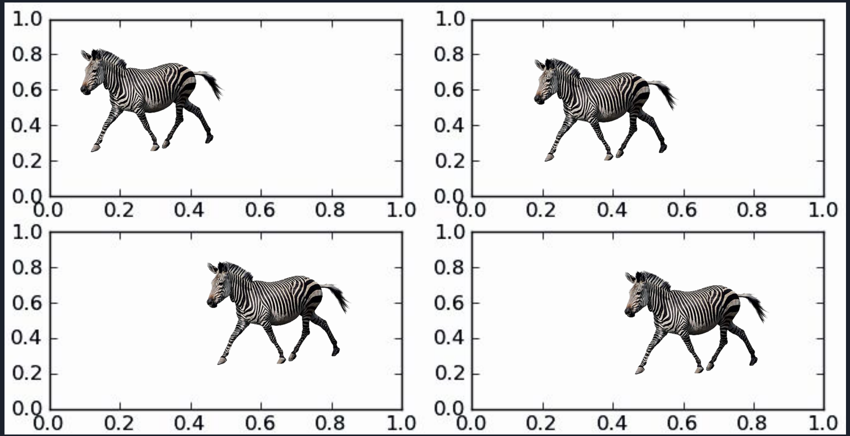
Harmonic neural networks

Why equivariance

- Data is limited
- Symmetries
  - Invariance: for a set of actions, the output shouldn't change
  - Example: translation group

Why equivariance

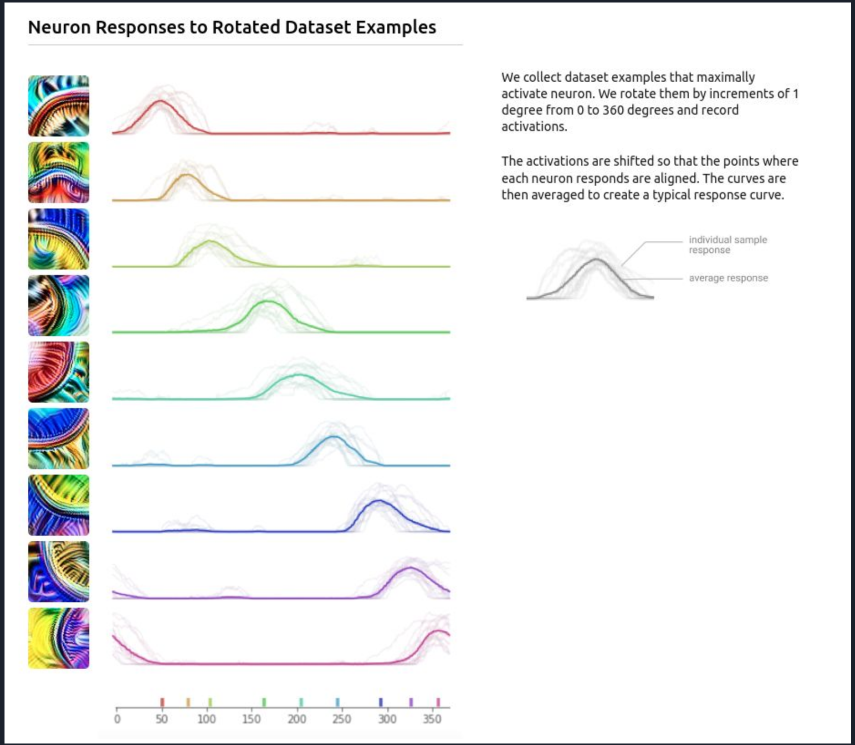
- Data is limited
- Symmetries
    - Invariance: for a set of actions, the output shouldn't change
    - Example: translation group
- Augmentation



Why equivariance

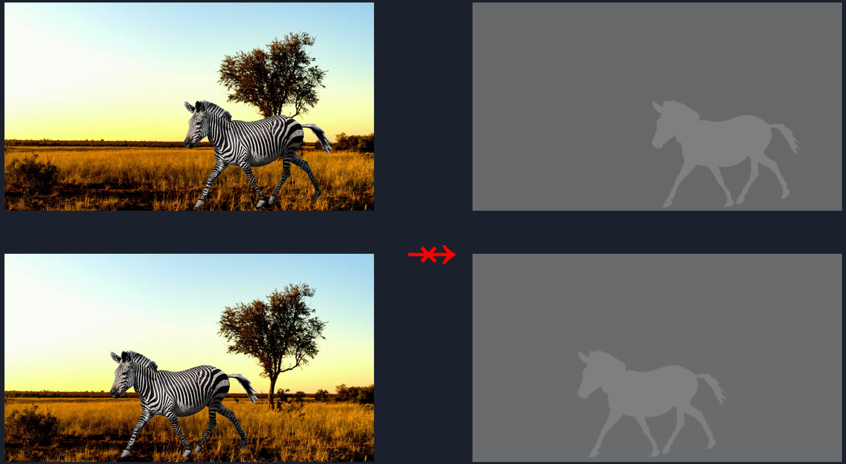
- Data is limited
- Symmetries
  - Invariance: for a set of actions, the output shouldn't change
  - Example: translation group
- Augmentation
  - Duplicated filters
    - Bad for interpretability!

[1]



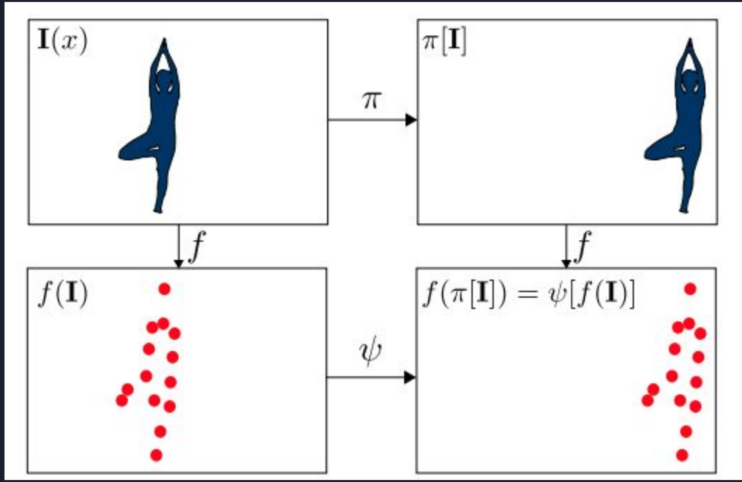
Why equivariance

- Data is limited
- Symmetries
    - Invariance: for a set of actions, the output shouldn't change
    - Example: translation group
- Augmentation
    - Duplicated filters
        - Bad for interpretability!
    - Features aren't equivariant



Why equivariance

- Data is limited
- Symmetries
  - Invariance: for a set of actions, the output shouldn't change
  - Example: translation group
- Augmentation
  - Duplicated filters
    - Bad for interpretability!
  - Features aren't equivariant
- Equivariance!
  - For a set of actions, the feature extractor and the action commute



…a feature mapping $f\colon X \to Y$ is equivariant to a group of transformations if we can associate every transformation $\pi \in \Pi$ of the input $x \in X$ with a transformation $\psi \in \Psi$ of the features; that is,

$$\psi[f(\mathbf{x})] = f(\pi[\mathbf{x}])$$

LaTeX at home:

we have LaTeX at home

Why equivariance

- Data is limited
- Symmetries
  - Invariance: for a set of actions, the output shouldn't change
  - Example: translation group
- Augmentation
  - Duplicated filters
    - Bad for interpretability!
  - Features aren't equivariant
- Equivariance!
  - For a set of actions, the feature extractor and the action commute
  - Better scaling laws!

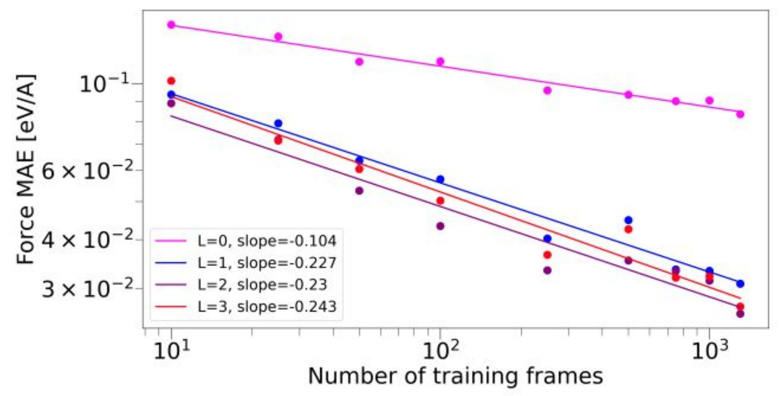
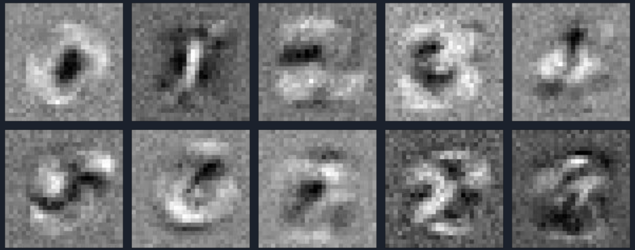


FIG. 7: Log-log plot of the predictive error on the water data set from [50] using NequIP with $l \in \{0, 1, 2, 3\}$ as a function of training set size, measured via the force MAE. The equivariant networks display a different scaling behavior than the invariant network.

[2]

Equivariance for linear layers

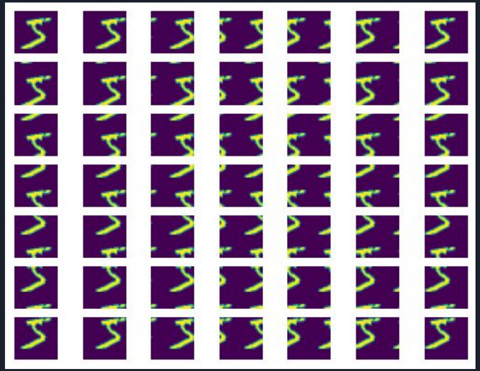
- Linear transformation

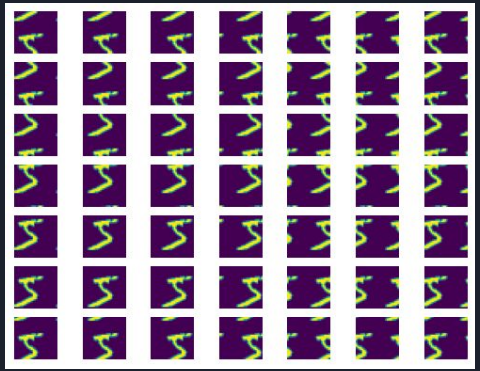


[3]

Equivariance for linear layers

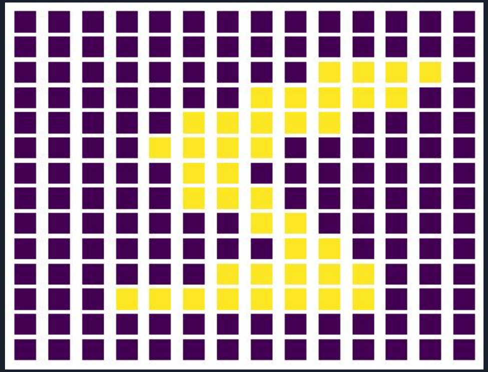
- Linear transformation
  - Lifted space!

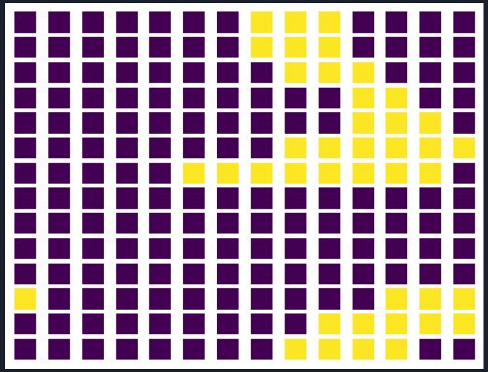




Equivariance for linear layers

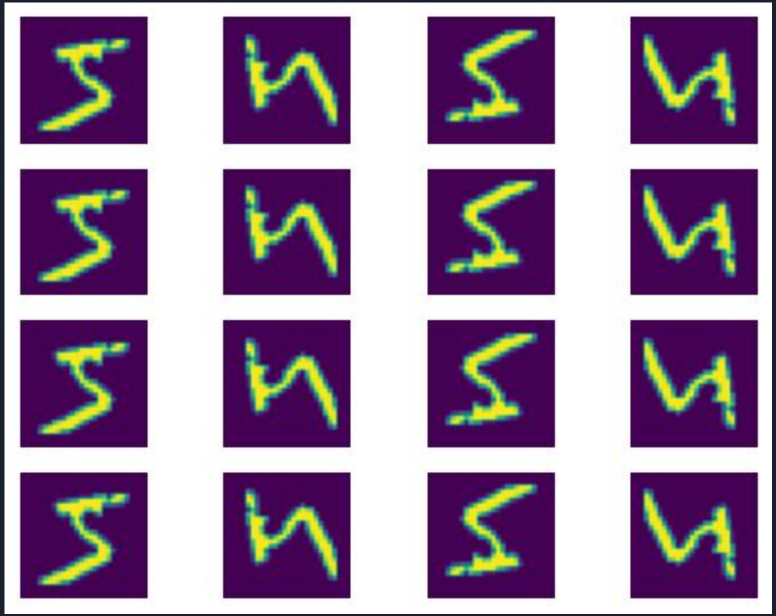
- Linear transformation
  - Lifted space!
  - 2D convolution





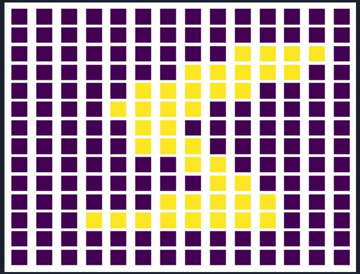
Equivariance for linear layers

- Linear transformation
    - Lifted space!
    - 2D convolution
- p4 group



Equivariance for linear layers

- Linear transformation
  - Lifted space!
  - 2D convolution
- p4 group



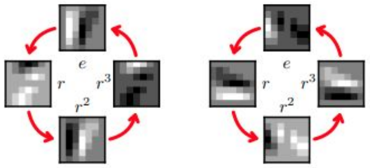


Figure 1. A p4 feature map and its rotation by r.

[4]

$$[f \star \psi](g) = \sum_{h \in G} \sum_{k} f_k(h) \psi_k(g^{-1}h).$$

Equivariance for linear layers

- Linear transformation
  - Lifted space!
  - 2D convolution
- p4 group
- All group equivariant linear transformations are group convolutions!

$$[f \star \psi](g) = \sum_{h \in G} \sum_{k} f_k(h) \psi_k(g^{-1}h).$$

Equivariance for linear layers

Proof. Since we are only interested in equivariant maps, we get a constraint on κ. For all $u, g \in G$:

$$\begin{aligned}
[\kappa \cdot [\pi_1(u) f]](g) &= [\pi_2(u)[\kappa \cdot f]](g) \\
\Leftrightarrow \quad \int_G \kappa(g, g') f(u^{-1} g') dg' &= \int_G \kappa(u^{-1} g, g') f(g') dg' \\
\Leftrightarrow \quad \int_G \kappa(g, u g') f(g') dg' &= \int_G \kappa(u^{-1} g, g') f(g') dg' \\
\Leftrightarrow \quad \kappa(g, u g') &= \kappa(u^{-1} g, g') \\
\Leftrightarrow \quad \kappa(u g, u g') &= \kappa(g, g')
\end{aligned} \tag{6}$$

Hence, without loss of generality, we can define the two-argument kernel $\kappa(\cdot, \cdot)$ in terms of a one-argument kernel: $\kappa(g^{-1} g') \equiv \kappa(e, g^{-1} g') = \kappa(ge, g g^{-1} g') = \kappa(g, g')$.

The application of κ to f thus reduces to a cross-correlation:

$$[\kappa \cdot f](g) = \int_G \kappa(g, g') f(g') dg' = \int_G \kappa(g^{-1} g') f(g') dg' = [\kappa \star f](g). \tag{7}$$

$\square$

[5]

Fourier transform

- Convolutions can be expensive!

Fourier transform

- Convolutions can be expensive!
- Use representations

$$\rho : G \to \mathrm{GL}(V)$$
$$\rho(st) = \rho(s)\rho(t)$$

Fourier transform

- Convolutions can be expensive!
- Use representations
  - Fourier transform;
  - Group convolution;
  - Convolution theorem;
  - Inverse Fourier transform

$$\widehat{f}(\varrho) = \sum_{a \in G} f(a)\varrho(a).$$

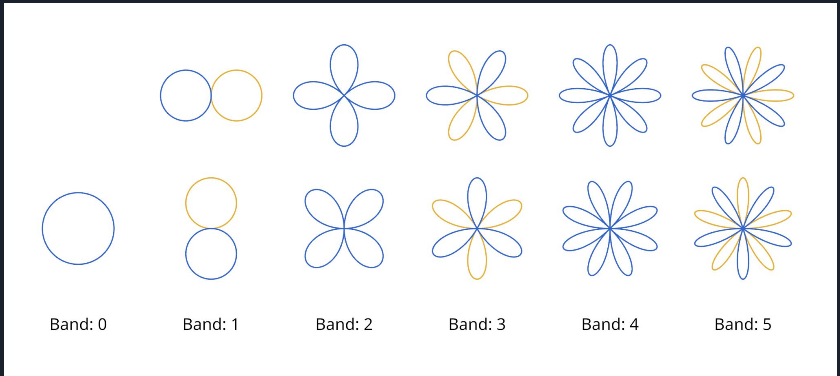
$$(f * g)(a) = \sum_{b \in G} f\left(ab^{-1}\right) g(b).$$

$$\widehat{f * g}(\varrho) = \hat{f}(\varrho)\hat{g}(\varrho).$$

$$f(a) = \frac{1}{|G|} \sum_i d_{\varrho_i} \operatorname{Tr}\left(\varrho_i(a^{-1})\widehat{f}(\varrho_i)\right).$$

Circular harmonics

- Fourier basis
    - Solutions to Laplace's equation
- Irreducible representation



[6]

$$m \in \mathbb{Z} : \mathbf{W}_m(x) = e^{2i\pi m x}$$

Circular harmonics

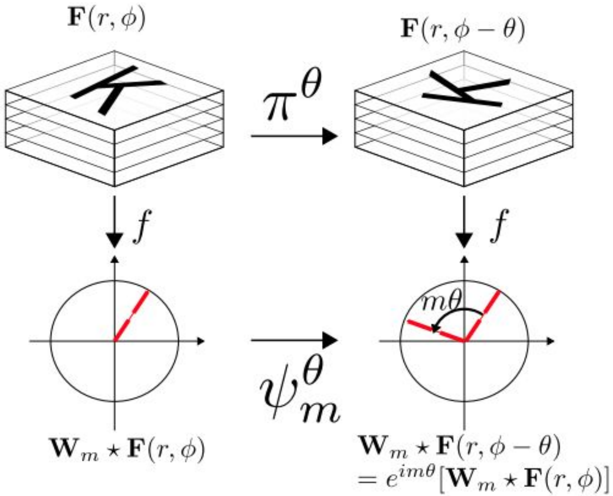


Figure 3. DOWN: Cross-correlation of the input patch with $\mathbf{W}_m$ yields a scalar complex-valued response. ACROSS-THEN-DOWN: Cross-correlation with the θ-rotated image yields another complex-valued response. BOTTOM: We transform from the unrotated response to the rotated response, through multiplication by $e^{im\theta}$.

Circular harmonics

Consider an image $\mathbf{F}(r,\phi)$, where r and ϕ are polar coordinates. We are interested in proving that there exists a filter $\mathbf{W}_m$, such that cross-correlation of $\mathbf{F}$ with $\mathbf{W}_m$ yields a rotationally equivariant feature map. We show that $\mathbf{W}_m$ takes the form of a circular harmonic, $\mathbf{W}_m = R(r)e^{i(m\phi+\beta)}$, for some $m \in \mathbb{Z}$. We begin from the fact that a rotation by θ about the origin leads to a new image $\mathbf{F}(r,\phi-\theta)$. In the notation of the paper, we also write this as $\mathbf{F}(r,\phi^\theta[\phi]) := \mathbf{F}(r,\phi-\theta)$. This means that the cross-correlation is

$$[\mathbf{W} \star \mathbf{F}(r,\pi^\theta[\phi])] = \int \mathbf{W}(r,\phi)\mathbf{F}(r,\pi^\theta[\phi])\,\mathrm{d}r\mathrm{d}\phi \tag{1}$$

$$= \int \mathbf{W}(r,\phi)\mathbf{F}(r,\phi-\theta)\,\mathrm{d}r\mathrm{d}\phi \tag{2}$$

$$= \int \mathbf{W}(r,\phi'+\theta)\mathbf{F}(r,\phi')\,\mathrm{d}r\mathrm{d}\phi', \tag{3}$$

$$\tag{4}$$

where we have used the substitution $\phi' := \phi - \theta$. If we assume that $\mathbf{W}_m$ is of the form $\mathbf{W}_m = R(r)e^{i(m\phi+\beta)}$, then the integral becomes

$$[\mathbf{W} \star \mathbf{F}(r,\pi^\theta[\phi])] = \int R(r)e^{i(m(\phi'+\theta)+\beta)}\mathbf{F}(r,\phi')\,\mathrm{d}r\mathrm{d}\phi' \tag{5}$$

$$= e^{im\theta}\int R(r)e^{i(m\phi'+\beta)}\mathbf{F}(r,\phi')\,\mathrm{d}r\mathrm{d}\phi' \tag{6}$$

$$= e^{im\theta}[\mathbf{W} \star \mathbf{F}(r,\pi^0[\phi])] = e^{im\theta}[\mathbf{W} \star \mathbf{F}(r,\phi)] \tag{7}$$

Thus we see that cross-correlation of the rotated signal $\mathbf{F}(r,\pi^\theta[\phi])$ with the circular harmonic filter $\mathbf{W}_m = R(r)e^{i(m\phi+\beta)}$ is equal to the response at zero rotation $[\mathbf{W} \star \mathbf{F}(r,\phi)]$, multiplied by a complex phase shift $e^{im\theta}$. In the notation of the paper, we denote this multiplication by $e^{im\theta}$ as $\psi_m^\theta[\bullet] = e^{im\theta}\cdot\bullet$. Thus cross-correlation with $\mathbf{W}_m$ yields a rotationally equivariant feature mapping.

Circular harmonics

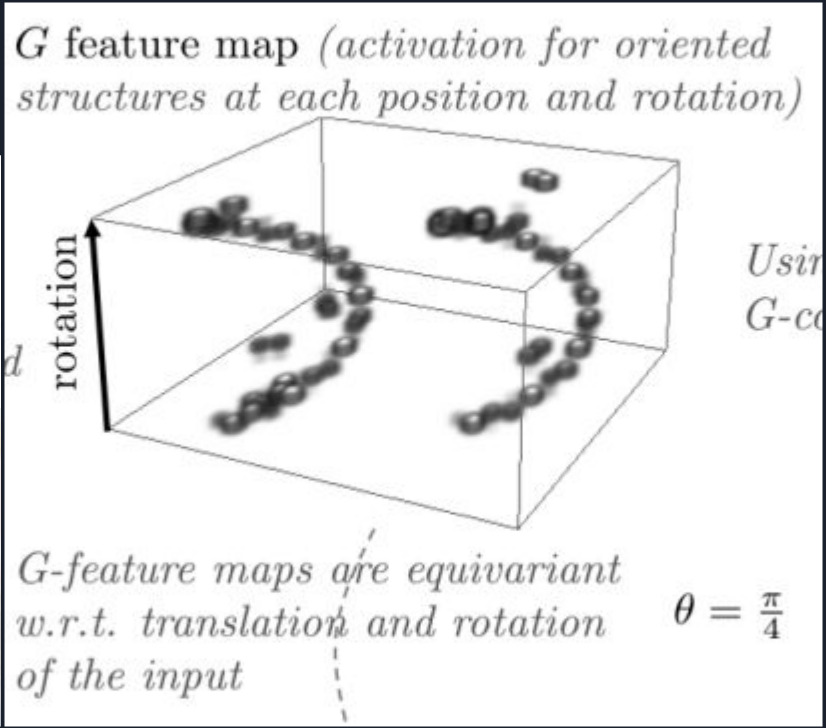
The summation of feature maps of the same rotation order is a new feature map of the same rotation order. Consider two feature maps $\mathbf{F}_1$ and $\mathbf{F}_2$ of rotation order m. Summation is a pointwise operation, so we only consider two corresponding points in the feature maps, which we denote $F_1 e^{i(m\theta+\beta_1)}$ and $F_2 e^{i(m\theta+\beta_2)}$, where β_1 and β_2 are phase offsets. The sum is

$$F_1 e^{i(m\theta+\beta_1)} + F_2 e^{i(m\theta+\beta_2)} = e^{im\theta}\left(F_1 e^{i\beta_1} + F_2 e^{i\beta_2}\right), \tag{13}$$

which for fixed $F_1, F_2, \beta_1, \beta_2$ is a function of m and θ only and so also rotationally equivariant with order m.

Harmonic networks

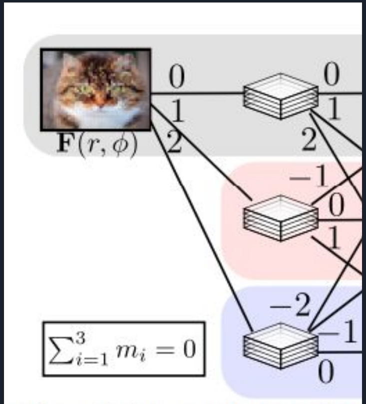
- Lifted representations



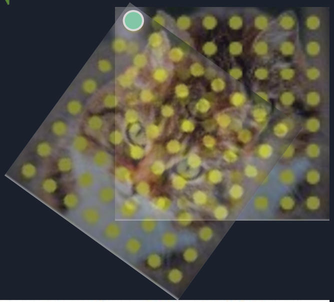
Harmonic networks

$$m \in \mathbb{Z}^+ : \mathbf{W}_m(x) = e^{2i\pi m x}$$

- Lifted representations
- HN does this indirectly



$$e^{2i\pi m\theta}$$



which is easier to analyze, but computationally infeasible. The introduction of the sampling defines *centers of equivariance* at pixel centers (yellow dots), about which a feature map is rotationally equivariant.

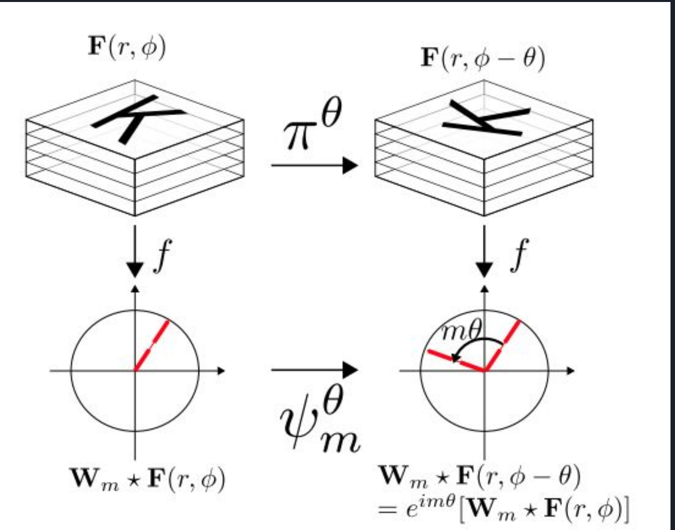


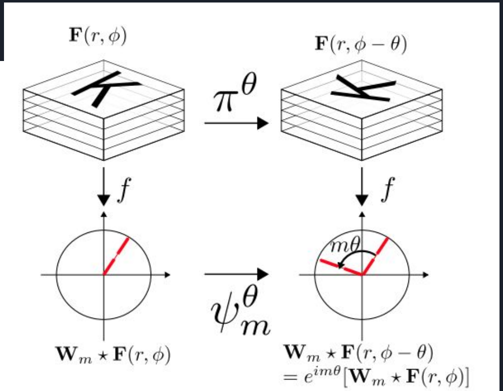
Figure 3. DOWN: Cross-correlation of the input patch with $\mathbf{W}_m$ yields a scalar complex-valued response. ACROSS-THEN-DOWN: Cross-correlation with the θ-rotated image yields another complex-valued response. BOTTOM: We transform from the unrotated response to the rotated response, through multiplication by $e^{im\theta}$.

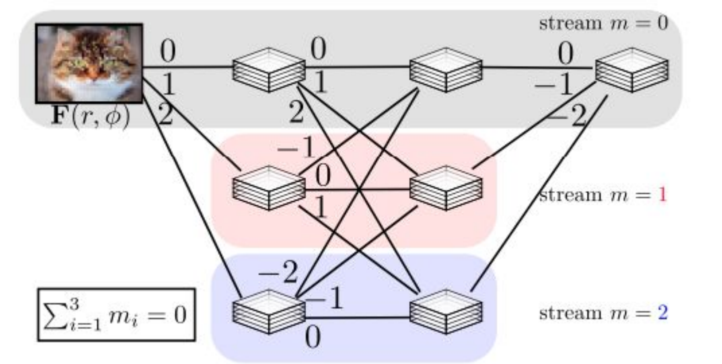
Circular harmonics

We claimed in **Arithmetic and Equivariance Condition**, that the rotation order of a feature map resulting from chained cross-correlations is equal to the sum of the the rotation orders of the filters in the chain. We prove this for a chain of two filters, and the rest follows by induction. Consider cross-correlation of a θ-rotated image $\mathbf{F}(r, \phi - \theta)$, with a filter $\mathbf{W}_m$, followed by cross-correlation with $\mathbf{W}_n$. We write the response of the first convolution as $e^{im\theta}[\mathbf{W}\star\mathbf{F}]$ when we are considering a single point in the plane. This is a function of θ only. If instead we considered the response at every point in the first layer feature map, then we would write the response as $e^{im\theta}[\mathbf{W}\star\mathbf{F}](\hat{r}, \hat{\phi} - \theta)$, where $\hat{r}, \hat{\phi}$ are the polar coordinates indexing the spatial dimensions of the first layer response. Then we can write the chained convolution as

$$\begin{aligned}
\mathbf{W}_n(\hat{r}, \hat{\phi}) \star [\mathbf{W}_m(r, \phi) \star \mathbf{F}(r, \phi - \theta)] &= \mathbf{W}_n(\hat{r}, \hat{\phi}) \star e^{im\theta}[\mathbf{W}_m \star \mathbf{F}](\hat{r}, \hat{\phi} - \theta) && (8)\\
&= e^{im\theta}\left(\mathbf{W}_n(\hat{r}, \hat{\phi}) \star [\mathbf{W}_m \star \mathbf{F}](\hat{r}, \hat{\phi} - \theta)\right) && (9)\\
&= e^{im\theta}e^{in\theta}\left[\mathbf{W}_n \star [\mathbf{W}_m \star \mathbf{F}]\right] && (10)\\
&= e^{i(m+n)\theta}\left[\mathbf{W}_n \star [\mathbf{W}_m \star \mathbf{F}]\right]. && (11)
\end{aligned}$$

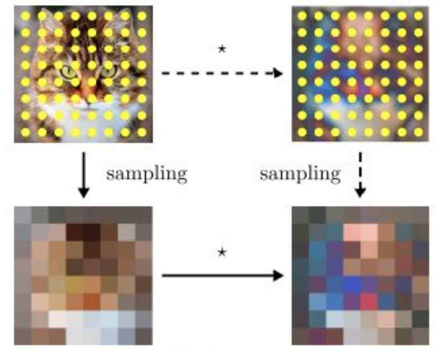
From the first to the second line we have used the property that the cross-correlation is linear and that we may pull the scalar factor $e^{im\theta}$ outside. Thus we see that the chained cross-correlation results in a summation of the rotation orders of the individual filters $\mathbf{W}_m$ and $\mathbf{W}_n$.





Harmonic networks

- Lifted representations
- HN does this indirectly



H-Net (magnitudes only). The solid arrows follow the path of the im-
plementation, while the dashed arrows follow the possible alternative,
which is easier to analyze, but computationally infeasible. The intro-

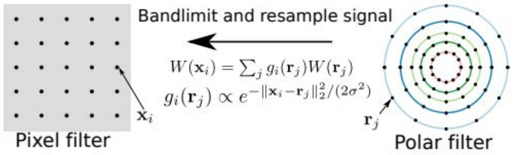


Figure 6. Images are sampled on a rectangular grid but our filters are defined in the polar domain, so we bandlimit and resample the data before cross-correlation via Gaussian resampling.

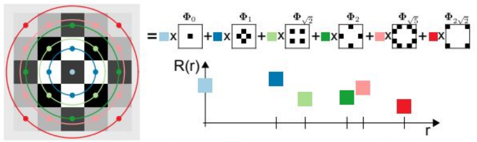
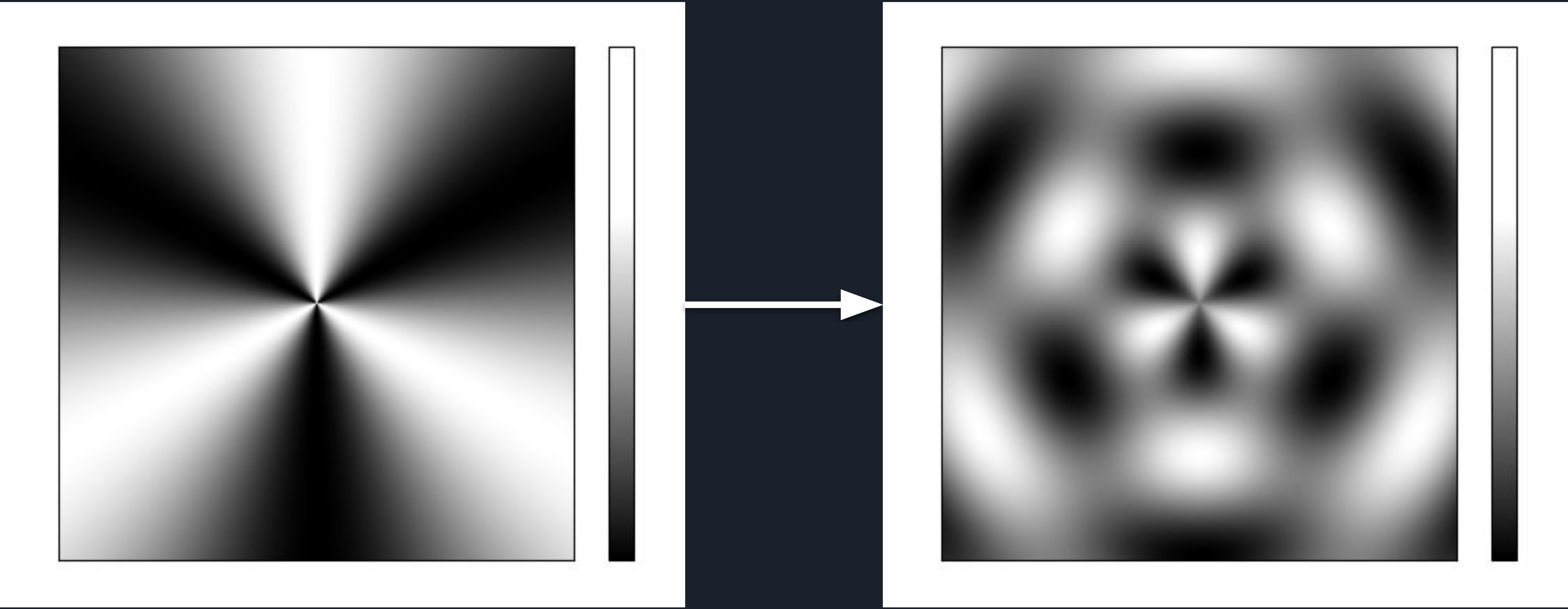


Figure 2. Mapping the radial profile to the elements of a square filter.

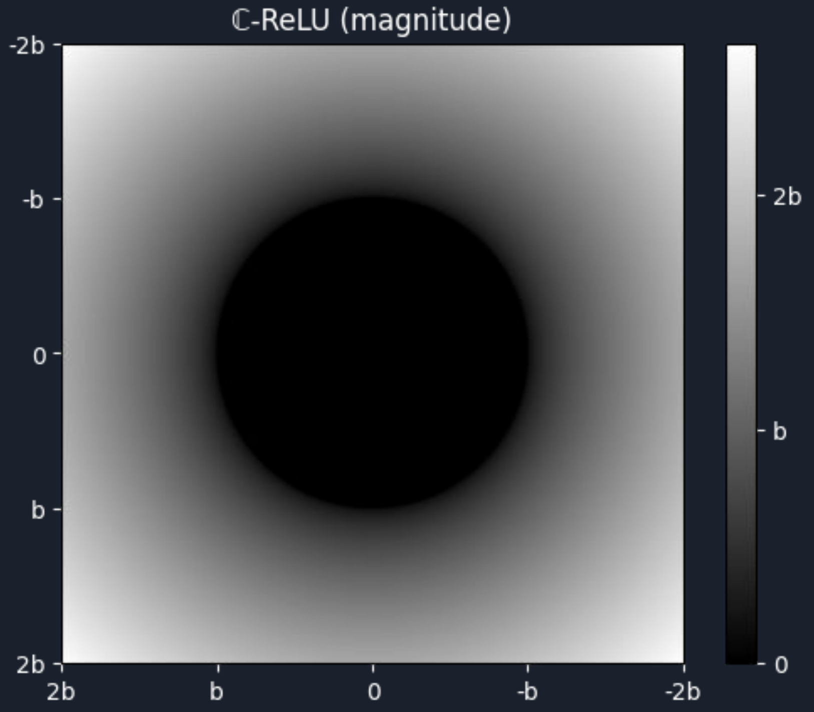
Filters



Activation functions

- ReLU messes with the phase
- ℂ-ReLU operates on magnitudes only

$\mathbb{C}\text{-ReLU}_b(Xe^{i\varphi})=\text{ReLU}(X+b)e^{i\varphi}$



Pooling

- No pooling!

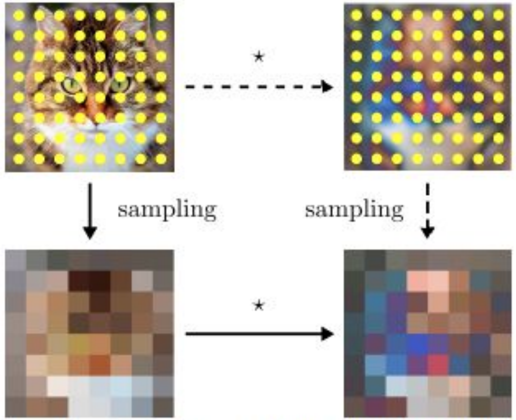


Figure 5. H-Nets operate in a continuous spatial domain, but we can implement them on pixel-domain data because sampling and cross-correlation commute. The schematic shows an example of a layer of an H-Net (magnitudes only). The solid arrows follow the path of the implementation, while the dashed arrows follow the possible alternative, which is easier to analyze, but computationally infeasible. The introduction of the sampling defines *centers of equivariance* at pixel centers (yellow dots), about which a feature map is rotationally equivariant.

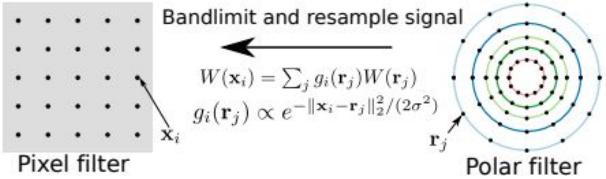


Figure 6. Images are sampled on a rectangular grid but our filters are defined in the polar domain, so we bandlimit and resample the data before cross-correlation via Gaussian resampling.

Conclusion

- More data-efficient
- Spherical harmonics

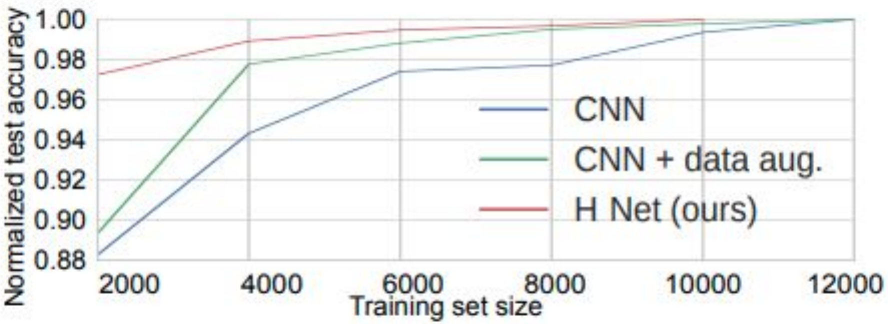


Figure 10. Data ablation study. On the rotated MNIST dataset, we experiment with test accuracy for varying sizes of the training set. We normalize the maximum test accuracy of each method to 1, for direct comparison of the falloff with training size. Clearly H-Nets are more data-efficient than regular CNNs, which need more data to discover rotational equivariance unaided.

Citations

1. Cammarata, et al., "Curve Detectors", Distill, 2020.
2. Batzner, S., Musaelian, A., Sun, L. et al., “E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials”, Nat Commun 13, 2453, 2022.
3. https://ml4a.github.io/ml4a/looking_inside_neural_nets/
4. Cohen, et al., "Group Equivariant Convolutional Networks", Proceedings of Machine Learning Research, 2016.
5. Cohen, Geiger, Weiler, “A General Theory of Equivariant CNNs on Homogeneous Spaces”, NeurIPS Proceedings, 2019.
6. https://valdes.cc/articles/ch.html
7. Worrall, et. al., “Harmonic Networks: Deep Translation and Rotation Equivariance”, CVF Open Access, 2017.

Don’t go past this slide

There’s nothing here