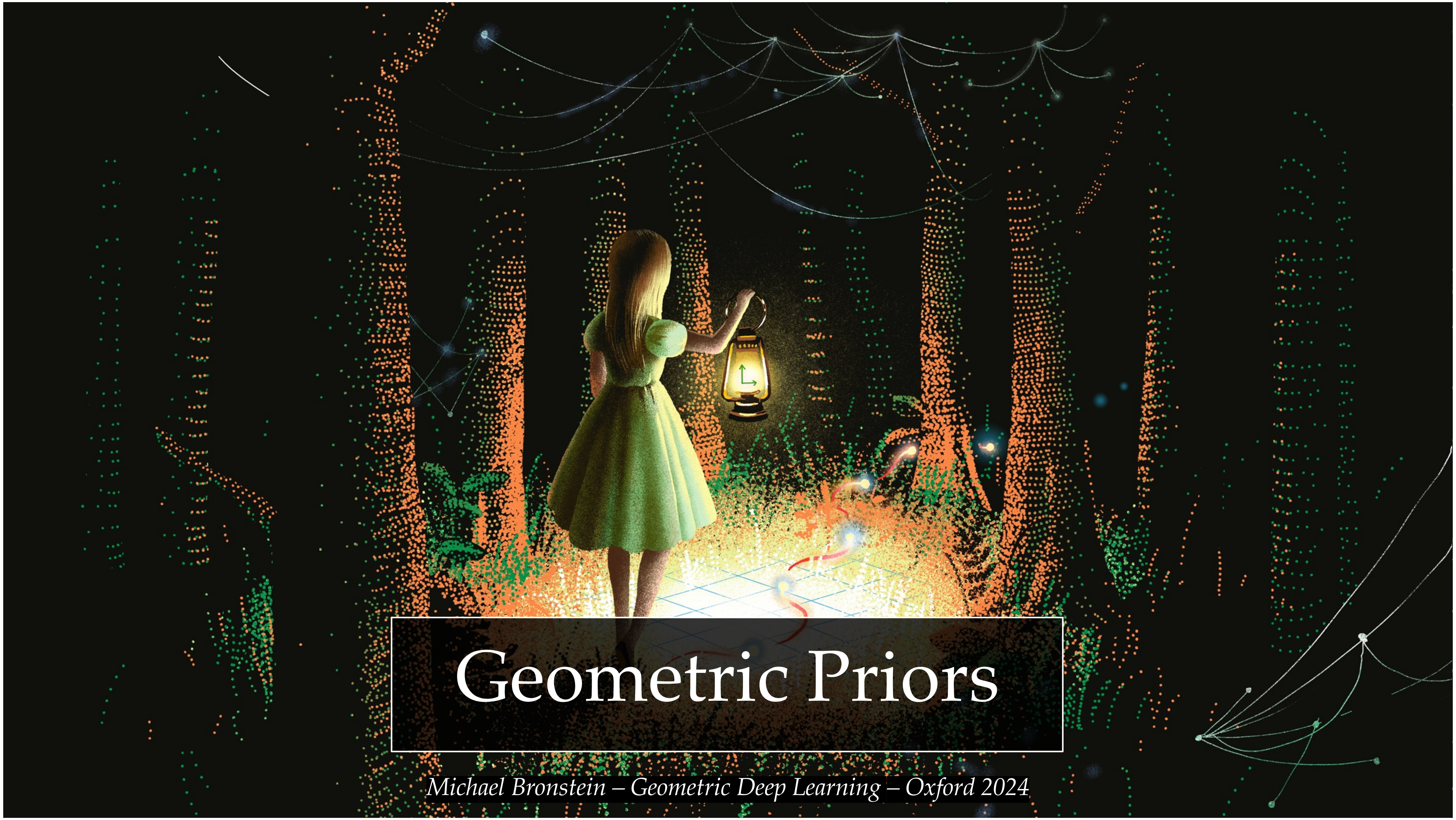


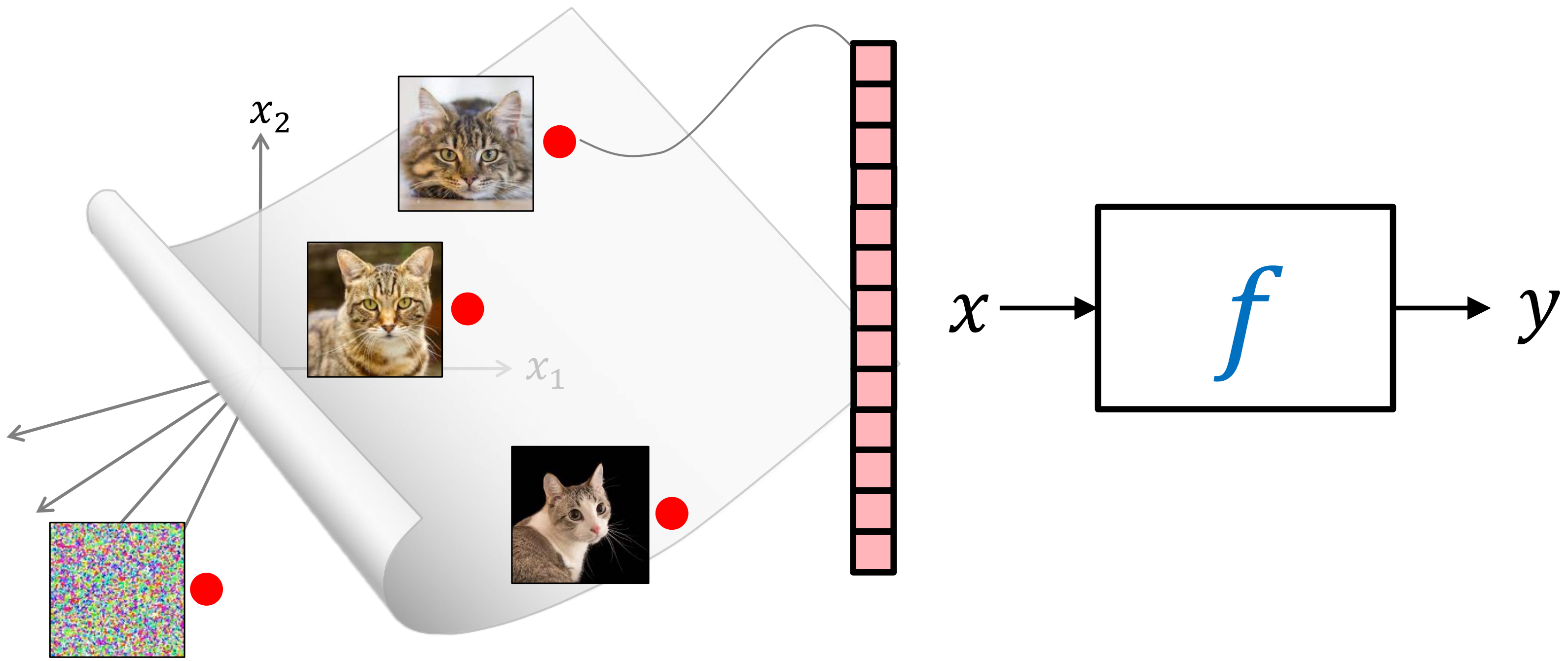
Geometric Priors

Michael Bronstein – Geometric Deep Learning – Oxford 2024

Outline

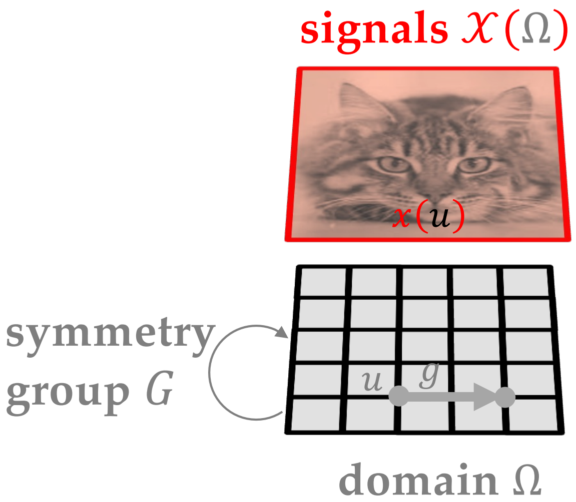
- Geometric priors in ML problems: transformations (symmetries) of the input space that leave the output *invariant*
- Mathematically, symmetries are structure-preserving transformations forming a *group* (a central object of study in Group Theory)
- Groups act on data via *group representations* (a central object of study in Representation Theory)
- To exploit symmetries in neural networks, we use *invariant* and *equivariant layers*

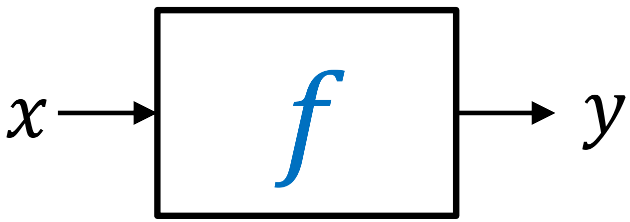
The Curse of Dimensionality



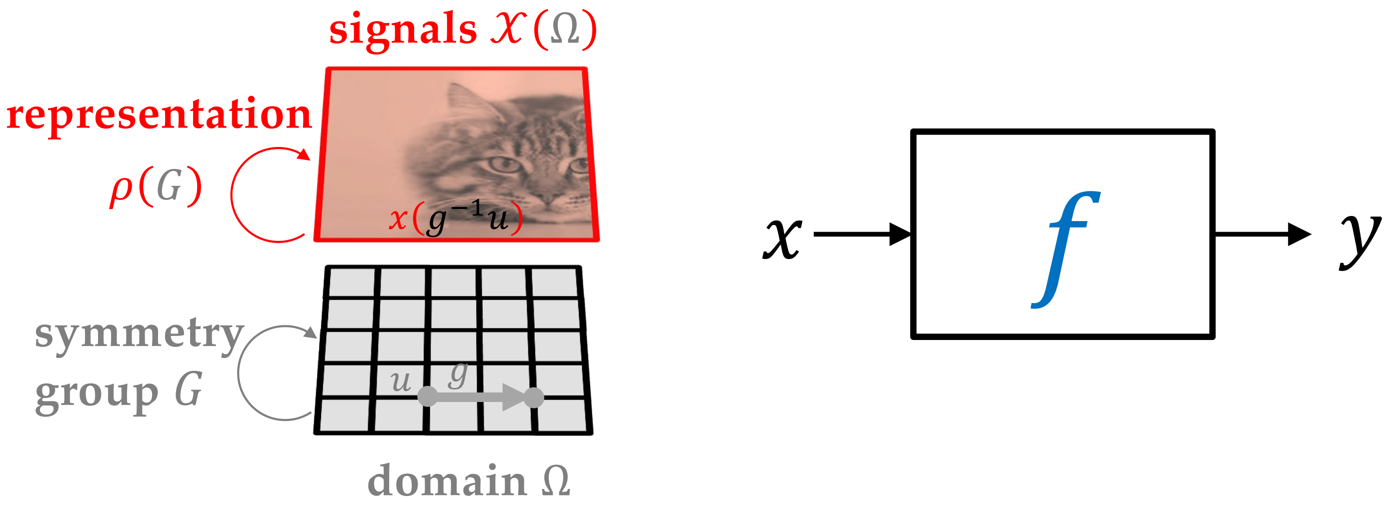
Approximating a d-dimensional Lipschitz function with accuracy ε requires $\mathcal{O}(\varepsilon^{-d})$ samples

Geometric priors

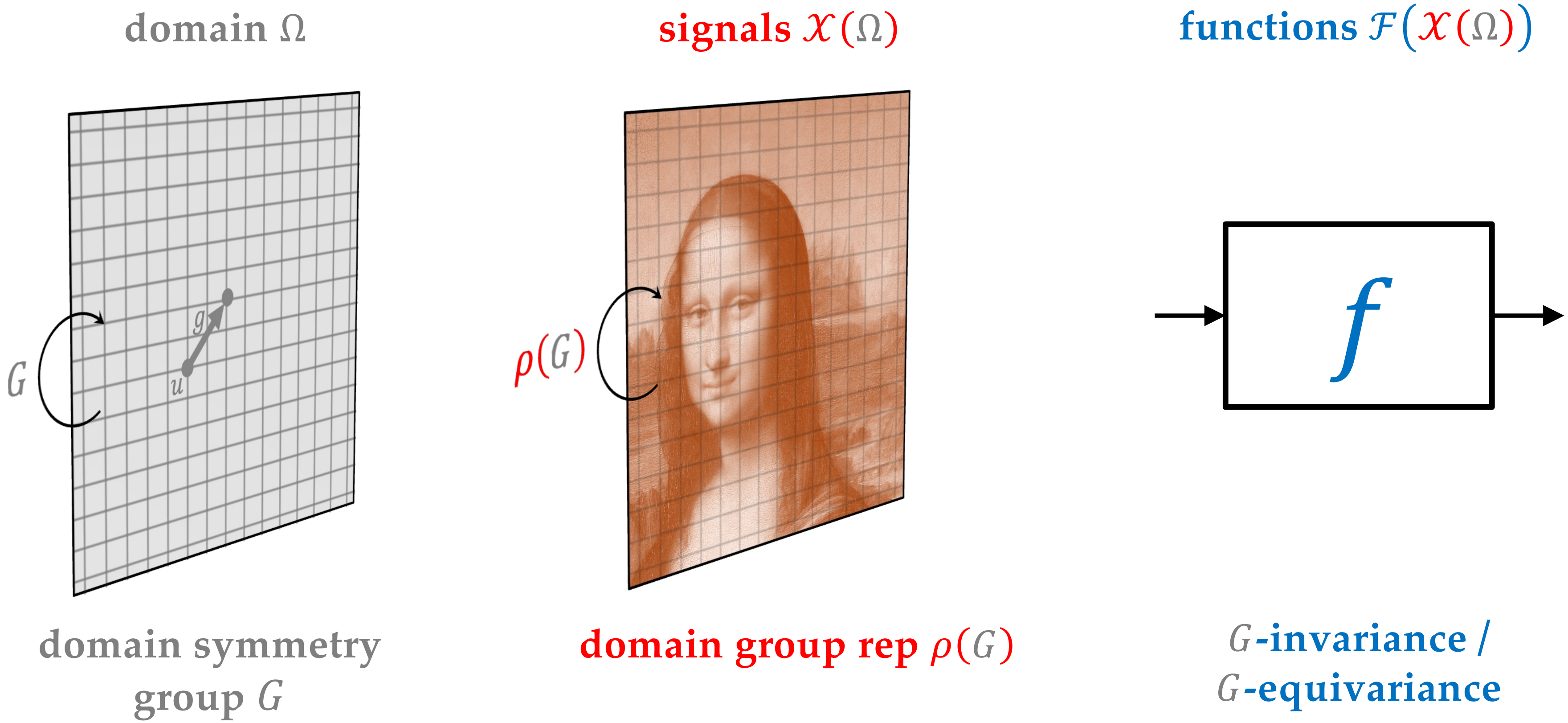




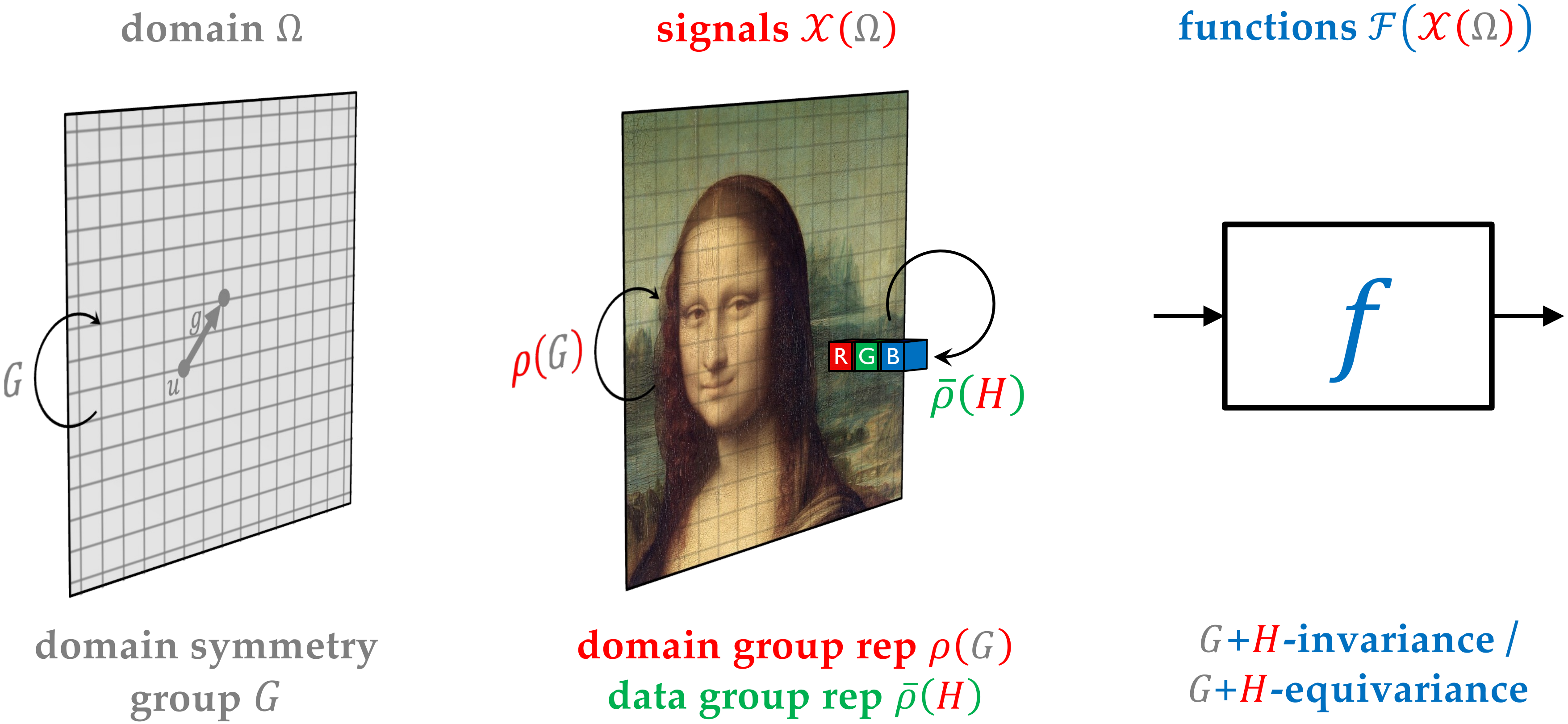
Geometric priors

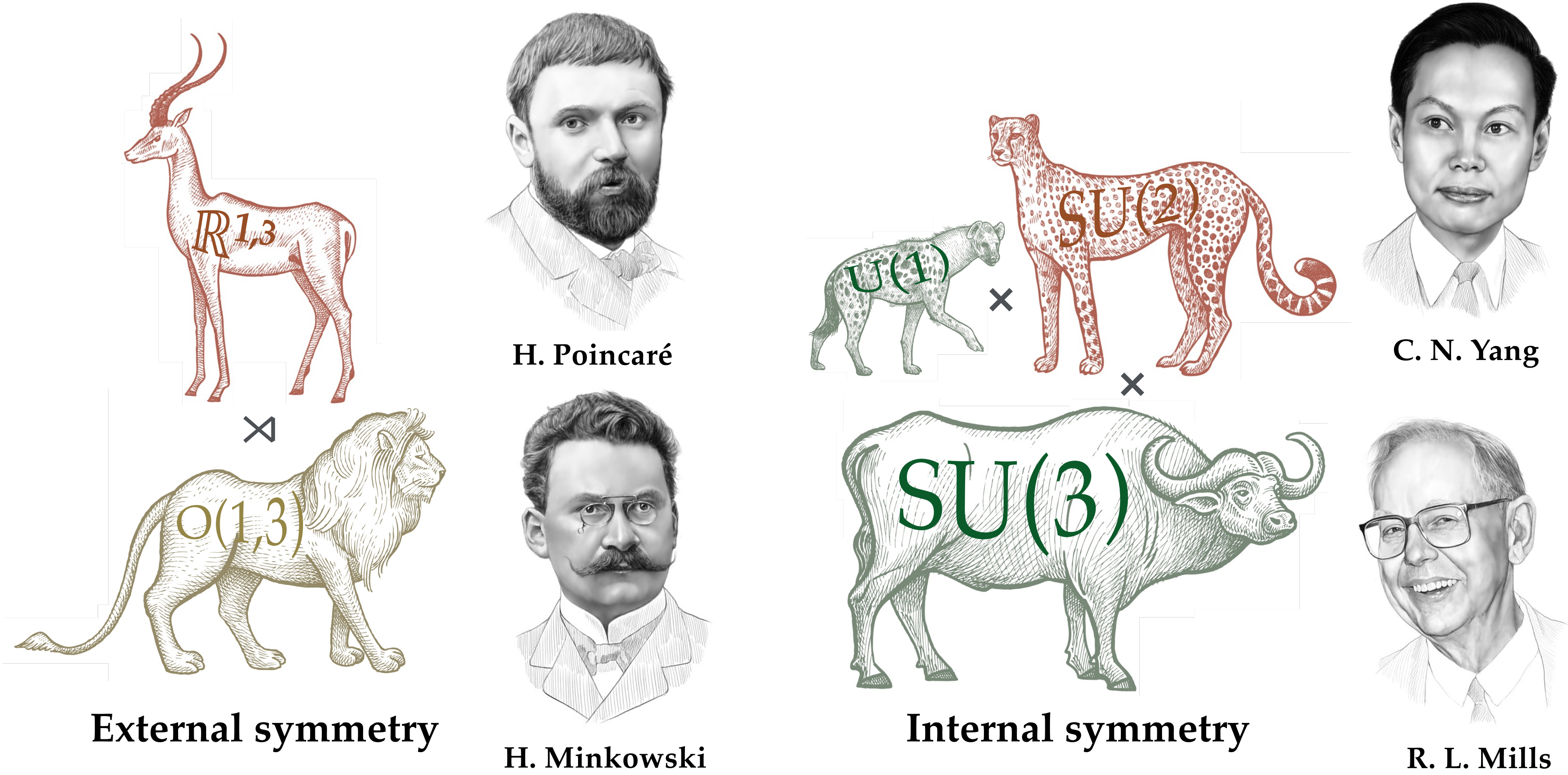


Key ingredients of Geometric Deep Learning



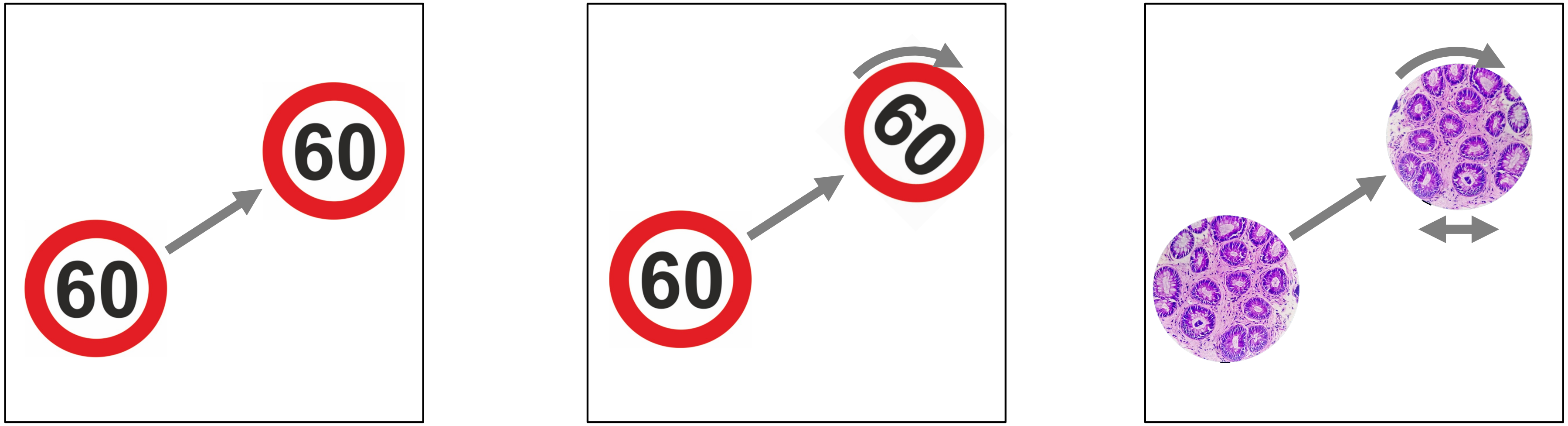
Key ingredients of Geometric Deep Learning





$\mathbb{R}^{1,3}$
⋊
O(1,3)
H. Poincaré
H. Minkowski
External symmetry
U(1)
×
SU(2)
×
SU(3)
C. N. Yang
R. L. Mills
Internal symmetry

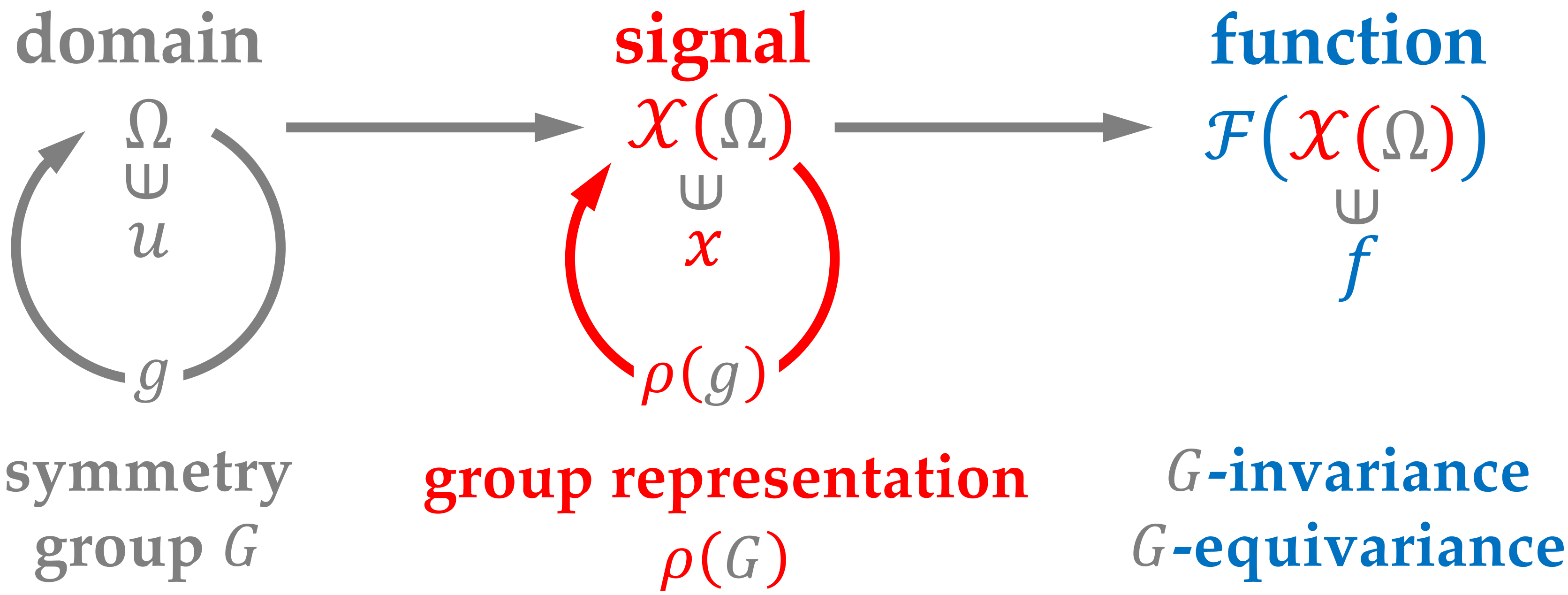
How to choose the symmetry group?



Self-driving car
Translation

Self-flying plane
Translation + Rotation

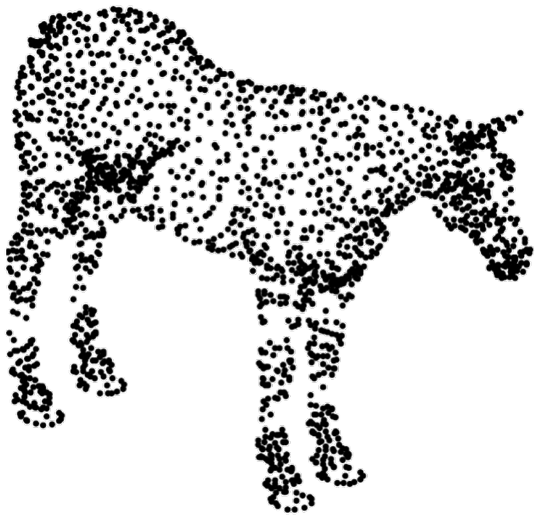
Pathology
Translation + Rotation + Reflection


domain
Ω
⋓
u
g
symmetry
group G
signal
X(Ω)
⋓
x
ρ(g)
group representation
ρ(G)
function
F(X(Ω))
⋓
f
G-invariance
G-equivariance

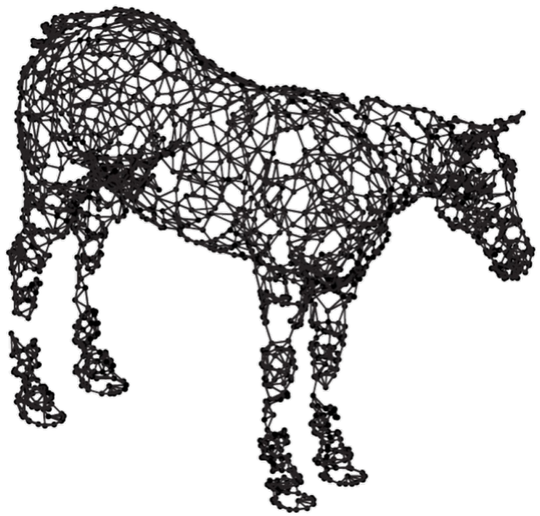
GEOMETRIC DOMAINS

Geometric domains

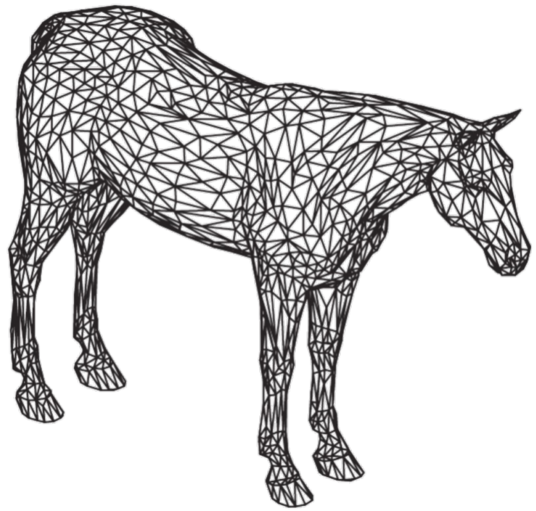
- **Domain** Ω = set + some structure



Point cloud
(bare set)



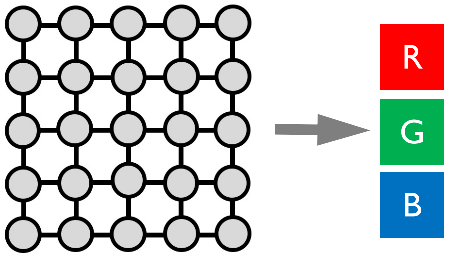
Graph
(local neighbourhood)



Mesh
(local metric)

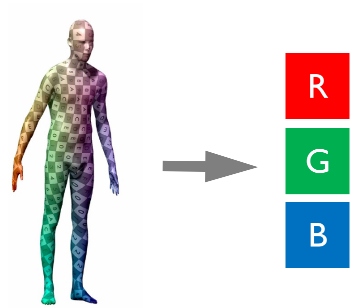
Signals on Geometric domains

- **Signal** $x \in \mathcal{X}(\Omega, \mathcal{C}) = \{x: \Omega \rightarrow \mathcal{C}\}$ "$\mathcal{C}$-valued functions on Ω"
  - *Domain* Ω
  - *Vector space* $\mathcal{C}$ (dimensions referred to as "channels")



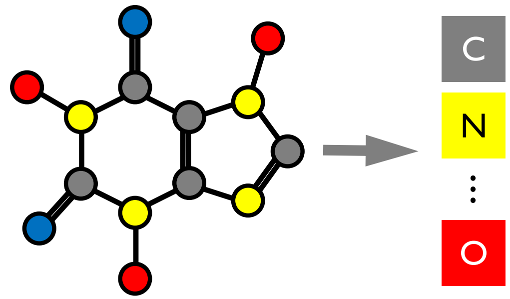
$\Omega = \mathbb{Z}_n \times \mathbb{Z}_n$ $\quad$ $\mathcal{C} = \mathbb{R}^3$

Image



$\Omega = \mathcal{M}$ $\quad$ $\mathcal{C} = \mathbb{R}^3$

Textured surface

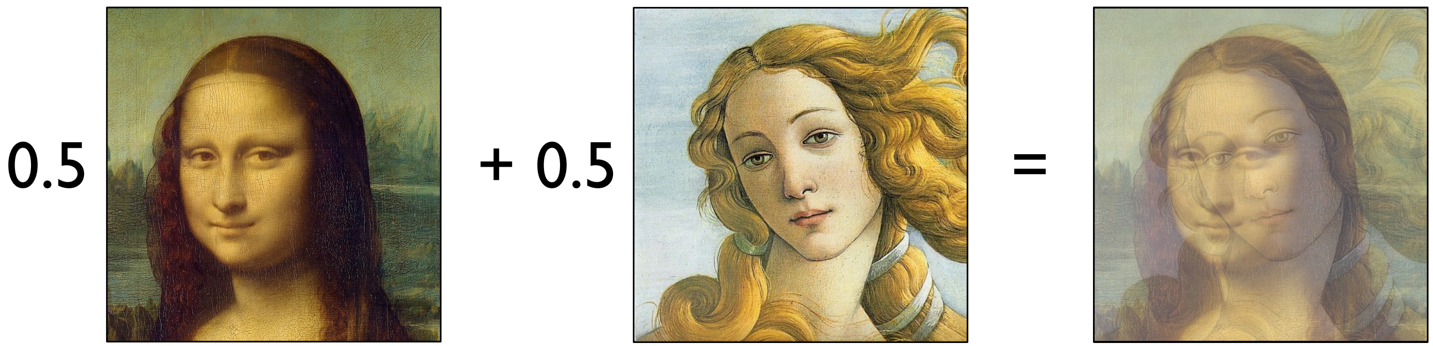


$\Omega = (V, E)$ $\quad$ $\mathcal{C} = \mathbb{R}^m$

Molecular graph

Signals on Geometric domains

- **Signal** $x \in \mathcal{X}(\Omega, \mathcal{C}) = \{x : \Omega \rightarrow \mathcal{C}\}$ "$\mathcal{C}$-valued functions on Ω"
  - *Domain* Ω (often no vector space structure, i.e., we cannot add points on Ω)
  - *Vector space* $\mathcal{C}$ (dimensions referred to as "channels")
- The **space of signals** $\mathcal{X}(\Omega, \mathcal{C})$ is a *vector space* (possibly infinite-dimensional)
  - We can add signals and multiply them by a scalar

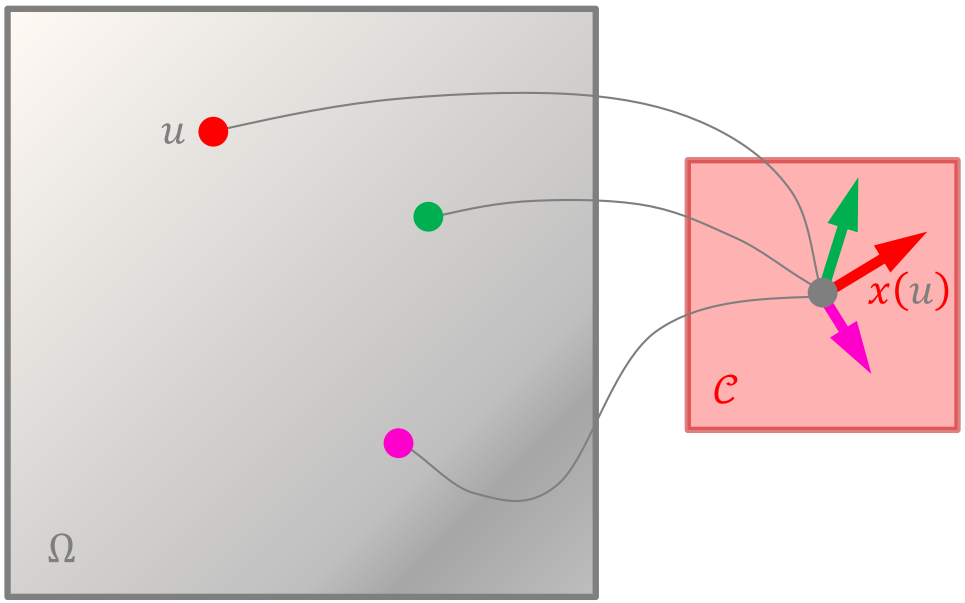


Signals on Geometric domains

- **Signal** $x \in \mathcal{X}(\Omega, \mathcal{C}) = \{x: \Omega \rightarrow \mathcal{C}\}$ "$\mathcal{C}$-valued functions on Ω"
  - *Domain* Ω
  - *Vector space* $\mathcal{C}$ (dimensions referred to as "channels")
- The **space of signals** $\mathcal{X}(\Omega, \mathcal{C})$ is a *vector space* (possibly infinite-dimensional)
- Given an *inner product* $\langle \cdot,\cdot \rangle_{\mathcal{C}}$ on $\mathcal{C}$ and a *measure* μ on Ω, we can define an inner product on $\mathcal{X}(\Omega, \mathcal{C})$ as

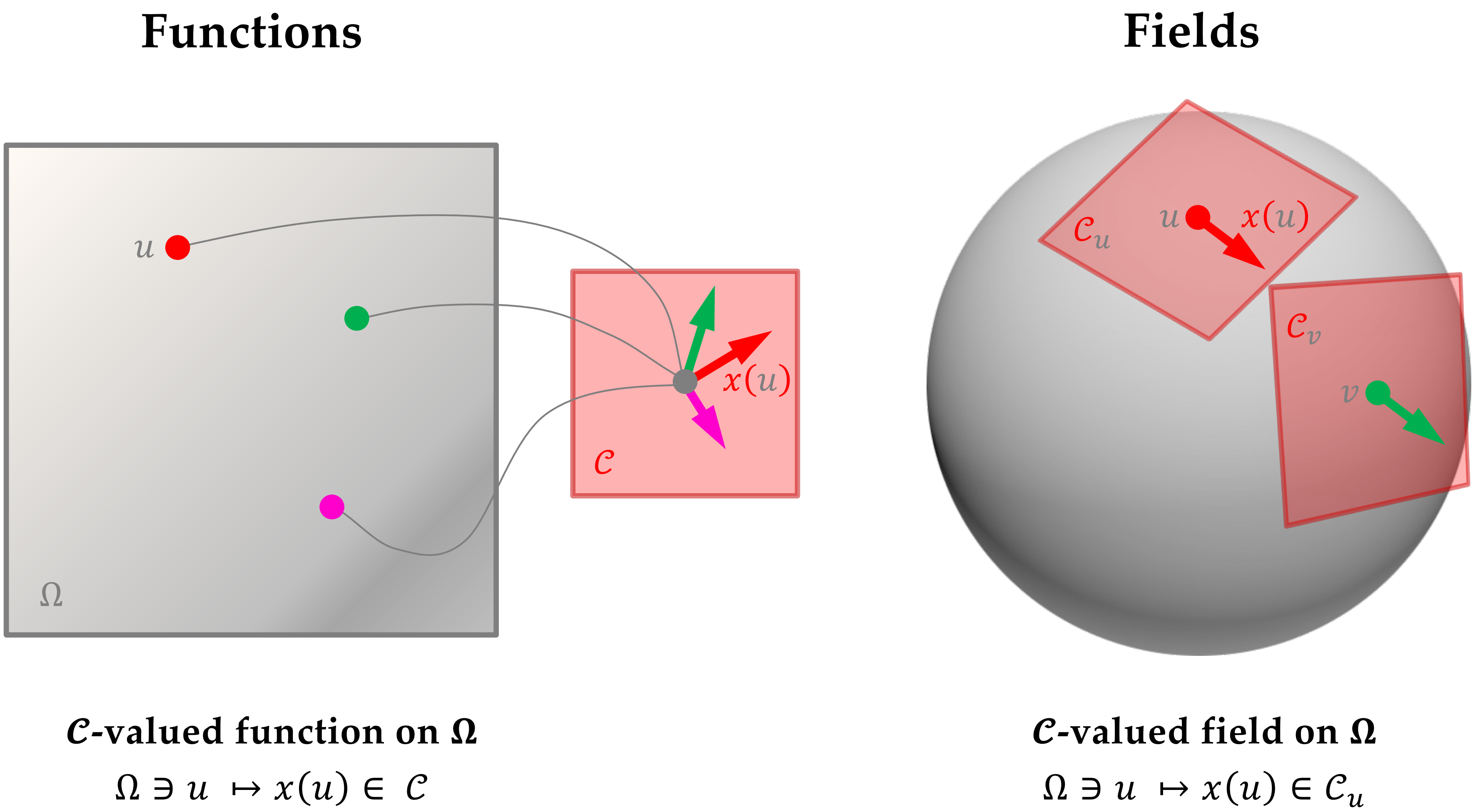
$$\langle x, y \rangle = \int_{\Omega} \langle x(u), y(u) \rangle_{\mathcal{C}} \, \mathrm{d}\mu(u)$$

Exercise: prove that $\langle x, y \rangle$ defined this way satisfies the axioms of an inner product



$\mathcal{C}$-valued function on Ω

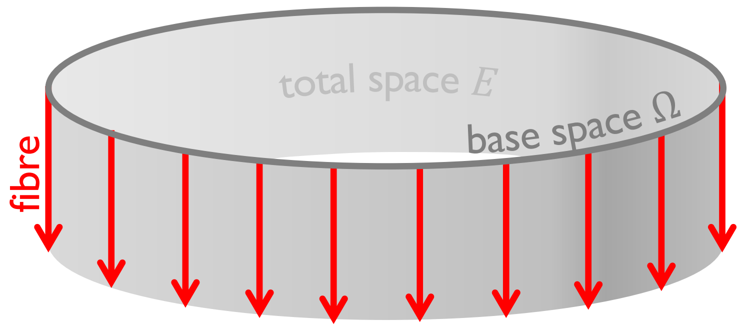
$$\Omega \ni u \mapsto x(u) \in \mathcal{C}$$



Functions
Fields
u
Ω
$\mathcal{C}$
$x(u)$
$\mathcal{C}_u$
u
$x(u)$
$\mathcal{C}_v$
v
$\mathcal{C}$-valued function on Ω
$\Omega \ni u \mapsto x(u) \in \mathcal{C}$
$\mathcal{C}$-valued field on Ω
$\Omega \ni u \mapsto x(u) \in \mathcal{C}_u$

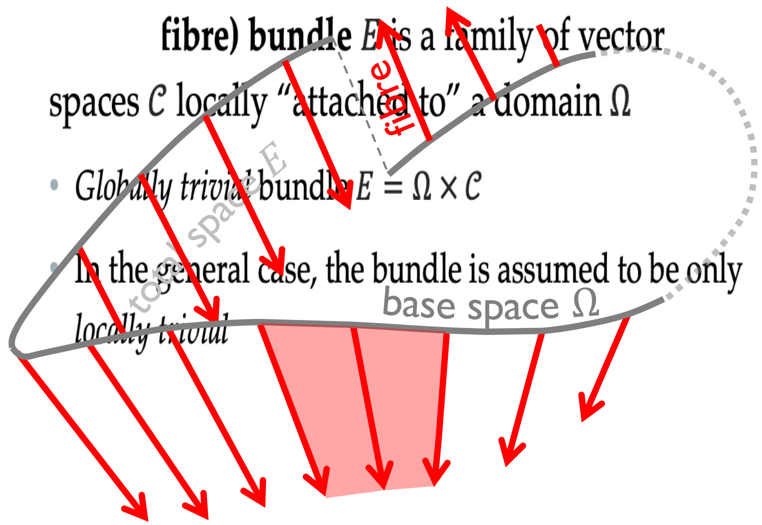
Fields on Geometric domains

- **Vector (fibre) bundle** E is a family of vector spaces $\mathcal{C}$ locally "attached to" a domain Ω
  - *Globally trivial* bundle $E = \Omega \times \mathcal{C}$



Fields on Geometric domains

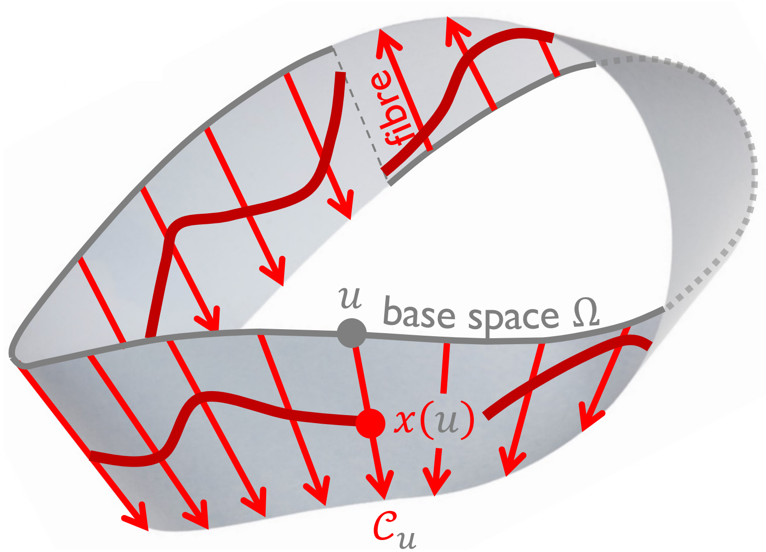
- **Vector (fibre) bundle** E is a family of vector spaces $\mathcal{C}$ locally "attached to" a domain Ω
  - *Globally trivial* bundle $E = \Omega \times \mathcal{C}$
  - In the general case, the bundle is assumed to be only *locally trivial*



Fields on Geometric domains

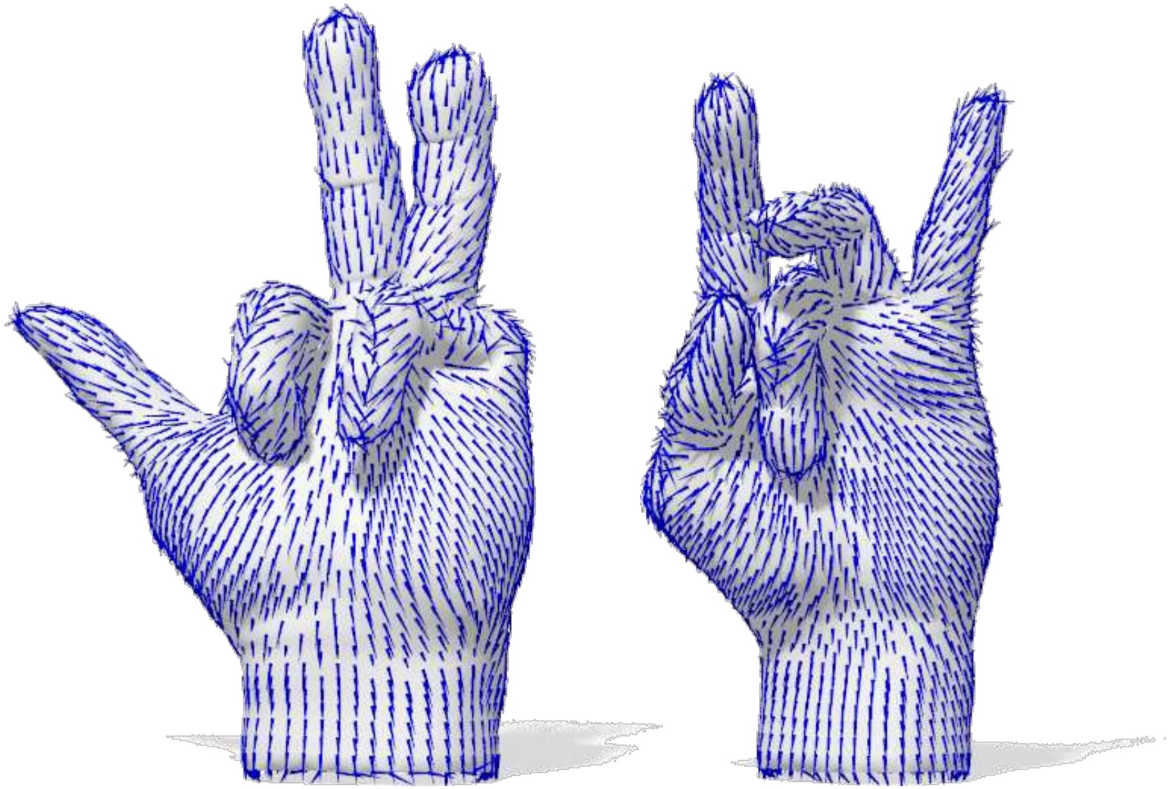
- **Vector (fibre) bundle** E is a family of vector spaces $\mathcal{C}$ locally "attached to" a domain Ω

- **Vector field (section of the bundle)** $x: \Omega \to E$ "continuously attaching to every point u a vector from $\mathcal{C}_u$ in a manner compatible with the bundle structure"



- Given an inner product $\langle \cdot,\cdot \rangle_u$ on $\mathcal{C}_u$ (**Riemannian metric** in differential geometry) we can define an inner product between vector fields as

$$\langle x, y \rangle = \int_{\Omega} \langle x(u), y(u) \rangle_u \, \mathrm{d}\mu(u)$$



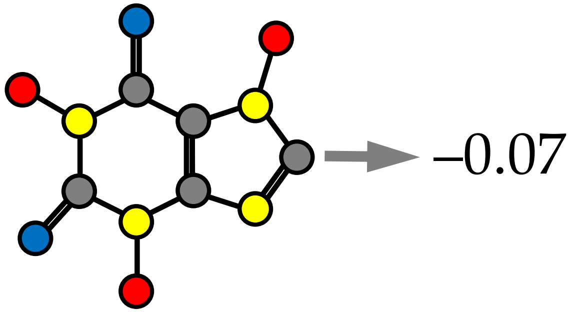
Tangent vector fields on a manifold

Domain as a Signal

- In some cases, there is no given signal defined on the domain Ω
- The *structure* of the domain can be considered as a signal, e.g.
  - Adjacency matrix of a graph $G = (V, E)$ is a signal on $V \times V$
  - Metric tensor of a Riemannian manifold $\mathcal{M}$ is a signal on $\mathcal{M}$

Functions on Signals defined on Geometric domains

- **Label function** $f \in \mathcal{F}\big(\mathcal{X}(\Omega, \mathcal{C})\big) = \{f : \mathcal{X}(\Omega, \mathcal{C}) \to \mathcal{Y}\}$

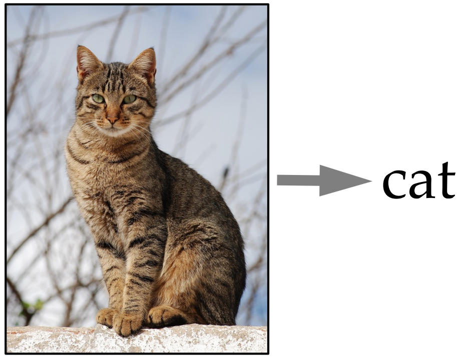


Regression

$\Omega = (V, E)$

$\mathcal{C} = \mathbb{R}^m$

$\mathcal{Y} = \mathbb{R}$

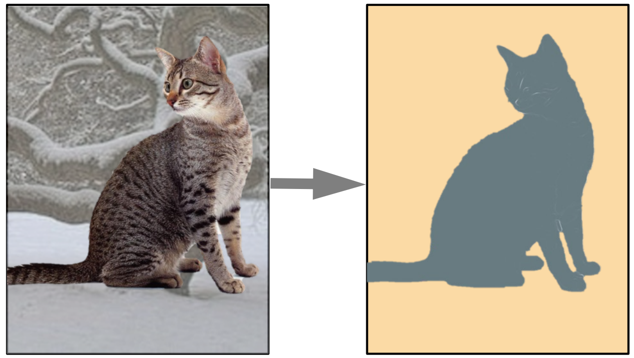


Classification

$\Omega = \mathbb{Z}_n \times \mathbb{Z}_n$

$\mathcal{C} = \mathbb{R}^3$

$\mathcal{Y} = \{1, \dots, K\}$

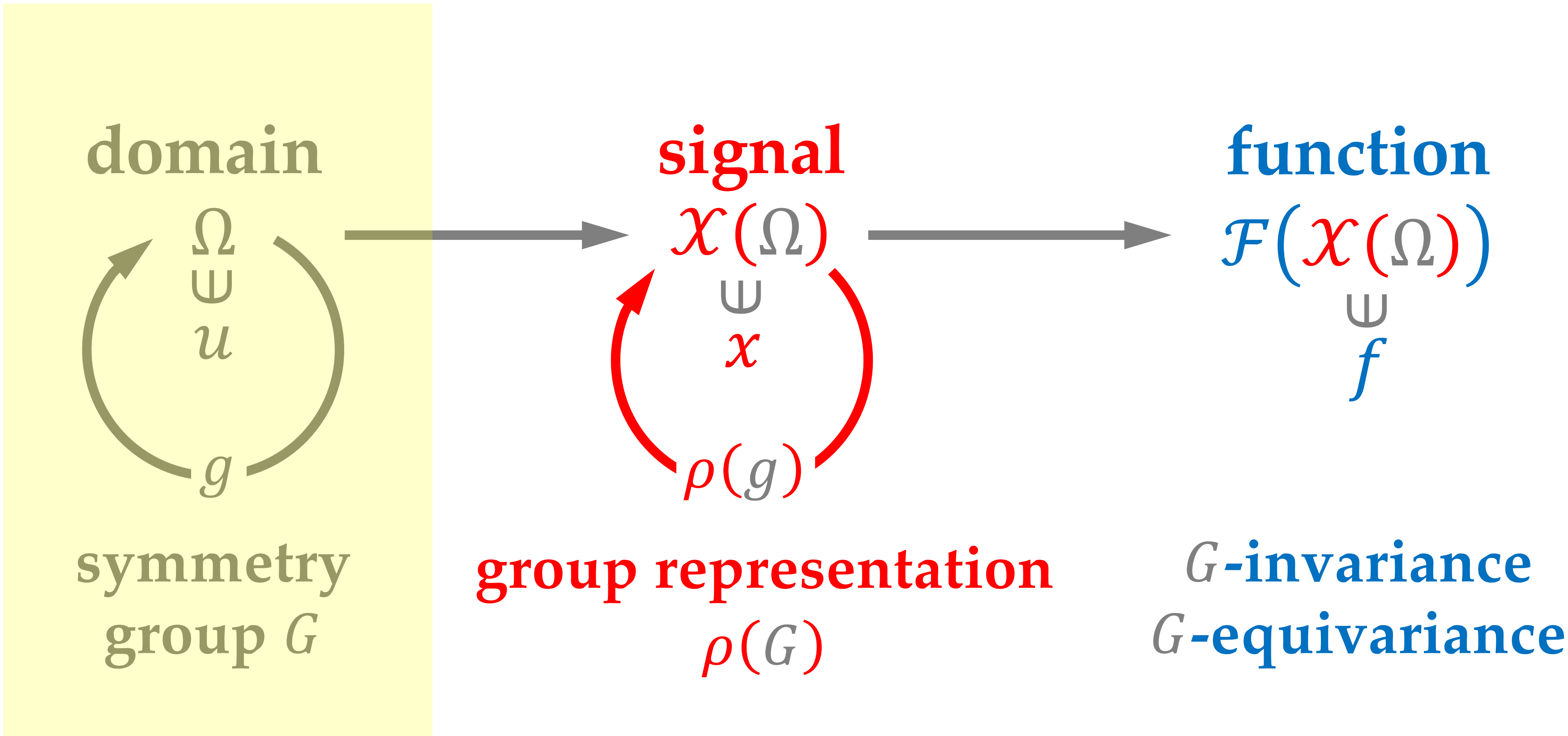


Structured prediction

$\Omega = \mathbb{Z}_n \times \mathbb{Z}_n$

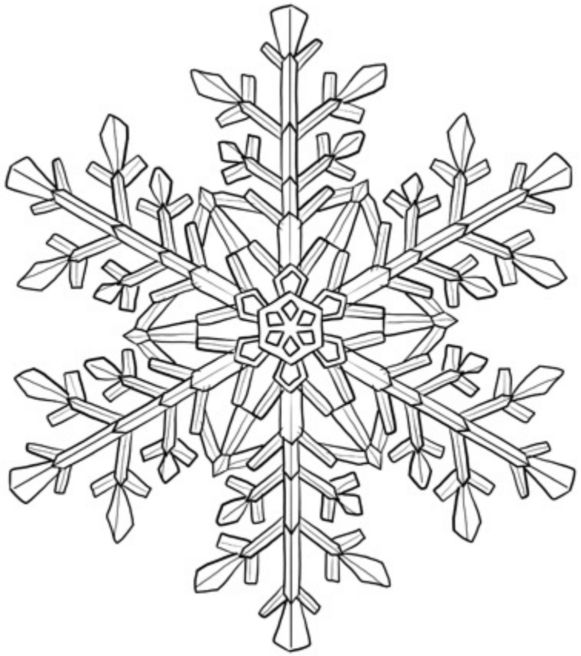
$\mathcal{C} = \mathbb{R}^3$

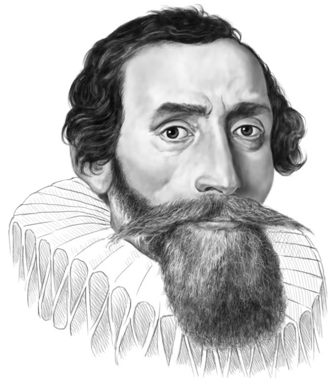
$\mathcal{Y} = \mathcal{X}(\mathbb{Z}_n \times \mathbb{Z}_n, \{0,1\})$


domain
Ω
u
g
symmetry group G
signal
X(Ω)
x
ρ(g)
group representation
ρ(G)
function
F(X(Ω))
f
G-invariance
G-equivariance

SYMMETRY GROUPS

Symmetry

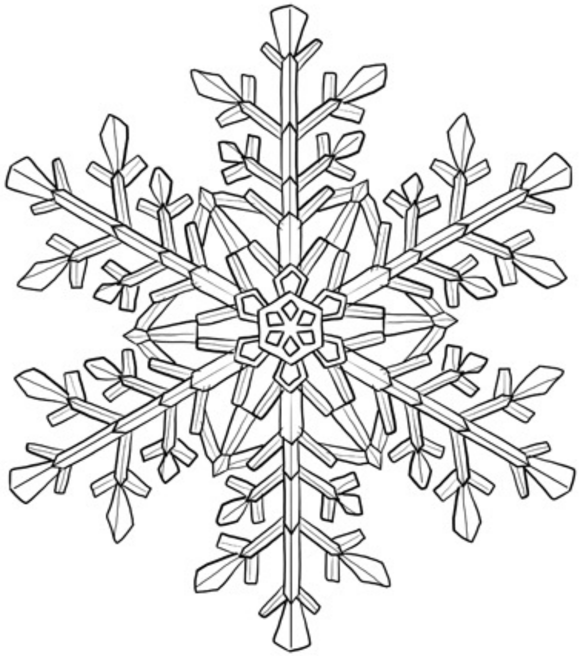


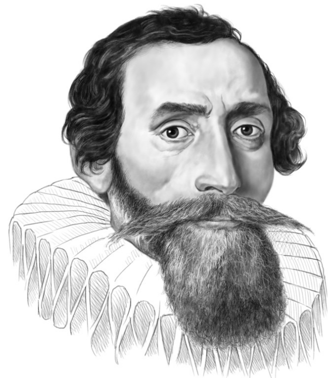


J. Kepler

“a transformation of an object leaving it unchanged”

Symmetry

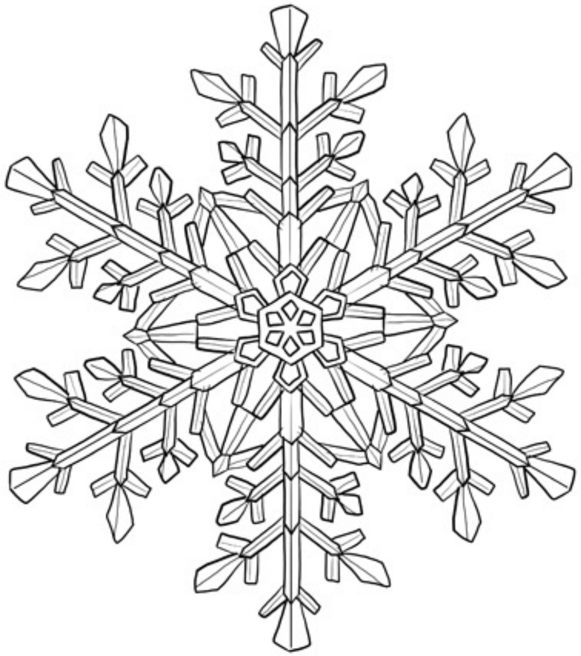


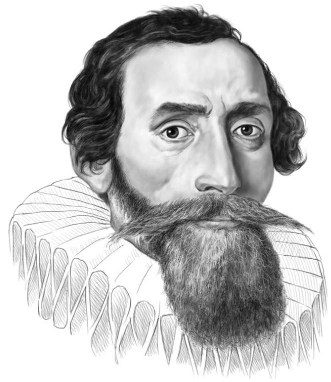


J. Kepler

"element of a symmetry group"

Symmetry

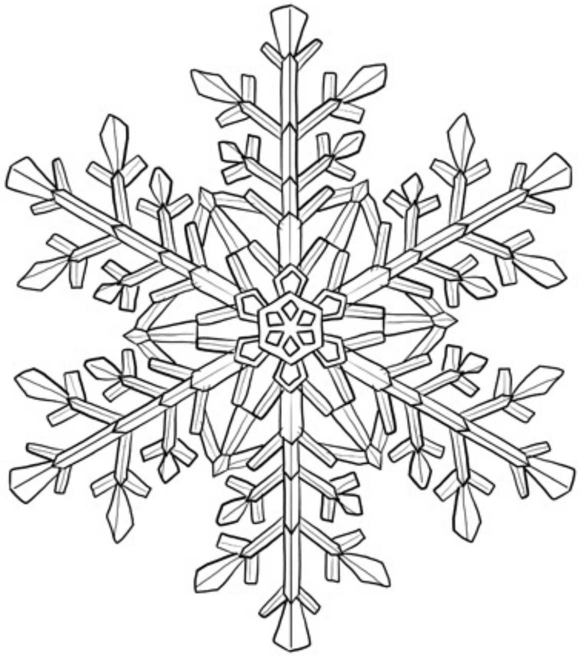


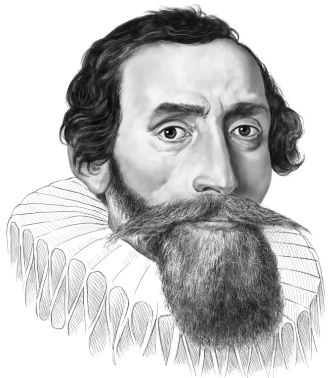


J. Kepler

"invertible structure-preserving map (isomorphism) from the object to itself"

Symmetry

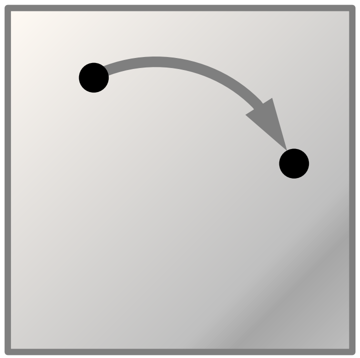




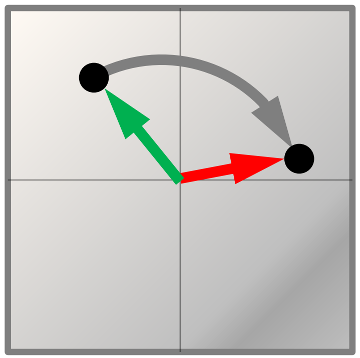
J. Kepler

"automorphism"

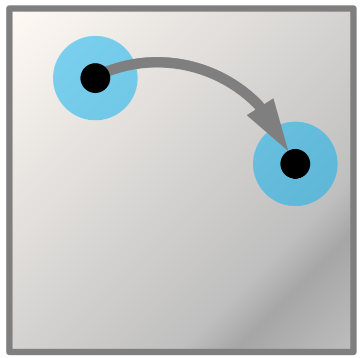
Examples of structure-preserving maps



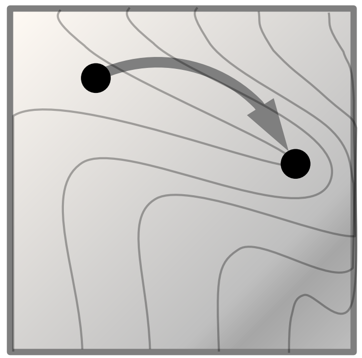
Set
bijections



Vector space
invertible matrices



Manifold
homeomorphisms

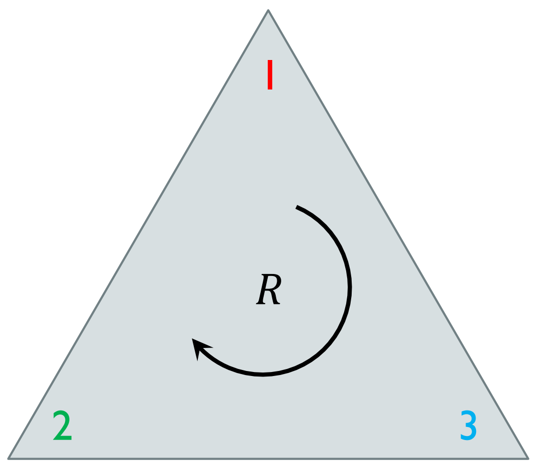


Differential manifold
diffeomorphism

Reminder:
Homeomorphism is a bijective continuous function (bicontinuous). It preserves topological structure.

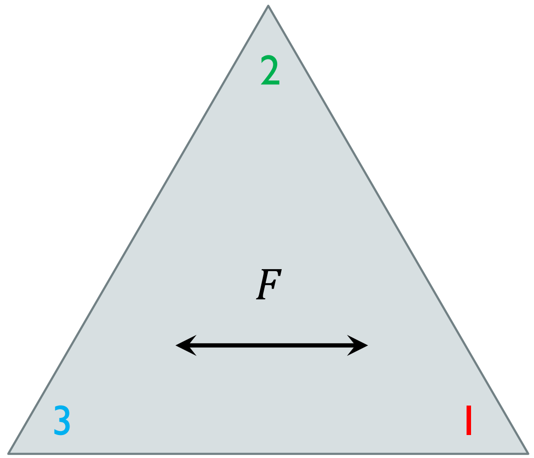
Diffeomorphism is a bijective differentiable function with differentiable inverse. It preserves differential structure on manifolds.

Symmetry of a triangle



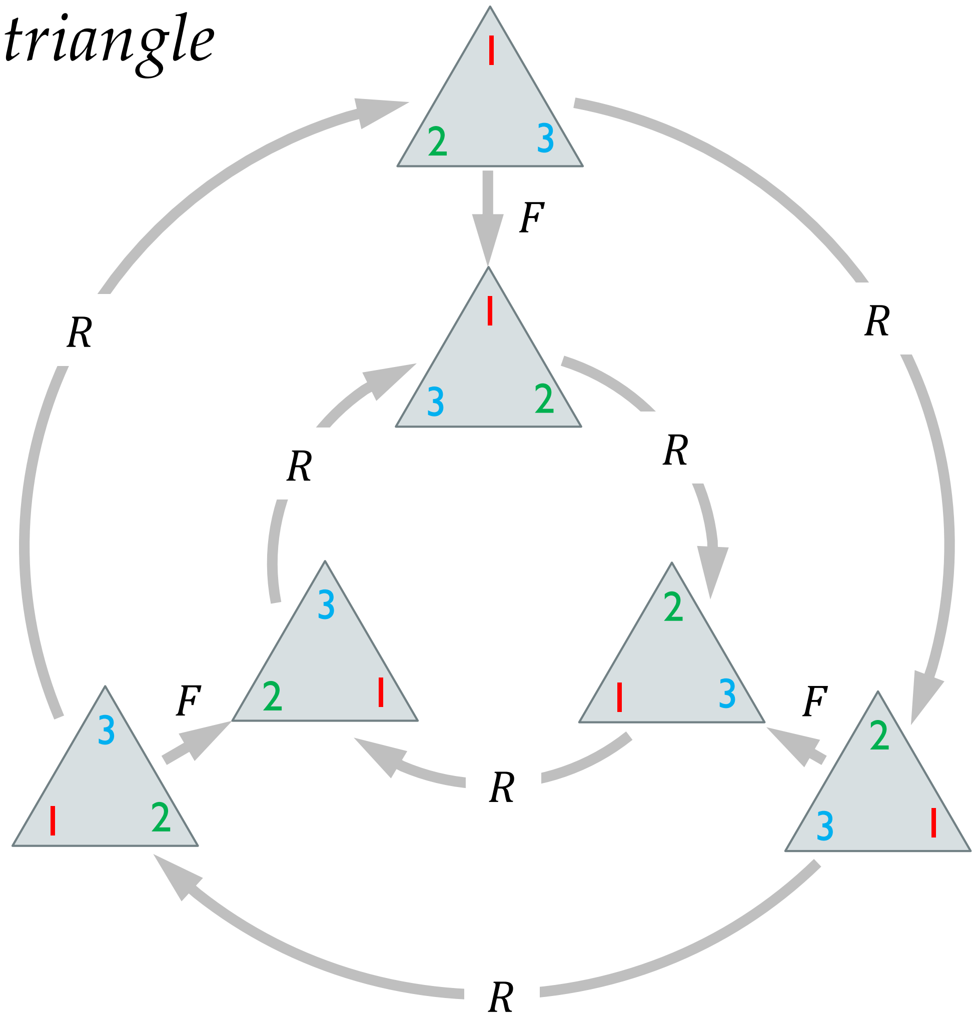
rotation by 120°

Symmetry of a triangle

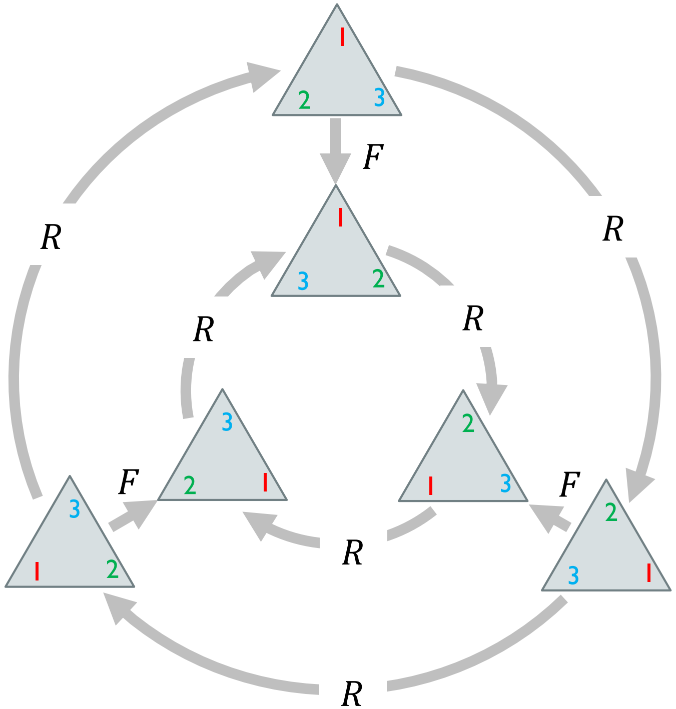


reflection

Symmetry of a triangle



Symmetry of a triangle



Cayley graph

$\circ$	I	R	R^2	F	FR	FR^2
I	I	R	R^2	F	FR	FR^2
R	R	R^2	I	FR^2	F	FR
R^2	R^2	I	R	FR	FR^2	F
F	F	FR	FR^2	I	R	R^2
FR	FR	FR^2	F	R^2	I	R
FR^2	FR^2	F	FR	R	R^2	I

Cayley table

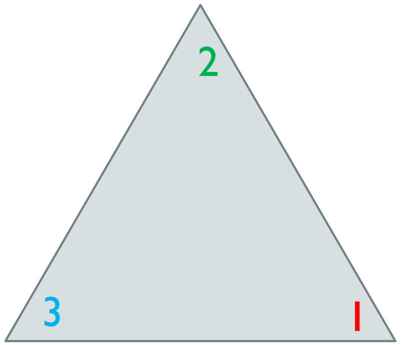
Groups

A **group** $(G,*)$ is a set G together with binary operation $* : G\times G \to G$ (denoted by juxtaposition $g * h = gh$ for brevity) satisfying the following axioms:

- *Associativity:* $(gh)k = g(hk)$ for all $g, h, k \in G$
- *Identity:* $\exists!\, e \in G$ satisfying $eg = ge = g$ for all $g \in G$
- *Inverse:* $\exists!\, g^{-1} \in G$ for each $g \in G$ satisfying $g^{-1}g = gg^{-1} = e$

- *Closure* $(gh \in G)$ follows from definition
- Not necessarily commutative (i.e., $gh \neq hg$). Commutative groups are called *Abelian*
- Groups can be finite, infinite, discrete, or continuous.
- *Lie groups* such as 3D rotations are smooth manifolds (we can do calculus on them)

Equivalent groups



$\{1, 2, 3\}$ $\{3, 1, 2\}$ $\{2, 3, 1\}$

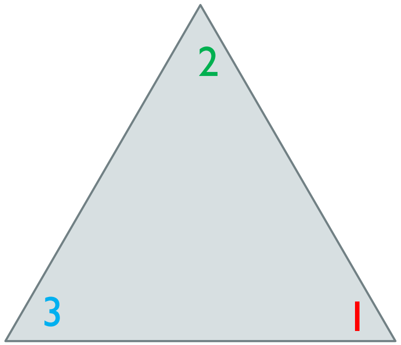
$\{3, 2, 1\}$ $\{2, 1, 3\}$ $\{1, 3, 2\}$

dihedral group $\cong$ **permutation group**

D_3 S_3

The group abstracts out the objects themselves and captures only how they compose

Equivalent groups



{1, 2, 3} {3, 1, 2} {2, 3, 1}

{3, 2, 1} {2, 1, 3} {1, 3, 2}

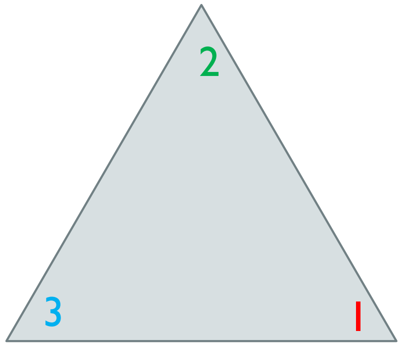
dihedral group $\cong$ **permutation group**

D_3 S_3

Two groups $(G,*)$ and $(H,\circ)$ are **isomorphic** (denoted by $(G,*) \cong (H,\circ)$) if there exists a bijection $\varphi: G \to H$ (called **group isomorphism**) satisfying for all $g, h \in G$

$$\varphi\,(g * h) = \varphi(g) \circ \varphi(h)$$

Equivalent groups



dihedral group

D_3

$\cong$

{1, 2, 3} {3, 1, 2} {2, 3, 1}

{3, 2, 1} {2, 1, 3} {1, 3, 2}

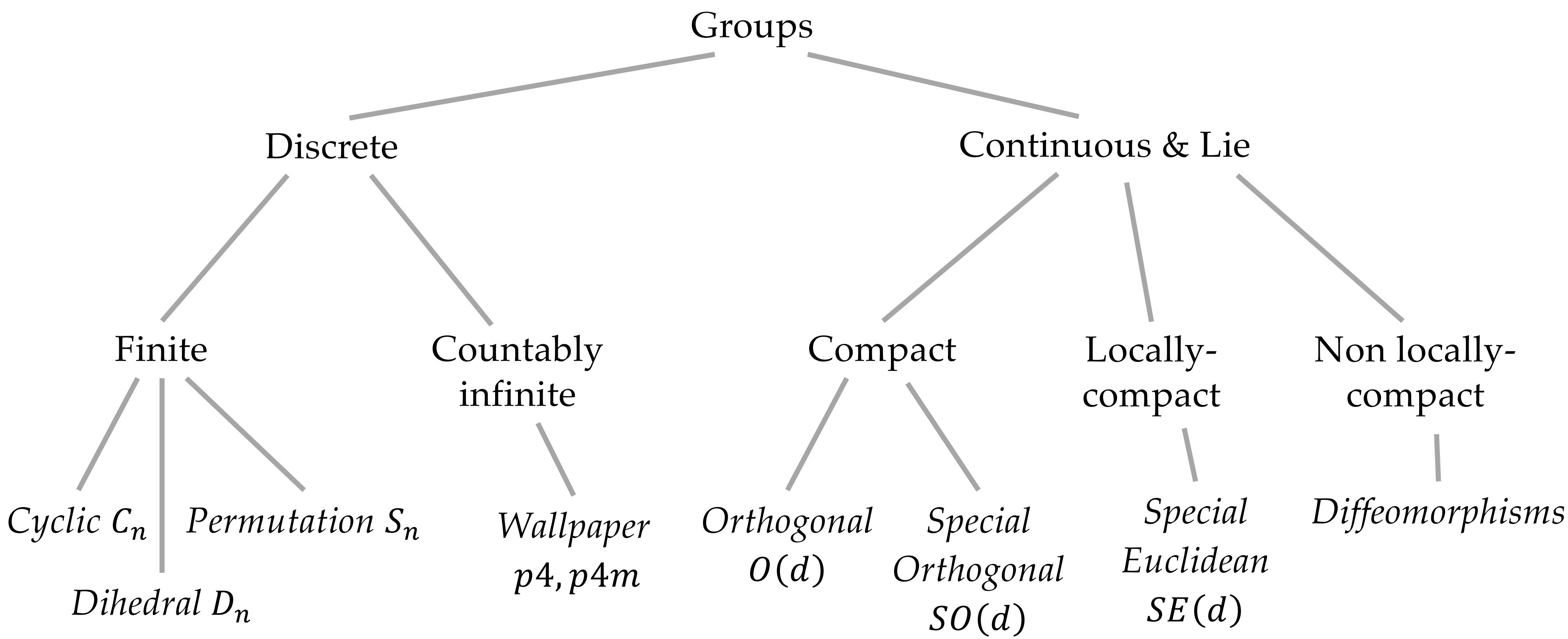
permutation group

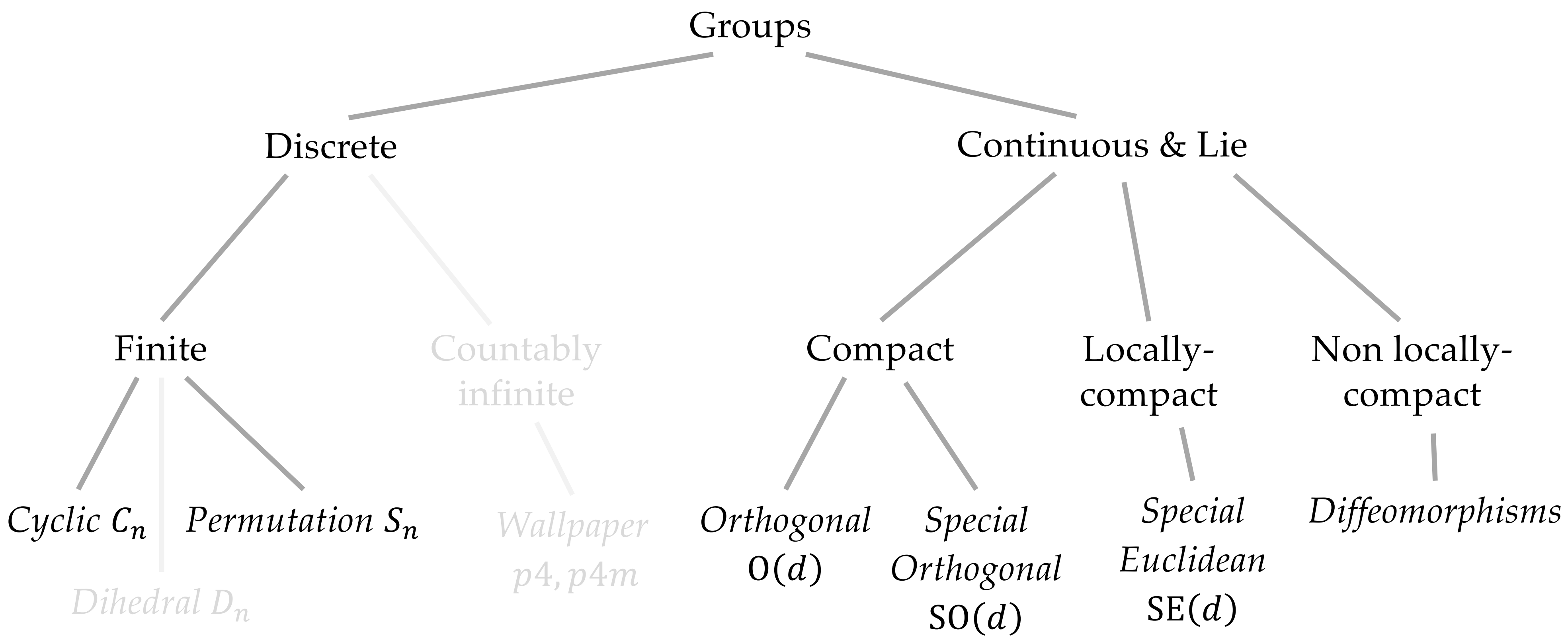
S_3

Group homomorphism (don't confuse with *homeomorphism*, which is a map between topological spaces) is a map $\varphi: (G,*) \rightarrow (H,\circ)$ satisfying $\varphi\ (g * h) = \varphi(g) \circ \varphi(h)$. It preserves group operations but not necessarily group structure.

Group isomorphism is a bijective group homomorphism. It preserves group structure.

Exercise: prove that group homomorphism maps the identity of G to the identity of H


Groups
Discrete
Continuous & Lie
Finite
Countably infinite
Compact
Locally-compact
Non locally-compact
Cyclic C_n
Dihedral D_n
Permutation S_n
Wallpaper $p4, p4m$
Orthogonal $O(d)$
Special Orthogonal $SO(d)$
Special Euclidean $SE(d)$
Diffeomorphisms

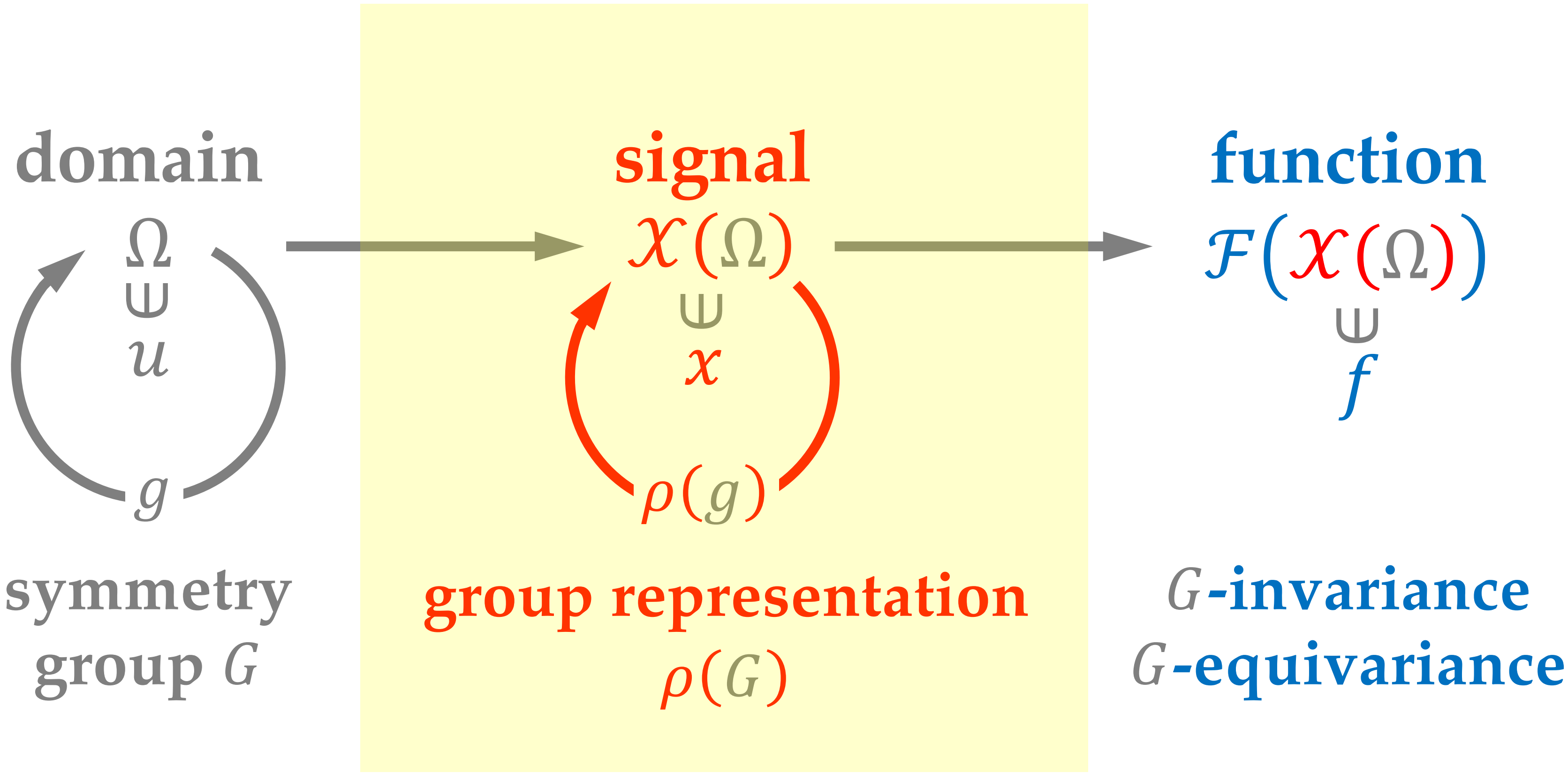


Groups
Discrete
Continuous & Lie
Finite
Countably infinite
Compact
Locally-compact
Non locally-compact
Cyclic C_n
Permutation S_n
Dihedral D_n
Wallpaper $p4, p4m$
Orthogonal $\mathrm{O}(d)$
Special Orthogonal $\mathrm{SO}(d)$
Special Euclidean $\mathrm{SE}(d)$
Diffeomorphisms

Examples of Important groups

- **Permutation** (**symmetric**) group S_n: reorder a set of n elements
- **Cyclic** group C_n: shift the order of n elements by one position modulo n
- **Groups of matrices** of size $d \times d$ with matrix multiplication operation
  - **General linear group** $\mathrm{GL}(d)$: invertible matrices
  - **Special linear group** $\mathrm{SL}(d)$: volume- and orientation-preserving matrices ($\det = 1$)
  - **Orthogonal group** $\mathrm{O}(d)$: angle-preserving (orthogonal) matrices
  - **Special orthogonal group** $\mathrm{SO}(d)$: volume-, orientation- and angle-preserving matrices

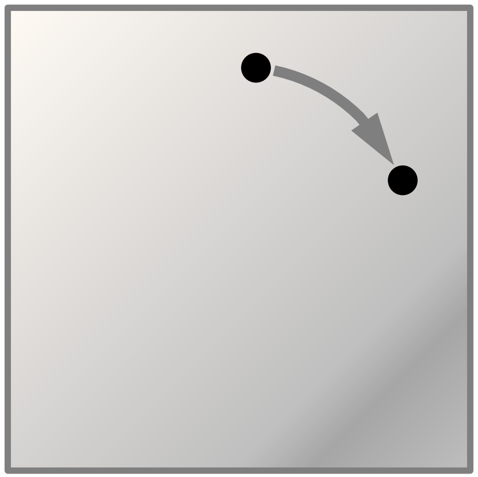
Exercise: show the above groups indeed satisfy the group axioms



domain
Ω
⋓
u
g
symmetry
group G
signal
X(Ω)
⋓
x
ρ(g)
group representation
ρ(G)
function
F(X(Ω))
⋓
f
G-invariance
G-equivariance

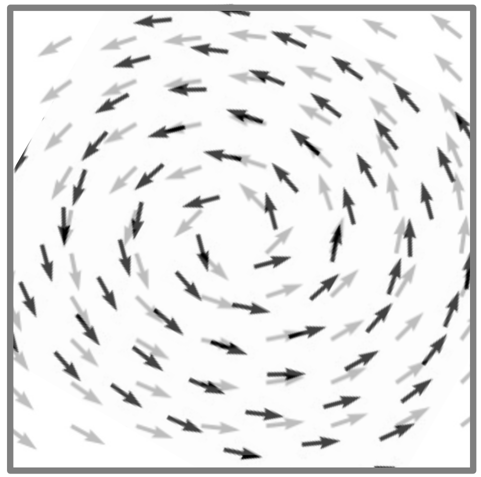
GROUP ACTIONS & REPRESENTATIONS

Group actions on objects



Point in a plane

Image (function)



Vector field

The type of an object can be defined by the way it transforms by a group

Group action

Let G be a group and X a set. A **(left) group action** of G on X (often denoted $gx = \alpha(g, x)$) is a mapping of the form $\alpha : G \times X \to X$ satisfying

- *Identity:* $\alpha(e, x) = x$ for all $x \in X$
- *Composition:* $\alpha(gh, x) = \alpha\big(g, \alpha(h, x)\big)$ for all $g, h \in G$ and $x \in X$

Group representation

A **representation** of G on X is a mapping of the form $\rho: G \to \{f: X \to X\}$ that assigns to each $g \in G$ a map $\rho(g): X \to X$ satisfying

- *Identity:* $\rho(e) = \mathrm{id}$
- *Composition:* $\rho(gh) = \rho(g) \circ \rho(h)$ for all $g, h \in G$

- Given a group action α, a representation can be defined as $\rho(g)x = \alpha(g, x)$
- Preserves *positive relations* (e.g., $g^{-1}g = gg^{-1} = e$) that hold in the group G
- *Negative relations* (of the form $gh \neq k$) may not be preserved
- *Trivial representation* $\rho \equiv \mathrm{id}$
- *Faithful representation* is *injective* ($g \neq h \Rightarrow \rho(g) \neq \rho(h)$) and preserves negative relations
- Additional structure of ρ (e.g. smoothness if G is a Lie group)

Linear Group representation

A **linear representation** of G on a vector space X is group homomorphism $\rho: G \to \mathrm{GL}(X)$ that assigns to each $g \in G$ an **invertible linear** map $\rho(g): X \to X$ satisfying

$$\rho(gh) = \rho(g)\rho(h) \text{ for all } g, h \in G$$

- $\dim(X)$ is called the dimension of the representation
- In finite-dimensional cases, ρ can be represented by *matrices*
- This turns group theory into linear algebra
- Efficient implementation on standard hardware

Linear Group representation

A d–**dimensional (linear) representation** of G is a map $\rho: G \rightarrow \mathbb{R}^{d\times d}$ assigning to each $g \in G$ an **invertible matrix** $\rho(g) \in \mathbb{R}^{n\times n}$ satisfying $\rho(gh) = \rho(g)\rho(h)$ for all $g, h \in G$.

Exercise I: show that $\rho(e) = \mathbf{I}$.

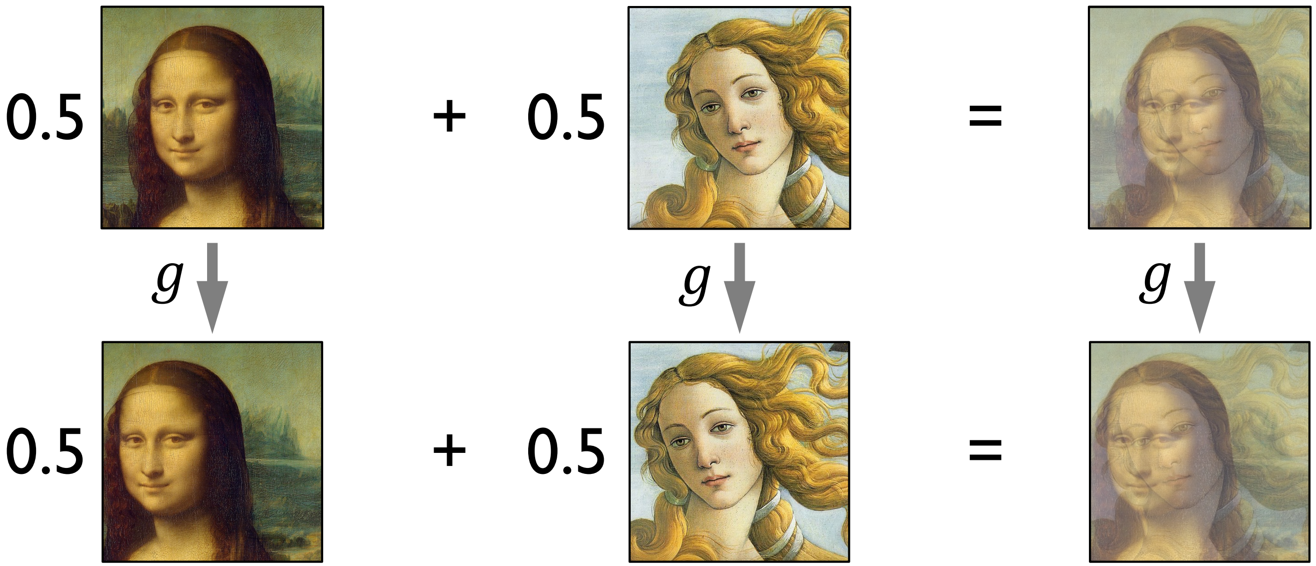
Note: such a representation is not unique! Given an invertible matrix $\mathbf{A}$ ("change of basis"), we can define a new representation $\bar{\rho}(g) = \mathbf{A}\rho(g)\mathbf{A}^{-1}$.

Exercise II: verify that $\bar{\rho}$ is indeed a representation

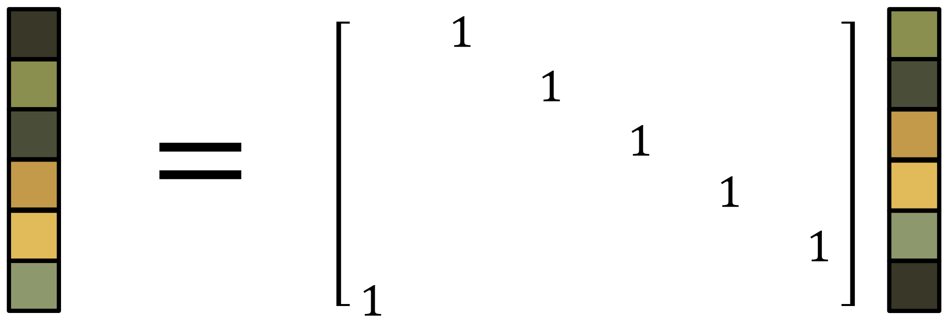
Group actions on Signals defined on geometric Domains

Given a group G acting on a **domain** Ω, we automatically obtain an action of G on the space of signals $\mathcal{X}(\Omega)$ through the **regular representation** $(\rho(g)x)(u) = x(g^{-1}u)$

Exercise: prove that this representation is linear

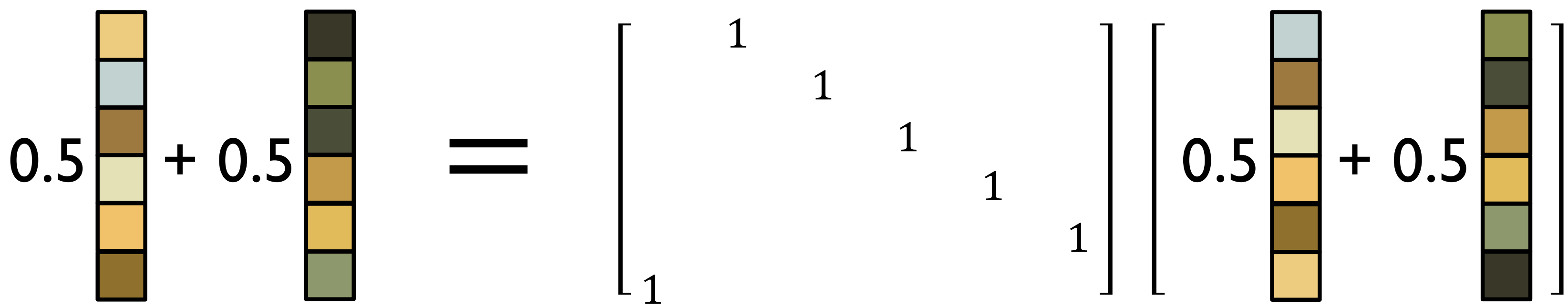


Intuition

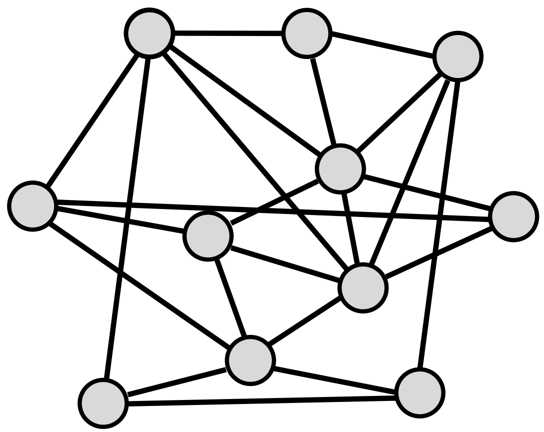


Note: a 2D shift can be represented as tensor (Kronecker) product $(\mathbf{S} \otimes \mathbf{S})\mathrm{vec}(\mathbf{X}) = \mathrm{vec}\left(\mathbf{S}\mathbf{X}\mathbf{S}^{\mathrm{T}}\right)$

Intuition

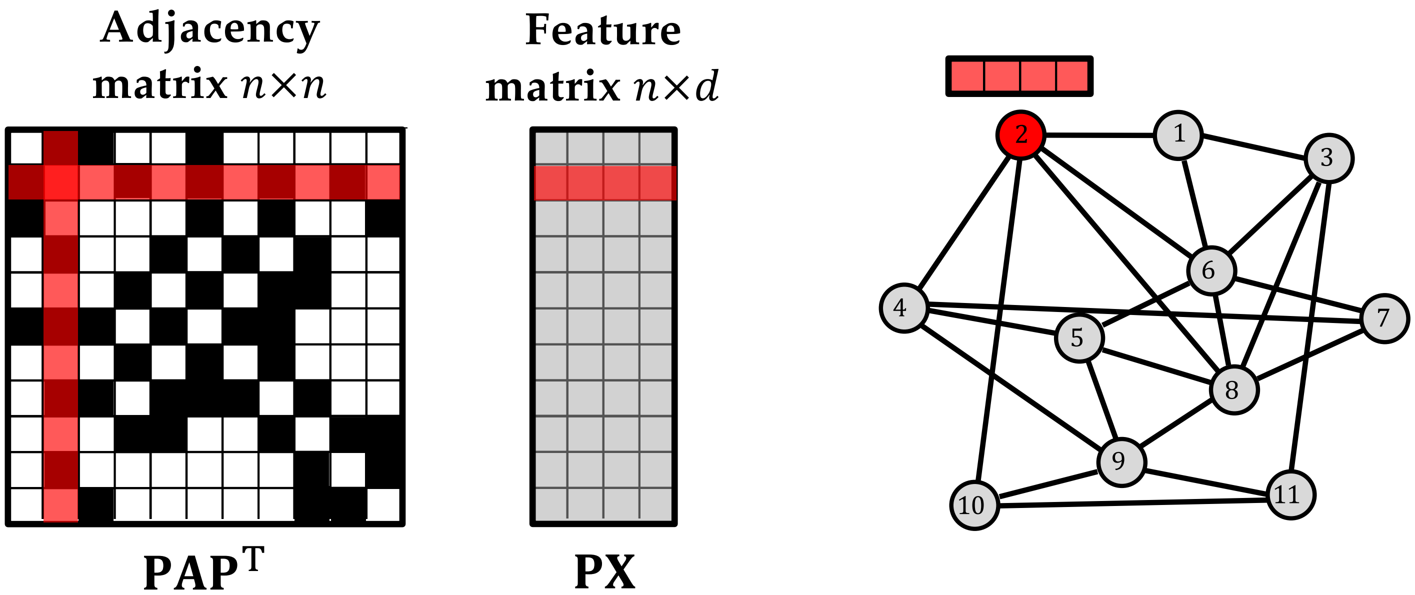


Example: Symmetries of Graphs



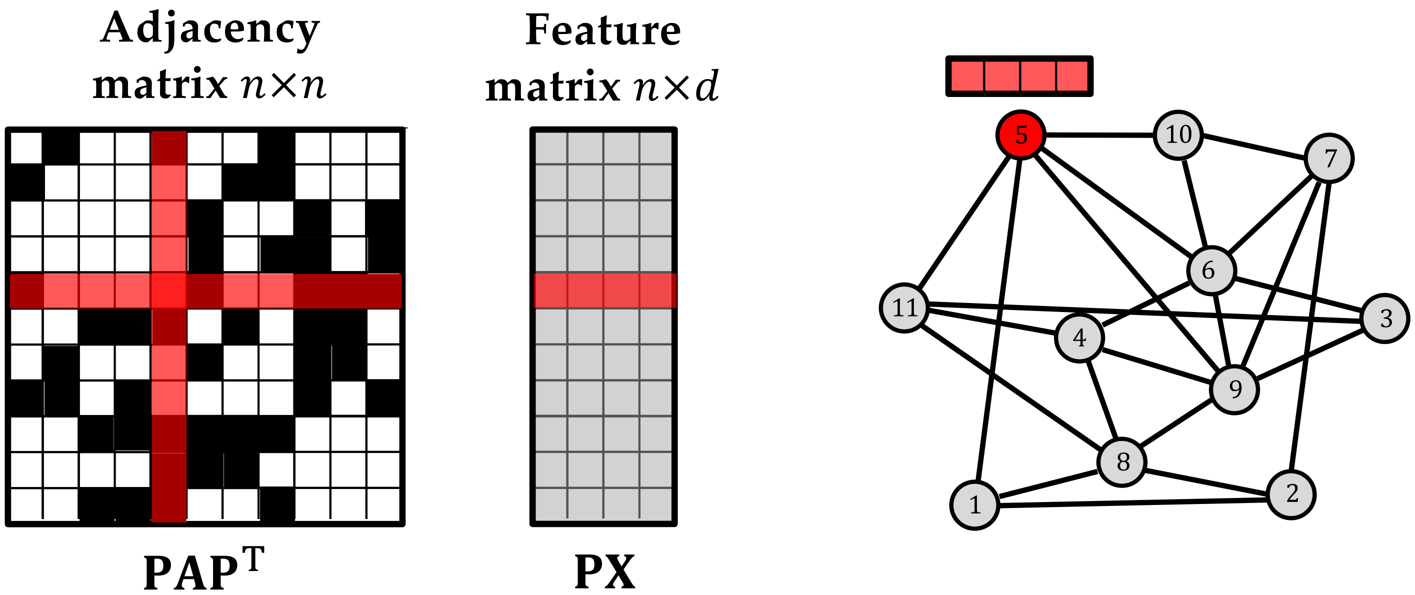
- A graph is an abstract object

Example: Symmetries of Graphs



- A graph is an abstract object
- Its *description* (adjacency / feature matrix) has “extrinsic” properties (order of nodes)

Example: Symmetries of Graphs



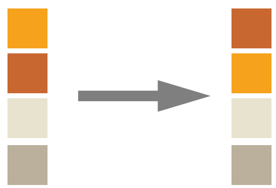
- A graph is an abstract object
- Its *description* (adjacency / feature matrix) has "extrinsic" properties (order of nodes)

Different representations of the permutation group on Graphs

- **Domain**: set of n graph vertices $\Omega = \{1, \ldots, n\}$
- **Group:** permutations $G = S_n$

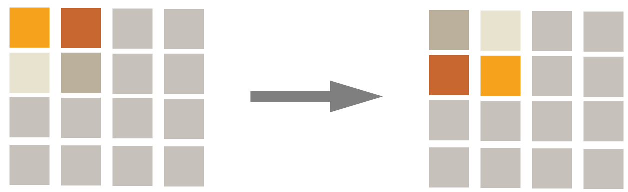
Scalar

$\rho_0(g)x = 1 \cdot x$



Vector

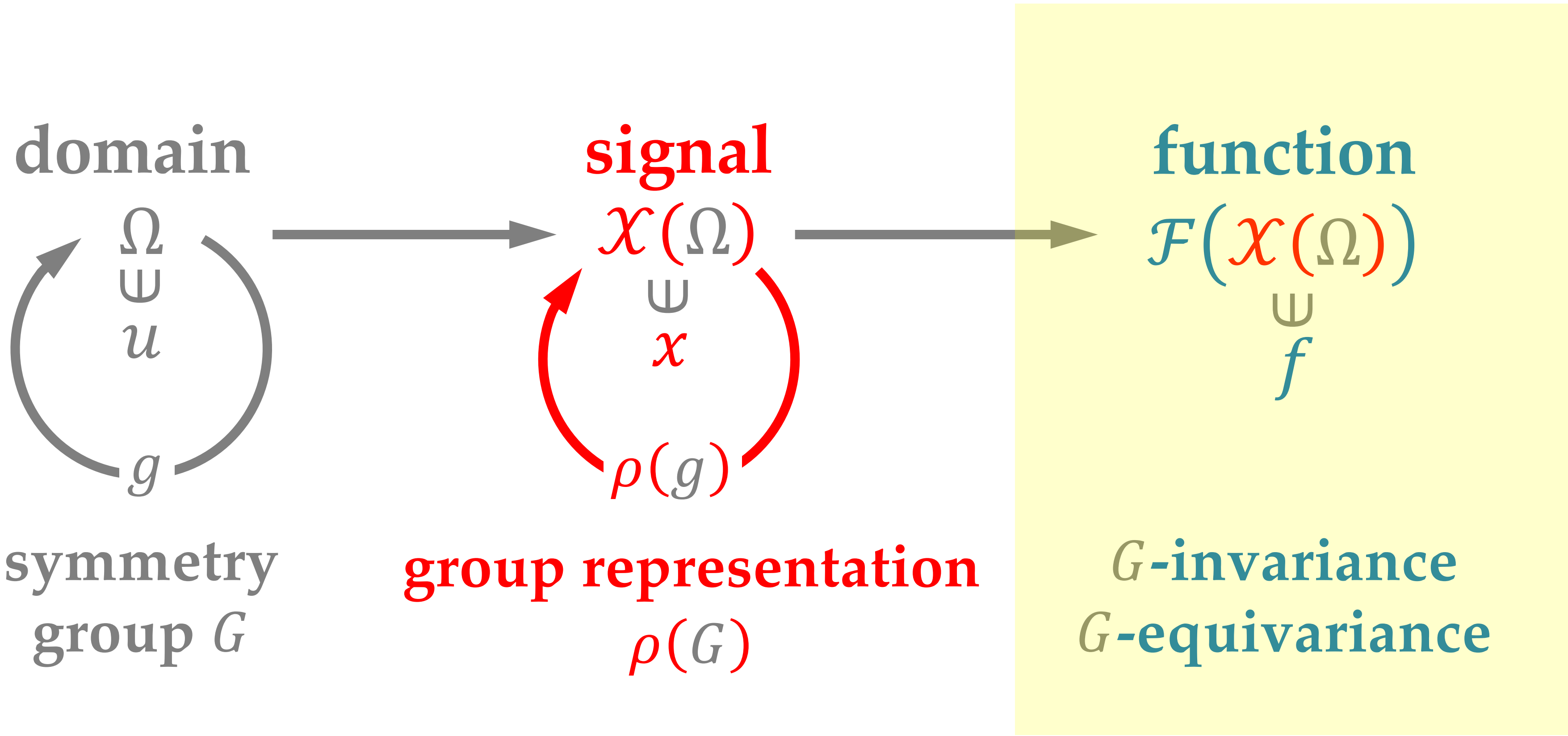
$\rho_1(g)\mathbf{v} = \mathbf{P}\mathbf{v}$



Matrix

$\rho_2(g)\mathbf{M} = \mathbf{P}\mathbf{M}\mathbf{P}^{\mathrm{T}}$

Exercise: verify that each of these are valid group representations



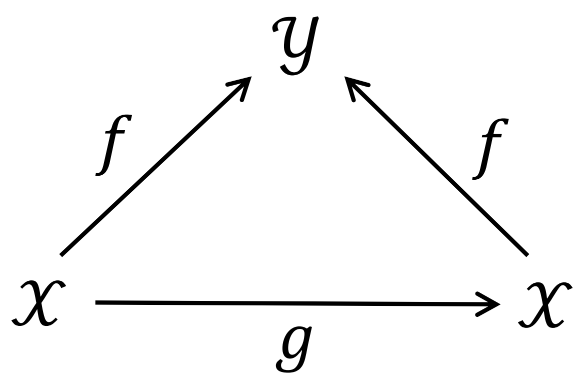
domain
Ω
$\ni$
u
g
symmetry
group G
signal
$\mathcal{X}(\Omega)$
$\ni$
x
$\rho(g)$
group representation
$\rho(G)$
function
$\mathcal{F}(\mathcal{X}(\Omega))$
$\ni$
f
G-invariance
G-equivariance

SYMMETRY IN LEARNING

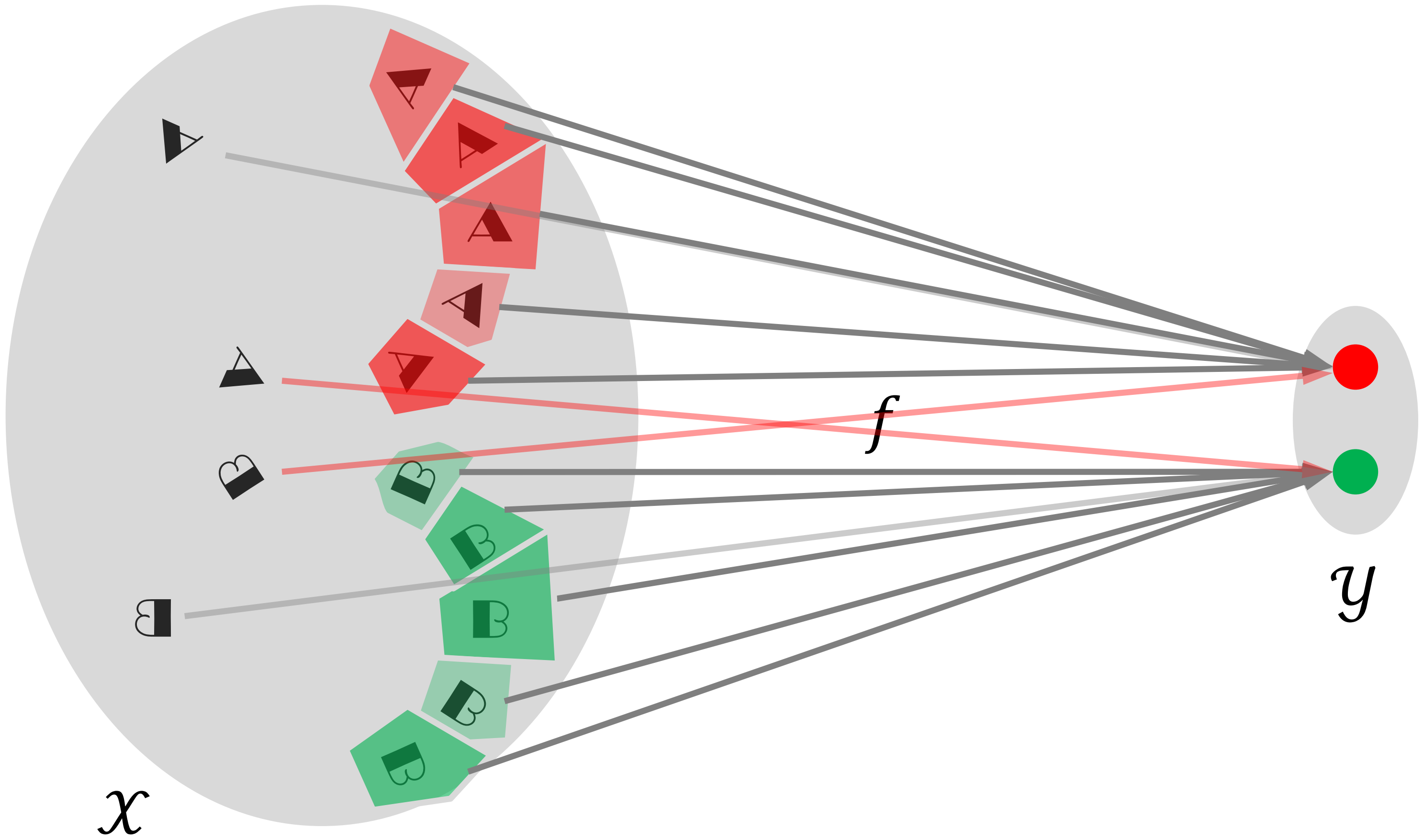
Symmetries of the Label Function

- **Label function** $f: \mathcal{X} \to \mathcal{Y}$ e.g., classification ($\mathcal{Y} = \{1, \ldots, K\}$)
- **Symmetry of a label function** is an invertible label-preserving map $g: \mathcal{X} \to \mathcal{X}$, i.e.

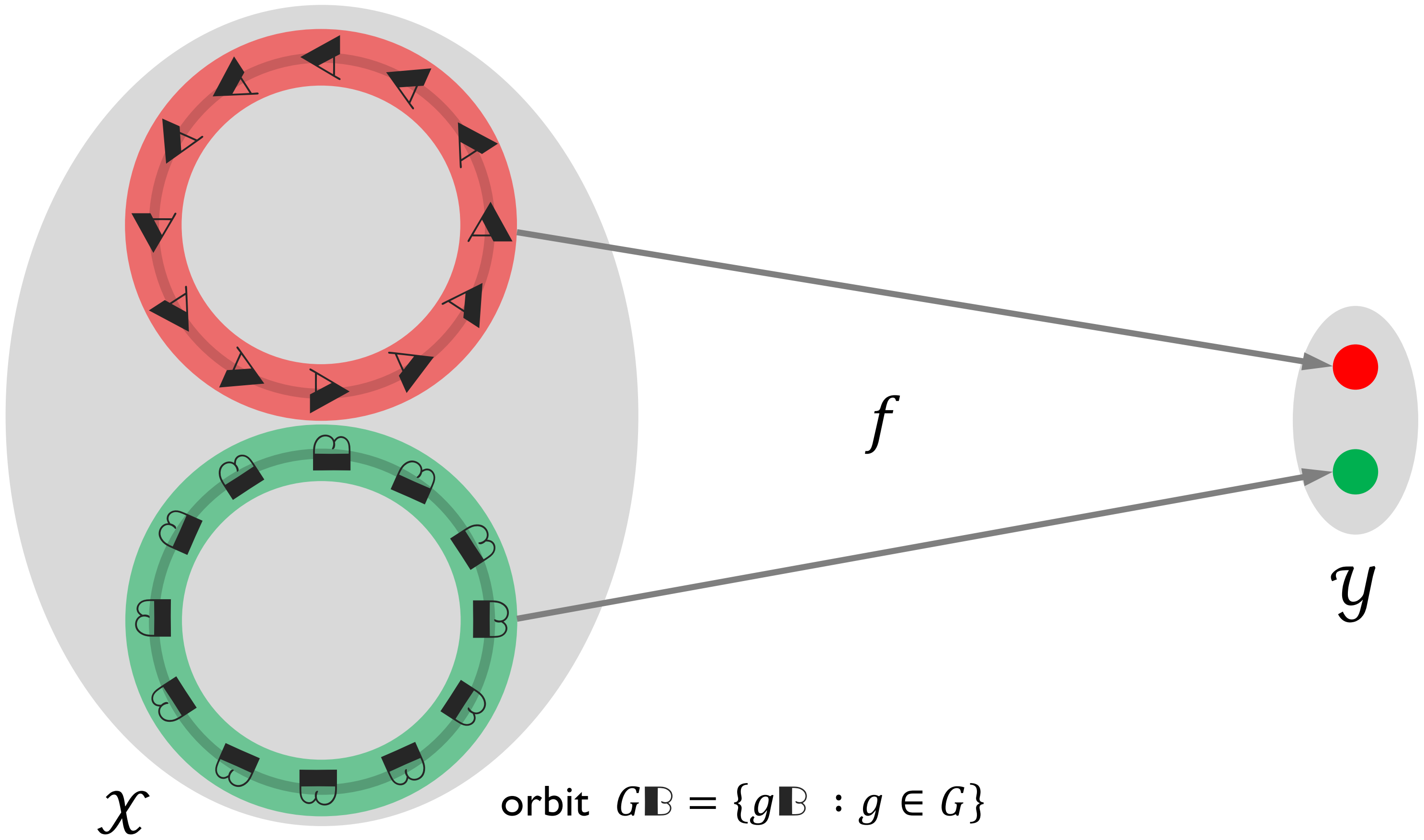
$$(f \circ g)(x) = f(x) \text{ for all } x \in \mathcal{X}$$



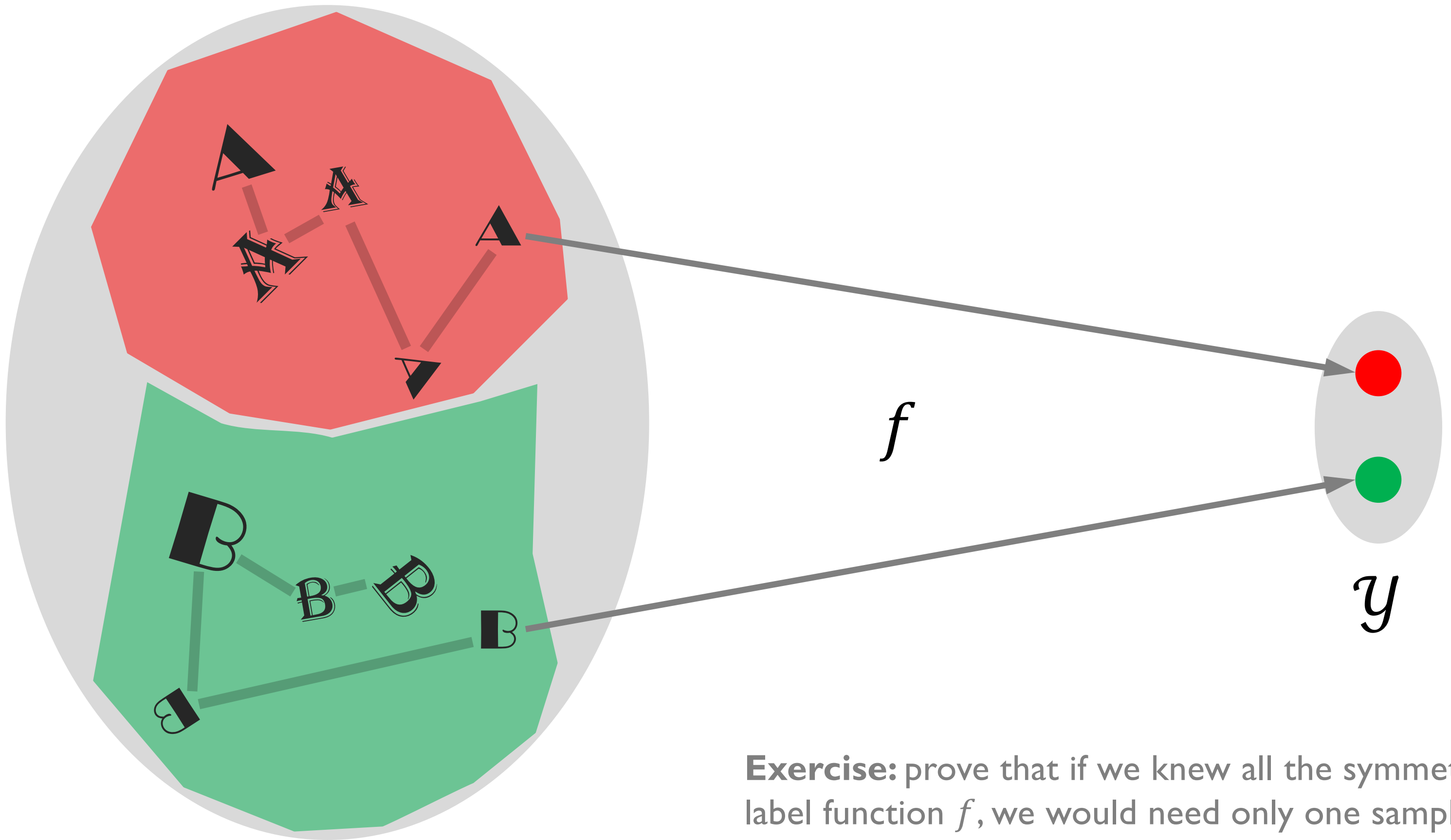
Symmetries of the Label Function



Symmetries of the Label Function



Symmetries of the Label Function

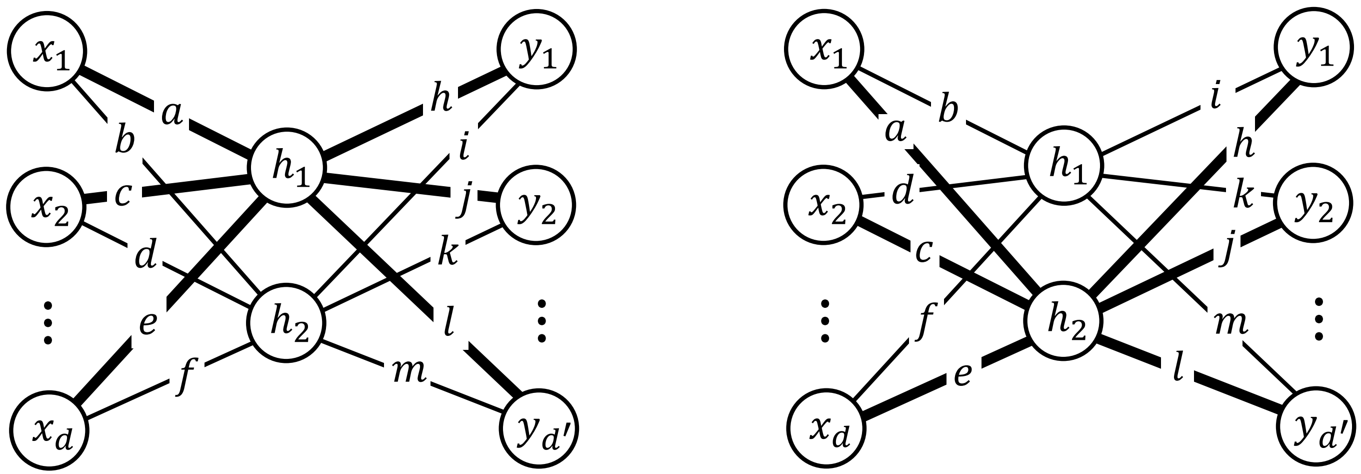


Exercise: prove that if we knew all the symmetries of the label function f, we would need only one sample per class

Symmetries of the Weights

- Let $f_\theta: \mathcal{X} \times \Theta \to \mathcal{Y}$ be a parametric model (neural network)
- A transformation $h: \Theta \to \Theta$ is a **symmetry of the weights** if, for all $x \in \mathcal{X}$ and $\theta \in \Theta$

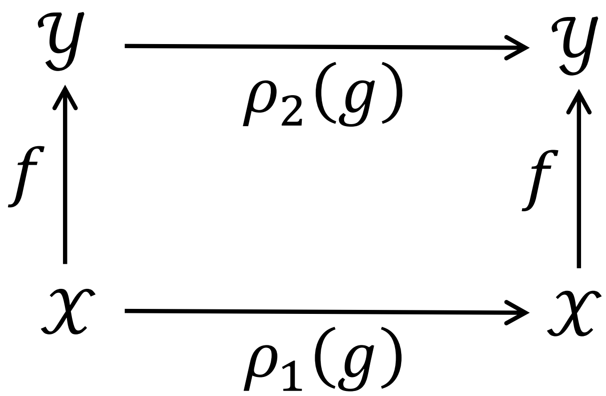
$$f_{h\theta}(x) = f_\theta(x)$$



Equivariance

Let $f: \mathcal{X} \rightarrow \mathcal{Y}$ and G a group acting on $\mathcal{X}$ and $\mathcal{Y}$ through representation ρ_1 and ρ_2, respectively. Then, f is G-**equivariant** if for all $g \in G$

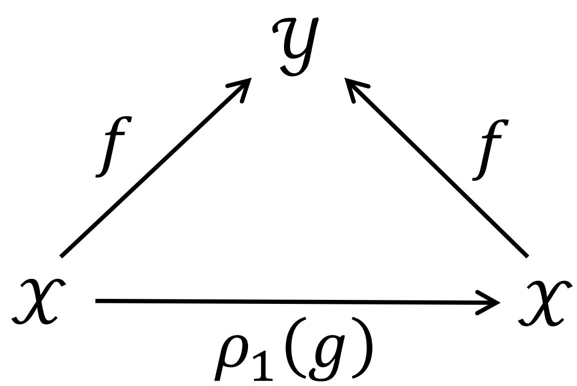
$$f(\rho_1(g)x) = \rho_2(g)f(x)$$



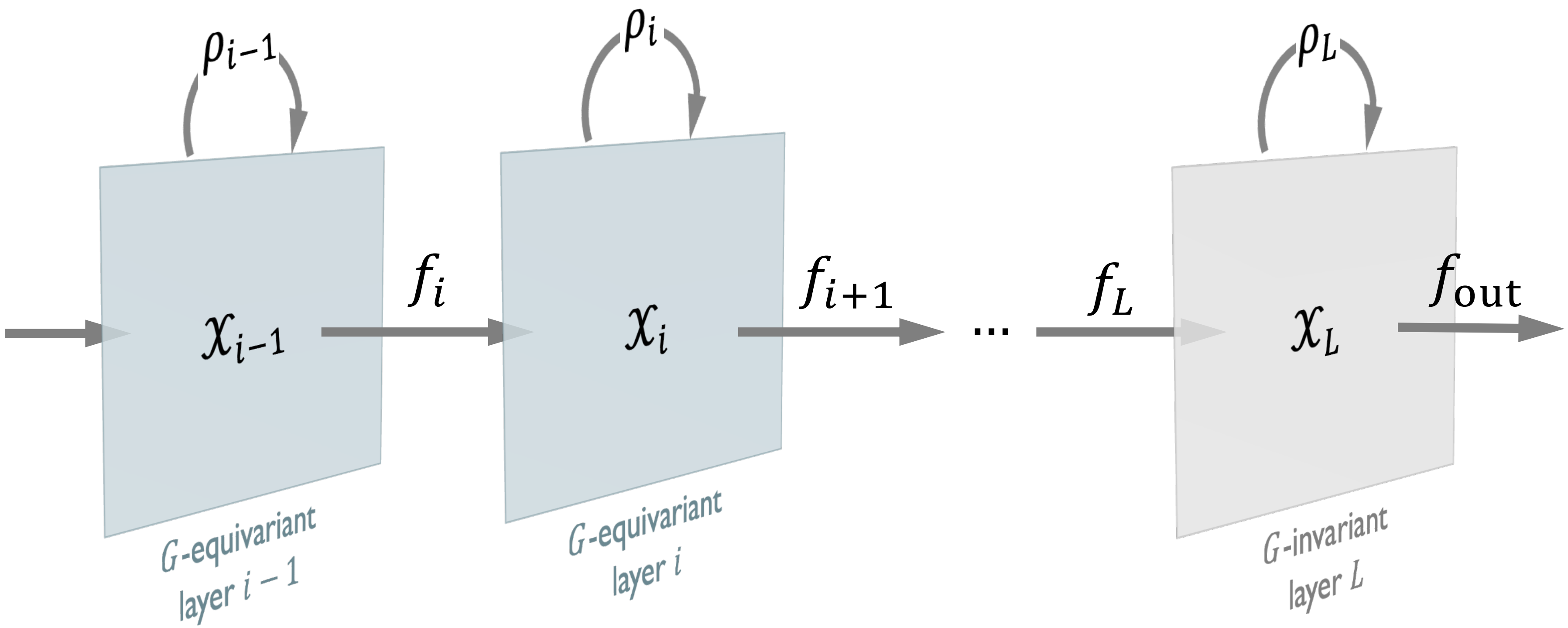
Invariance: special case of Equivariance with trivial representation

Let $f: \mathcal{X} \rightarrow \mathcal{Y}$ and G a group acting on $\mathcal{X}$ and $\mathcal{Y}$ through representation ρ_1 and $\rho_2 = \text{id}$, respectively. Then, f is G-**invariant** if for all $g \in G$

$$f(\rho_1(g)x) = f(x)$$

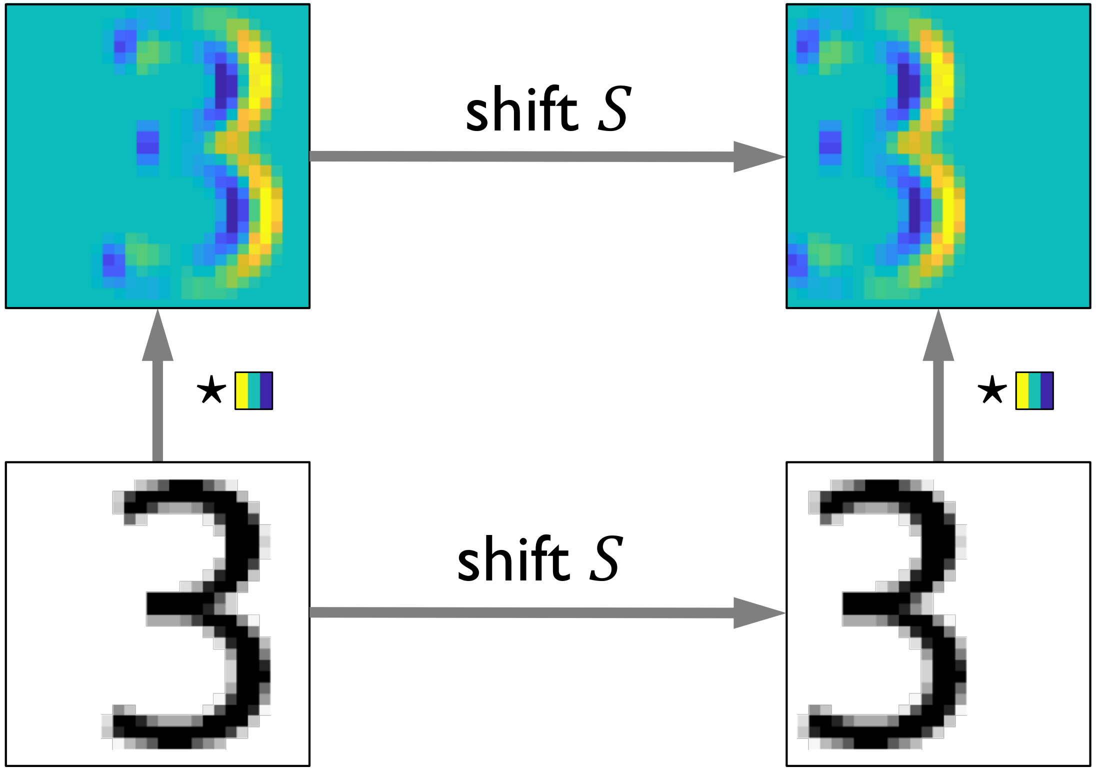


Geometric Deep Learning Blueprint (so far)

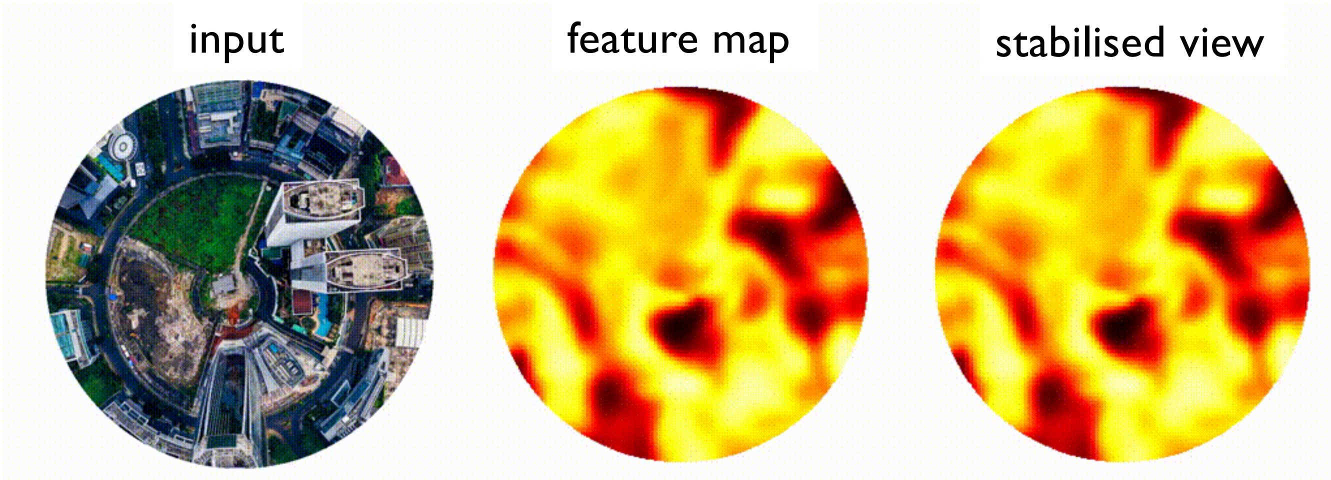


$f_i \circ \rho_{i-1}(g) = \rho_i(g) \circ f_i$ $\qquad$ $f_L \circ \rho_L(g) = f_L$

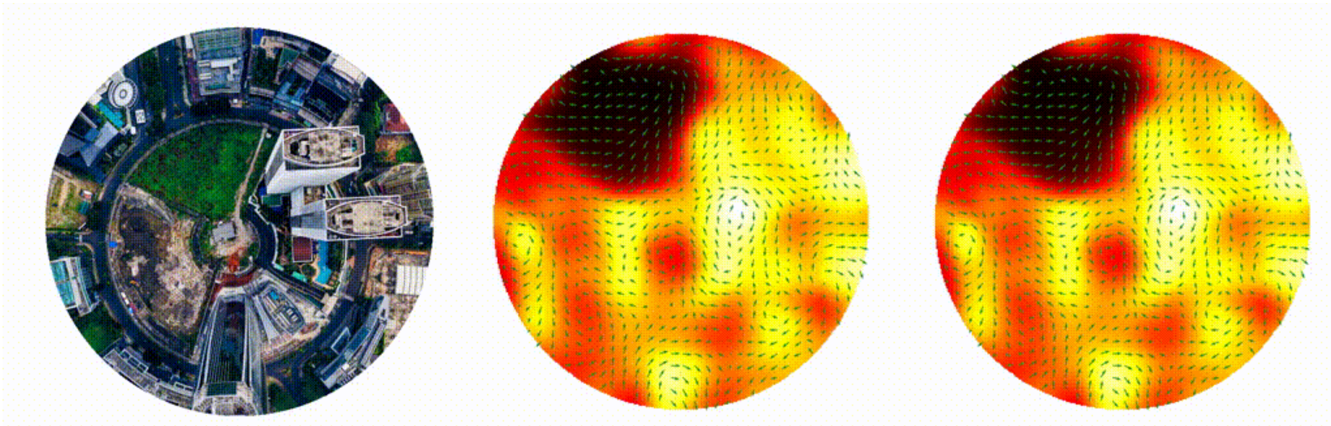
Example: Convolutional Neural Networks



CNN



Rotation-equivariant CNN



Weiler, Cesa 2019

Examples of equivariance in Geometric Deep Learning

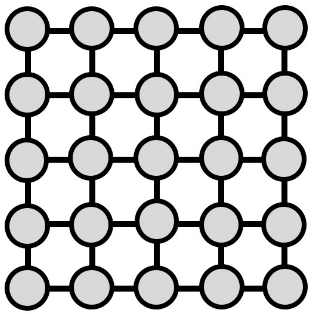
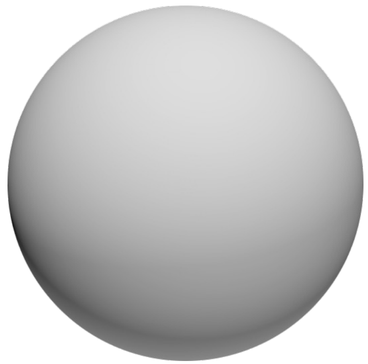
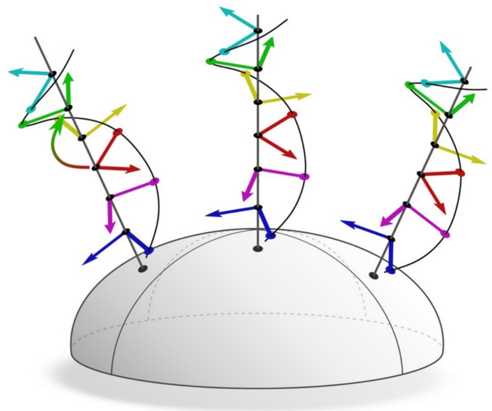
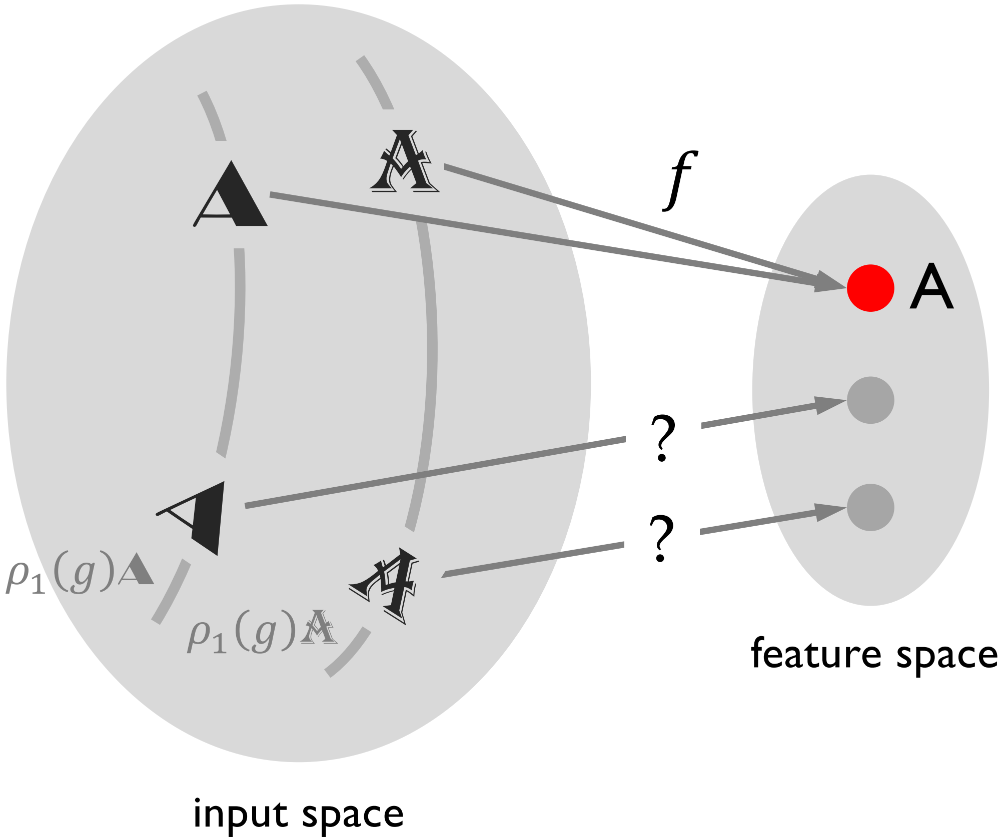
Domain Ω:	**Grid**	**Sphere**	**Manifold / Mesh**
Group:	Translation (Rotation, Reflection)	SO(3)	Gauge transformations e.g. SO(2)
			

Image: M. Weiler

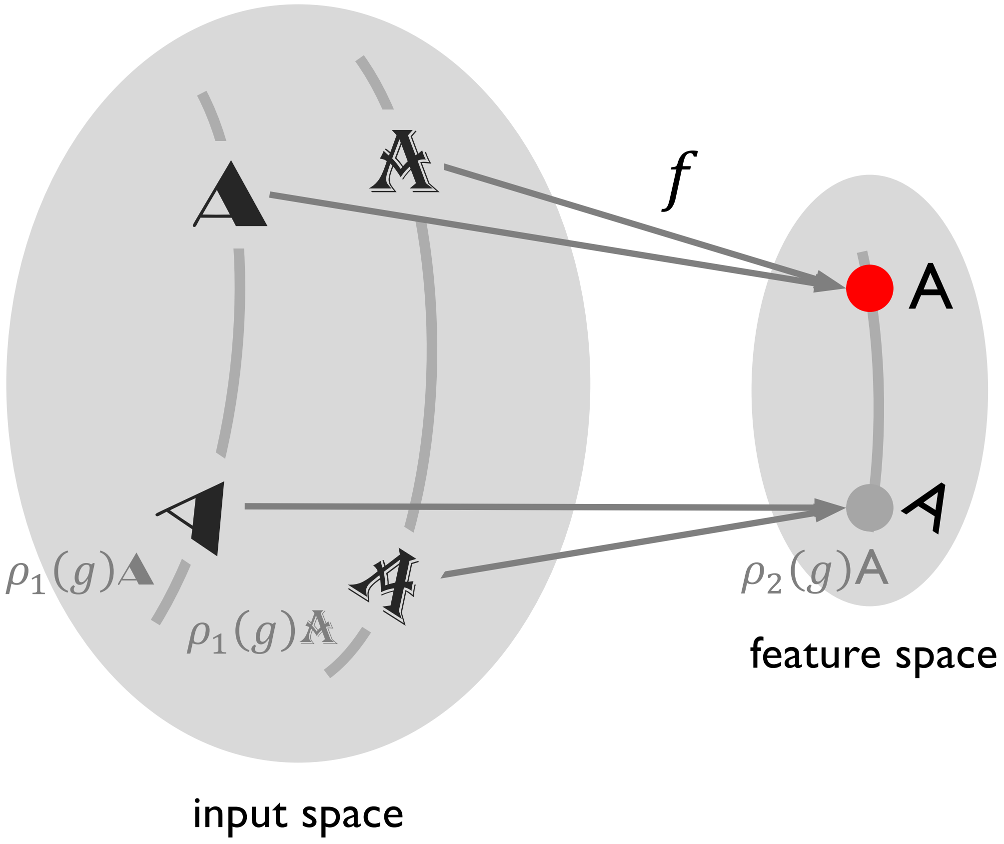
Equivariance = Symmetry-consistent generalisation



$$f(\rho_1(g)\mathbf{A}) = \rho_2(g)\, f(\mathbf{A})$$

$$f(\rho_1(g)\mathfrak{A}) = \rho_2(g)\, f(\mathfrak{A})$$

Equivariance = Symmetry-consistent generalisation

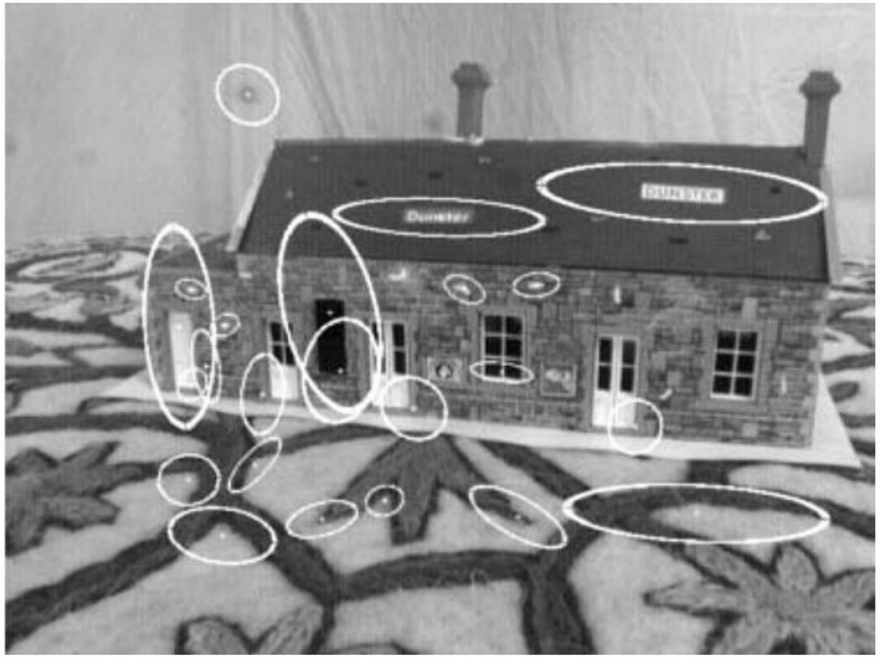


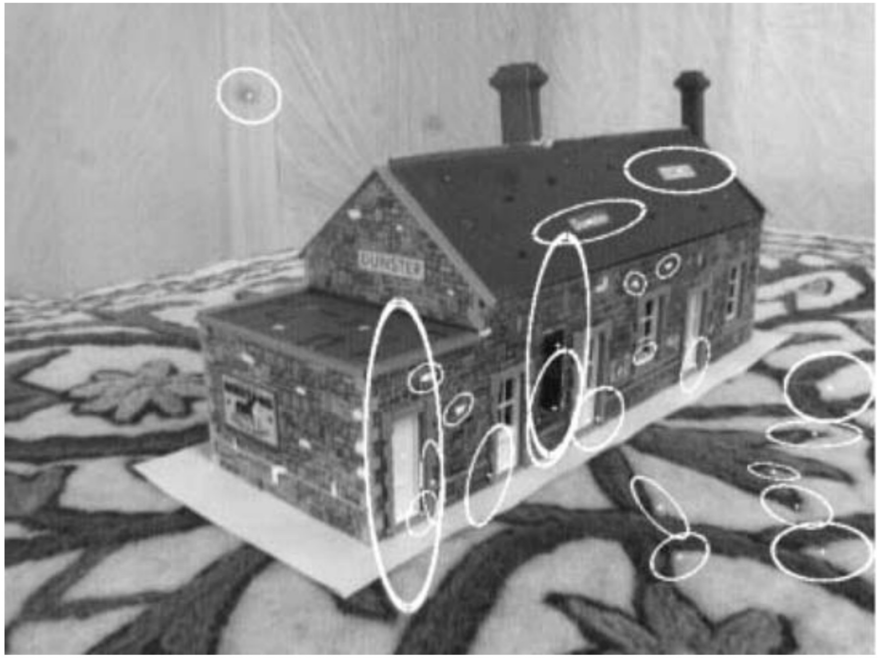
$$f(\rho_1(g)\mathbf{A}) = \rho_2(g)\, f(\mathbf{A})$$

$$\|$$

$$f(\rho_1(g)\mathfrak{A}) = \rho_2(g)\, f(\mathfrak{A})$$

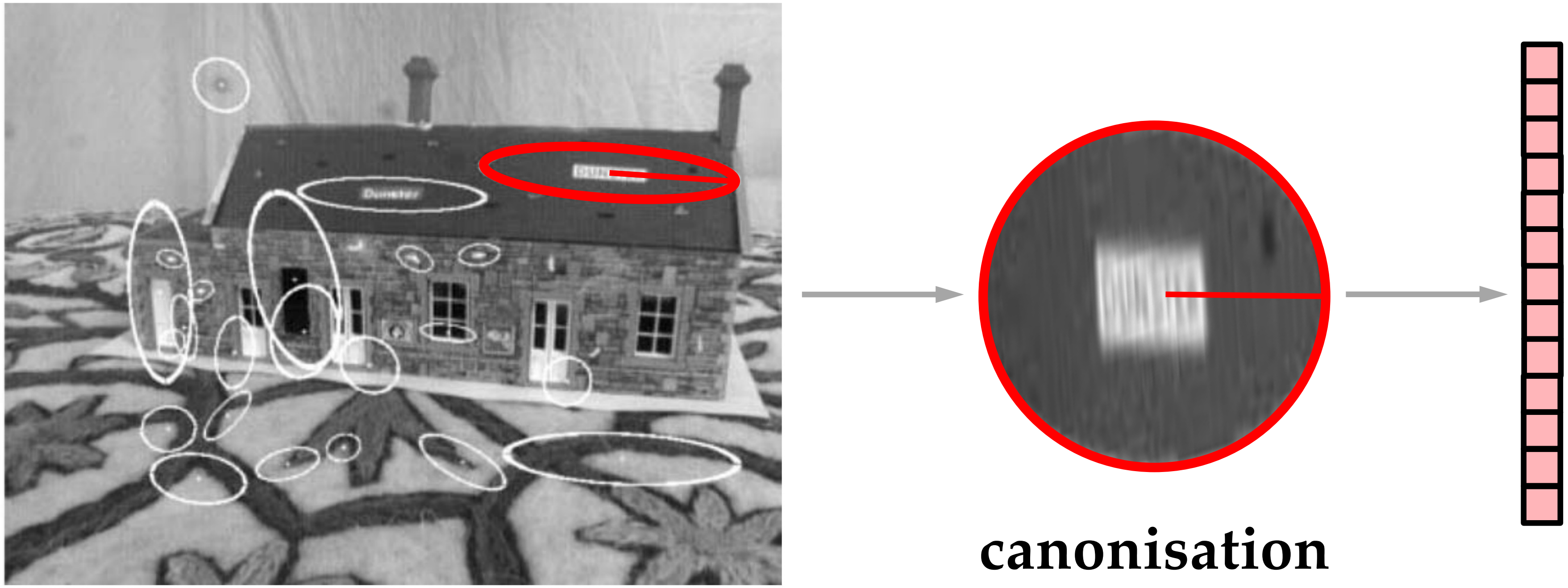
Canonisation



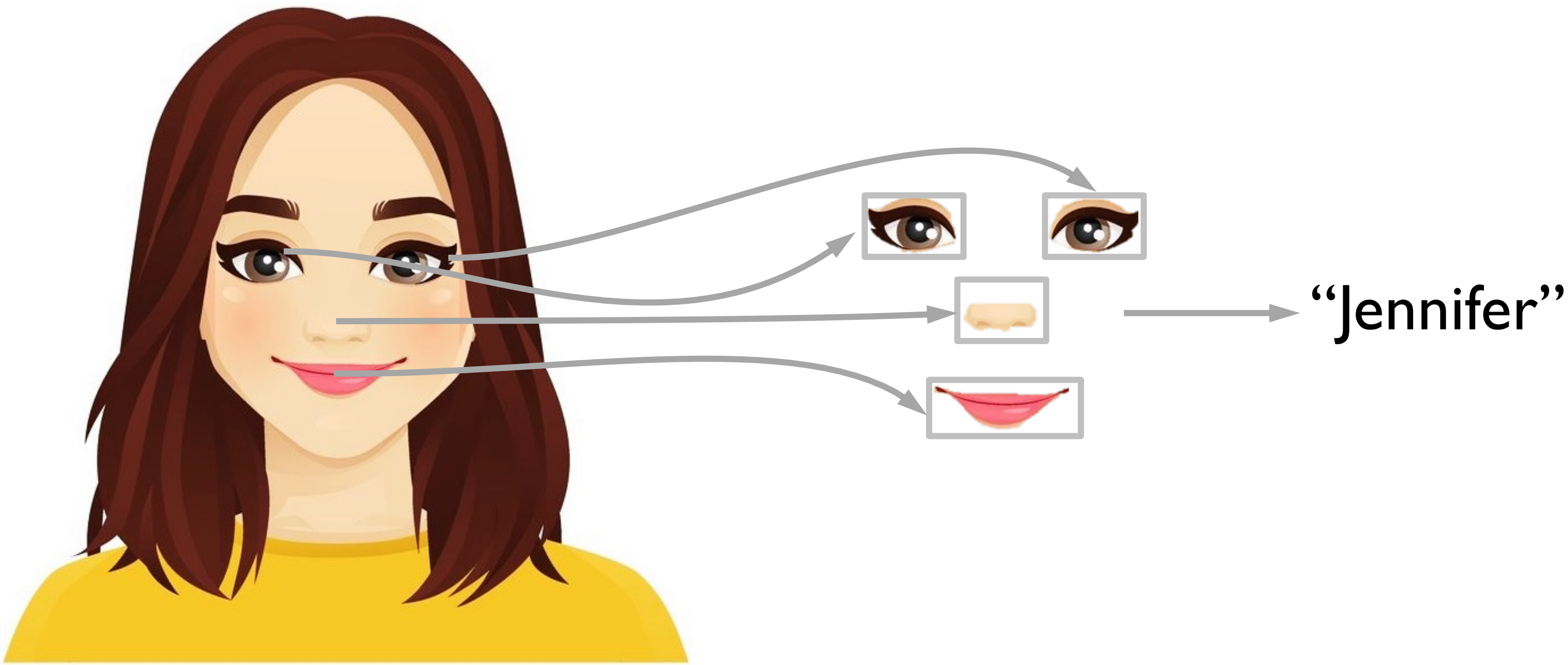


Mikolajczyk, Schmid 2004

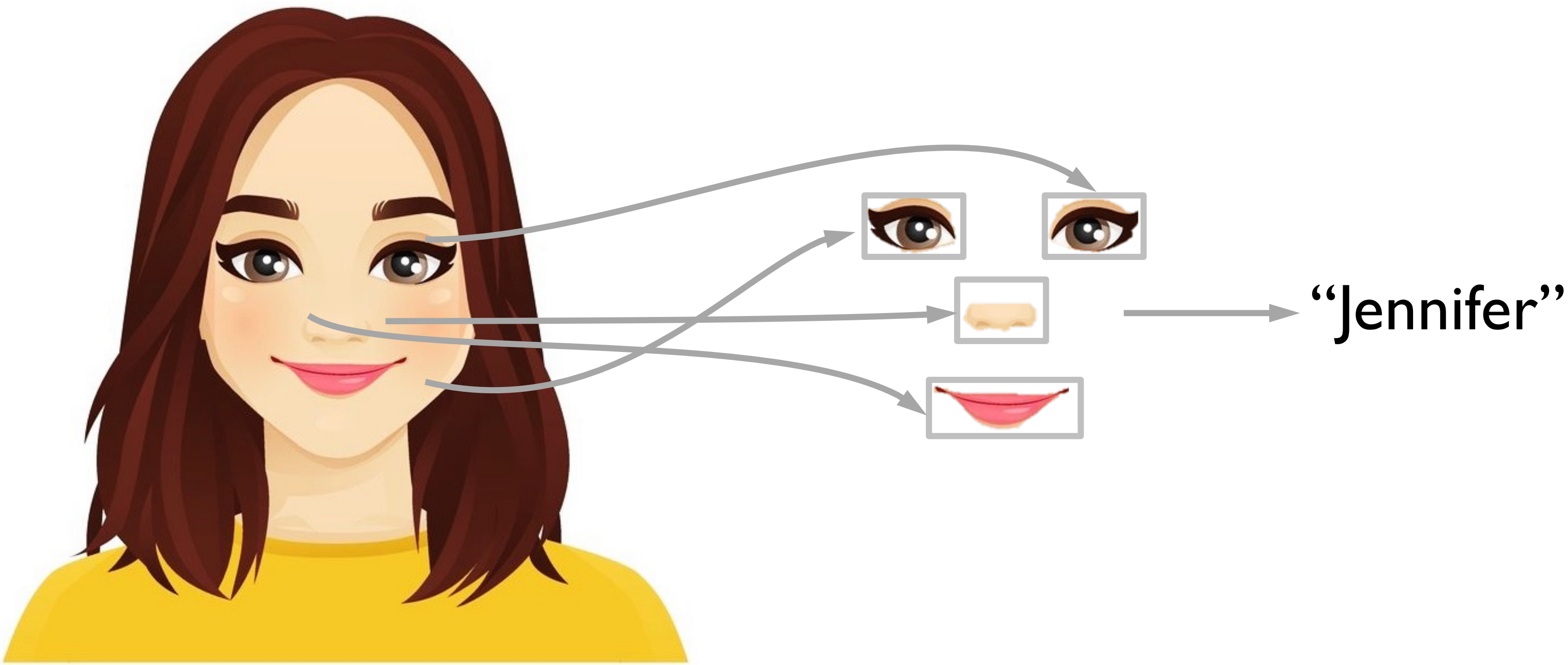
Canonisation



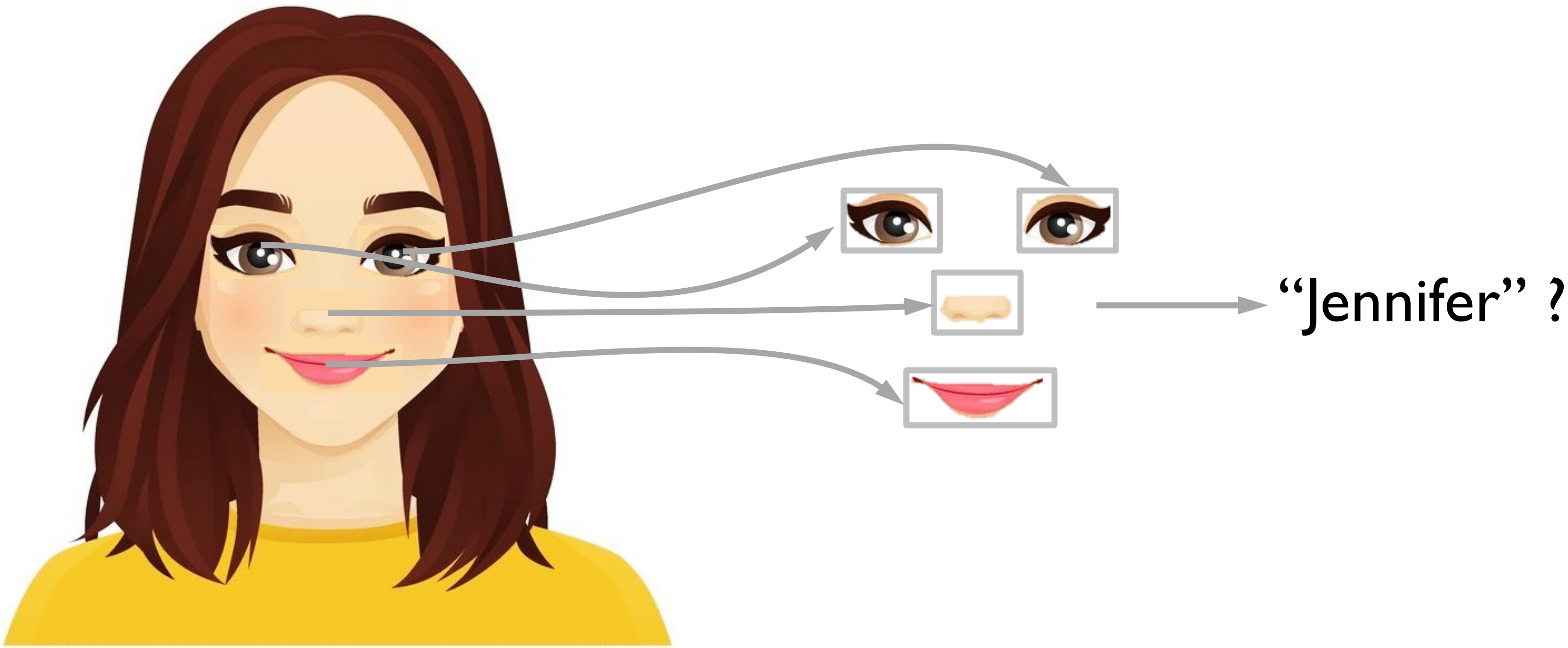
Canonisation vs Equivariance



Canonisation vs Equivariance



Canonisation vs Equivariance



Takeaways

- *Symmetries* are transformations leaving the object *invariant*
- In general ML, we care about symmetries of the *label function* and its *parameters* (neural network weights)
- In Geometric Deep Learning, we care about symmetries of a *geometric domain*, signals on which are inputs into a neural network
- Symmetry is exploited in deep learning in the form of *equivariant neural networks*
- In an equivariant neural network, each feature space is associated with a *group representation* and each layer is equivariant w.r.t. this representation
- *Invariance* is a special case of equivariance where the trivial representation is used
- Next lecture: learning under Invariance and Scale Separation geometric priors

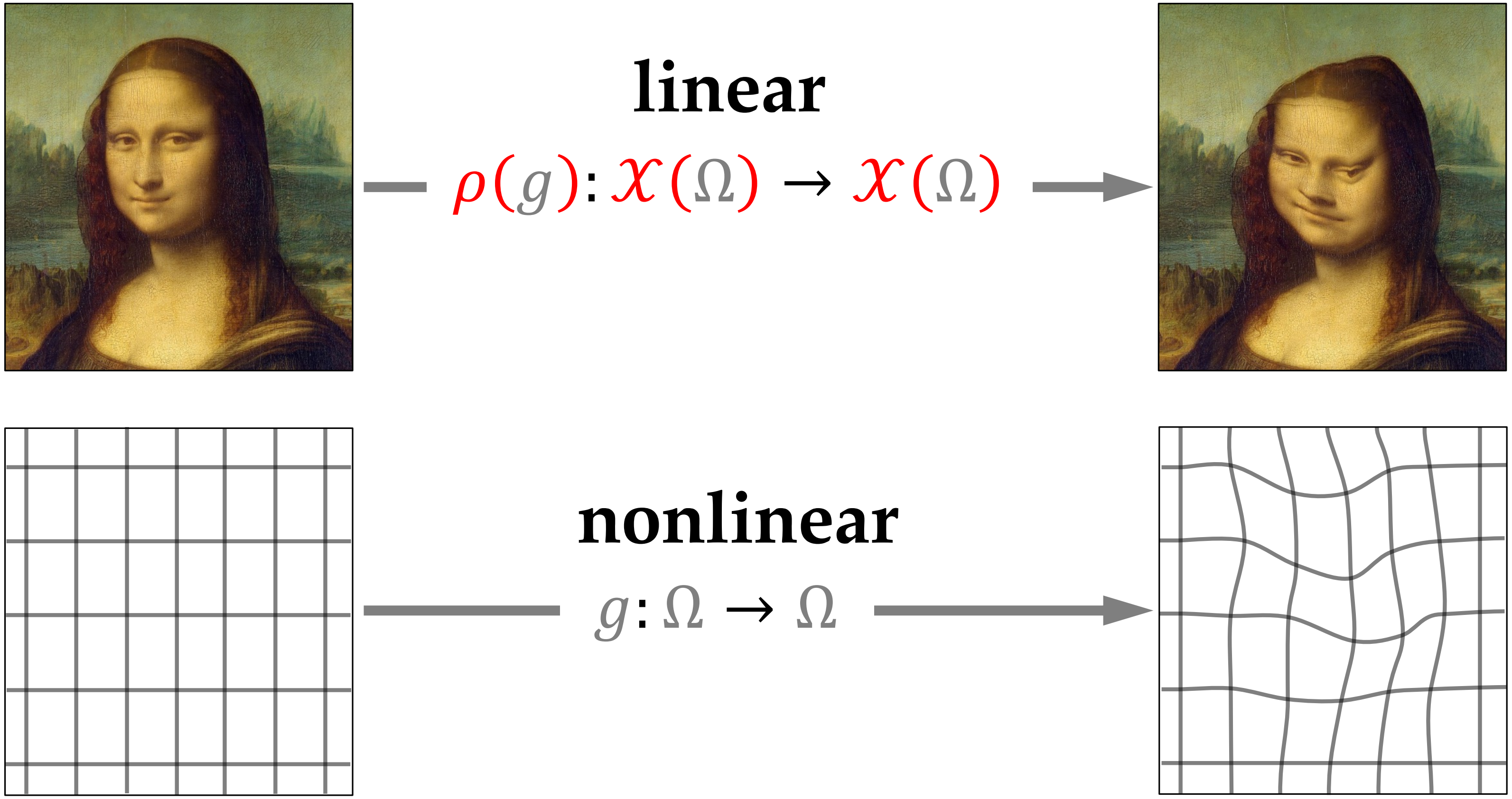
Key Concepts

- Symmetry Groups
- Group Actions and Representations
- Invariance and Equivariance

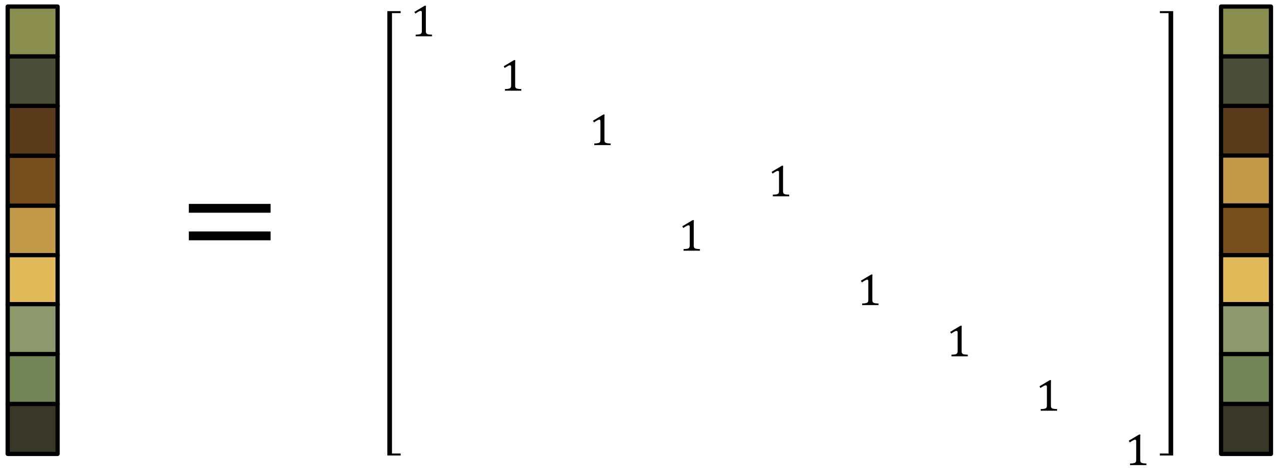
Outline

- Group theory provides the math language to describe symmetries in ML problems
- *Equivariant neural networks* are constructed such that each layer is equivariant w.r.t. the action of a symmetry group
- Symmetry prior leads to a new model class that however on its own may not tame the curse of dimensionality
- Symmetry prior is often combined with *Scale Separation*, typically implemented in the form of *pooling*
- These two geometric priors are the core of Geometric Deep Learning, a principled blueprint of highly expressive architectures that defy the curse of dimensionality

“Lifting”

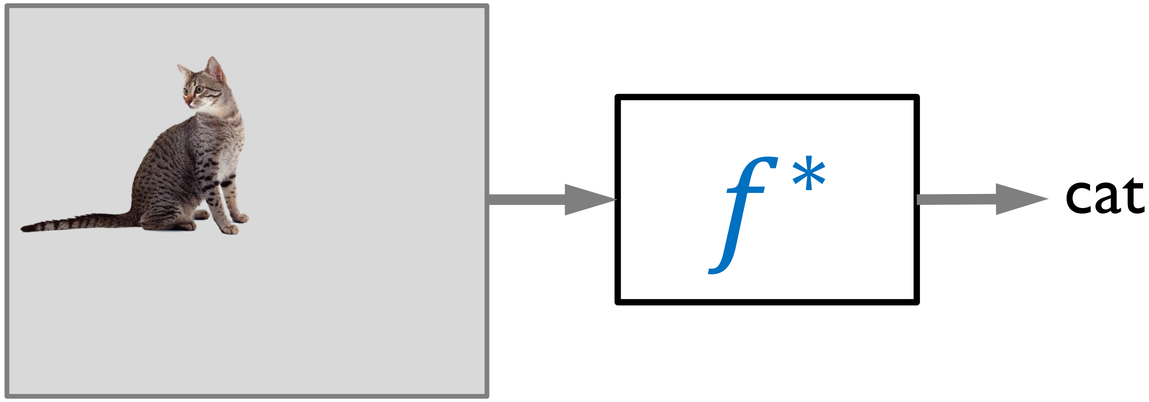

linear
$\rho(g) : \mathcal{X}(\Omega) \to \mathcal{X}(\Omega)$
nonlinear
$g : \Omega \to \Omega$

“Lifting”



“pixel permutation”

Invariant learning tasks



The function is *a priori* assumed to be shift-invariant,

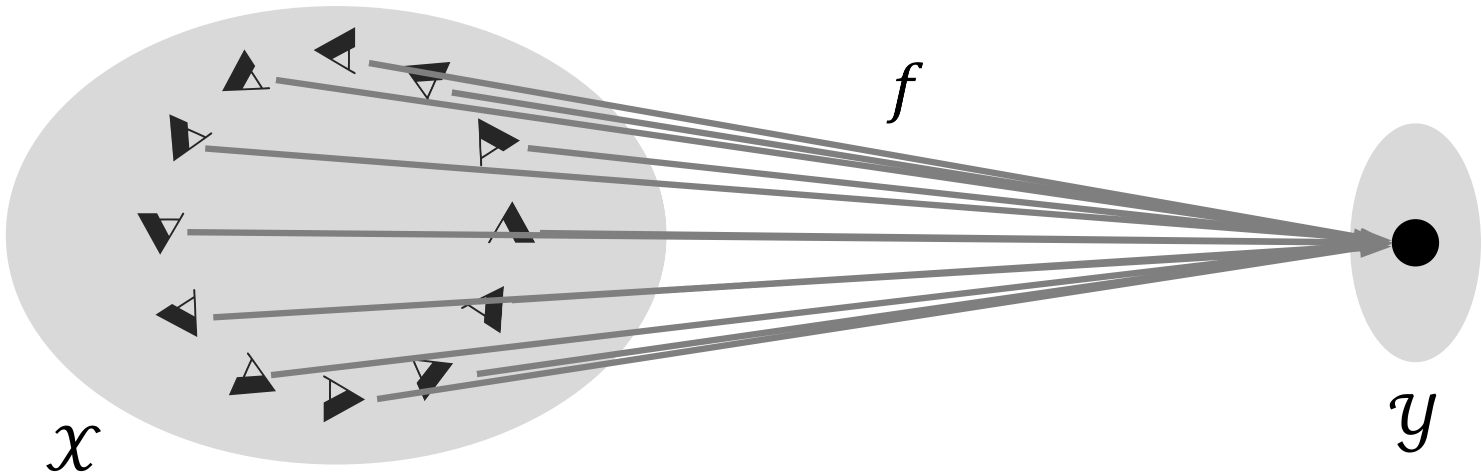
only one sample necessary per image

Data augmentation

The function is generic,

training set contains multiple shifted versions of each image

Group-invariant function classes



- G-**invariant** model class

$$\mathcal{F}_G = \{f: \mathcal{X} \to \mathcal{Y} \;\; \text{s.t.} \;\; f(gx) = f(x) \;\; \forall x \in \mathcal{X}, g \in G\}$$

  - How to leverage invariant function classes in learning?
  - Is this generally sufficient to break the curse of dimensionality?

Group averaging

- Assume G is discrete of finite size
- **Group averaging** (or **smoothing**) **operator** S_G (defined with abuse of notation as either $S_G\colon \mathcal{X} \to \mathcal{X}$ or $S_G\colon \mathcal{F}(\mathcal{X}) \to \mathcal{F}(\mathcal{X})$) averaging along group orbits

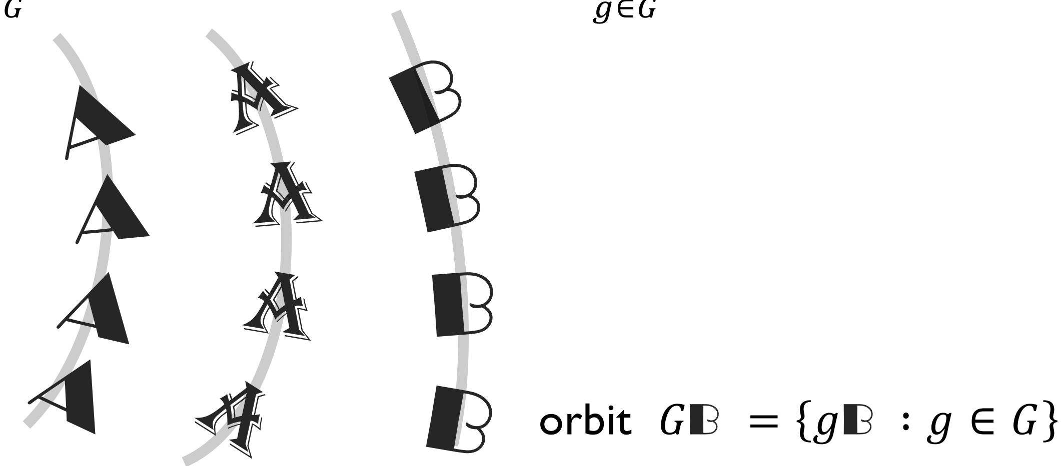
$$S_G x = \frac{1}{|G|} \sum_{g \in G} gx \qquad\qquad S_G f(x) = \frac{1}{|G|} \sum_{g \in G} f(gx)$$

Note: More generally, we can define $S_G f(x) = \frac{1}{\mu(G)} \int_G f(gx) \mathrm{d}\mu(g)$, where μ is the *Haar measure* on the group

Group averaging

- Assume G is discrete of finite size
- **Group averaging** (or **smoothing**) **operator** S_G (defined with abuse of notation as either $S_G\colon \mathcal{X} \to \mathcal{X}$ or $S_G\colon \mathcal{F}(\mathcal{X}) \to \mathcal{F}(\mathcal{X})$) averaging along group orbits

$$S_G x = \frac{1}{|G|}\sum_{g\in G} gx \qquad\qquad S_G f(x) = \frac{1}{|G|}\sum_{g\in G} f(gx)$$



Group averaging

- Assume G is discrete of finite size
- **Group averaging** (or **smoothing**) **operator** S_G (defined with abuse of notation as either $S_G\colon \mathcal{X} \to \mathcal{X}$ or $S_G\colon \mathcal{F}(\mathcal{X}) \to \mathcal{F}(\mathcal{X})$) averaging along group orbits

$$S_G x = \frac{1}{|G|} \sum_{g \in G} gx \qquad\qquad S_G f(x) = \frac{1}{|G|} \sum_{g \in G} f(gx)$$

Assume f is G-**invariant**. Then, $f(Gx) = \text{const.}$

Group averaging

- Assume G is discrete of finite size
- **Group averaging** (or **smoothing**) **operator** S_G (defined with abuse of notation as either $S_G: \mathcal{X} \to \mathcal{X}$ or $S_G: \mathcal{F}(\mathcal{X}) \to \mathcal{F}(\mathcal{X})$) averaging along group orbits

$$S_G x = \frac{1}{|G|} \sum_{g \in G} gx \qquad\qquad S_G f(x) = \frac{1}{|G|} \sum_{g \in G} f(gx)$$

Assume f is G**-invariant**. Then, $S_G f = f$.

- Given a hypothesis class $\mathcal{F}$, we can make it G-invariant by applying the group averaging operator, $S_G \mathcal{F} = \{S_G f,\ f \in \mathcal{F}\}$.

Exercise: Let $\Omega = \{1, \dots, d\}$ a grid, $G = C_d$ cyclic group, and $\mathcal{F}$ =polynomials of degree k. Write $S_G \mathcal{F}$.

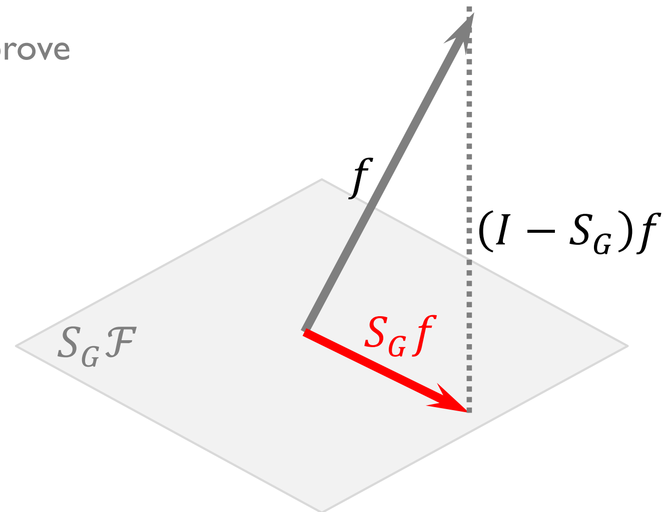
Learning under invariance

Approximation error is unaffected by group smoothing, i.e.,

$$\inf_{f \in \mathcal{F}} \|f - f^*\|^2 = \inf_{f \in S_G \mathcal{F}} \|f - f^*\|^2$$

- Since S_G is an orthogonal projection in L_2: **Exercise:** prove

$$\|f - f^*\|^2 = \|S_G(f - f^*)\|^2 + \|(I - S_G)(f - f^*)\|^2$$
$$= \|S_G f - f^*\|^2 + \|(I - S_G)f\|^2$$



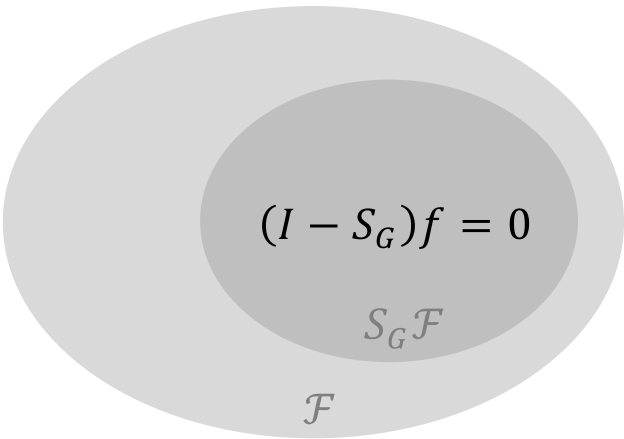
Learning under invariance

Approximation error is unaffected by group smoothing, i.e.,

$$\inf_{f \in \mathcal{F}} \|f - f^*\|^2 = \inf_{f \in S_G \mathcal{F}} \|f - f^*\|^2$$

- Since S_G is an orthogonal projection in L_2:

$$\|f - f^*\|^2 = \|S_G(f - f^*)\|^2 + \|(I - S_G)(f - f^*)\|^2$$
$$= \|S_G f - f^*\|^2 + \|(I - S_G)f\|^2$$



- Statistical error is reduced… but by how much?

Learning invariant Lipschitz functions

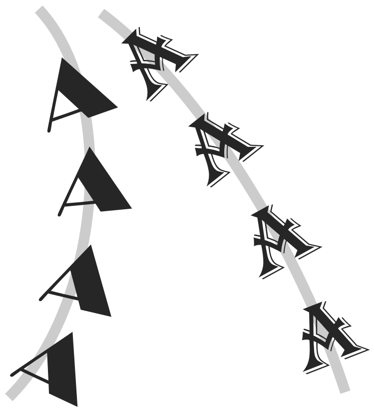
- Consider the class of **Lipschitz functions**

$$\mathcal{F} = \{f: \mathcal{X} \subseteq \mathbb{R}^d \to \mathbb{R} \quad \text{s.t.} \quad |f(x) - f(x')| \leq \beta \|x - x'\| \quad \forall x, x' \in \mathcal{X}\}$$

- Group-averaged Lipschitz class

$$S_G\mathcal{F} = \left\{f: \mathcal{X} \subseteq \mathbb{R}^d \to \mathbb{R} \quad \text{s.t.} \quad |f(x) - f(x')| \leq \beta \inf_{g\in G} \|x - gx'\| \quad \forall x, x' \in \mathcal{X}\right\}$$

"points in nearby orbits are not mapped too far away"



Learning invariant Lipschitz functions

Theorem: Using G-invariant kernel ridge regression, the generalisation error of learning a G-invariant d-dimensional Lipschitz function from N samples is bounded by

$$\mathbb{E}\left(R(\hat{f}) - R(f^*)\right) \lesssim (|G|N)^{-1/d}$$

- Sharp gains w.r.t. non-invariant kernels
- Group size $|G|$ can be exponential in dimension
- Rate can still be dimensionality-cursed, suggesting invariance alone is insufficient

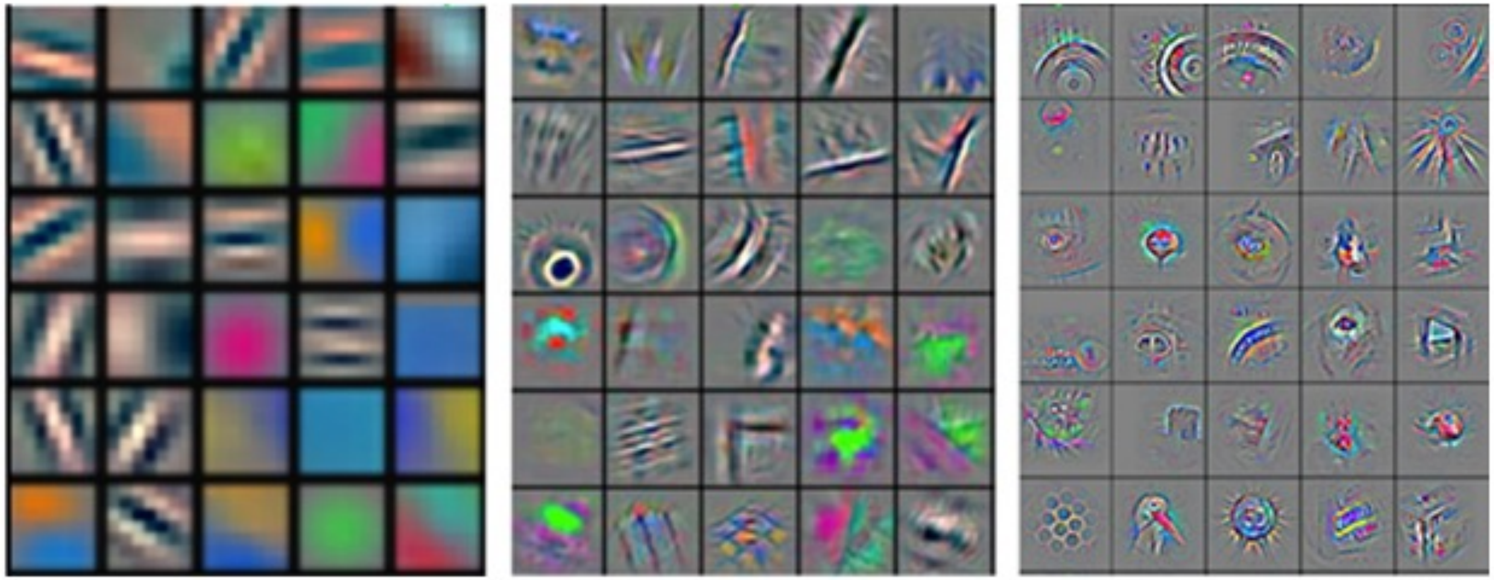
Bietti, Venturi, Bruna 2021

Conclusions so far

- Using known global symmetries in hypothesis class is a *no-brainer*: guaranteed improvement in sample complexity
- Might not break the curse of dimensionality. What else is missing?
- How to build such invariant classes efficiently? I.e., we need an algorithmic recipe

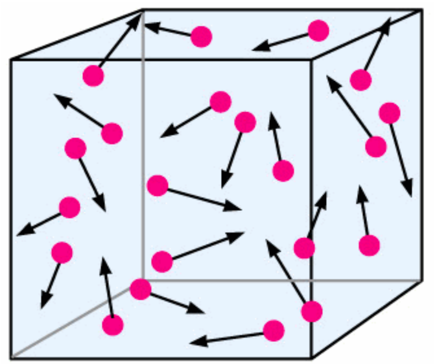
SCALE SEPARATION

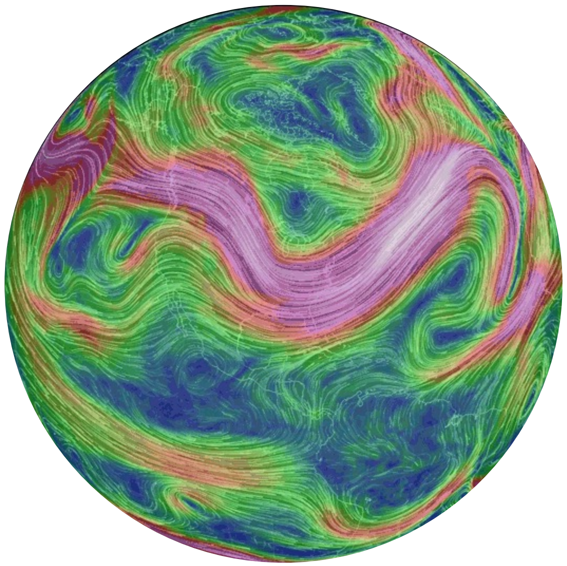
Compositionality in Deep Learning

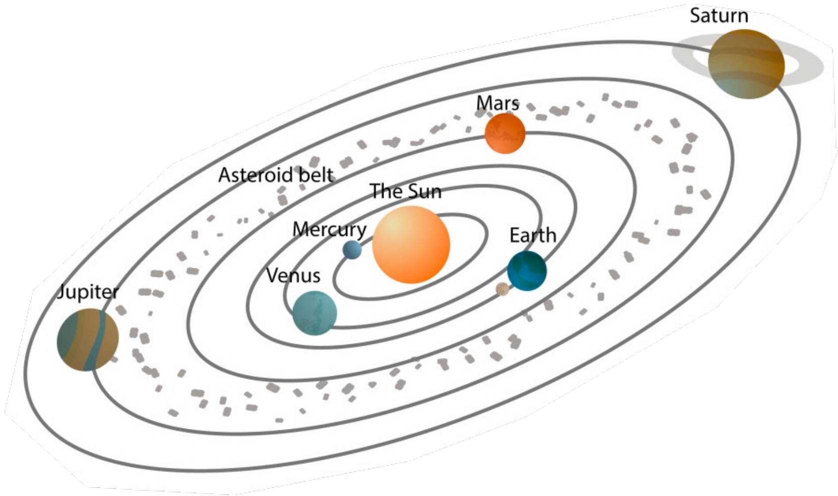


Increasingly complex features in deeper layers of a convolutional neural network

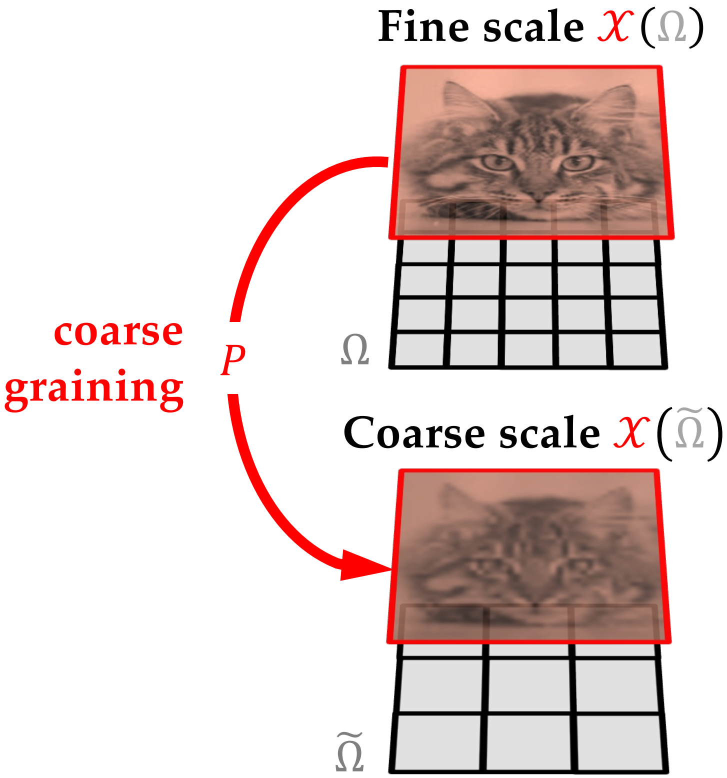
Compositionality in Physics







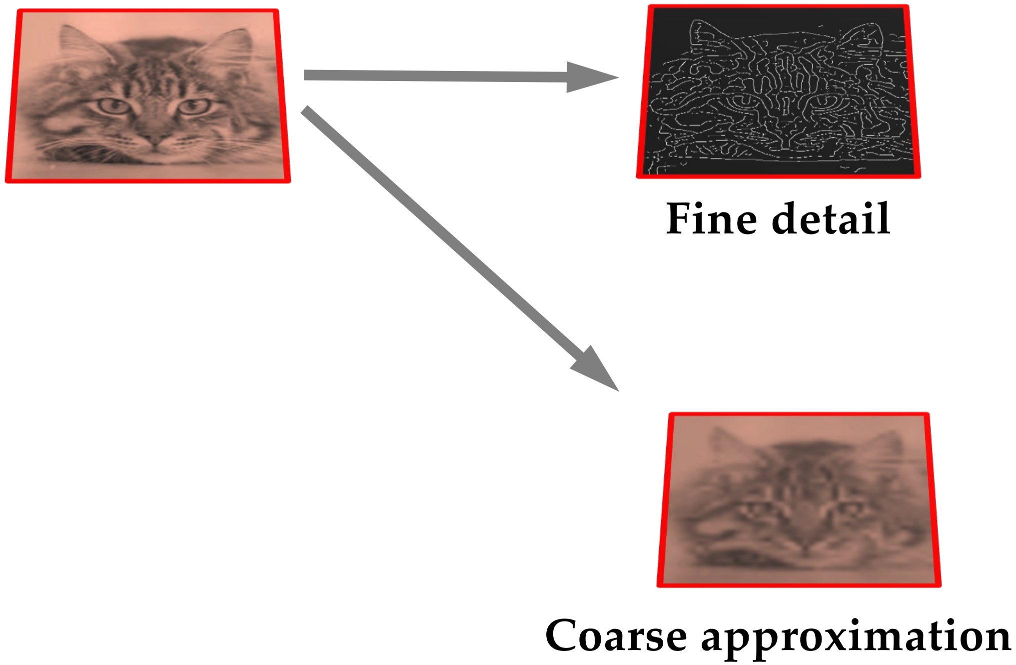
Multiresolution Analysis



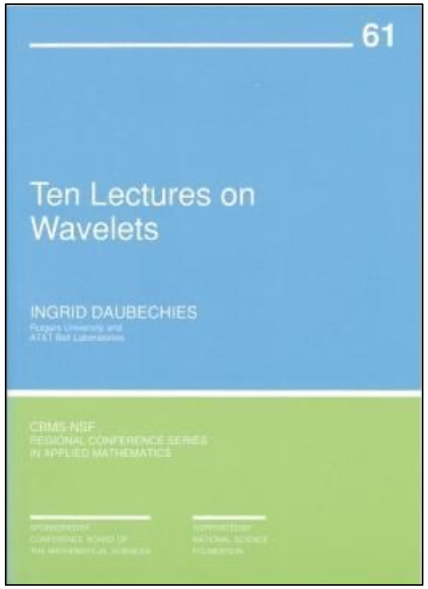
- Hierarchy of domains $\ldots \subset \tilde{\Omega} \subset \Omega$
- Hierarchy signal spaces $\mathcal{X}(\Omega), \mathcal{X}(\tilde{\Omega}), \ldots$
- Coarse graining operator

$$P: \mathcal{X}(\Omega) \to \mathcal{X}(\tilde{\Omega})$$

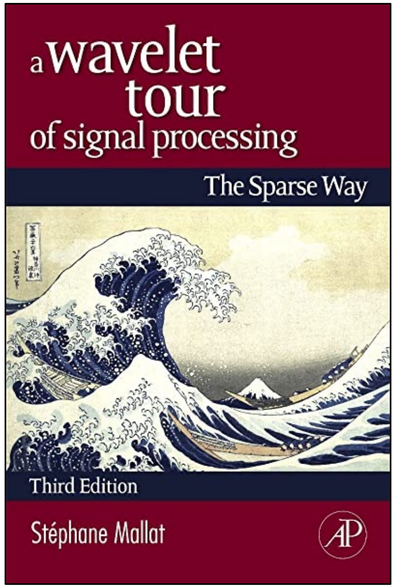
Multiresolution Analysis



Wavelets



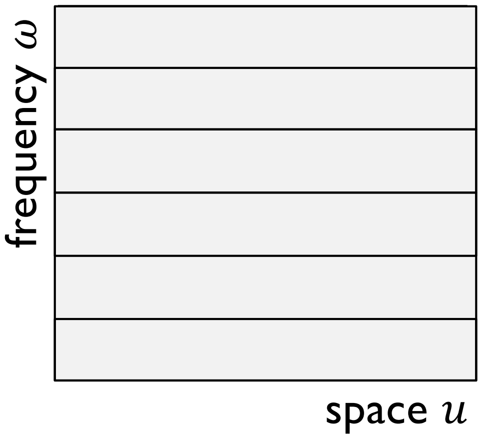
I. Daubechies



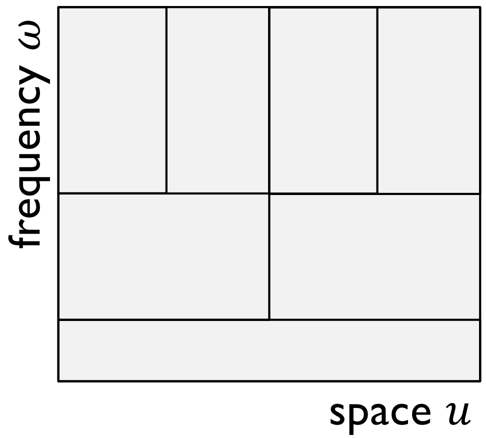
S. Mallat

Wavelets vs Fourier

Wavelets vs Fourier



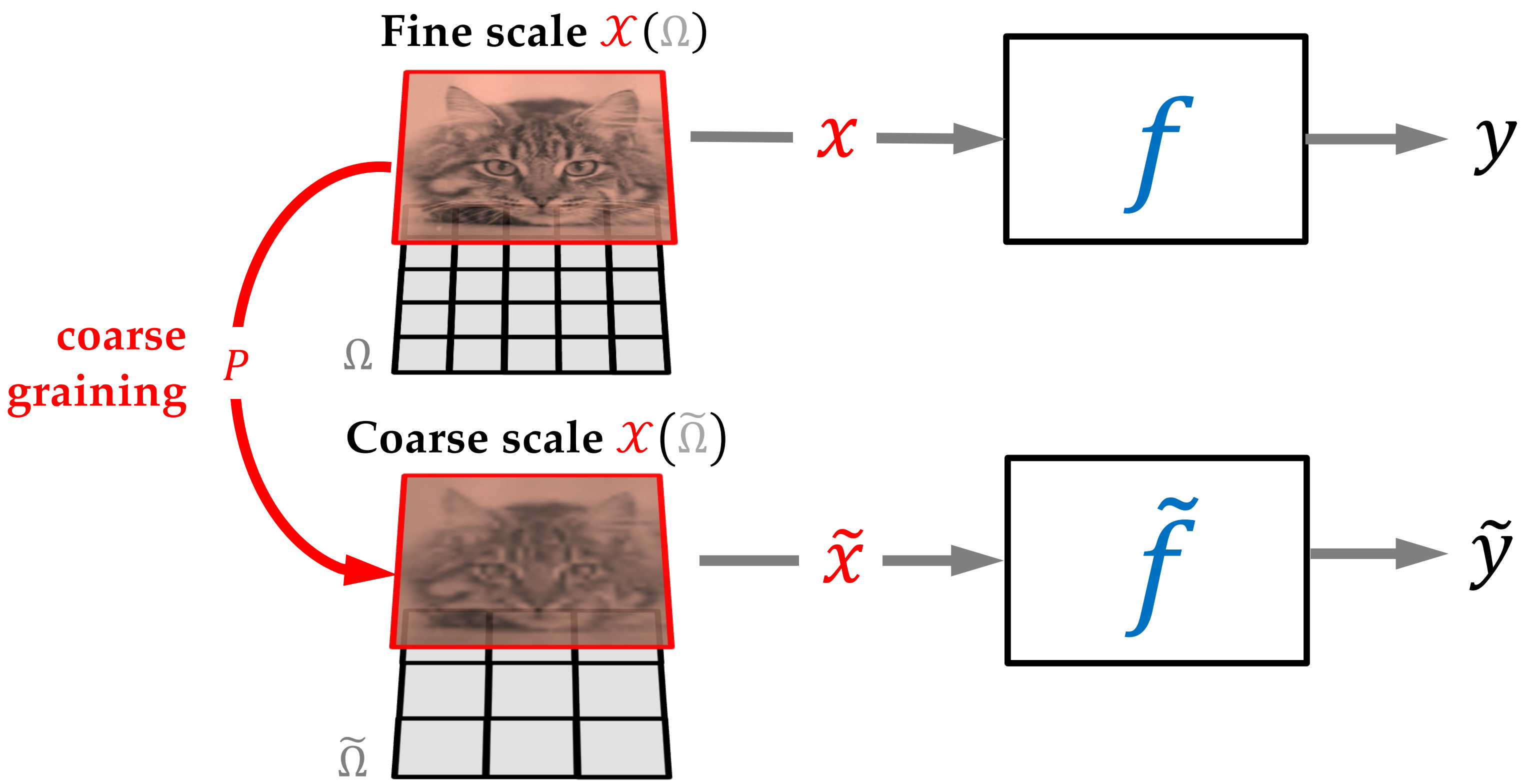
J. Fourier

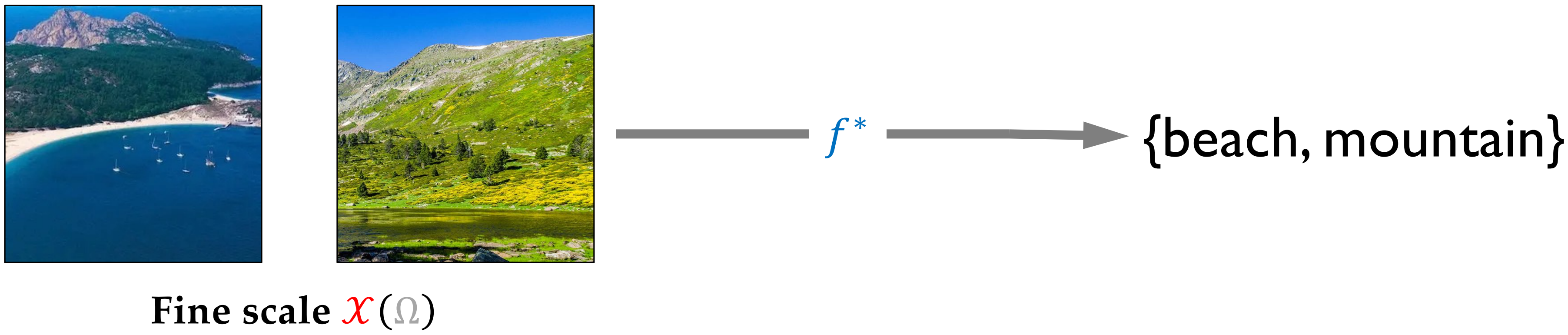


I. Daubechies

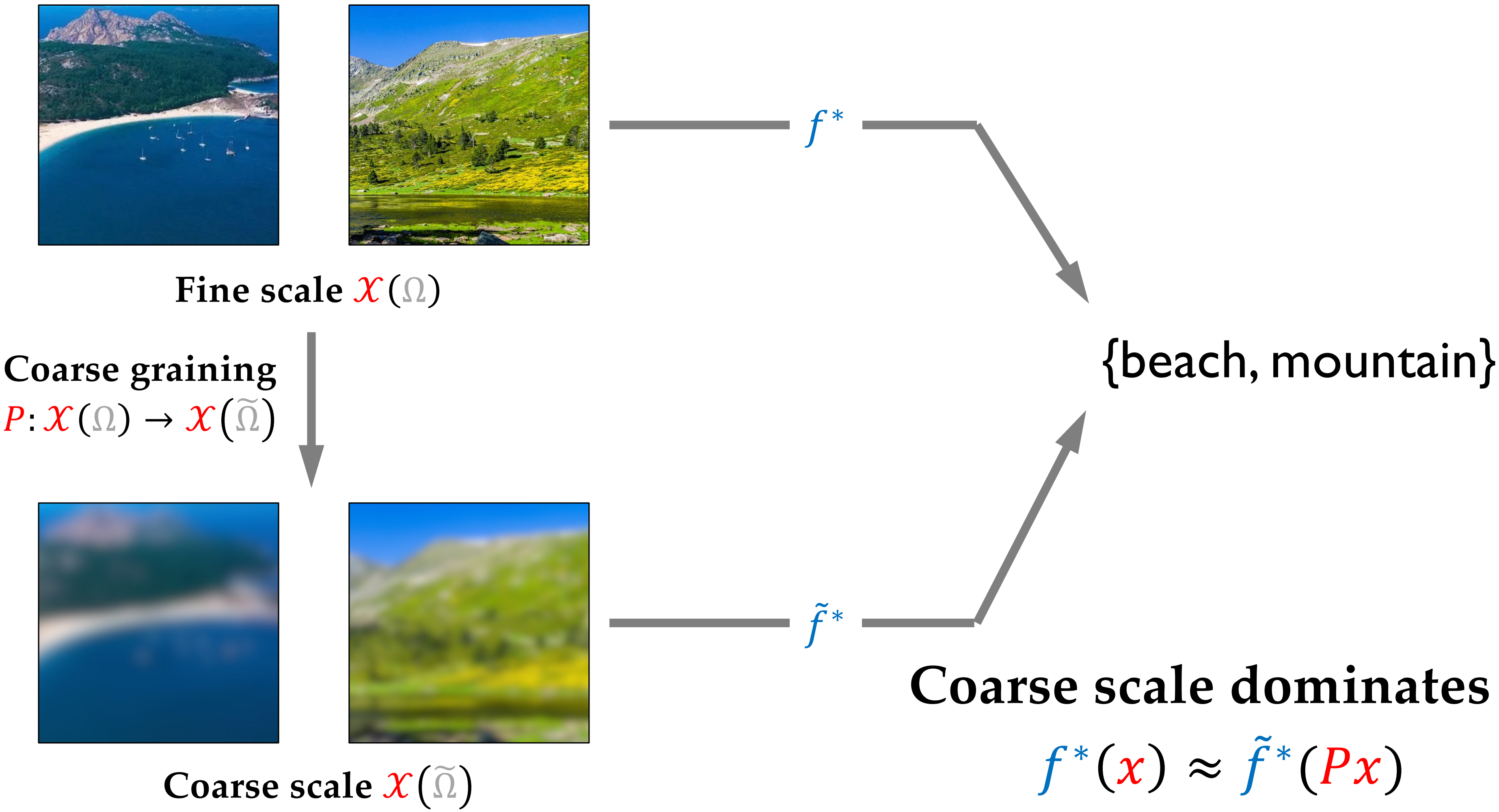
S. Mallat

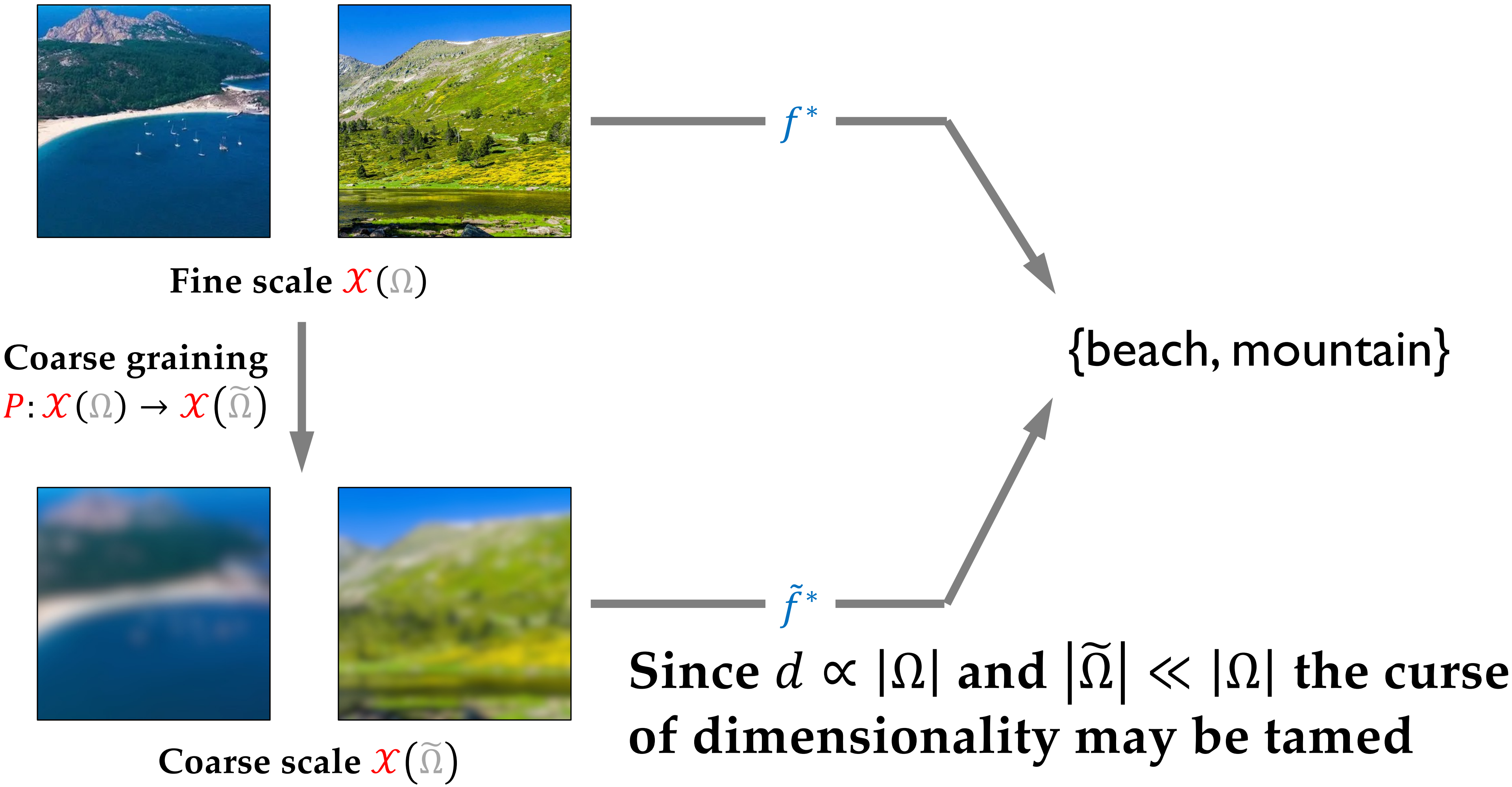
Multiresolution Analysis in ML

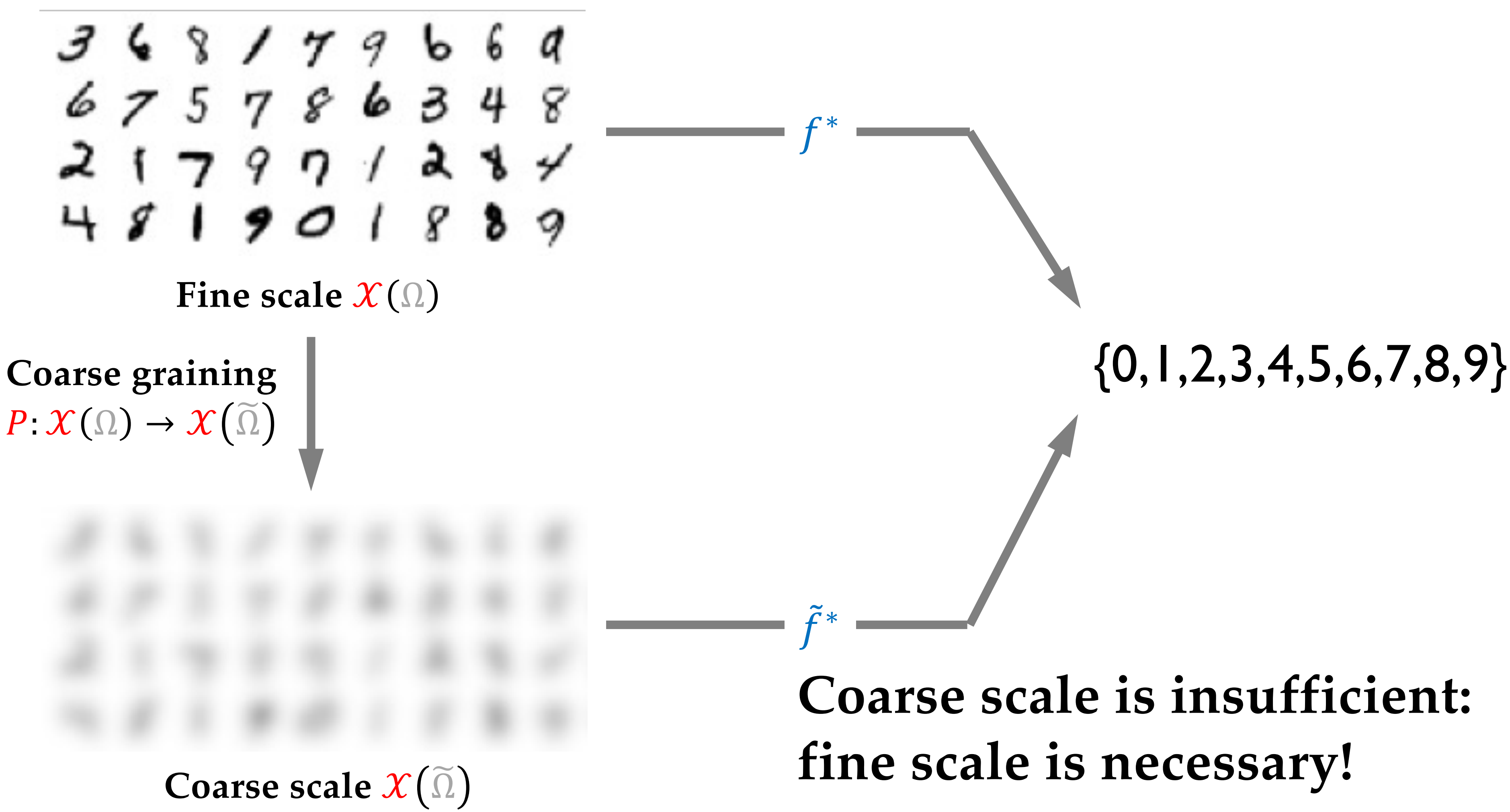


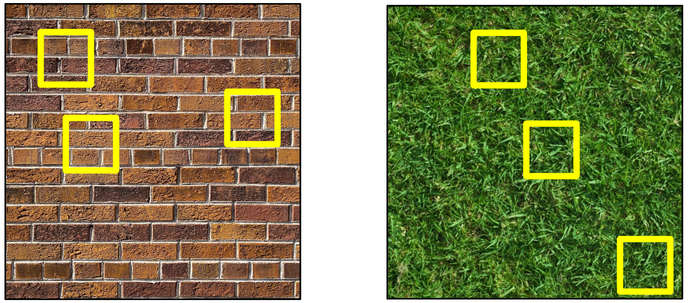


f^*
{beach, mountain}
Fine scale $\mathcal{X}(\Omega)$


Fine scale $\mathcal{X}(\Omega)$
Coarse graining
$P: \mathcal{X}(\Omega) \to \mathcal{X}(\tilde{\Omega})$
Coarse scale $\mathcal{X}(\tilde{\Omega})$
f^*
$\tilde{f}^*$
{beach, mountain}
Coarse scale dominates
$f^*(x) \approx \tilde{f}^*(Px)$


Fine scale $\mathcal{X}(\Omega)$
Coarse graining
$P: \mathcal{X}(\Omega) \to \mathcal{X}(\tilde{\Omega})$
Coarse scale $\mathcal{X}(\tilde{\Omega})$
f^*
$\tilde{f}^*$
{beach, mountain}
Since $d \propto |\Omega|$ and $|\tilde{\Omega}| \ll |\Omega|$ the curse of dimensionality may be tamed


Fine scale $\mathcal{X}(\Omega)$
Coarse graining
$P: \mathcal{X}(\Omega) \to \mathcal{X}(\tilde{\Omega})$
Coarse scale $\mathcal{X}(\tilde{\Omega})$
f^*
$\tilde{f}^*$
{0,1,2,3,4,5,6,7,8,9}
Coarse scale is insufficient:
fine scale is necessary!

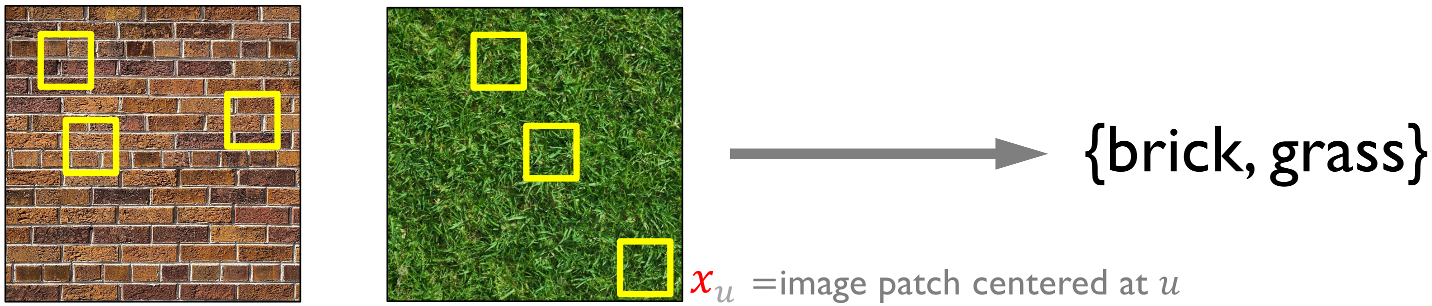


→ {brick, grass}

x_u =image patch centered at u

Fine scale dominates

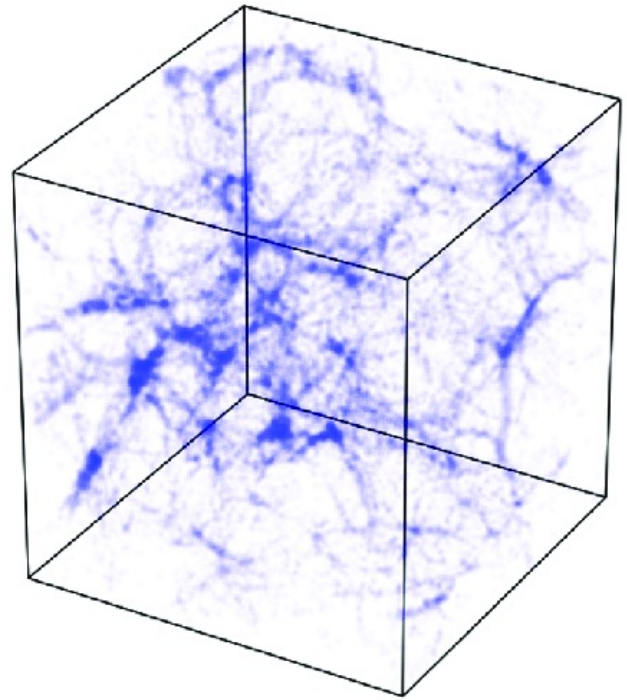
$$f^*(x) \approx \sum_u g(x_u)$$



Since $d \propto$ patch size, the curse of dimensionality can be avoided

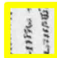
Local phenomena in Physics

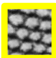
$$\frac{\mathrm{d}^2\mathbf{x}_i}{\mathrm{d}t^2} = \sum_{\substack{j=1 \\ j\neq i}}^{N} G m_j \frac{(\mathbf{x}_i - \mathbf{x}_j)}{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^3}$$



N-body system

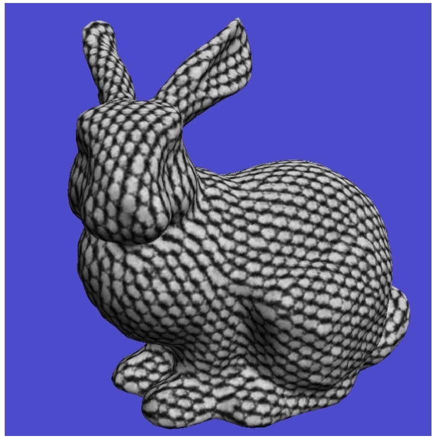
Local vs Global

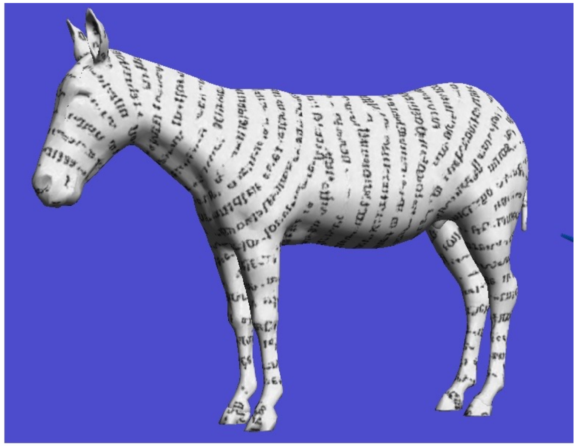


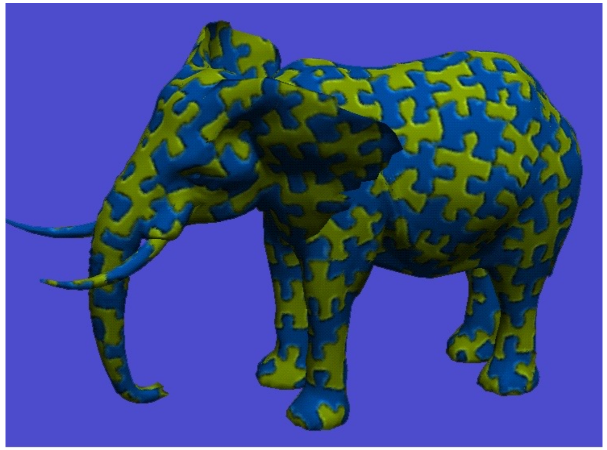


?

Local vs Global



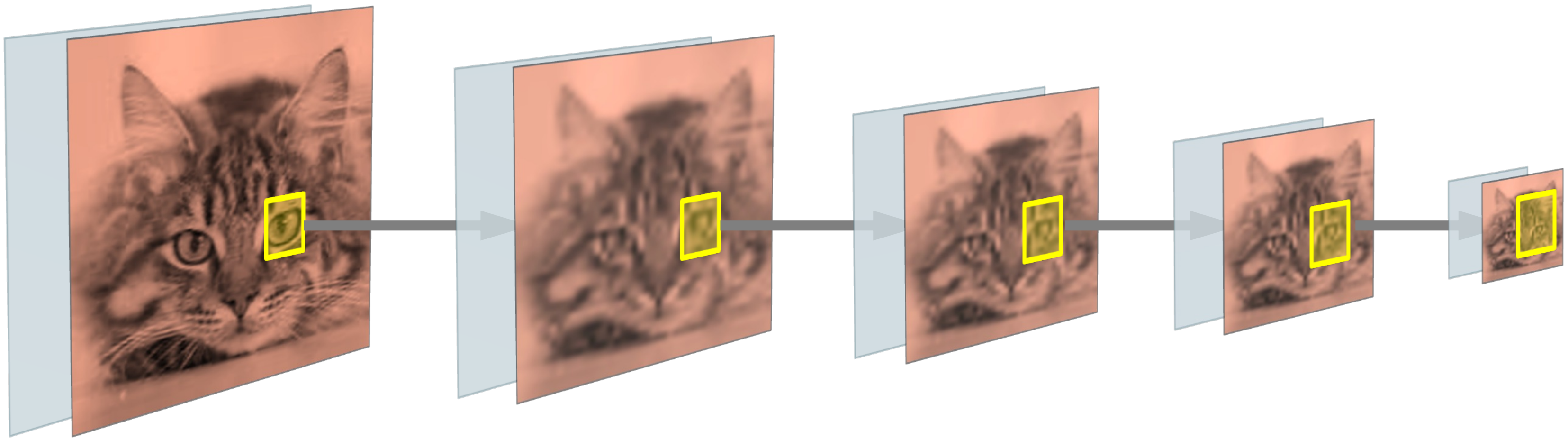




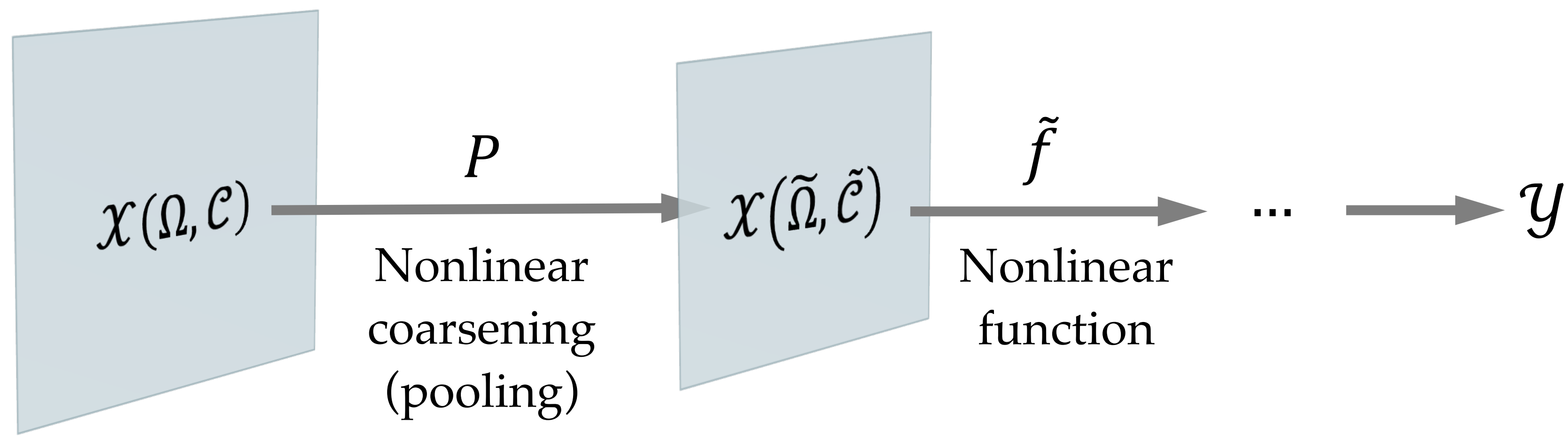
Local patches do not convey information about global structure

Turk 2001

Multiscale compositional priors

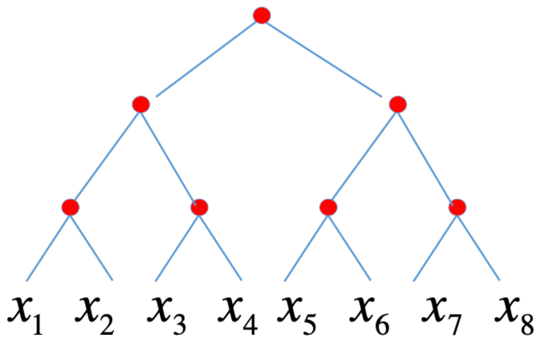


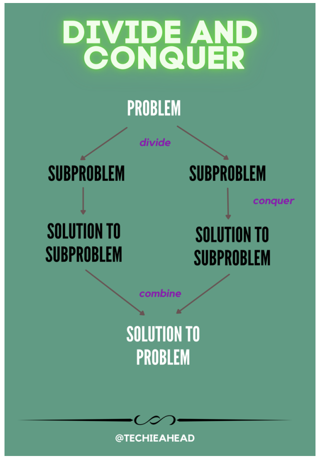
Multiscale compositional priors

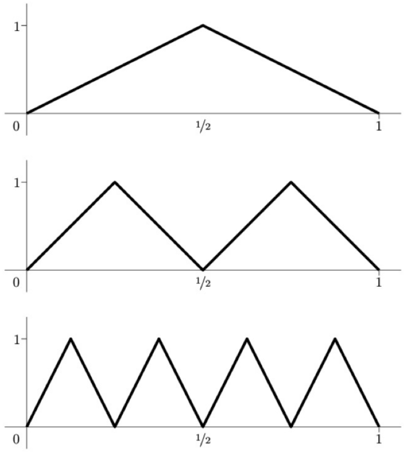


Benefits of composition

- Provable approximation, estimation, and computational benefits in specific contexts
- General structure of multiscale hypothesis spaces is still not completely understood theoretically
- Combining Symmetry and Scale Separation priors gives powerful model from first principles







Telgarsky 2015; Cohen & Shashua 2016

THE BLUEPRINT

Combining Invariance with Scale Separation

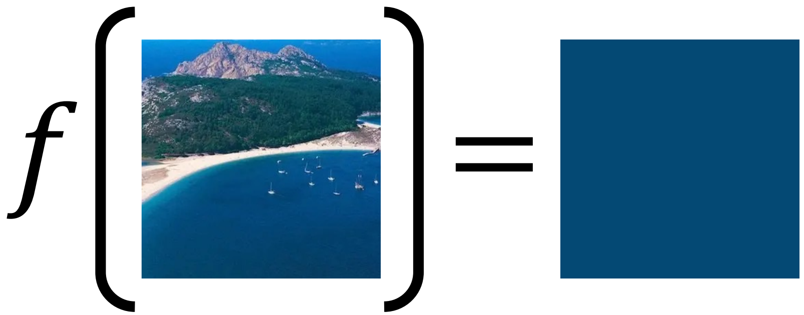
- Our hypothesis class wish list:
  - Group invariance
  - Multiscale structure
  - Expressivity
- What neural network architecture can satisfy these desiderata?

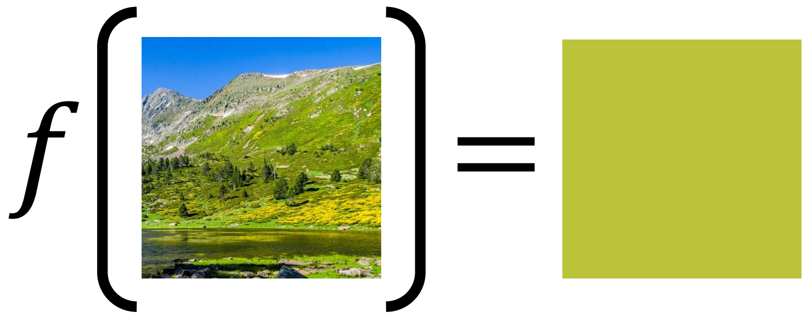
Linear group invariants

> Let $f : \mathcal{X} \to \mathbb{R}$ be **linear G-invariant**. Then $f(x) = f(S_G x)$ for all $x \in \mathcal{X}$, i.e., group average is the only linear group invariant.

Exercise: prove

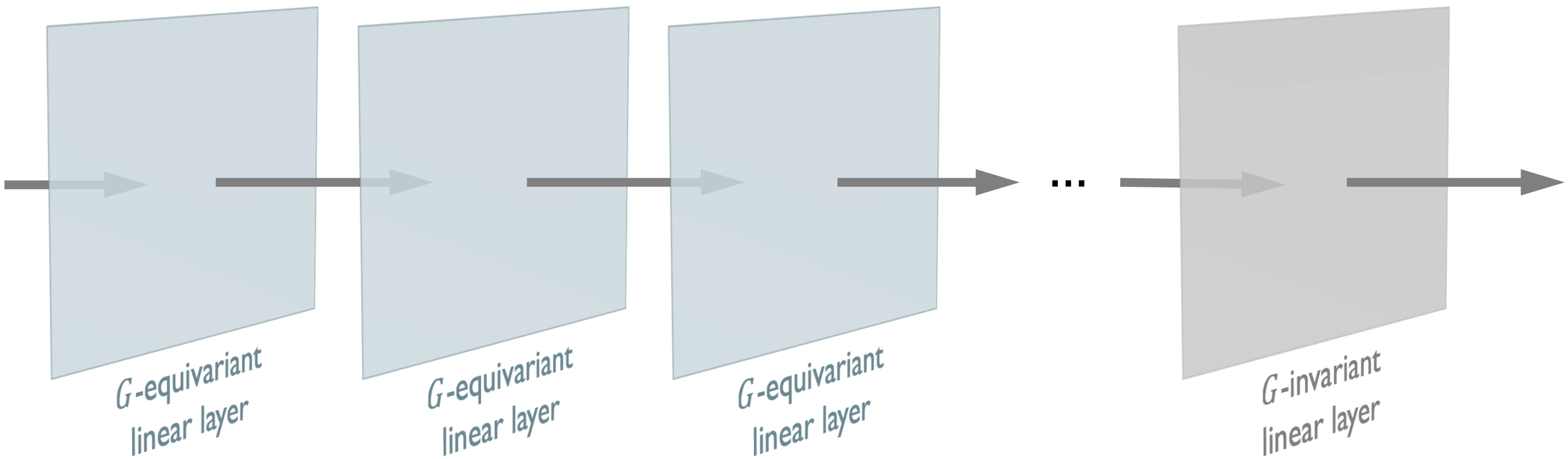
- Linear invariants are **not expressive**: f depends on x through the group average $S_G x$
- In case of images with translation, it would amount to using only the average colour!





Linear group equivariants

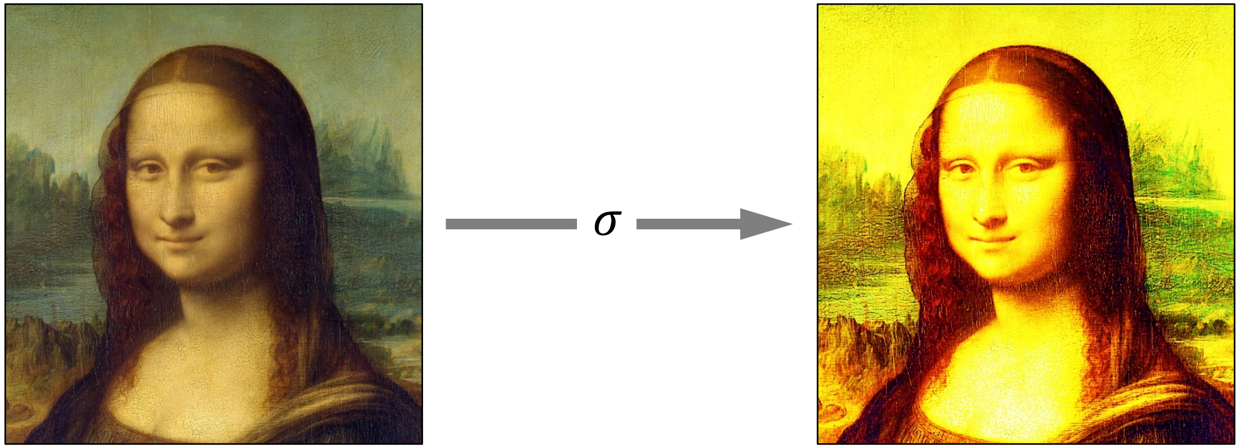
- Assume $f\colon \mathcal{X} \to \mathcal{X}'$ is **linear G-equivariant**, i.e., is linear and satisfies $f(gx) = gf(x)$ for all $x \in \mathcal{X}$ and $g \in G$
- Many examples in deep learning:
  - Convolutions in CNNs (equivariant w.r.t. translation)
  - Message passing in GNNs (equivariant w.r.t. permutation)
- Can we combine linear equivariants with a linear invariant?



G-equivariant linear layer
G-equivariant linear layer
G-equivariant linear layer
...
G-invariant linear layer

G-invariant
linear layer

Element-wise nonlinearity



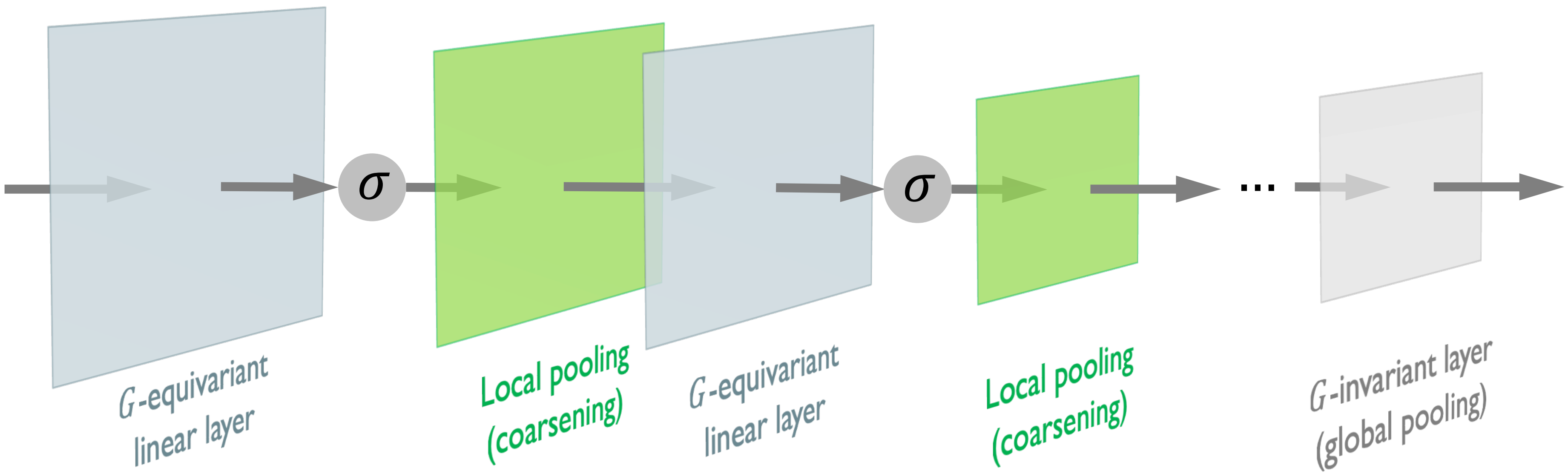
- **Element-wise nonlinear** function $\sigma: \mathcal{X} \to \mathcal{X}$ defined as $(\sigma x)(u) = \sigma\big(x(u)\big)$
- Allows to make nonlinear equivariants out of linear ones by composition: if $f: \mathcal{X} \to \mathcal{X}$ is linear G-equivariant, then the composition $\sigma \circ f$ is nonlinear G- equivariant

Exercise: prove

Geometric Deep Learning Building Blocks

- **Linear equivariant:** $B: \mathcal{X}(\Omega) \to \mathcal{X}'(\Omega)$ satisfying $B(gx) = gB(x)$
- **Nonlinearity:** $\sigma: \mathcal{X} \to \mathcal{X}$ applied element-wise, $(\sigma x)(u) = \sigma\big(x(u)\big)$
- **Local pooling (coarsening):** $P: \mathcal{X}(\Omega) \to \mathcal{X}\big(\widetilde{\Omega}\big)$
- **Invariant layer (global pooling):** $A: \mathcal{X} \to \mathcal{Y}$ satisfying $A(gx) = A(x)$

Geometric Deep Learning Blueprint

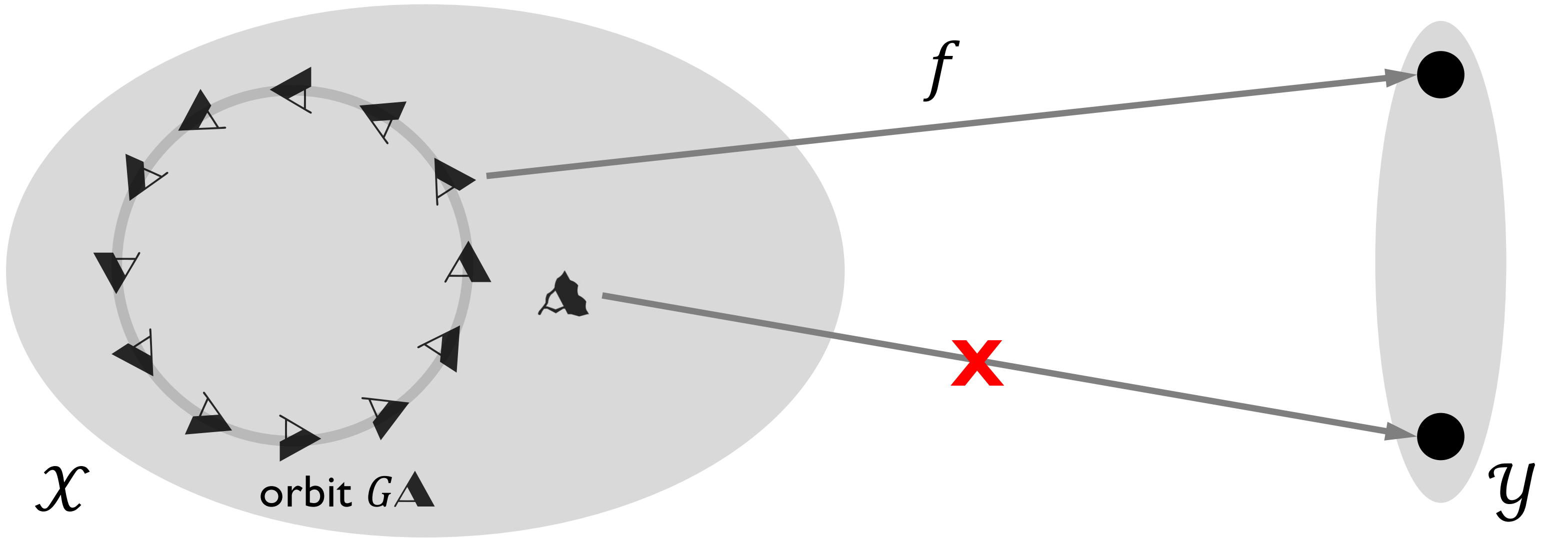


Popular architectures as instances of the Blueprint

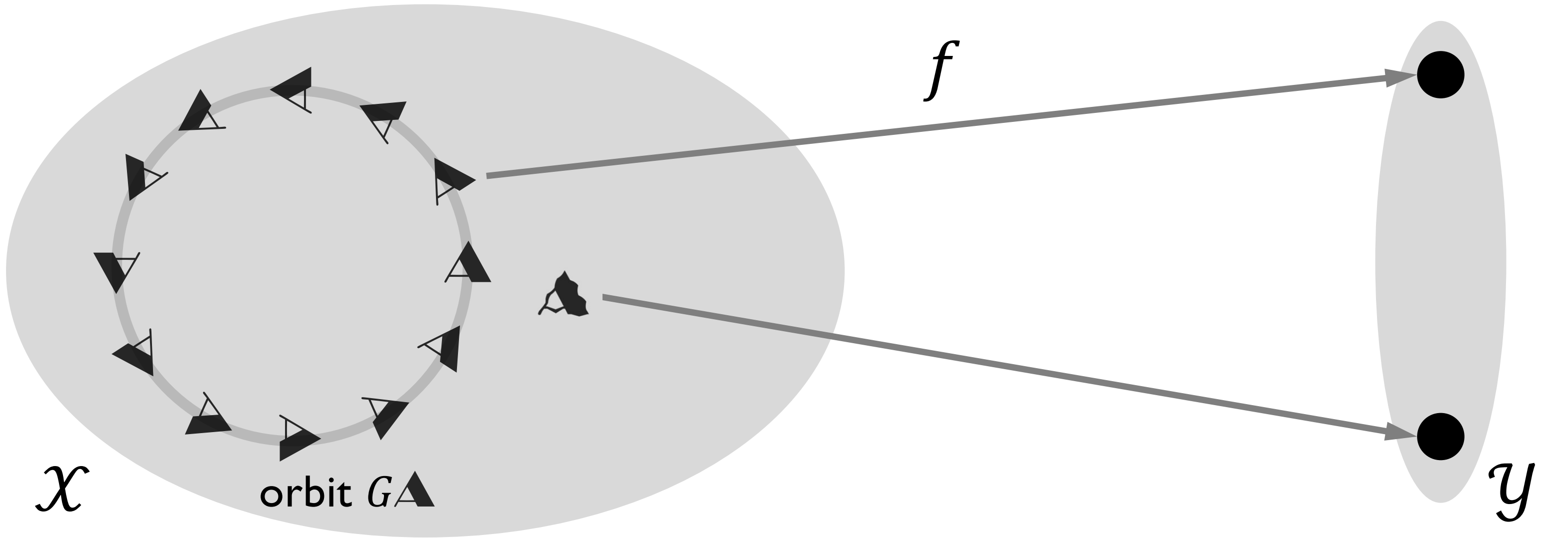
Architecture	**Domain Ω**	**Symmetry Group $\mathfrak{G}$**
CNN	Grid	Translation
Spherical CNN	Sphere / $\mathrm{SO}(3)$	Rotation $\mathrm{SO}(3)$
Intrinsic / Mesh CNN	Manifold / Mesh	Isometry $\mathrm{Iso}(\Omega)$ / Gauge Symmetry $\mathrm{SO}(2)$
GNN	Graph	Permutation S_n
Deep Sets	Set	Permutation S_n
Transformer	Complete Graph	Permutation S_n
LSTM	1D Grid	Time warping

APPROXIMATE INVARIANCE & GEOMETRIC STABILITY

Approximate group invariance

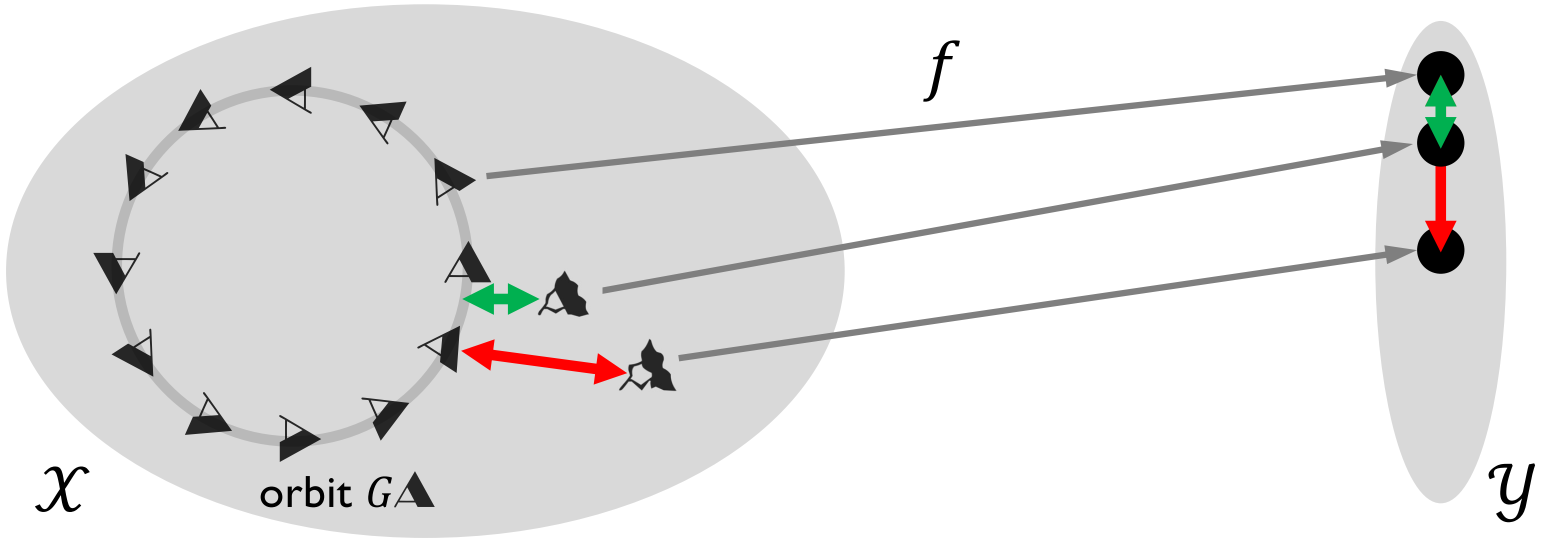


Approximate group invariance



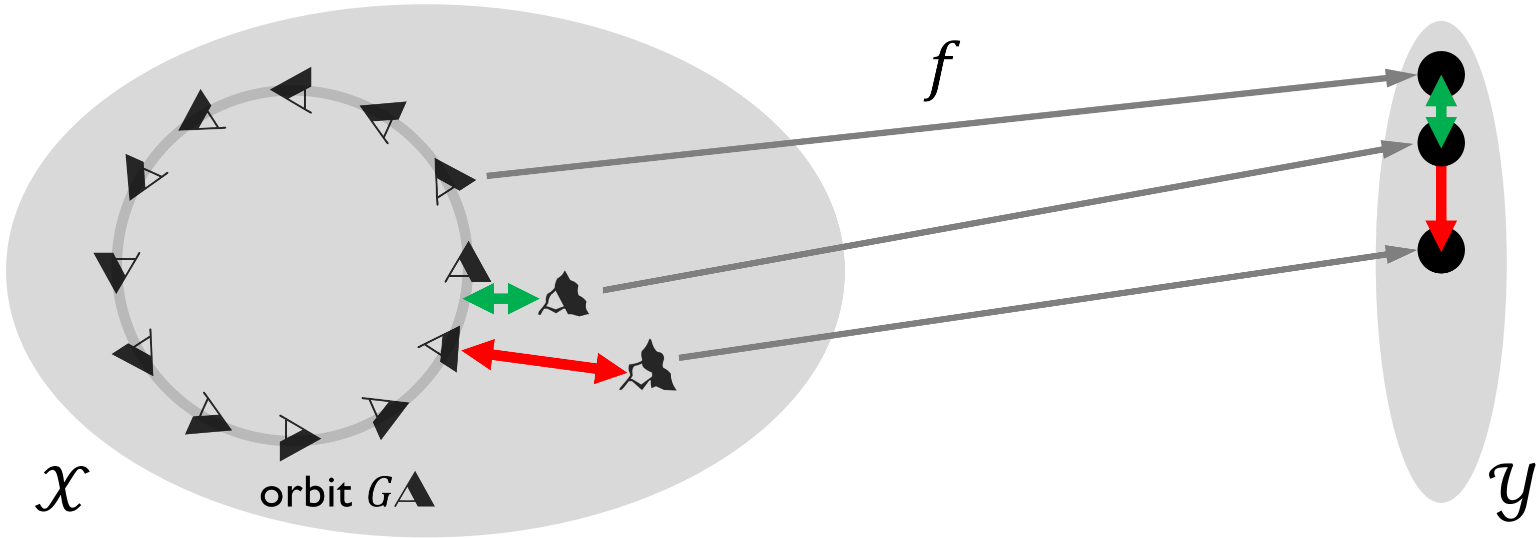
"Approximate invariance to transformations approximately in the group G"

Approximate group invariance



"Approximate invariance to transformations approximately in the group G"

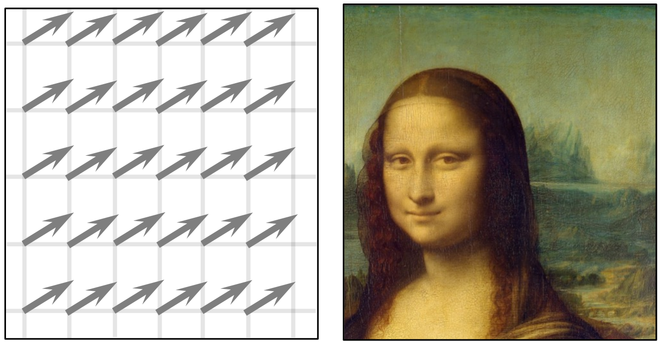
Approximate group invariance



- A function f is said to be **geometrically stable** if for a general deformation $\tau: \Omega \to \Omega$ and some distance d on the space of transformations

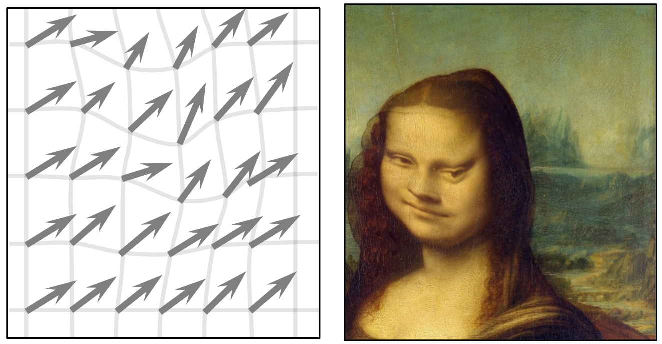
$$\|f(x \circ \tau^{-1}) - f(x)\| \leq d(\tau, G)\|x\|$$

Example: 2D warping



Translation

$$\|\nabla\tau\|^2 = \int_{\mathbb{R}^2} \|\nabla\tau(u)\|^2 \mathrm{d}u = 0$$



Warping

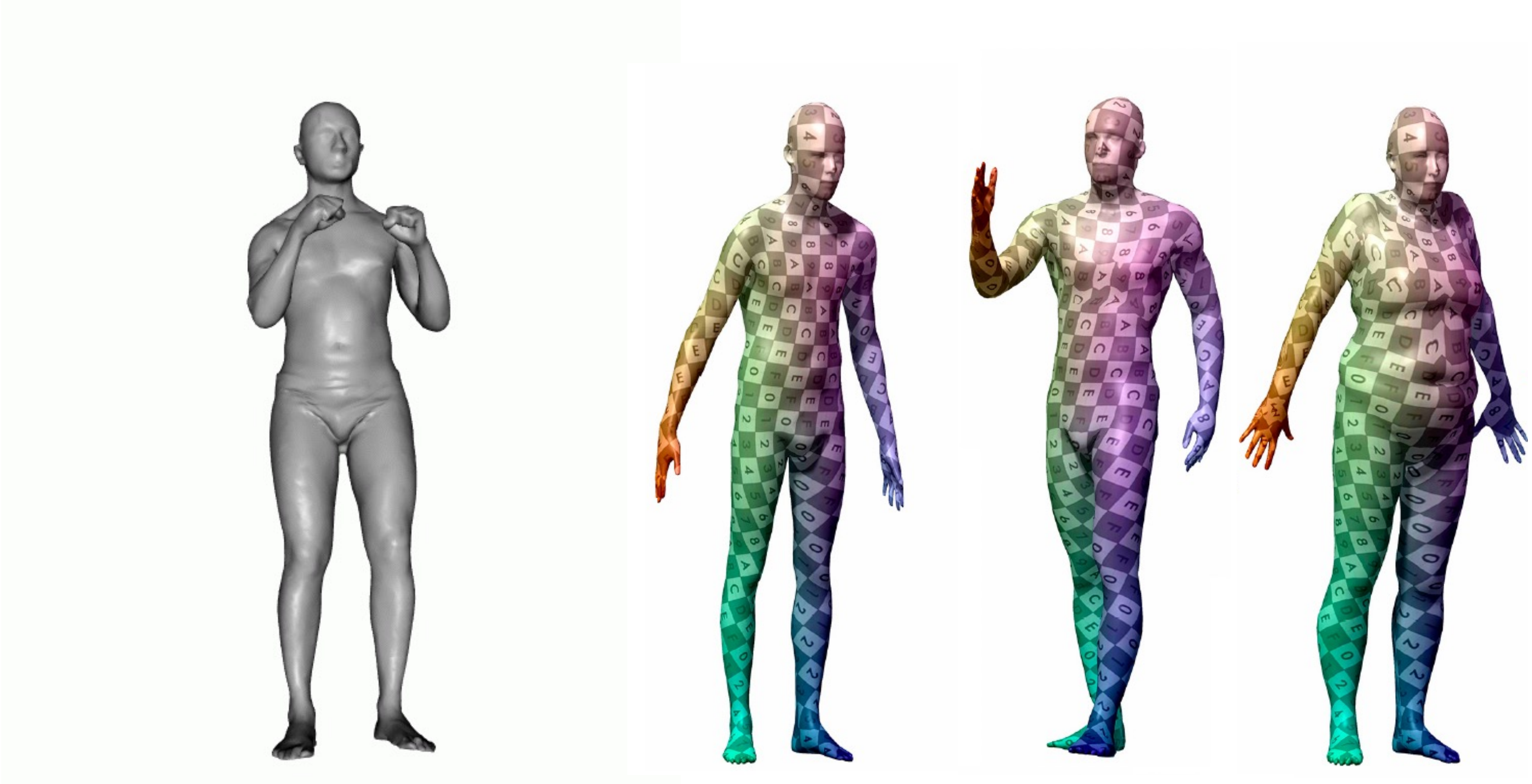
$$\|\nabla\tau\|^2 > 0$$

A **geometrically stable** function obeys a bound of the form

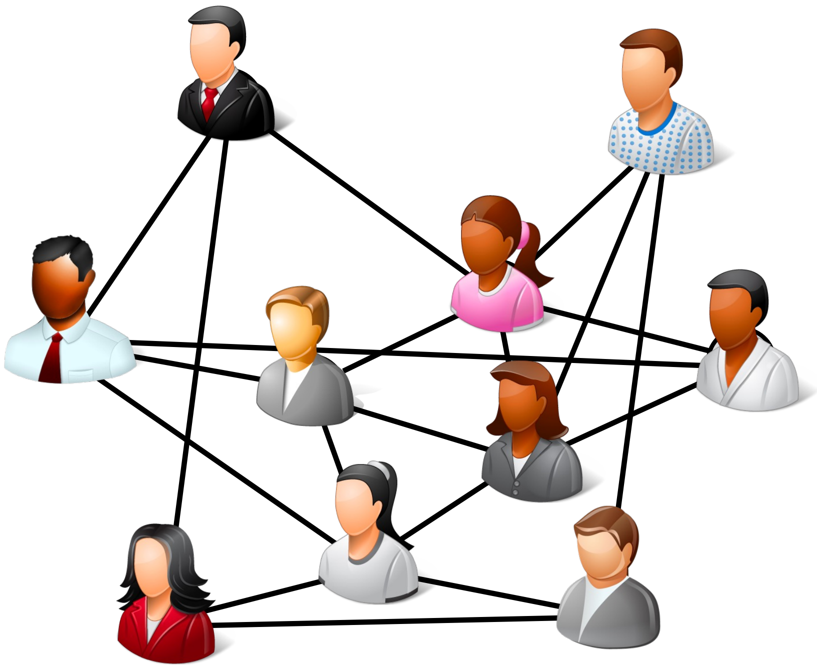
$$\|f(x \circ \tau^{-1}) - f(x)\| \leq \|\nabla\tau\| \, \|x\|$$

Bruna, Mallat 2012

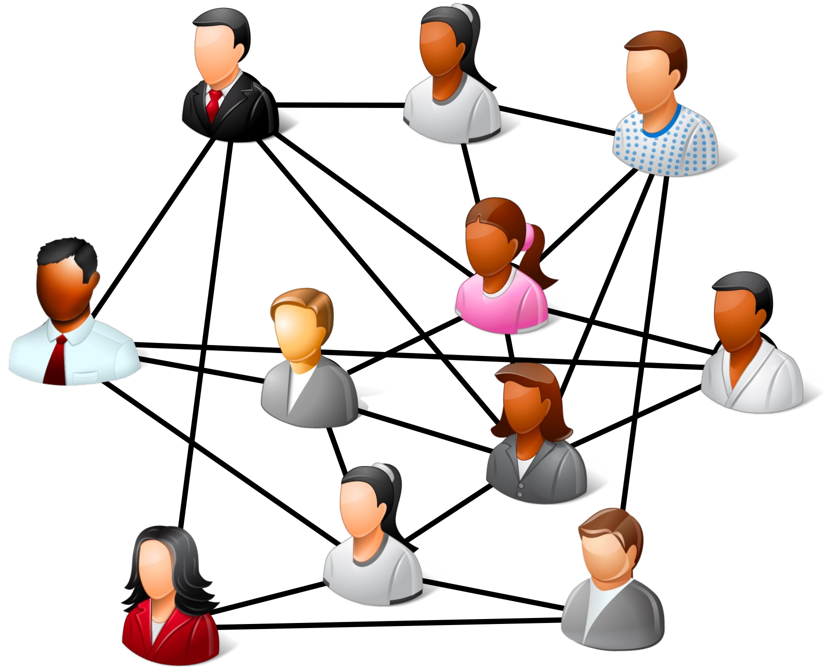
Stability under domain deformation



Stability under domain deformation



Stability under domain deformation



Takeaways

- Invariance reduces the sampling complexity but on its own might not be sufficient to tame the curse of dimensionality
- Symmetry prior must be combined with *Scale Separation*
- *Linear invariants* are not sufficiently expressive
- Instead, one may use nonlinear equivariants obtained by combining *linear equivariants* with *element-wise nonlinearities*
- Combination of these principles leads to a *novel hypothesis class* that is expressive and able to tame the curse of dimensionality.
- Its implementation in the form of neural networks is what we call the *Geometric Deep Learning blueprint*
- Next lectures: examples of instances of the Geometric Deep Learning blueprint on different domains / symmetry groups

Key Concepts

- Scale separation and multiresolution analysis
- Linear equivariants and invariants
- Geometric Deep Learning blueprint

Main References

- M. Bronstein et al., Geometric deep learning, *arXiv*:2104.13478, 2021. Section 3 "Geometric priors"
- N. Carter, *Visual group theory*, 2009. Textbook introducing main concepts of group theory
- C. Esteves, Theoretical aspects of group equivariant neural networks, *arXiv*:2004.05154, 2020. Group representations, harmonic analysis, equivariant networks