Equivariant & coordinate independent CNNs on Riemannian manifolds

Maurice Weiler AMLab, QUVA Lab University of Amsterdam

😏 @maurice_weiler







Principle of Covariance (Einstein, 1916)

"Universal **laws of nature** are to be expressed by equations which hold good for all systems of coordinates [...] "





"The *inference of CNNs* is to be expressed by equations which hold good for all systems of coordinates [...] "

Equivariant convolutions



active transformations - acting on the data itself:





equivariant NNs

passive transformations - acting on coordinatization of data:





$$\mathbb{N}$$

> coordinate independent CNNs

active transformations - acting on the data itself:



Outline

Data gauges & gauge independent neural networks

Coordinate independent CNNs on Riemannian manifolds

physical systems or data often have no canonical mathematical description

a gauge has to be chosen arbitrarily

gauge theories ensure that their prediction is equivalent for any gauge

space / coordinates



potential / reference potential

graph / node indexing















physical mass has no canonical units

gauging: introduce units by choosing some *reference weight*

different gauges are equally valid - within some set of structurally distinguished gauges gauge transformations translate between these choices (= unit conversions) determines *qauge* group $\mathbb{R}_{>0}$ ψ^B $\mathbb{R}_{>0}$ ψ^A M $1/\frac{1}{k_{a}}$ $1/\frac{1}{100}$ scale free physical mass 1kg 1lbs unit conversion $q^{BA} := \psi^B \circ (\psi^A)^{-1}$ $\mathbb{R}_{>0},*)$

physical mass has no canonical units

gauging: introduce units by choosing some *reference weight*



Euclidean affine space gauging

Euclidean affine spaces \mathbb{E}_d are *coordinate free*

gauges = coordinate charts $x^A : \mathbb{E}_d \to \mathbb{R}^d$

gauge transformations = *chart transition maps* $x^B \circ (x^A)^{-1}$: $\mathbb{R}^d \to \mathbb{R}^d$

additional geometric structure map prefer a subset of charts and reduce the gauge group



(multi) sets have no canonical order

a *gauge* introduces an order of set elements, representing the set by a *tuple*

gauge transformations are permutations

the (multi) set can be viewed as an *equivalence class* (orbit) of tuples



Coordinate independent NN layers



Coordinate independent NN layers + equivariance



Coordinate independent NN layers + equivariance



on manifolds: independence from *local reference frames* (observers)

local gauge equivariance



on manifolds: independence from *local reference frames* (**observers**)

local gauge equivariance



Coordinate independent CNNs on Riemannian manifolds



Convolutions on Riemannian manifolds







spherical CNNs



artery wall stress estimation





general relativity

Euclidean CNNs



Image adapted from Konakovic-Lukovic et al.



how to ...

- ... define *feature fields* on M ?
- ... define *convolution kernels* on M ?

 \dots 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)



Weight sharing - via parallel transport

sharing weights by "shifting" kernel over manifold ?

1 parallel transport in general path dependent



Weight sharing - approaches in the literature

↔ topological obstructions to the existence of G-structures

solution approaches in the literature:

1) gauge invariant features 1 low expressiveness



- 2) heuristic gauges
- **1** instable under deformations



3) spectral approaches

gauge independent but instable under deformations

the kernel alignment ("gauge") on manifolds is inherently ambiguous!



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 $\ensuremath{\mathbb{R}}^2$



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



G-structures

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

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



G-structures

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

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



GM-coordinate independence - tangent vectors

all frames of the G-structure are equally valid

 \implies 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

 \implies any object or morphism should be expressible relative to any frame in GM

coordinate independent feature vectors transform according to group representation ρ :

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

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	

GM-coordinate independence - feature vector fields

all frames of the G-structure are equally valid

 \implies any object or morphism should be expressible relative to any frame in GM

coordinate independent feature vectors transform according to group representation ρ :

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

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



GM-coordinate independence - feature vector fields

all frames of the G-structure are equally valid

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

coordinate independent feature vectors transform according to group representation ρ :

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

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

feature vector fields are bundle sections == a choice of feature vector at each point $p \in M$

GM-coordinate independence - linear maps on $T_{p}M_{p}$

all frames of the G-structure are equally valid

 \implies any object or morphism should be expressible relative to any frame in GM

- example: linear maps $\mathcal{M}: T_p M \to T_p M$ 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-convolutions

we want to design a convolution operation that is

- 1) mapping a $ho_{\mathrm{in}} ext{-field}$ to a $ho_{\mathrm{out}} ext{-field}$
- 2) parameterized by a spatially shared kernel
- 3) GM-coordinate independent

how to define kernels / kernel fields?

in coordinates, a convolution kernel is a map

$$\Rightarrow$$
 coordinate free, at $p \in M$: $\mathcal{K}_p : T_pM \to \operatorname{Hom}(\mathcal{A}_{\operatorname{in},p}, \mathcal{A}_{\operatorname{out},p})$

coordinate free, global (kernel field):



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





($\mathcal{A}_{\mathrm{in/out}}$ are the feature bundles)

(allowing potentially for a different kernel at each $p \in M$, i.e. here pre-weight-sharing)

convolutional kernel fields should have the same kernel at each position

GM-convolutions

relation between coordinate free kernels \mathcal{K}_p and their coordinate expressions \mathcal{K}_p^A and \mathcal{K}_p^B :



in which gauge should we share a given template kernel? Ambiguous, need to share in all gauges X: $\mathcal{K}_n^X = K$

GM-coordinate independence spatial weight sharing

requires G-steerability
$$K=rac{1}{|\det q|}ig(
ho_{ ext{in}}^{- op}\otimes
ho_{ ext{out}}ig)(g)\circ K\circ g^{-1}$$

given ...

- ... a Riemannian manifold M
- ... a G-structure GM
- ... a G-compatible connection
- ... a G-steerable convolution kernel K

the corresponding *GM*-convolution is performed by:

- 1) applying "transporter pullbacks" Exp_p^* of the feature field to the tangent spaces
- 2) matching this pullback with the kernel $\ K$

$$f_{\text{out}}^{A}(p) := \int_{\mathbb{R}^{d}} \mathrm{d}v^{A} K(v^{A}) \big[\operatorname{Exp}_{p}^{*} f_{\text{in}} \big]^{A}(v^{A})$$

the kernel's G-steerability ensures that the chosen gauge / kernel alignment is irrelevant



(a local observer's viewpoint)

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



Isometry equivariance - *GM*-convolutions

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

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

 \implies GM-convolutions are Isom_{GM} -equivariant



- horizontal translations

- horizontal translations
- vertical translations
- horizontal reflections

Isometry equivariance - *GM*-convolutions



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



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



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



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



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



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





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



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



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

M = "Suzanne", G = SO(2)



M =Möbius, $G = \mathcal{R}$

locally flat geometry

github: https://github.com/mauriceweiler/MobiusCNNs

regular pixel grid

Levi-Civita connection has holonomy group \mathbb{Z}_2 (reflections)

admits \mathbb{Z}_2 -structure



cut and flatten the strip

parallel transport: - trivial within strip

- "reflection padding" over the cut

after transport padding: Euclidean convolution with \mathbb{Z}_2 -steerable kernels



transport pad $\rho(s)$

Möbius CNNs - experiments

Möbius MNIST toy dataset



performance depends on field types

perfect isometry equivariance

model	f	ield types	$s ho_i$	params	test error (%)	
	trivial	sign-flip	regular		shifted train digits	centered train digits
CNN (channels)				$1501\mathrm{k}$	1.97 ± 0.11	42.99 ± 2.65
CNN (params)				$832\mathrm{k}$	2.08 ± 0.10	43.68 ± 2.85
gauge CNN (scalar)	\checkmark	×	×	902 k	1.60 ± 0.10	1.60 ± 0.09
gauge CNN (sign-flip)	×	\checkmark	×	820 k	4.27 ± 0.24	4.89 ± 0.36
gauge CNN (regular)	×	×	\checkmark	$752\mathrm{k}$	1.24 ± 0.08	1.23 ± 0.07
gauge CNN (irreps)	\checkmark	\checkmark	×	$752\mathrm{k}$	1.65 ± 0.09	1.64 ± 0.12
gauge CNN (mixed)	\checkmark	\checkmark	\checkmark	$752\mathrm{k}$	1.43 ± 0.09	1.42 ± 0.10

Icosahedral CNNs - geometry

Platonic solid

locally flat approximation of the sphere

hexagonal grid

Levi-Civita connection has holonomy group \mathbb{Z}_6

admits $\mathbb{Z}_6\text{-structure}$



flatten icosahedron by cutting it at north and south pole

parallel transport non-trivial over cut edges



Icosahedral CNNs - implementation

flatten icosahedron by cutting it at north and south pole

parallel transport non-trivial over cut edges



Icosahedral CNNs - experiments

kernel	N/N	N/I	I/ I
non-steerable	99.42	25.41	98.67
\mathbb{Z}_6 -steerable	99.43	99.43	99.38

climate pattern segmentation

Model	BG	TC	AR	Mean	mAP
Mudigonda et al.	97	74	65	78.67	-
Jiang et al.	97	94	93	94.67	-
Ours (S2R)	97.3	97.8	97.3	97.5	0.686
Ours (R2R)	97.4	97.9	97.8	97.7	0.759



omnidirectional RGB-D image segmentation

	mAcc	mIoU
(Jiang et al., 2018) Ours (B2B -U-Net)	0.547	0.383 0 30 4
Ours (R2R-U-Net)	0.559	0.394



Image credits: Jiang et al., 2018

Gauge Equivariant Mesh CNNs

Gauge Equivariant Mesh CNNs Anisotropic convolutions on geometric graphs

Pim de Haan *123 Maurice Weiler *23 Taco Cohen 1 Max Welling 13



Pim de Haan

Mesh Convolutional Neural Networks for Wall Shear Stress Estimation in 3D Artery Models

Julian Suk¹, Pim de Haan^{3,2}, Phillip Lippe², Christoph Brune¹, and Jelmer M. Wolterink¹





manifold structure group global symmetry representation citation Literature review MGAffen or Isomen 0 \mathbb{E}_d $\{e\}$ \mathcal{T}_d [130, 253] + any conventional CNN trivial \mathbb{E}_1 $\mathcal{T}_1 \rtimes S$ S regular [186] R $\mathcal{T}_2 \rtimes \mathcal{R}$ [234] regular [244, 234, 231] irreps [51, 33, 257, 34, 236, 8, 93, 192] [234, 79, 125, 210, 232, 185, 158] regular **Euclidean steerable CNNs** [201, 7, 67, 227, 83, 159, 231, 92] Coordinate independent CNNs SO(2)SE(2)[50, 206, 19, 207, 208, 164, 29, 86] [34, 234] quotients [33, 143, 234] unify a wide range of related work: [144, 234] 8 \mathbb{E}_2 trivial [110, 234] irreps [234] [51, 33, 93, 34, 234] regular [159, 79, 201] O(2)E(2)quotients [34] 12 * L IL [234] induced SO(2)-irreps [234] 14 L L regular [243, 212, 7, 258] punctured Euclidean $\mathcal{T}_2 \rtimes S$ regular pool trivial 16 [77] L irreps [235, 224, 156, 120, 2, 6] quaternion [250] 18 SO(3)**SE(3)** regular [67, 241, 242] regular mool trivial [3] regular [241] \mathbb{E}_3 O(3)E(3)quotient O(3)/O(2)[103] 22 irrep morm trivial [174] C_4 $\mathcal{T}_3 \rtimes C_4$ regular [219] 24 D_4 $\mathcal{T}_3 \rtimes D_4$ regular [219] spherical / icosahedral $\mathcal{T}_d \rtimes \mathrm{SO}(d-1,1)$ $\mathbb{E}_{d-1,1}$ SO(d-1, 1)irreps [205] 26 SO(2)[30, 67] $\mathbb{E}_2 \setminus \{0\}$ $\{e\}$ trivial $SO(2) \times S$ [62, 67] 28 O(2)O(3)trivial [178] $\mathbb{E}_3 \setminus \{0\}$ $\{e\}$ $\{e\}$ trivial [13] irreps [122, 64] SO(2)SO(3) S^2 regular [35, 111] 32 general O(2)O(3)trivial [61, 169, 245] [39, 222, 254, 149, 105] 2d surfaces / $S^2 \setminus \text{poles}$ $\{e\}$ SO(2)trivial [217, 218, 55, 131] C_6 icosahedron $I \approx SO(3)$ regular [38] meshes ico \ poles $\{e\}$ $C_5 (\approx SO(2))$ trivial [251, 139] [238] irreps [173, 220, 246, 48] SO(2) $\operatorname{Isom}_{+}(M)$ regular 3.8 surface (d=2)[150, 151, 160, 220] (e.g. meshes) D_4 $Isom_{D_4M}$ trivial [98] Möbius $\{e\}$ trivial [160, 194, 106, 221, 133] Isom_{(e)M} irreps R Möbius strip SO(2)Section 5 regular

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 3.1 Gauges, gauge transformations and G-structures 3.2 Coordinate independent feature vector fields 3.3 Parallel transport of feature vectors 3.4 Isometry actions and induced gauge transformations Coordinate independent networks and GM-convolutions	 22 30 34 37 40
	4.1 Pointwise gauge equivariant operations 4.2 Kernel field transforms and <i>GM</i> -convolutions 4.3 Isometry equivariance	41 45 54
5	Toy model: reflection equivariant Möbius convolutions 5.1 Geometry of the Möbius strip . 5.2 Orientation independent feature fields . 5.3 Orientation independent convolutional networks . 5.4 Numerical implementation and evaluation of Möbius convolutions .	57 58 60 61 66
п	Theory of coordinate independent CNNs	73
6	Associated bundles and coordinate free feature fields 6.1 A brief introduction to fiber bundles 6.2 The tangent bundle TM and frame bundle FM 6.3 G-structures GM and associated feature vector bundles A 6.4 Local bundle trivializations of TM, FM, GM and A 6.5 Parallel transporters on associated bundles	73 74 80 83 86 92
7	Coordinate free kernel field transforms and GM-convolutions 7.1 1×1 GM-convolutions 7.2 Kernel field transforms and GM-convolutions	96 96 101
8	Isometry equivariance 8.1 Isometries and their action on manifolds, bundles and fields	109 110 123 129

п	I A literature review on coordinate independent CNNs	140
9	Euclidean coordinate independent CNNs 9.1 Classical formulation of G -steerable CNNs on \mathbb{R}^d 9.2 Affine geometry of Euclidean spaces \mathbb{E}_d 9.3 Affine group equivariant CNNs on Euclidean spaces \mathbb{E}_d 9.4 Euclidean CNNs in the literature	145 146 149 155 158
10	Rotation equivariant CNNs on punctured Euclidean spaces	161
11	Spherical coordinate independent CNNs 11.1 Geometry of the 2-sphere S ² 11.2 Fully rotation equivariant spherical CNNs 11.3 Azimuthal rotation equivariant spherical CNNs on cylindrical topologies 11.4 Icosahedral approximations of spherical CNNs Coordinate independent CNNs	169 171 174 179 184
12	12.1 Geometry of embedded surfaces 12.2 Rotation-steerable surface convolutions 12.3 {e}-steerable surface convolutions	199 [207]
IX	/ Conclusion	0.1.0
13	Conclusion	212
v	Appendix	212
V A	Appendix Relation to the coordinate chart formalism of differential geometry A.1 Tangent spaces, cotangent spaces and dual bases A.2 Differentials, gradients and Jacobians A.3 Chart induced coordinate bases A.4 Coordinate bases as local bundle trivializations A.5 G-structures and vielbein fields	212 216 216 . 217 . 218 . 220 . 224 . 224
V A B	Appendix Relation to the coordinate chart formalism of differential geometry A.1 Tangent spaces, cotangent spaces and dual bases A.2 Differentials, gradients and Jacobians A.3 Chart induced coordinate bases A.4 Coordinate bases as local bundle trivializations A.5 G-structures and vielbein fields Coordinate independent weight sharing and G-steerable kernels	212 216 217 217 218 218 220 220 224 224 226 230
V A B C	Appendix Relation to the coordinate chart formalism of differential geometry A.1 Tangent spaces, cotangent spaces and dual bases A.2 Differentials, gradients and Jacobians A.3 Chart induced coordinate bases A.4 Coordinate bases as local bundle trivializations A.5 G-structures and vielbein fields Coordinate independent weight sharing and G-steerable kernels Integration over tangent spaces	212 216 217 217 217 218 220 224 224 226 230 231
V A B C D	Appendix Relation to the coordinate chart formalism of differential geometry A.1 Tangent spaces, cotangent spaces and dual bases A.2 Differentials, gradients and Jacobians A.3 Chart induced coordinate bases A.4 Coordinate bases as local bundle trivializations A.5 G-structures and vielbein fields Coordinate independent weight sharing and G-steerable kernels Integration over tangent spaces Equivariant convolutions on homogeneous spaces	212 216 217 217 218 220 220 220 220 230 230 231 232
V A B C D E	Appendix Relation to the coordinate chart formalism of differential geometry A.1 Tangent spaces, cotangent spaces and dual bases A.2 Differentials, gradients and Jacobians A.3 Chart induced coordinate bases A.4 Coordinate bases as local bundle trivializations A.5 G-structures and vielbein fields Coordinate independent weight sharing and G-steerable kernels Integration over tangent spaces Equivariant convolutions on homogeneous spaces Quotient representative kernel fields – proofs	212 216 216 217 217 218 220 220 224 230 231 232 232 245
V A B C D E F	Appendix Relation to the coordinate chart formalism of differential geometry A.1 Tangent spaces, cotangent spaces and dual bases A.2 Differentials, gradients and Jacobians A.3 Chart induced coordinate bases A.4 Coordinate bases as local bundle trivializations A.5 G-structures and vielbein fields Coordinate independent weight sharing and G-steerable kernels Integration over tangent spaces Equivariant convolutions on homogeneous spaces Quotient representative kernel fields – proofs Spherical steerable convolutions as GM-convolutions – proofs	212 216 217 217 217 218 220 220 220 220 230 230 231 232 232 232 232

258

H Regular feature fields as scalar functions on G-structure