Equivariance II: Steerable CNNs

Gabriele Cesa

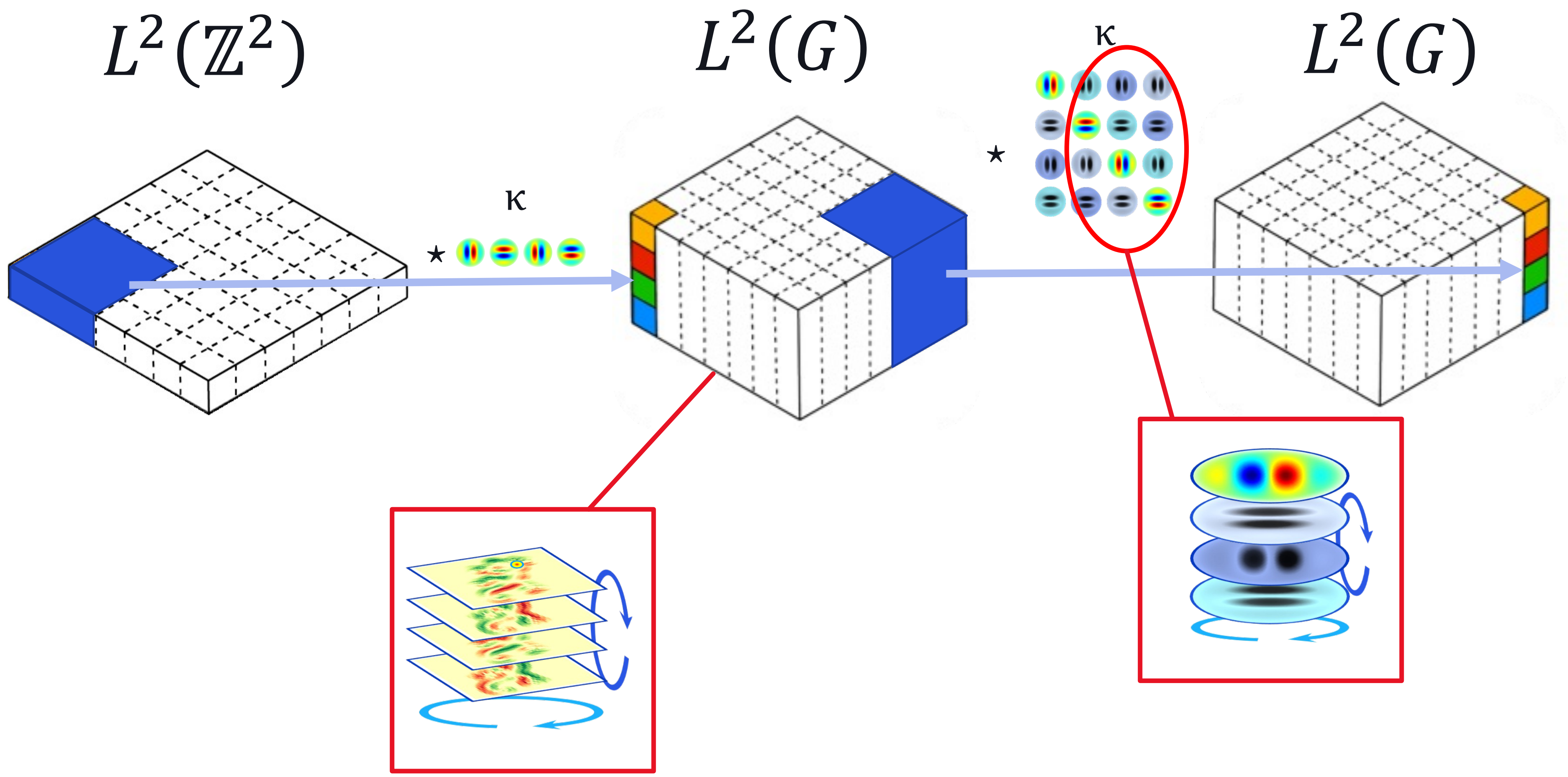
Senior Engineer, Qualcomm AI Research*

University of Amsterdam, AMLab

13/06/2024

*Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

Brief Recap: Lifting Convolution and Group Convolution



Brief Recap: Group Convolution is all you need

We have seen that Group Convolution provides an **equivariant linear map**

$$[\kappa \star \cdot]: L^2(X) \to L^2(G)$$

Group-CNNs are *universal approximators* for equivariant functions

Siamak Ravanbakhsh. Universal Equivariant Multilayer Perceptrons. *International Conference on Machine Learning (ICML), 2020.*

Haggai Maron, Ethan Fetaya, Nimrod Segol, and Yaron Lipman. On the Universality of Invariant Networks. *International Conference on Machine Learning (ICML), 2019*

Wataru Kumagai and Akiyoshi Sannai. Universal Approximation Theorem for Equivariant Maps by Group CNNs, *2020*

Sho Sonoda, Isao Ishikawa, and Masahiro Ikeda. Universality of group convolutional neural networks based on ridgelet analysis on groups. *Advances in Neural Information Processing Systems (NeurIPS), 2022*

Brief Recap: Group Convolution is all you need

We have seen that Group Convolution provides an **equivariant linear map**

$$[\kappa \star \cdot]: L^2(X) \rightarrow L^2(G)$$

Any equivariant linear map of this form is a convolution! (typically, under some minor conditions)

- For any **compact** group G (Kondor and Trivedi, 2018), (Cohen et al., 20120)
- For any **Lie** group (Bekker, 2020)
- Unimodular Lie groups (Aronsson, 2022)

Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. *International Conference on Machine Learning (ICML), 2018.*

Taco S. Cohen, Mario Geiger, and Maurice Weiler. A general theory of equivariant CNNs on homogeneous spaces. *Advances in Neural Information Processing Systems (NeurIPS), 2019*

Erik J Bekkers. B-spline CNNs on Lie groups. *International Conference on Learning Representations (ICLR), 2020.*

Jimmy Aronsson. Homogeneous vector bundles and G-equivariant convolutional neural networks. *Sampling Theory, Signal Processing, and Data Analysis, 2022*

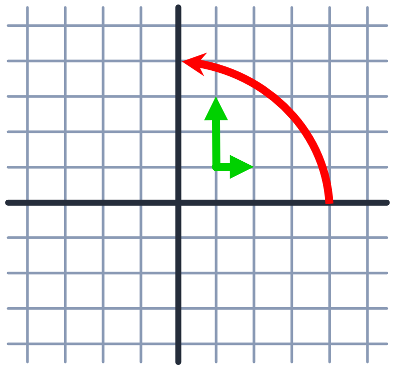
Group Convolution: Challenges

$$[\kappa \star f](g) = \int_{x \in X} \kappa(g^{-1}\boldsymbol{x})\ f(\boldsymbol{x}) d\mu(x)$$

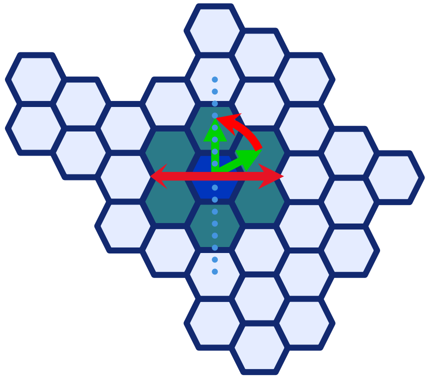
Integrate over X

Transform filter $\kappa \in L^2(X)$
by any group element
$g \in G$

Parameterize
features f
and filters $\kappa \in L^2(X)$



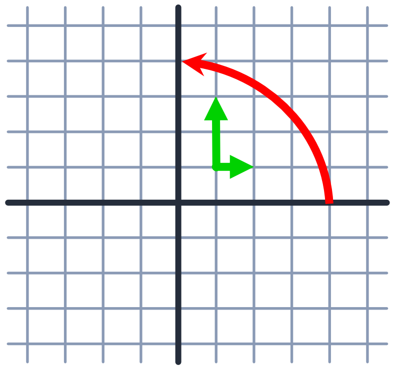
$p4m = (\mathbb{Z}^2, +) \rtimes D_4$



$p6m = (\mathbb{Z}^2, +) \rtimes D_6$

- Often some form of discretization is required, causing artifacts and harming equivariance

Group Convolution: Challenges



$p4m = (\mathbb{Z}^2, +) \rtimes D_4$

$$[\kappa \star f](g) = \int_{x \in X} \kappa(g^{-1}\,\boldsymbol{x})\ f(\boldsymbol{x}) d\mu(x)$$

- Often some form of discretization is required, causing artifacts and harming equivariance
- What if input or output are not in $L^2(\mathbb{R}^n)$ or $L^2(G)$?
  - e.g. vector fields?
- The group G might be infinit or too large to fully enumerate
  - $G = SO(3)$ rotations of a point cloud
  - $n!$ permutations of a set or nodes of a graph

Agenda

Steerable Filters

Steerable Fields and Representation Theory

Group Convolution to Steerable Convolution

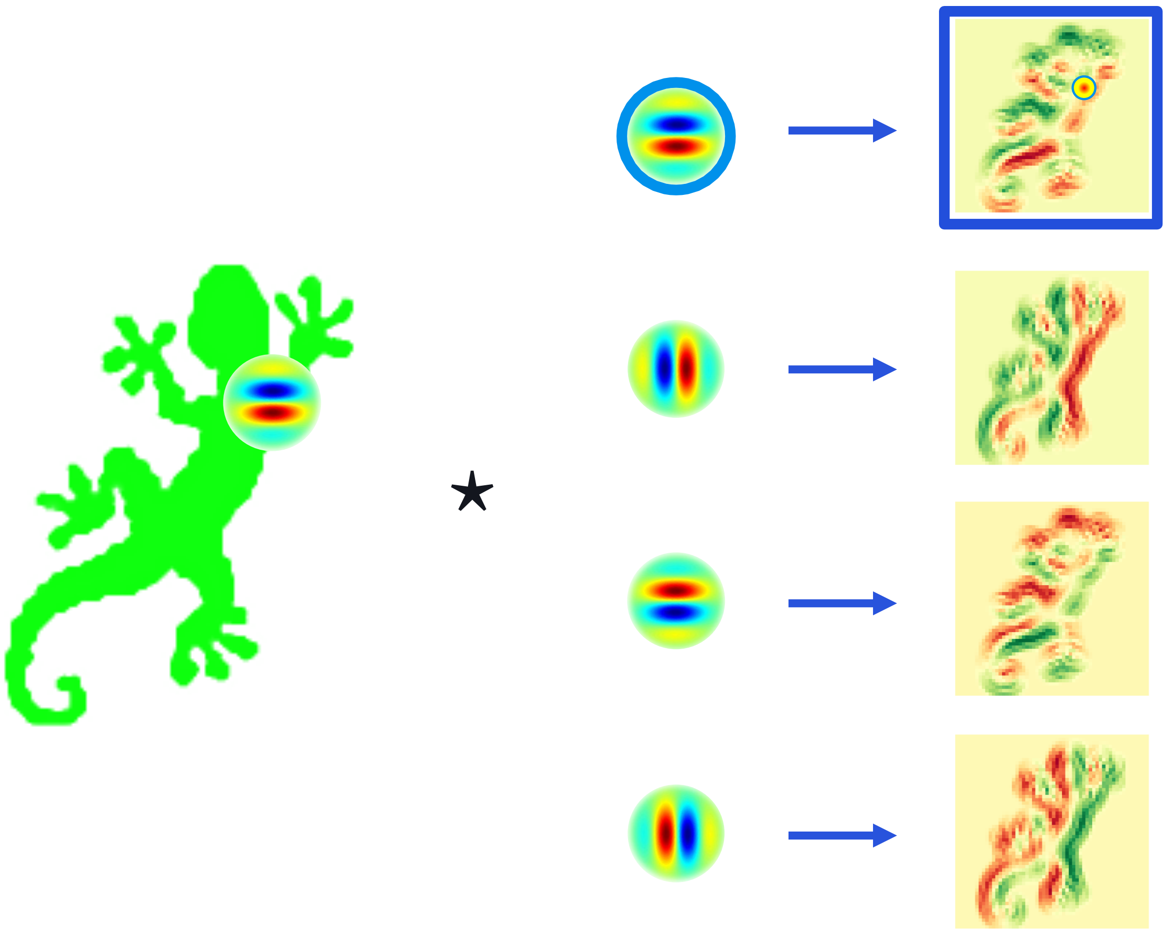
Steerable CNNs

Solving the kernel constraint

Instances and Applications

Group Convolution

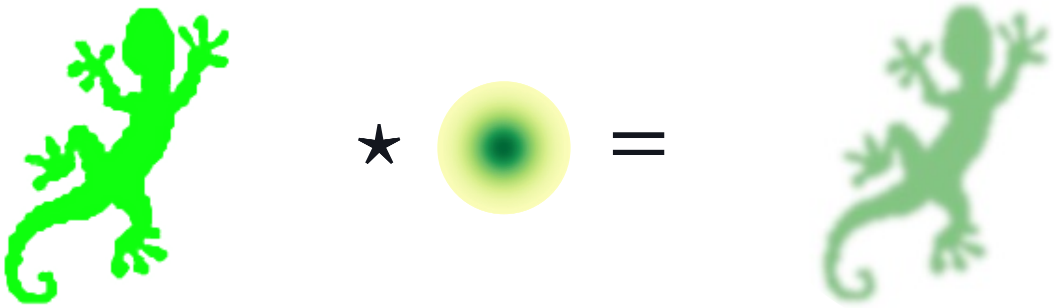
$$[\kappa \star \cdot]: L^2(X) \to L^2(G)$$



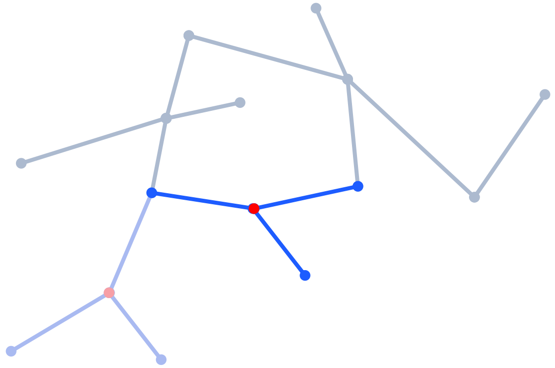
Isotropic Filters

$$[\kappa \star \cdot]: L^2(X) \to L^2(X)$$

- Isotropic filters: the response does not change when rotated



- Not very expressive
- Analogous to Graph Message Passing: no directional dependence!



Steerable Filters

- Filter can be rotated via linear combination

$$\kappa(r, \phi) = w(r)\cos(\phi + \beta)$$

- e.g. filters used for edge-detection in classical Computer Vision

William T. Freeman and Edward H. Adelson. The design and use of steerable filters.
IEEE Transactions on Pattern Analysis & Machine Intelligence, 1991

Steerable Filters

$$
\begin{aligned}
[R_\theta . \kappa](r,\phi) &= \kappa(r, \phi - \theta) \\
&= w(r)\cos(\phi + \beta - \theta) \\
&= w(r)\cos(\phi+\beta)\cos\theta - w(r)\sin(\phi+\beta)\sin\theta \\
&= \cos\theta\, \kappa(r,\phi) - \sin\theta\, \kappa\left(r, \phi - \frac{\pi}{2}\right) \\
&= \cos\theta\, \kappa(r,\phi) - \sin\theta\, R_{\frac{\pi}{2}} . \kappa(r,\phi)
\end{aligned}
$$

- Filter can be rotated via linear combination

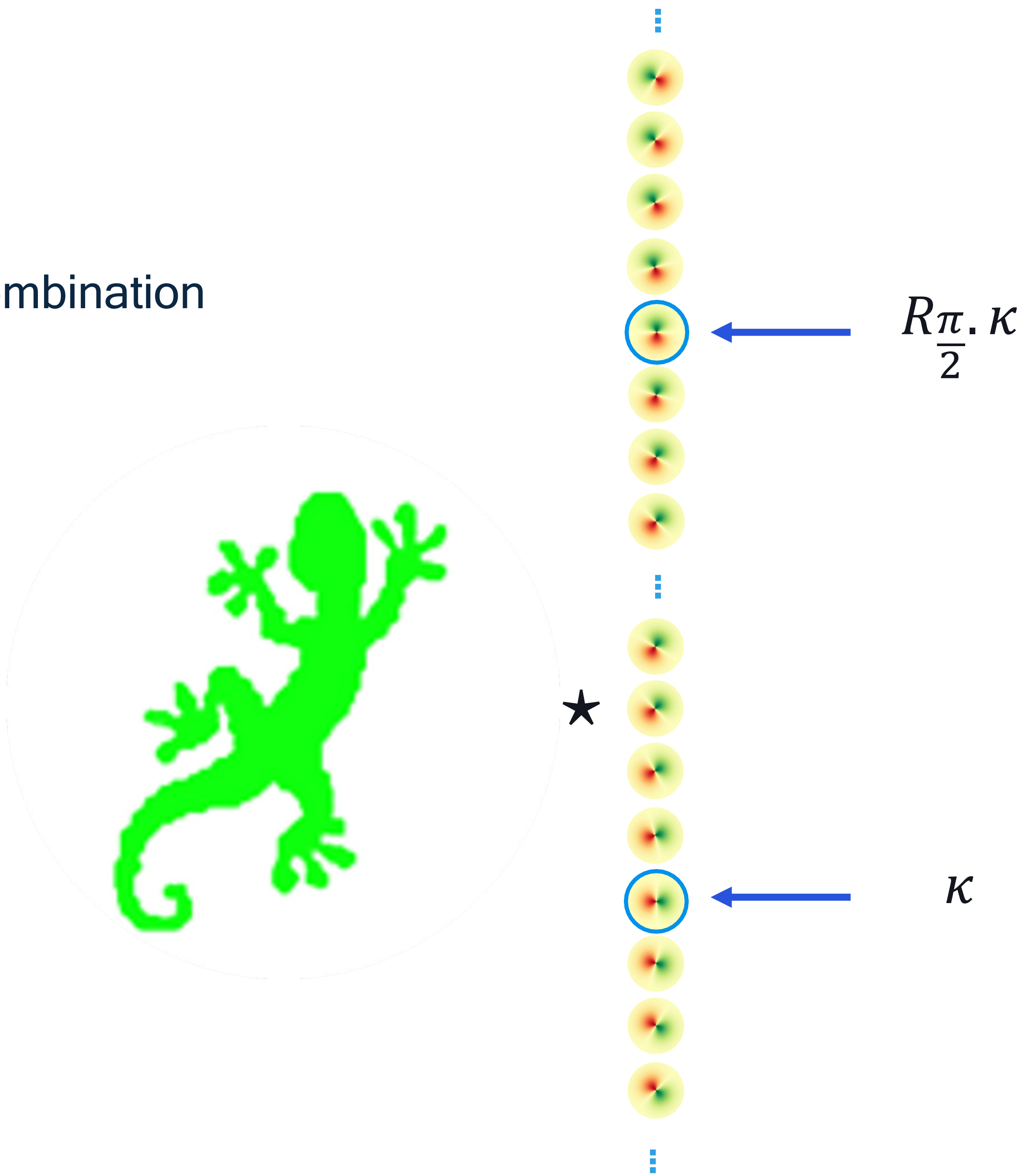
$$\kappa(r,\phi) = w(r)\cos(\phi+\beta)$$

- e.g. filters used for edge-detection in classical Computer Vision

William T. Freeman and Edward H. Adelson. The design and use of steerable filters.
IEEE Transactions on Pattern Analysis & Machine Intelligence, 1991

Steerable Filters

- Filter can be rotated via linear combination



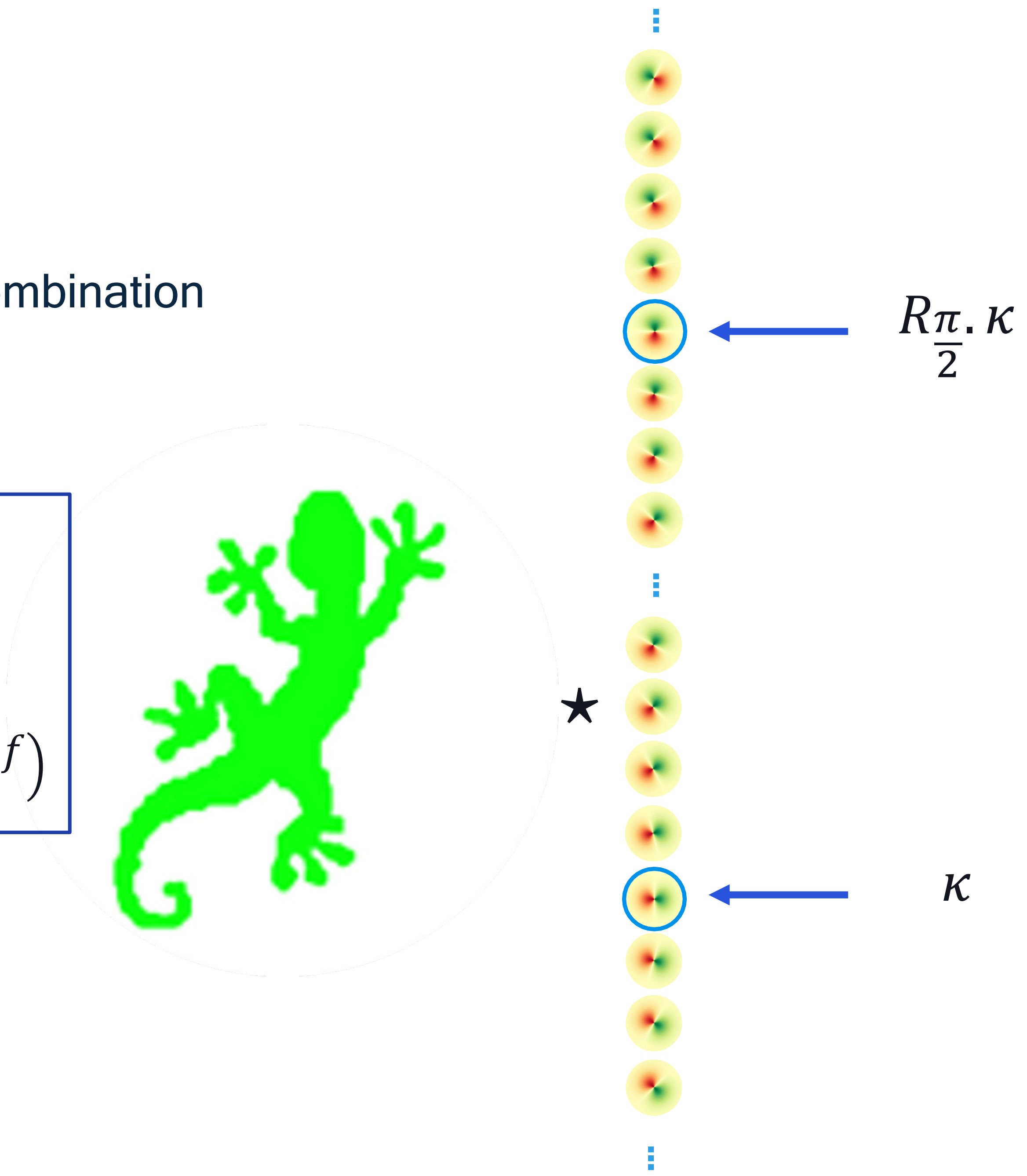
Steerable Filters

- Filter can be rotated via linear combination

Thanks to *linearity* of convolution operator:

$$R_\theta . \kappa \star f = \left(\cos\theta\, \kappa - \sin\theta\, R_{\frac{\pi}{2}} . \kappa\right) \star f$$

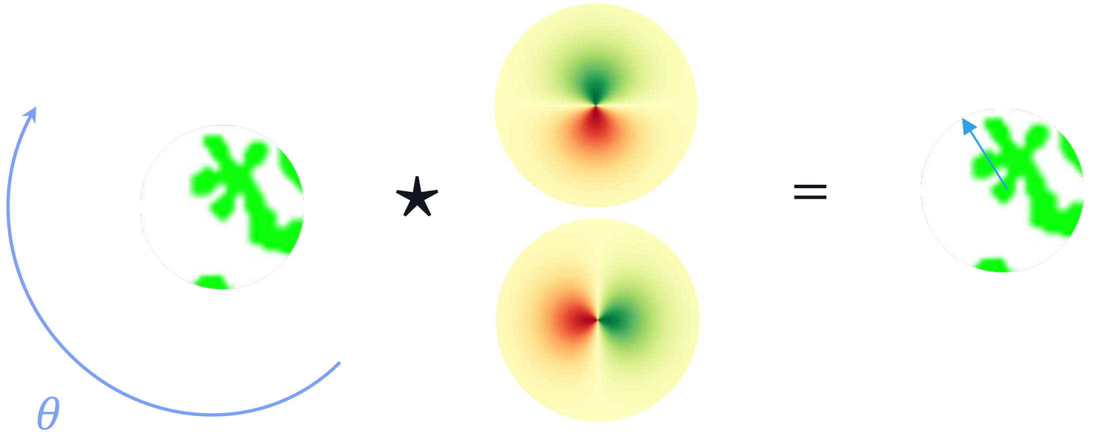
$$= \cos\theta\, (\kappa \star f) - \sin\theta \left(R_{\frac{\pi}{2}} . \kappa \star f\right)$$



Steerable Filters

- Filter can be rotated via linear combination

$$\rho(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

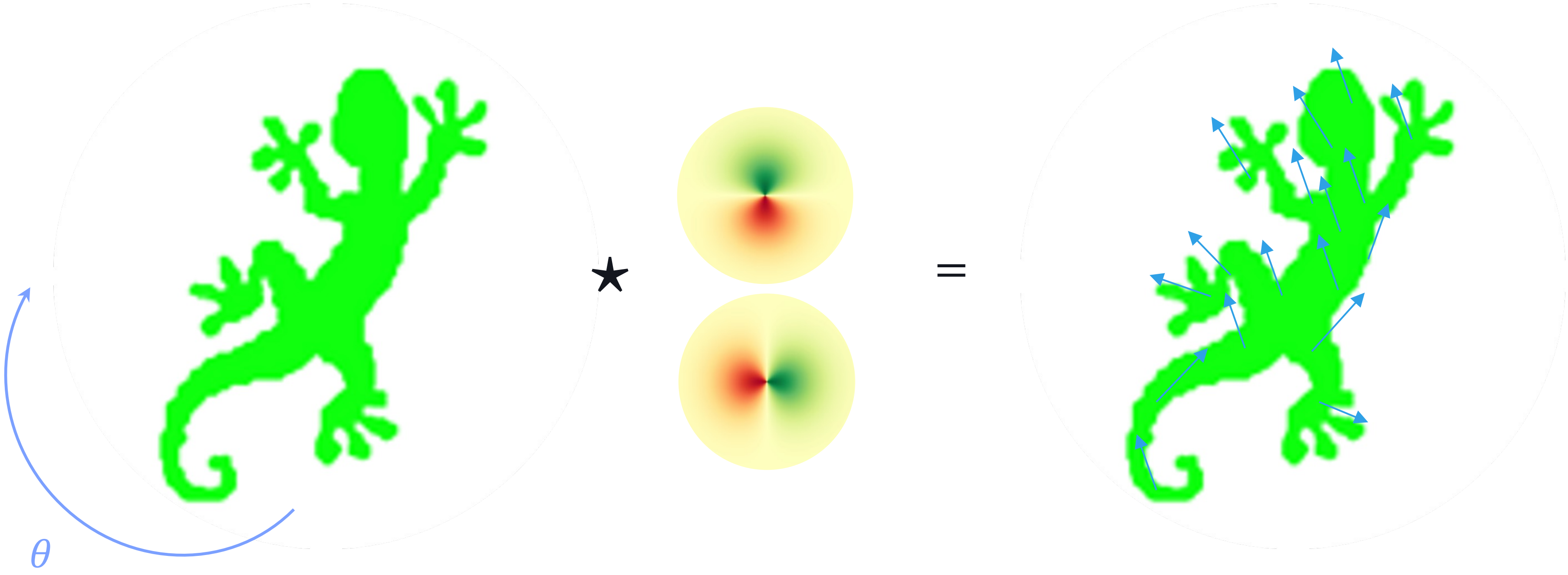


$$\boxed{R_\theta . \kappa \star f = \cos\theta\,(\kappa \star f) - \sin\theta\,\left(R_{\frac{\pi}{2}} . \kappa \star f\right)}$$

Steerable Filters

- Filter can be rotated via linear combination

$$\rho(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

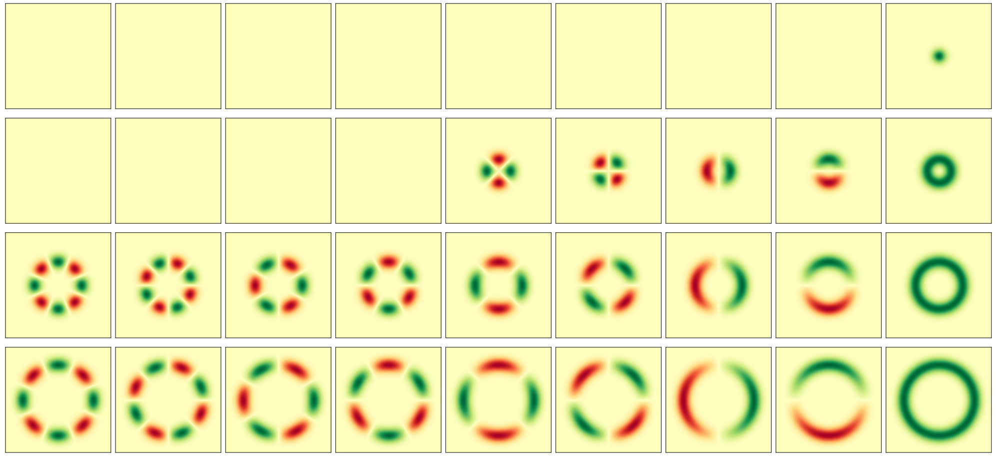


$$\boxed{R_\theta . \kappa \star f = \cos\theta\,(\kappa \star f) - \sin\theta\left(R_{\frac{\pi}{2}} . \kappa \star f\right)}$$

Steerable Filters: Other examples

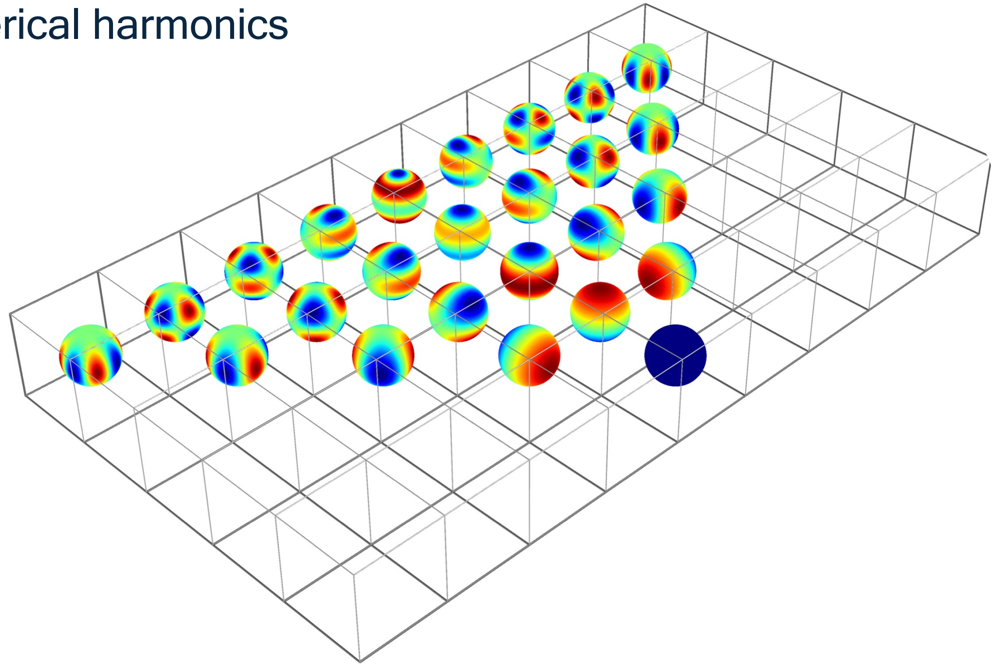
- Filter can be rotated via linear combination

Circular harmonics



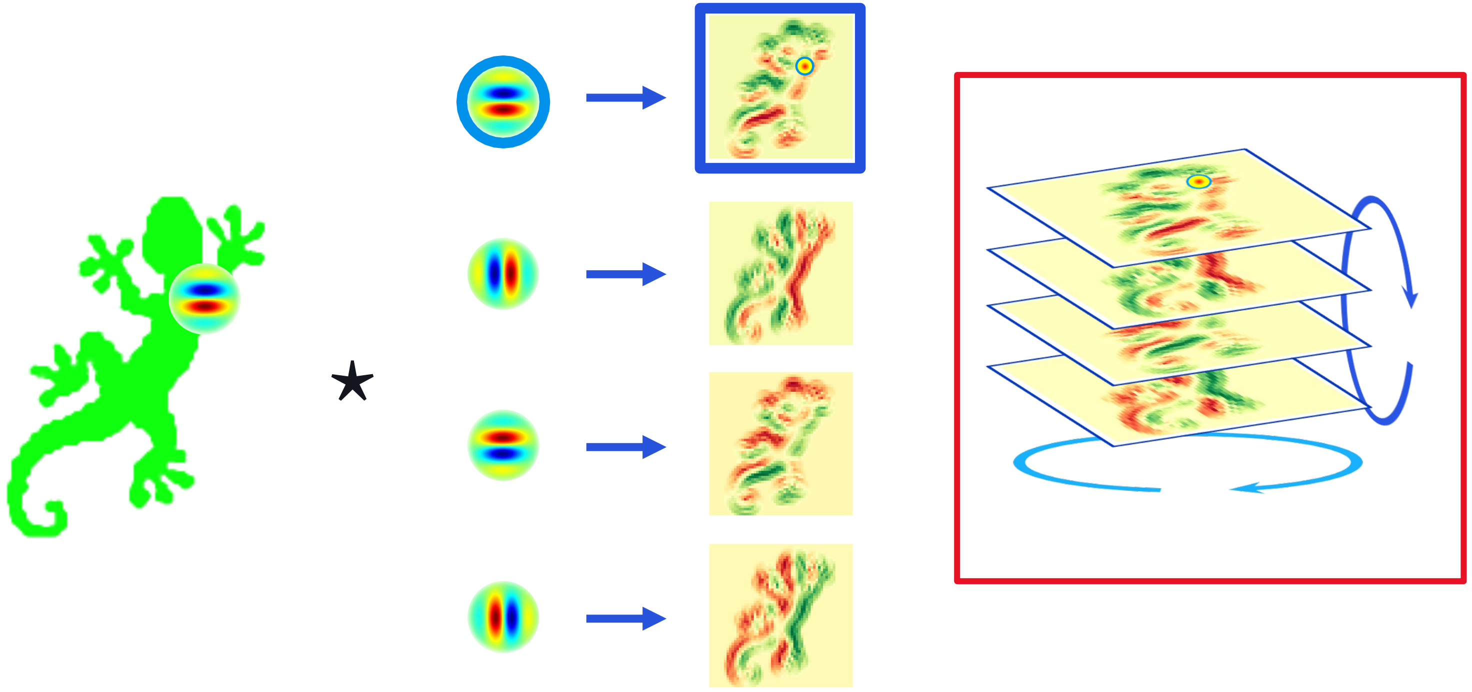
$$\rho_k(\theta) = \begin{bmatrix} \cos k\theta & -\sin k\theta \\ \sin k\theta & \cos k\theta \end{bmatrix}$$

Spherical harmonics



Steerable Filters: Output of Group Convolution

- Filter can be rotated via linear combination



$$\rho(r^0) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\rho(r^1) = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\rho(r^2) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\rho(r^3) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Agenda

Steerable Filters

Steerable Fields and Representation Theory

Group Convolution to Steerable Convolution

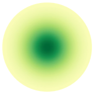
Steerable CNNs

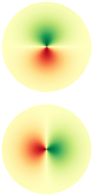
Solving the kernel constraint

Instances and Applications

Feature Fields

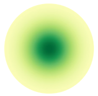
- Interpret features as a multi-channels signal $f: \mathbb{R}^n \to \mathbb{R}^d$

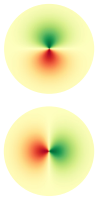
[ $\star$ ]: $\mathbb{R}^2 \to \mathbb{R}^1$

[ $\star$ ]: $\mathbb{R}^2 \to \mathbb{R}^2$

Feature Fields

- Interpret features as a multi-channels signal $f: \mathbb{R}^n \to \mathbb{R}^d$
- Signal transforms under *(point) symmetry group* G according to a *transformation law*
  - symmetry group G : e.g. rotations or reflections
  - N.B.: before we used G to also indicate translations
  - For now, we will implicitly consider equivariance to $(\mathbb{R}^n, +) \rtimes G$

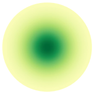
[ $\star$ ]: $\mathbb{R}^2 \to \mathbb{R}^1$

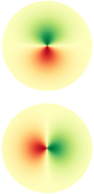
[ $\star$ ]: $\mathbb{R}^2 \to \mathbb{R}^2$

Feature Fields and Steerable CNNs

- Interpret features as a multi-channels signal $f: \mathbb{R}^n \rightarrow \mathbb{R}^d$
- Signal transforms under *(point) symmetry group* G according to a *transformation law*
  - symmetry group G : e.g. rotations or reflections
- Type of transformation identified by a **representation** of G $\quad \rho: G \rightarrow \mathbb{R}^{d \times d}$

$$[g.f](x) = \rho(g) f(g^{-1}x)$$

[ $\star$ ]: $\mathbb{R}^2 \rightarrow \mathbb{R}^1$

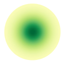
[ $\star$ ]: $\mathbb{R}^2 \rightarrow \mathbb{R}^2$

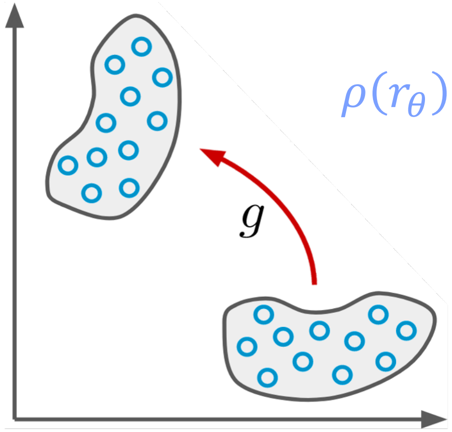
Taco S. Cohen and Max Welling, Steerable CNNs,
International Conference on Learning Representations (ICLR), 2017

Feature Fields and Steerable CNNs

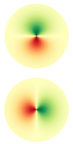
- Interpret features as a multi-channels signal $f: \mathbb{R}^n \rightarrow \mathbb{R}^d$
- Signal transforms under *(point) symmetry group* G according to a *transformation law*
  - symmetry group G : e.g. rotations or reflections
- Type of transformation identified by a **representation** of G $\quad \rho: G \rightarrow \mathbb{R}^{d \times d}$

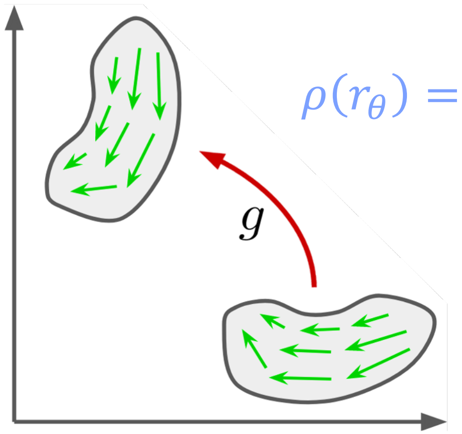
$$[g.f](x) = \rho(g)f(g^{-1}x)$$

$f = [$  $\star$  $]: \mathbb{R}^2 \rightarrow \mathbb{R}^1$



$\rho(r_\theta) = 1$

$f = [$  $\star$  $]: \mathbb{R}^2 \rightarrow \mathbb{R}^2$



$\rho(r_\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

Source: e2cnn

Definition: **Representation** of a Compact Group

A *(real)* *representation* $\rho: G \to GL(\mathbb{R}^d)$ of a *compact group* G is a map which associates to each element $g \in G$ an *invertible* $d \times d$ matrix s.t.:

- $\forall a, b \quad \rho(a)\rho(b) = \rho(ab)$
- $\forall a \ \ \rho(a)^{-1} = \rho(a^{-1})$
- $\rho(e) = \mathrm{Id}_{d \times d}$

Definition: **Representation** of a Compact Group

A *(real) representation* $\rho : G \to GL(\mathbb{R}^d)$ of a *compact group* G is a map which associates to each element $g \in G$ an *invertible* $d \times d$ matrix s.t.:

- $\forall a, b \quad \rho(a)\rho(b) = \rho(ab)$
- $\forall a \ \ \rho(a)^{-1} = \rho(a^{-1})$
- $\rho(e) = \mathrm{Id}_{d \times d}$

Can assume w.l.o.g. *orthogonal* representations, i.e. that $\rho(g)^{-1} = \rho(g)^T$

Describes the action of a group G on a *vector space* $V = \mathbb{R}^d$

E.g. representations of the rotation group $SO(2)$

- *Trivial representation* $\rho(r_\theta) = 1$
- *Standard representation* $\rho(r_\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

Agenda

Steerable Filters

Steerable Fields and Representation Theory

Group Convolution to Steerable Convolution

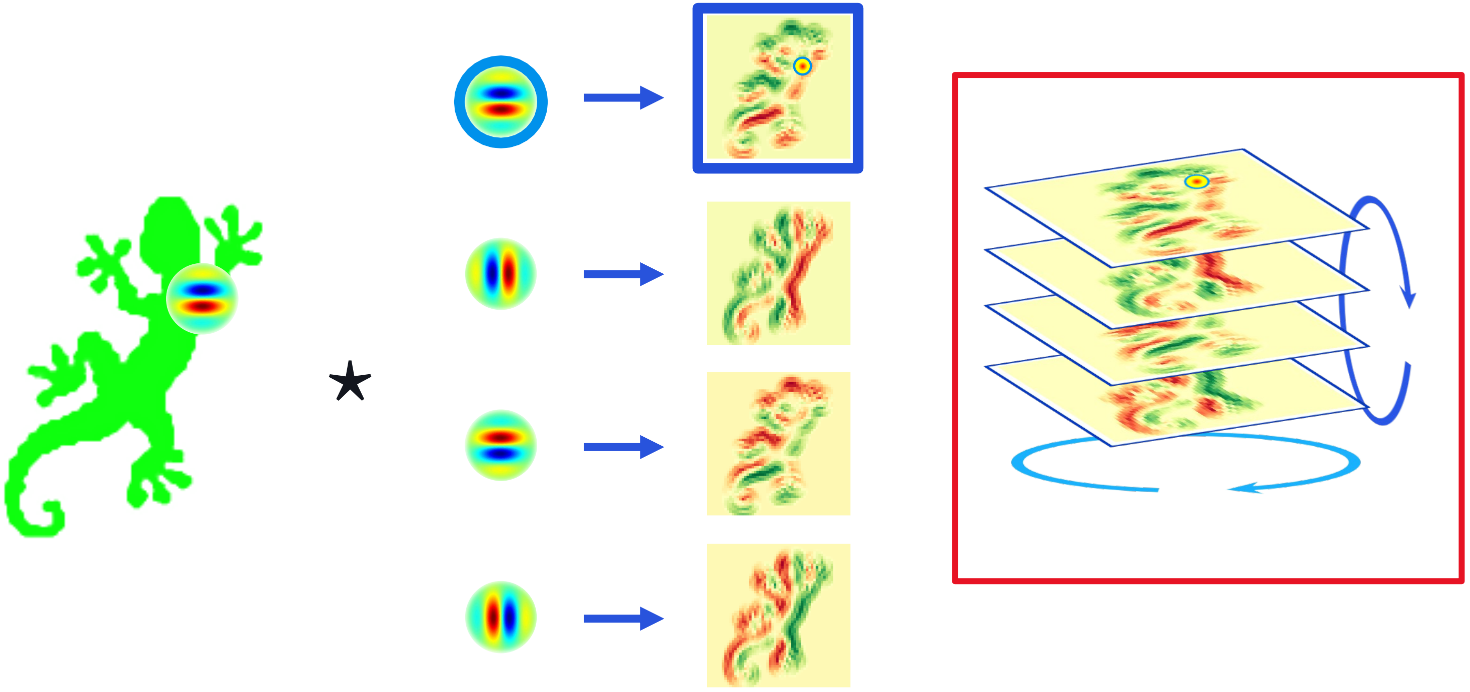
Steerable CNNs

Solving the kernel constraint

Instances and Applications

From Group Convolution to Steerable Filters

- Filter can be rotated via linear combination



$$\rho(r^0) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

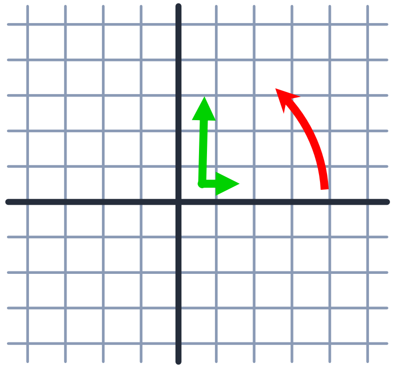
$$\rho(r^1) = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\rho(r^2) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\rho(r^3) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Semi-Direct Product group convolution

Equivariance group $(\mathbb{R}^n, +) \rtimes G$



$SE(2) = (\mathbb{R}^2, +) \rtimes SO(2)$

$$[\kappa \star f](t, g) = \int_{x \in \mathbb{R}^n} \kappa\left(g^{-1}(\boldsymbol{x} - t)\right) f(\boldsymbol{x})\, dx$$

$$[\kappa \star f](t, g) = \int_{x \in \mathbb{R}^n,\, h \in G} \kappa(g^{-1}(\boldsymbol{x} - t), g^{-1}h)\ f(x, h)\, dx$$

Semi-Direct Product group convolution

Equivariance group $(\mathbb{R}^n, +) \rtimes G$

$f: (\mathbb{R}^n, +) \rtimes G \to \mathbb{R}$

$f(t, g) \in \mathbb{R}$

Semi-Direct Product group convolution

Equivariance group $(\mathbb{R}^n, +) \rtimes G$

$f: (\mathbb{R}^n, +) \rtimes G \to \mathbb{R}$

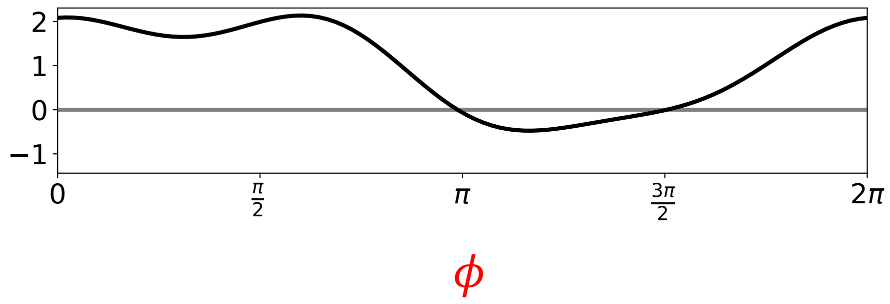
$f(t, \cdot): G \to \mathbb{R}$

Semi-Direct Product group convolution

Equivariance group $(\mathbb{R}^2, +) \rtimes SO(2)$

$f: (\mathbb{R}^2, +) \rtimes SO(2) \rightarrow \mathbb{R}$

$f(t, \cdot): [0, 2\pi) \rightarrow \mathbb{R}$



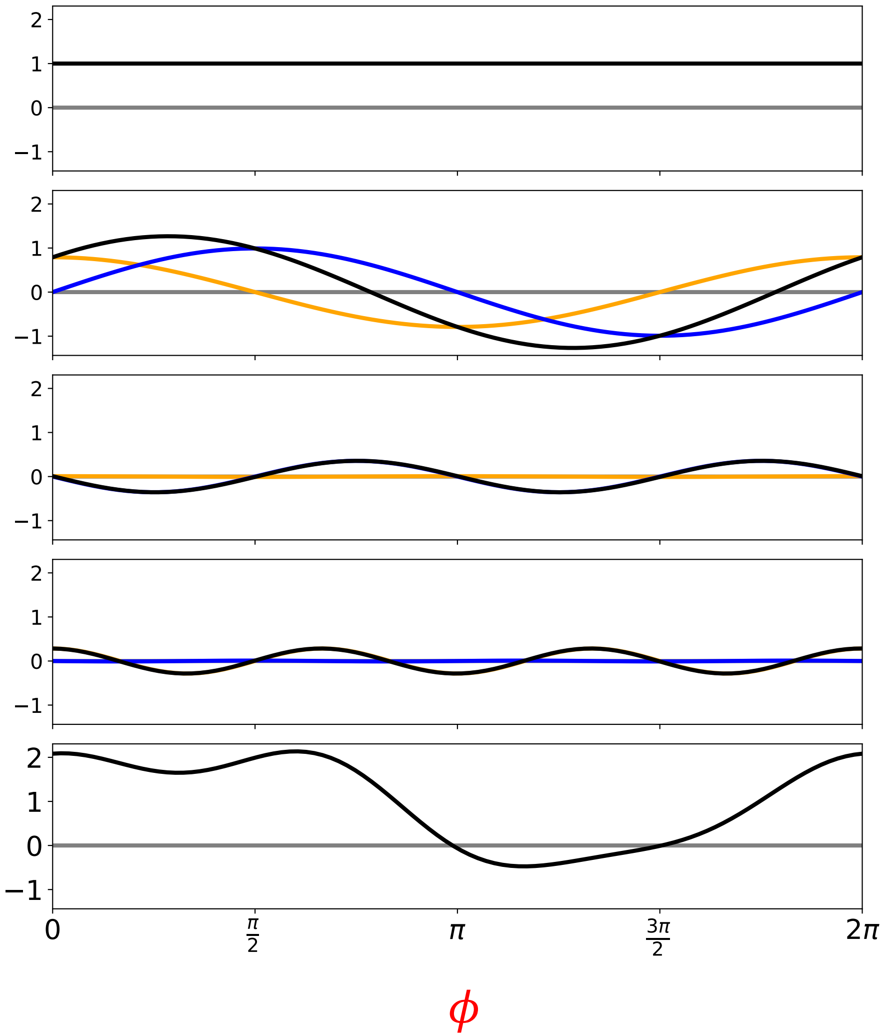
Semi-Direct Product group convolution

Equivariance group $(\mathbb{R}^2, +) \rtimes SO(2)$

$$f: (\mathbb{R}^2, +) \rtimes SO(2) \to \mathbb{R}$$

$$f(t, \cdot): [0, 2\pi) \to \mathbb{R}$$

$$f(t, r_\phi) = w_0(t) + \sum_{k=1} (\cos k\,\phi, \sin k\,\phi) \begin{pmatrix} w_{2k}(t) \\ w_{2k+1}(t) \end{pmatrix}$$



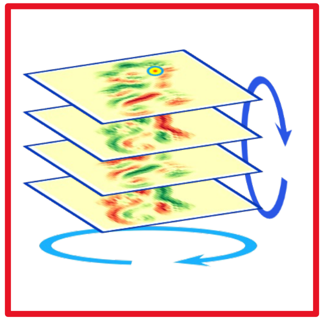
Semi-Direct Product group convolution

Equivariance group $(\mathbb{R}^2, +) \rtimes SO(2)$

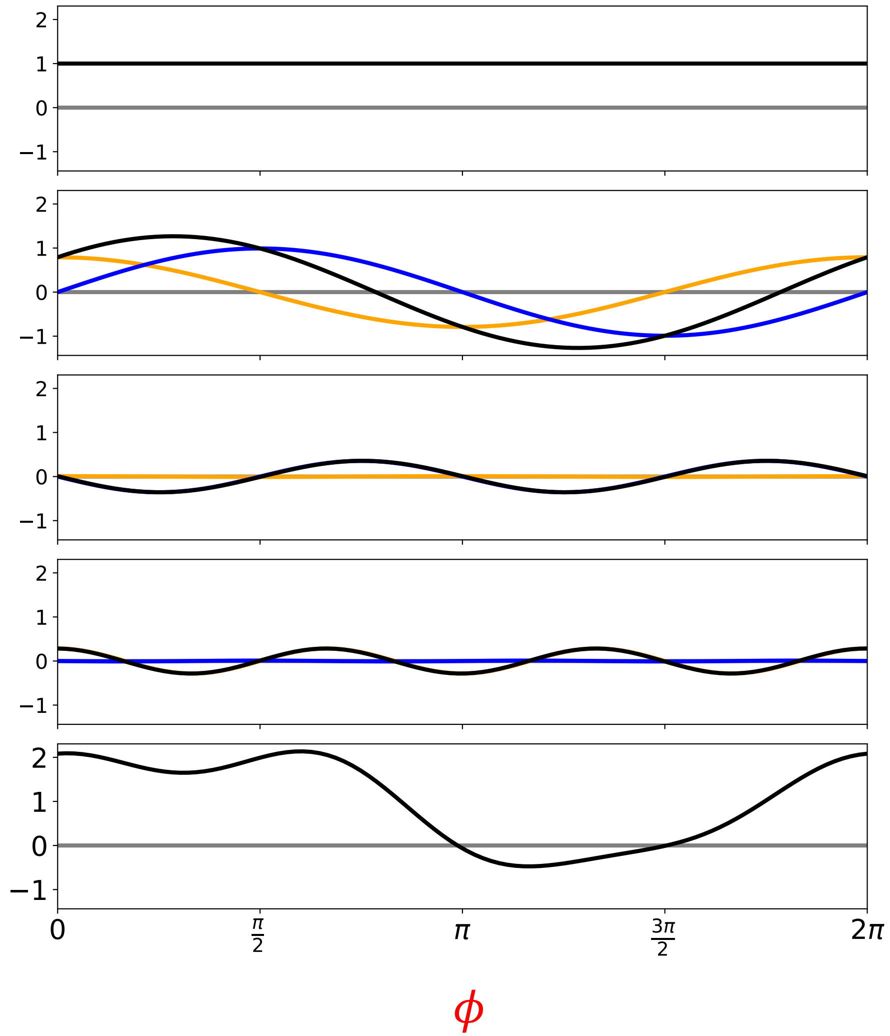
$$f: (\mathbb{R}^2, +) \rtimes SO(2) \rightarrow \mathbb{R}$$

$$f(t, \cdot): [0, 2\pi) \rightarrow \mathbb{R}$$

$$f(t, r_\phi) = w_0(t) + \sum_{k=1} (\cos k\,\phi\,, \sin k\,\phi) \begin{pmatrix} w_{2k}(t) \\ w_{2k+1}(t) \end{pmatrix}$$



$$[r_\theta . f](t, r_\phi) = f(r_\theta^{-1} t, r_\theta^{-1} r_\phi)$$



Semi-Direct Product group convolution

Equivariance group $(\mathbb{R}^2, +) \rtimes SO(2)$

$f: (\mathbb{R}^2, +) \rtimes SO(2) \to \mathbb{R}$

$f(t, \cdot): [0, 2\pi) \to \mathbb{R}$

$$[r_\theta . f](t, r_\phi) = w_0\left(r_\theta^{-1} t\right) + \sum_{k=1} \left(\cos k\,(\phi - \theta), \sin k\,(\phi - \theta)\right) \begin{pmatrix} w_{2k}\left(r_\theta^{-1} t\right) \\ w_{2k+1}\left(r_\theta^{-1} t\right) \end{pmatrix}$$

Semi-Direct Product group convolution

Equivariance group $(\mathbb{R}^2, +) \rtimes SO(2)$

$$f: (\mathbb{R}^2, +) \rtimes SO(2) \to \mathbb{R}$$

$$f(t, \cdot): [0, 2\pi) \to \mathbb{R}$$

$$[r_\theta . f](t, r_\phi) = w_0\left(r_\theta^{-1} t\right) + \sum_{k=1} (\cos k\,\phi, \sin k\,\phi) \begin{bmatrix} \cos k\theta & -\sin k\theta \\ \sin k\theta & \cos k\theta \end{bmatrix} \begin{pmatrix} w_{2k}\left(r_\theta^{-1} t\right) \\ w_{2k+1}\left(r_\theta^{-1} t\right) \end{pmatrix}$$

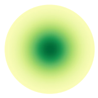
Semi-Direct Product group convolution

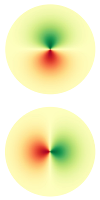
Equivariance group $(\mathbb{R}^2, +) \rtimes SO(2)$

$f : (\mathbb{R}^2, +) \rtimes SO(2) \to \mathbb{R}$

$f(t, \cdot) : [0, 2\pi) \to \mathbb{R}$

$$[r_\theta . f](t, r_\phi) = \underbrace{w_0(r_\theta^{-1} t)}\ + \sum_{k=1} (\cos k\,\phi\,, \sin k\,\phi) \underbrace{\begin{bmatrix} \cos k\theta & -\sin k\theta \\ \sin k\theta & \cos k\theta \end{bmatrix} \begin{pmatrix} w_{2k}(r_\theta^{-1} t) \\ w_{2k+1}(r_\theta^{-1} t) \end{pmatrix}}$$

[ $\star$ ]: $\mathbb{R}^2 \to \mathbb{R}^1$

[ $\star$ ]: $\mathbb{R}^2 \to \mathbb{R}^2$

$[g . f](x) = \rho(g) f(g^{-1} x)$

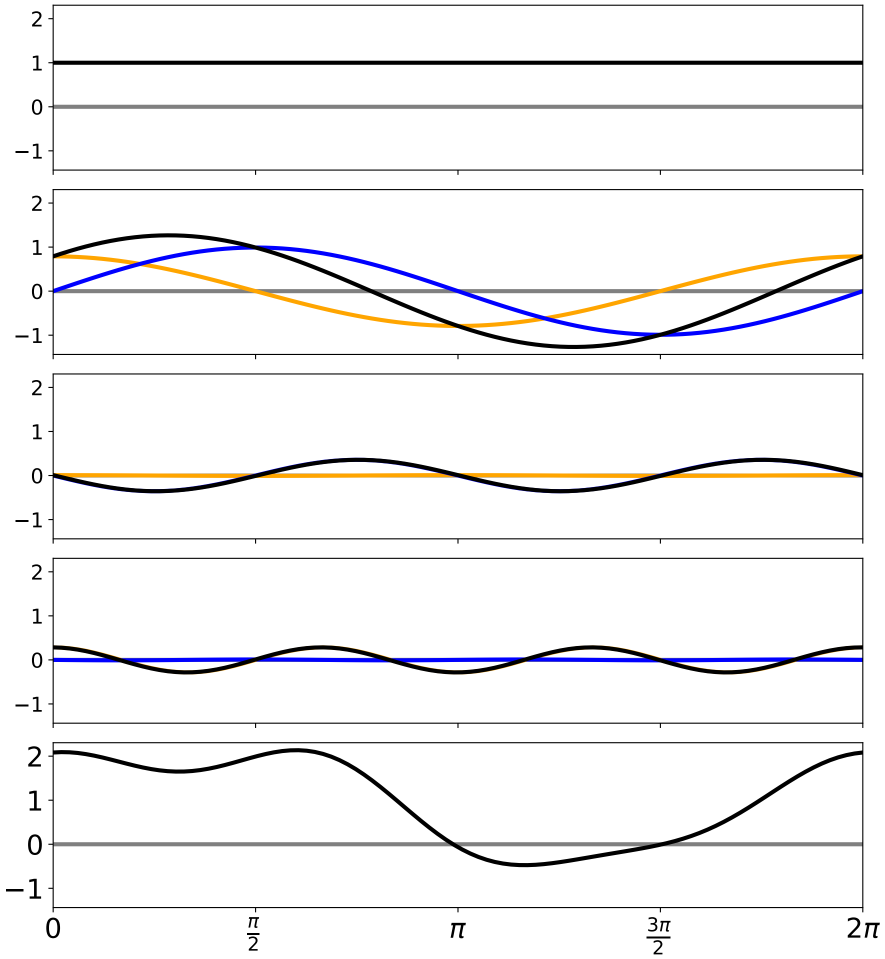
From Group Convolution to Steerable Filters

$$\rho_0(\theta) = 1 \qquad\qquad w_0(t)$$

$$\rho_1(\theta) = \begin{bmatrix} \cos 1\theta & -\sin 1\theta \\ \sin 1\theta & \cos 1\theta \end{bmatrix} \qquad\qquad \begin{pmatrix} w_1(t) \\ w_2(t) \end{pmatrix}$$

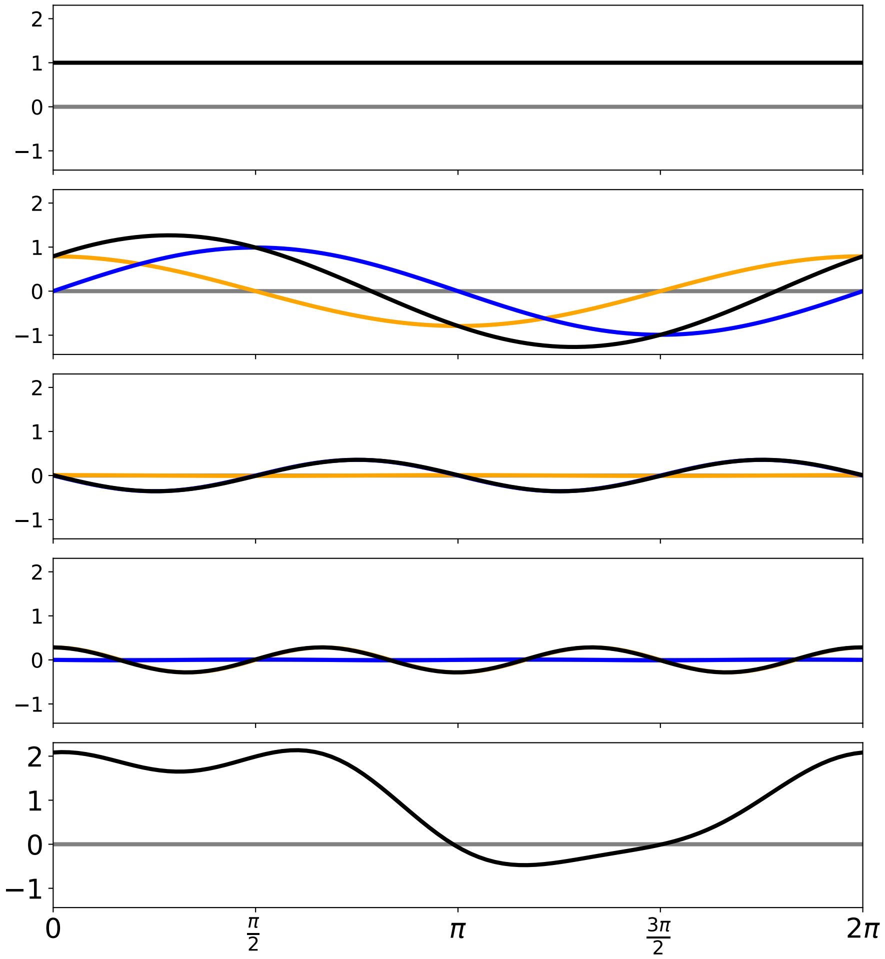
$$\rho_2(\theta) = \begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix} \qquad\qquad \begin{pmatrix} w_3(t) \\ w_4(t) \end{pmatrix}$$

$$\rho_3(\theta) = \begin{bmatrix} \cos 3\theta & -\sin 3\theta \\ \sin 3\theta & \cos 3\theta \end{bmatrix} \qquad\qquad \begin{pmatrix} w_5(t) \\ w_6(t) \end{pmatrix}$$



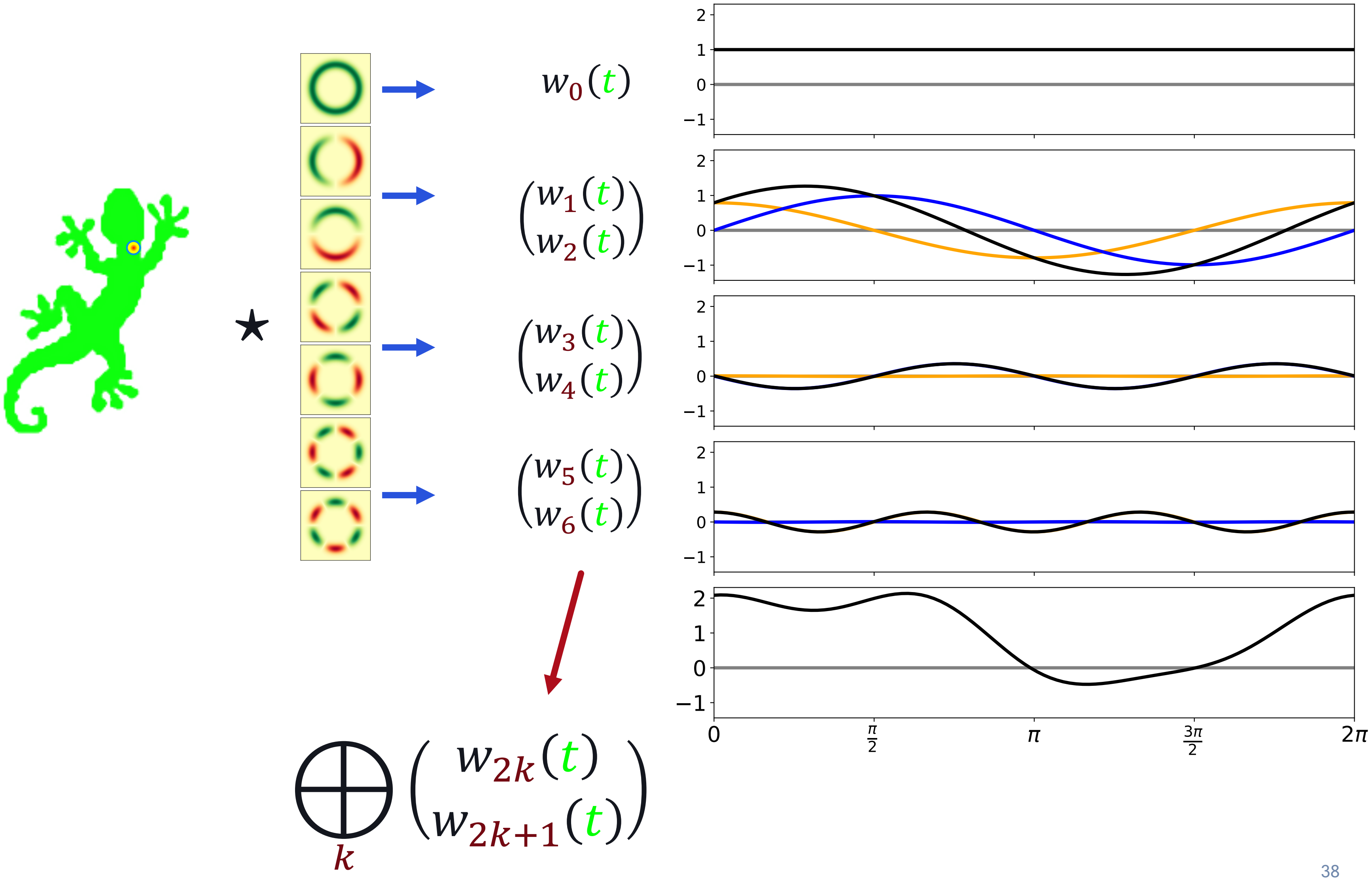
Direct Sum

$$\rho(\theta) = \begin{bmatrix} 1 & & & \\ & \begin{bmatrix} \cos 1\theta & -\sin 1\theta \\ \sin 1\theta & \cos 1\theta \end{bmatrix} & & \\ & & \begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix} & \\ & & & \begin{bmatrix} \cos 3\theta & -\sin 3\theta \\ \sin 3\theta & \cos 3\theta \end{bmatrix} \end{bmatrix} \qquad \begin{matrix} w_0(t) \\ \begin{pmatrix} w_1(t) \\ w_2(t) \end{pmatrix} \\ \begin{pmatrix} w_3(t) \\ w_4(t) \end{pmatrix} \\ \begin{pmatrix} w_5(t) \\ w_6(t) \end{pmatrix} \end{matrix}$$

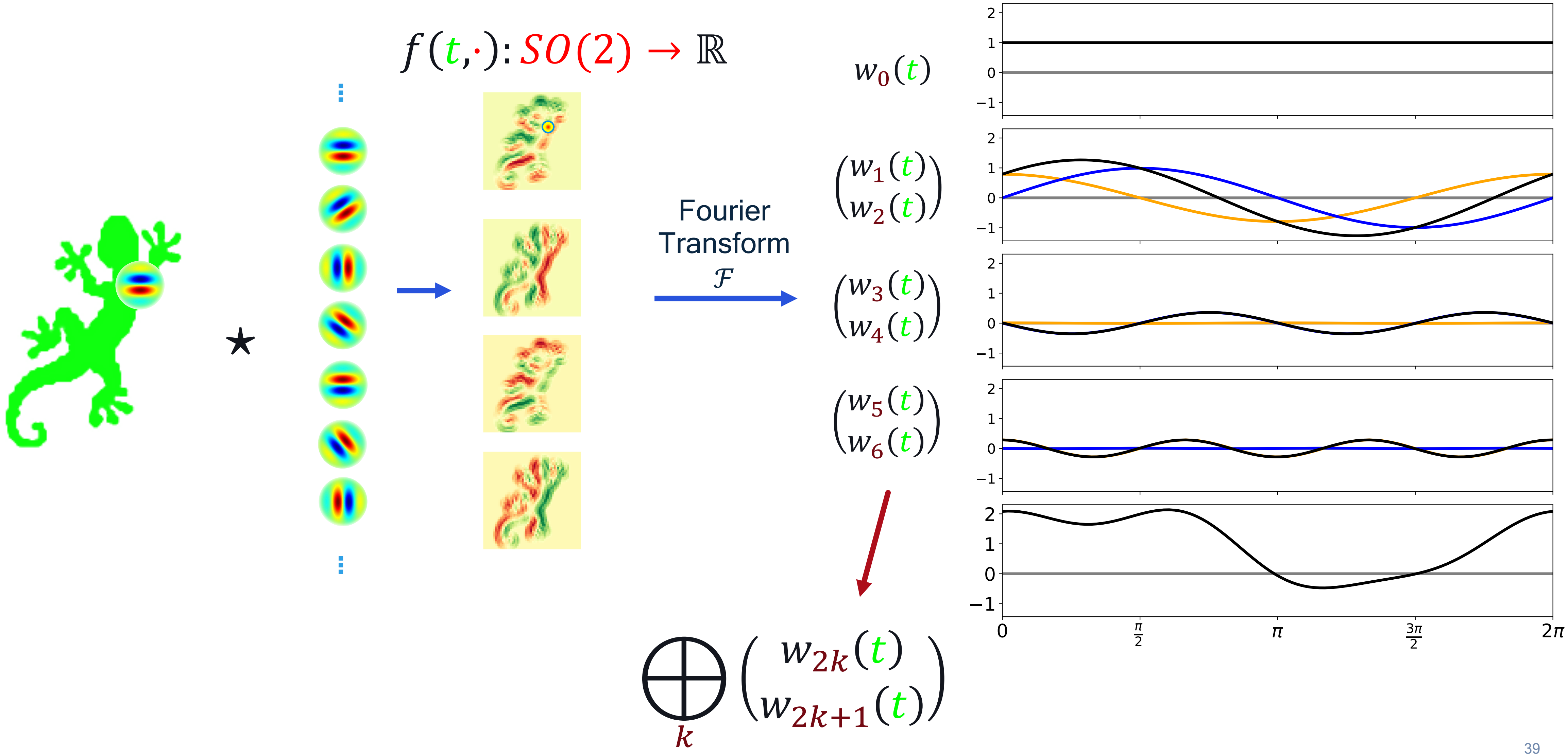


$$\rho(\theta) = \bigoplus_k \rho_k(\theta) \qquad\qquad \bigoplus_k \begin{pmatrix} w_{2k}(t) \\ w_{2k+1}(t) \end{pmatrix}$$

Recap


★
$w_0(t)$
$\begin{pmatrix} w_1(t) \\ w_2(t) \end{pmatrix}$
$\begin{pmatrix} w_3(t) \\ w_4(t) \end{pmatrix}$
$\begin{pmatrix} w_5(t) \\ w_6(t) \end{pmatrix}$
$\bigoplus_k \begin{pmatrix} w_{2k}(t) \\ w_{2k+1}(t) \end{pmatrix}$
2
1
0
−1
0
$\frac{\pi}{2}$
π
$\frac{3\pi}{2}$
2π

Recap



$f(t,\cdot): SO(2) \to \mathbb{R}$
$\star$
Fourier Transform
$\mathcal{F}$
$w_0(t)$
$\begin{pmatrix} w_1(t) \\ w_2(t) \end{pmatrix}$
$\begin{pmatrix} w_3(t) \\ w_4(t) \end{pmatrix}$
$\begin{pmatrix} w_5(t) \\ w_6(t) \end{pmatrix}$
$\bigoplus_k \begin{pmatrix} w_{2k}(t) \\ w_{2k+1}(t) \end{pmatrix}$
0
$\frac{\pi}{2}$
π
$\frac{3\pi}{2}$
2π

Formally: the Regular Representation

- Recall that $L^2(G)$ is the vector space of square integrable functions on G
- $L^2(G)$ carries an **orthogonal action** of G

$$g: L^2(G) \to L^2(G), \qquad f \mapsto g.f$$
$$[g.f](\boldsymbol{x}) := f(g^{-1}\boldsymbol{x})$$

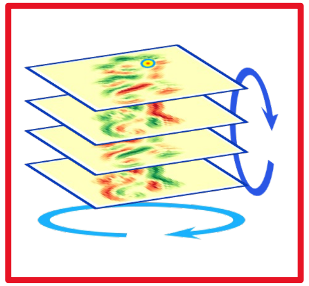
- This is the *Regular Representation* of G

Formally: the Regular Representation

- Recall that $L^2(G)$ is the vector space of square integrable functions on G
- $L^2(G)$ carries an **orthogonal action** of G

$$g: L^2(G) \to L^2(G), \qquad f \mapsto g.f$$
$$[g.f](\boldsymbol{x}) \coloneqq f(g^{-1}\,\boldsymbol{x})$$

- This is the *Regular Representation* of G
- When G is a **finite group** it looks like permutation matrices (e.g. C_4)



$$\rho(r^0) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \rho(r^1) = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \rho(r^2) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \rho(r^3) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Formally: the Fourier Transform (Peter-Weyl theorem)

- Let G be a *compact* group
- There is a set of *irreducible representations* denoted $\hat{G}$
  - analogous to frequencies in classical Fourier Transform
  - e.g. circular harmonics or Wigner-D Matrices
- The matrix coefficients form an *orthogonal basis* for $L^2(G)$

$$f(g) = \sum_{\rho \in \hat{G}} \sqrt{d_\rho} \langle \hat{f}(\rho), \rho(g) \rangle$$

$$\rho_0(\theta) = 1 \qquad \rho_1(\theta) = \begin{bmatrix} \cos 1\theta & -\sin 1\theta \\ \sin 1\theta & \cos 1\theta \end{bmatrix} \qquad \rho_2(\theta) = \begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix} \qquad \rho_3(\theta) = \begin{bmatrix} \cos 3\theta & -\sin 3\theta \\ \sin 3\theta & \cos 3\theta \end{bmatrix}$$

Formally: the Fourier Transform (Peter-Weyl theorem)

- Let G be a *compact* group
- There is a set of *irreducible representations* denoted $\hat{G}$
  - analogous to frequencies in classical Fourier Transform
  - e.g. circular harmonics or Wigner-D Matrices
- The matrix coefficients form an *orthogonal basis* for $L^2(G)$

$$f(g) = \sum_{\rho \in \hat{G}} \sqrt{d_\rho} \langle \hat{f}(\rho), \rho(g) \rangle$$

$\hat{f}(\rho) \in \mathbb{R}^{d_\rho \times d_\rho}$
Contains the weights

$$\langle A, B \rangle = \sum_{ij} A_{ij} B_{ij} = Tr(AB^T)$$

inner product between matrices

Formally: the Fourier Transform (Peter-Weyl theorem)

- Let G be a *compact* group
- There is a set of *irreducible representations* denoted $\hat{G}$
  - analogous to frequencies in classical Fourier Transform
  - e.g. circular harmonics or Wigner-D Matrices
- The matrix coefficients form an *orthogonal basis* for $L^2(G)$

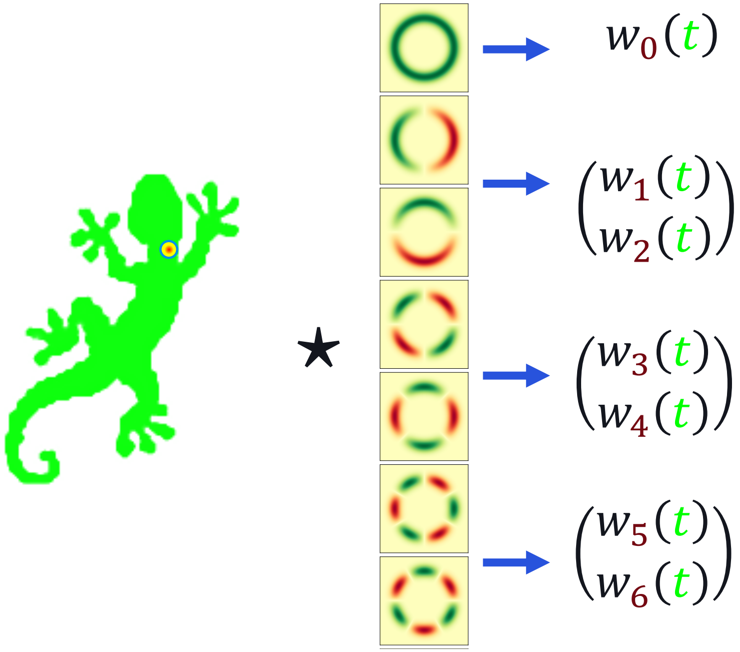
$$f(g) = \sum_{\rho \in \hat{G}} \sqrt{d_\rho} \langle \hat{f}(\rho), \rho(g) \rangle$$

Caveat: this is actually a basis only in $\mathbb{C}$, but sometimes it has redundant entries in $\mathbb{R}$, e.g.

$$\rho_1(\theta) = \begin{bmatrix} \cos 1\theta & -\sin 1\theta \\ \sin 1\theta & \cos 1\theta \end{bmatrix}$$

What did we find?

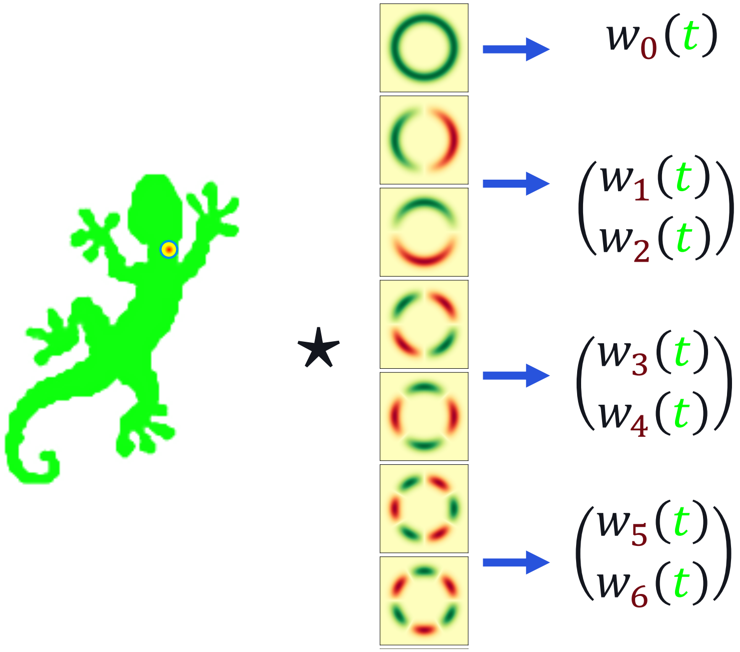
- Group convolution with steerable filters produces smaller steerable features
  - No need to store redundant activations
- So far, only studied lifting convolution $L^2(\mathbb{R}^n) \to L^2\big((\mathbb{R}^n,+) \rtimes G\big)$



$$[g.f](x) = \rho(g) f(g^{-1}x)$$

What did we find?

- Group convolution with steerable filters produces smaller steerable features
  - No need to store redundant activations
- So far, only studied lifting convolution $L^2(\mathbb{R}^n) \to L^2\big((\mathbb{R}^n,+) \rtimes G\big)$
  - Input is a **scalar field**
  - Recover lifting convolution when using output regular representation
- What about other input fields? Intermediate layers?



$$[g.f](x) = \rho(g) f(g^{-1}x)$$

Problem

Given a choice of input and output **steerable feature types**, what convolution do we need to use?

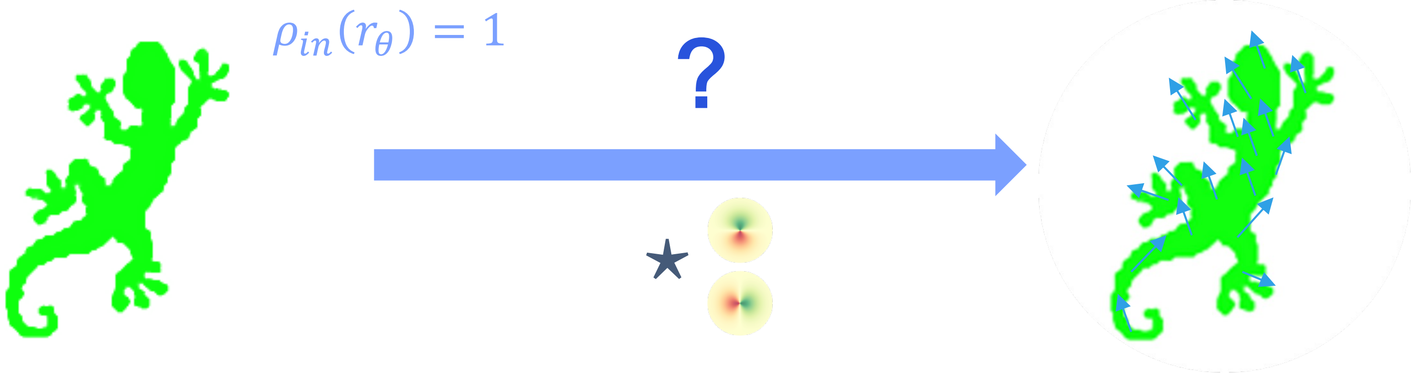
- What kind of filters produce an output feature map with the *desired transformation type*?

$f_{in}: \mathbb{R}^2 \to \mathbb{R}^1$

$f_{out}: \mathbb{R}^2 \to \mathbb{R}^2$

$\rho_{in}(r_\theta) = 1$

$\rho_{out}(r_\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$



Agenda

Steerable Filters

Steerable Fields and Representation Theory

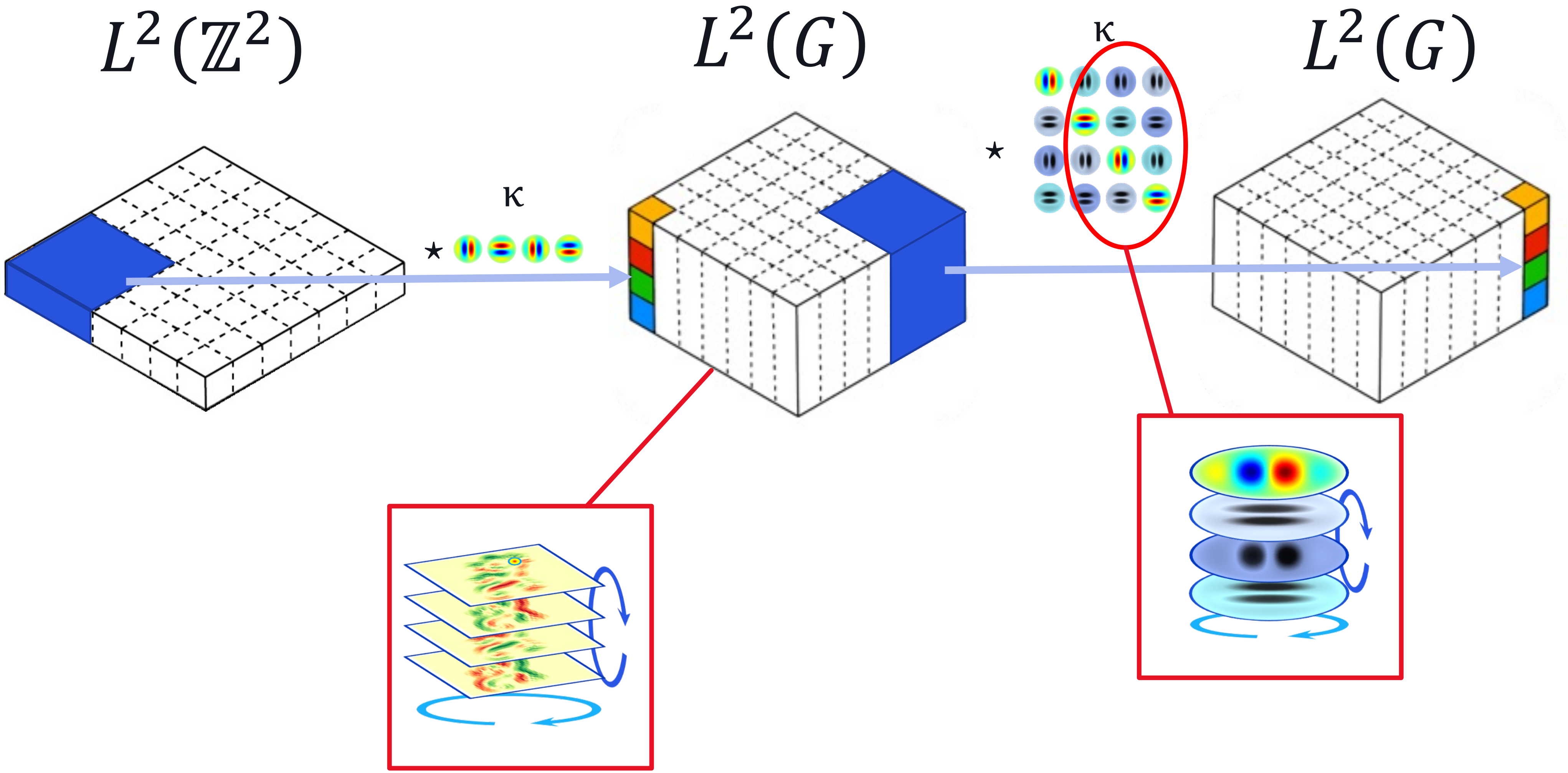
Group Convolution to Steerable Convolution

Steerable CNNs

Solving the kernel constraint

Instances and Applications

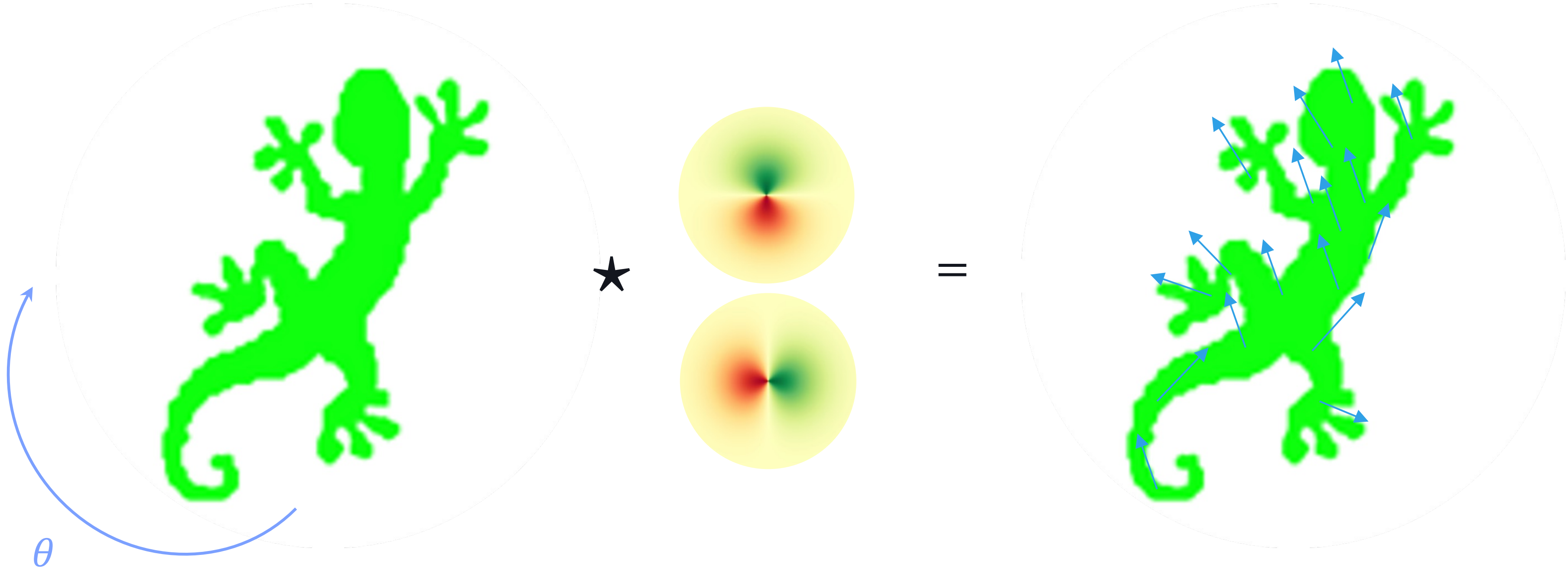
Brief Recap: Lifting Convolution and Group Convolution



Brief recap: Steerable Filters

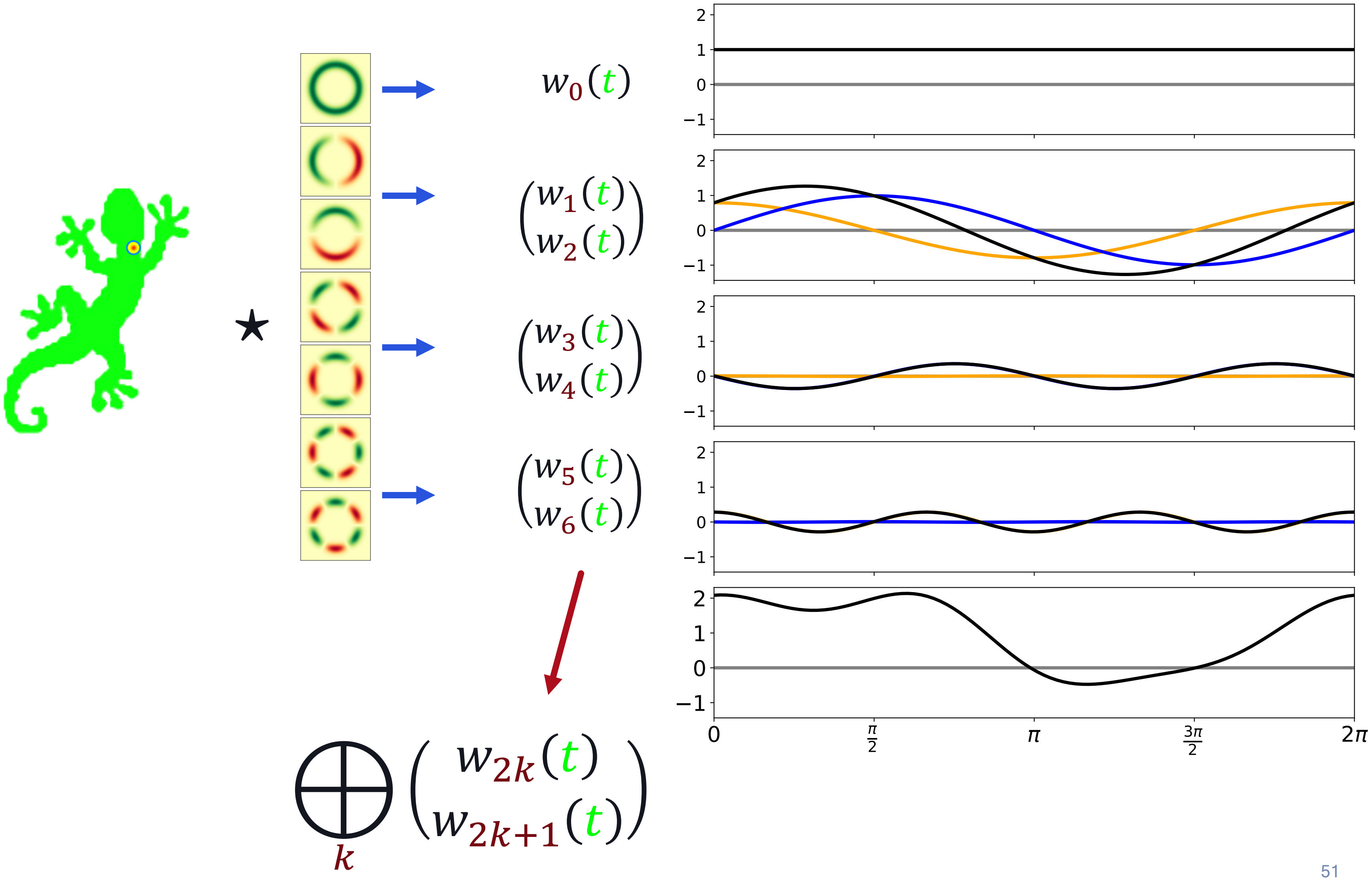
- Filter can be rotated via linear combination

$$\rho(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

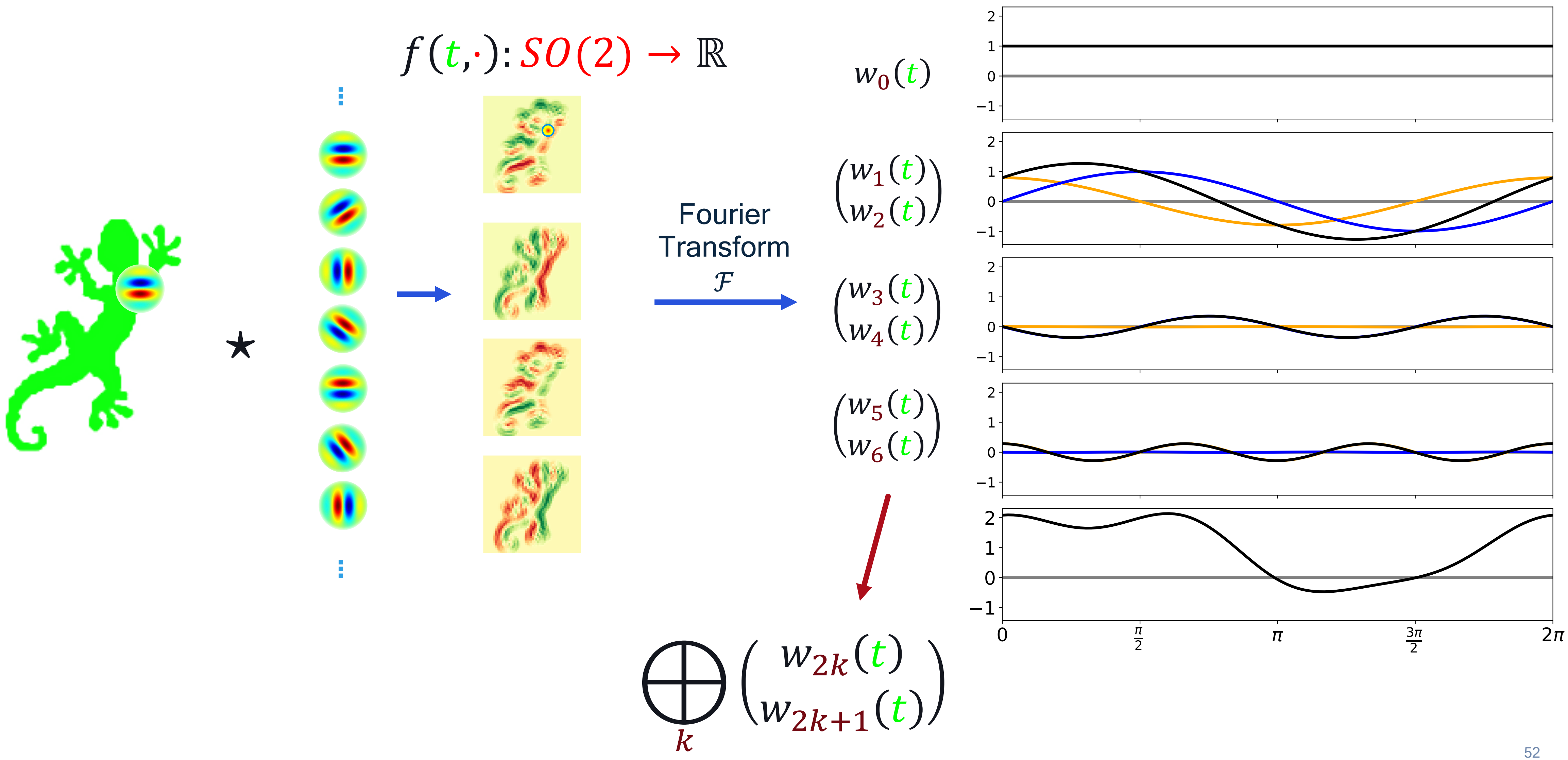


$$\boxed{R_\theta.\kappa \star f = \cos\theta\,(\kappa \star f) - \sin\theta\left(R_{\frac{\pi}{2}}.\kappa \star f\right)}$$

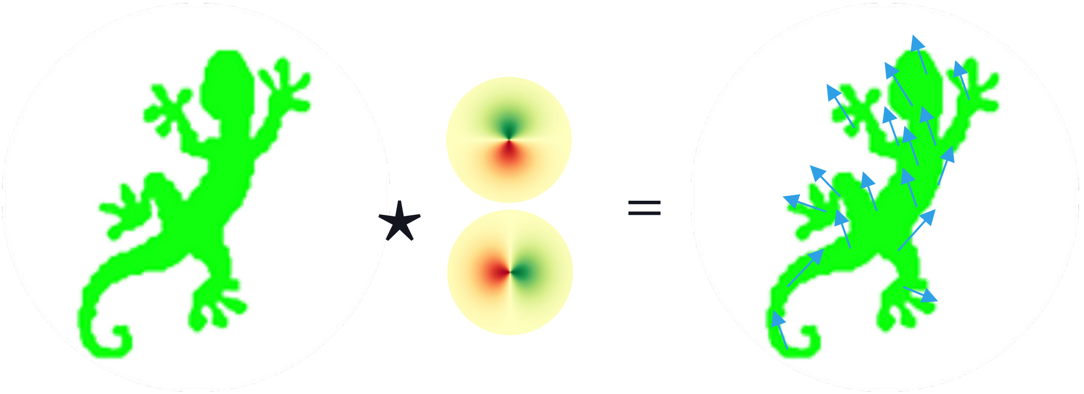
Brief Recap: steerable convolution



Recap: steerable convolution

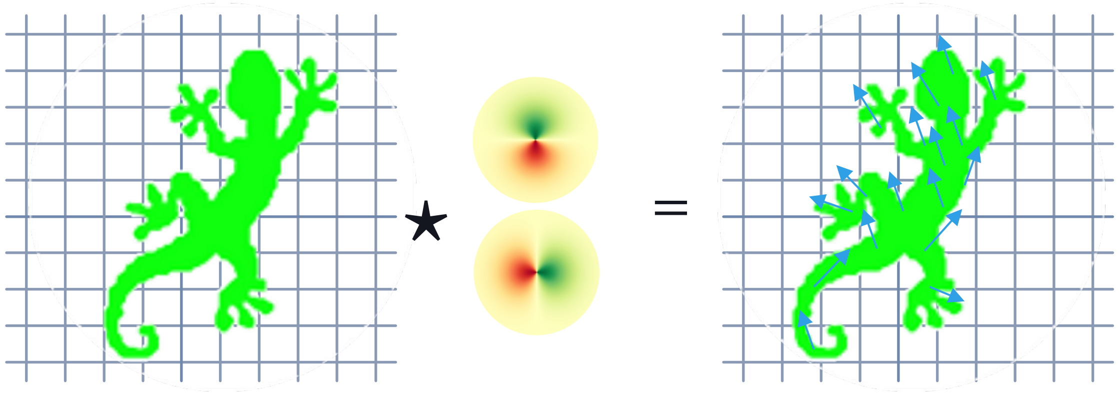


Convolution and Message Passing



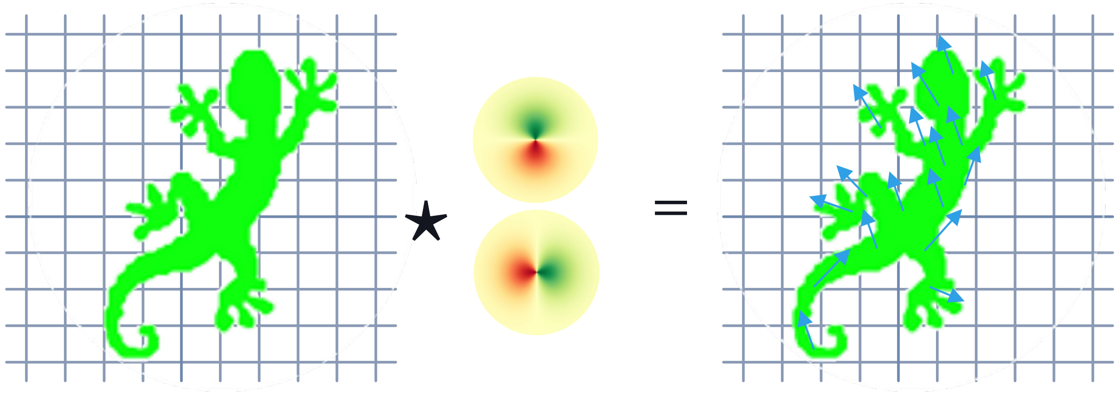
$$[\kappa \star f](y) = \sum_{x \in \mathbb{Z}^n} \kappa(x - y)\ f(x)$$

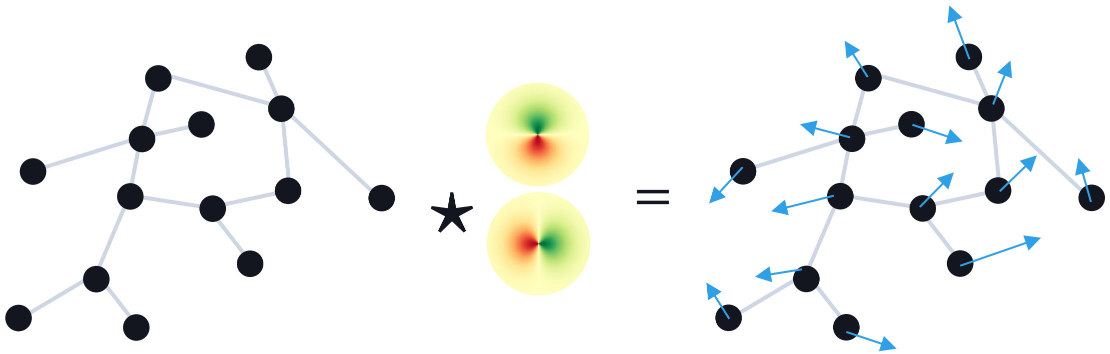
Convolution and Message Passing



$$[\kappa \star f](y) = \sum_{x \in \mathbb{Z}^n} \kappa(x - y)\ f(x)$$

Convolution and Message Passing





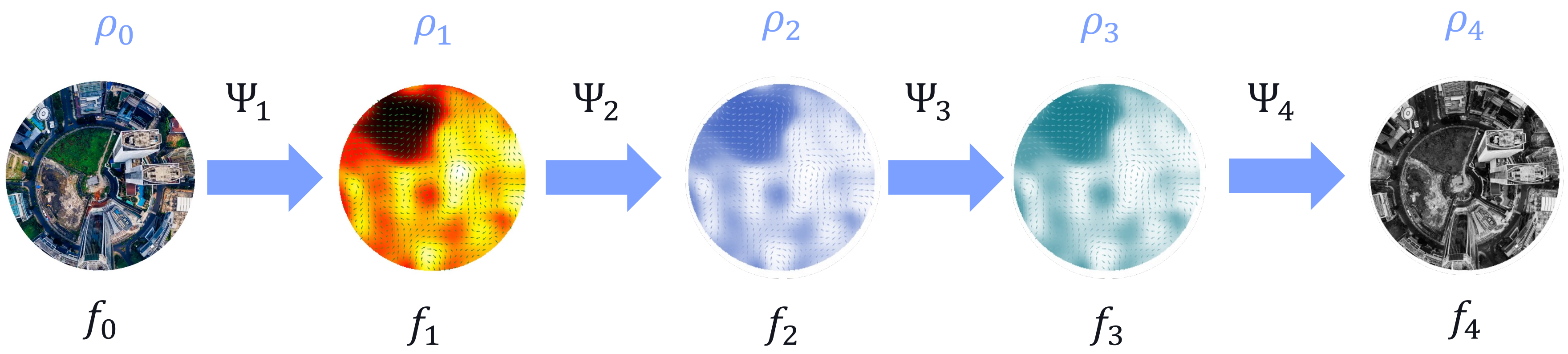
$$[\kappa \star f](y) = \sum_{x \in \mathbb{Z}^n} \kappa(x - y)\ f(x)$$

$$[\kappa \star f](i) = \sum_{j \in N_i} \kappa\left(x_i - x_j\right)\ f_j$$

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. (2018)

Feature Fields and Steerable CNNs

- Symmetry group G
- An intermediate feature is a multi-channels signal $f_l \colon \mathbb{R}^n \to \mathbb{R}^d$
- Associated with its own transformation law ρ_l
- Steerable CNN is equivariant when each layer Ψ_l commutes with its input and output transformations



Steerable CNNs

- Standard ***convolution*** with G-*steerable filter* K guarantees also G **equivariance**

$$K: \mathbb{R}^n \to \mathbb{R}^{d_{out} \times d_{in}}$$

$$K(g.x) = \rho_{out}(g) K(x) \rho_{in}(g)^T$$

Steerability Constraint

Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma and Taco Cohen,
3D Steerable CNNs: Learning Rotationally Equivariant Features in Volumetric Data,
Conference on Neural Information Processing Systems (NIPS), 2018

Taco Cohen, Mario Geiger and Maurice Weiler,
A General Theory of Equivariant CNNs on Homogeneous Spaces,
Conference on Neural Information Processing Systems (NeurIPS), 2019

Steerable CNNs

- Standard *convolution* with G *-steerable filter* K guarantees also G **equivariance**

$$K: \mathbb{R}^n \to \mathbb{R}^{d_{out} \times d_{in}}$$

$$K(g.x) = \rho_{out}(g) K(x) \rho_{in}(g)^T$$

Steerability Constraint

Q: How do we parameterize G-steerable filters?

Agenda

Steerable Filters

Steerable Fields and Representation Theory

Group Convolution to Steerable Convolution

Steerable CNNs

Solving the kernel constraint

Instances and Applications

Solving the steerability constraint

- Consider the vectorized kernel $\kappa(x) = vec(K(x))$

$$K: \mathbb{R}^n \to \mathbb{R}^{d_{out} \times d_{in}}$$

$$K(g.x) = \rho_{out}(g) K(x)\, \rho_{in}(g)^{\mathrm{T}}$$

Solving the steerability constraint

- Consider the vectorized kernel $\kappa(x) = vec(K(x))$
  - Column-wise vectorization of the matrix $K(x)$

$$K: \mathbb{R}^n \to \mathbb{R}^{d_{out} \times d_{in}}$$

$$K(g.x) = \rho_{out}(g) K(x)\, \rho_{in}(g)^{\mathrm{T}}$$

$$vec(B\ K\ A^T) = (A \otimes B) vec(K)$$

Kronecker product

$$A \otimes B = \begin{bmatrix} a_{11} \cdot B & \cdots & a_{1n} \cdot B \\ \vdots & \ddots & \vdots \\ a_{n1} \cdot B & \cdots & a_{nn} \cdot B \end{bmatrix}$$

Solving the steerability constraint

- Consider the vectorized kernel $\kappa(x) = vec(K(x))$

$$\kappa: \mathbb{R}^n \rightarrow \mathbb{R}^{d_{out} \cdot d_{in}}$$

$$\kappa(g.x) = [(\rho_{in} \otimes \rho_{out})(g)]\kappa(x)$$

Vectorized Steerability Constraint

Kronecker product

$$A \otimes B = \begin{bmatrix} a_{11} \cdot B & \cdots & a_{1n} \cdot B \\ \vdots & \ddots & \vdots \\ a_{n1} \cdot B & \cdots & a_{nn} \cdot B \end{bmatrix}$$

$$\rho_{in} \otimes \rho_{out} : G \rightarrow \mathbb{R}^{(d_{out} \cdot d_{in}) \times (d_{out} \cdot d_{in})}$$

Steerable Filters

- Consider the vectorized kernel $\kappa(x) = vec(K(x))$
- Interpret filters as **linear map** $L^2(\mathbb{R}^n) \to \mathbb{R}^{d_{out} \cdot d_{in}}$

$$f \mapsto [\kappa \star f](y) = \int_{\mathbb{R}^n} \kappa(x) f(x+y)\, dx$$

Steerable Filters

- Consider the vectorized kernel $\kappa(x) = vec(K(x))$
- Interpret filters as **linear map** $L^2(\mathbb{R}^n) \to \mathbb{R}^{d_{out} \cdot d_{in}}$

$$f \mapsto [\kappa \star f](y) = \int_{\mathbb{R}^n} \kappa(x) f(x+y)\, dx$$

Inner product of functions in $L^2(\mathbb{R}^n)$

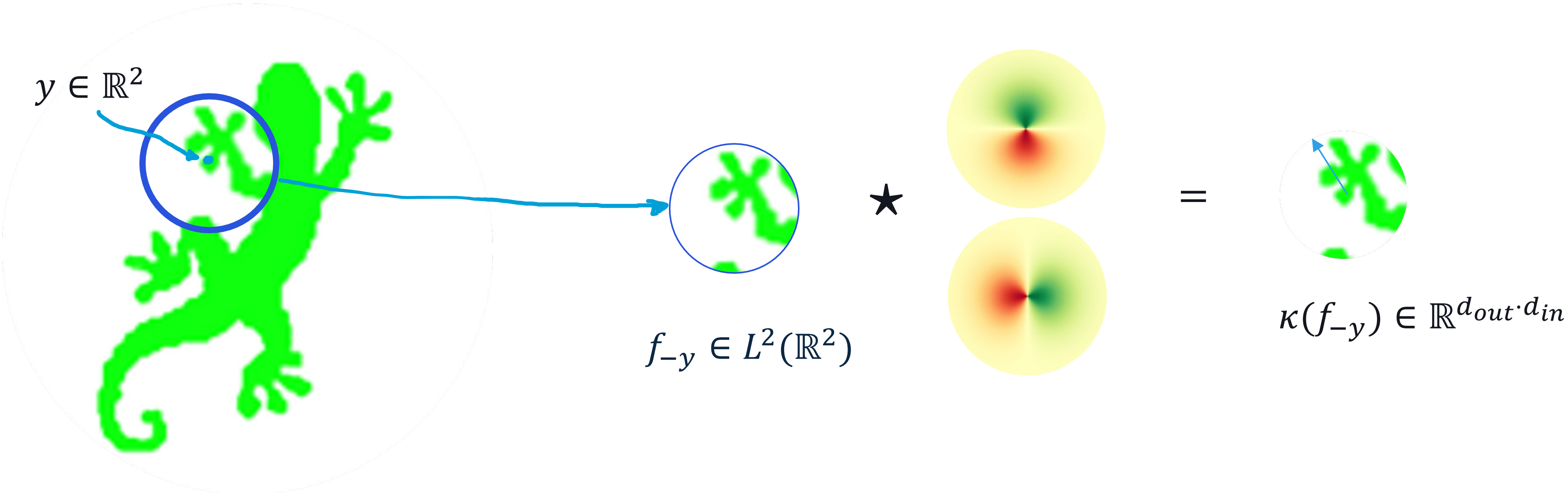
$$= \int_{\mathbb{R}^n} \kappa(x) f_{-y}(x)\, dx$$

Translated function by $-y$
$f_{-y}(x) := f(x+y)$

Steerable Filters

- Consider the vectorized kernel $\kappa(x) = vec(K(x))$
- Interpret filters as **linear map** $\kappa: L^2(\mathbb{R}^n) \to \mathbb{R}^{d_{out} \cdot d_{in}}$

$$\kappa: \; f_{-y} \mapsto \kappa\big(f_{-y}\big) = \; [\kappa \star f](y)$$



Solving the steerability constraint

- Consider the vectorized kernel $\kappa(x) = vec(K(x))$
- Interpret filters as **linear map** $\kappa: L^2(\mathbb{R}^n) \to \mathbb{R}^{d_{out} \cdot d_{in}}$

$$\kappa(g.x) = [(\rho_{in} \otimes \rho_{out})(g)]\kappa(x)$$

Action of G on $L^2(\mathbb{R}^n)$

Action of G on $\mathbb{R}^{d_{out} \cdot d_{in}}$

Steerable Filters

- Consider the vectorized kernel $\kappa(x) = vec(K(x))$
- Interpret filters as linear map $\kappa: L^2(\mathbb{R}^n) \to \mathbb{R}^{d_{out} \cdot d_{in}}$

$$\int_{\mathbb{R}^n} \kappa(g.x) f(x+y)\, dx = \int_{\mathbb{R}^n} \kappa(x) \underbrace{[g.f_{-y}]}(x)\, dx = [\kappa \star g.f](g.y)$$

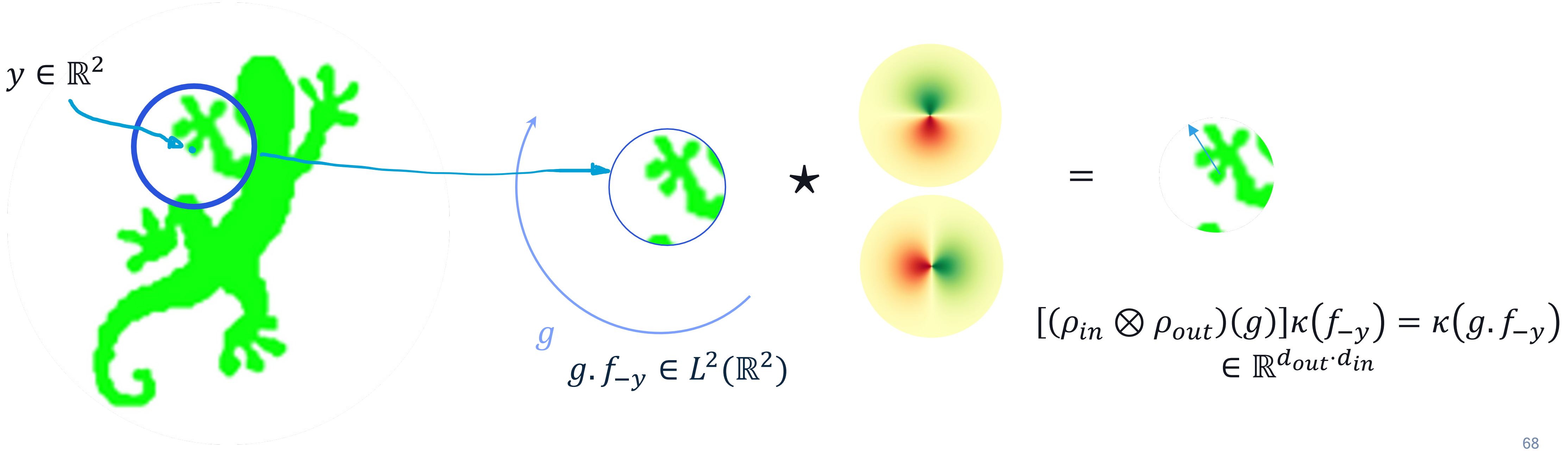
Function translated by $-y$ and rotated by g

$g.f_{-y}(x) \coloneqq f(g^{-1}.x + y)$

Solving the steerability constraint

- Consider the vectorized kernel $\kappa(x) = vec(K(x))$
- Interpret filters as **linear map** $\kappa: L^2(\mathbb{R}^n) \to \mathbb{R}^{d_{out} \cdot d_{in}}$

$$\kappa(g.x) = [(\rho_{in} \otimes \rho_{out})(g)]\kappa(x)$$



Solving the steerability constraint

- Consider the vectorized kernel $\kappa(x) = vec(K(x))$
- Interpret filters as **linear map** $\kappa\colon L^2(\mathbb{R}^n) \to \mathbb{R}^{d_{out} \cdot d_{in}}$
- IDEA: simply *find intertwiners* between two *(orthogonal)* representations!

$$\kappa(g.x) = [(\rho_{in} \otimes \rho_{out})(g)]\kappa(x)$$

Action of G on $L^2(\mathbb{R}^n)$

Action of G on $\mathbb{R}^{d_{out} \cdot d_{in}}$

Formally: Peter-Weyl Theorem (again)

Any *(orthogonal)* representation ρ can be decomposed in a *block-diagonal* form

- change-of-basis Q
- each block contains a copy of an *irreducible representation* ρ_j

$$\rho(g) = Q^T \underbrace{\begin{pmatrix} \rho_{j_1}(g) & & \\ & \rho_{j_2}(g) & \\ & & \ddots \end{pmatrix}}_{\bigoplus_{j \in I} \rho_j(g)} Q$$

Irreducible Representations (irreps): special set of representations $\hat{G} = \{\rho_j\}_j$ of G

Formally: Schur's Lemma

The space of matrices (*intertwiners*) commuting with two *irreps* $\rho_i, \rho_j \in \hat{G}$

1. is a vector space

2. is empty unless ρ_i and ρ_j are *isomorphic*, i.e. they only differ by a change of basis:

$$\rho_i(g) = Q\,\rho_j(g) Q^T$$

3. If not empty, the space is 1, 2 or 4 dimensional

Solving the steerability constraint

- Consider the vectorized kernel $\kappa(x) = vec(K(x))$
- Interpret filters as **linear map** $\kappa: L^2(\mathbb{R}^n) \rightarrow \mathbb{R}^{d_{out} \cdot d_{in}}$
- IDEA: simply *find intertwiners* between two *(orthogonal)* representations!

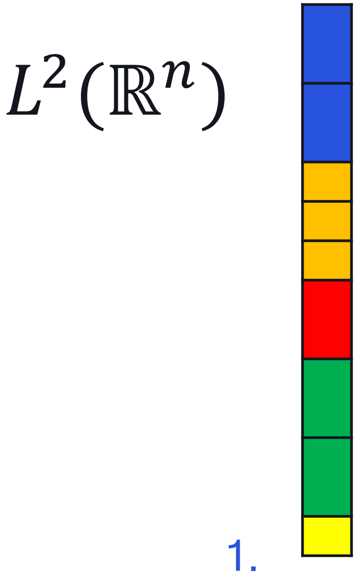
$$\kappa(g.x) = [(\rho_{in} \otimes \rho_{out})(g)]\kappa(x)$$

Action of G on $L^2(\mathbb{R}^n)$

Action of G on $\mathbb{R}^{d_{out} \cdot d_{in}}$

Solving the steerability constraint

1. decompose G's action on $L^2(\mathbb{R}^n)$

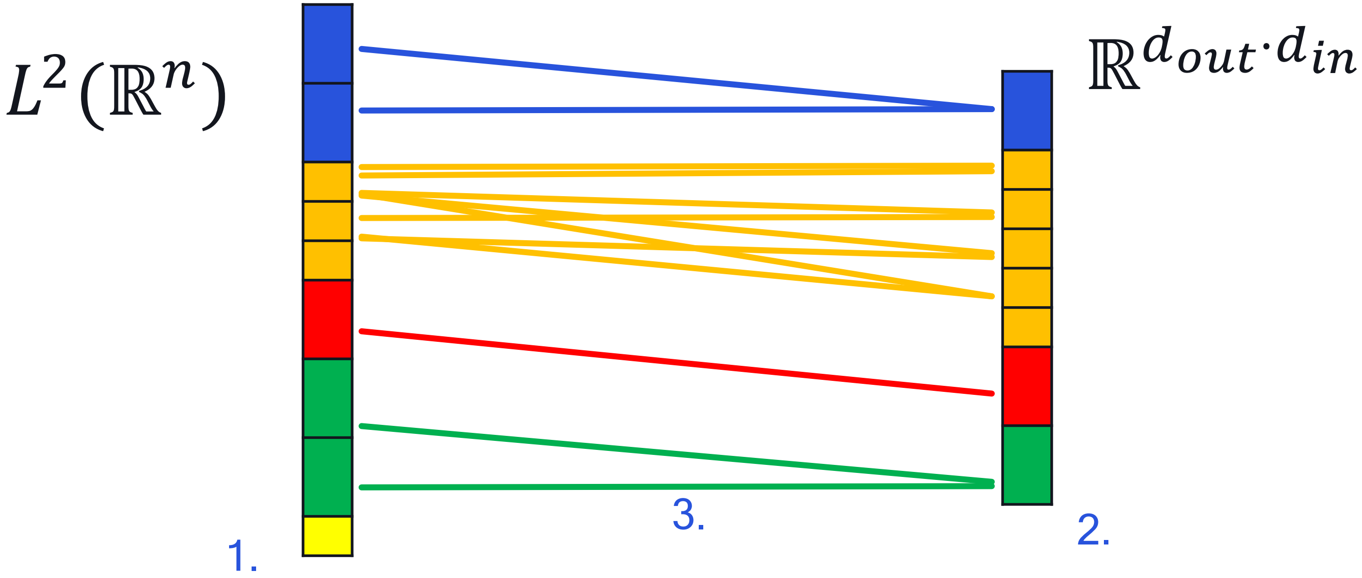


Solving the steerability constraint

1. decompose G's action on $L^2(\mathbb{R}^n)$
2. decompose G's action on $\mathbb{R}^{d_{out} \cdot d_{in}}$

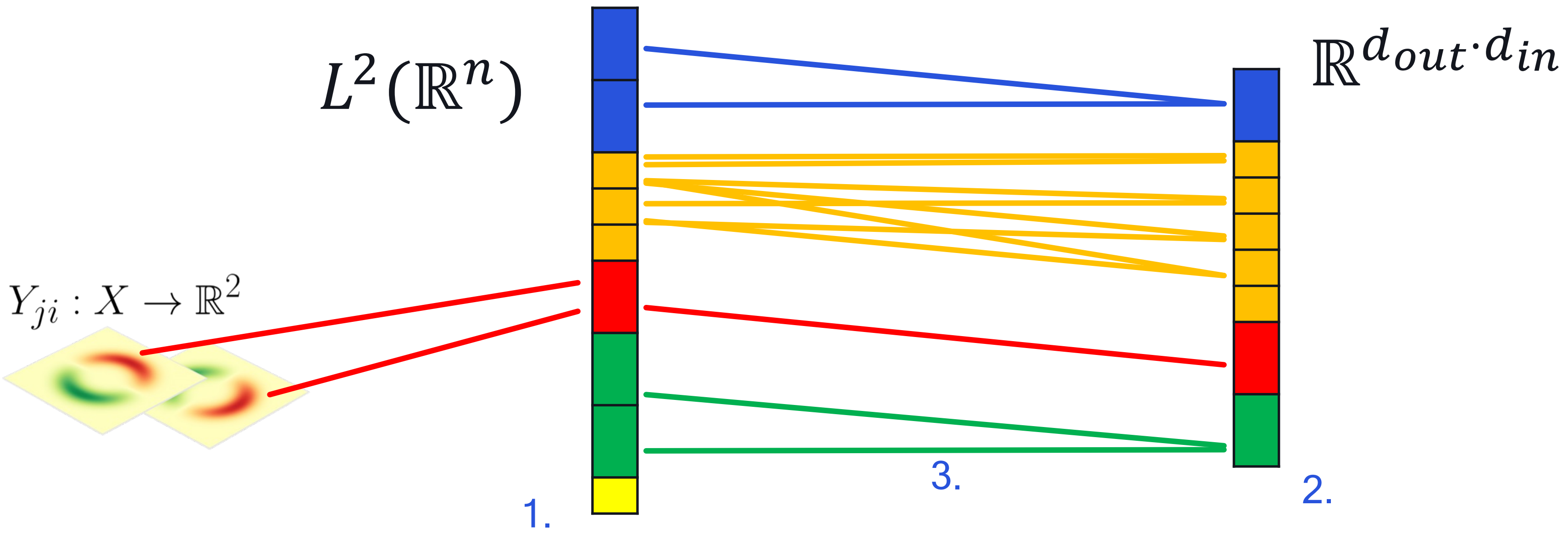
Solving the steerability constraint

1. decompose G's action on $L^2(\mathbb{R}^n)$
2. decompose G's action on $\mathbb{R}^{d_{out} \cdot d_{in}}$
3. Schur's Lemma: linear equivariant maps only between irreps of the same kind



Solving the steerability constraint

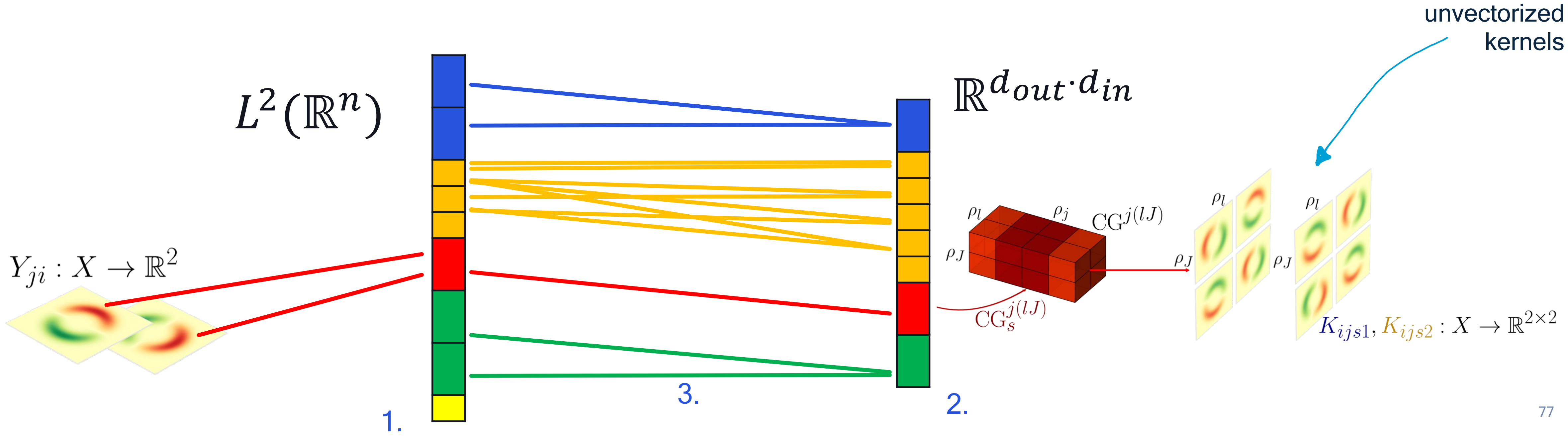
1. decompose G's action on $L^2(\mathbb{R}^n)$
   - Given by a G-steerable basis $\mathcal{B} = \{Y_{ji}: \mathbb{R}^n \to \mathbb{R}^{d_j}\}_{ji}$ for $L^2(\mathbb{R}^n)$ as in (Freeman & Adelson, 1991)
2. decompose G's action on $\mathbb{R}^{d_{out} \cdot d_{in}}$
3. Schur's Lemma: linear equivariant maps only between irreps of the same kind



William T. Freeman and Edward H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 1991

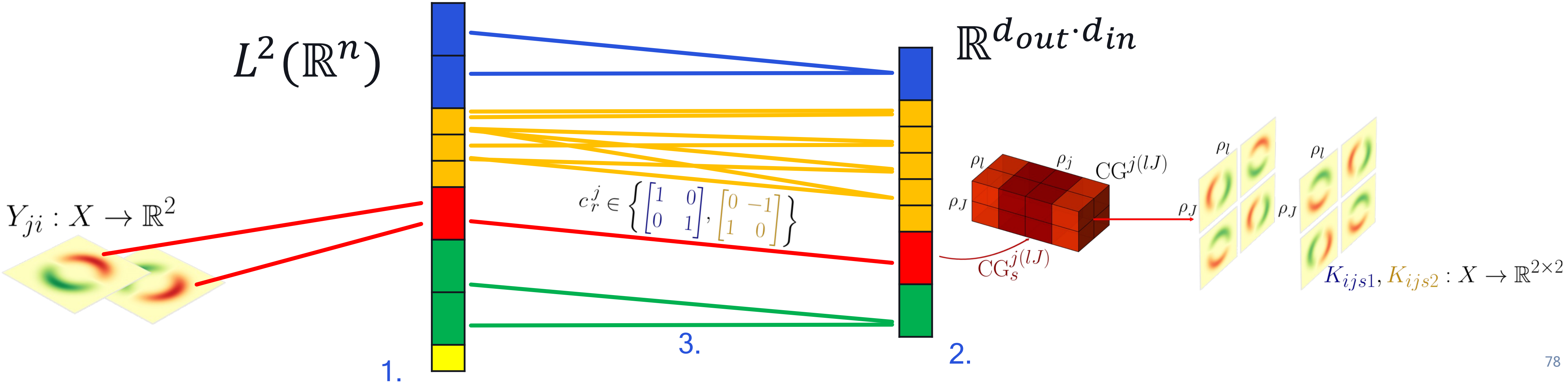
Solving the steerability constraint

1. decompose G's action on $L^2(\mathbb{R}^n)$
2. decompose G's action on $\mathbb{R}^{d_{out} \cdot d_{in}}$
   - Clebsh-Gordan coefficients CG^{lJ} to decompose $[\rho_l \otimes \rho_J](g) = [\mathrm{CG}^{lJ}]^T \left(\oplus_j \oplus_s^{[j(lJ)]} \rho_j(g) \right) \mathrm{CG}^{lJ}$
3. Schur's Lemma: linear equivariant maps only between irreps of the same kind



Solving the steerability constraint

1. decompose G's action on $L^2(\mathbb{R}^n)$
2. decompose G's action on $\mathbb{R}^{d_{out} \cdot d_{in}}$
3. Schur's Lemma: linear equivariant maps only between irreps of the same kind
   - The space of matrices M s.t. $\rho_j(g)M = M\rho_j(g)$ is spanned by a set of matrices $\left\{c_r^j\right\}_r$



Basis for G-steerable kernels

Theorem 2.1 (Basis for G-Steerable Kernels). *Let G be a compact group acting on a space X. Let $\mathcal{B} = \{Y_{ji}\}_{ji}$ be a G-steerable basis for $L^2(X)$. Assume $\rho_{\text{in}} = \rho_l$ and $\rho_{\text{out}} = \rho_J$ in $\widehat{G}$. Under minor conditions, a basis for (vectorized) G-steerable kernels over X is given by $\mathcal{K} = \{\kappa_{jisr}\}_{jisr}$, with:*

$$\kappa_{jisr}(x) = [\mathrm{CG}_s^{j(lJ)}]^T \cdot c_r^j \cdot Y_{ji}(x)$$

Decomposes $\rho_l \otimes \rho_J$
(Clebsh-Gordan coefficients)

Linear maps
equivariant to ρ_j
(Schur's Lemma)

Decomposes $L^2(\mathbb{R}^n)$
(Steerable filters)

Gabriele Cesa, Leon Lang, and Maurice Weiler.
A program to build E (N)-equivariant steerable CNNs.
International conference on learning representations. 2021.

Basis for G-steerable kernels

Theorem 2.1 (Basis for G-Steerable Kernels). *Let G be a compact group acting on a space X. Let $\mathcal{B}=\{Y_{ji}\}_{ji}$ be a G-steerable basis for $L^2(X)$. Assume $\rho_{\text{in}}=\rho_l$ and $\rho_{\text{out}}=\rho_J$ in $\widehat{G}$. Under minor conditions, a basis for (vectorized) G-steerable kernels over X is given by $\mathcal{K}=\{\kappa_{jisr}\}_{jisr}$, with:*

$$\kappa_{jisr}(x) = [\mathrm{CG}_s^{j(lJ)}]^T \cdot c_r^j \cdot Y_{ji}(x)$$

Generalize Wigner-Eckart theorem of (Lang & Weiler, 2020)
from *homogeneous spaces* to *general G-spaces*

Leon Lang and Maurice Weiler.
A Wigner-Eckart theorem for group equivariant convolution kernels.
International Conference on Learning Representations, 2020

Building G-steerable basis for $L^2(\mathbb{R}^n)$

- Building a G-steerable basis $\mathcal{B}$ for $L^2(\mathbb{R}^n)$ requires an ad-hoc solution for each G

Building G-steerable basis for $L^2(\mathbb{R}^n)$

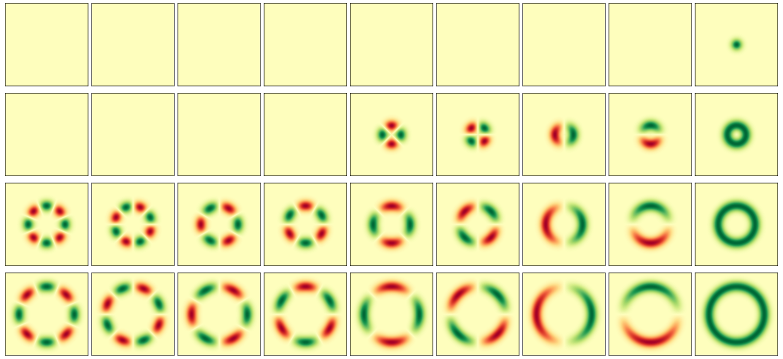
- Building a G-steerable basis $\mathcal{B}$ for $L^2(\mathbb{R}^n)$ requires an ad-hoc solution for each G
- If $G < G'$, a G'-steerable basis $\mathcal{B}' = \{Y_{j'i'}\}_{j'i'}$ is also G-steerable

Building G-steerable basis for $L^2(\mathbb{R}^n)$

- Building a G-steerable basis $\mathcal{B}$ for $L^2(\mathbb{R}^n)$ requires an ad-hoc solution for each G
- If $G < G'$, a G'-steerable basis $\mathcal{B}' = \{Y_{j'i'}\}_{j'i'}$ is also G-steerable
- Can always choose $G' = O(n)$
- e.g. $\mathcal{B}'$ contains circular/spherical harmonics combined with a radial profile

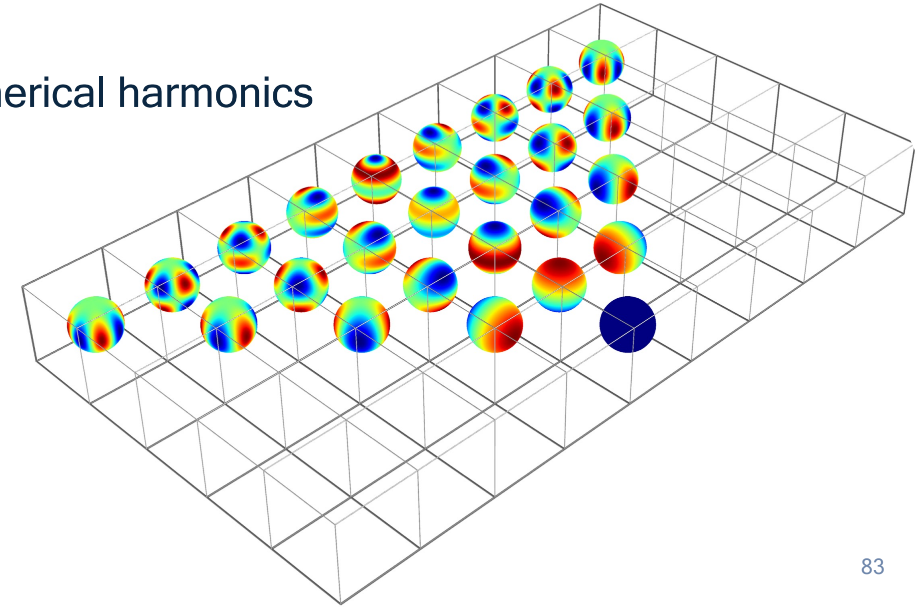
Circular harmonics

$\mathcal{B}' =$



Spherical harmonics

$\mathcal{B}' =$



Agenda

Steerable Filters

Steerable Fields and Representation Theory

Group Convolution to Steerable Convolution

Steerable CNNs

Solving the kernel constraint

Instances and Applications

Question: Equivariant Non-Linearities

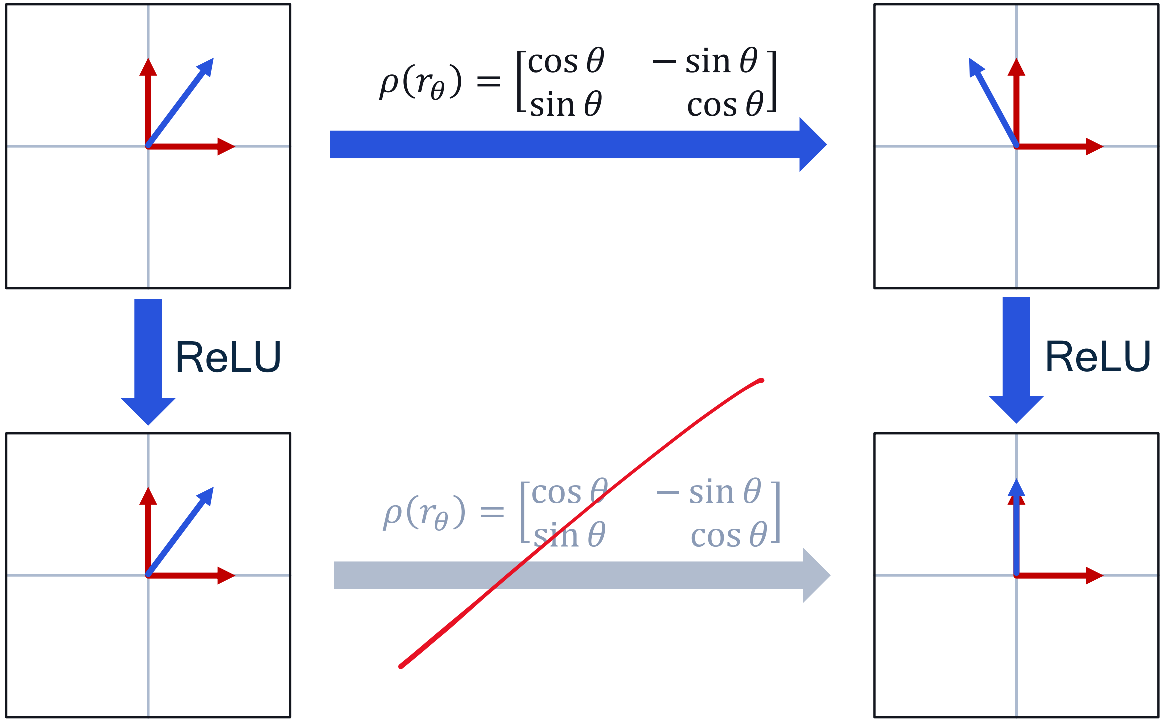
- Intermediate feature $f: \mathbb{R}^n \to \mathbb{R}^d$
- Transforms under representation of G $\quad \rho: G \to \mathbb{R}^{d \times d}$

$$[g.f](x) = \rho(g) f(g^{-1}x)$$

Equivariant Non-Linearities

- Intermediate feature $f: \mathbb{R}^n \rightarrow \mathbb{R}^d$
- Transforms under representation of G $\quad \rho: G \rightarrow \mathbb{R}^{d \times d}$
- We can NOT always use point-wise non-linearities (e.g ReLU)

$$[g.f](x) = \rho(g) f(g^{-1} x)$$



Equivariant Non-Linearities: Other Choices

- Let the intermediate feature $f: \mathbb{R}^n \to \mathbb{R}^d$ transform under $\rho: G \to O(d)$
- The quantity $|f(x)|_2^2 \in \mathbb{R}^+$ is invariant

Daniel E. Worrall, Stephan J. Garbin, Daniyar Turmukhambetov, and Gabriel J. Brostow. Harmonic networks: Deep translation and rotation equivariance. *Conference on Computer Vision and Pattern Recognition (CVPR), 2017.*

Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco S. Cohen. 3D steerable CNNs: Learning rotationally equivariant features in volumetric data. *Conference on Neural Information Processing Systems (NeurIPS), 2018*

Equivariant Non-Linearities: Other Choices

- Let the intermediate feature $f: \mathbb{R}^n \to \mathbb{R}^d$ transform under $\rho: G \to O(d)$
- The quantity $|f(x)|_2^2 \in \mathbb{R}^+$ is invariant
- **Norm Non-Linearity:** $f(x) \mapsto \sigma(|f(x)|_2^2)\, f(x)$ (Worrall et al., 2017)
- **Gated Non-Linearity:** $f(x), f_g(x) \mapsto \sigma\left(f_g(x)\right) f(x)$ (Weiler et al., 2018)
  - where $f_g(x)$ is another, invariant, feature field transforming under $\rho(g) = 1$

Potentially less expressive

- Not always universal (Finzi et al., 2021)

Daniel E. Worrall, Stephan J. Garbin, Daniyar Turmukhambetov, and Gabriel J. Brostow. Harmonic networks: Deep translation and rotation equivariance. *Conference on Computer Vision and Pattern Recognition (CVPR), 2017.*

Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco S. Cohen. 3D steerable CNNs: Learning rotationally equivariant features in volumetric data. *Conference on Neural Information Processing Systems (NeurIPS), 2018*

Marc Finzi, Max Welling, and Andrew Gordon Wilson. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. International conference on machine learning (ICML), 2021.

Equivariant Non-Linearities: Other Choices

- Let the intermediate feature $f: \mathbb{R}^n \to \mathbb{R}^d$ transform under $\rho: G \to O(d)$
- The quantity $|f(x)|_2^2 \in \mathbb{R}^+$ is **invariant**
- Can also use other quadratic invariants: (Kondor et al., 2018)
- **Tensor-Product**: $f(x) \mapsto f(x) \otimes f(x) \in \mathbb{R}^{d^2}$
  - output transforms under $\rho_{out} = \rho \otimes \rho$

$$(x, y, z) \otimes (x, y, z) = vec\begin{pmatrix} x^2 & xy & xz \\ xy & y^2 & yz \\ xz & yz & z^2 \end{pmatrix}$$

Kronecker product

$$\boldsymbol{a} \otimes \boldsymbol{b} = vec(\boldsymbol{a}\,\boldsymbol{b}^T)$$

$$A \otimes B = \begin{bmatrix} a_{11} \cdot B & \cdots & a_{1n} \cdot B \\ \vdots & \ddots & \vdots \\ a_{n1} \cdot B & \cdots & a_{nn} \cdot B \end{bmatrix}$$

Risi Kondor, Zhen Lin, and Shubhendu Trivedi. Clebsch-gordan nets: a fully Fourier space spherical convolutional neural network. Advances in Neural Information Processing Systems (NeurIPS), 2018.
Nadav Dym, and Haggai Maron. On the Universality of Rotation Equivariant Point Cloud Networks. International Conference on Learning Representations (ICML). 2020.

Equivariant Non-Linearities: Other Choices

- Let the intermediate feature $f: \mathbb{R}^n \to \mathbb{R}^d$ transform under $\rho: G \to O(d)$
- The quantity $|f(x)|_2^2 \in \mathbb{R}^+$ is **invariant**
- Can also use other quadratic invariants: (Kondor et al., 2018)
- **Tensor-Product**: $f(x) \mapsto f(x) \otimes f(x) \in \mathbb{R}^{d^2}$
  - output transforms under $\rho_{out} = \rho \otimes \rho$
  - universality (Dym and Maron, 2020)
  - More expensive

Kronecker product

$$\boldsymbol{a} \otimes \boldsymbol{b} = vec(\boldsymbol{a}\,\boldsymbol{b}^T)$$

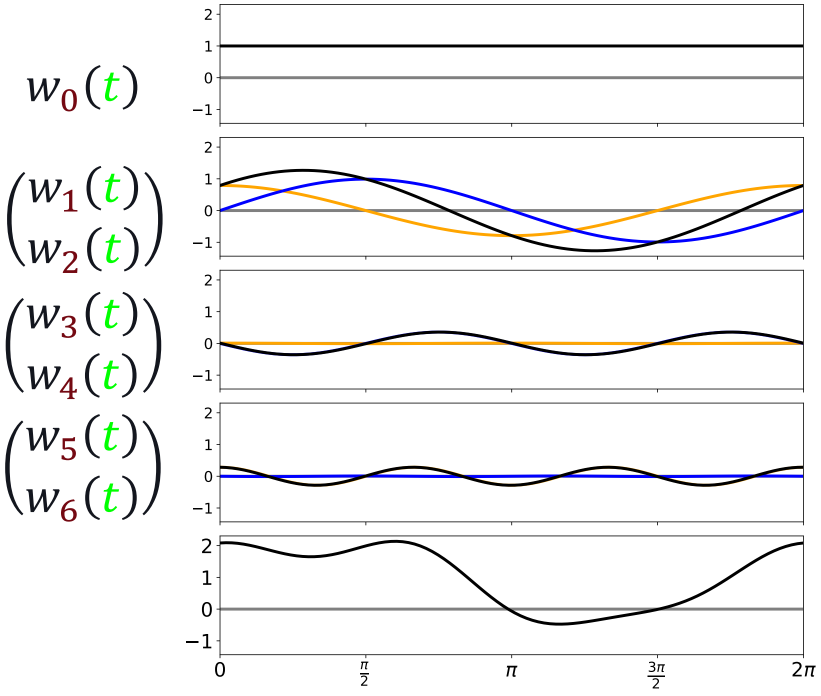
$$A \otimes B = \begin{bmatrix} a_{11} \cdot B & \cdots & a_{1n} \cdot B \\ \vdots & \ddots & \vdots \\ a_{n1} \cdot B & \cdots & a_{nn} \cdot B \end{bmatrix}$$

Risi Kondor, Zhen Lin, and Shubhendu Trivedi. Clebsch-gordan nets: a fully Fourier space spherical convolutional neural network. Advances in Neural Information Processing Systems (NeurIPS), 2018.

Nadav Dym, and Haggai Maron. On the Universality of Rotation Equivariant Point Cloud Networks. International Conference on Learning Representations (ICML). 2020.

Equivariant Non-Linearities: Fourier Transform based

- Imitate a GCNN
- Choose a **band-limited** subset of irreps $\{\rho_i\}_{i\in I} \subset \hat{G}$
- A feature vector $f(x) \in \mathbb{R}^d$ represents a bandlimited signal in $L^2(G)$
- Apply point-wise non-linearity σ (e.g. ReLU) by:
  - Sampling the signal $f(x)$ on a finite subset $\mathcal{G} \subset G$
  - Applying σ on each sample
  - Reconstruct a band-limited signal from the samples



Equivariant Non-Linearities: Fourier Transform based

- Imitate a GCNN
- Choose a **band-limited** subset of irreps $\{\rho_i\}_{i\in \mathrm{I}} \subset \hat{G}$
- A feature vector $f(x) \in \mathbb{R}^d$ represents a bandlimited signal in $L^2(G)$
- Apply point-wise non-linearity σ (e.g. ReLU) by:
  - Sampling the signal $f(x)$ on a finite subset $\mathcal{G} \subset G$ (discrete Inverse Fourier Transform)
  - Applying σ on each sample
  - Reconstruct a band-limited signal from the samples (discrete Fourier Transform)

- Band-limit + sufficient samples to control *reconstruction error*

Equivariant Non-Linearities: Fourier Transform based

- Imitate a GCNN
- Choose a **band-limited** subset of irreps $\{\rho_i\}_{i \in \mathrm{I}} \subset \hat{G}$
- A feature vector $f(x) \in \mathbb{R}^d$ represents a bandlimited signal in $L^2(G)$
- Apply point-wise non-linearity σ (e.g. ReLU) by:
  - Sampling the signal $f(x)$ on a finite subset $\mathcal{G} \subset G$ (discrete Inverse Fourier Transform)
  - Applying σ on each sample
  - Reconstruct a band-limited signal from the samples (discrete Fourier Transform)

- Can also consider functions on homogeneous space X rather than G for reduced complexity.

  Recall Spherical CNNs

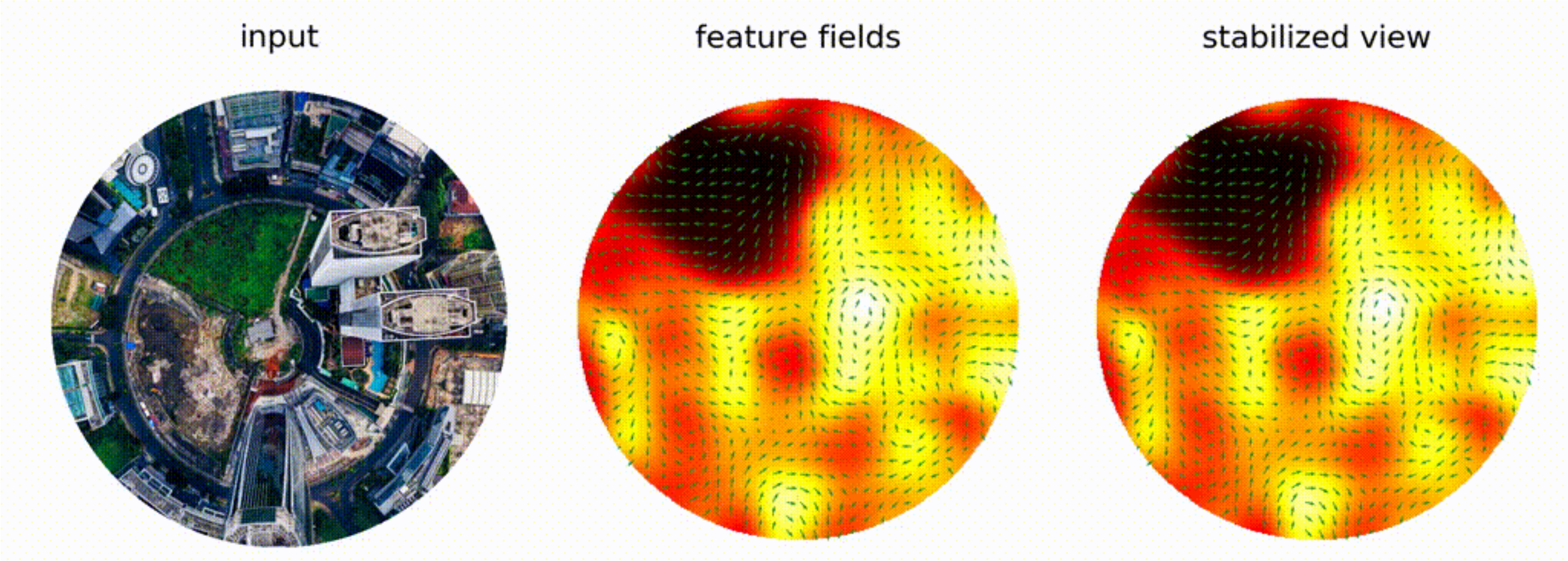
Equivariant Non-Linearities: Take Away

A useful framework to look at the literature

Most models out there differ by *choice of*

- intermediate feature types
- activation functions
- parameterization of convolution filters / messages
- discretization

Outcome: equivariance to continuous rotations



Source: e2cnn

General Program to implement G-equivariance: 2D images

2D rotational symmetries

$(\mathbb{R}^2, +) \rtimes G < E(2)$

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

MNIST Variations

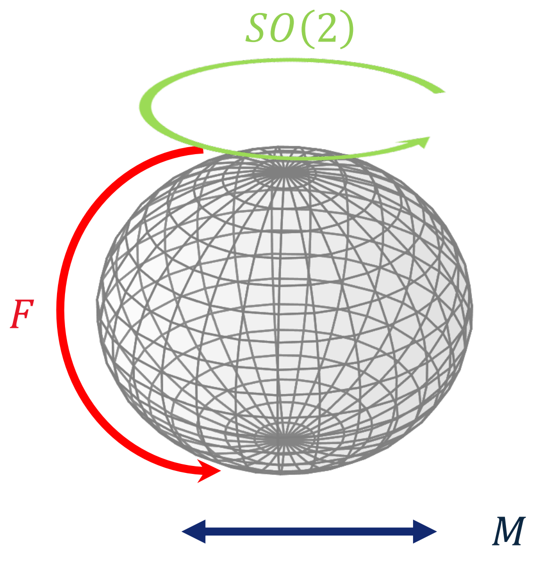
Maurice Weiler* and Gabriele Cesa*.
General E(2)-Equivariant Steerable CNNs,
Neural Information Processing Systems (NeurIPS), 2019

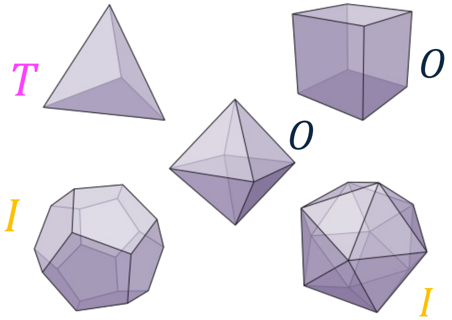
General Program to implement G-equivariance: 3D voxel data

Table 1: Rotated ModelNet10 (O(3) symmetry). $*$ indicates wider models to fix the computational cost.

G	Description	Accuracy
$\{e\}$	Conventional CNN	82.5 ± 1.4
SO(2)	Axial Symmetry	86.9 ± 1.9
$\mathrm{SO}(2) \rtimes \mathrm{F} \cong \mathrm{O}(2)$	Dihedral Symmetry	87.5 ± 0.7
$\mathrm{SO}(2) \rtimes \mathrm{M} \cong \mathrm{O}(2)$	Conical Symmetry	88.5 ± 0.8
$\mathrm{Inv} \times \mathrm{SO}(2)$	Cylindrical Symmetry	86.8 ± 0.7
$\mathrm{Inv} \times \mathrm{SO}(2) \rtimes \mathrm{F}$	Full Cylindrical Symmetry	87.0 ± 1.0
O	Octahedral Symmetry (Winkels & Cohen, 2018)	89.7 ± 0.6
I	Icosahedral Symmetry	90.0 ± 0.6
I	Icosahedral Symmetry (finite orbits basis)	88.2 ± 1.0
SO(3)	Chiral (Tensor product) (Anderson et al., 2019)	86.3 ± 1.0
SO(3)	Chiral (Gated non-linearity)(Weiler et al., 2018b)	88.8 ± 1.2
SO(3)	Chiral (Regular, $\lvert\mathcal{G}\rvert = 96$)	89.1 ± 1.2
SO(3)	Chiral (Regular, $\lvert\mathcal{G}\rvert = 192$)*	89.4 ± 1.4
SO(3)	Chiral (Quotient $\mathcal{S}^2 = \mathrm{SO}(3)/\mathrm{SO}(2)$, $\lvert\mathcal{X}\rvert = 30$)	89.5 ± 1.0
O(3)	Achiral (Regular, $\lvert\mathcal{G}\rvert = 120$)	89.2 ± 0.6
O(3)	Achiral (Regular, $\lvert\mathcal{G}\rvert = 144$)*	89.4 ± 0.7
O(3)	Achiral (Quotient $\mathrm{Inv} \times \mathcal{S}^2 = \mathrm{O}(3)/\mathrm{SO}(2)$, $\lvert\mathcal{X}\rvert = 60$)	88.6 ± 0.9

3D rotational symmetries





Axial rotational symmetries in 3D

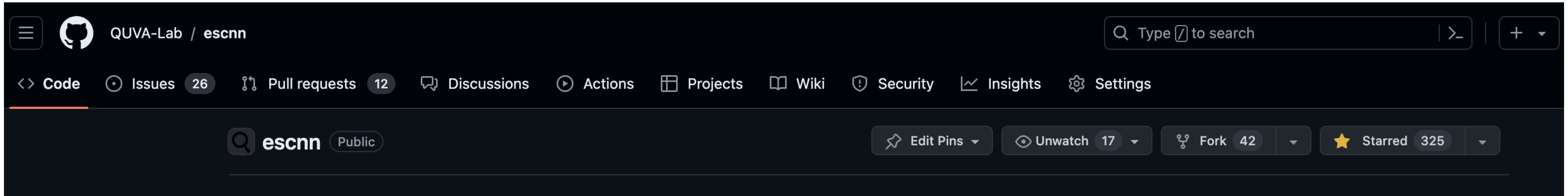
Table 2: ModelNet10 (O(2) symmetry)

G	Description	Accuracy
$\{e\}$	Conventional CNN	91.2 ± 0.5
SO(2)	Azimuthal Symmetry	91.9 ± 0.8
SO(3)	Chiral (Regular, $\lvert\mathcal{G}\rvert = 72$)	89.8 ± 0.6
O(2)	Full Azimuthal Symmetry	92.3 ± 0.4
O(3)	Achiral (Regular, $\lvert\mathcal{G}\rvert = 120$)	89.9 ± 1.0
$\mathrm{C}_2 \rtimes \mathrm{F}$	Klein Group (dihedral symmetry)	91.0 ± 0.6
	VOXNet (Maturana & Scherer, 2015)	92.0
$\mathrm{C}_2 \rtimes \mathrm{F}$	Klein Group (Worrall & Brostow, 2018)	94.2

Gabriele Cesa, Leon Lang, Maurice Weiler,
A Program to build E(n)-Equivariant Steerable CNNs,
International Conference on Representation Learning, 2022

Conclusion

- Complete theoretical description of the space of G-steerable filters
  - For any compact G and any transformation laws ρ_{in}, ρ_{out}
- Algorithm to explicitly construct the steerable convolution layers
- Generic implementation in the form of a library: github.com/QUVA-Lab/escnn



Thank you

Nothing in these materials is an offer to sell any of the components or devices referenced herein.

References in this presentation to "Qualcomm" may mean Qualcomm Incorporated, Qualcomm Technologies, Inc., and/or other subsidiaries or business units within the Qualcomm corporate structure, as applicable. Qualcomm Incorporated includes our licensing business, QTL, and the vast majority of our patent portfolio. Qualcomm Technologies, Inc., a subsidiary of Qualcomm Incorporated, operates, along with its subsidiaries, substantially all of our engineering, research and development functions, and substantially all of our products and services businesses, including our QCT semiconductor business.

Snapdragon and Qualcomm branded products are products of Qualcomm Technologies, Inc. and/or its subsidiaries. Qualcomm patented technologies are licensed by Qualcomm Incorporated.

Follow us on:

For more information, visit us at qualcomm.com & qualcomm.com/blog

Semi-Direct Product group convolution

Equivariance group $(\mathbb{R}^n, +) \rtimes SO(2)$

$f: (\mathbb{R}^n, +) \rtimes SO(2) \to \mathbb{R}$

$f(t, \cdot): [0, 2\pi) \to \mathbb{R}$

$$f(t, r_\phi) = (0, \cos 1\,\phi, \sin 1\,\phi, \cos 2\,\phi, \dots) \begin{pmatrix} w_0(t) \\ w_1(t) \\ w_2(t) \\ \dots \end{pmatrix}$$

$$\bigoplus_k \begin{pmatrix} \cos k\,\phi \\ \sin k\,\phi \end{pmatrix} \qquad \bigoplus_k \begin{pmatrix} w_{2k}(t) \\ w_{2k+1}(t) \end{pmatrix}$$

Semi-Direct Product group convolution

Equivariance group $(\mathbb{R}^n, +) \rtimes SO(2)$

$f : (\mathbb{R}^n, +) \rtimes SO(2) \to \mathbb{R}$

$f(t, \cdot) : [0, 2\pi) \to \mathbb{R}$

$$[r_\theta . f](t, r_\phi) = (0, \cos 1\,\phi, \sin 1\,\phi, \cos 2\,\phi, \ldots)\,\rho(\theta) \begin{pmatrix} w_0(t) \\ w_1(t) \\ w_2(t) \\ \ldots \end{pmatrix}$$

$$\bigoplus_k \begin{pmatrix} \cos k\,\phi \\ \sin k\,\phi \end{pmatrix} \qquad \bigoplus_k \rho_k(\theta) \qquad \bigoplus_k \begin{pmatrix} w_{2k}(t) \\ w_{2k+1}(t) \end{pmatrix}$$