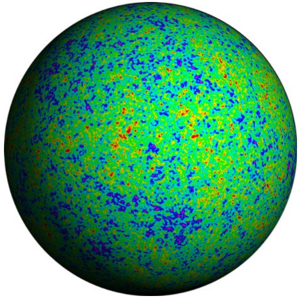
shape correspondence



deformations (metric field)

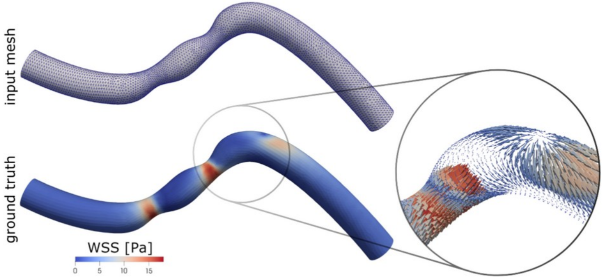


spherical CNNs

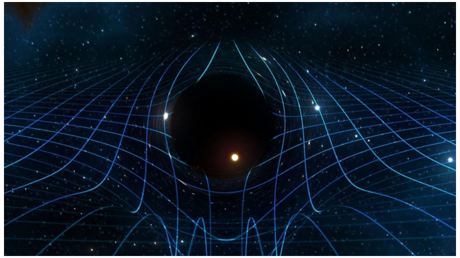


intrinsic convolutions, not in *embedding space !*

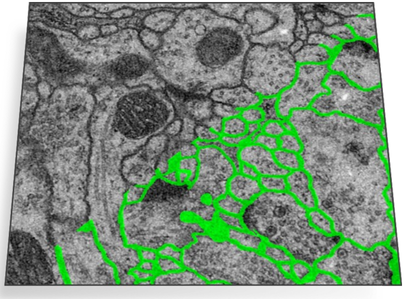
artery wall stress estimation



general relativity

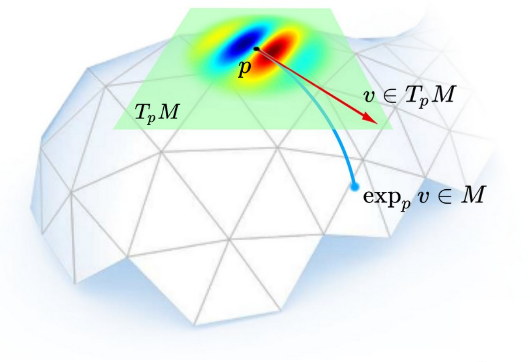


Euclidean CNNs



Design questions

Image adapted from Konakovic-Lukovic et al.



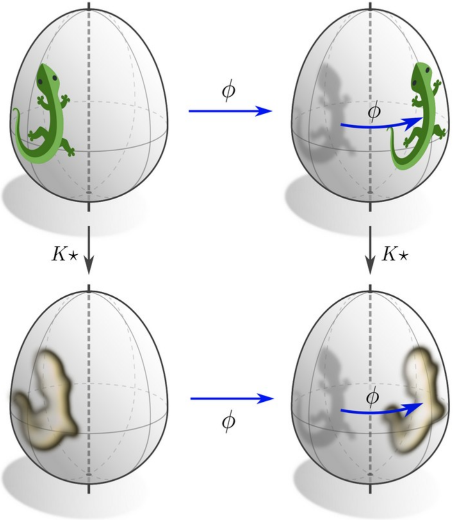
how to ...

... define *feature fields* on M?

... define *convolution kernels* on M ?

... *share weights* over M ?

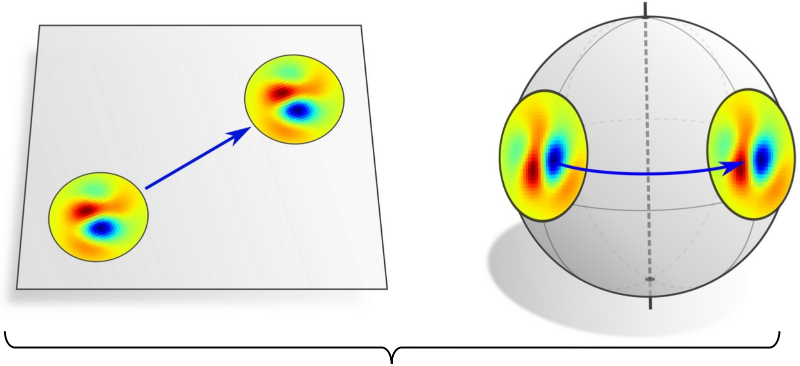
... guarantee *isometry equivariance* ?



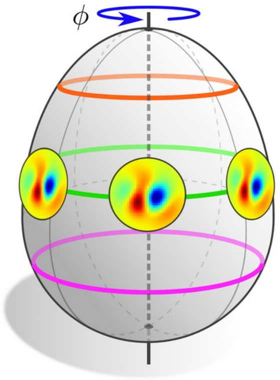
Weight sharing - via global symmetries

weight sharing by *demanding equivariance* w.r.t. global symmetries (isometries)

↯ can only share over *symmetry orbits* (in general non-transitive)



homogeneous spaces,
transitive orbits



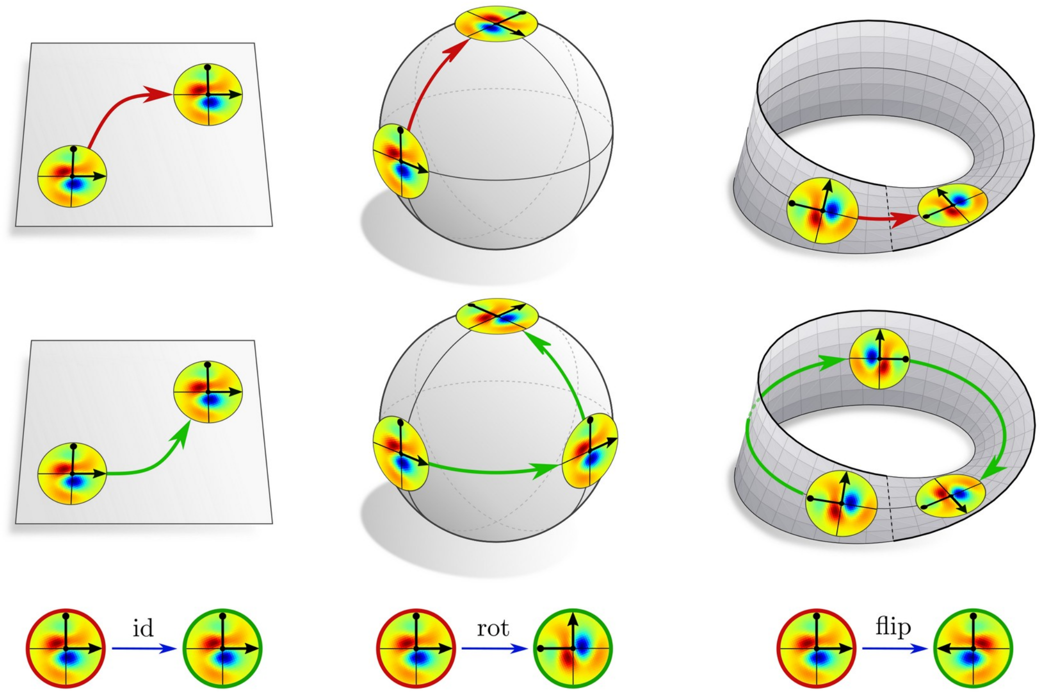
SO(2) orbits

trivial orbits

Weight sharing - via parallel transport

sharing weights by “shifting” kernel over manifold ?

⚡ parallel transport in general path dependent

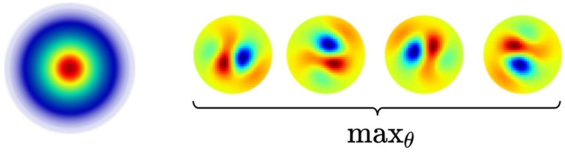


Weight sharing - approaches in the literature

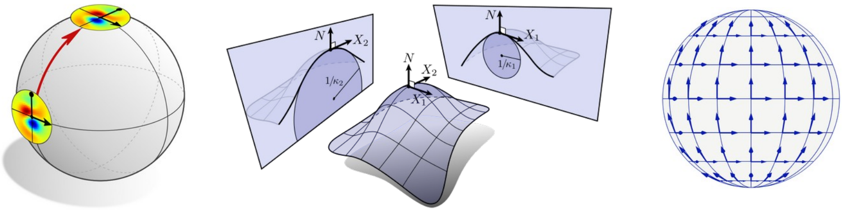
the kernel alignment (“*gauge*”) on manifolds is *inherently ambiguous*! ⇔ topological obstructions to the existence of *G*-structures

solution approaches in the literature:

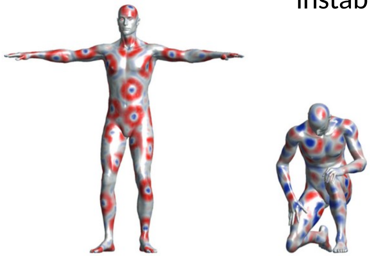
1) **gauge *invariant*** features ↯ low expressiveness



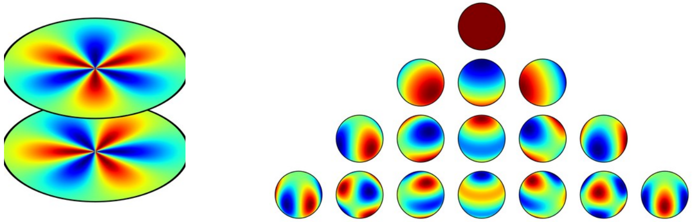
2) **heuristic** gauges ↯ instable under deformations



3) **spectral** approaches ↯ gauge independent but instable under deformations

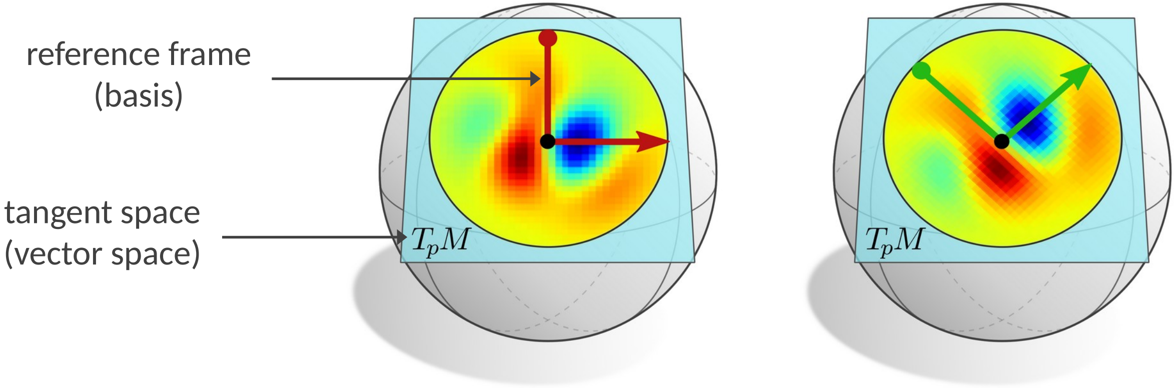


4) **gauge *equivariant*** features (ours, covers 1,2 as special cases)



Reference frames and kernel alignments

identify kernel alignment with a choice of reference frame

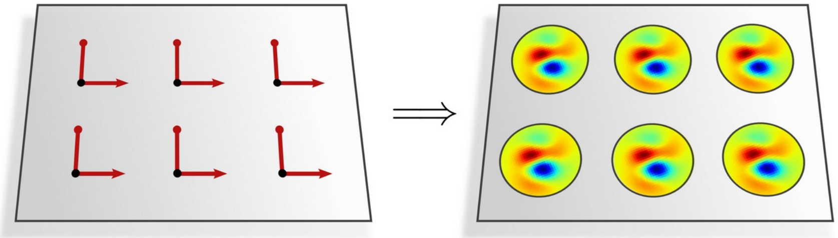


Reference frames and kernel alignments

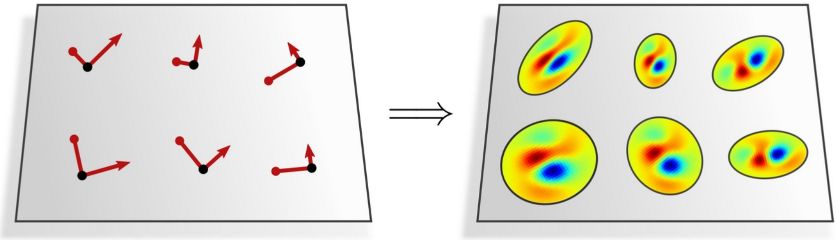
identify kernel alignment with a choice of reference frame

frame field ⟷ kernel field

standard (canonical)
frame / kernel field / CNN
on $\mathbb{R}^2$



alternative
frame / kernel field / CNN
on $\mathbb{R}^2$



Reference frames and kernel alignments

identify kernel alignment with a choice of reference frame

frame field ⟷ kernel field

ambiguity of kernel alignments == ambiguity of reference frames

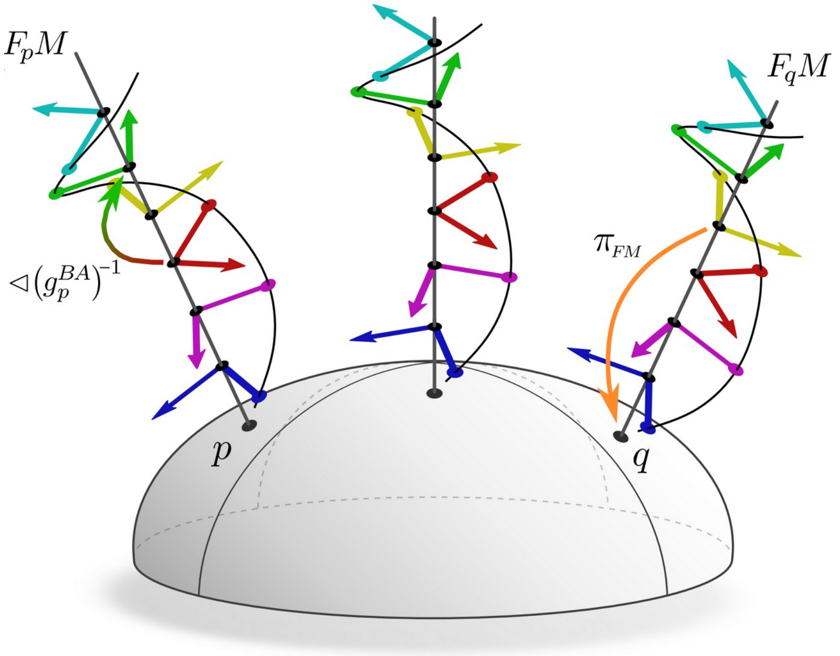
↑

“G-structure”

G-structures

frame bundle *FM* = “set” (bundle) of all frames (GL(*d*)-valued transition functions)

G*-structures *GM = sub-bundles of frames with $G \leq \mathrm{GL}(d)$ valued transition functions



G-structures

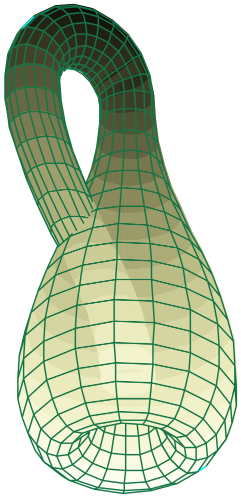
frame bundle *FM* = "set" (bundle) of all frames (GL(*d*)-valued transition functions)

G-structures *GM* = sub-bundles of frames with $G \leq \mathrm{GL}(d)$ valued transition functions

G-structures encode **additional geometric structure on *M*** in a unified way:

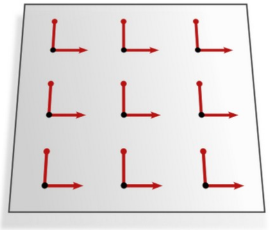
structure on M	distinguished frames	structure group $G \leq \mathrm{GL}(d)$
smooth structure only	all reference frames	$\mathrm{GL}(d)$
orientation of M	positively oriented frames	$\mathrm{GL}^+(d)$
volume form	unit volume frames	$\mathrm{SL}(d)$
Riemannian metric	orthonormal frames	$\mathrm{O}(d)$
pseudo-Riemannian metric	pseudo-orthonormal frames	$\mathrm{O}(d-n,n)$
global trivialization	global frame field	$\{e\}$

Klein bottle
non-orientable

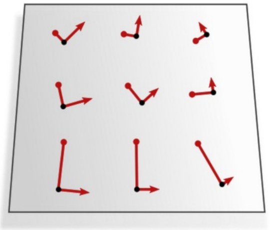


topological obstructions may prevent the existence of (continuous) *G*-structures

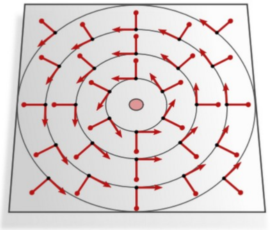
G-structures



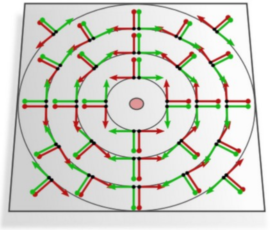
$M = \mathbb{R}^2,\ \ G = \{e\}$



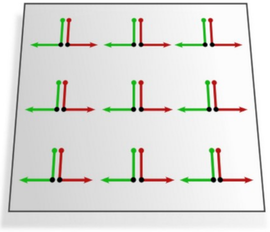
$M = \mathbb{R}^2,\ \ G = \{e\}$



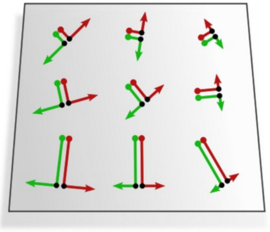
$M = \mathbb{R}^2 \backslash \{0\},\ \ G = \{e\}$



$M = \mathbb{R}^2 \backslash \{0\},\ \ G = \mathcal{R}$

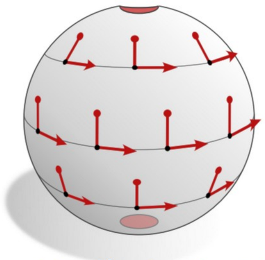


$M = \mathbb{R}^2,\ \ G = \mathcal{R}$

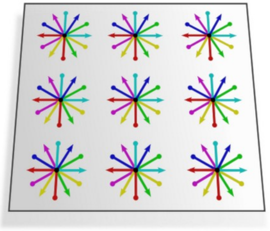


$M = \mathbb{R}^2,\ \ G = \mathcal{R}$

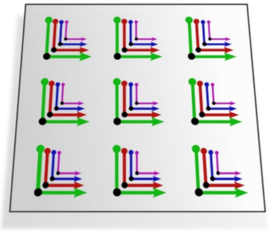
$M = S^2,\ \ G = \mathrm{SO}(2)$



$M = S^2 \backslash \text{poles},\ \ G = \{e\}$

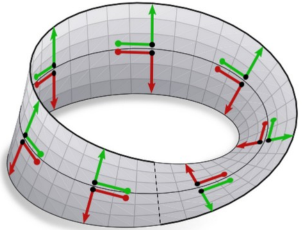


$M = \mathbb{R}^2,\ \ G = \mathrm{SO}(2)$



$M = \mathbb{R}^2,\ \ G = \mathcal{S}$

$M = \text{“Suzanne”},\ \ G = \mathrm{SO}(2)$

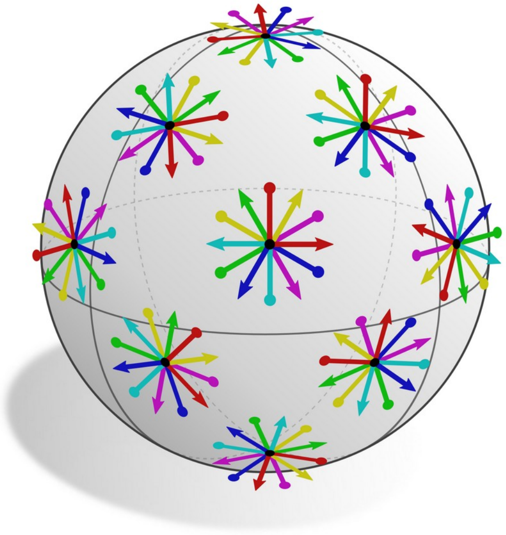


$M = \text{Möbius},\ \ G = \mathcal{R}$

GM-coordinate independence - tangent vectors

all frames of the *G*-structure are equally valid

$\Longrightarrow$ any *object* or *morphism* should be expressible relative to any frame in *GM*



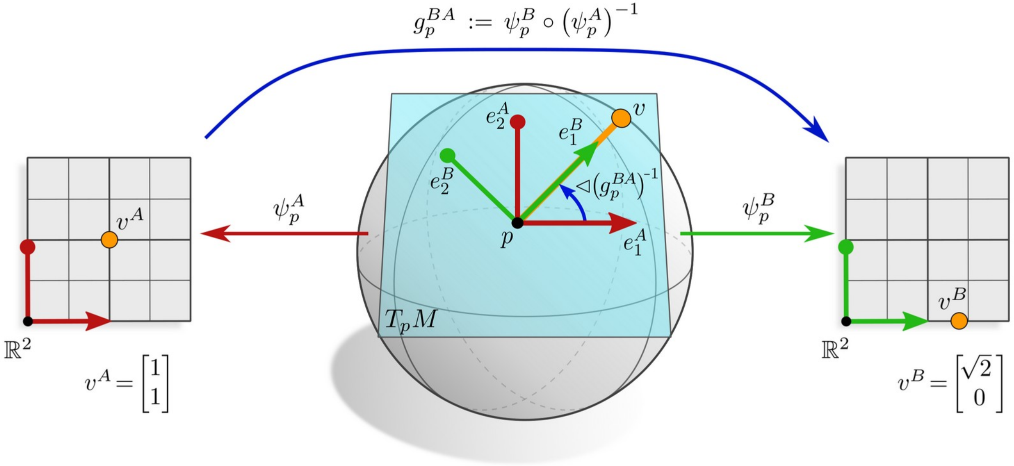
GM-coordinate independence - tangent vectors

all frames of the *G*-structure are equally valid

$\Longrightarrow$ any *object* or *morphism* should be expressible relative to any frame in *GM*

example:
- tangent vectors $v \in T_pM$ are *coordinate free*
- in gauge *A*, v is expressed by *coefficients* $v^A \in \mathbb{R}^d$
- in gauge *B*, v is expressed by *coefficients* $v^B \in \mathbb{R}^d$
- *gauge trafos* $g^{BA} \in G$ relate coefficients: $v^B = g^{BA} v^A$

different coefficients,
same information content!



GM-coordinate independence - feature vector fields

all frames of the *G*-structure are equally valid

$\Longrightarrow$ any *object* or *morphism* should be expressible relative to any frame in *GM*

coordinate independent *feature vectors* transform according to *G-representation* ρ:

$$f^A,\, f^B \in \mathbb{R}^c \qquad\qquad f^B = \rho(g^{BA})\, f^A$$

scalar field	trivial representation	$\rho(g) = id$
tangent vector field	standard representation	$\rho(g) = g$
tensor field	tensor representation	$\rho(g) = (g^{-T})^{\otimes s} \otimes g^{\otimes r}$
irrep field	irreducible representation	
regular feature field	regular representation	

formally, feature vectors are elements of a *G-associated feature vector bundle* $(GM \times \mathbb{R}^c)/\sim_\rho$

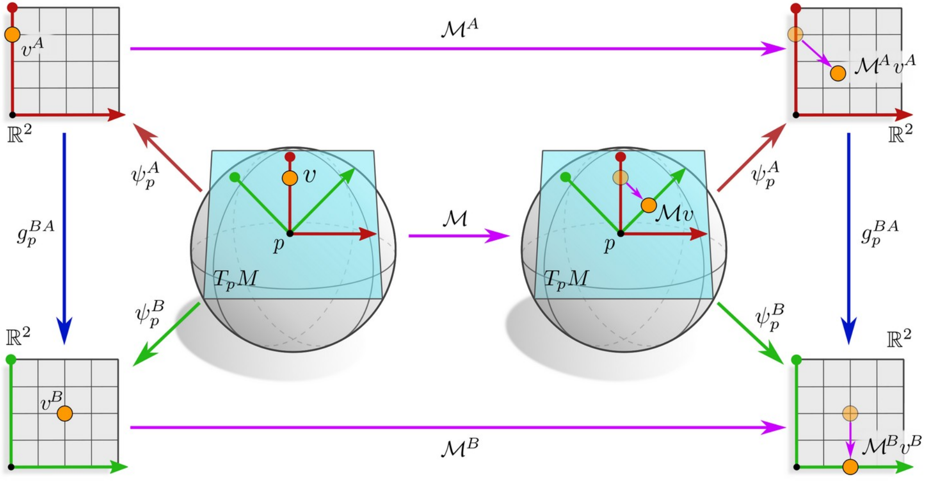
GM-coordinate independence - linear maps on T_pM

all frames of the *G*-structure are equally valid

$\Longrightarrow$ any *object* or *morphism* should be expressible relative to any frame in *GM*

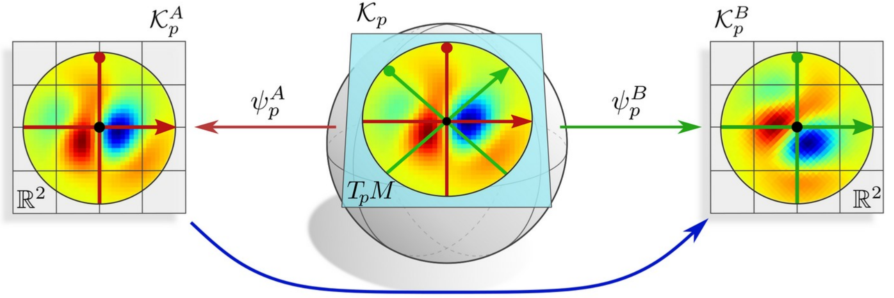
example: - linear maps $\mathcal{M} : T_pM \to T_pM$ are *coordinate free*

- in gauge *A*, $\mathcal{M}$ is expressed by *coefficients* $\mathcal{M}^A \in \mathbb{R}^{d\times d}$
- in gauge *B*, $\mathcal{M}$ is expressed by *coefficients* $\mathcal{M}^B \in \mathbb{R}^{d\times d}$
- *gauge trafos* $g^{BA} \in G$ relate coefficients: $\mathcal{M}^B = g^{BA}\mathcal{M}^A(g^{BA})^{-1}$



GM-coordinate independence - kernels

coordinate independence of kernels:



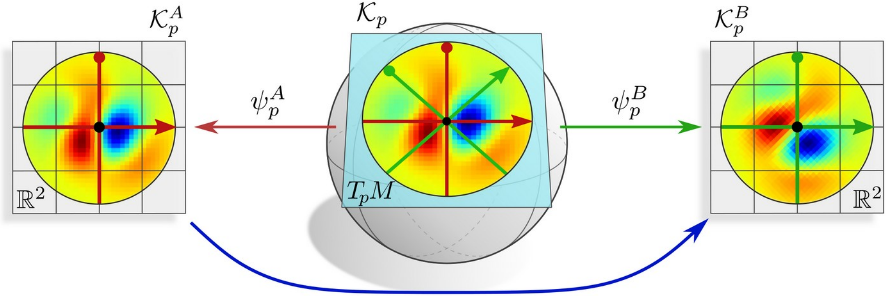
G-steerability constraint $\neq$ gauge trafo: $\boxed{\mathcal{K}_p^B} = \frac{1}{|\det g_p^{BA}|}\big(\rho_{\text{in}}^{-\top} \otimes \rho_{\text{out}}\big)\big(g_p^{BA}\big) \circ \boxed{\mathcal{K}_p^A} \circ \big(g_p^{BA}\big)^{-1}$

$$\begin{tikzcd}
\mathbb{R}^d \arrow[r, "K"] \arrow[d, "g\,\cdot"'] & \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \arrow[d, "\frac{\rho_{\text{in}}^* \otimes \rho_{\text{out}}}{|\det|}(g)"] \\
\mathbb{R}^d \arrow[r, "K"'] & \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}
\end{tikzcd}$$

$$\begin{tikzcd}
\mathbb{R}^d \arrow[r, "\boxed{\mathcal{K}_p^A}"] \arrow[d, "g_p^{BA}\cdot"'] & \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}} \arrow[d, "\frac{\rho_{\text{in}}^* \otimes \rho_{\text{out}}}{|\det|}\left(g_p^{BA}\right)"] \\
\mathbb{R}^d \arrow[r, "\boxed{\mathcal{K}_p^B}"'] & \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}
\end{tikzcd}$$

GM-coordinate independence - kernels

coordinate independence of kernels:



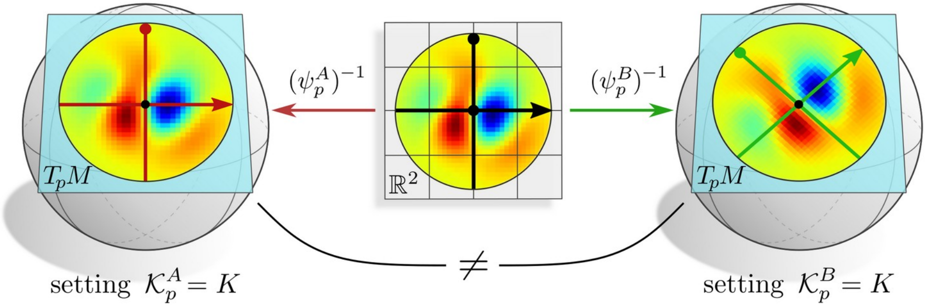
equivalence constraint $\neq$ gauge trafo: $\boxed{\mathcal{K}_p^B} = \frac{1}{|\det g_p^{BA}|} \left(\rho_{\text{in}}^{-\top} \otimes \rho_{\text{out}}\right)(g_p^{BA}) \circ \boxed{\mathcal{K}_p^A} \circ (g_p^{BA})^{-1}$

weight sharing of kernels:

depends in general on chosen gauge!

coordinate independent iff:

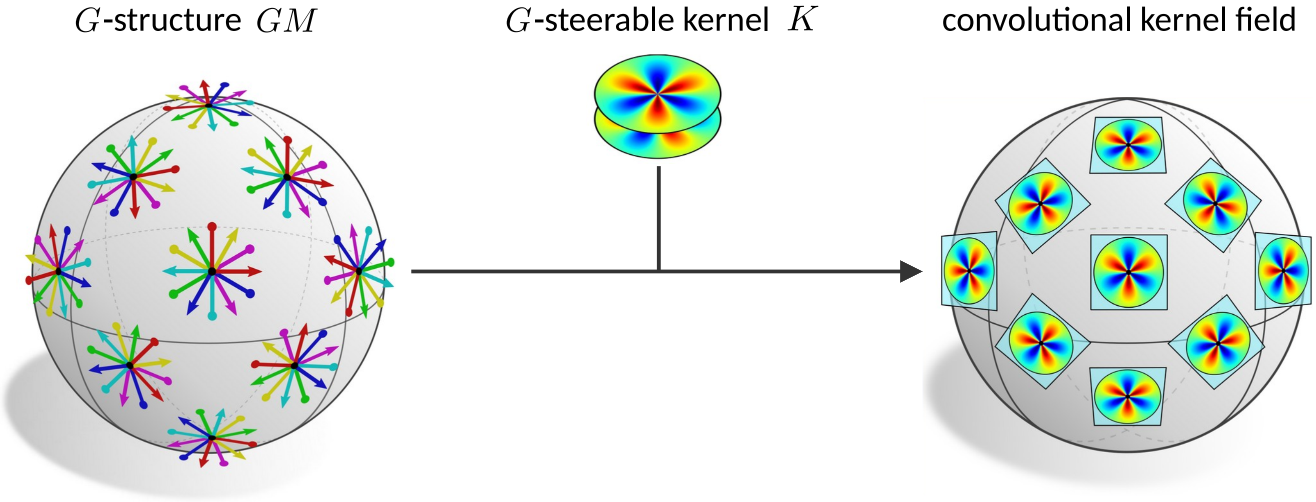
$\mathcal{K}_p^X = K$ for *any* gauge X



$$\left.\begin{array}{r}\text{coordinate independence}\\ \text{weight sharing}\end{array}\right\} \longrightarrow G\text{-steerability} \qquad K = \frac{1}{|\det g|}\left(\rho_{\text{in}}^{-\top} \otimes \rho_{\text{out}}\right)(g) \circ K \circ g^{-1} \quad \forall\, g \in G$$

GM-convolutions

GM*-coordinate independent convolutions* operate by applying a *convolutional kernel field* to a feature field



G-structure GM

G-steerable kernel K

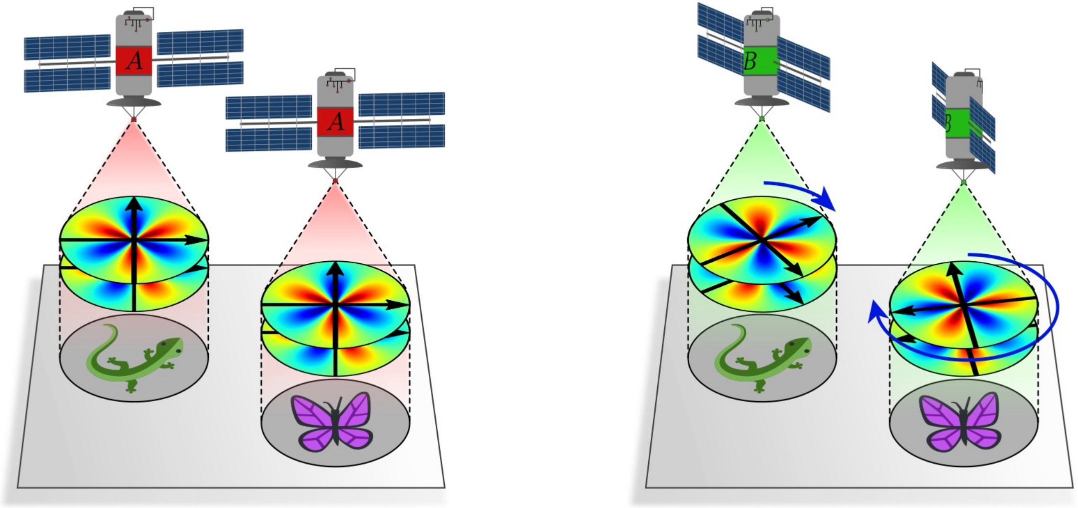
convolutional kernel field

symmetry properties:

1) *GM*-coordinate independence (passive)
2) local gauge equivariance (active)
3) global isometry equivariance (active)

GM-convolutions

GM*-coordinate independent convolutions* operate by applying a *convolutional kernel field* to a feature field

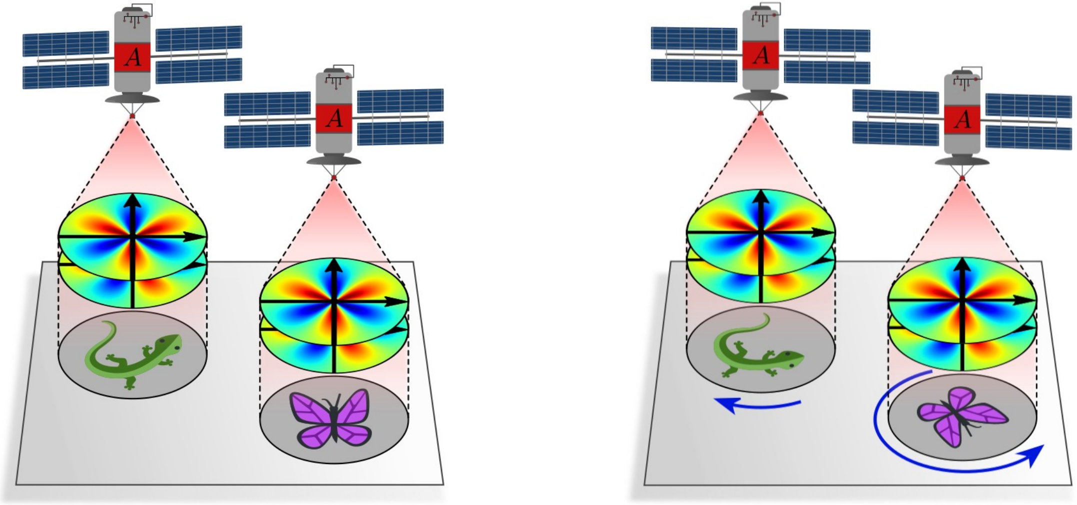


symmetry properties:

1) *GM*-coordinate independence (passive)
2) local gauge equivariance (active)
3) global isometry equivariance (active)

GM-convolutions

GM-*coordinate independent convolutions* operate by applying a *convolutional kernel field* to a feature field



symmetry properties:

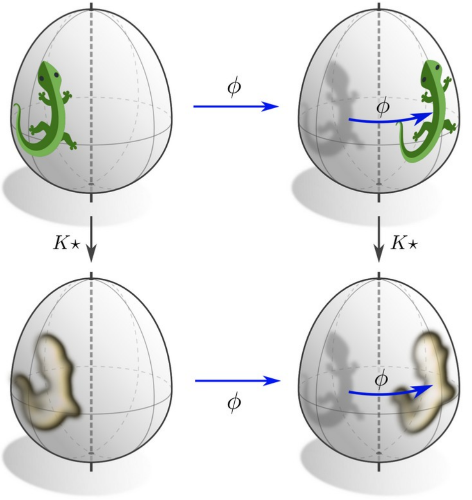
1) *GM*-coordinate independence (passive)

2) local gauge equivariance (active)

3) global isometry equivariance (active)

GM-convolutions

GM-coordinate independent convolutions operate by applying a *convolutional kernel field* to a feature field



symmetry properties:

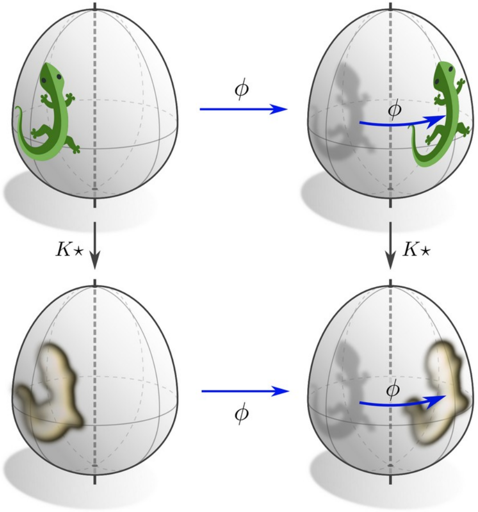
1) *GM*-coordinate independence (passive)
2) local gauge equivariance (active)
3) global isometry equivariance (active)

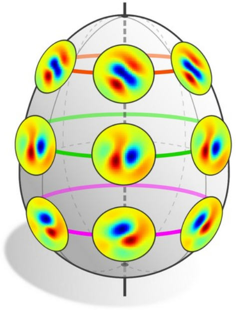
Isometry equivariance

"*kernel field transform*": similar to convolution, but not assuming weight sharing

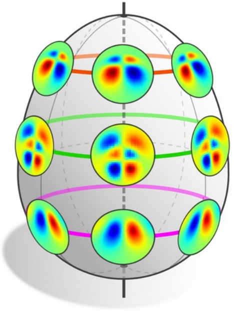
parameterized by a *kernel field*

Theorem: let $\mathcal{I} \leq \mathrm{Isom}(M)$, then: $\mathcal{I}$-*equivariant* kernel field transform $\iff$ $\mathcal{I}$-*invariant* kernel field





SO(2)-invariant kernel field



O(2)-invariant kernel field

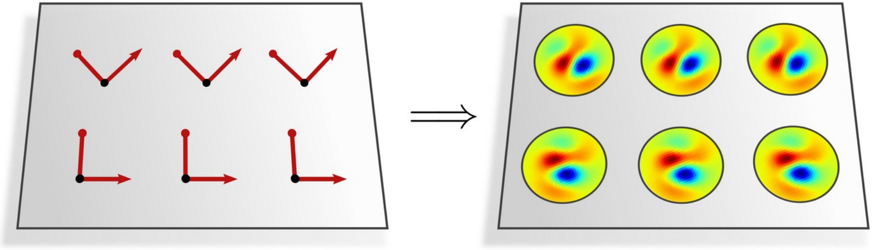
Isometry equivariance - *GM*-convolutions

Let $\mathrm{Isom}_{GM} \leq \mathrm{Isom}(M)$ be the subgroup of isometries that are symmetries of *GM*

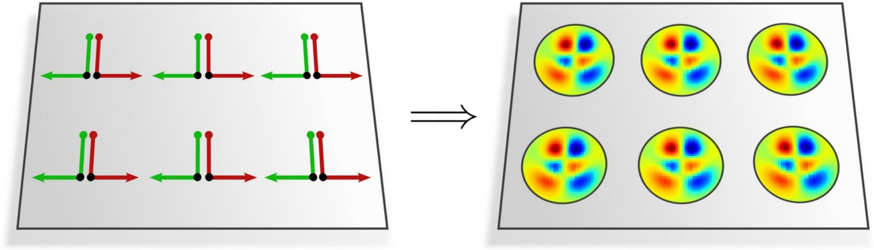
G-steerable (convolutional) kernel fields inherit this Isom_{GM}-invariance

$\Longrightarrow$ *GM*-convolutions are Isom_{GM}-equivariant

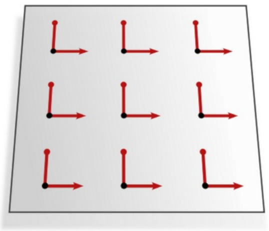
- horizontal translations



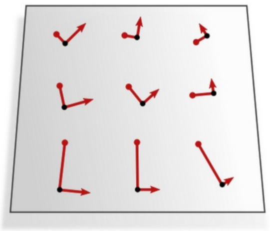
- horizontal translations
- vertical translations
- horizontal reflections



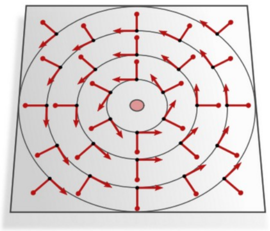
Isometry equivariance - *GM*-convolutions



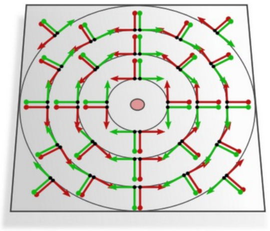
$M = \mathbb{R}^2,\ G = \{e\}$



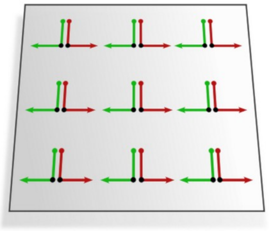
$M = \mathbb{R}^2,\ G = \{e\}$



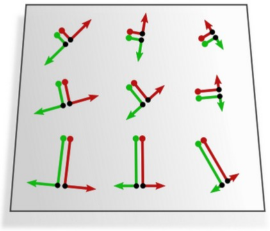
$M = \mathbb{R}^2 \backslash \{0\},\ G = \{e\}$



$M = \mathbb{R}^2 \backslash \{0\},\ G = \mathscr{R}$

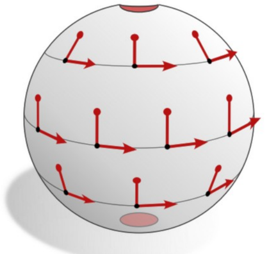


$M = \mathbb{R}^2,\ G = \mathscr{R}$

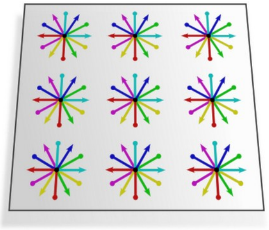


$M = \mathbb{R}^2,\ G = \mathscr{R}$

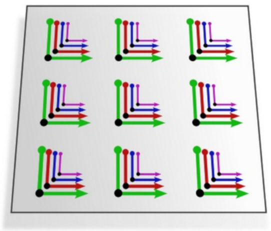
$M = S^2,\ G = \mathrm{SO}(2)$



$M = S^2 \backslash \text{poles},\ G = \{e\}$

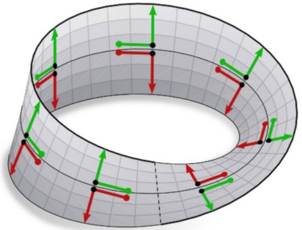


$M = \mathbb{R}^2,\ G = \mathrm{SO}(2)$



$M = \mathbb{R}^2,\ G = \mathscr{S}$

$M = \text{“Suzanne”},\ G = \mathrm{SO}(2)$



$M = \text{Möbius},\ G = \mathscr{R}$

COORDINATE INDEPENDENT CONVOLUTIONAL NETWORKS

ISOMETRY AND GAUGE EQUIVARIANT CONVOLUTIONS ON RIEMANNIAN MANIFOLDS

1 Introduction 2

2 Overview and visual intuition 10

I An introduction to coordinate independent CNNs 21

3 Coordinate independent feature fields 22
- 3.1 Gauges, gauge transformations and G-structures . . . 22
- 3.2 Coordinate independent feature vector fields . . . 30
- 3.3 Parallel transport of feature vectors . . . 34
- 3.4 Isometry actions and induced gauge transformations . . . 37

4 Coordinate independent networks and *GM*-convolutions 40
- 4.1 Pointwise gauge equivariant operations . . . 41
- 4.2 Kernel field transforms and GM-convolutions . . . 45
- 4.3 Isometry equivariance . . . 54

5 Toy model: reflection equivariant Möbius convolutions 57
- 5.1 Geometry of the Möbius strip . . . 58
- 5.2 Orientation independent feature fields . . . 60
- 5.3 Orientation independent convolutional networks . . . 61
- 5.4 Numerical implementation and evaluation of Möbius convolutions . . . 66

II Theory of coordinate independent CNNs 73

6 Associated bundles and coordinate free feature fields 73
- 6.1 A brief introduction to fiber bundles . . . 74
- 6.2 The tangent bundle TM and frame bundle FM . . . 80
- 6.3 G-structures GM and associated feature vector bundles $\mathcal{A}$ . . . 83
- 6.4 Local bundle trivializations of TM, FM, GM and $\mathcal{A}$ . . . 86
- 6.5 Parallel transporters on associated bundles . . . 92

7 Coordinate free kernel field transforms and *GM*-convolutions 96
- 7.1 1×1 GM-convolutions . . . 96
- 7.2 Kernel field transforms and GM-convolutions . . . 101

8 Isometry equivariance 109
- 8.1 Isometries and their action on manifolds, bundles and fields . . . 110
- 8.2 Isometry equivariance of kernel field transforms and GM-convolutions . . . 123
- 8.3 Quotient kernel fields . . . 129

III A literature review on coordinate independent CNNs 140

9 Euclidean coordinate independent CNNs 145
- 9.1 Classical formulation of G-steerable CNNs on $\mathbb{R}^d$ . . . 146
- 9.2 Affine geometry of Euclidean spaces $\mathbb{E}_d$ . . . 149
- 9.3 Affine group equivariant CNNs on Euclidean spaces $\mathbb{E}_d$ . . . 155
- 9.4 Euclidean CNNs in the literature . . . 158

10 Rotation equivariant CNNs on punctured Euclidean spaces 161

11 Spherical coordinate independent CNNs 169
- 11.1 Geometry of the 2-sphere S^2 . . . 171
- 11.2 Fully rotation equivariant spherical CNNs . . . 174
- 11.3 Azimuthal rotation equivariant spherical CNNs on cylindrical topologies . . . 179
- 11.4 Icosahedral approximations of spherical CNNs . . . 184

12 Coordinate independent CNNs on general surfaces 189
- 12.1 Geometry of embedded surfaces . . . 191
- 12.2 Rotation-steerable surface convolutions . . . 199
- 12.3 $\{e\}$-steerable surface convolutions . . . 207

IV Conclusion 212

V Appendix 216

A Relation to the coordinate chart formalism of differential geometry 216
- A.1 Tangent spaces, cotangent spaces and dual bases . . . 217
- A.2 Differentials, gradients and Jacobians . . . 218
- A.3 Chart induced coordinate bases . . . 220
- A.4 Coordinate bases as local bundle trivializations . . . 224
- A.5 G-structures and vielbein fields . . . 226

B Coordinate independent weight sharing and *G*-steerable kernels 230

C Integration over tangent spaces 231

D Equivariant convolutions on homogeneous spaces 232

E Quotient representative kernel fields – proofs 245

F Spherical steerable convolutions as *GM*-convolutions – proofs 251

G Existence and smoothness of kernel field transforms 255

H Regular feature fields as scalar functions on *G*-structure 258

mathematical structure

(group/representation theory & differential geometry)

physics	deep learning
tensor fields	feature fields
Minkowski space + global Poincaré symmetry	Euclidean space + global Aff(G) symmetry
curved spacetime + local Lorentz transformations	Riemannian manifold + local gauge transformations
invariant laws of nature (relativity)	invariant neural connectivity
equivariant system dynamics	equivariant inference
scalar / vector / tensor operators in QM	G-steerable kernels
quantum state transition rules	feature transition rules

Thank you

Maurice Weiler

AMLab, QUVA Lab

University of Amsterdam

@maurice_weiler