Supplementary Material
Harmonic Networks: Deep Translation and Rotation Equivariance

Daniel E. Worrall, Stephan J. Garbin, Daniyar Turmukhambetov and Gabriel J. Brostow
{d.worrall, s.garbin, d.turmukhambetov, g.brostow}@cs.ucl.ac.uk
Computer Science Department, University College London.

Abstract

We include some proofs and derivations of the rotational equivariance properties of the circular harmonics, along with a demonstration of how we calculate the number of parameters for various network architectures.

1. Equivariance properties

In Section 3.2 we mentioned that cross-correlation with the circular harmonics is a $360°$-rotation equivariant feature transform. Here we provide the proof, and some of the properties mentioned in **Arithmetic and Equivariance Condition**.

1.1. Equivariance of the Circular Harmonics

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. The proof requires us to introduce two different kinds of transformation: rotation $\mathcal{R}$ and translation $\mathcal{T}$. To simplify the math, we use vector notation, so the spatial domain of the filter/image is $\mathbb{R}^2$. We write the filter as $\mathbf{W}_m(\mathbf{x})$ and image as $\mathbf{F}(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^2$. We define the transformation operators $\mathcal{R}_\theta$ and $\mathcal{T}_\mathbf{t}$, such that $\mathcal{R}_\theta \mathbf{F} = \mathbf{F}(\mathbf{R}_{-\theta}\mathbf{x})$ and $\mathcal{T}_\mathbf{t}\mathbf{F} = \mathbf{F}(\mathbf{x}-\mathbf{t})$, where $\mathbf{R}_\theta$ is a 2D rotation matrix for a θ counter-clockwise rotation. We introduce rotational cross-correlation $\star$. This is defined as

$$[\mathbf{W}_m \star \mathbf{F}] = \int_\Phi \int_R \mathbf{W}_m(r\mathbf{R}_\phi \hat{\mathbf{x}})\mathbf{F}(r\mathbf{R}_\phi \hat{\mathbf{x}}) \, \mathrm{d}r\mathrm{d}\phi, \tag{1}$$

where we have used the decomposition $\mathbf{x} = r\hat{\mathbf{x}}$, with $r = \|\mathbf{x}\|_2 \geq 0$ and $\hat{\mathbf{x}} = \mathbf{x}/r$. The rotational cross-correlation is performed about the origin of the image. If we rotate the image, then we have

$$\begin{aligned}
[\mathbf{W}_m \star \mathcal{R}_\theta \mathbf{F}] &= \int_\Phi \int_R \mathbf{W}_m(r\mathbf{R}_\phi \hat{\mathbf{x}})\mathbf{F}(r\mathbf{R}_\phi \mathbf{R}_{-\theta}\hat{\mathbf{x}}) \, \mathrm{d}r\mathrm{d}\phi & (2)\\
&= \int_\Phi \int_R \mathbf{W}_m(r\mathbf{R}_\phi \hat{\mathbf{x}})\mathbf{F}(r\mathbf{R}_{\phi-\theta}\hat{\mathbf{x}}) \, \mathrm{d}r\mathrm{d}\phi & (3)\\
&= \int_\Phi \int_R \mathbf{W}_m(r\mathbf{R}_{\phi'+\theta} \hat{\mathbf{x}})\mathbf{F}(r\mathbf{R}_{\phi'}\hat{\mathbf{x}}) \, \mathrm{d}r\mathrm{d}\phi'. & (4)
\end{aligned}$$

If we define $\mathbf{W}_m(\mathbf{x}) = \mathbf{W}_m(r\hat{\mathbf{x}}) = R(r)e^{i(m\phi+\beta)}$, where $\phi = \angle\hat{\mathbf{x}}$, then

$$\begin{aligned}
[\mathbf{W}_m \star \mathcal{R}_\theta \mathbf{F}] &= \int_\Phi \int_R \mathbf{W}_m(r\mathbf{R}_{\phi'+\theta} \hat{\mathbf{x}})\mathbf{F}(r\mathbf{R}_{\phi'}\hat{\mathbf{x}}) \, \mathrm{d}r\mathrm{d}\phi' & (5)\\
&= \int_\Phi \int_R R(r)e^{i(m(\phi'+\theta)+\beta)}\mathbf{F}(r\mathbf{R}_{\phi'}\hat{\mathbf{x}}) \, \mathrm{d}r\mathrm{d}\phi' & (6)\\
&= e^{im\theta} \int_\Phi \int_R R(r)e^{i(m\phi'+\beta)}\mathbf{F}(r\mathbf{R}_{\phi'}\hat{\mathbf{x}}) \, \mathrm{d}r\mathrm{d}\phi' & (7)\\
&= e^{im\theta}[\mathbf{W}_m \star \mathbf{F}]. & (8)
\end{aligned}$$

And so rotational cross-correlation is rotationally equivariant about the origin of rotation. In the next part, we build up to a result needed for proving the chained cross-correlation result.

Cross-correlation about t To perform the rotational cross-correlation about another point $\mathbf{t}$, we first have to translate the image such that $\mathbf{t}$ is the new origin, so $F_{\mathbf{t}}(\mathbf{x}) = \mathbf{F}(\mathbf{x} - \mathbf{t})$, then perform the rotational cross-correlation, so

$$[\mathbf{W}_m \star \mathcal{T}_{\mathbf{t}}\mathbf{F}] = [\mathbf{W}_m \star \mathbf{F}_{\mathbf{t}}] \tag{9}$$

$$= \int_\Phi \int_R \mathbf{W}_m(r\mathbf{R}_\phi \hat{\mathbf{x}})\mathbf{F}_{\mathbf{t}}(r\mathbf{R}_\phi \hat{\mathbf{x}}) \, \mathrm{d}r_x \mathrm{d}\phi \tag{10}$$

$$= \int_\Phi \int_R \mathbf{W}_m(r\mathbf{R}_\phi \hat{\mathbf{x}})\mathbf{F}(r\mathbf{R}_\phi \hat{\mathbf{x}} - \mathbf{t}) \, \mathrm{d}r_x \mathrm{d}\phi. \tag{11}$$

Cross-correlation about t with rotated F about t In general, for every $\mathbf{t}$ this expression returns a different value. The response of a θ-rotated image about $\mathbf{t}$ is then

$$[\mathbf{W}_m \star \mathcal{R}_\theta \mathcal{T}_{\mathbf{t}}\mathbf{F}] = [\mathbf{W}_m \star \mathcal{R}_\theta \mathbf{F}_{\mathbf{t}}] \tag{12}$$

$$= \int_\Phi \int_R \mathbf{W}_m(r\mathbf{R}_\phi \hat{\mathbf{x}})\mathbf{F}_{\mathbf{t}}(r\mathbf{R}_{-\theta}\mathbf{R}_\phi \hat{\mathbf{x}}) \, \mathrm{d}r_x \mathrm{d}\phi \tag{13}$$

$$= \int_\Phi \int_R \mathbf{W}_m(r\mathbf{R}_\phi \hat{\mathbf{x}})\mathbf{F}_{\mathbf{t}}(r\mathbf{R}_{\phi-\theta} \hat{\mathbf{x}}) \, \mathrm{d}r_x \mathrm{d}\phi \tag{14}$$

$$= \int_\Phi \int_R \mathbf{W}_m(r\mathbf{R}_{\phi'+\theta} \hat{\mathbf{x}})\mathbf{F}_{\mathbf{t}}(r\mathbf{R}_{\phi'} \hat{\mathbf{x}}) \, \mathrm{d}r_x \mathrm{d}\phi' \tag{15}$$

$$= e^{im\theta} \int_\Phi \int_{R_z} R(r_z) e^{i(m\phi'+\beta)} \mathbf{F}_{\mathbf{t}}(r\mathbf{R}_{\phi'} \hat{\mathbf{x}})) \, \mathrm{d}r \mathrm{d}\phi' \tag{16}$$

$$= e^{im\theta} [\mathbf{W}_m \star \mathcal{T}_{\mathbf{t}}\mathbf{F}]. \tag{17}$$

Cross-correlation about t with rotated F about origin Say we wish to perform the rotational cross-correlation about a point $\mathbf{t}$, when the image has been rotated about the origin. Denoting $\mathbf{F}^\theta = \mathcal{R}_\theta \mathbf{F}$, then the response is

$$[\mathbf{W}_m \star \mathcal{T}_{\mathbf{t}} \mathcal{R}_\theta \mathbf{F}] = [\mathbf{W}_m \star \mathcal{T}_{\mathbf{t}} \mathbf{F}^\theta] \tag{18}$$

$$= \int_\Phi \int_R \mathbf{W}_m(r\mathbf{R}_\phi \hat{\mathbf{x}})\mathbf{F}^\theta(r_x\mathbf{R}_\phi \hat{\mathbf{x}} - \mathbf{t}) \, \mathrm{d}r_x \mathrm{d}\phi \tag{19}$$

$$= \int_\Phi \int_R \mathbf{W}_m(r\mathbf{R}_\phi \hat{\mathbf{x}})\mathbf{F}(r_x\mathbf{R}_{-\theta}\mathbf{R}_\phi \hat{\mathbf{x}} - \mathbf{R}_{-\theta}\mathbf{t}) \, \mathrm{d}r_x \mathrm{d}\phi \tag{20}$$

$$= \int_\Phi \int_R \mathbf{W}_m(r\mathbf{R}_\phi \hat{\mathbf{x}})\mathbf{F}(r_x\mathbf{R}_{\phi-\theta} \hat{\mathbf{x}} - \mathbf{R}_{-\theta}\mathbf{t}) \, \mathrm{d}r_x \mathrm{d}\phi \tag{21}$$

$$= \int_\Phi \int_R \mathbf{W}_m(r\mathbf{R}_{\phi'+\theta} \hat{\mathbf{x}})\mathbf{F}(r_x\mathbf{R}_{\phi'} \hat{\mathbf{x}} - \mathbf{R}_{-\theta}\mathbf{t}) \, \mathrm{d}r_x \mathrm{d}\phi' \tag{22}$$

$$= e^{im\theta} \int_\Phi \int_R \mathbf{W}_m(r\mathbf{R}_{\phi'} \hat{\mathbf{x}})\mathbf{F}(r_x\mathbf{R}_{\phi'} \hat{\mathbf{x}} - \mathbf{R}_{-\theta}\mathbf{t}) \, \mathrm{d}r_x \mathrm{d}\phi' \tag{23}$$

$$= e^{im\theta} [\mathbf{W}_m \star \mathcal{T}_{\mathbf{R}_{-\theta}\mathbf{t}}\mathbf{F}]. \tag{24}$$

Thus we see that cross-correlation of the rotated signal $\mathbf{F}^\theta$ with the circular harmonic filter $\mathbf{W}_m = R(r)e^{i(m\phi+\beta)}$ is equal to the response at zero rotation $[\mathbf{W} \star \mathbf{F}]$, 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.

1.2. Properties

1.2.1 Chained cross-correlation

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 taking a θ-rotated image $\mathbf{F}$ about the origin, then cross-correlating it with a filter $\mathbf{W}_m$ as every point in the image plane $\mathbf{t} \in \mathbb{R}^2$, followed by cross-correlation with $\mathbf{W}_n$ as a point $\mathbf{s} \in \mathbb{R}^2$. We already know that the response to the rotation is $[\mathbf{W}_m \star \mathcal{T}_{\mathbf{t}}\mathcal{R}_\theta \mathbf{F}] = e^{im\theta}[\mathbf{W}_m \star \mathcal{T}_{\mathbf{R}_{-\theta}\mathbf{t}}\mathbf{F}]$, for all rotations θ of the input and all points $\mathbf{t}$ in the response plane, so we can write the chained convolution as

$$
\begin{aligned}
[\mathbf{W}_n \star \mathcal{T}_{\mathbf{s}}[\mathbf{W}_m \star \mathcal{T}_{\mathbf{t}}\mathcal{R}_\theta \mathbf{F}]] &= [\mathbf{W}_n \star \mathcal{T}_{\mathbf{s}} e^{im\theta}[\mathbf{W}_m \star \mathcal{T}_{\mathbf{R}_{-\theta}\mathbf{t}}\mathbf{F}]] && (25)\\
&= e^{im\theta}\left[\mathbf{W}_n \star \mathcal{T}_{\mathbf{s}}[\mathbf{W}_m \star \mathcal{T}_{\mathbf{R}_{-\theta}\mathbf{t}}\mathbf{F}]\right] && (26)
\end{aligned}
$$

We have used the property that the cross-correlation is linear and that we may pull the scalar factor $e^{im\theta}$ outside. If we write $\mathbf{G}(\mathbf{t}) = [\mathbf{W}_m \star \mathcal{T}_{\mathbf{t}}\mathbf{F}]$ then $[\mathbf{W}_m \star \mathcal{T}_{\mathbf{R}_{-\theta}\mathbf{t}}\mathbf{F}] = \mathbf{G}(\mathbf{R}_{-\theta}\mathbf{t}) = [\mathcal{R}_\theta \mathbf{G}](\mathbf{t})$, so

$$
\begin{aligned}
[\mathbf{W}_n \star \mathcal{T}_{\mathbf{s}}[\mathbf{W}_m \star \mathcal{T}_{\mathbf{t}}\mathcal{R}_\theta \mathbf{F}]] &= e^{im\theta}\left[\mathbf{W}_n \star \mathcal{T}_{\mathbf{s}}[\mathbf{W}_m \star \mathcal{T}_{\mathbf{R}_{-\theta}\mathbf{t}}\mathbf{F}]\right] && (27)\\
&= e^{im\theta}[\mathbf{W}_n \star \mathcal{T}_{\mathbf{s}}\mathcal{R}_\theta \mathbf{G}] && (28)\\
&= e^{im\theta}e^{in\theta}\left[\mathbf{W}_n \star \mathcal{T}_{\mathbf{R}_{-\theta}\mathbf{s}}\mathbf{G}\right]. && (29)
\end{aligned}
$$

Thus we see that the chained cross-correlation results in a summation of the rotation orders of the individual filters $\mathbf{W}_m$ and $\mathbf{W}_n$. Setting $\mathbf{s} = \mathbf{0}$, such that we evaluate the cross-correlation at the center of rotation, we regain an equation similar to 8.

1.2.2 Magnitude nonlinearities

Point-wise nonlinearities acting on the magnitude of a feature map maintain rotational equivariance. Consider that we have a point on a feature map of rotation order m, which we can write as $Fe^{im\theta}$, where $F \geq 0$ is the magnitude of the feature map and $e^{im\theta}$ is the phase component. The output of the nonlinearity $g : \mathbb{R}_+ \to \mathbb{R}$ is

$$
g(Fe^{im\theta}) = g(F)e^{im\theta}, \tag{30}
$$

since g only acts on magnitudes. Since for fixed F the output is a function of m and θ only, the point-wise magnitude-acting nonlinearity preserves rotational equivariance.

1.2.3 Summation of feature maps

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

$$
F_1 e^{i(m\theta+\beta_1)} + F_2 e^{i(m\theta+\beta_2)} = e^{im\theta}\left(F_1 e^{i\beta_1} + F_2 e^{i\beta_2}\right), \tag{31}
$$

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.

2. Number of parameters

Here we list a break down of how we computed the number of parameters for the various network architectures in the experiments section. The networks architectures used are in Figure 1. Red boxes are cross-correlations, blue boxes are pooling (average for H-Nets, max for regular CNNs), green boxes are 1×1-cross-correlations.

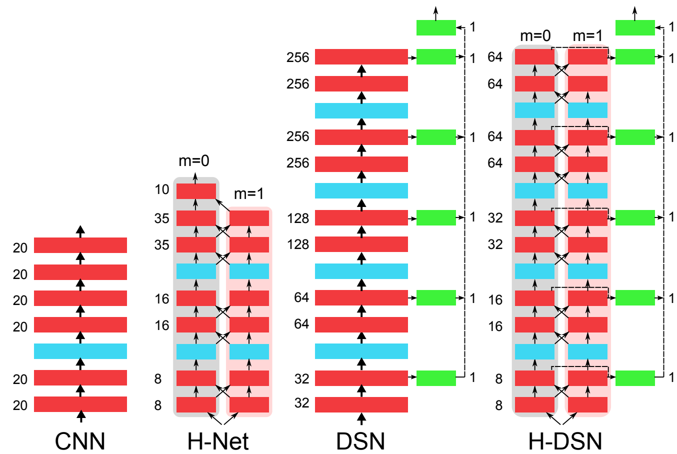

Figure 1. Networks used

2.1. Standard CNN

For a standard CNN layer with i input channels and o output channels, and $k \times k$ sized weights, the number of learnable parameters is iok^2. Since there is one bias per output layer, this increases to $iok^2 + o$. If using batch normalization, then there is an extra per-channel scaling factor, which increases the number of learnable parameters to $iok^2 + 2o$. The standard CNN for the rotated MNIST experiments has 6 layers of 3×3 cross-correlations, and 1 layer of 4×4-cross-correlations, with 20 feature maps per layer and 3 batch normalization layers so the number of learnable parameters is 21570. The calculations are shown in Table 1.

Layer	Weights	Batch Norm/Bias	#Params
1	$3 \cdot 3 \cdot 1 \cdot 20$	20	200
2	$3 \cdot 3 \cdot 20 \cdot 20$	$2 \cdot 20$	3640
3	$3 \cdot 3 \cdot 20 \cdot 20$	20	3620
4	$3 \cdot 3 \cdot 20 \cdot 20$	$2 \cdot 20$	3640
5	$3 \cdot 3 \cdot 20 \cdot 20$	20	3620
6	$3 \cdot 3 \cdot 20 \cdot 20$	$2 \cdot 20$	3640
7	$4 \cdot 4 \cdot 20 \cdot 10$	10	3210
Total			21570

Table 1. Number of parameters for a regular CNN.

2.2. Harmonic networks

The learnable parameters of a Harmonic Network are the radial profile and the the per-filter phase offset. For a $k \times k$ filter, the number of radial profile elements is equal to the number of rings of equal distance from the center of the filter. For example, consider the Figure 2, which is an excerpt from the main paper. This is a 5×5 filter, with 6 rings of equal distance from the center of the filter (the smallest ring is just a single point). So this filter has 6 radial profile terms and 1 phase offset to learn. This contrasts with a regular filter, which would have 25 learnable parameters. Note, that for filters with rotation order $m \neq 0$, the center pixel of the filter is in fact always zero, and so for a 5×5 rotation order $m \neq 0$ filter, the number of radial profile terms is $6 - 1 = 5$. So for the H-Net in the main paper with 5×5 filters and batch normalization in layers 2, 4, & 6, the number of learnable parameters is 33347. The calculations are in Table 2. Note that the final layer contains just one set of biases and no phase offsets. A similar set of calculations can be performed for the deeply supervised networks.

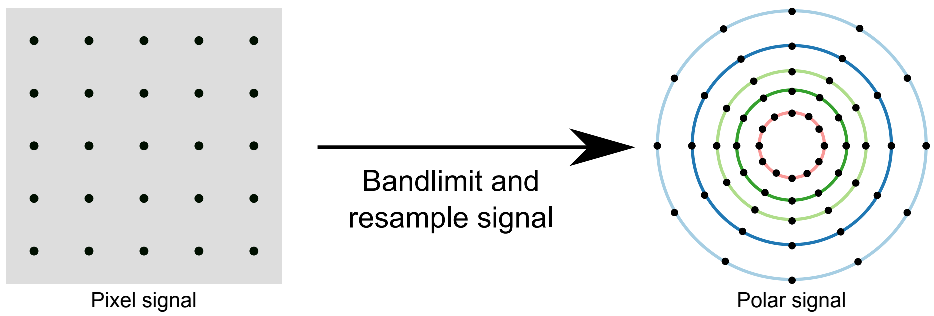

Figure 2. Each radius has a single learnable weight. Then there is a bias for the whole filter.

Layer	$m = 0$	$m = 1$	Batch Norm/Bias	#Params
1	$6 \cdot 1 \cdot 8 + 1 \cdot 8$	$5 \cdot 1 \cdot 8 + 1 \cdot 8$	$2 \cdot 8$	120
2	$6 \cdot 8 \cdot 8 + 8 \cdot 8$	$5 \cdot 8 \cdot 8 + 8 \cdot 8$	$2 \cdot 16$	864
3	$6 \cdot 8 \cdot 16 + 8 \cdot 16$	$5 \cdot 8 \cdot 16 + 8 \cdot 16$	$2 \cdot 16$	1696
4	$6 \cdot 16 \cdot 16 + 16 \cdot 16$	$5 \cdot 16 \cdot 16 + 16 \cdot 16$	$2 \cdot 32$	3392
5	$6 \cdot 16 \cdot 35 + 16 \cdot 35$	$5 \cdot 16 \cdot 35 + 16 \cdot 35$	$2 \cdot 35$	7350
6	$6 \cdot 35 \cdot 35 + 35 \cdot 35$	$5 \cdot 35 \cdot 35 + 35 \cdot 35$	$2 \cdot 70$	16065
7	$6 \cdot 35 \cdot 10$	$5 \cdot 35 \cdot 10$	10	3860
Total				33347

Table 2. Number of parameters for H-Net.