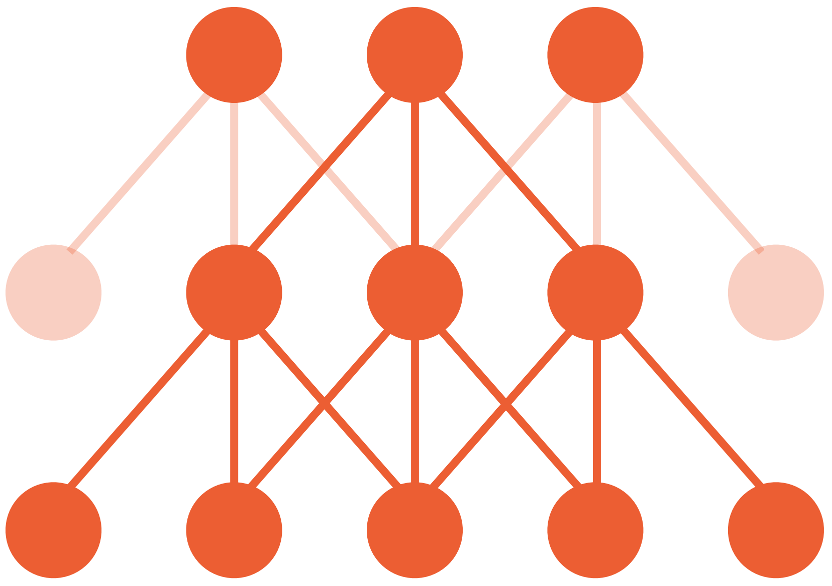
Understanding And Generalising The Convolution

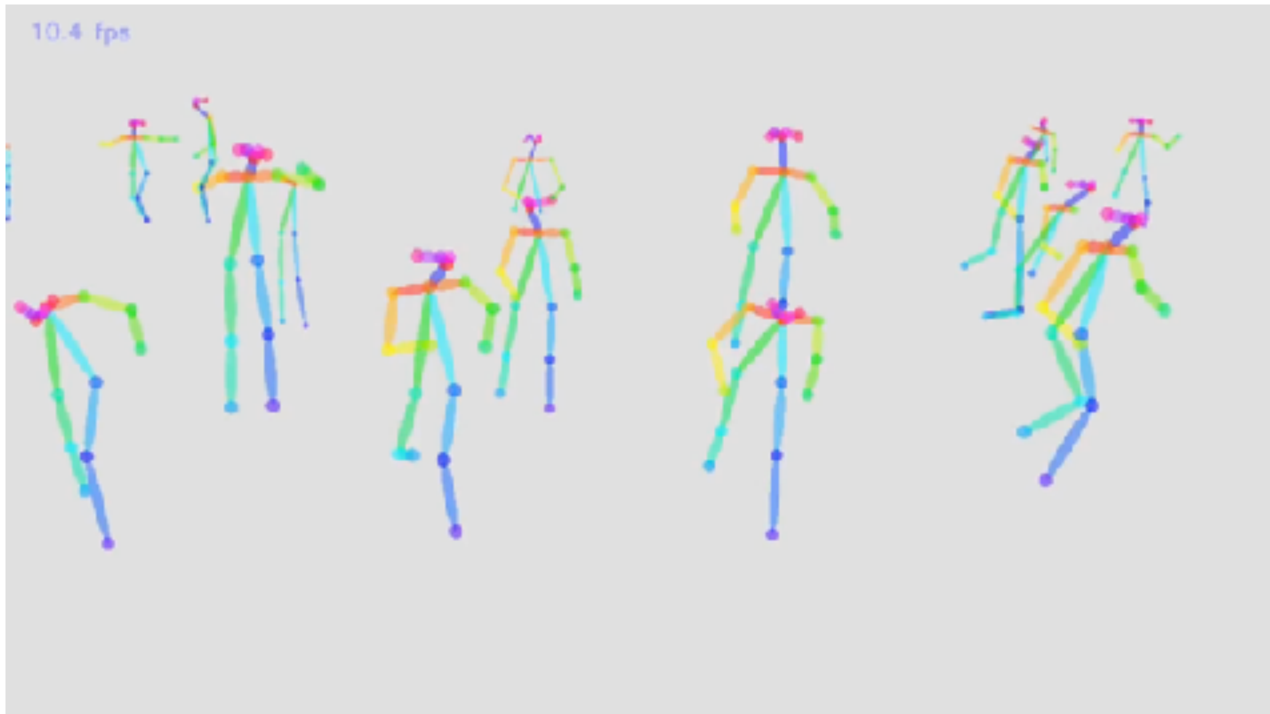
Daniel Worrall
Universiteit van Amsterdam
Instituut voor Informatica
Philips Lab, AMLab

University College London



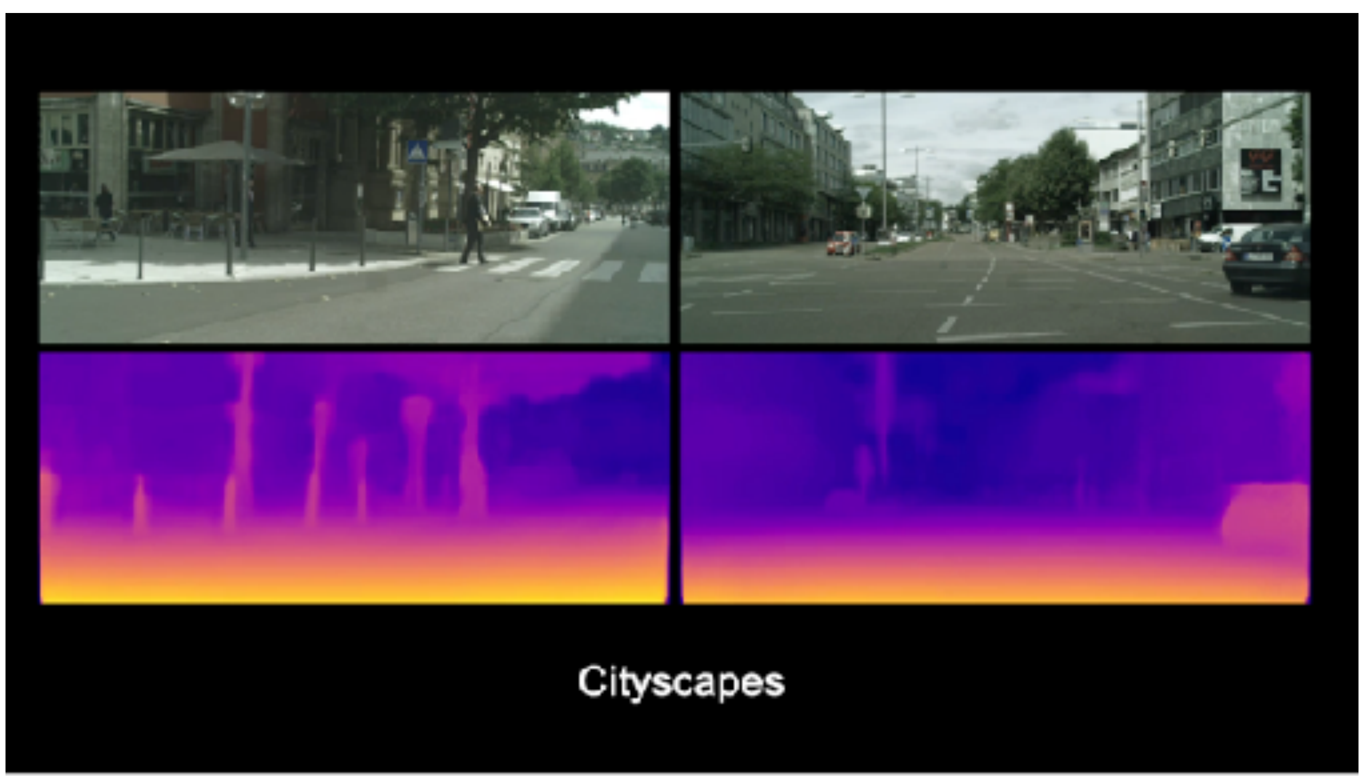
What makes CNNs special?

CNNs Are Responsible For Many Successes

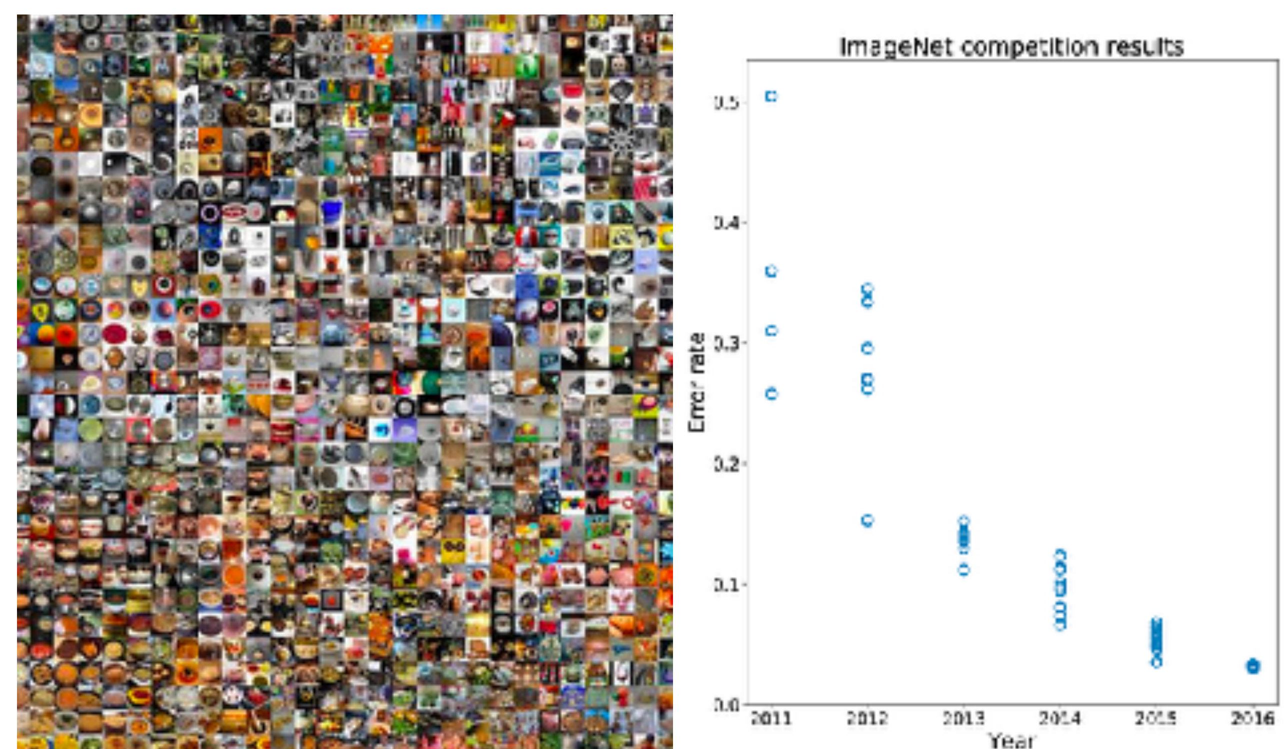


Cao et al. (2017)

Karras et al. (2017)



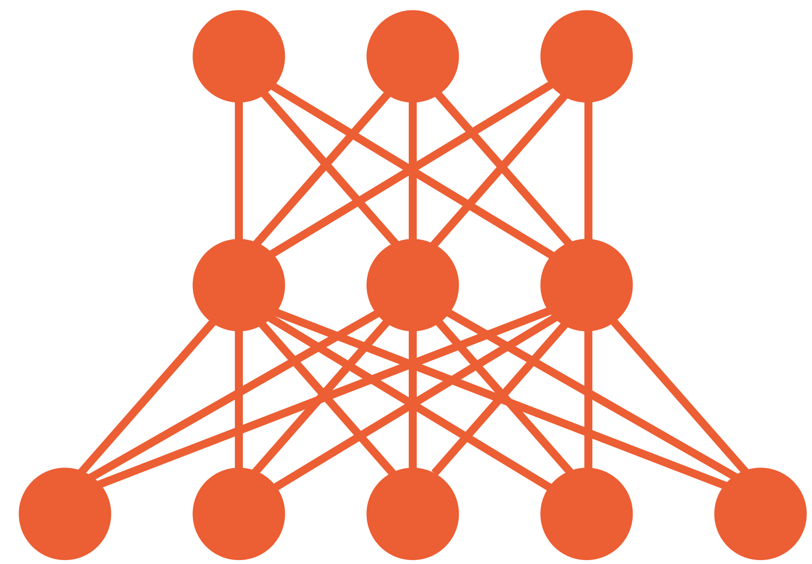
Godard et al. (2017)



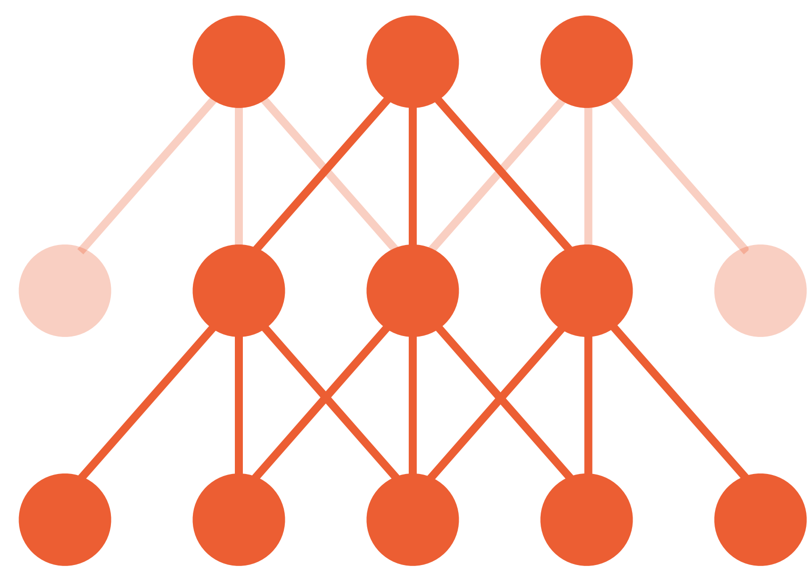
Ongoing [figure from Wikipedia]

WHAT MAKES CNNS SPECIAL?

- **Multilayer perceptrons (MLPs)**
  1. Global connectivity
  2. Seldom used in computer vision

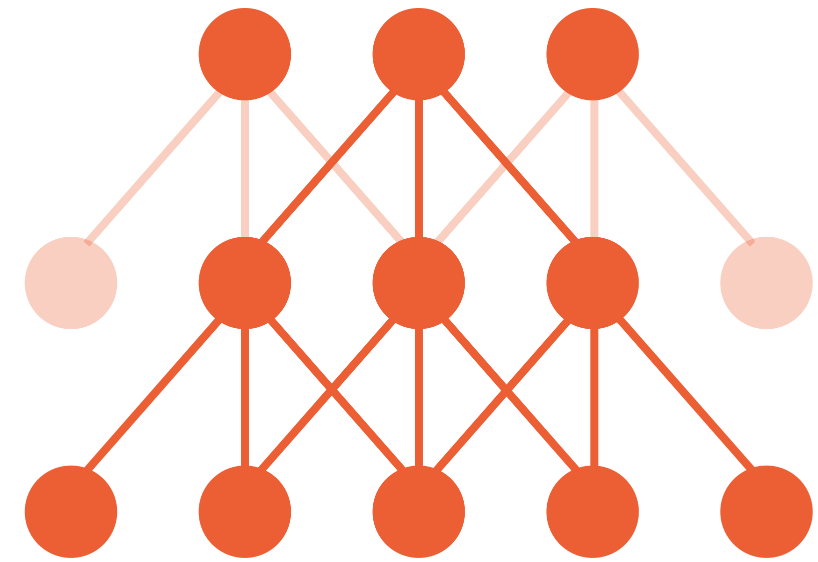


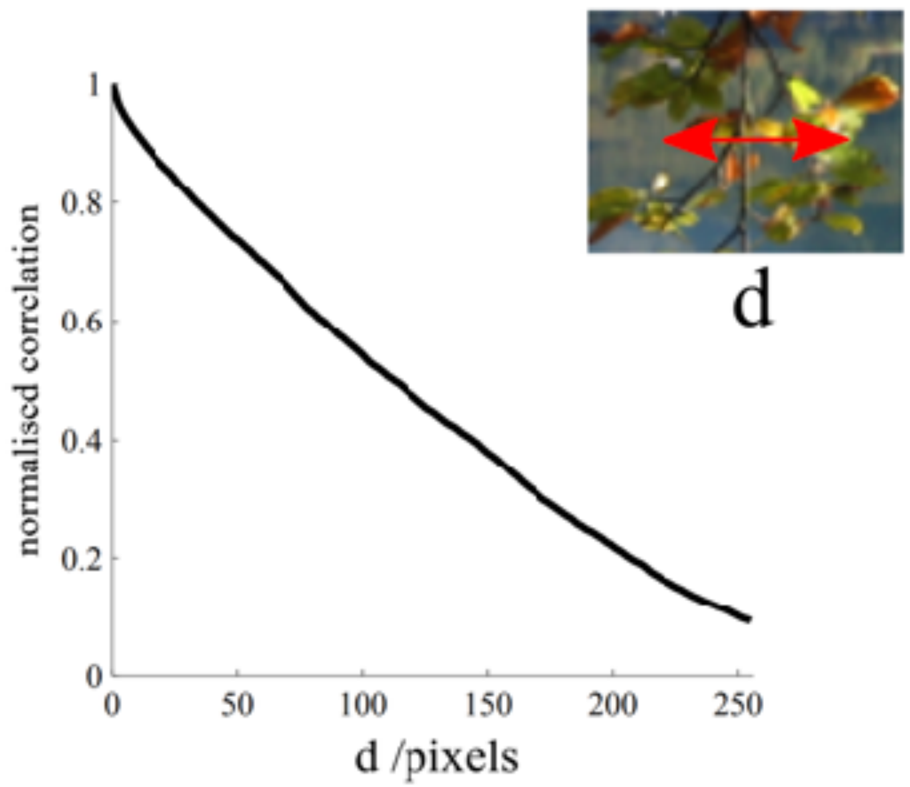
- **Convolutional neural networks (CNNs)**
  1. Local "receptive field" connectivity
  2. Translational weight tying ***of convolution***
     - Basis of ***many*** state-of-the-art models



WHAT MAKES CNNS SPECIAL?

- **Convolutional neural networks (CNNs)**
  1. Local "receptive field" connectivity
  2. Translational weight tying ***of convolution***
  - Basis of ***many*** state-of-the-art models

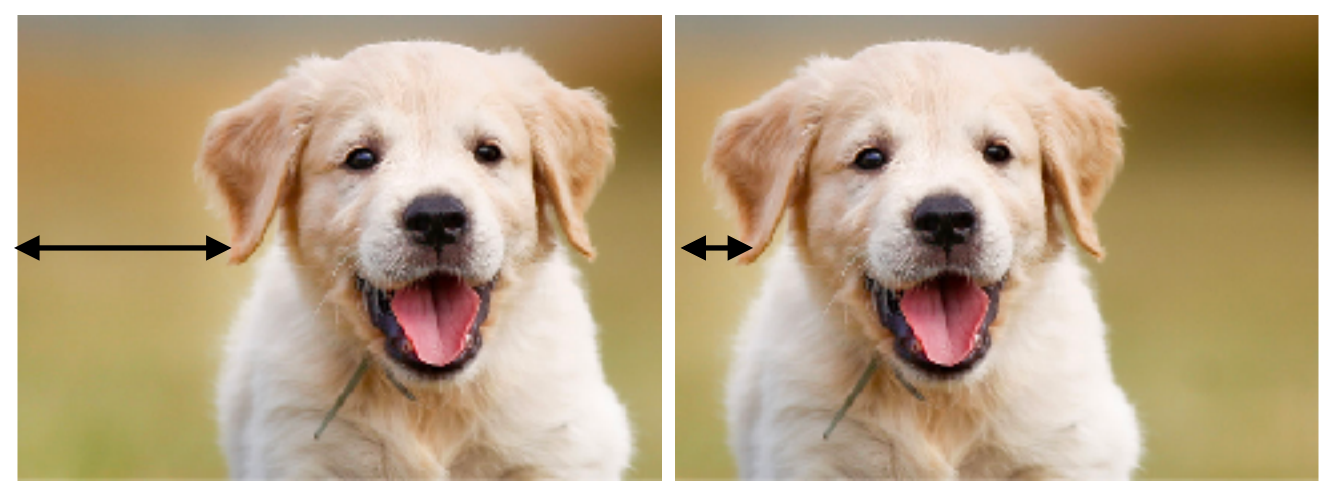




Locality **of *pixel statistics***

Property of the data

http://www.gatsby.ucl.ac.uk/~turner/teaching/4g3/2013/stats-recep-field.pdf

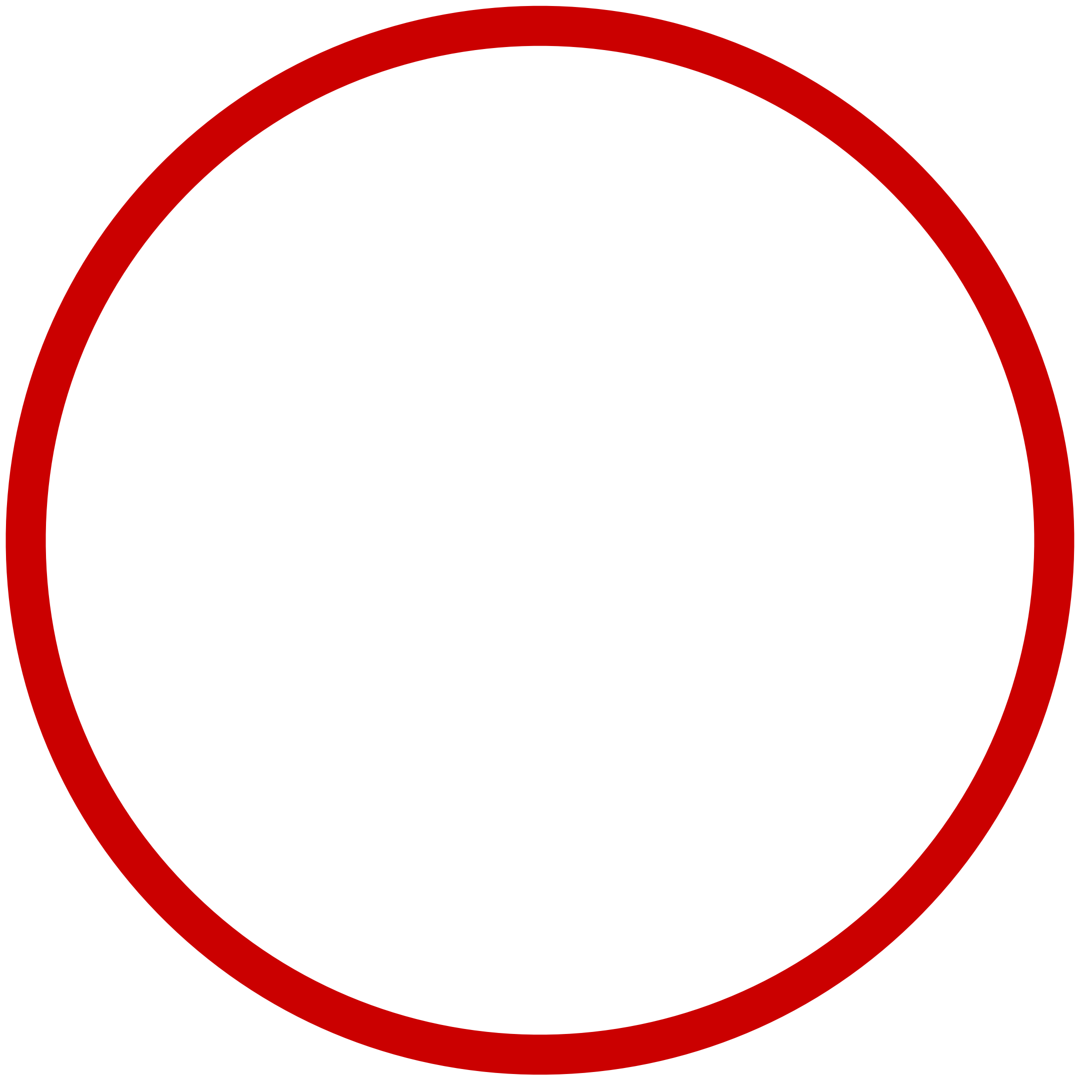


Translational invariance **of classification**

Property of the task

https://www.healthypawspetinsurance.com/Images/V3/DogAndPuppyInsurance/Dog_CTA_Desktop_HeroImage.jpg

OVERVIEW



What is symmetry?

What Is Symmetry?

Set of input transformations leaving f invariant

$$f(\mathbf{I}) = f(\mathcal{T}_\theta[\mathbf{I}])$$

image

function/
feature mapping

transformation

Notational aside:

e.g. Geometric translation

$$\mathcal{T}_\theta[\mathbf{I}](\mathbf{x}) = \mathbf{I}(\mathbf{x} - \boldsymbol{\theta})$$

e.g. Geometric rotation

$$\mathcal{T}_\theta[\mathbf{I}](\mathbf{x}) = \mathbf{I}(\mathbf{R}_{\boldsymbol{\theta}}^{-1}\mathbf{x})$$

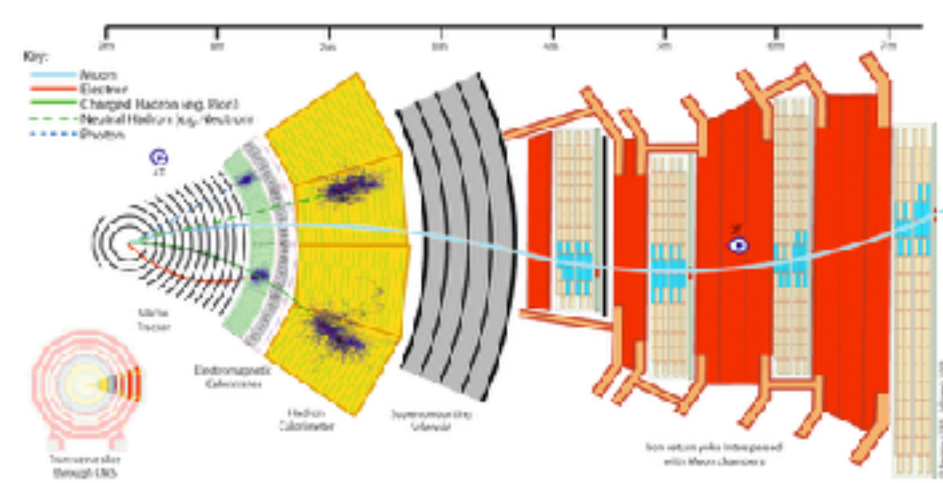
e.g. Pixel normalisation

$$\mathcal{T}[\mathbf{I}] = (\mathbf{I} - \boldsymbol{\mu})/.\boldsymbol{\sigma}^{-1}$$

Symmetry is a property of functions/tasks, e.g.

Classification

Disentangling
(cocktail party)



Signal discovery/detection

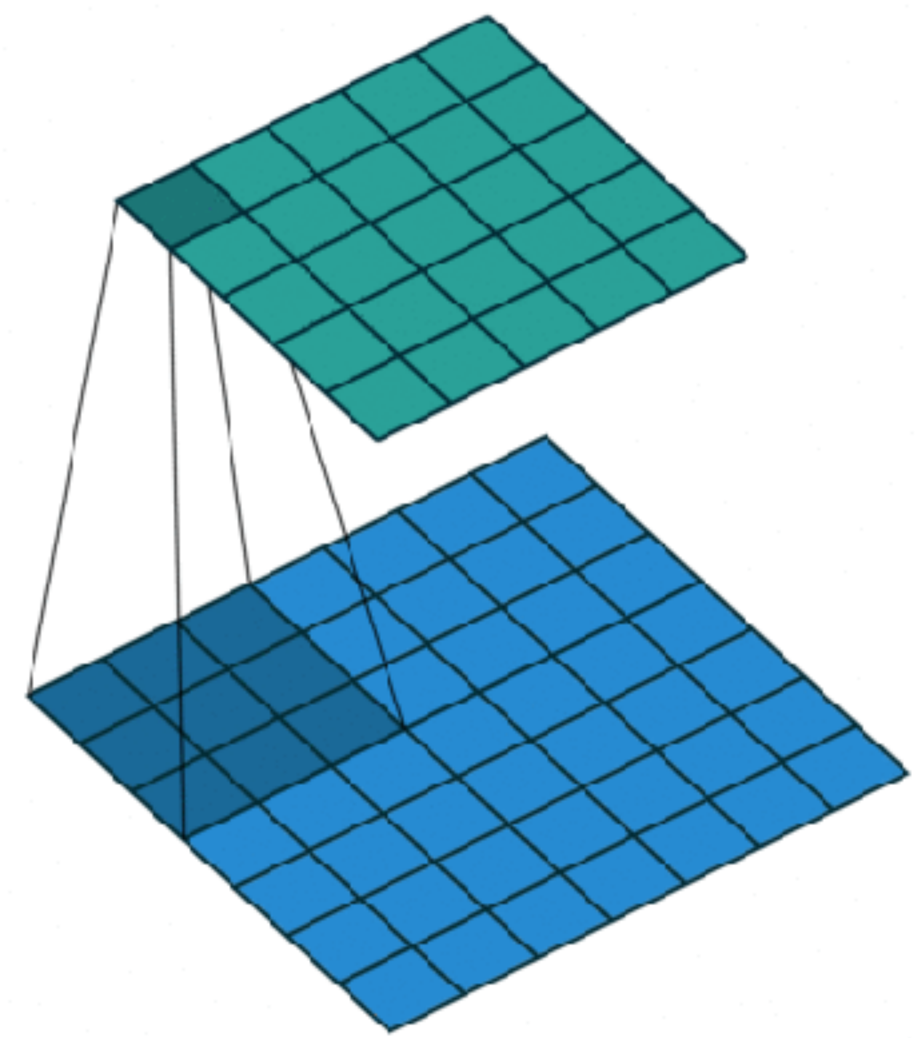
EQUIVARIANCE

Different *representations* of same transformation

Mapping preserves algebraic structure of transformation

$$\mathcal{S}_\theta[f](\mathbf{I}) = f(\mathcal{T}_\theta[\mathbf{I}])$$

transformation in feature space



Invariance

$$\mathcal{S}_{\boldsymbol{\theta}} = \mathrm{Id}$$

Convolution (and correlation)

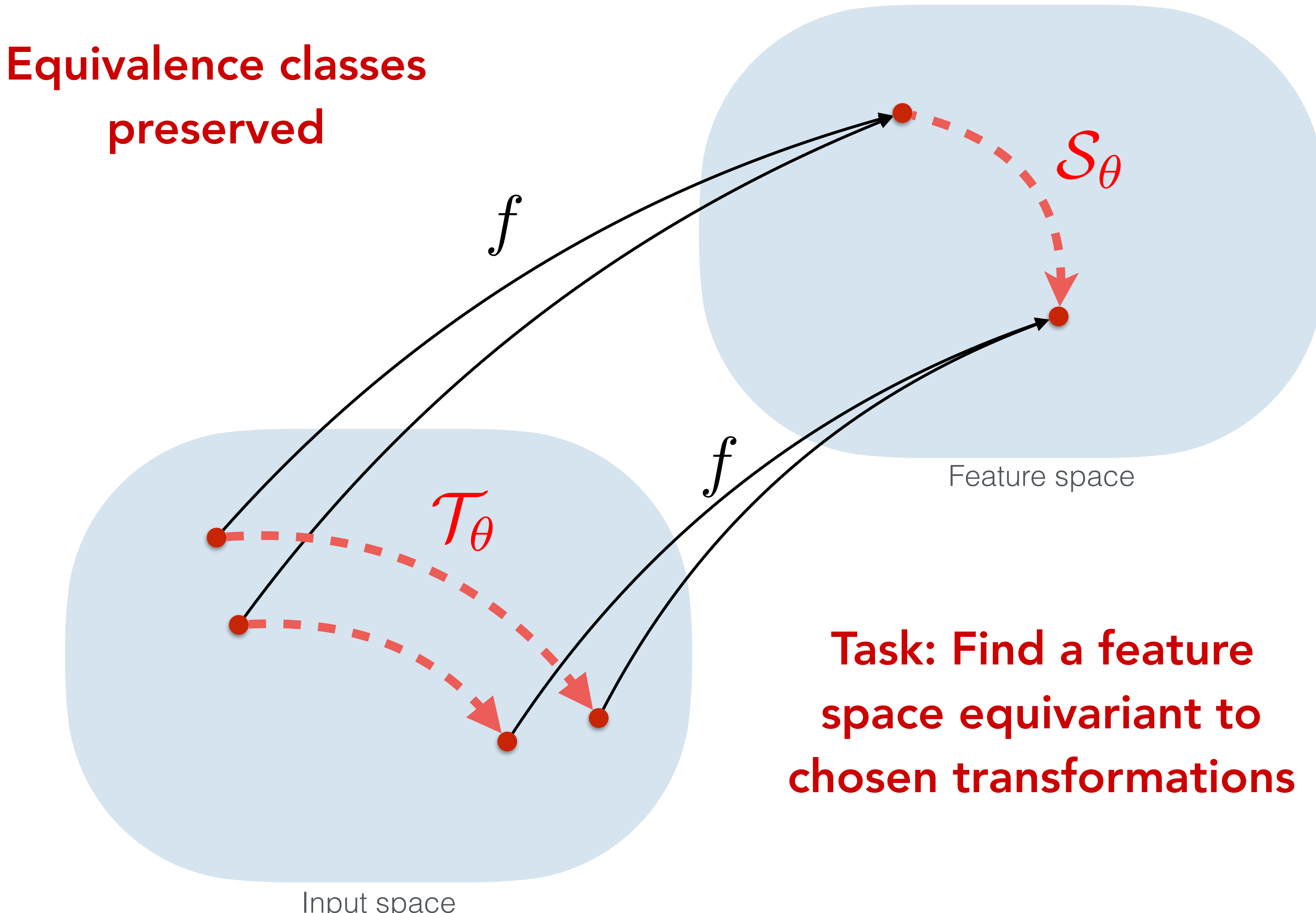
$$[\mathbf{I} * \mathbf{W}](\mathbf{x} - \boldsymbol{\theta}) = \mathcal{T}_{\boldsymbol{\theta}}[\mathbf{I}] * \mathbf{W}(\mathbf{x})$$

Fourier spectrum phase

$$e^{-\iota\boldsymbol{\omega}^\top\boldsymbol{\theta}} \int_{\mathbb{R}^N} \mathbf{I}(\mathbf{x}) e^{-\iota\boldsymbol{\omega}^\top\mathbf{x}}\, \mathrm{d}\mathbf{x} = \int_{\mathbb{R}^N} \mathbf{I}(\mathbf{x} - \boldsymbol{\theta}) e^{-\iota\boldsymbol{\omega}^\top\mathbf{x}}\, \mathrm{d}\mathbf{x}$$

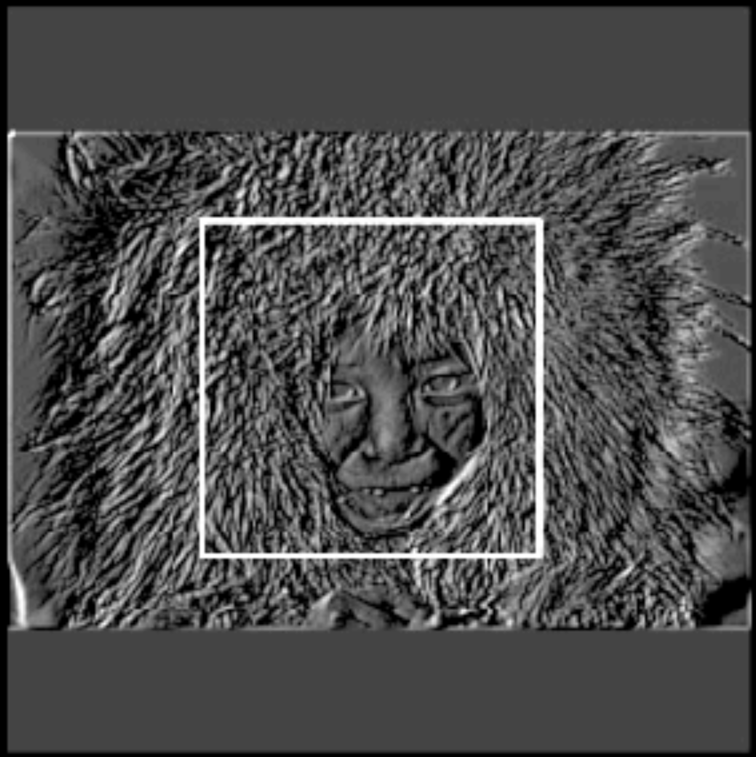
https://github.com/vdumoulin/conv_arithmetic

EQUIVARIANCE INTUITION

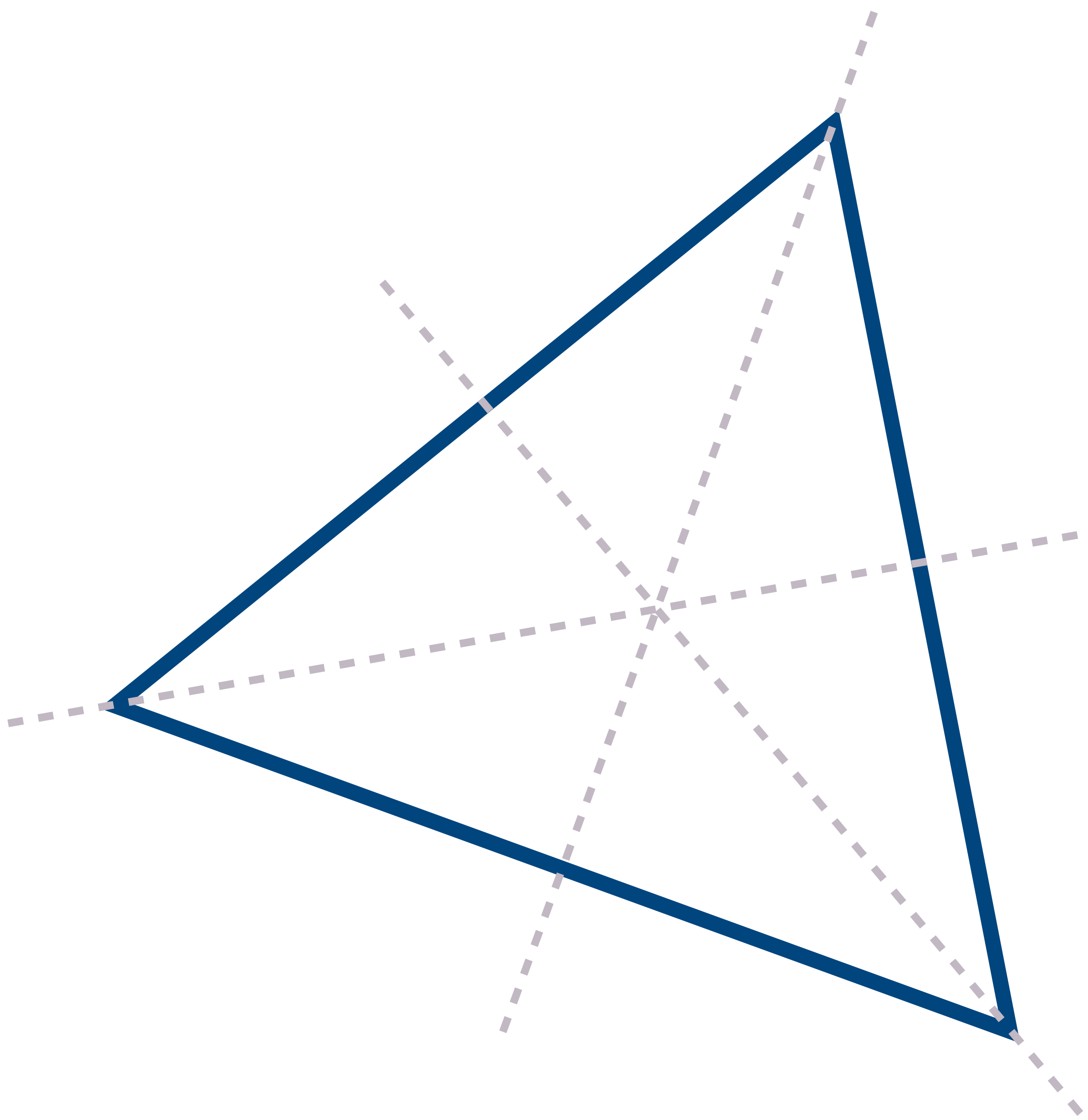


Thanks to Andrea Vedaldi for this insight

LACK OF KNOWN ROTATION **EQUI**VARIANCE:



Overview



Group symmetries

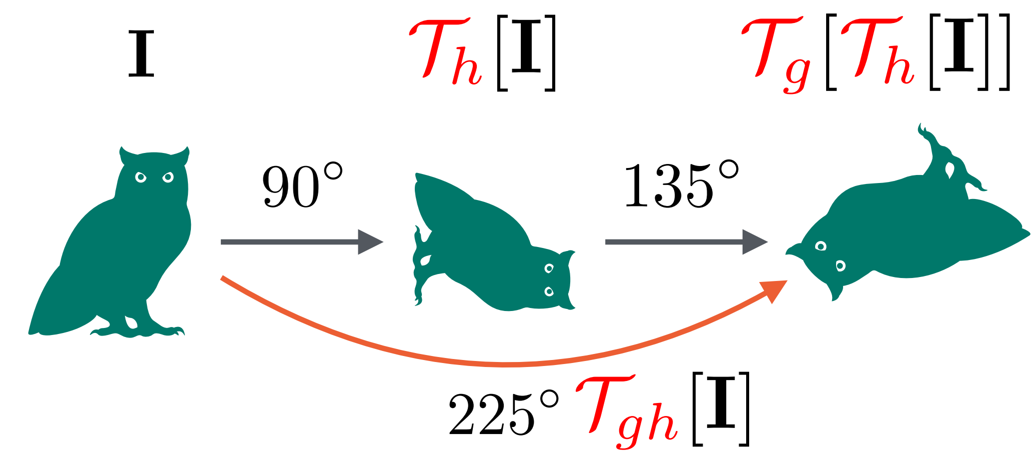
What Is Group Theory?

- **Field of abstract mathematics**
  - Language to describe invertible transformations

- **Criteria to be a group G**

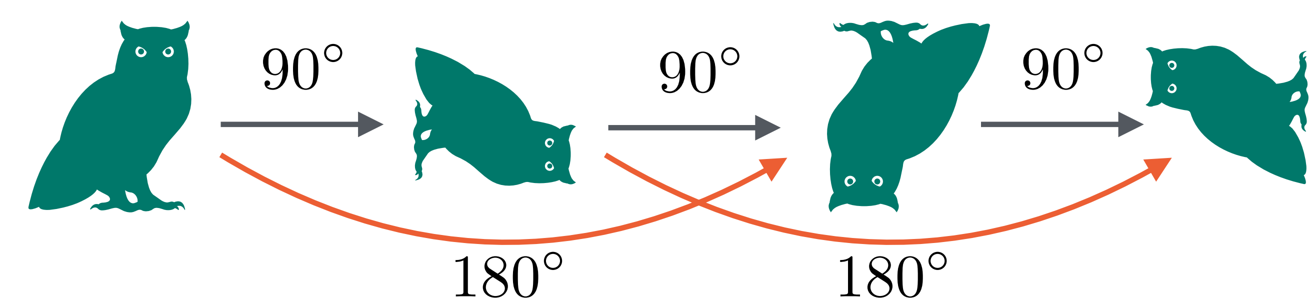
1. **Closure:** two transformations in a row is a transformation

$$gh \in G$$



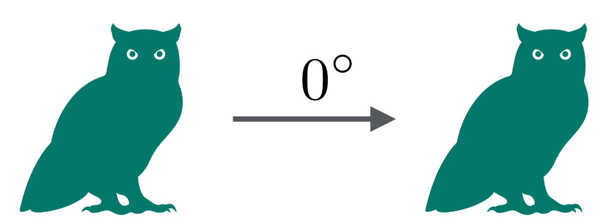
2. **Associativity:** transforming twice same as single compound transformation

$$g(hk) = (gh)k \in G$$



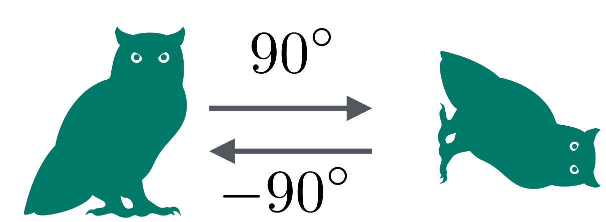
3. **Identity:** there is a “do nothing” transformation

$$eg = ge = g$$

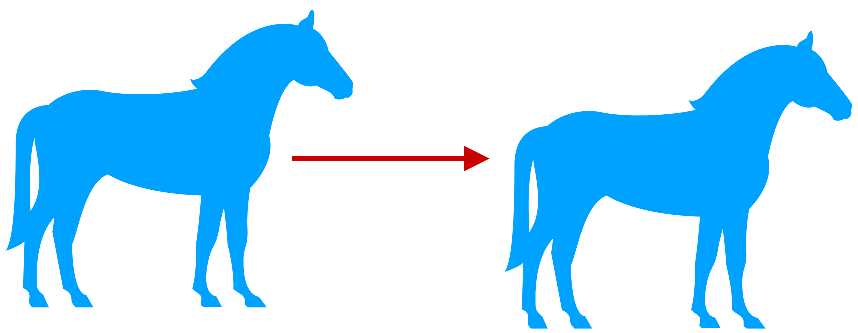


4. **Invertibility:** every transformation can be reversed perfectly

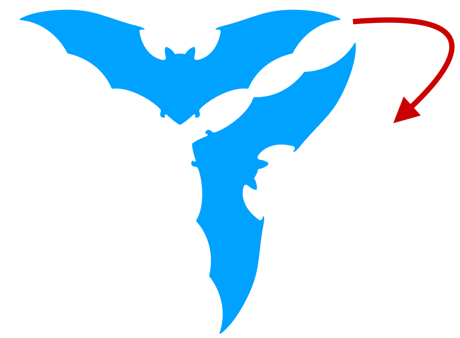
$$g^{-1}g = gg^{-1} = e$$



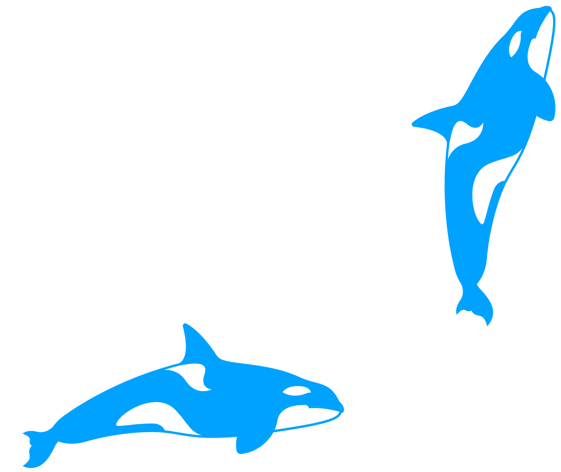
Group Examples



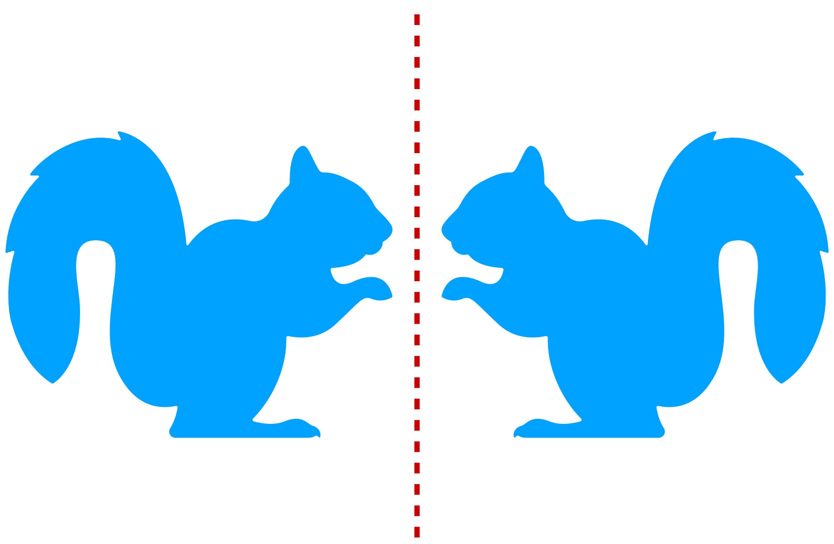
Translation



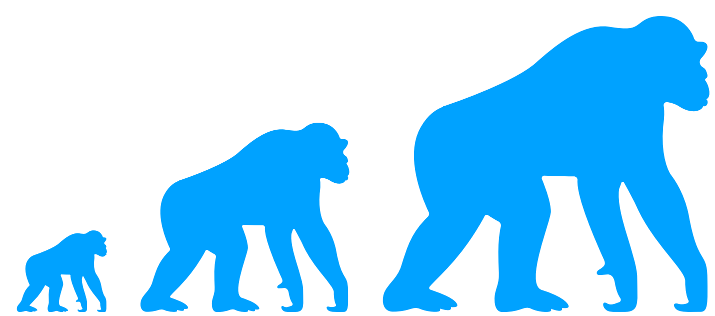
Rotation



Roto-translation

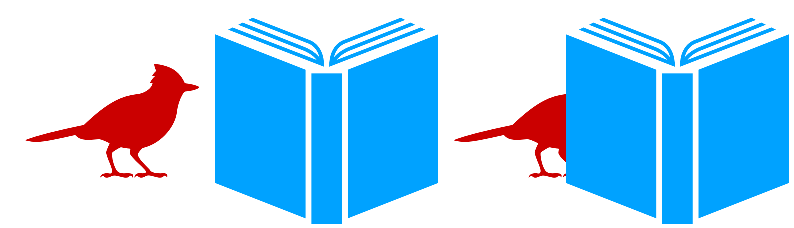


Reflections



Scalings*

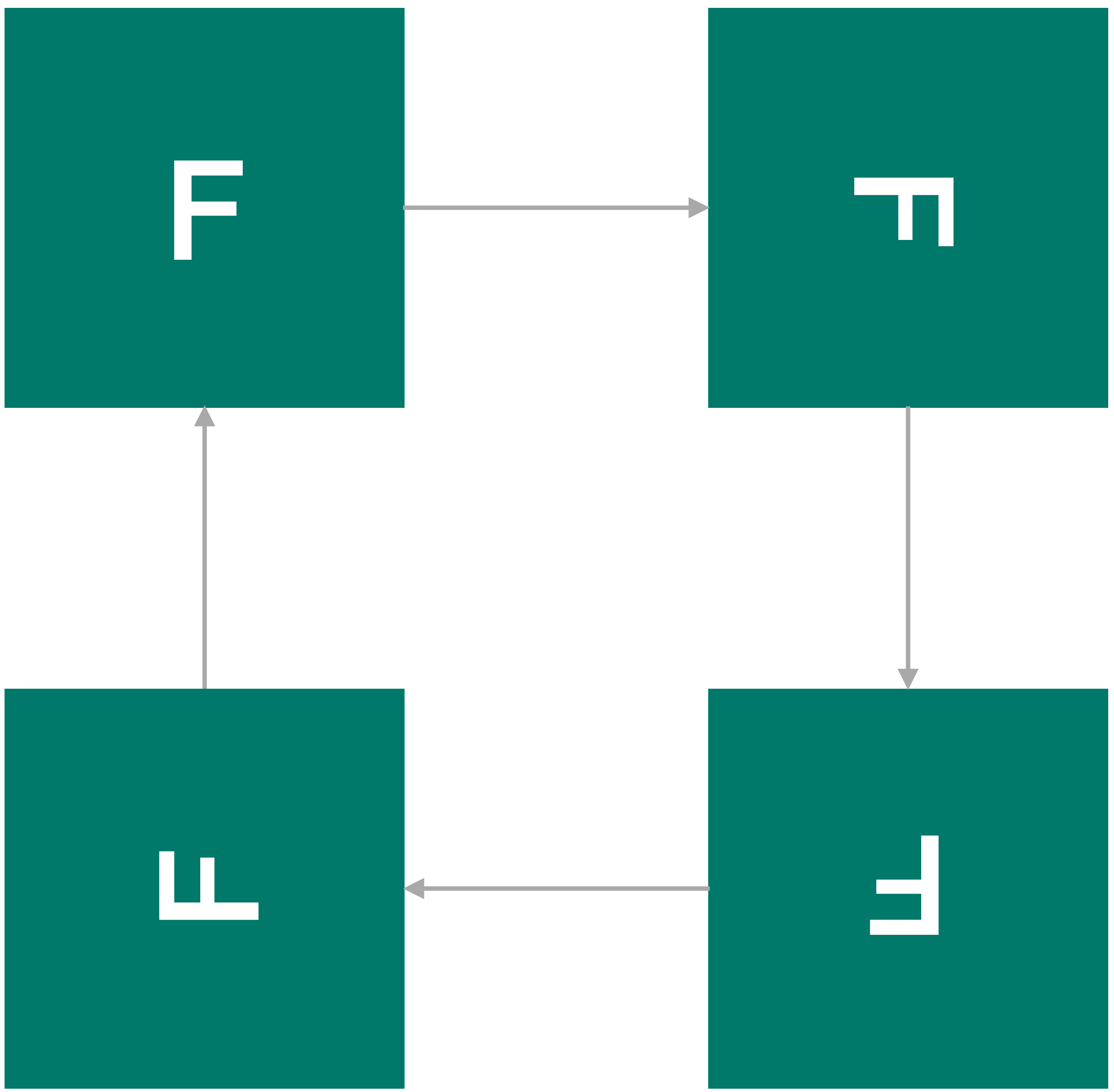
Non-example



Occlusions

*Scalings are probably better modelled as semigroups, i.e. groups without the invertibility condition

OVERVIEW



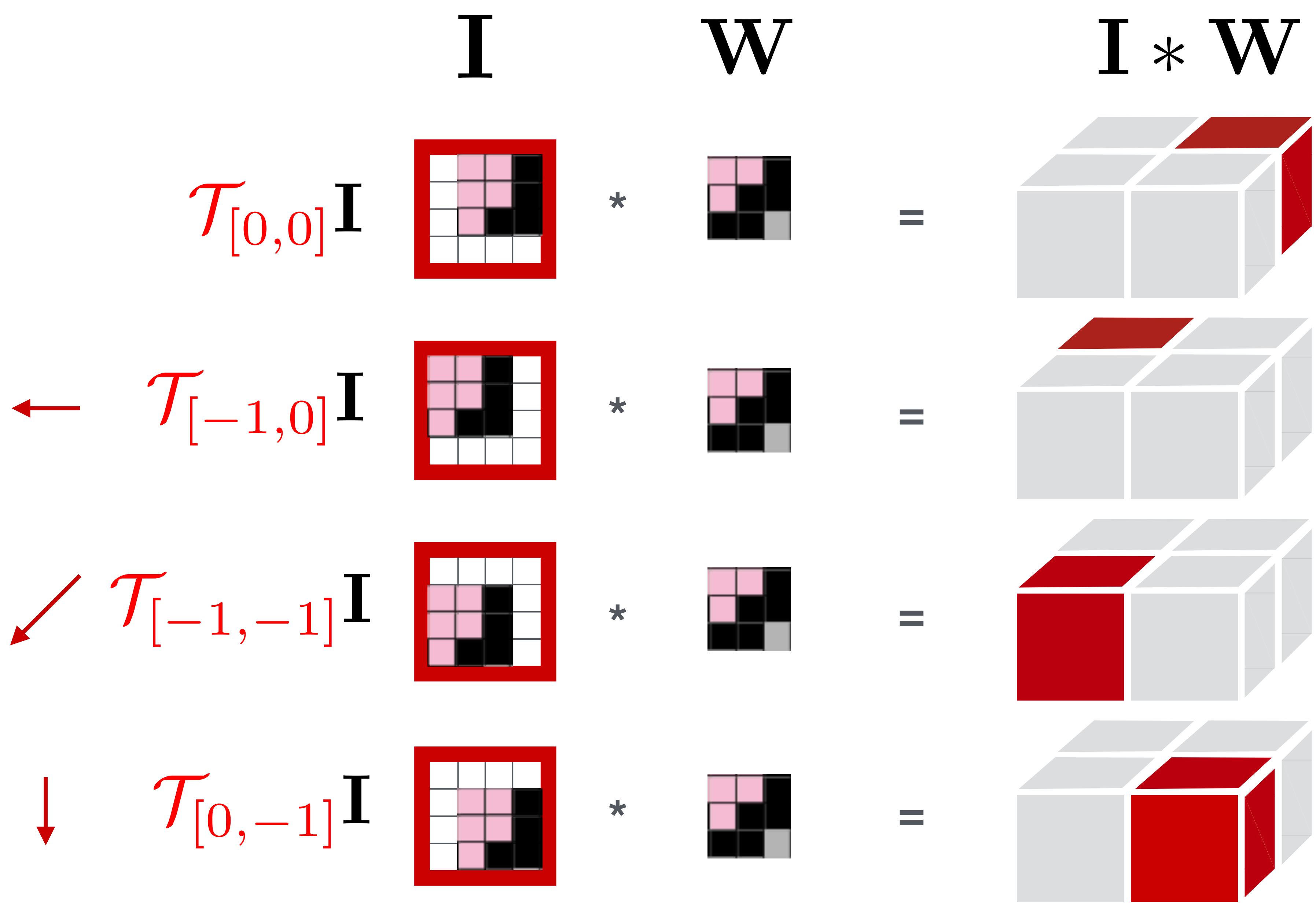
Group convolutions

CONVOLUTIONS ON TRANSLATION GROUP

Standard (yet much beloved) convolution

$$[\mathbf{I} * \mathbf{W}](\mathbf{y}) = \sum_{\mathbf{x}\in\mathbb{Z}^2} \mathbf{I}(\mathbf{x})\mathbf{W}(\mathbf{x}-\mathbf{y})$$

Translating kernel



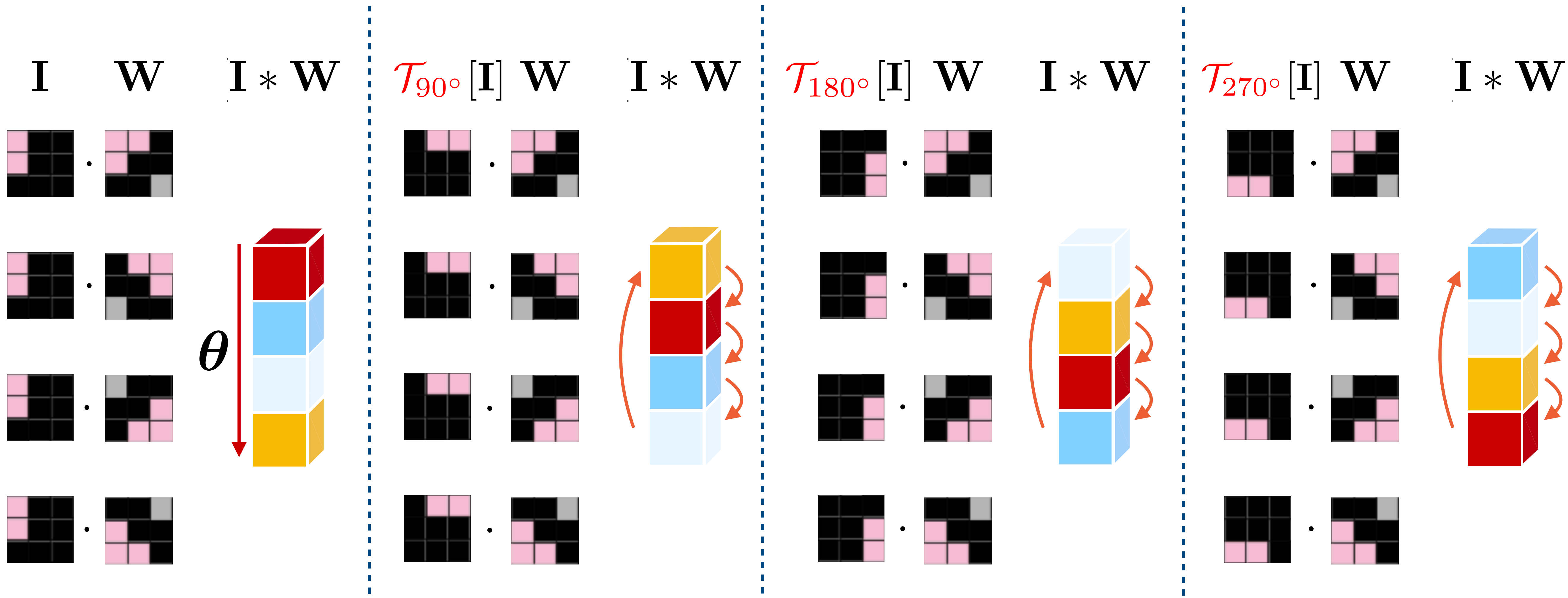
Convolutions on Rotation Group

Group convolution **Notice index of convolution**

Ref.
$$[\mathbf{I} * \mathbf{W}](\mathbf{y}) = \sum_{\mathbf{x}\in\mathbb{Z}^2} \mathbf{I}(\mathbf{x})\mathbf{W}(\mathbf{x}-\mathbf{y})$$

$$[\mathbf{I} * \mathbf{W}](\boldsymbol{\theta}) = \sum_{\mathbf{x}\in\mathbb{Z}^2} \mathbf{I}(\mathbf{x})\mathbf{W}(\mathbf{R}_{\theta}^{-1}\mathbf{x})$$

Rotating kernel



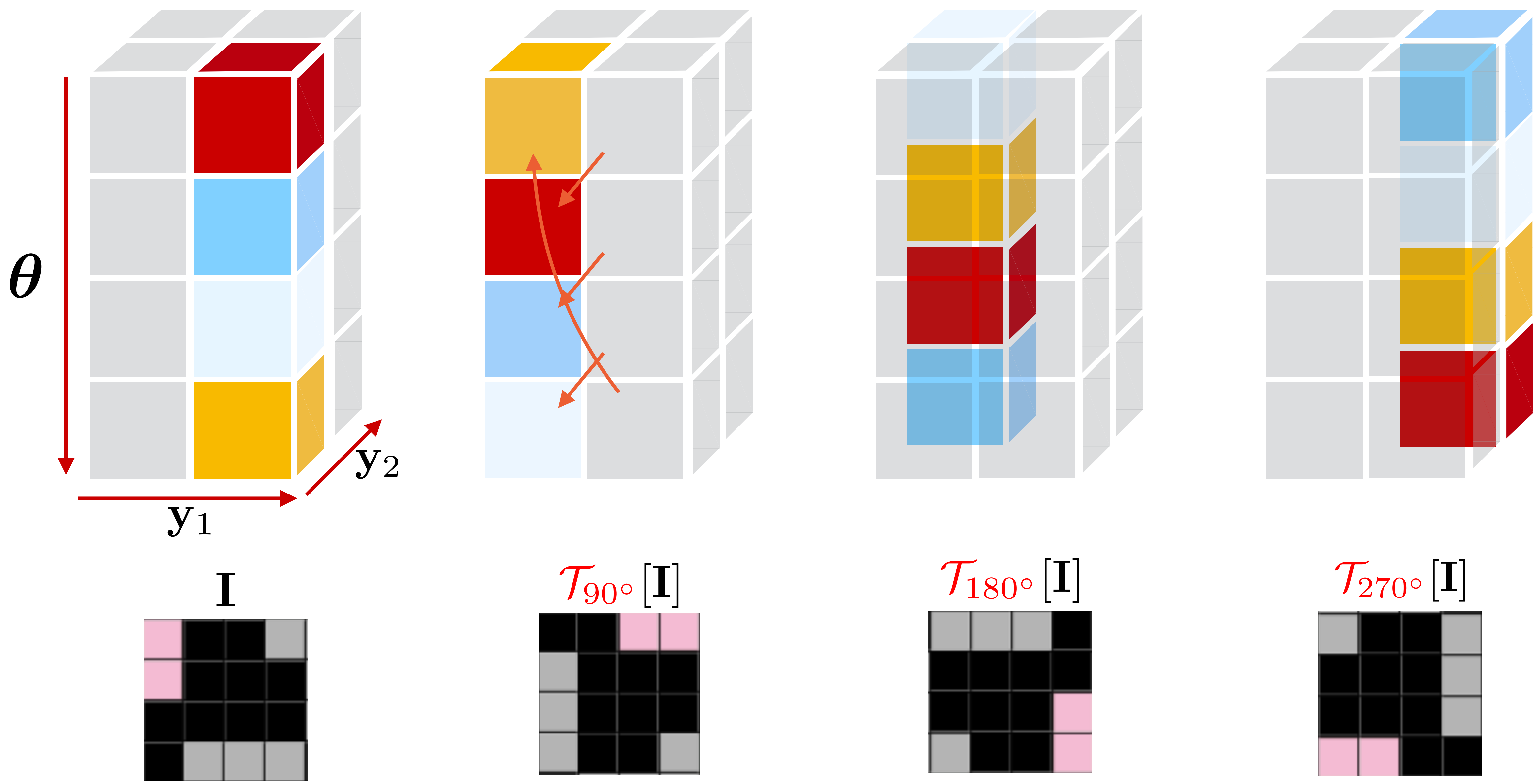
@ hidden layer $$[\mathbf{I} * \mathbf{W}](\psi) = \sum_{\boldsymbol{\theta}\in\Theta} \mathbf{I}(\boldsymbol{\theta})\mathbf{W}(\theta - \psi)$$

Convolutions on Roto-Translation Group

Group convolution

$$[\mathbf{I} * \mathbf{W}](\boldsymbol{\theta}, \mathbf{y}) = \sum_{\mathbf{x} \in \mathbb{Z}^2} \mathbf{I}(\mathbf{x}) \mathbf{W}(\mathbf{R}_{\theta}^{-1}\mathbf{x} - \mathbf{y})$$

Roto-translating kernel



@ hidden layer $$[\mathbf{I} * \mathbf{W}](\boldsymbol{\psi}, \mathbf{z}) = \sum_{(\boldsymbol{\theta}, \mathbf{Z}) \in (\Theta \times \mathbb{Z}^2)} \mathbf{I}(\boldsymbol{\theta}, \mathbf{y}) \mathbf{W}(\mathbf{R}_{\psi}^{-1}\mathbf{y} - \mathbf{z}, \boldsymbol{\theta} - \boldsymbol{\psi})$$

CONVOLUTIONS ON DISCRETE GROUPS

Convolution examples

$$[\mathbf{I} * \mathbf{W}](\mathbf{y}) = \sum_{\mathbf{x} \in \mathbb{Z}^2} \mathbf{I}(\mathbf{x}) \mathbf{W}(\mathbf{x} - \mathbf{y})$$

$$[\mathbf{I} * \mathbf{W}](\boldsymbol{\theta}) = \sum_{\mathbf{x} \in \mathbb{Z}^2} \mathbf{I}(\mathbf{x}) \mathbf{W}(\mathbf{R}_\theta^{-1} \mathbf{x})$$

$$[\mathbf{I} * \mathbf{W}](\boldsymbol{\theta}, \mathbf{y}) = \sum_{\mathbf{x} \in \mathbb{Z}^2} \mathbf{I}(\mathbf{x}) \mathbf{W}(\mathbf{R}_\theta^{-1} \mathbf{x} - \mathbf{y})$$

For general discrete groups

@ input layer

$$[\mathbf{I} * \mathbf{W}](\boldsymbol{\theta}) = \sum_{\mathbf{x} \in \mathbb{Z}^2} \mathbf{I}(\mathbf{x}) \mathcal{T}_\theta[\mathbf{W}](\mathbf{x})$$

@ hidden layer

$$[\mathbf{I} * \mathbf{W}](\psi) = \sum_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathbf{I}(\boldsymbol{\theta}) \mathcal{T}_\psi[\mathbf{W}](\boldsymbol{\theta})$$

$$= \sum_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathbf{I}(\boldsymbol{\theta}) \mathbf{W}(\psi^{-1} \boldsymbol{\theta})$$

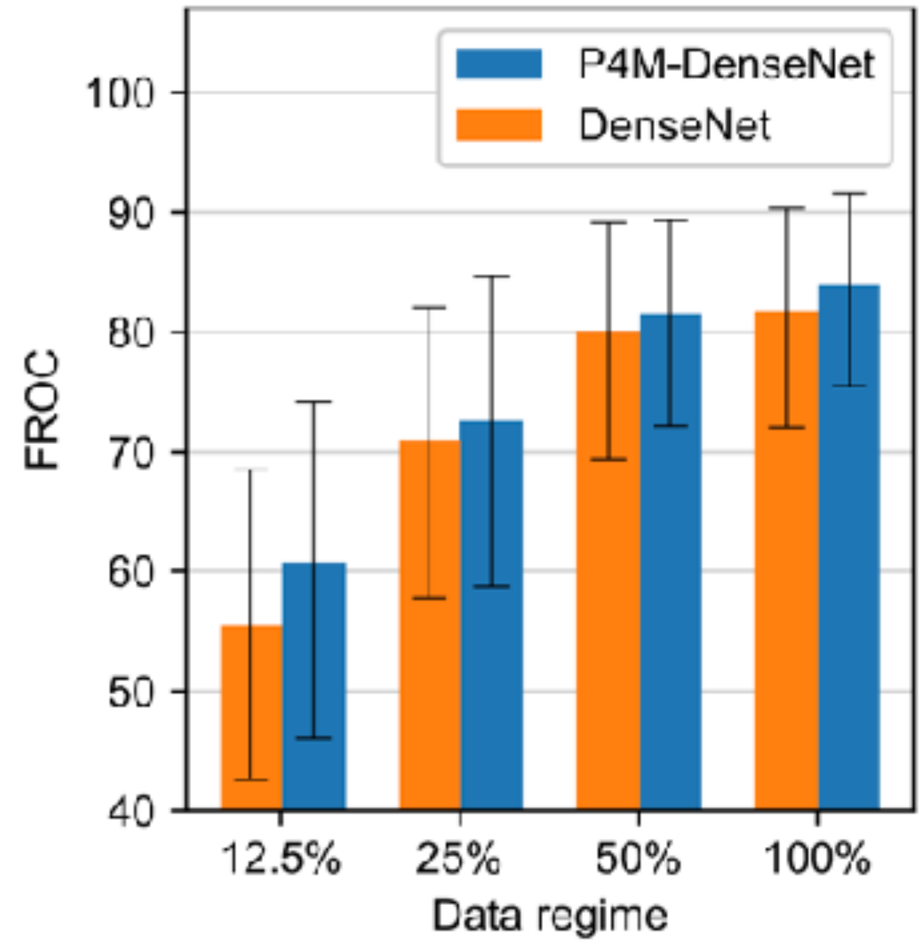
Read: Group Equivariant Convolutional Networks, Cohen & Welling, 2016

EXAMPLES

Rotation Equivariant CNNs for Digital Pathology

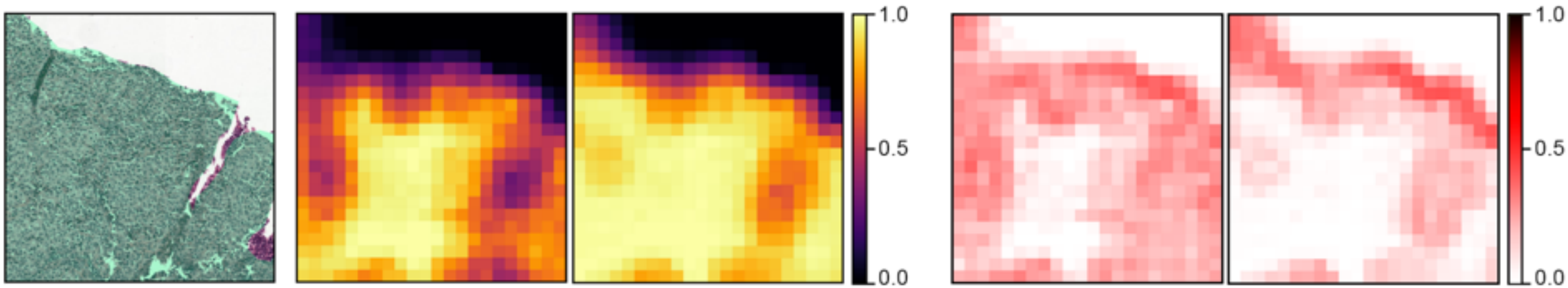
Bastiaan S. Veeling*, Jasper Linmans*, Jim Winkens*, Taco Cohen, and Max Welling

University of Amsterdam, The Netherlands



Mean prediction

Standard deviation



Input

DenseNet

Rotation equivariant DenseNet

DenseNet

Rotation equivariant DenseNet

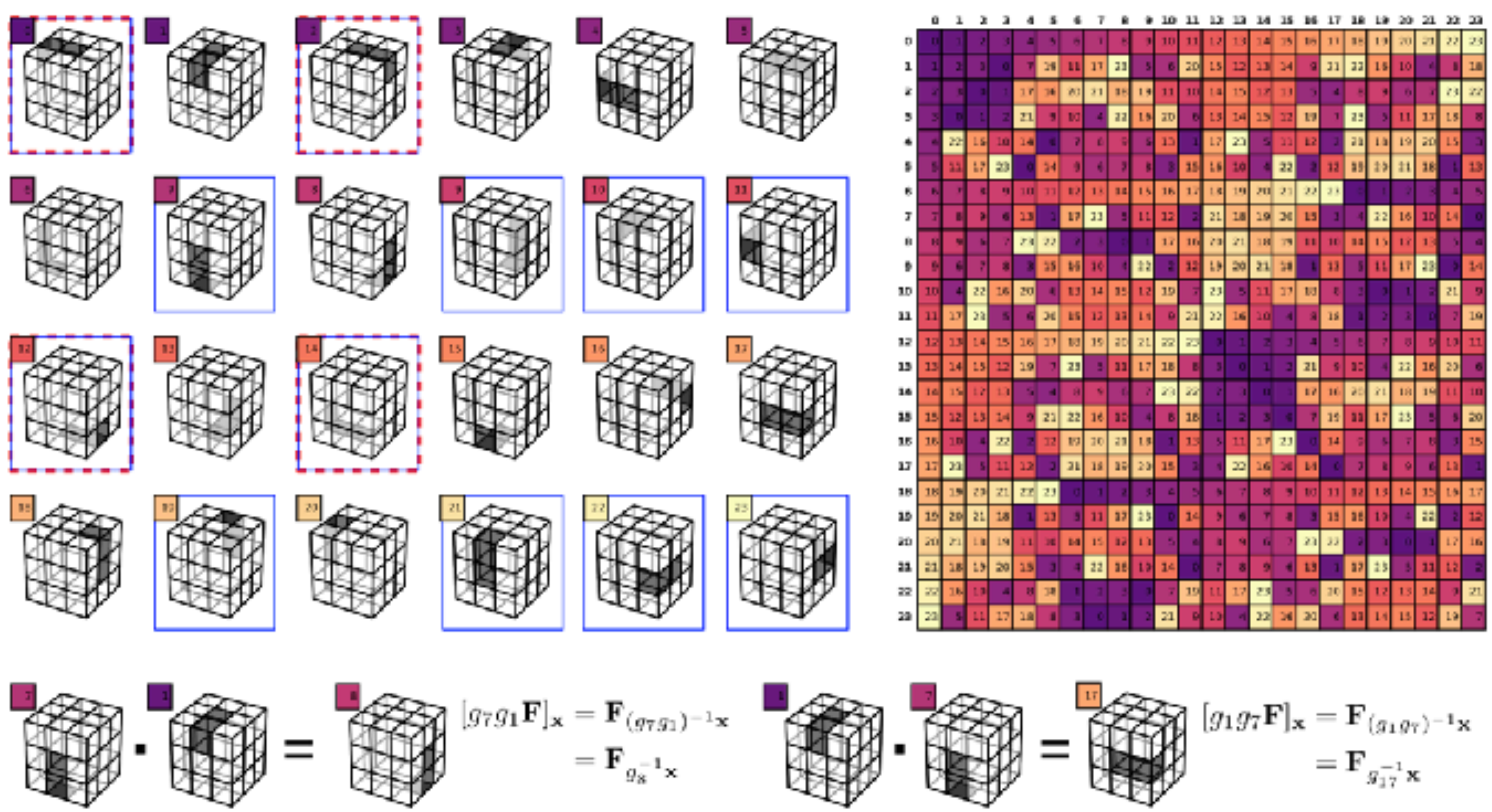
EXAMPLES

CubeNet: Equivariance to 3D Rotation and Translation

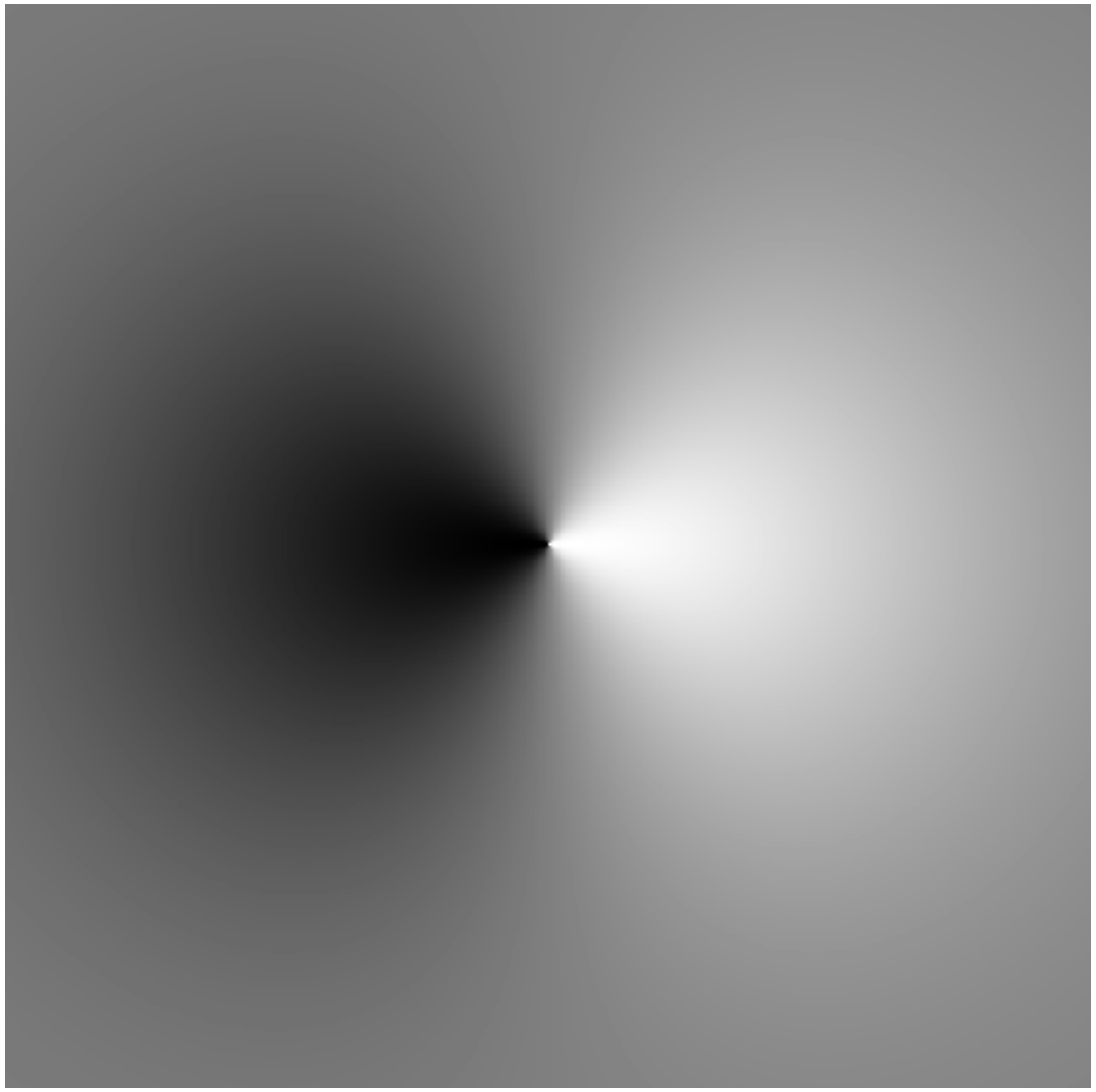
Daniel Worrall and Gabriel Brostow

Computer Science Department, University College London, UK
{d.worrall,g.brostow}@cs.ucl.ac.uk

Method	ModelNet10	# params ($\times 10^6$, 2 s.f.)
3D ShapeNets [41]	0.8354	12
Xu & Todovoric [48]	0.8800	0.080
3D-GAN [49]	0.9100	11
VRN [42]	0.9133	18
VoxNet [40]	0.9200	0.92
Fusion-Net [50]	0.9311	120
ORION [43]	0.9380	0.91
Ours T_4	0.9127	4.5
Our V (average)	0.9372	4.5
Ours V (best model single-view)	0.9120	4.5
Ours V (best model rotation averaged)	**0.9460**	4.5
VRN Ensemble [42]	0.9714	108



Overview



Harmonic/steerable convolutions

ROTATION EQUIVARIANCE METHODS

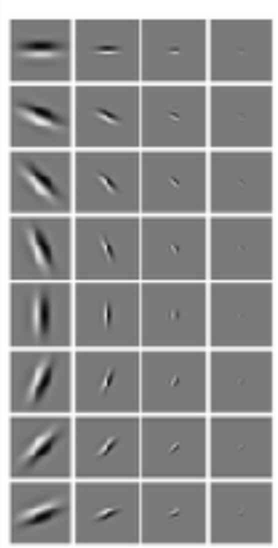
Rotate in network

Rotate filters

Rotate activations

Exploiting Cyclic Symmetry
(Dieleman et al., 2016)

Store full response



Roto-Translation Scattering
(Oyallon et al., 2014)
Group Equivariant CNNs
(Cohen et al., 2016)
Oriented Response Networks
(Zhou et al., 2017)
Steerable CNNs
(Cohen et al., 2017)
Steerable Filters G-CNNs
(Weiler et al. 2018)

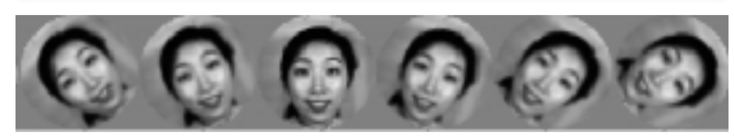
Store max/mean response

Roto-Translation Invariant Scattering
(Sifre et al., 2013)
Rotation Equivariant Vector Fields
(Marcos et al., 2017)
And more…

Rotate input

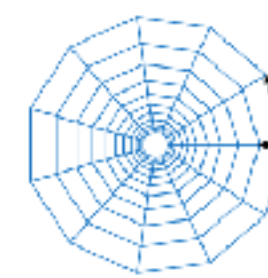
Brute-force

Rotation-Invariant Neoperceptron
(Fasel et al., 2006)
TI-Pooling
(Laptev et al., 2016)



Log-polar trans.

Warped Convolutions
(*Henriques et al., 2017)*
Polar Transformer Nets
(Esteves et al., 2017)



The Fourier Shift Theorem

Interestingly this generalises beyond 1D shifts to any transformation if considering groups

$$\int_{-\pi}^{\pi} I(\phi - \theta) e^{-\iota m \phi} \, \mathrm{d}\phi = \int_{-\pi-\theta}^{\pi-\theta} I(\phi') e^{-\iota m (\phi' + \theta)} \, \mathrm{d}\phi'$$

$$m \in \mathbb{Z} \qquad = \underbrace{e^{-\iota m \theta}}_{\text{Equivariant}} \underbrace{\int_{-\pi}^{\pi} I(\phi') \boxed{e^{-\iota m \phi'}} \, \mathrm{d}\phi'}_{\text{Invariant}}$$

Constraint on angular component of weights

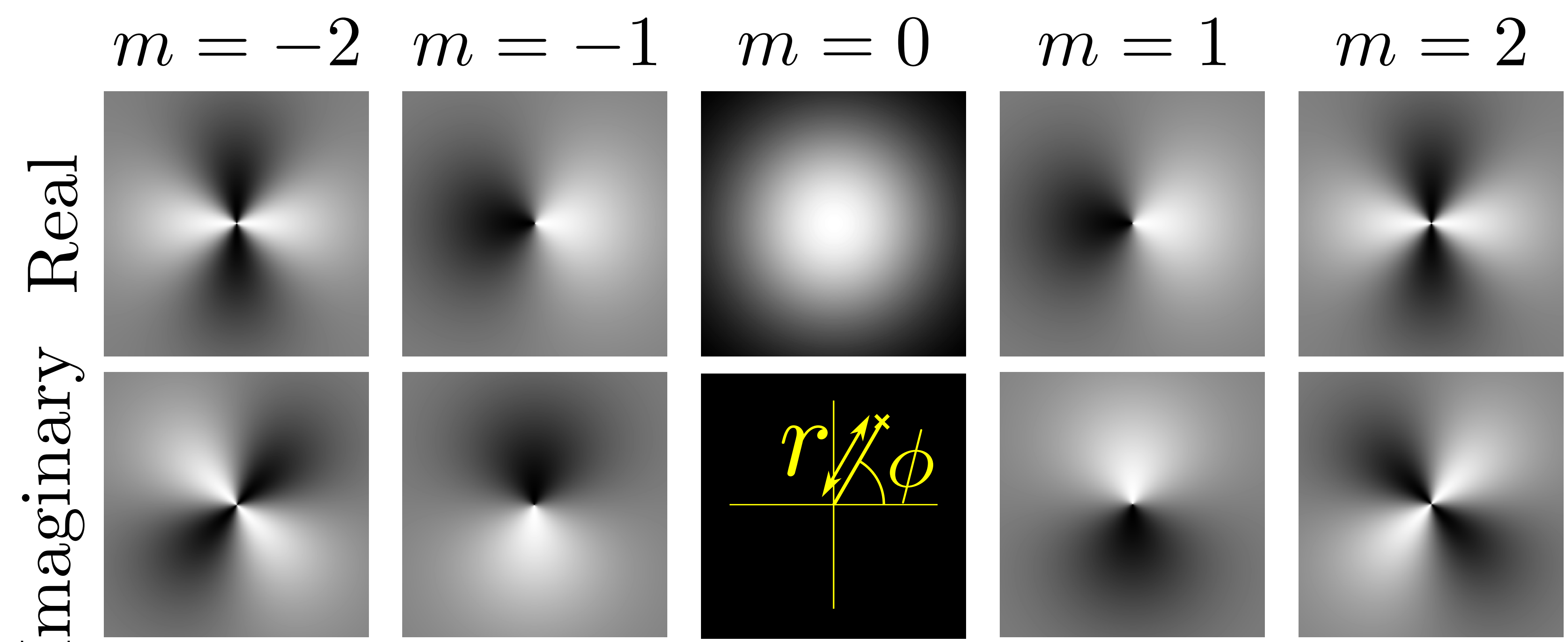
Circular Harmonics

Note: the sign flip in the angular component

Live in orthogonal spaces, so don't interact

$$W_m(r;\phi) = R(r)e^{\iota(m\phi+\beta)}$$

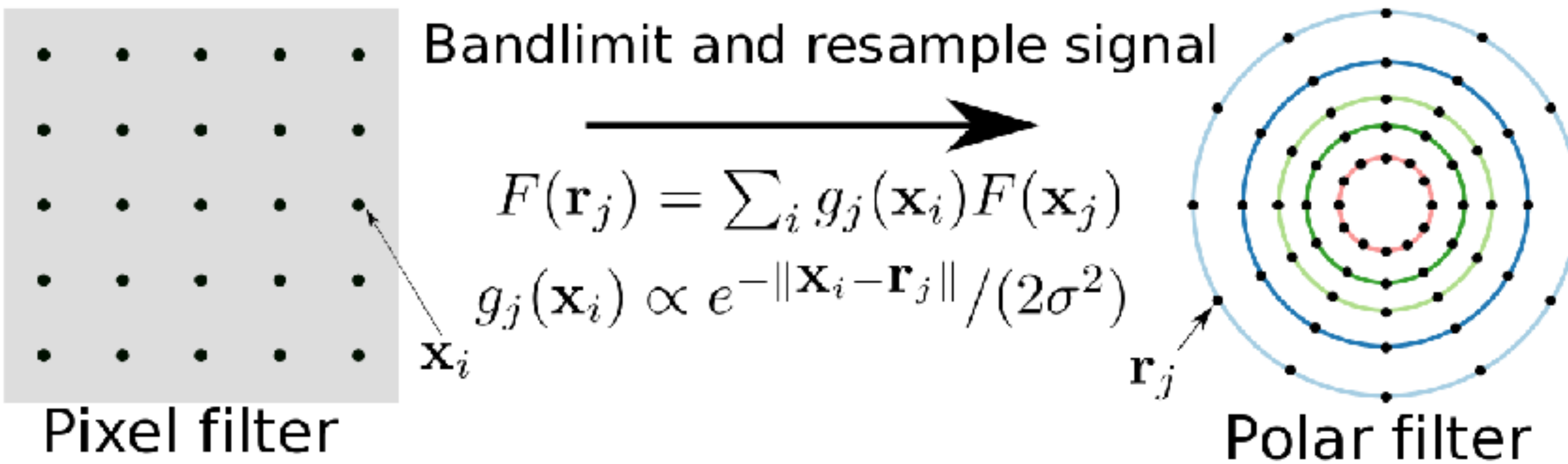
Free Constrained



Example with Gaussian radial component

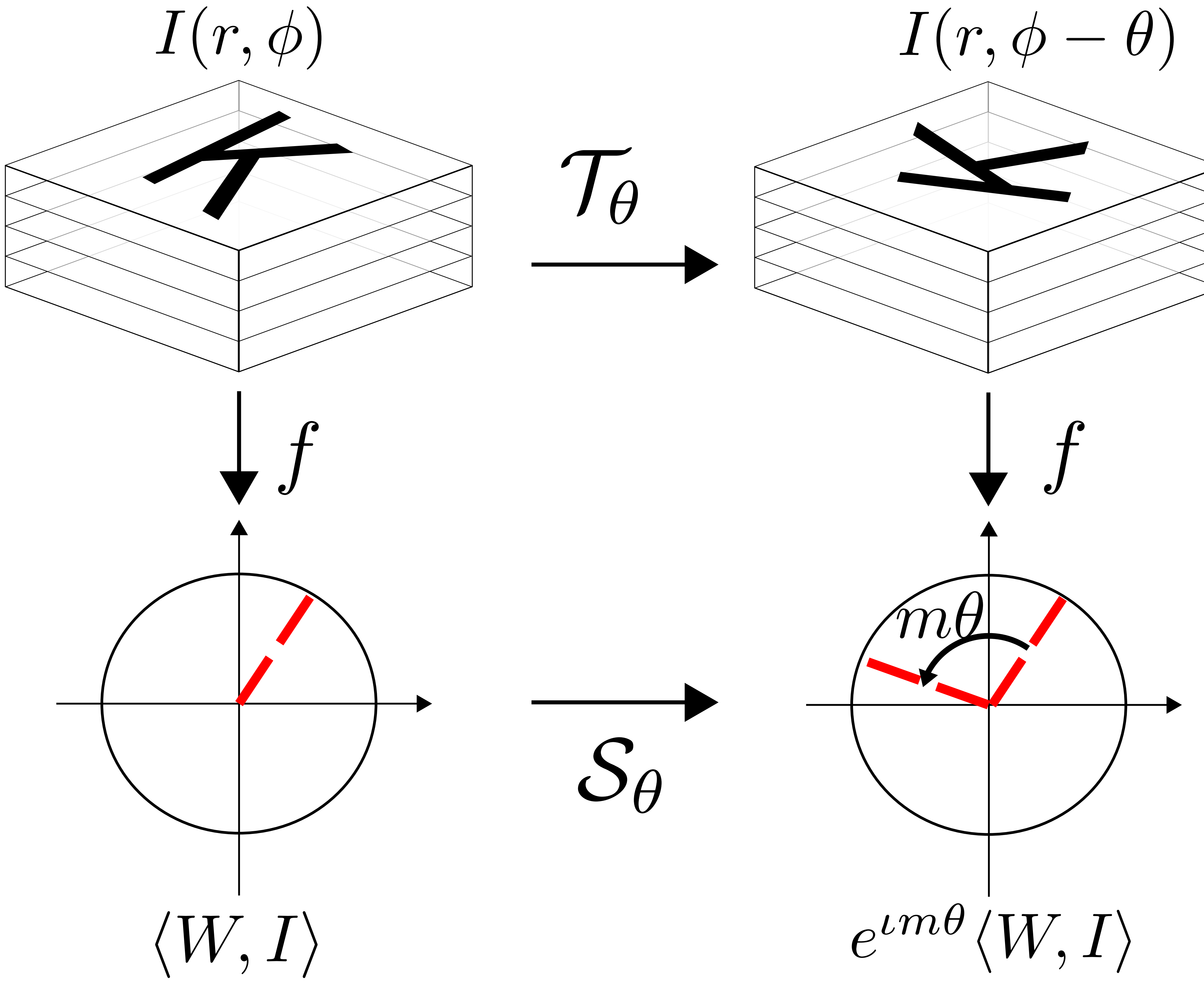
CIRCULAR HARMONIC WEIGHTS: IN PRACTICAL TERMS

Circular harmonics are defined in polar coordinates, so **resample** with Gaussian **anti-aliasing filter**.



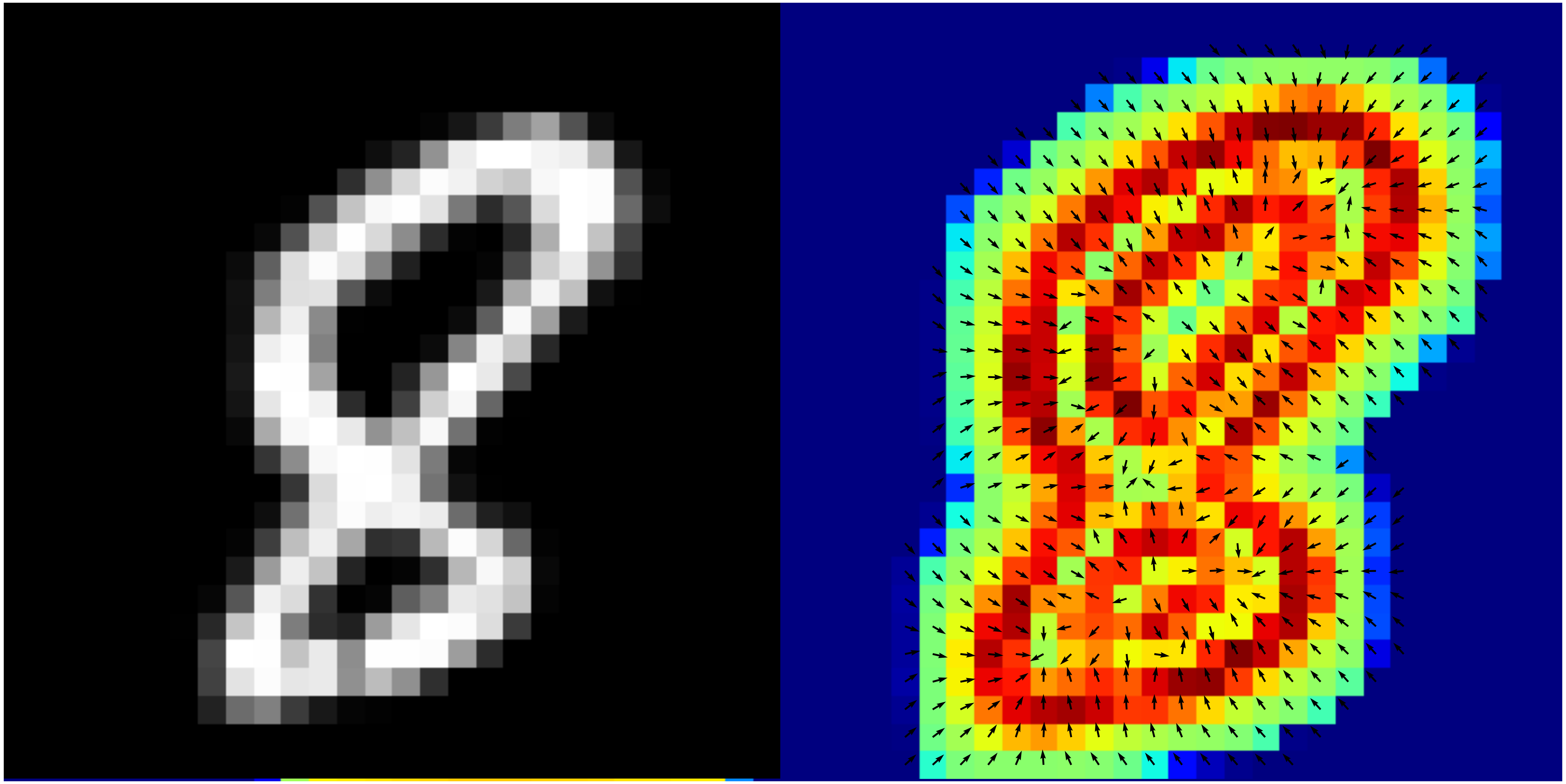
$$\mathbf{I} * \mathbf{W}_m = \underbrace{(\mathbf{I}^{\mathrm{Re}} * \mathbf{W}_m^{\mathrm{Re}} - \mathbf{I}^{\mathrm{Im}} * \mathbf{W}_m^{\mathrm{Im}})}_{\text{Real}} + \iota\underbrace{(\mathbf{I}^{\mathrm{Im}} * \mathbf{W}_m^{\mathrm{Re}} - \mathbf{I}^{\mathrm{Re}} * \mathbf{W}_m^{\mathrm{Im}})}_{\text{Imaginary}}$$

Circular Harmonic Weights



Circular Harmonics

X-correlation w. circular harmonics returns orientation field



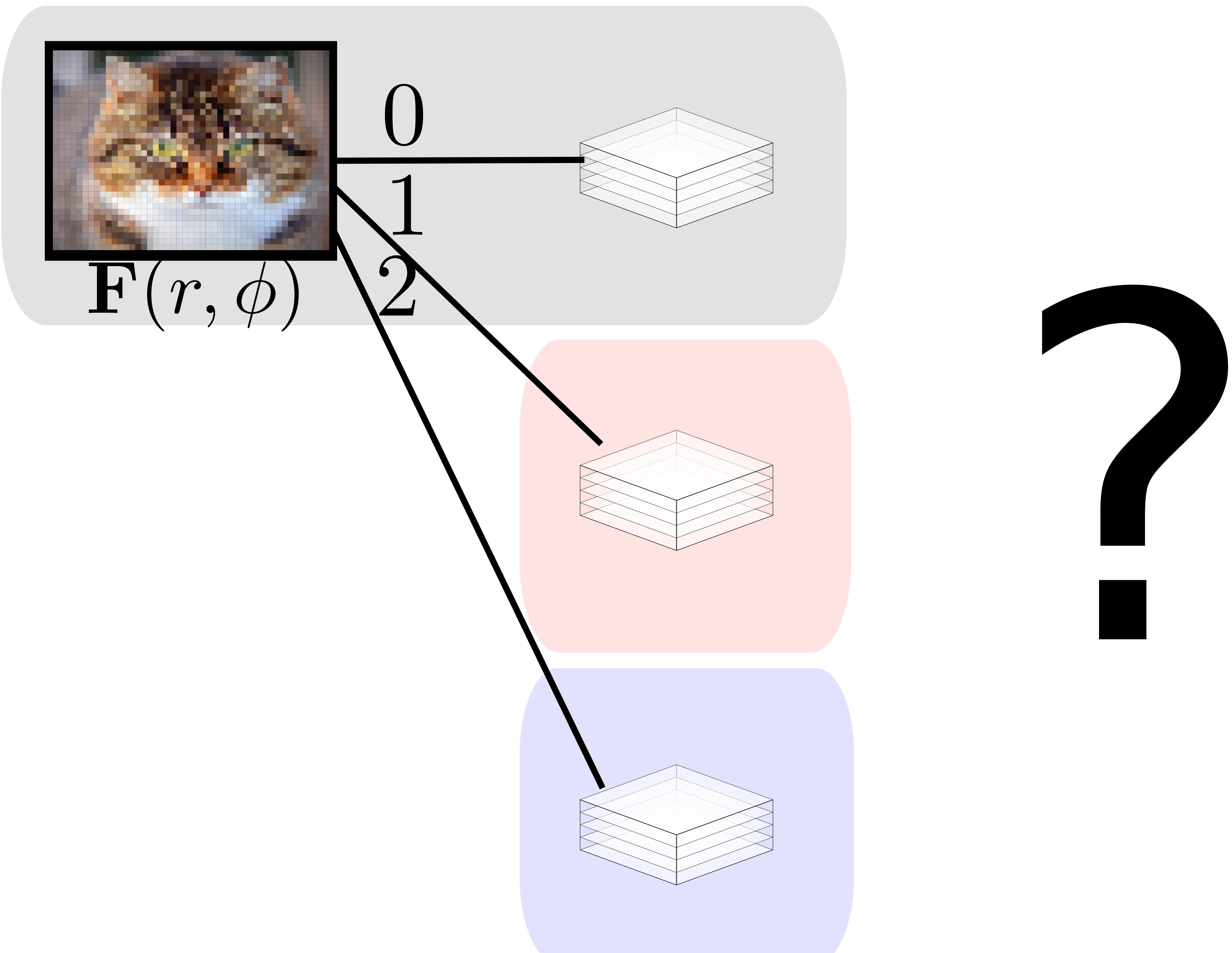
CIRCULAR HARMONIC ALGEBRA

So far we have a mapping $W\star : \mathbf{R}^{M\times N\times K} \to \mathbf{C}^{M\times N\times 1}$

But what about subsequent layers?

$$g(\mathcal{T}_\theta[\mathbf{I}] * \mathbf{W}_{m1}) * \mathbf{W}_{m_2}$$

What About Subsequent Layers?

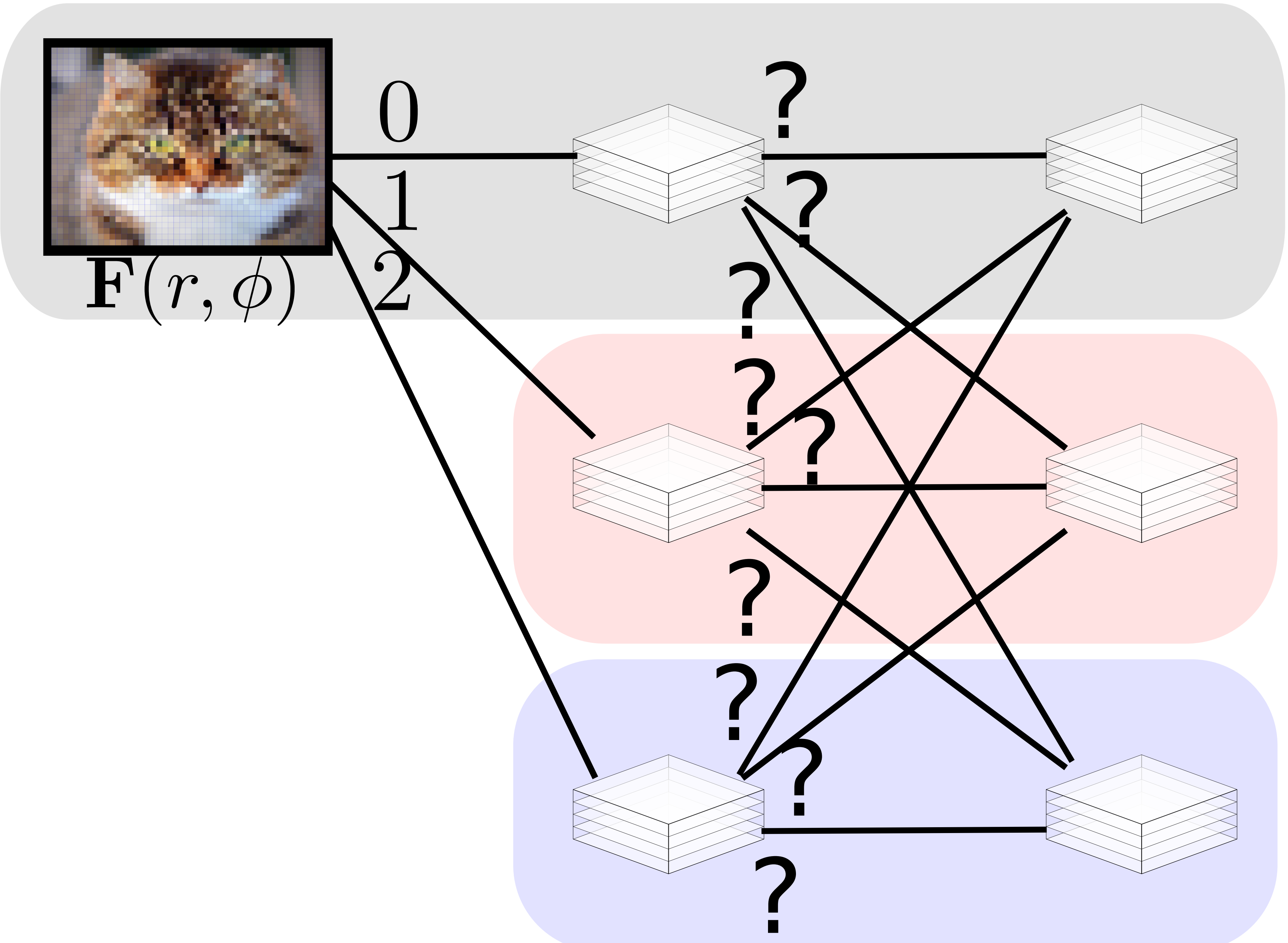


Circular Harmonic Algebra

Sum orders along chained cross-correlations

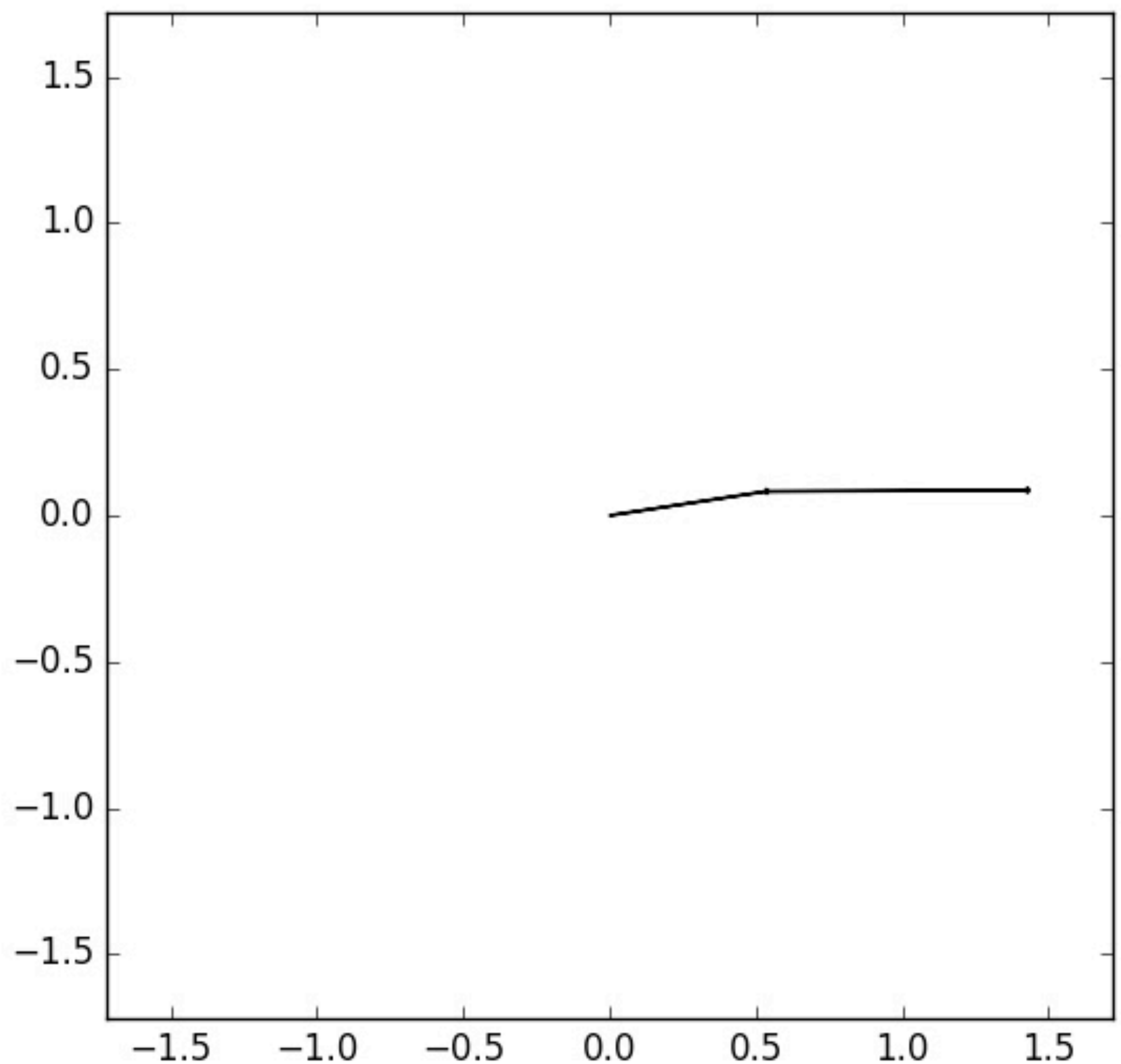
$$g(\mathcal{T}_\theta[\mathbf{I}] * \mathbf{W}_{m1}) * \mathbf{W}_{m_2} = Ye^{\iota(m_1+m_2)\theta}$$

Circular Harmonic Algebra



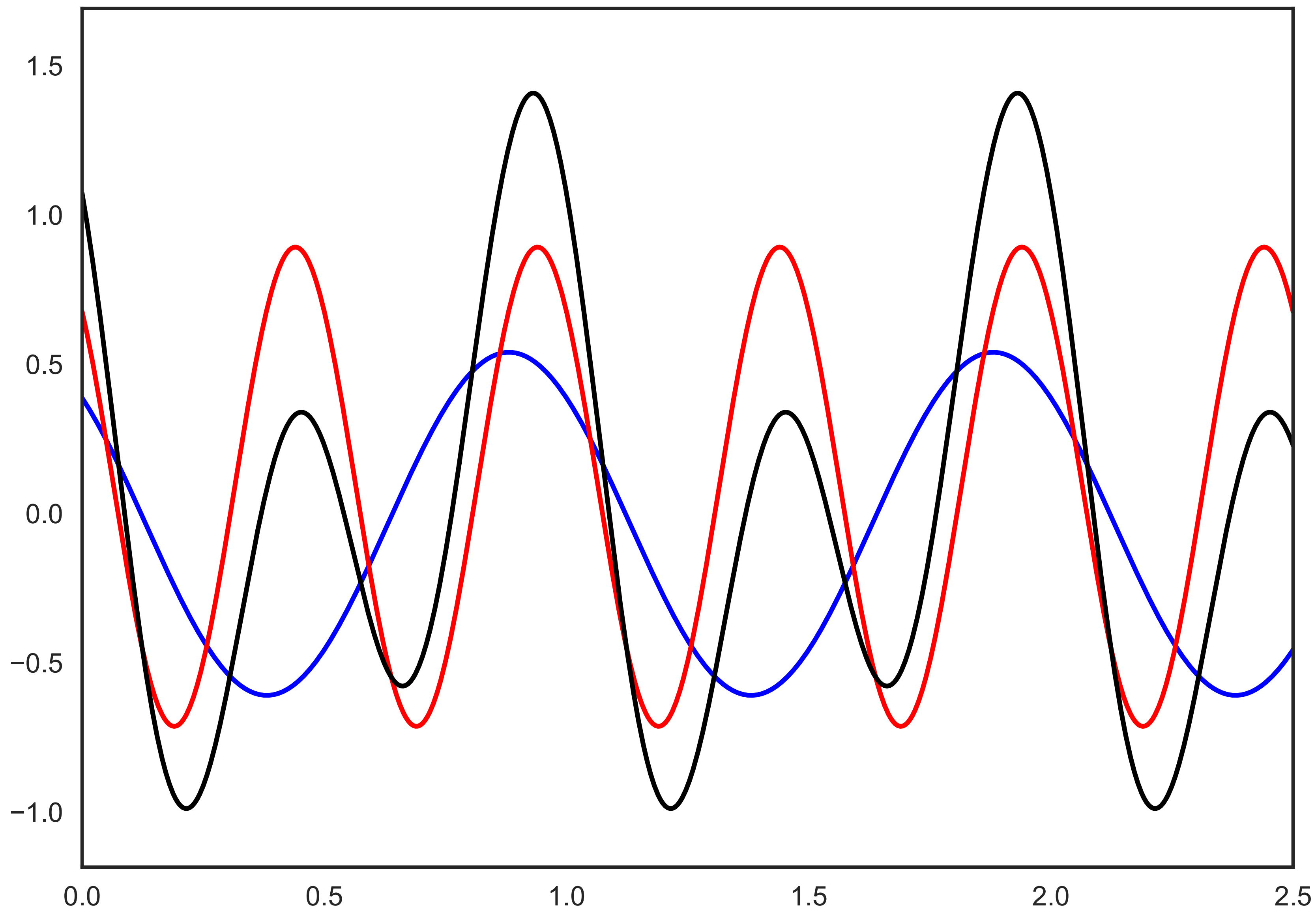
Circular Harmonic Algebra

At a layer, we sum over incoming feature maps, but need to pay special attention to rotation order



Circular Harmonic Algebra

At a layer, we sum over incoming feature maps, but need to pay special attention to rotation order



CIRCULAR HARMONIC ALGEBRA

The Equivariance/Coherence Condition

Rotation order of filter

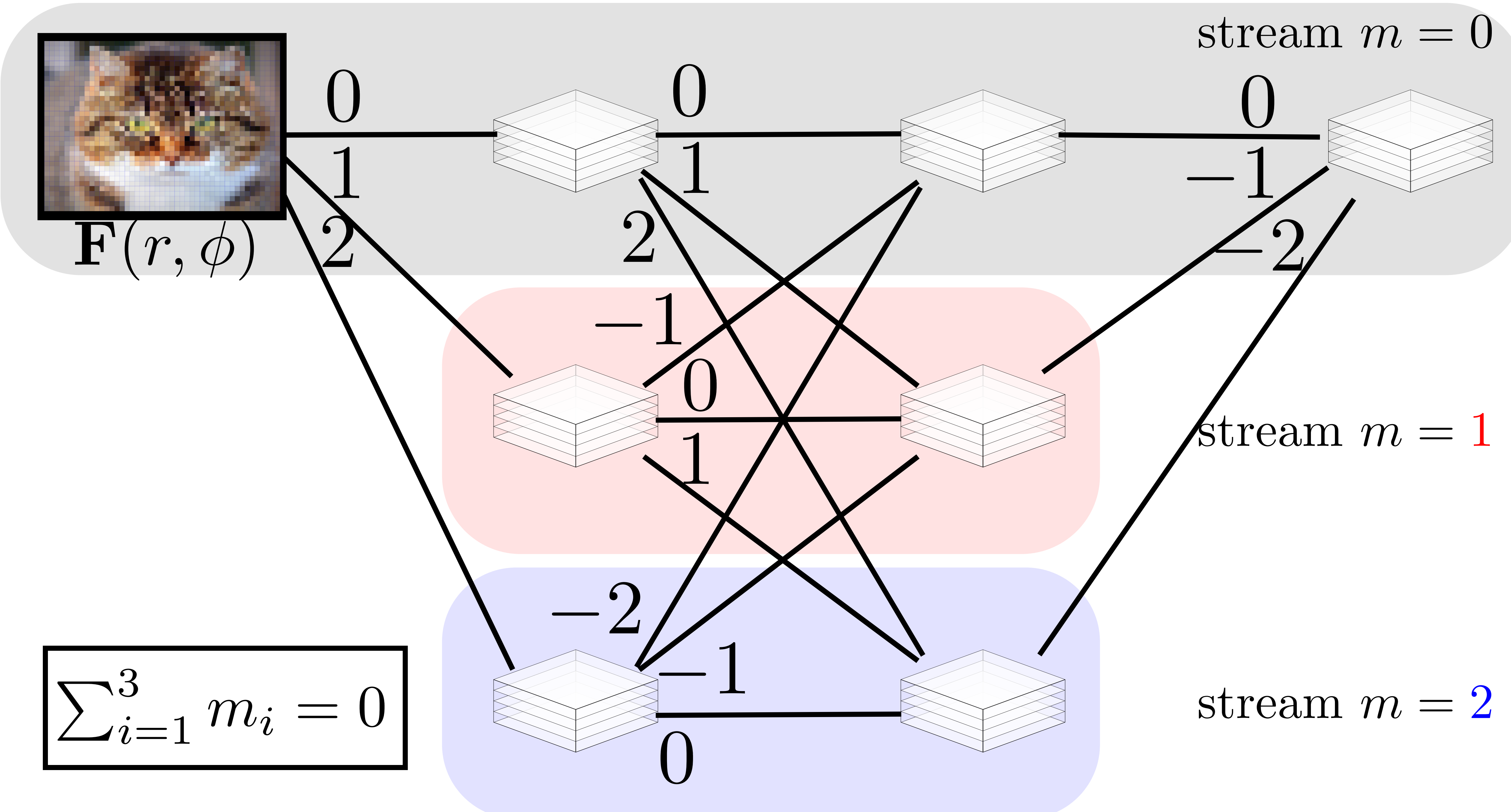
Constant at feature map i

$$\sum_{p \in \mathcal{P}_i} m_p = \kappa_i$$

Path through i

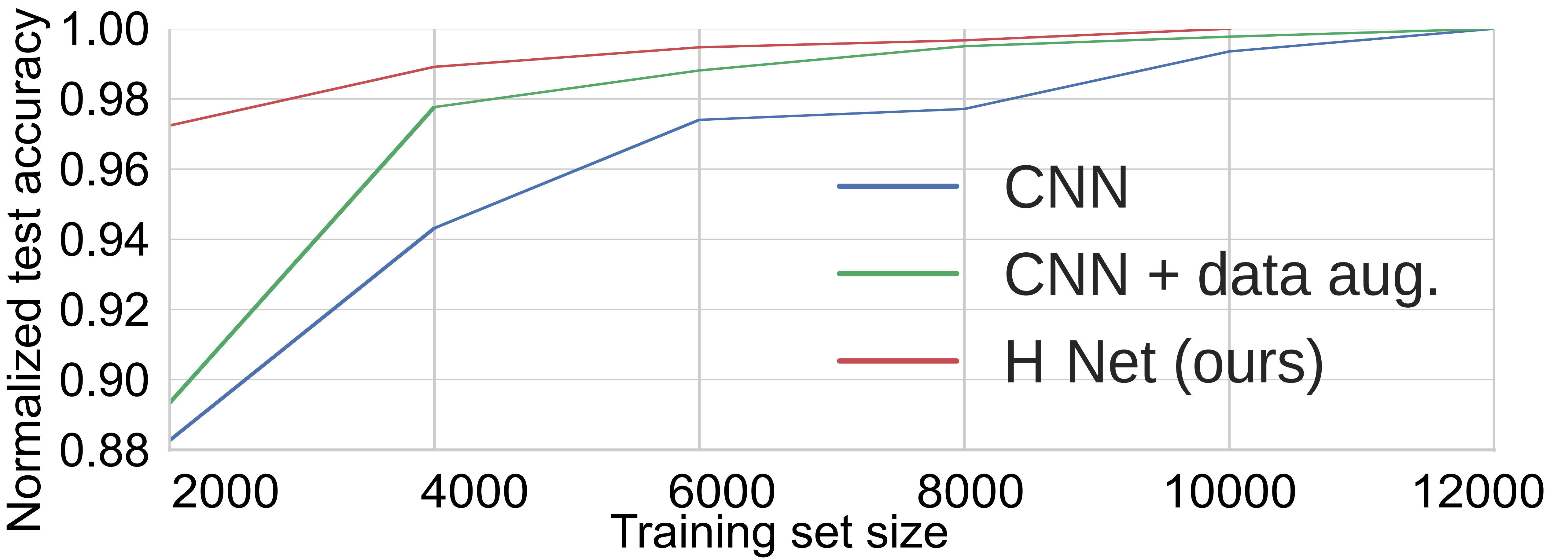
My conjecture: For Lie Groups, streams identified with basis vectors of Lie Algebra, and harmonic filters act as raising/lowering operators, jumping between basis vectors

CIRCULAR HARMONIC ALGEBRA



Harmonic Networks: Deep Rotation And Translation Equivariance, Worrall et al. (2017)

DATA ABLATION

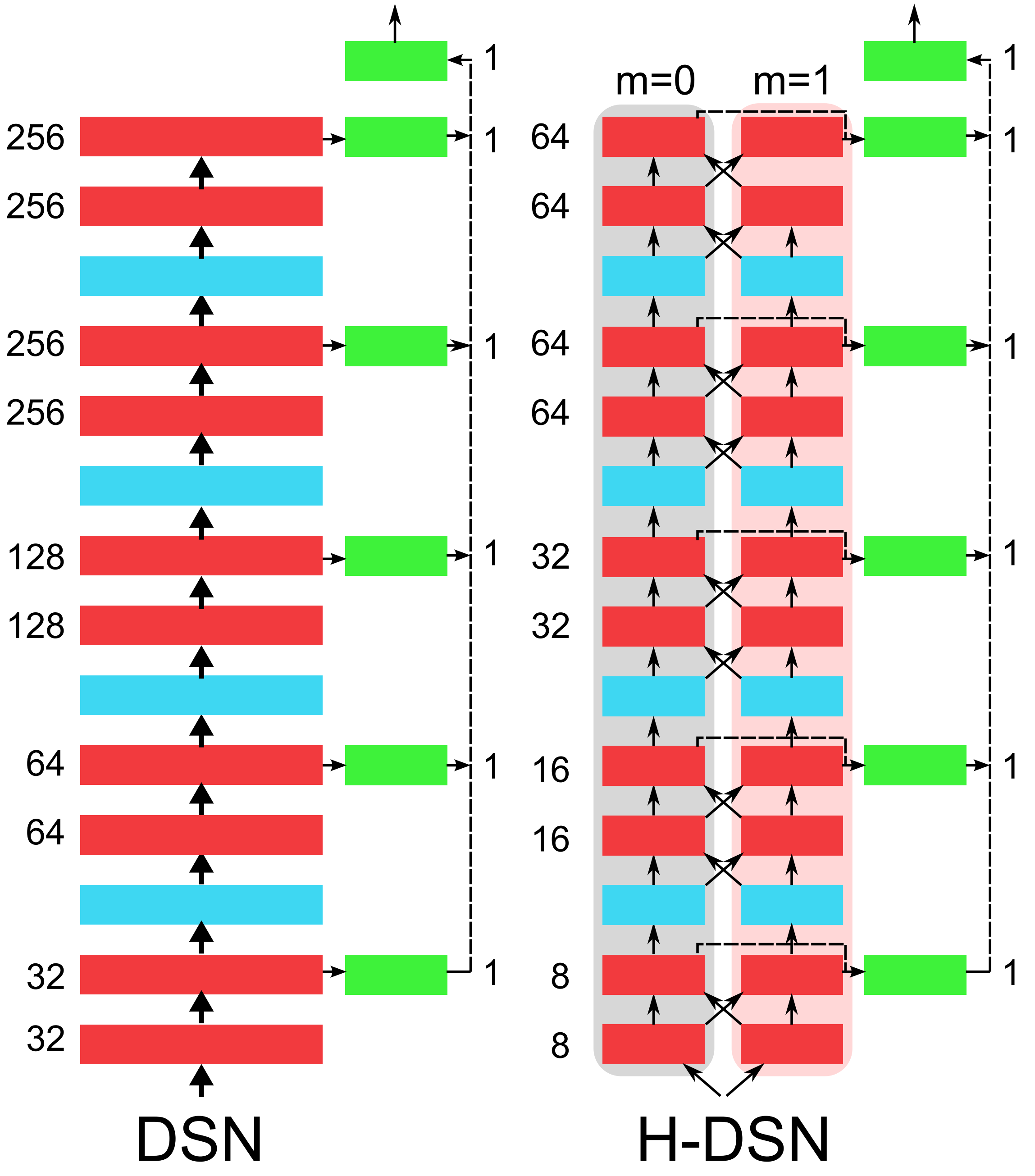


BOUNDARY SEGMENTATION

Boundary segmentation: BSD500

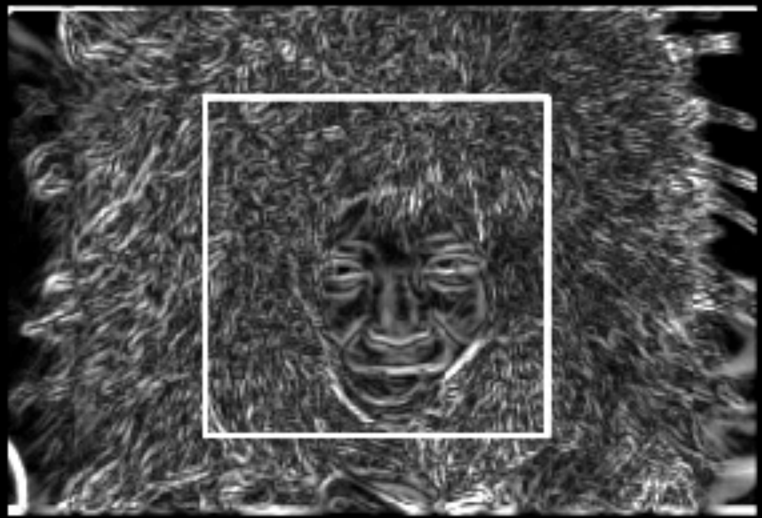
Recent results

Dynamic Steerable ResNet **0.732 0.751**



ORIENTED BOUNDARY SEGMENTATION

ROTATING FACE



ROTATING FACE

Follow Up Work

- G-convs and H-convs both in unified theory *The Quite General Theory of Equivariant Convolutional Networks Cohen et al., (under review), 2018*

- *3D Generalisation: Tensor Field Networks: Rotation- and Translation-Equivariant Neural Networks For 3D Point Clouds, Thomas et al., 2017*

OVERVIEW

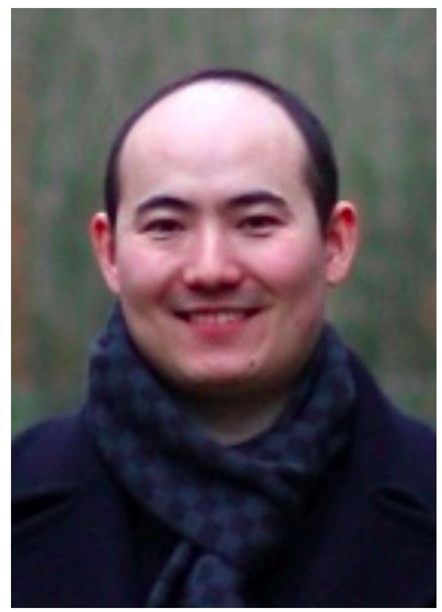
Future challenges

SUMMARY, LIMITATIONS, & FUTURE

- G-convs (good discriminativity, okay equivariance)
- H-convs (good equivariance, okay discriminativity)
- No generalisation guarantees as of yet
- Is group theory too restrictive? I think it is…
- Despite theory have only explored rotations and flips: *what about scale, affine transformation, occlusions…?*

COLLABORATORS

Stephan Garbin



Daniyar Turmukhambetov

Gabriel Brostow

Max Welling