Machine Learning with Group Theory

Kenji Fukumizu
The Institute of Statistical Mathematics

Machine Learning Summer School 2024
March 11, 2024 @ OIST

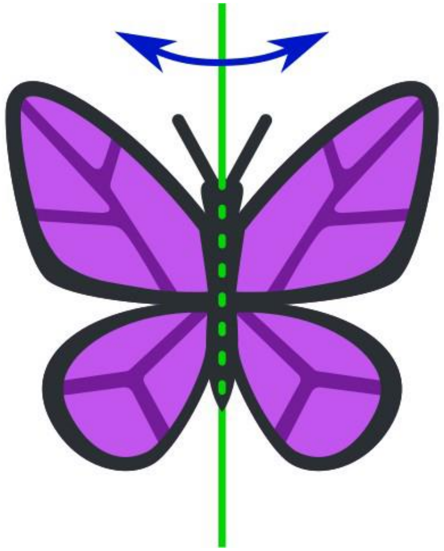
Symmetry is everywhere

In Design

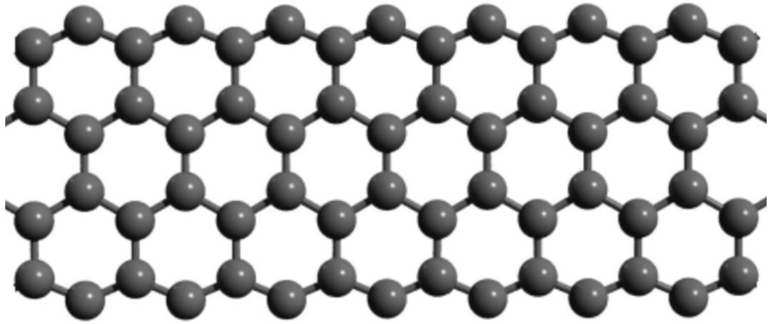
Aoi, Mon (Shogun family)

Cloisters, University of Glasgow

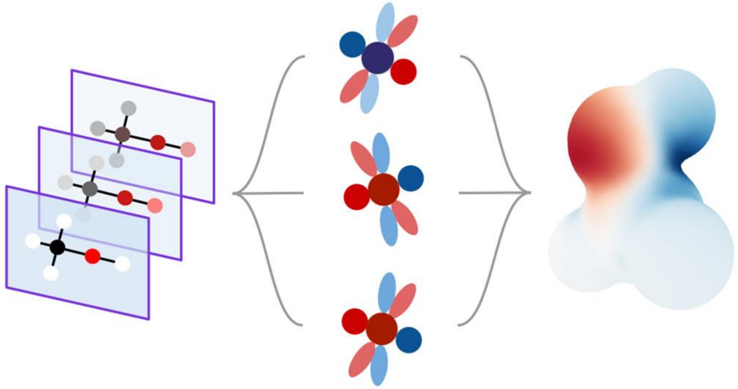
In Nature



Reflection



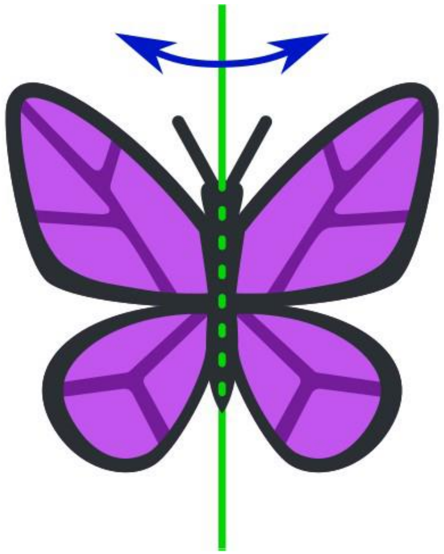
Crystal lattice



[Thürlemann et al. *J. Chem. Theory Comput.* 2022]

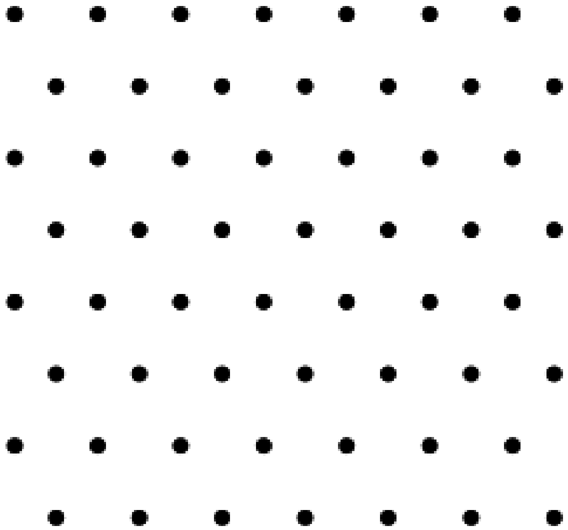
Atomic configuration/static potential

Transformations for symmetry



Reflection/flip

Rotation



Shift

Symmetry and Geometry

- Symmetry or the associated transforms is mathematically formulated as a “group” or “symmetry group” as a part of algebra.

- Core of geometry
  - Geometry
    Euclidean geometry
    Non-Euclidean geometry
    Projective geometry

➡

1872 Klein’s Erlangen program

A unified characterization of geometries based on group transformations.

→ Strong impact on geometry as well as researches of mathematics

Symmetry and Machine Learning

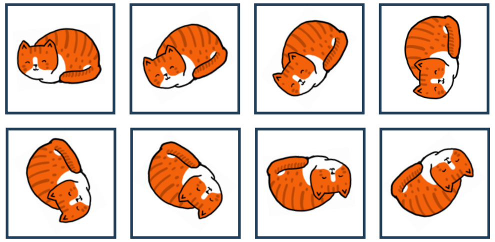
- Symmetry in the geometry of data

“cat”

Shifts and rotations

Invariant classification

- ML methods should make use of symmetry/groups



Approaches
  - Data augmentation: training with transformed data
  - Embed the symmetry in the architecture → Convolutional Neural Networks

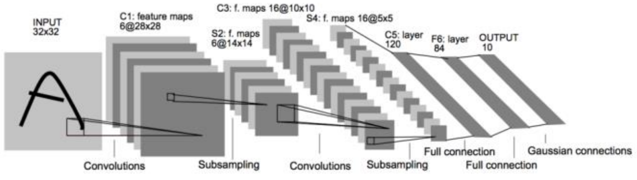


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

LeCun 1998

Aims of this lecture

- Selected topics on ML methods to handle the symmetry in data through group theory.
(Not a comprehensive review of the existing researches.)

  - Basics of group theory as the foundation.

  - NN architecture to use symmetry: Group equivariant convolutional neural networks.

  - Representation learning through group theory.

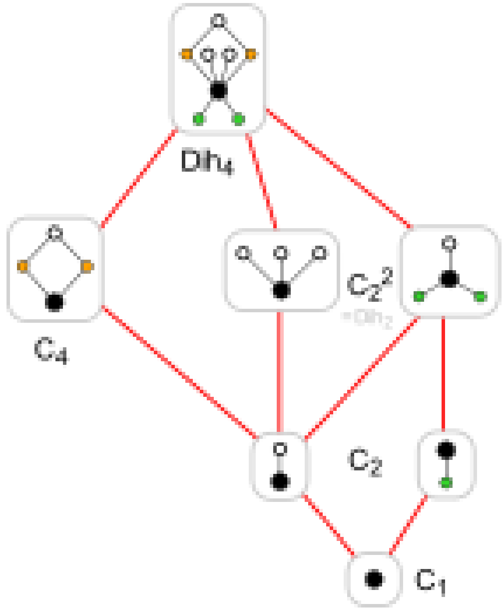
Outline of This Lecture

1. Introduction: symmetry and machine learning ✓
2. Group theory I: Basics
3. Equivariant architecture: Group convolutional NN
4. Group theory II: Group representation and Fourier Transform
5. Representation Learning through Group Action

Outline of This Lecture

1. Introduction: symmetry and machine learning
2. Group theory I: Basics
3. Equivariant architecture: Group convolutional NN
4. Group theory II: Group representation and Fourier Transform
5. Representation Learning through Group Action

Group



- Symmetry/Symmetry transformations are formulated by "group" mathematically.

- <u>Def.</u> Group

  A non-empty set G with is a binary operation (denoted by $\cdot$) is a <u>group</u> if the following three properties hold:

  (1) (Associativity) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for any $a, b, c \in G$.

  (2) (Identity) There is $e \in G$ such that $a \cdot e = e \cdot a = a$ for any $a \in G$.

  (3) (Inverse) For any $a \in G$, there is $b \in G$ such that $a \cdot b = b \cdot a = e$.

  Such b is denoted by a^{-1}.

  * ' $\cdot$ ' is often omitted, and ab is used.

- Example 1. The non-zero real numbers $\mathbb{R}^{\times} = \{ a \in \mathbb{R} \mid a \neq 0 \}$ is a group under multiplication. This is a commutative group, i.e. $ab = ba$.

  A commutative group is also called an Abelian group.

- Example 2. The integers with addition $(\mathbb{Z}, +)$, the product $(\mathbb{Z}^n, +)$, Euclidian space $(\mathbb{R}^n, +)$ are all Abelian groups.

$$a + b = b + a.$$

  If the operation is addition (and thus commutative), the group is called an additive group.

  The identity and inverse are denoted by 0 and $-a$, respectively, for additive groups.

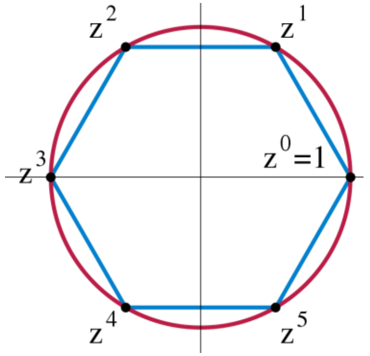
- Example 3. Cyclic group $\mathbb{Z}_n := \mathbb{Z}/n\mathbb{Z}$ (integers mod n with additive operation).

$$\mathbb{Z}_n = \{\bar{0}, \bar{1}, \bar{2}, \bar{3}, \ \dots\ , \overline{n-1}\}$$ (n elements). Additive Abelian group.

$$\bar{1} \in \mathbb{Z}_n. \quad \bar{1}+\bar{1}=\bar{2},\ \bar{2}+\bar{1}=\bar{3},\ \bar{3}+\bar{1}=\bar{4},\ \dots\ , \overline{n-1}+\bar{1}=\bar{0}$$

  - $\mathbb{Z}_n$ is also denoted by C_n, which is often considered as a multiplicative group: $C_n = \{1, \zeta, \zeta^2, \dots, \zeta^{n-1}\}$

$$\zeta := e^{i\frac{2\pi}{n}}$$



- Example 4. The $n \times n$ nonsingular matrices is a group under multiplication. This is called the general linear group and denoted by $GL(n)$.

  - $GL(n)$ is non-commutative, if $n \geq 2$.
    For two matrices A and B, generally $AB \neq BA$.

Some basic definitions on groups

- Subgroup

  Let G be a group. A subset H in G is a subgroup if H is also a group under the binary operation of G. Denoted by $H < G$.

  - Example 1: $m\mathbb{Z}$ (the multiples of m) is a subgroup of $\mathbb{Z}$.
  - Example 2: If $m\ell = n$, $\mathbb{Z}_m < \mathbb{Z}_n$.

    $\mathbb{Z}_6 = \{\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}\}$, $H \coloneqq \{\bar{0}, \bar{3}\} < \mathbb{Z}_6$. $H \cong \mathbb{Z}_2$, $\bar{3} + \bar{3} = \bar{0}$.
  - Example 3: $O(n)$ (orthogonal group) and $SO(n)$ are subgroups of $GL(n)$.

- Direct product

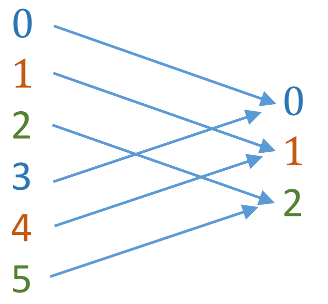
  The direct product $G_1 \times G_2$ of groups G_1 and G_2 is $\{(g_1, g_2) \mid g_1 \in G_1, g_2 \in G_2\}$ with the operation

$$(g_1, g_2) \cdot (h_1, h_2) = (g_1 \cdot h_1, g_2 \cdot h_2).$$

- Homomorphism

  Let G_1 and G_2 be groups. A map $\varphi: G_1 \to G_2$ is a homomorphism if
  $$\varphi(g \cdot h) = \varphi(g) \cdot \varphi(h)$$
  for any $g, h \in G$.

  Example. $\varphi: \mathbb{Z}_6 \to \mathbb{Z}_3$, $m \text{ (mod 6)} \mapsto m \text{ (mod 3)}$



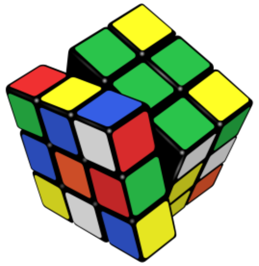
- Isomorphism

  A group homomorphism $\varphi: G_1 \to G_2$ is an isomorphism if it is bijective.

  If there is an isomorphism $\varphi: G_1 \to G_2$, we write $G_1 \cong G_2$. They are essentially the same group.

  Example.
  $$\mathbb{Z}_6 \cong \mathbb{Z}_2 \times \mathbb{Z}_3, \quad n \bmod 6 \mapsto (n \bmod 2, n \bmod 3)$$
  Excercise: Confirm this.

Symmetric group



<u>Symmetric group</u> $\mathfrak{S}_n$ is the group of the permutations on n items.

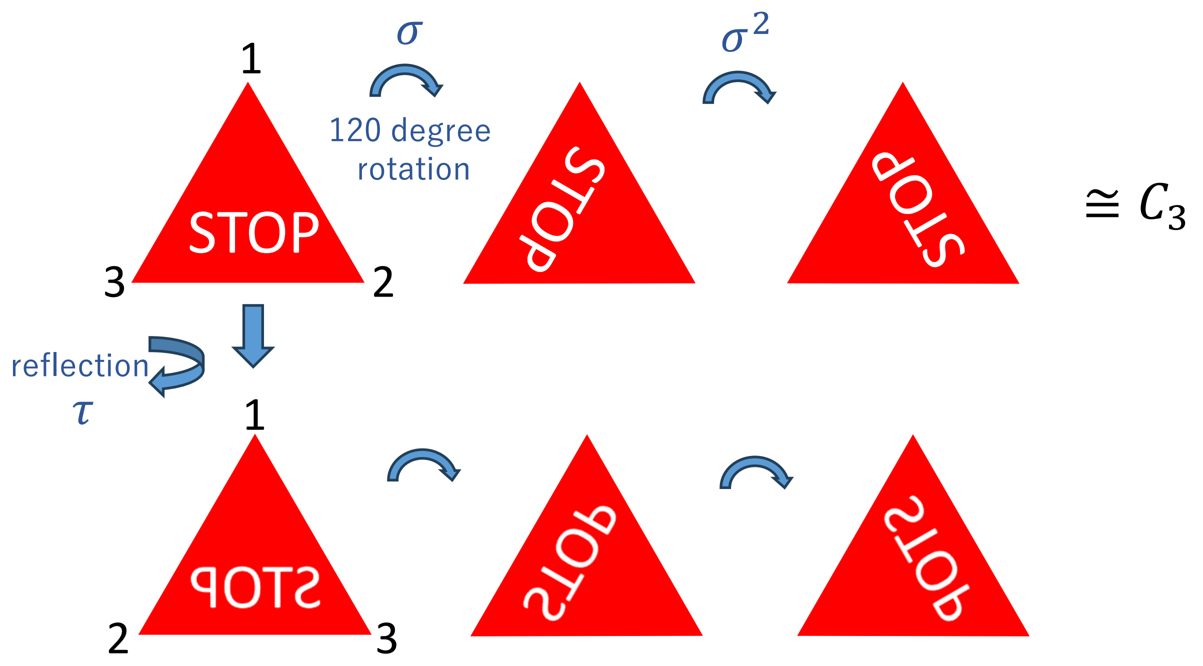
- $\mathfrak{S}_n$ is non-commutative if $n \geq 3$, and $|\mathfrak{S}_n| = n!$

- <u>Example</u>. $\mathfrak{S}_3$

  $\mathfrak{S}_3$ contains C_3 or $\mathbb{Z}_3$ (cyclic group of order 3) as a subgroup.

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad \tau = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$$

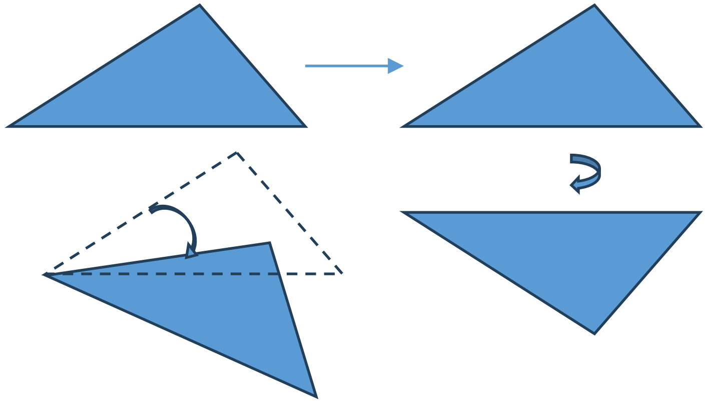
$$\tau\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \quad \sigma\tau = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$



Euclidean motions

Euclidean group $E(n)$: the isometries of a Euclidean space $\mathbb{R}^n$.

- consists of the composition of rotations, translations, and reflections.
- can be written by a pair (R, a), where $R \in O(n)$ (rotation and reflections) and shift $a \in \mathbb{R}^n$.



Special Euclidean group $SE(n)$:

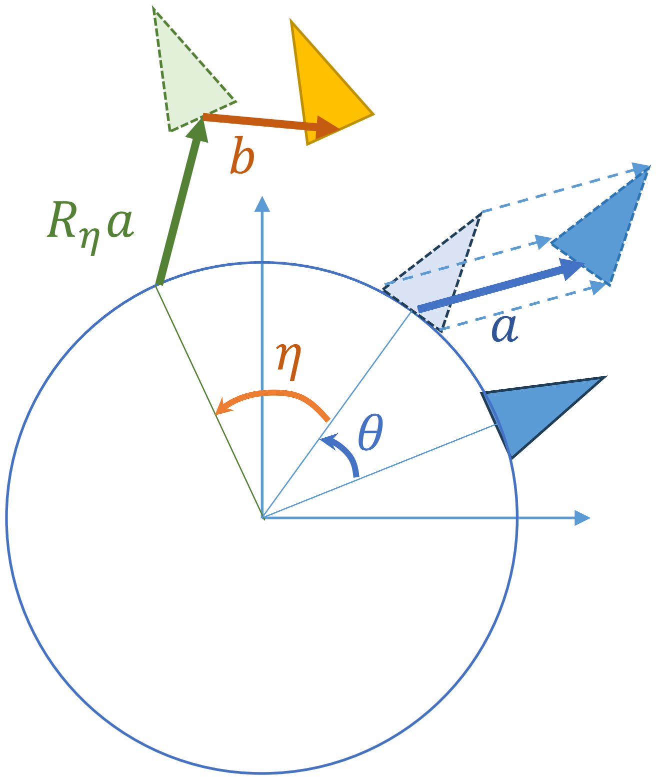
$SE(n)$ consists of the rigid motions in $\mathbb{R}^n$.
An element is given by an arbitrary composition of translations and rotations (but not reflections).

- $SE(2)$

Rotation: $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in SO(2)$

Shift: $a \in \mathbb{R}^2$

$$(R_\eta, b)\cdot(R_\theta, a) = (R_{\eta+\theta}, R_\eta a + b)$$



- $E(n)$ $(n \geq 2)$ is non-commutative.
  (Note: $SO(2)$ is commutative)

- $E(n)$ is realized by $(n+1)\times(n+1)$ matrices

$$\left\{ \begin{pmatrix} A & a \\ 0 & 1 \end{pmatrix} : A \in O(n), b \in \mathbb{R}^n \right\}$$

$$\begin{pmatrix} B & b \\ 0 & 1 \end{pmatrix}\begin{pmatrix} A & a \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} BA & Ba+b \\ 0 & 1 \end{pmatrix}$$

Matrix multiplication is compatible with the group operation.

Normal subgroup

A subgroup N of G is called normal if $ghg^{-1} \in N$ for any $h \in N$ and $g \in G$.
Often denoted by $N \lhd G$.

- Conjugate operation gNg^{-1} does not change N.
- G/N (cosets) has a natural group structure. (See Appendix)
- Any subgroup in a commutative group is normal.

- Example:

  $\mathbb{R}^n \lhd E(n)$ (normal subgroup).

$$\begin{pmatrix} A & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I_n & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A & a \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} A & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I_n & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A^{-1} & -A^{-1}a \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} I_n & Ab \\ 0 & 1 \end{pmatrix}$$

  - $E(n)/\mathbb{R}^n \cong O(n)$
  - $E(n) \neq \mathbb{R}^n \times O(n)$ In the direct product $G_1 \times G_2$, G_1 and G_2 must be commutative.
  - $E(n) \cong \mathbb{R}^n \rtimes O(n)$ (Semidirect product) → Explained in the next slide.

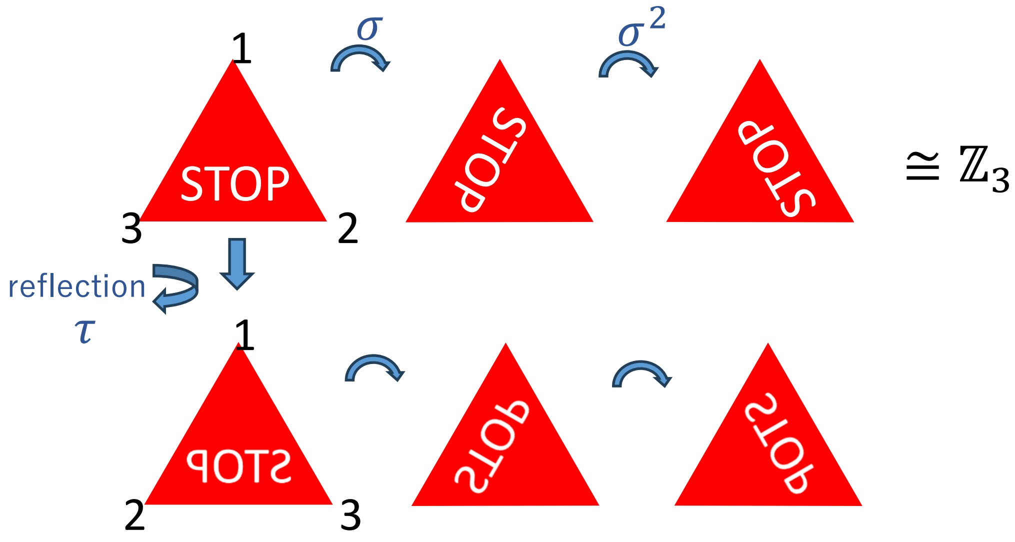
Semidirect Product (Intuition)

- Example 1: $\mathfrak{S}_3 \cong \mathbb{Z}_3 \rtimes \mathbb{Z}_2$

  $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ generates $\mathbb{Z}_3$; $\tau = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ generates $\mathbb{Z}_2$

  $\mathfrak{S}_3 = \{1, \sigma, \sigma^2, \tau, \sigma\tau, \sigma^2\tau\}$ $\quad \sigma^\ell \tau^m \quad (\ell = 0,1,2; m = 0,1)$

  $\tau\sigma\tau^{-1} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} = \sigma^2$ ⟵ Change in $\mathbb{Z}_3$



- Example 2: $SE(2) \cong \mathbb{R}^2 \rtimes SO(2)$

  $SE(2) = \{(a, R_\theta) \mid a \in \mathbb{R}^2, R_\theta \in SO(2)\}$

  $(0, R_\theta)(a, I_2)(0, R_\theta)^{-1} = (R_\theta a, I_2)$ ⟵ Change in $\mathbb{R}^2$

- Semidirect product: $G = N \rtimes K$

  For $N \lhd G$ (normal) and $K < G$ (subgroup), specify $\phi_k\colon H \ni h \mapsto khk^{-1} \in H$.
  Then, we have a unique expression hk or (h, k) $(h \in N, k \in K)$.
  (See Appendix for more rigorous definition.)

A little bit coming back to ML or geometry…
Group action

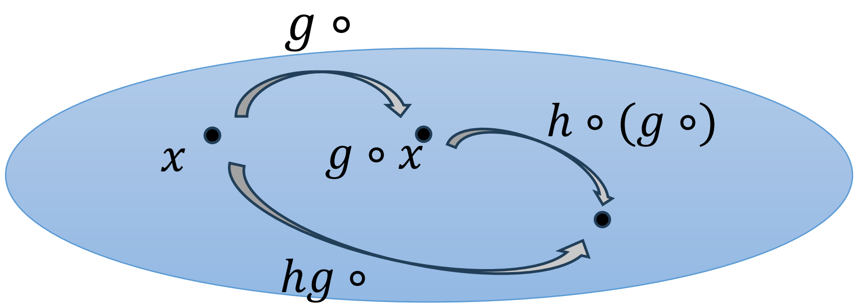
Def.

G: group, X: set. An action of G on X is a mapping $\alpha: G \times X \to X$ such that

i) $\alpha(e, x) = x$

ii) $\alpha(hg, x) = \alpha(h, \alpha(g, x))$

for any $x \in X$ and $g, h \in G$.



- We often use the notation $g \circ x \coloneqq \alpha(g, x)$.

i') $e \circ x = x$

ii') $h \circ (g \circ x) = (hg) \circ x$

- A group representation $\rho: G \to GL(V)$ defines a linear group action: $V = X$ and $g \circ x = \rho(g)x$.

Invariance and Equivariance

Def. Invariance and Equivariance

A group G acts on two sets X and Y. $\varphi: X \to Y$.

- φ is invariant to the group action if

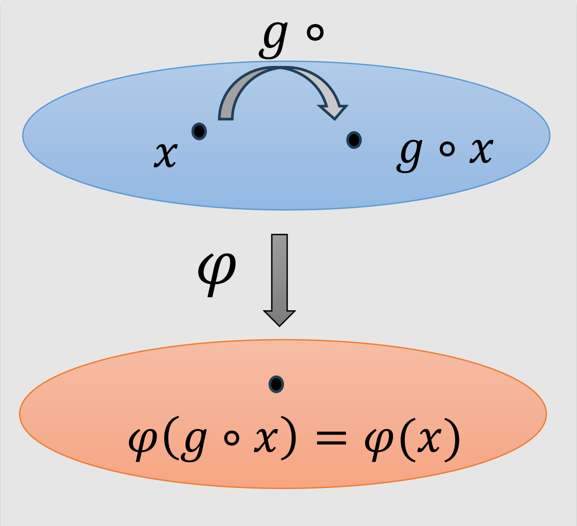
$$\varphi(g \circ x) = \varphi(x)$$

for any $g \in G$ and $x \in X$.

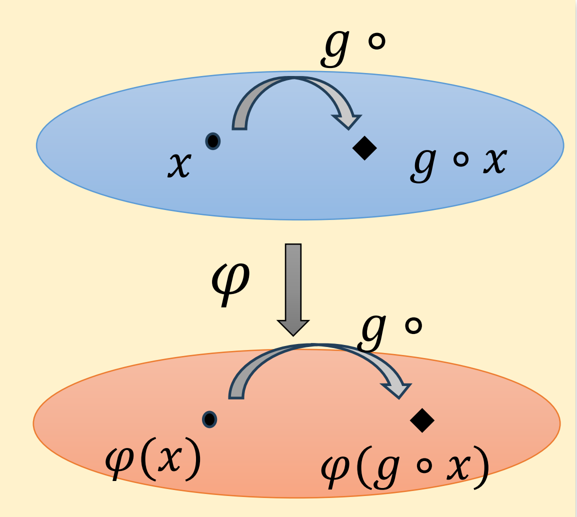
- φ is equivariant to the group action if

$$\varphi(g \circ x) = g \circ \varphi(x)$$

for any $g \in G$ and $x \in X$.



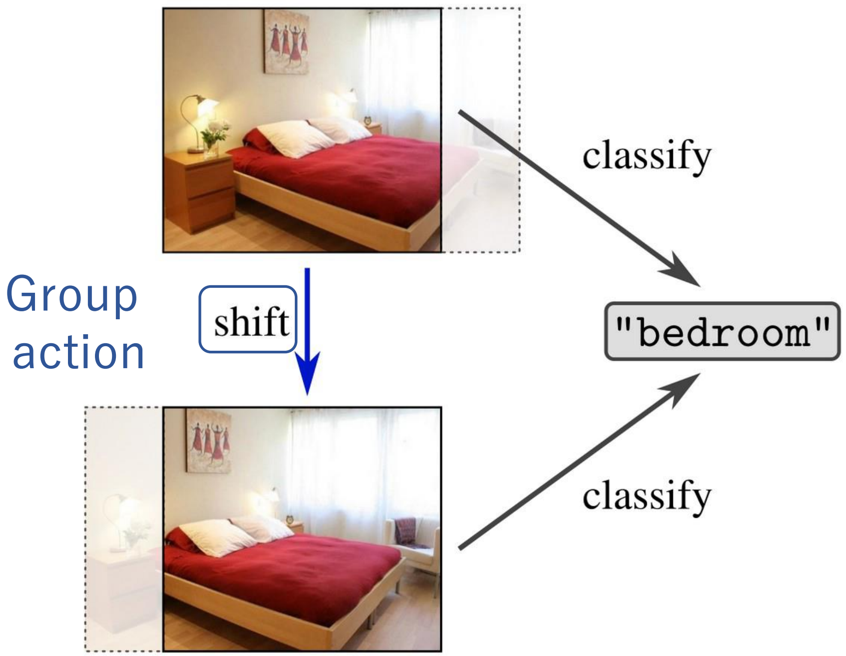
Invariance



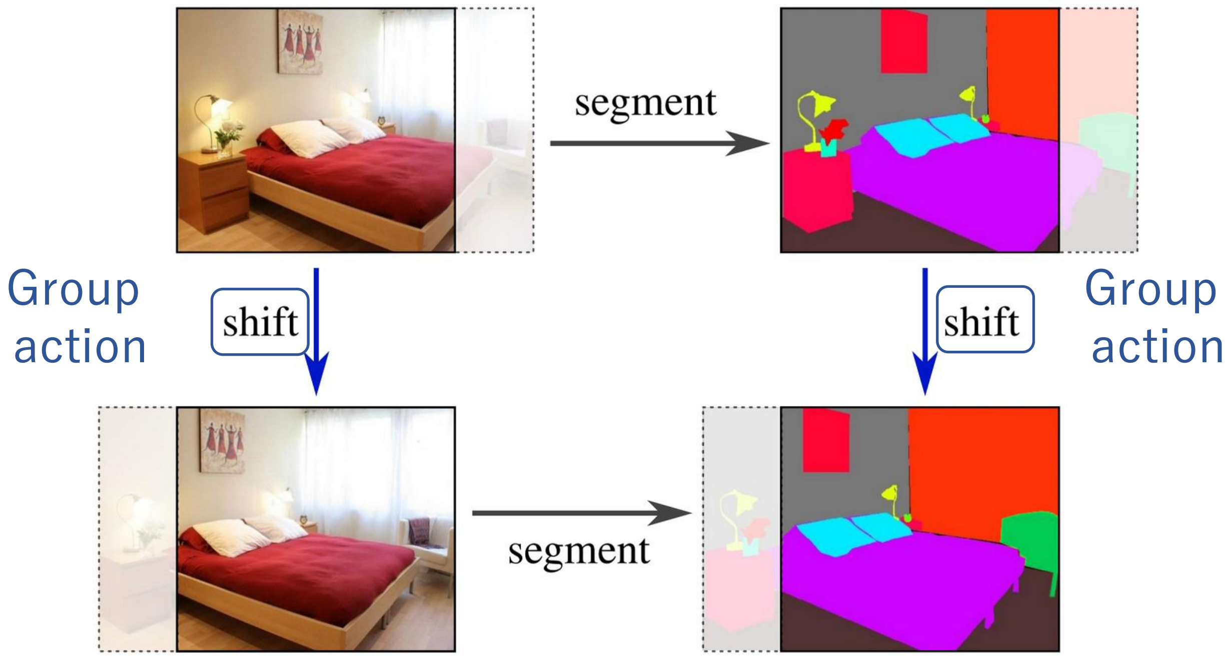
Equivariance

Invariance and equivariance in ML

Invariant
object classification

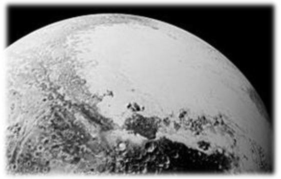


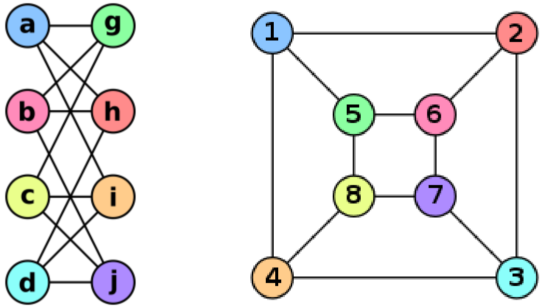
Equivariant
segmentation



From "Groups, Representations & Equivariant maps" by Maurice Weiler (University of Amsterdam)

- What are the advantages?
  - Various data has symmetry and group actions.
    - Image: shifts, SO(2)-rotation, ⋯
    - Spherical data: SO(3)-rotations
    - Graphs: permutation/graph isomorphism
  - Incorporating such group actions should be useful for the compact representation:
    - Data: Low dimensional expression
    - Model: Smaller models, efficient learning





- Approaches to invariance/equivariance ML
  - Data augmentation:
    For invariant/equivariant learning, transform the training data with the known group actions.
    - Easily extendable to non-group cases.
    - Needs many training data.
  - Architecture:
    Equivariant Neural Networks （CNN, G-CNN, etc）
    Impose the symmetry in the architecture of the networks.
  - Representation learning:
    Learn the symmetry in the latent representation automatically
    - Group representation/Fourier transform

Invariance vs Equivariance

- Equivariance is usually more focused.
  - Invariance can be added only at the end.

    If Ψ_ℓ's are all equivariant, adding an invariant layer Φ in final layer

    $$\Phi \circ \Psi_L \circ \cdots \circ \Psi_1(x)$$

    makes an invariant mapping.

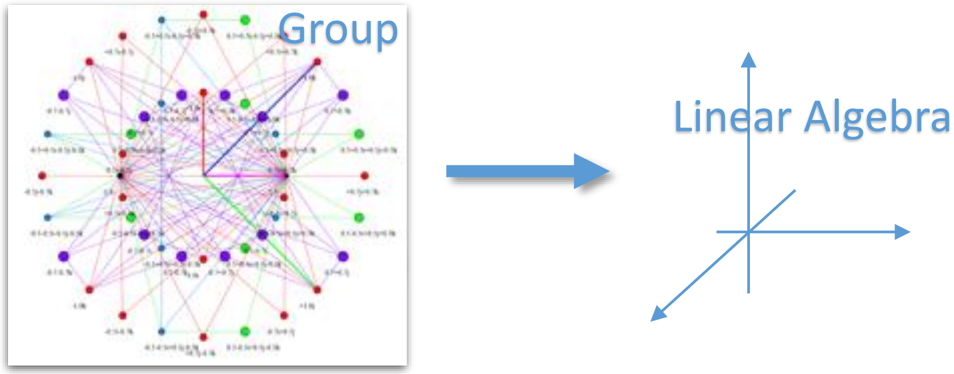
    $$\because) \ \Phi \circ \Psi(gx) = \Phi\big(g\Psi(x)\big) = \Phi \circ \Psi(x)$$

  - Invariance is a special type of equivariance.

    $\Phi: X \to Y$ $\quad G$ acts on X and Y, but action on Y is trivial: $g \circ y = y$ $g \in G$ and $y \in Y$.

    Then, equivariance means invariance: $\Phi(g \circ x) = \Phi(x)$.

Group representation
Linear group action



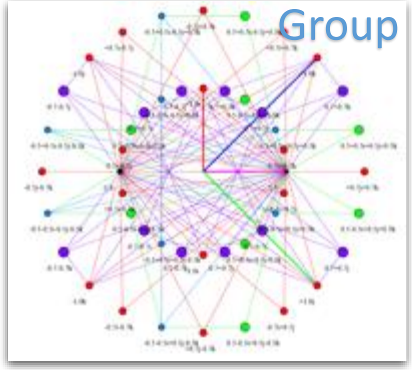
- Group representation: overview
  - The group representation is a mathematical tool to describe a group in terms of linear transformations of a vector space.
  - It reduces various group-theoretic problems to linear algebra/matrix theory.

  - Typical cases
    - Finite groups: Uses finite dimensional linear algebra, so developed most.
    - Abelian groups: This corresponds to the Fourier analysis.
    - Compact group: Extension of Fourier analysis is possible.
    - Lie groups: Used often in physics and chemistry (not covered in this lecture).

- Representation
  - <u>Def</u>. G: group. V: vector space. $GL(V)$: invertible linear transformations of V.

    $\rho: G \to GL(V)$ is a <u>representation</u> of G on V if ρ is a group homomorphism, i.e., $\rho(ab) = \rho(a)\rho(b)$ for any $a, b \in G$.

    - We often write (ρ, V) to specify a representation.
    - When dim$V$ is finite,
      V can be $\mathbb{R}^n$ and $GL(V)$ is $GL(n; \mathbb{R})$.
    - dim$V$ is called <u>dimensionality</u> or <u>degree</u>
      of the representation.
    - V may be a vector space over $\mathbb{C}$.
      It may cause a simpler expression.



$a \cdot b$

ρ ↓

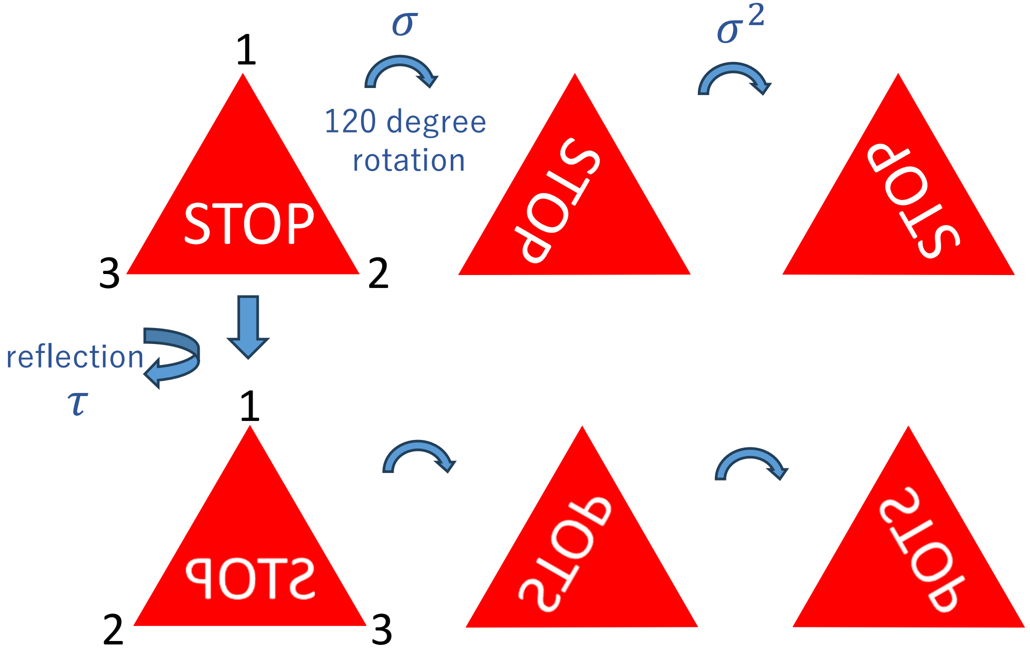
$\big(\ \rho(a)\ \big)\big(\ \rho(b)\ \big)$

Matrices

- Example: $\mathfrak{S}_3$

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = (123) \quad \text{cycle of order 3}$$

$$\tau = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = (23) \quad \text{reflection of order 2}$$



  - σ and τ generate $\mathfrak{S}_3$

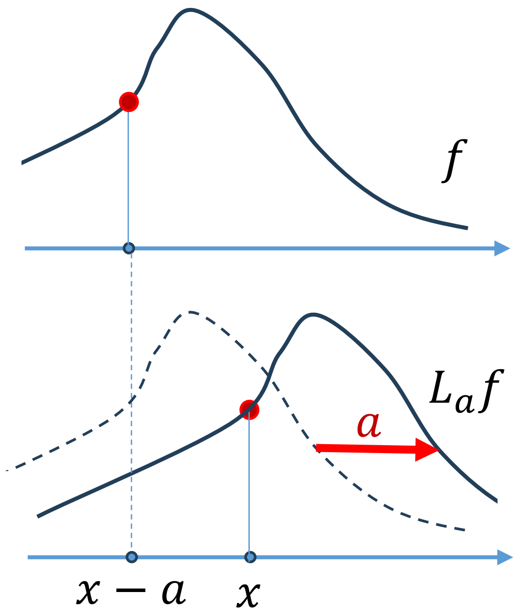
    $\mathfrak{S}_3 = \{e, \sigma, \sigma^2, \tau, \tau, \sigma^2\tau\}$

  - A representation ρ of dim 2 is defined by

$$\begin{cases} \rho(\sigma) = \begin{pmatrix} \cos(-2/3\pi) & -\sin(-2/3\pi) \\ \sin(-2/3\pi) & \cos(-2/3\pi) \end{pmatrix} = \begin{pmatrix} -\sqrt{3}/2 & 1/2 \\ -1/2 & -\sqrt{3}/2 \end{pmatrix} \\ \rho(\tau) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \end{cases}$$

Regular representation

Simply, shift of functions. A building-block of CNN.



Def. G: group. $V \coloneqq \{f: G \to \mathbb{R}\}$ functions on G.

$$L_g: V \to V, \quad f \mapsto f(g^{-1}\,\cdot)$$

L_g is a representation, i.e. linear and $L_{gh} = L_g \circ L_h$.
This group representation is called regular representation.

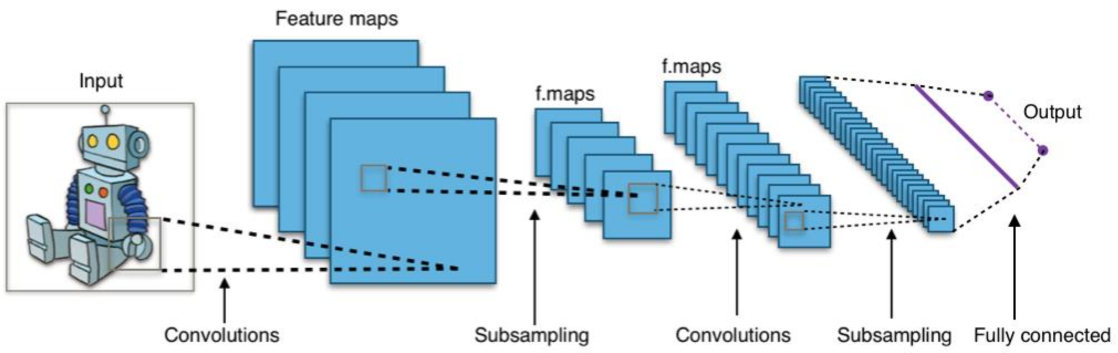
$$\because) \ \left(L_{gh} f\right)(x) = f\left((gh)^{-1}x\right) = f(h^{-1}g^{-1}x) = (L_h f)(g^{-1}x) = L_g(L_h f)(x).$$

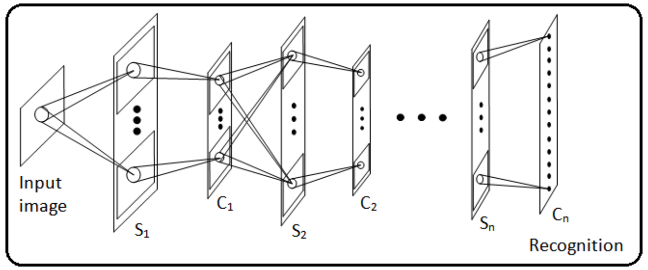
Outline of This Lecture

1. Introduction: symmetry and machine learning
2. Group theory I: Basics
3. Equivariant architecture: Group convolutional NN
4. Group theory II: Group representation and Fourier Transform
5. Representation Learning through Group Action

Convolutional Neural Networks

- CNN (Wei Zhang et al 1988; Yann LeCun 1989)
  - The earliest NN model to realize translation-equivariance.
  - Convolution layer + Pooling layer trained by back-propagation.
  - Inspired by the biological neural networks of early visual cortex.
  - Origin: Neocognitron (Fukushima 1980)
    Backprop was not used.





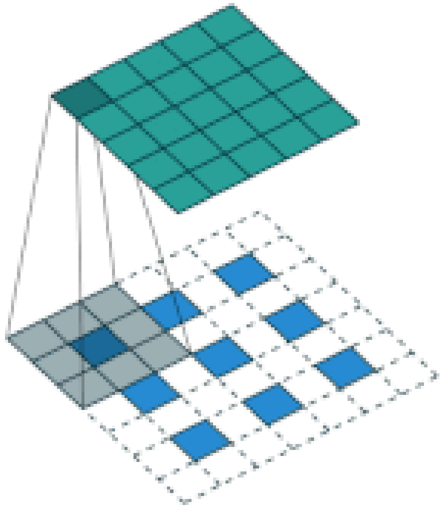
Kunihiko Fukushima 1979

Signal is expressed by a function $f[i,j]$ on the pixels

- Convolutional layer
  - 2D gray-scale images $(W \times H)$ (for simplicity).
    A spatial filter of small size $(3 \times 3$ or $5 \times 5)$ $\psi_{[a,b]}$ is used.

$$h^{Out}_{[i,j]} = \sum_{i-a\in\{0,\pm1\}} \sum_{j-b\in\{0,\pm1\}} \psi_{[i-a,j-b]} f^{In}_{[a,b]}$$

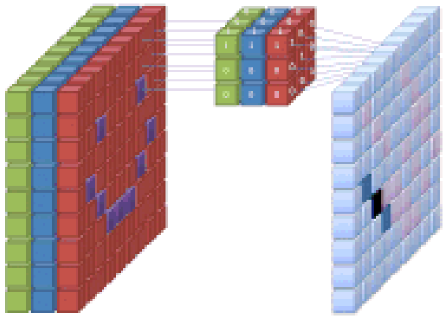
$$f^{Out}_{[i,j]} = \phi\left(h^{Out}_{[i,j]} + \theta\right)$$ ϕ: activation function



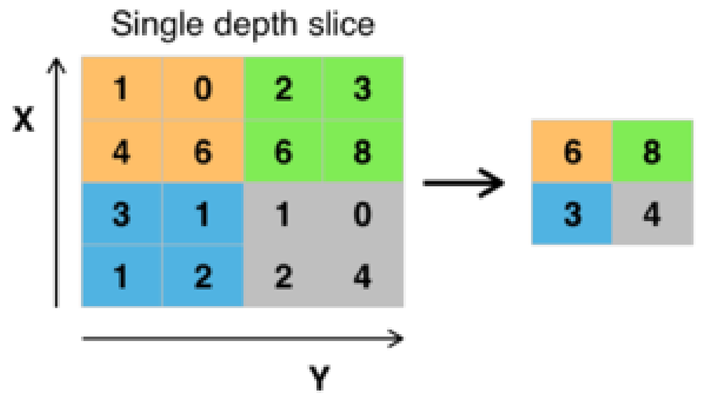
Often, 0 is padded beyond the boundary.
$\sum_{a\in\mathbb{Z}^2}\sum_{b\in\mathbb{Z}^2}$ is okay.

  - 2D Color images, “channel” dimension is added.
    In the first layer, RGB makes 3 channels.
    Each output channel k uses its own filter $\psi^k_{[a,b,c]}$

$$h^{Out}_{[i,j,k]} = \sum_a \sum_b \sum_{c=1}^{m_C} \psi^k_{[i-a,j-b,c]} f^{In}_{[a,b,c]}$$



- Pooling layer (Subsampling/down-sampling)
  - Spatial size is reduced.
  - Take a representative value in a small neighbor.
    - Max pooling is the most popular
    - Average pooling, ℓ_2-norm pooling, etc.
    - Usually done for each channel (#channels unchanged)



- Fully connected layer
  - After several convolutional and pooling layers, a fully connected layer is used in the last layer.
  - Invariance/equivariance can be achieved

Translation Equivariance of CNN

- Convolutional layer
  - Ψf: mapping of f with convolution kernel ψ
  - Shift by $s = [s_W, s_H]$. $(L_s f)_{[i,j]} \coloneqq f(i - s_W, j - s_H)$.

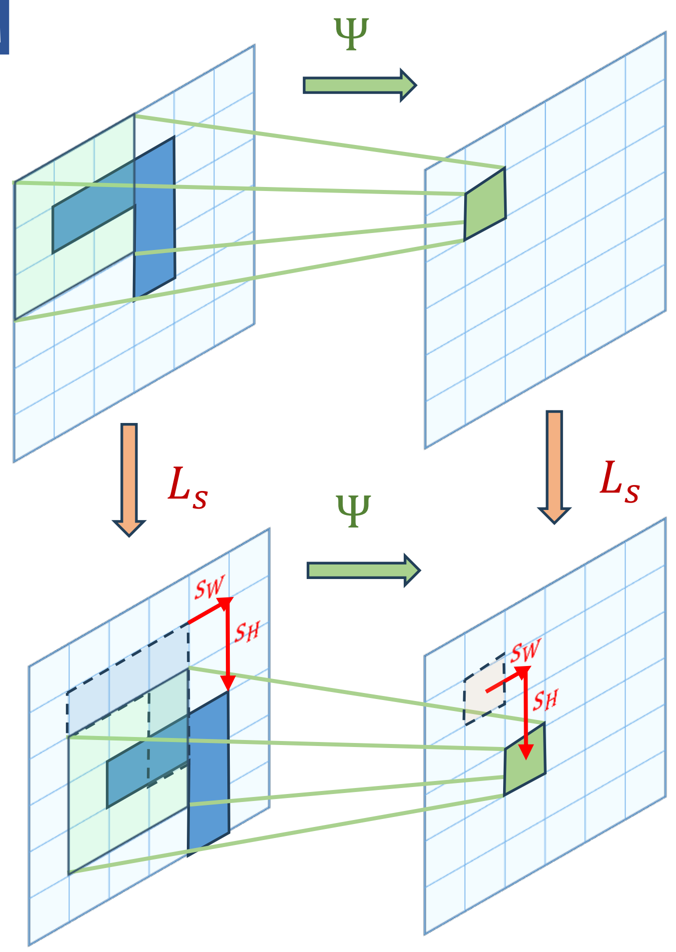
Prop. $L_s(\Psi f) = \Psi(L_s f)$

- Equivariance: Convolution and translation are commutative.

(Proof) Shown for the case of single channel (gray-scale).

$$L_s(\Psi f)[i,j] = (\Psi f)(i - s_W, j - s_H)$$
$$= \sum_a \sum_b \psi_{[i-s_W-a,j-s_W-b]} f_{[a,b]}$$
$$= \sum_{a'} \sum_{b'} \psi_{[i-a',j-b']} f_{[a'-s_W,b'-s_H]}$$
$$= \sum_{a'} \sum_{b'} \psi_{[i-a',j-b']} (L_s f)[a',b'] \quad = \Psi(L_s f)[i,j]$$

$a' \coloneqq a + s_W,\ b' \coloneqq b + s_H$



- Activation

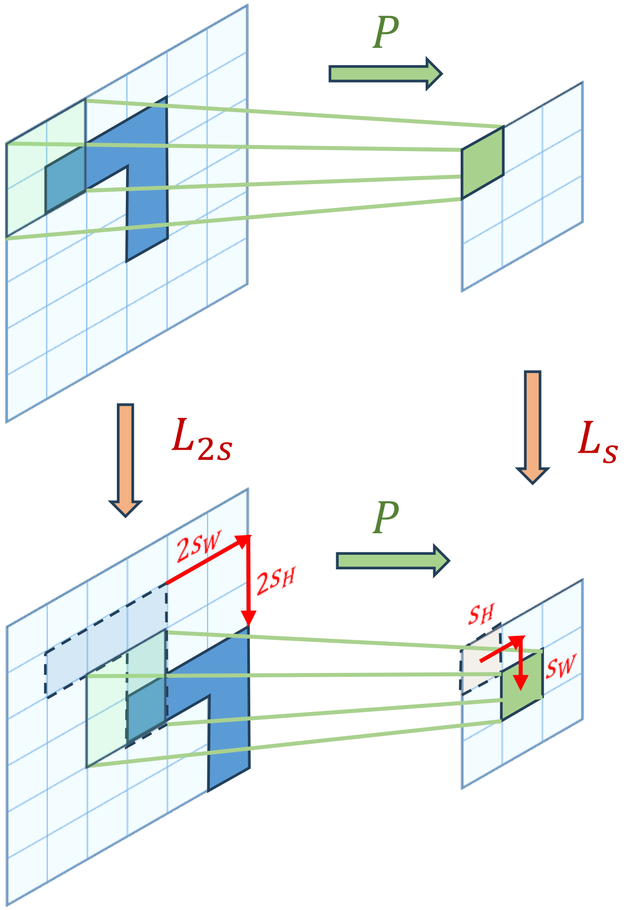
  Prop. $\phi(L_s f) = L_s \phi(f)$. [Equivariant]

  - Applying activation ϕ is just a change of the value, so it is obvious.

- Pooling layer
  - Pf: Pooling of f.
  - After pooling on 2×2 regions, the translation for Pf should be the half for f:

$$P\left(L_s^{(2)} f\right) = L_s(Pf)$$

$L_s^{(2)}$: shift of $2 \times [s_W, s_H]$.



Group equivariant Convolutional Networks
(Cohen & Welling ICML 2016)

- G-CNN: Generalization of CNN to general group action.
  - CNN (recap)
    - An image can be looked at a function
      $$f\colon \mathbb{Z}^2(\text{pixels}) \to \mathbb{R}^3(\text{RGB}).$$
    - Equivariant to the group operation (shift) of $\mathbb{Z}^2$.
      $$\Psi_{\text{Conv}}\big(f(\cdot - s)\big) = (\Psi_{\text{Conv}} f)(\cdot - s).$$
  - G-CNN
    - Considers more general groups for the signal.
      $$f\colon G \to \mathbb{R}^K.$$
    - Equivariant to G.
      $$\Psi_{\text{Conv}}\big(L_g f\big) = L_g(\Psi_{\text{Conv}} f).$$

G-CNN: Motivation

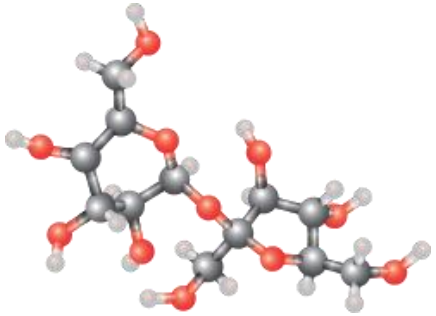
- Symmetries
  - Many image properties are invariant to Euclidean motions (E(2), SE(2)).
    - Medical images: orientation or translation is not relevant.

      c.f., Orientation may be meaningful in natural images (rooms, road, etc), characters (alphabets, numbers, etc).

  - 3D data/360-degree images (SO(3), E(3))
    - 3D scanner, rendering, estimated.
    - 360-degree camera.

      https://github.com/QUVA-Lab/e2cnn?tab=readme-ov-file

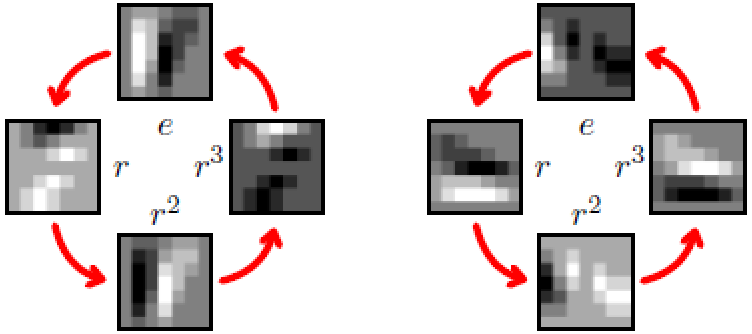
  - Graphs (not covered in this lecture)
    - Permutation invariance
    - Graphs in 3D space, e.g., molecules (SE(3))



G-CNN: Model

- Examples of group G
  - $p4$: 90º rotations and translations $(< SE(2))$

$$\left\{ \begin{pmatrix} \cos\frac{r\pi}{2} & -\sin\frac{r\pi}{2} & s_x \\ \sin\frac{r\pi}{2} & \cos\frac{r\pi}{2} & s_y \\ 0 & 0 & 1 \end{pmatrix} \middle| \; r \in \{0,1,2,3\}; s_x, s_y \in \mathbb{Z} \right\}$$

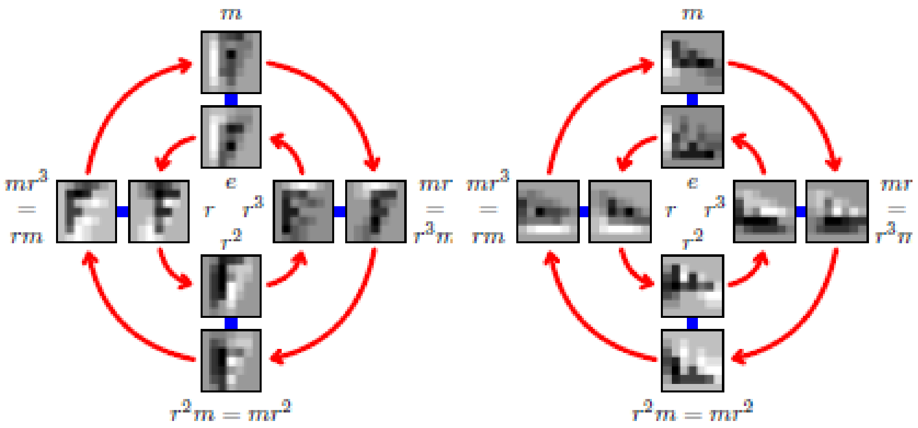


$$p4 \cong \mathbb{Z}^2 \rtimes SO(2,4)$$

$SO(2,N)$: N-discretization of $SO(2)$.

  - $p4m$: p4 + reflection.

$$p4m \cong \mathbb{Z}^2 \rtimes O(2,4)$$



- Convolution layer in G-CNN
  - General ℓ-th layer: $G \to G$ equivariant

$$f \in \mathcal{F}_\ell \coloneqq \left\{ f: G \to \mathbb{R}^{K^{(\ell)}} \right\} \quad \text{signal at the } \ell\text{-th layer}$$

$$\Psi_{\mathrm{Gconv}}: \mathcal{F}_\ell \to \mathcal{F}_{\ell+1} \coloneqq \left\{ \tilde{f}: G \to \mathbb{R}^{K^{(\ell+1)}} \right\}$$

$$(\Psi_{\mathrm{Gconv}} f)_{k'}(g) = \sum_{h \in G} \sum_{k=1}^{K^{(\ell)}} \psi_{k'k}(g^{-1}h) f_k(h) = \sum_{h \in G} \sum_{k=1}^{K^{(\ell)}} \left(L_g \psi_{k'k} \right)(h) f_k(h)$$

$$(k' = 1, \dots, \ell + 1)$$

  - Input layer: $\mathbb{Z}^2 \to G$ equivariant

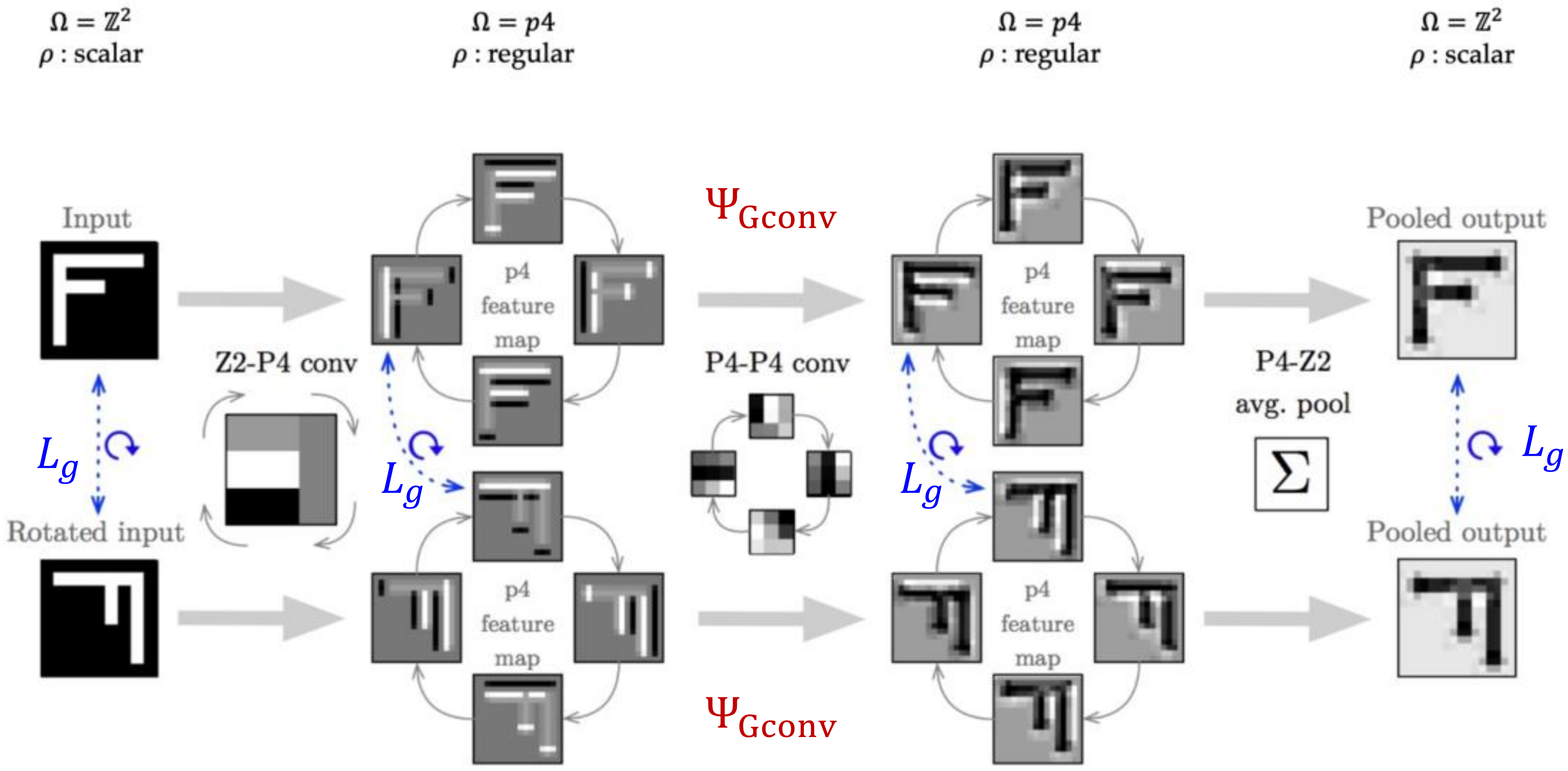
signal $f: \mathbb{Z}^2 \to \mathbb{R}^3$ (Assume $G \curvearrowright \mathbb{Z}^2$), filter $\psi: \mathbb{Z}^2 \to \mathbb{R}^{K' \times K}$

$$(\Psi_{\mathrm{in}} f)_{k'}(g) = \sum_{x \in \mathbb{Z}^2} \sum_{k=1}^{K} \psi_{k'k}(g^{-1} \circ x) f_k(x)$$

  - Final layer : $G \to \mathbb{Z}^2$ equivariant

Output $f: \mathbb{Z}^2 \to \mathbb{R}^K$ (Assume $G \curvearrowright \mathbb{Z}^2$), filter $\psi: G \to \mathbb{R}^{K' \times K}$

$$(\Psi_{\mathrm{fin}} f)_{k'}(x) = \sum_{g \in G} \sum_{k=1}^{K} \psi_{k'k}\left(g'^{-1} g_x\right) f_k(g')$$



Veeling et al. Rotation Equivariant CNNs for Digital Pathology, MICCAI 2018

- Equivariance of G-conv layer

Prop. $L_g(\Psi_{\mathrm{Gconv}} f) = \Psi_{\mathrm{Gconv}}(L_g f)$

Recall $(\Psi f)_{k'}(g) = \sum_{h\in G} \sum_{k=1}^{K} (L_g \psi_{k'k})(h) f_k(h)$
$= \sum_{h\in G} \sum_{k=1}^{K} \psi_{k'k}(g^{-1}h) f_k(h)$

Proof)

Define

$$(f,\varphi) := \sum_{h\in G} f(h)\varphi(h).$$

Then, we have $\Psi_{Gconv}(f)(u) = (L_u\psi, f)$,

and

$$\Psi_{Gconv}(L_g f)(u) = (L_u\psi, L_g f).$$

From Lemma,

$$(L_u\psi, L_g f) = (L_{g^{-1}} L_u \psi, f) = (L_{g^{-1}u}\psi, f)$$
$$= \Psi_{Gconv}(f)(g^{-1}u) = L_g(\Psi_{Gconv}(f))(u).$$

q.e.d.

Lemma

$$(\varphi, L_g f) = (L_{g^{-1}}\varphi, f).$$

$\because$) $(\varphi, L_g f) = \sum_h \varphi(h) f(g^{-1}h)$
$= \sum_{h'} \varphi(gh') f(h')$ $\quad h' := g^{-1}h$
$= \sum'_{h'} (L_{g^{-1}}\varphi)(h') f(h')$
$= (L_{g^{-1}}\varphi, f).$

Applications of G-CNN

- Rotation MNIST

  Random rotated MNIST images.

Network	Test Error (%)
Larochelle et al. (2007)	10.38 ± 0.27
Sohn & Lee (2012)	4.2
Schmidt & Roth (2012)	3.98
Z2CNN	5.03 ± 0.0020
P4CNNRotationPooling	3.21 ± 0.0012
P4CNN	**2.28 ± 0.0004**

Cohen & Welling ICML 2016

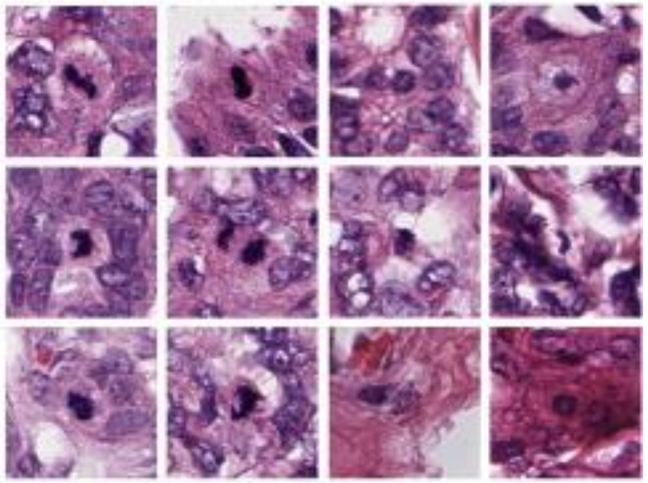
From "Exploring Strategies for Training Deep Neural Networks", Larochelle et al JMLR 2009

- # Medical images

Lafarge et al.: Roto-translation equivariant convolutional networks: Application to histopathology image analysis. *Medical Image Analysis* (2021)

Application of SE(2)-equivariant CNN to medical image analysis.

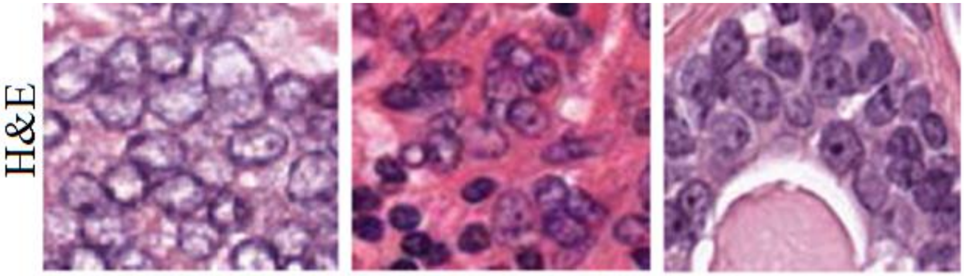
Mitosis detection



AMIDA13. Veta et al, Medical Image Analysis 2015

Eight cases (458 mitotic figures) were used to train the models and four cases (92 mitoses) for validation. Evaluation is performed on a test set of 11 independent cases (533 mitoses), 23 cancer cases

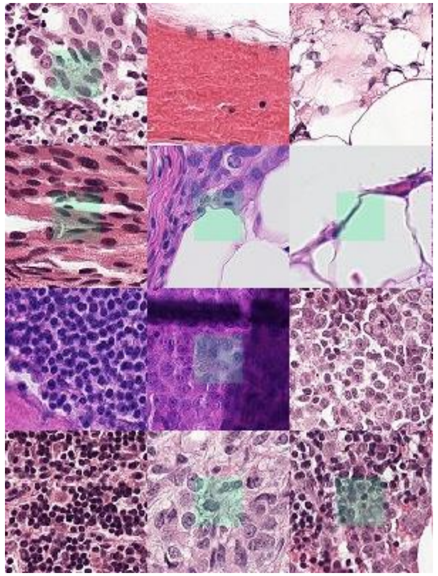
Multi-organ nuclei segmentation



Kumar et al (2017) IEEE Trans MI.

4 ×3 HPF images for training (7337 nuclei), 4 ×1 HPF images for validation (1474 nuclei) and 4 ×2 HPF images for testing (4130 nuclei).

Patch-based tumor detection



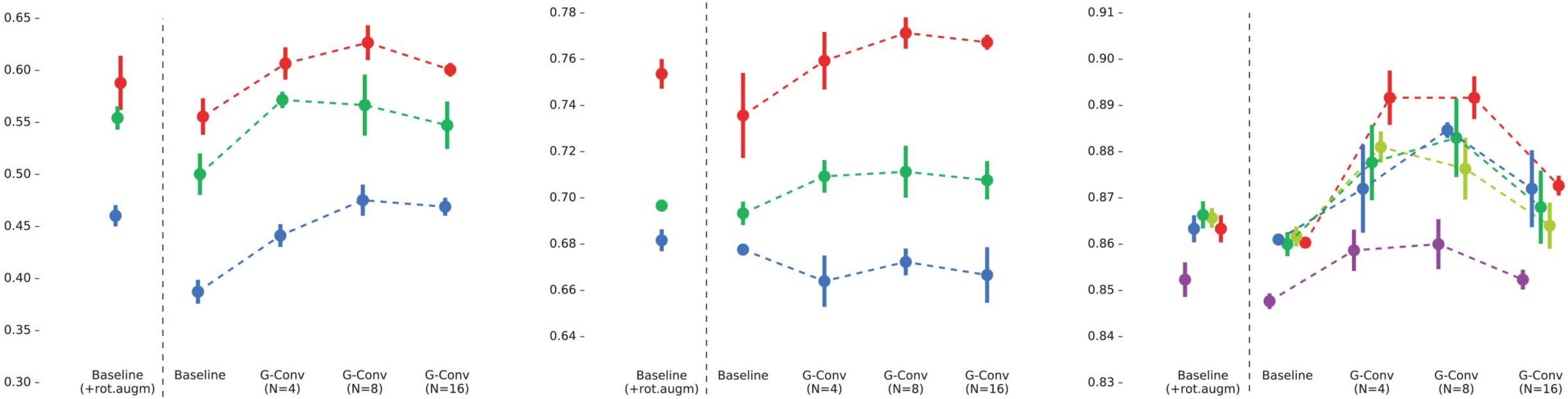
PCam. Veeling et al MICCAI 2018

327,680 image patches, benign/malignant

$SE(2,N) \coloneqq \mathbb{R}^2 \rtimes SO(2,N)$, where $SO(2,N)$ is N discretization of $SO(2)$

$|G| = O(N)$, #weights = $O(N^2)$

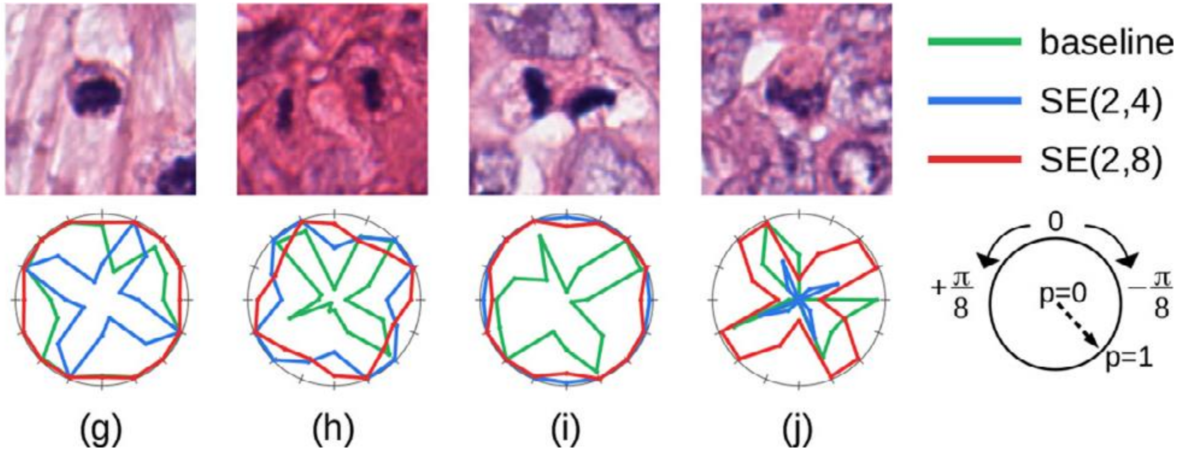
Mitosis detection — Nuclei segmentation — Tumor detection



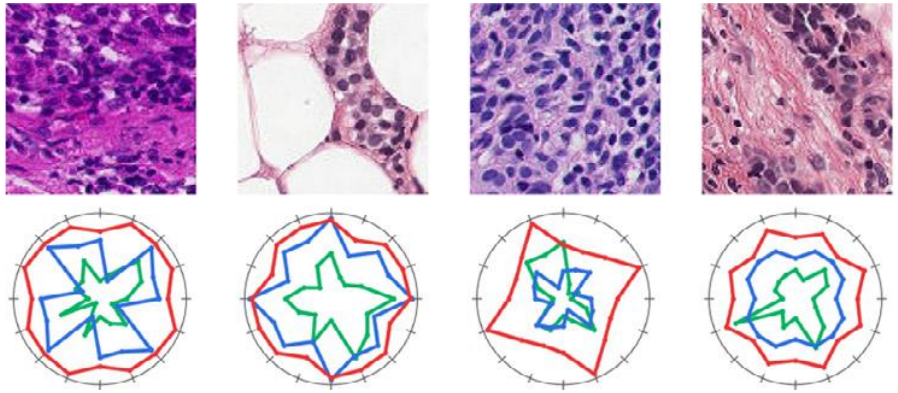
Baseline (+rot.augm) (leftmost): data augmentation by rotations
Baseline: CNN (translation)
Different colors correspond to variations of training data.

- Invariance to rotations
  - Test images (positive) are rotated and the prediction outputs by NN are shown.
  - $SE(2,N)$-CNN achieves much better rotation invariance in the prediction.

Mitosis detection



Tumor detection



Steerable CNN

(Cohen & Welling ICLR 2017; Weiler&Cesa NeurIPS 2019; Weiler et al NeurIPS 2018)

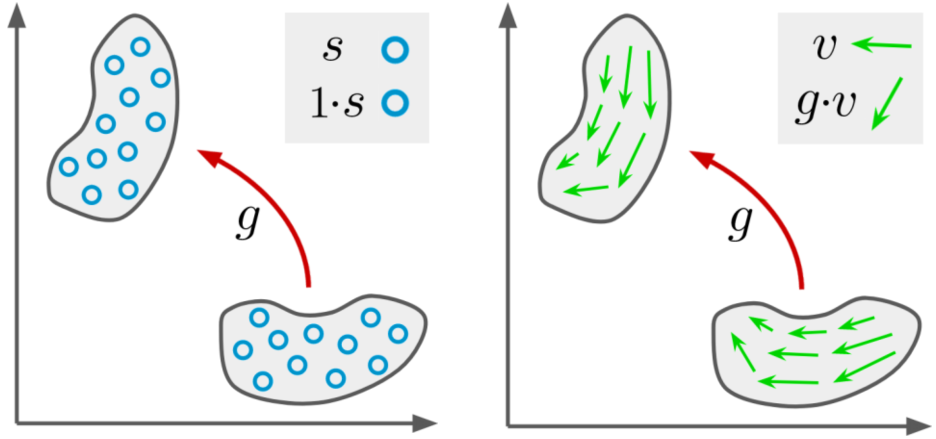
- Further symmetry!
  - In G-CNN, the channels do not consider any symmetry so far.
  - In some cases, symmetry in the channels should be considered.

<u>Example.</u>

Vector field and Euclid motions.

Input signal: $v: \mathbb{R}^2 \to \mathbb{R}^K \quad (K = 2)$

It is natural that when $x \mapsto R(\theta)x$ is applied, the vector field $v(x)$ is also rotated by $R(\theta)$.



scalar field/RGB $\rho(g) = 1$ vector field $\rho(g) = g$

Weiler & Cesa. General E(2)-Equivariant Steerable CNNs, NeurIPS 2021

- Steerable convolution layer

$$E(2) = (\mathbb{R}^2, +) \rtimes O(2), \quad g = (s, R) \in E(2),\ s \in \mathbb{R}^2, R \in O(2),$$

  - *c.f.* Standard G-CNN

$$(\Psi f)_{k'}(g) = \sum_{x\in\mathbb{R}^2} \sum_{k=1}^{K} (L_g \psi_{k'k})(x)(f_k)(x) = \sum_x \sum_k \psi_{k'k}(R^{-1}(x-s)) f_k(x)$$

  - Steerable convolution

$$(\Psi_{\mathrm{Steer}} f)_{k'}(g) = \sum_{x\in\mathbb{R}^2} \sum_{k=1}^{K} \psi_{k'k}(R^{-1}(x-s)) \sum_{c=1}^{K} \rho_{kc}(R) f_c(x)$$

$\rho: O(2) \to GL(\mathbb{R}^K)$ some representation of $O(2)$

    - Condition of ψ for equivariance:

$$\Psi_{\mathrm{Steer}}(L_g f) = L_g(\Psi_{\mathrm{Steer}} f) \quad \text{(Equivariance)}$$

if and only if

$$\psi(gx) = \rho_{out}(g)\psi(x)\rho_{in}(g^{-1}) \qquad \forall g \in E(2), x \in \mathbb{R}^2.$$

- Applications of $SE(2,N)$-Steerable CNN

Weiler, Hamprecht, Storath, "Learning Steerable Filters for Rotation Equivariant CNNs" CVPR 2018

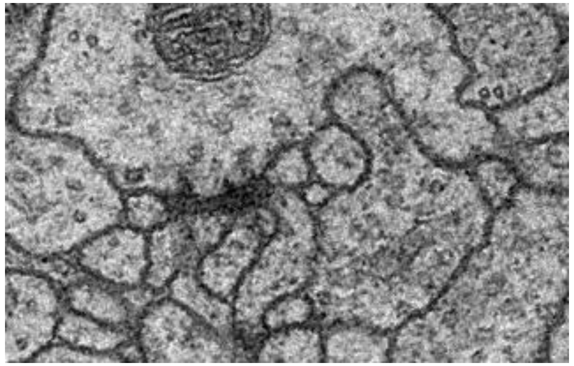
1) rot MNIST

Method	Test Error (%)
Ours – CoeffInit, train time augmentation	$\mathbf{0.714 \pm 0.022}$
Ours – CoeffInit	0.880 ± 0.029
Ours – HeInit	0.957 ± 0.025
Marcos et al. [23] – test time augmentation	1.01
Marcos et al. [23]	1.09
Laptev et al. [21]	1.2
Worrall et al. [24]	1.69
Cohen and Welling [2] - G-CNN	2.28 ± 0.0004
Schmidt and Roth [29]	4.0
Sohn and Lee [11]	4.2
Cohen and Welling [2] - conventional CNN	5.03 ± 0.0020
Larochelle et al. [30]	10.4 ± 0.27

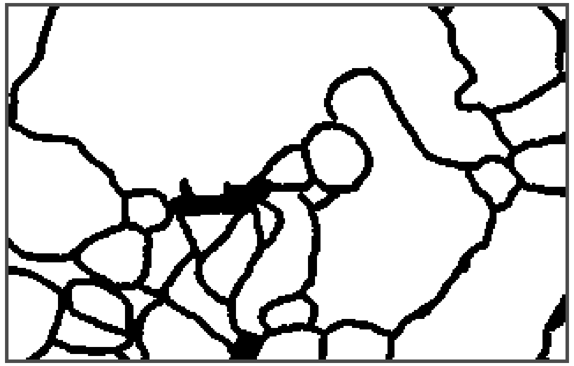
Table 1: Test errors on the rotated MNIST dataset. We distinguish He initialization (HeInit) from the proposed initialization scheme (CoeffInit).

2) ISBI 2012 electron microscopy segmentation challenge

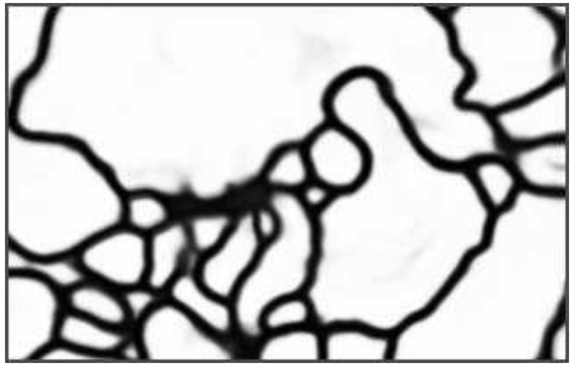
- Prediction of the locations of the cell boundaries in the Drosophila ventral nerve cord from EM images.
- 30 train and test slices of size 512 × 512 px with a binary segmentation ground truth



Raw EM image



Ground truth segmentation



Probability map by the proposed network

Accuracy. Top 6 of more than 100 entries of the leaderboard, as of Nov 13, 2017.

Method	V^{Rand}	V^{Info}
IAL MC/LMC	**0.98792**	**0.99183**
CASIA_MIRA	0.98788	0.99072
Ours	0.98680	0.99144
Quan et al. [26]	0.98365	0.99130
Beier et al. [27]	0.98224	0.98845
Drozdzal et al. [28]	0.98058	0.98816

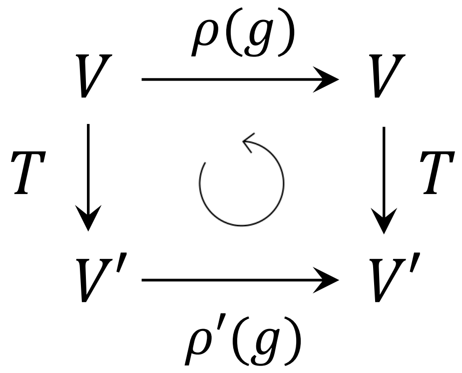
Outline of This Lecture

1. Introduction: symmetry and machine learning
2. Group theory I: Basics
3. Equivariant architecture: Group convolutional NN
4. Group theory II: Group representation and Fourier Transform
5. Representation Learning through Group Action

Mapping of Group Representations

- Def.

  G: group. $(\rho, V), (\rho', V')$: representations of G.
  A linear map $T: V \to V'$ is a G-linear map (or G-map) if the following diagram commutes for any $g \in G$;
  i.e. $\rho'(g)(Tv) = T(\rho(g)v)$ for any $g \in G$.



- Some representations are essentially the same

  Def. A G-map T is an isomorphism if it is invertible.
  Def. Two representations are isomorphic if there is an isomorphism.

Decomposition of representation

- In some cases, a group representation can be decomposed into a direct sum of "irreducible representations".
- In matrices, it corresponds to the simultaneous (w.r.t. g) block-diagonalization of the representation matrix.

- Def. If a subspace $W \subset V$ of a representation (ρ, V) satisfies $\rho(g)W \subset W$ for any $g \in G$, then the restriction $\rho|_W$ defines a representation $(\rho|_W, W)$. This is called a subrepresentation of (ρ, V).

$$\begin{pmatrix} A(g) & B(g) \\ O & D(g) \end{pmatrix} \begin{pmatrix} w \\ 0 \end{pmatrix} = \begin{pmatrix} A(g)w \\ 0 \end{pmatrix}$$

- Def. A representation (ρ, V) is reducible if there is a non-trivial subrepresentation (i.e., there is $W \subset V$ such that $\rho(g)W \subset W$ and $W \neq 0, V$.)
  - If a representation is *not* reducible, then it is called irreducible.

- Def.
  The direct product of representations (ρ_1, V_1) and (ρ_2, V_2) is defined by
  $$\rho_1 \oplus \rho_2 : G \to V_1 \oplus V_2, \qquad g \mapsto \rho_1(g) \oplus \rho_2(g).$$
  $$\begin{pmatrix} \rho_1(g) & O \\ O & \rho_2(g) \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \rho_1(g) v_1 \\ \rho_2(g) v_2 \end{pmatrix}$$

- Maschke's theorem
  Let (ρ, V) be a finite-dimensional representation of a *finite* group G. Let $(\rho|_W, W)$ $(W \subset V)$ be any subrepresentation. Then there exists a subspace $U \subset V$ such that $V = W \oplus U$ and $\rho = \rho|_W \oplus \rho|_U$.

  $$\begin{pmatrix} \rho|_W(g) & B(g) \\ O & D(g) \end{pmatrix} \Rightarrow \begin{pmatrix} \rho|_W(g) & O \\ O & \rho|_U(g) \end{pmatrix}$$

  - For a finite group, if we have a subrepresentation, we can always find a complimentary subrepresentation to give a decomposition as a direct sum.

- A representation is called completely reducible (or semisimple) if it is isomorphic to a direct sum of irreducible representations.

$$\rho \cong \rho_1 \oplus \rho_2 \oplus \cdots \oplus \rho_k$$

- In terms of matrices, we can make it simultaneously block-diagonal so that each block corresponds to an irreducible representation:

$$\rho(g) = P \begin{pmatrix} A_1(g) & O & \cdots & O \\ O & A_2(g) & \ddots & O \\ \vdots & & & \vdots \\ O & & \cdots & A_k(g) \end{pmatrix} P^{-1}$$

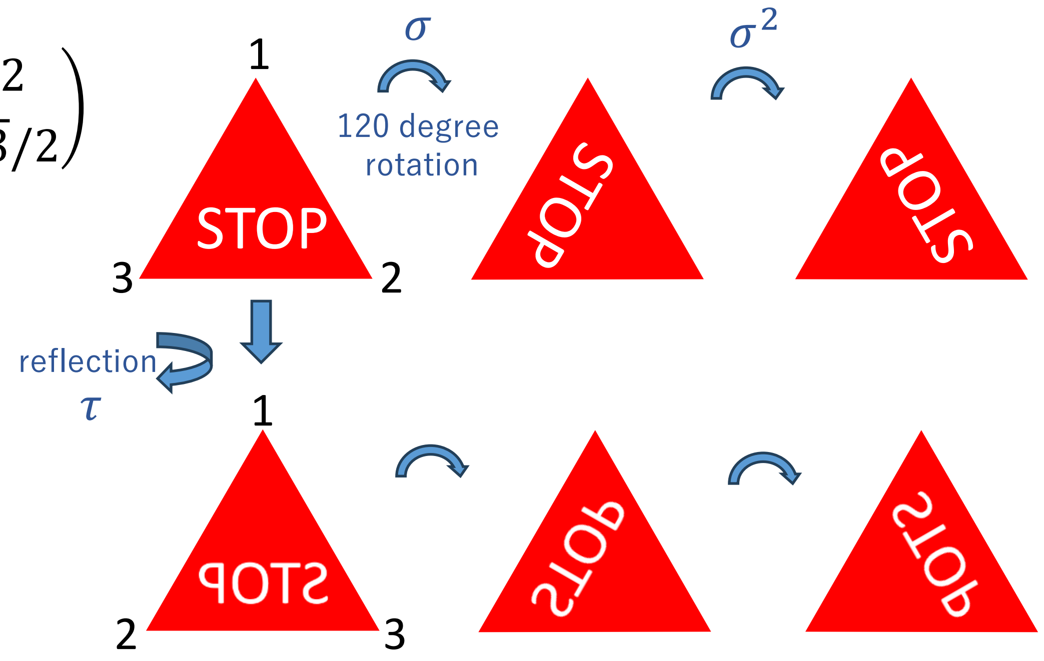
Basis matrix P does not depend on g

- For the following three classes, any finite dimensional representation is completely reducible:
  - Finite group (Maschke's theorem)
  - Locally compact Abelian group ($SO(2)$, etc)
  - Compact Lie groups ($O(n)$, etc).

- Example: irreducible representations of $\mathfrak{S}_3$

  1) Standard representation (2 dim)

$$\begin{cases} \rho_{st}(\sigma) = \begin{pmatrix} \cos(-2/3\pi) & -\sin(-2/3\pi) \\ \sin(-2/3\pi) & \cos(-2/3\pi) \end{pmatrix} = \begin{pmatrix} -\sqrt{3}/2 & 1/2 \\ -1/2 & -\sqrt{3}/2 \end{pmatrix} \\ \rho_{st}(\tau) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \end{cases}$$



  2) Alternating representation (1 dim)

$$\rho_{alt}(g) = \mathrm{sgn}(g)$$

  3) trivial (1 dim)

$$\rho_e(g) = 1$$

It is known that the above three are all the irreducible representations of $\mathfrak{S}_3$. (See, e.g., Fulton Harris)

Fourier transform and group

- Recall: Classical Discrete Fourier transform on function $\left\{0, \frac{1}{N}, \dots, \frac{N-1}{N}\right\} \subset \mathbb{S}^1$.

  Fourier Transform: $\hat{f}_n = \Phi(f)(n) = \sum_{k=0}^{N-1} e^{-i2\pi\frac{nk}{N}} f\left(\frac{k}{N}\right)$,

  Inversion: $f\left(\frac{k}{N}\right) = \sum_{n=0}^{N-1} e^{i2\pi\frac{nk}{N}} \hat{f}_n$.

- $\left\{0, \frac{1}{N}, \dots, \frac{N-1}{N}\right\} \subset \mathbb{S}^1$ is a group $\cong \mathbb{Z}_N$!
- Equivariance by shift (well-known):

$$\Phi\left(f\left(\cdot - \frac{a}{N}\right)\right) = \left(e^{i2\pi\frac{a}{N}n} \hat{f}_n\right)_{n=0}^{N}$$

or

$$\Phi(L_{a/N} f) = \tilde{L}_{a/N} \Phi(f).$$

Equivariance!

$\mathbb{Z}_N$ acts on $\mathbb{C}^N$ $(\hat{f}_n)$ by

$$\left(\frac{a}{N}\right) \circ \begin{pmatrix} \hat{f}_0 \\ \vdots \\ \hat{f}_{N-1} \end{pmatrix} := \begin{pmatrix} 1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & e^{2\pi i \frac{a}{N}(N-1)} \end{pmatrix} \begin{pmatrix} \hat{f}_0 \\ \vdots \\ \hat{f}_{N-1} \end{pmatrix}.$$

Fourier Transform on Group

- Fourier transform is generalized to locally compact Abelian groups (See e.g. Rudin)

What serves as the frequencies? -- The irreducible representations.

<u>Theorem</u>. Any irreducible representation of an Abelian group on $\mathbb{C}$-vector space is 1 dimensional.

$\hat{G} := \{\rho_n : G \to \mathbb{C}\}$: irreducible representations of G.

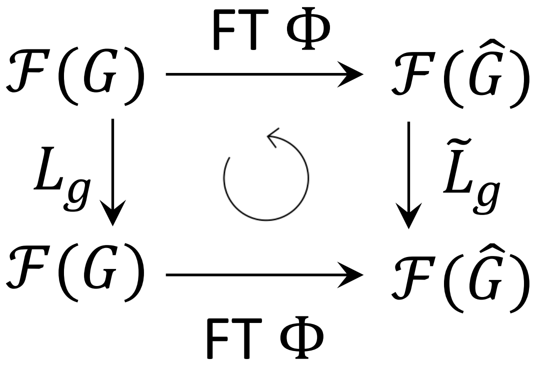
$$\rho_n\left(\frac{k}{N}\right) = e^{i2\pi\frac{nk}{N}}$$

$(n = 0,1,\ldots,N-1)$ are the irreducible representation of $\mathbb{Z}_N$

Fourier transform $$\hat{f}_n = \Phi(f)_n := \sum_{g\in G} \overline{\rho_n(g)}\, f(g)$$

Fourier inversion formula $$f(g) = \sum_n \rho_n(g)\hat{f}_n$$

Fourier Transform is equivariant



Prop. $\Phi\left(L_g f\right)_n = \rho_n(g)\Phi(f)_n$

So, if we define an action $\tilde{L}_g$ of G on $\left(\hat{f}_n\right)$ by $\hat{f}_n \mapsto \rho_n(g)\hat{f}_n$,

$$\Phi(L_g f) = \tilde{L}_g \Phi(f).$$ Equivariance

More generally, we define the convolution of f and g by

$$(f * \varphi)(g) = \sum_{h \in G} f(h)\varphi(h^{-1}g).$$

Prop. $\Phi(f * \varphi) = \Phi(f)\Phi(\varphi) = \left(\hat{f}_n \, \hat{\varphi}_n\right)_n$

Outline of This Lecture

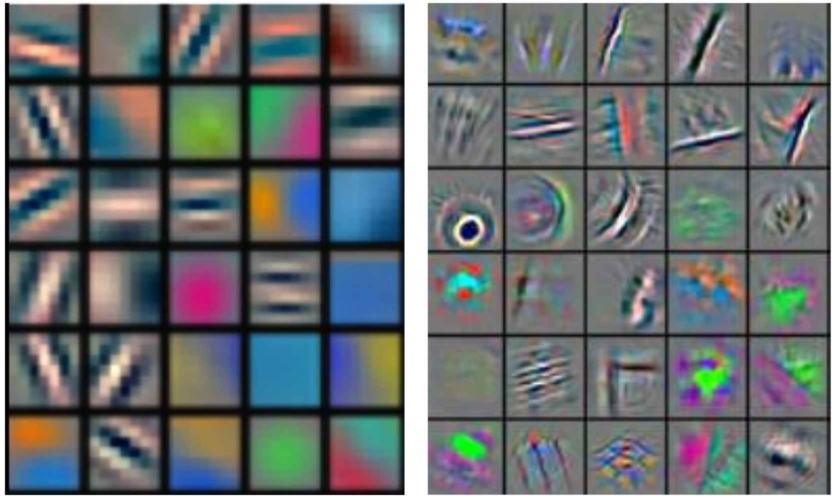
1. Introduction: symmetry and machine learning
2. Group theory I: Basics
3. Equivariant architecture: Group convolutional NN
4. Group theory II: Group representation and Fourier Transform
5. Representation Learning through Group Action

Representation Learning through Group Action

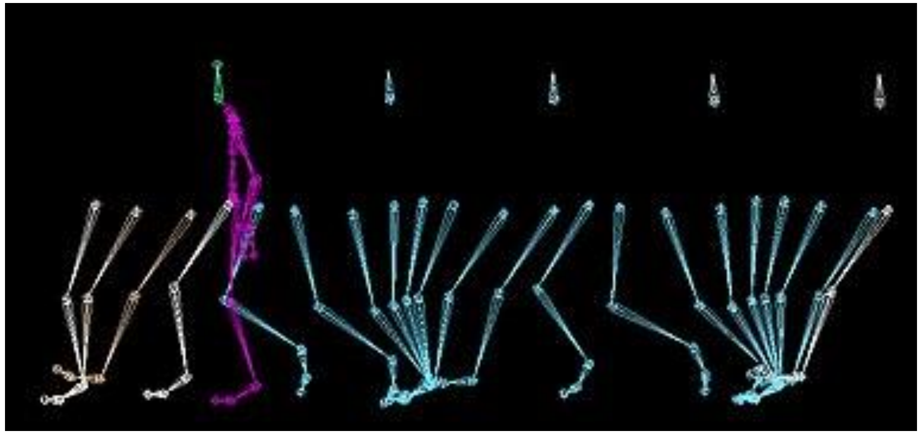
Miyato, Koyama, Fukumizu. NeurIPS 2022;
Koyama, Fukumizu., Hayashi, Miyato. ICLR 2024

Representation Learning

- Deep learning is expected to provide a good representation, which is effective to various downstream tasks.
- Group actions cause structured motions, which should be useful in extracting representations.



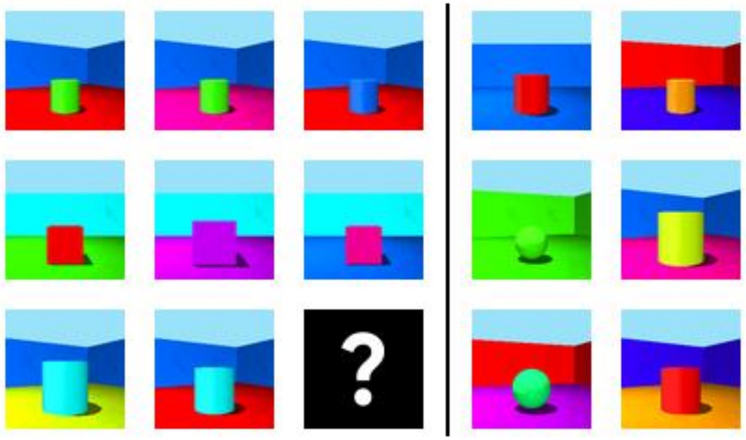
Feature visualization of convolutional net trained on ImageNet [Zeiler & Fergus 2013]



Driven by implicit group action.

Decomposed representation

- Decomposed/disentangled representations
  - Easier interpretation
  - Control of each factor



Locatello et al ICML2020

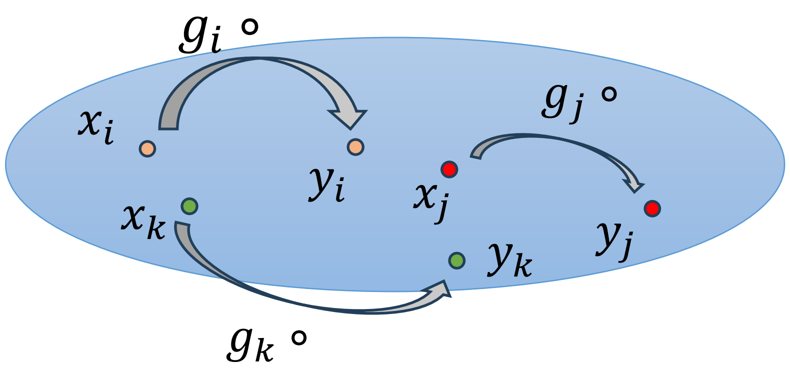
- Aim: Learning representations from group action data

  $(x_1, g_1 \circ x_1), (x_2, g_2 \circ x_2), \ldots$

  - "Group representation" gives decomposed representation.
  - Related to Fourier transformation: learning-based Fourier transform

Equivariant Representation Learning

(Miyato, Koyama, F. NeurIPS 2022; Koyama, F., Hayashi, Miyato ICLR 2024)



- General Problem setting
  - Some group G acts on data space $\mathcal{X}$.
  - Observation: many examples of group action
    - Paired data: $(x_i, g_i \circ x_i)$ $x_i \in \mathcal{X}$, $g_i \in G$
    - Sequences: $(x_i, g_i \circ x_i, g_i^2 \circ x_i, g_i^3 \circ x_i, \ldots)$
    - Triplet: $(x_i, g_i \circ x_i, g_i^2 \circ x_i)$

  - Various knowledge level on G or g
    - G and g_i may be known or may be unknown.
      → different types of settings.

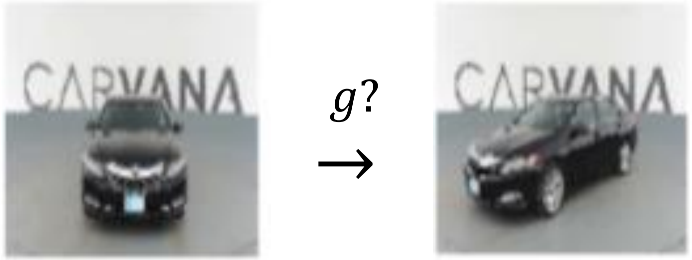
Existing approaches to equivariant learing

- Group G and its action are explicitly known and applied to data space
  - (Group) CNN: architecture adapted to a specific group
  - Data augmentation： augmentation using the group action

This work

- Group action does exist, but is not known explicitly
  - Not acting on the data space
  - May be observed with unknown nonlinearity

- Approach:
  Learn the group representation from data by equivariance constraint.



Rotation in the *latent* space

Nonlinear observation（fisheye lens)

Hardest setting: Unsupervised Learning of Equivariant Structure from Sequences (Miyato et al. NeurIPS 2022. https://github.com/takerum)

- Unsupervised learning: Neither G or $g \in G$ is known

- Data: many sequences $\{\mathbf{s}^{(i)}\}$

$$s_t^{(i)} = (g^{(i)})^t \circ s_0^{(i)}$$

- Generative model: sequences driven by group actions

  G: group (unknown) acting on $\mathcal{X}$.

  Sequence: $\mathbf{s} = (s_0, s_1, s_2, \ldots, s_T)$.

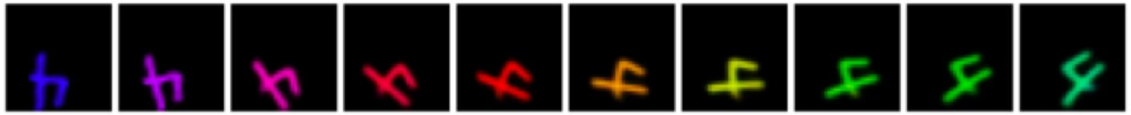
  $s_t = g^t \circ s_0$. $\quad s_t \in \mathcal{X}$

  Each seq $\mathbf{s}^{(i)}$ has its own $g^{(i)} \in G$, but unknown.

  Stationarity:
  $g^{(i)} \in G$ is the same in a sequence.

$\mathbf{s}^{(1)} = \mathbf{s}\left(g^{(1)}, s_0^{(1)}\right)$ e.g. $g^{(1)} = (15^\circ$ rot, (-2,1)shift, ΔRGB)

$\mathbf{s}^{(2)} = \mathbf{s}\left(g^{(2)}, s_0^{(2)}\right)$



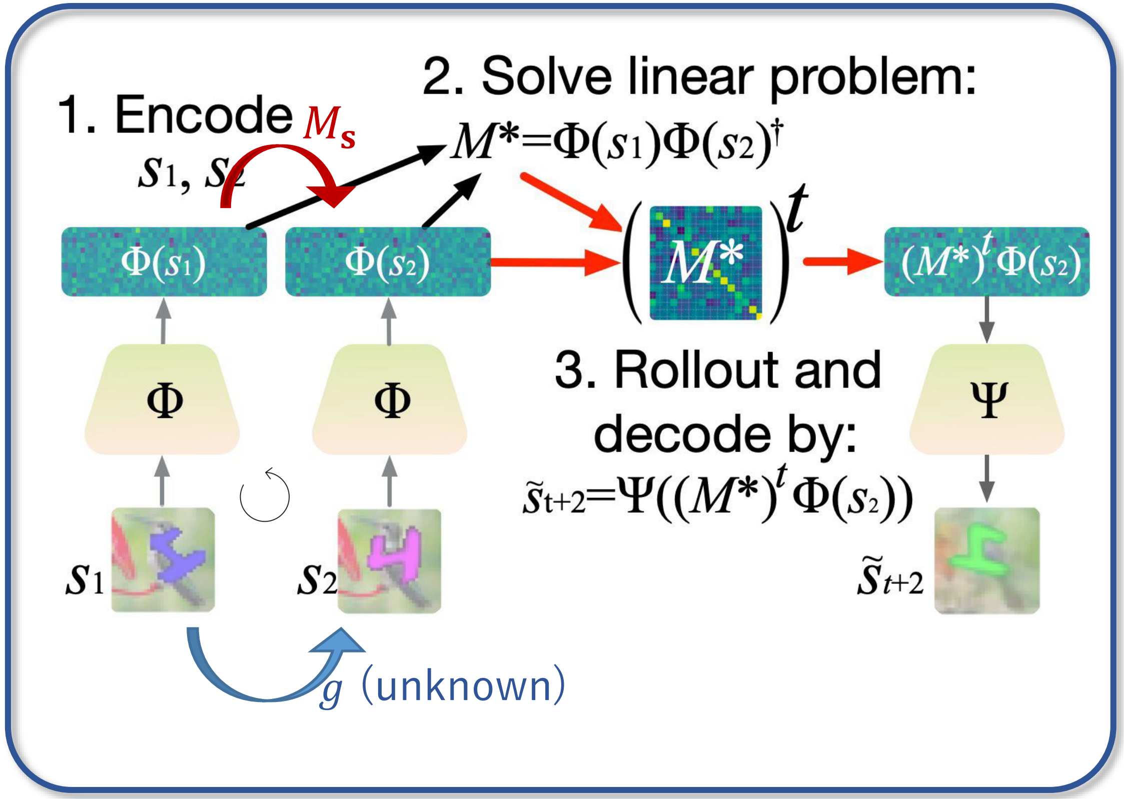
Rotation, Shifts, Color rotation

A sequence $\mathbf{s}$ is generated given an initial image s_0 and a group element $g \in G$.

Meta Sequential Prediction (MSP)

- Learning with Autoencoder
  Enc. $\Phi: \mathcal{X} \to \mathbb{R}^{m \times a}$,
  Dec. $\Psi: \mathbb{R}^{m \times a} \to \mathcal{X}$
  any NN

- **Learning group representation**
  Linear transform in the latent space
  $$\Phi\left(s_{t+1}^{(i)}\right) = M^{(i)}\Phi(s_t^{(i)})$$
  $M^{(i)}$: sequence-dependent matrix.



MSP Model

- Least square learning of $M_{\mathbf{s}}, \Phi, \Psi$

$$E\|M_{\mathbf{s}}\Phi(s_t) - \Phi(s_{t+1})\|^2 \quad \text{[Equivariance constrant]}$$
$$E\left\|\Psi\left(M_{\mathbf{s}}^{\ell}\Phi(s_t)\right) - s_{t+\ell}\right\|^2 \quad \text{[Pred./Reconst.]}$$

Disentanglement by Irreducible Decomposition

- Transition matrix $M_{\mathbf{s}}$ can depend on $\mathbf{s}$ (on $g \in G$, $M(g)$)

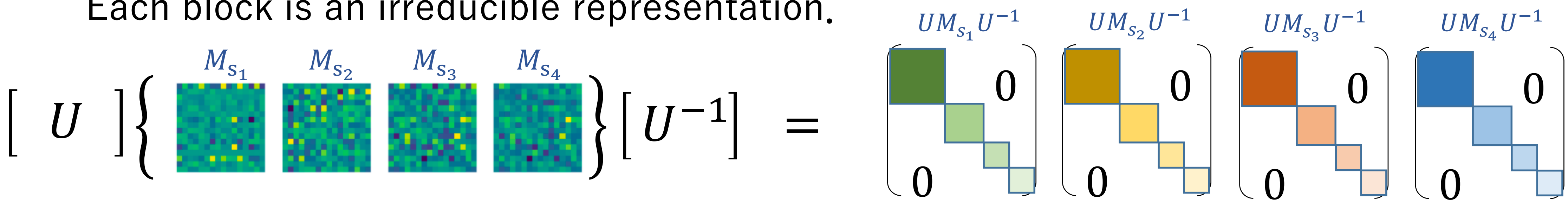
$$G \ni g \mapsto M(g) \in GL(\mathbf{R}^a)$$ Group representation G

(under some conditions)

- Irreducible decomposition → Simultaneous block-digonalization of $\{M_{\mathbf{s}}\}_{\mathbf{s}}$

  Common change of bases

  Each block is an irreducible representation.



Common matrix U

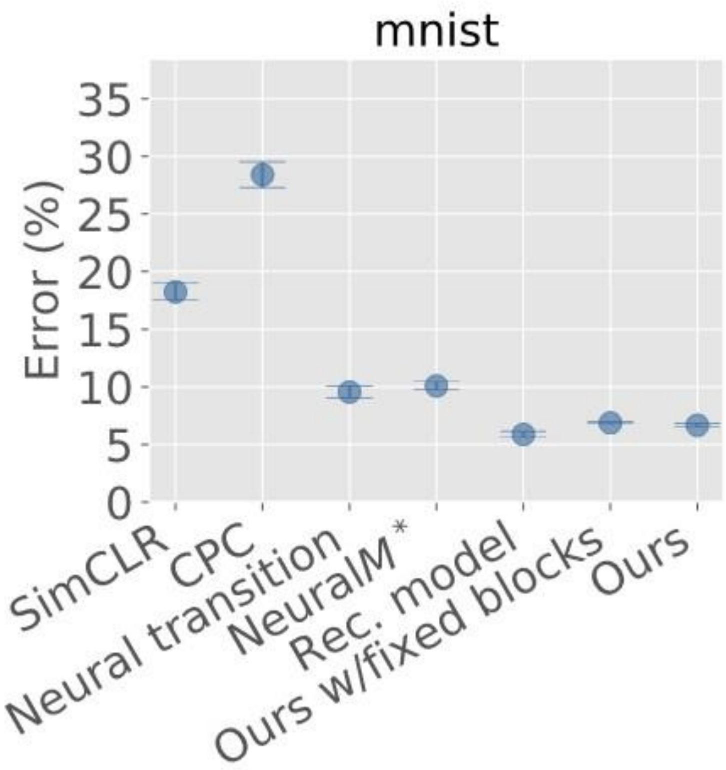
Disentanglement by irreducible repr.

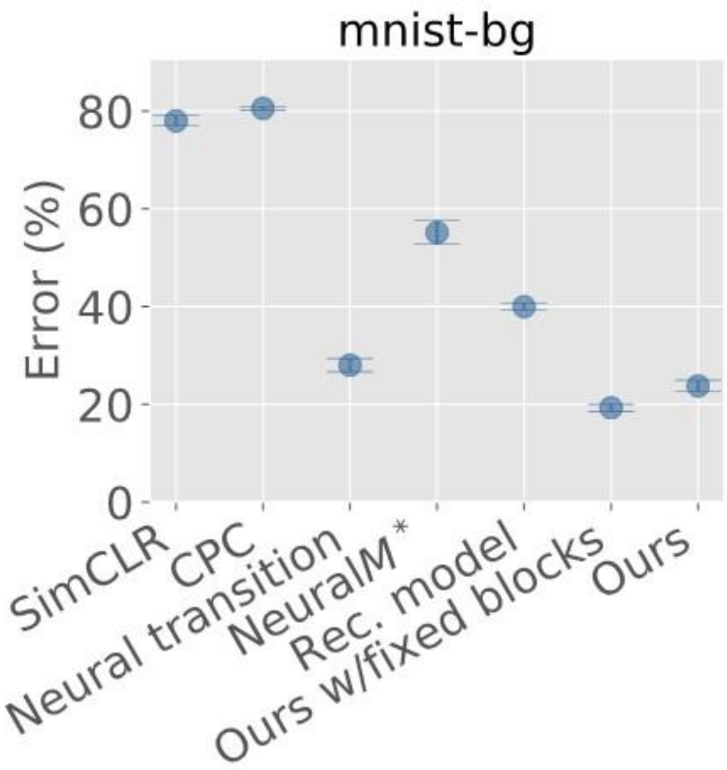
＊Spectral clustering method is applied to SBD.

Experiment 1: Effective representation

- Linear classification with learned $\Phi(x)$

  Trained with only “4” → 10 class classification for “0”,⋯,”9”.



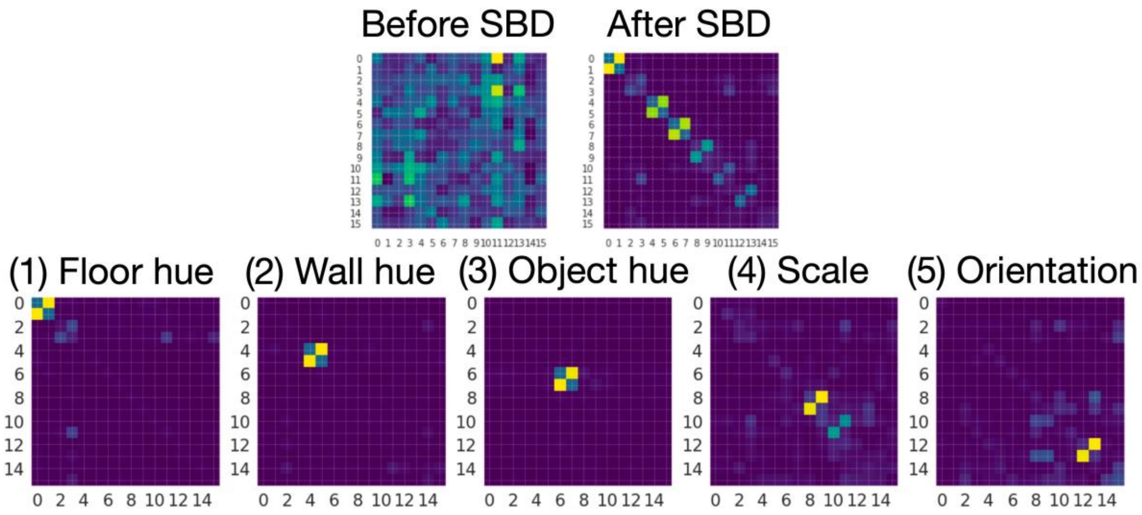


* SimCLR, CPC: standard methods of self-supervised learning

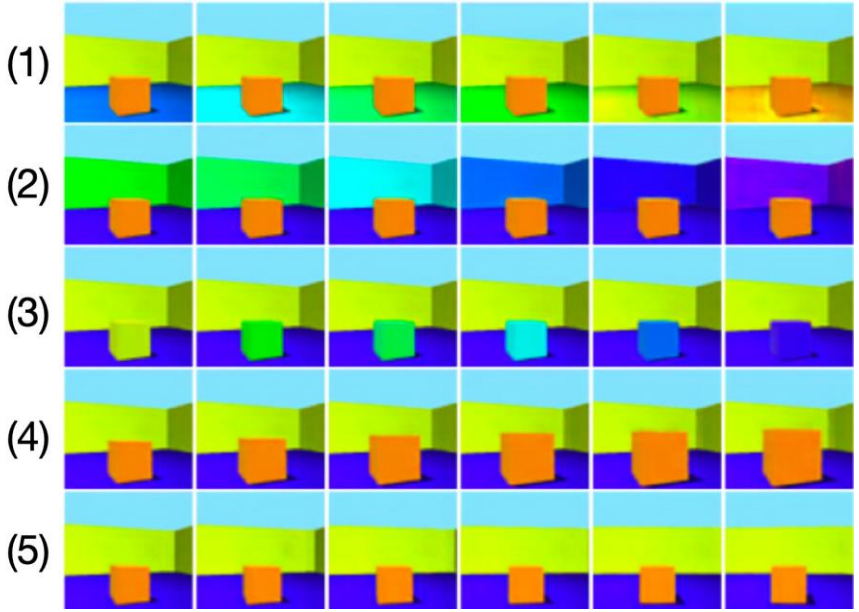
Experiment 2: disentanglement

Rendered image sequences:

Product group of 5 types of changes



Blocks obtained by simultaneous block diagonalization



Reconstruction by each block

$$\hat{s}_t := \Psi\left(M_b^t \Phi(s_0)\right)$$

Neural Fourier Transform (Koyama et al ICLR 2024)

- Proposed method: a nonlinear extension of Fourier transform
  - Learning by the equivariance constraint.

$$L_g x \approx \Psi \circ M(g) \circ \Phi(x) = \underbrace{\Psi \circ P^{-1}}_{\text{Inv. Fourier Transform}} \circ \underbrace{\begin{pmatrix} \rho_1(g) & & O \\ & \rho_2(g) & \\ O & & \ddots \end{pmatrix}}_{\text{Irreducible representations}} \circ \underbrace{P \circ \Phi(x)}_{\text{Fourier Transform}}.$$

- *c.f.* Classical Fourier transform on function $\left\{0, \frac{1}{N}, \ldots, \frac{N-1}{N}\right\} \subset \mathbb{S}^1$.

$$\hat{f}_n = \Phi(f)(n) = \sum_{k=0}^{N-1} e^{-i2\pi\frac{k}{N}} f\left(\frac{k}{N}\right)$$

- Equivariance: $\Phi(f(\cdot - a/N)) = e^{i2\pi\frac{a}{N}n} \hat{f}_n$

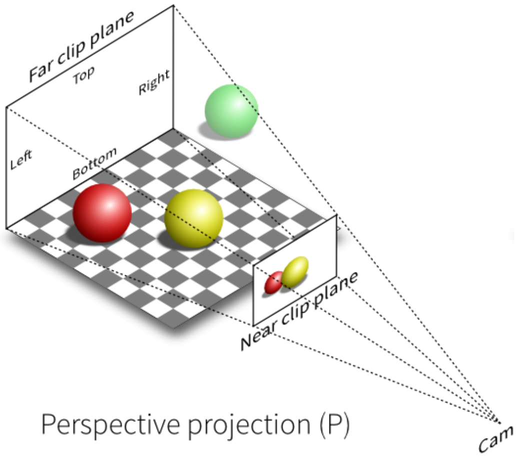
  or $L_{a/N} f = \Phi^{-1} \circ \begin{pmatrix} 1 & \cdots & O \\ \vdots & \ddots & \vdots \\ O & \cdots & e^{i2\pi a \frac{N-1}{N}} \end{pmatrix} \circ \Phi(f)$

- Neural Fourier Transform:

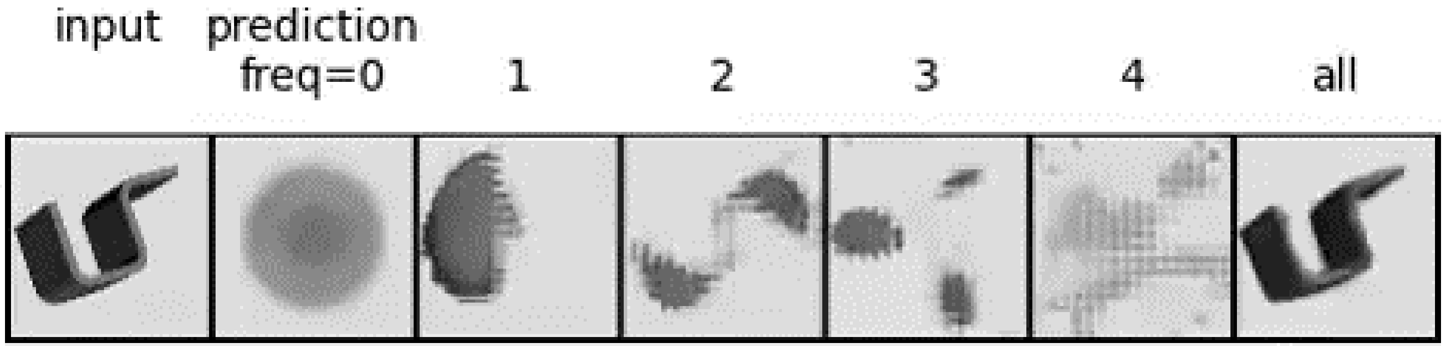
  Trained Φ and Ψ can be interpreted as nonlinear FT and Inv FT.

  - NFT works for “data”, while standard FT works for functions.
  - NFT learns the transforms though data (examples of actions) without knowing the group or group actions.
  - It uses only necessary frequencies based on data.

- g-NFT：3D image synthesis from 2D training images
  - Data: Paired 2D images and 3D rotation (S, S', R). $R \in SO(3)$, S, S': 2D images, $S' = RS$
  - $M(g)$ Spherical harmonics. Only encoder Φ and decoder Ψ are trained.
  - Testing: Provide a 2D image X_0 (not in the training data) and apply arbitrary 3D rotation g by $\hat{\Psi}\left(M(g)\hat{\Phi}(X_0)\right)$.



Perspective projection (P)



Wrap-up

- Group theory
  - Extract symmetry in nature
  - Well developed mathematics
  - Group representation: approach with linear algebra

- Machine leaning with group actions
  - Three approaches
    - Data augmentation
    - Built-in architecture: (Group) convolutional neural networks
    - Representation learning with group representation,

- Group CNN: Realize equivariant mapping
  - Guaranteed equivariance (*c.f.* data augmentation)
  - Compact representation
    - Data: Low dimensional expression
    - Model: Smaller models, efficient learning

- Equivariant representation learning
  - Implicit group action
    - Group or group action may not be known
    - Paired or sequential data are required.
    - Learning through equivariance constraint
  - Achieves nonlinear extension of Fourier transform
    - Data-based Fourier transform
    - Extract necessary frequencies


End
Enjoy Okinawa!

References

- Online materials
  - U Amasterdam - An Introduction to Group Equivariant Deep Learning https://uvagedl.github.io/

- Textbooks on Group Theory
  - [Group Theory] J.J. Rothman. An Introduction to the Theory of Groups. Springer 1995.
  - [Representation Theory] W. Fulton, J. Harris. Representation Theory. Springer 2004
  - [Fourier Analysis] W. Rudin. Fourier Analysis on Groups. John Wiley& Sons, 1991

- Papers

  - Fukushima, Kunihiko (1980). "Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position". *Biological Cybernetics* **36** (4): 193–202. doi:10.1007/BF00344251
  - Zhang, Wei (1988). "Shift-invariant pattern recognition neural network and its optical architecture". *Proceedings of annual conference of the Japan Society of Applied Physics*.
  - Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, Backpropagation Applied to Handwritten Zip Code Recognition Archived 2020-01-10 at the Wayback Machine; AT&T Bell Laboratories
  - T. Cohen and M. Welling. Group Equivariant Convolutional Networks. ICML 2016
  - Lafarge, M. W., Bekkers, E. J., Pluim, J. P. W., Duits, R., & Veta, M. (2021). Roto-translation equivariant convolutional networks: Application to histopathology image analysis. *Medical Image Analysis*, *68*, 101849. https://doi.org/https://doi.org/10.1016/j.media.2020.101849
  - T. Cohen and M. Welling. Steerable CNNs. ICLR 2017
  - M. Weiler and G. Cesa. General E(2) - Equivariant Steerable CNNs. NeurIPS 2019

- M. Weiler, M. Geiger, M. Welling, W. Boomsma, T. Cohen. 3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data. NeurIPS 2018 https://papers.nips.cc/paper_files/paper/2018/hash/488e4104520c6aab692863cc1dba45af-Abstract.html
- Weiler, Hamprecht, Storath, "Learning Steerable Filters for Rotation Equivariant CNNs" CVPR 2018
- T. Miyato, M. Koyama, K. Fukumizu. Unsupervised Learning of Equivariant Structure from Sequences. NeurIPS 2022
- M. Koyama, K. Fukumizu, K. Hayashi, T. Miyato. Neural Fourier Transform: A General Approach to Equivariant Representation Learning. ICLR 2024 https://openreview.net/forum?id=eOCvA8iwXH

[Theoretical analysis on invariant/equivariant neural networks]

- Clare Lyle, Mark van der Wilk, Marta Kwiatkowska, Yarin Gal, Benjamin Bloem-Reddy. On the Benefits of Invariance in Neural Networks. https://arxiv.org/abs/2005.00178
- Akiyoshi Sannai, Masaaki Imaizumi, and Makoto Kawano. Improved generalization bounds of group invariant/equivariant deep networks via quotient feature spaces. In Uncertainty in Artificial Intelligence, pages 771–780. PMLR, 2021.
- Arash Behboodi, Gabriele Cesa, Taco Cohen. A PAC-Bayesian Generalization Bound for Equivariant Networks. NeurIPS 2022 https://openreview.net/forum?id=6dfYc2IUj4
- J. Shawe-Taylor. Building symmetries into feedforward networks. In 1989 First IEE International Conference on Artificial Neural Networks, (Conf. Publ. No. 313), page 158–162, Oct 1989.

Appendix

Semidirect Product of Groups

- Def. G: group. $N \lhd G$ (normal subgroup), $K < G$ (subgroup).
  G is a semidirect product of N and K (denoted by $G = N \rtimes K$) if
  $NK = G$ and $N \cap H = \{e\}$.

  - There are several other equivalent definitions.
  - $g \in G$ is uniquely written as $g = hk$, where $h \in N$, $k \in K$.
  - Take $g_1 = h_1 k_1$, $g_2 = h_2 k_2$. $g_1 g_2 = h_1 k_1 h_2 k_2 = h_1 k_1 h_2 k_1^{-1} k_1 k_2$.
    From $N \lhd G$, there is $\tilde{h} \in N$ such that $k_1 h_2 k_1^{-1} = \tilde{h}$.
    Semidirect product specifies the conjugate $\phi(k): N \to N$, $h \mapsto khk^{-1}$.
    For 'direct' product $khk^{-1} = h$.

Coset and Quotient Group

- Coset

  G: group. $H < G$: subgroup.

  $gH := \{gh \in G \mid h \in H\}$ for any $g \in G$. (left coset)

  - Fact: $gH = g'H$ if and only if $g^{-1}g' \in H$.

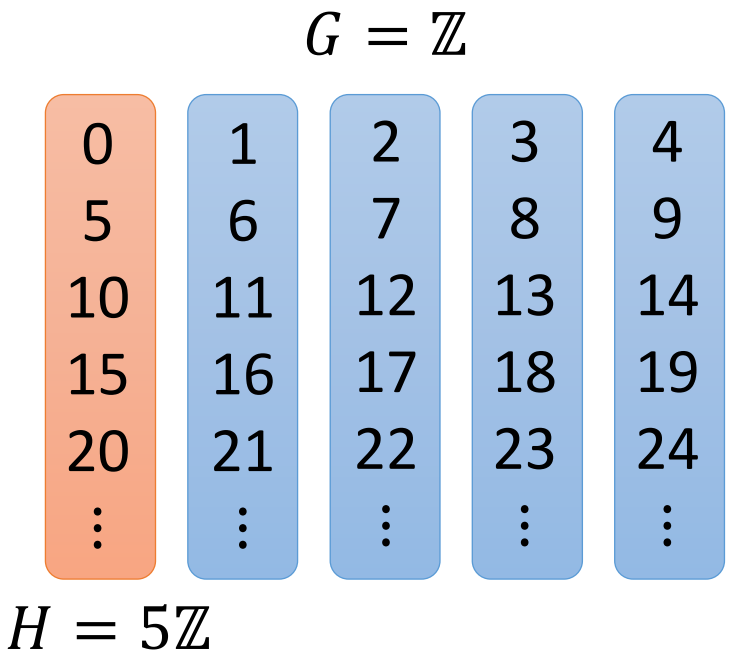
    [Exercise: Check this.]

  - The cosets $\{\, gH \mid g \in G \,\}$ gives a partition of G.

    Aka residue class.

  - Example. $G = \mathbb{Z},\ H = 5\mathbb{Z}$.

    The cosets are $H, 1 + H, 2 + H, 3 + H, 4 + H$. Partioned by the residues.



- Normal subgroup

  A subgroup N of G is called normal if $ghg^{-1} \in N$ for any $h \in N$ and $g \in G$. Often denoted by $N \lhd G$.

- Quotient group
  - For a subgroup $H < G$, the cosets $\{gH \mid g \in G\}$ may not form a group.

  Proposition. For a normal subgroup N of G, the cosets $\{gN \mid g \in G\}$ is a group with the operation $g_1N \cdot g_2N = g_1g_2N$.

  This is called a quotient group and denoted by G/N.
  Example: $\mathbb{Z}/n\mathbb{Z}$

  (Proof Sketch) Because N is normal, for any $h_1 \in N$, there is $h' \in N$ such that $g_2^{-1}h_1g_2 = h'$. Then, for any $h_2 \in H$, $g_1h_1g_2h_2 = g_1g_2h'h_2 \in g_1g_2H$. The multiplication is thus well defined. N is the identity and $g^{-1}N$ is the inverse of gH.

- Example

  $E(n)$ has a subgroup $\mathbb{R}^n$(translations).

  - $\mathbb{R}^n \lhd E(n)$ (normal subgroup).

    $$\begin{pmatrix} A & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I_n & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A & a \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} A & a \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I_n & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A^{-1} & -A^{-1}a \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} I_n & Ab \\ 0 & 1 \end{pmatrix}$$

  - $E(n)/\mathbb{R}^n \cong O(n)$

  - $E(n) \neq \mathbb{R}^n \times O(n)$

    In the direct product $G_1 \times G_2$, G_1 and G_2 must be commutative.

  - $E(n) \cong \mathbb{R}^n \rtimes O(n)$ (Semidirect product) → Explained in the next slide.

Finite Abelian groups

A group G is a finite group if it has finite elements.
The number of elements is called the order of G and denoted by $|G|$.

- Recall. Cyclic group $\mathbb{Z}_n := \mathbb{Z}/n\mathbb{Z}$. $|\mathbb{Z}_n| = n$

  $\mathbb{Z}_6 \cong \mathbb{Z}_2 \times \mathbb{Z}_3$.

- Theorem. A **finite Abelian** group G is isomorphic to a direct product of cyclic groups of the form

  $$G \cong \mathbb{Z}_{p_1^{a_1}} \times \cdots \times \mathbb{Z}_{p_k^{a_k}}$$

  where p_i's are prime numbers. $n = p_1^{a_1} \cdots p_k^{a_k}$.

Note: For $|G| = 4$, there are two cases:
$\mathbb{Z}_4$ and $\mathbb{Z}_2 \times \mathbb{Z}_2$, which are not isomorphic.