Harmonic Networks: Deep Rotation and Translation Equivariance

Daniel E. Worrall, Stephan J. Garbin, Daniyar Turmukhambetov, Gabriel J. Brostow
University College London

http://visual.cs.ucl.ac.uk/pubs/harmonicNets/index.html

1) What is equivariance and why you should use it

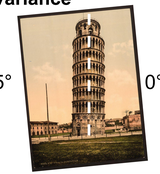
Rotational invariance

Dog

Dog (as well)

Rotational equivariance

5.5°



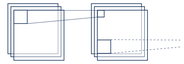
0°

feature-space transformation, input, black-box function, image transformation

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

Task: π is rotation, design equivariant layers f and feature-space transformation ψ

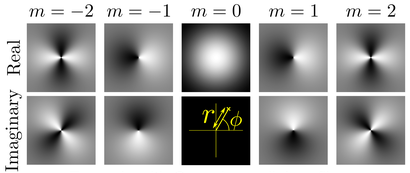
MLP 

CNN 

- Free-form weights
- Highly expressive
- Few data assumptions
- Need data augmentation
- Works on small images
- Translational weight-tying
- Fewer parameters
- Translationally equivariant
- No need to learn trans. equi.
- High performance on images

2) Our solution: Harmonic Networks

Convolve with learnable combination of circular harmonics to preserve rotational equivariance

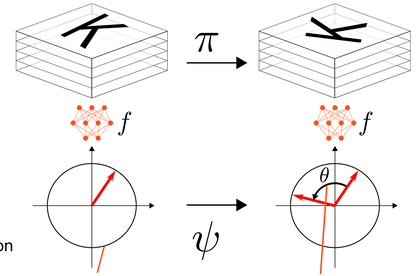


Example with Gaussian radial profile

$$\mathbf{W}_m(r,\phi) = R(r)e^{i(m\phi+\beta)}$$

$\sqrt{-1}$, Learnable filter orientation, Learnable radial profile, Frequency/rotation order

Complex-valued feature channels



Invariant magnitude, Equivariant orientation

Response rotates m times over input rotation

3) Results

Sanity check: MNIST-*rot*

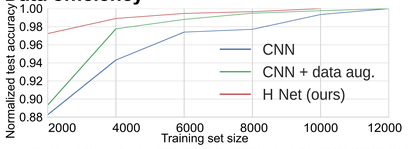
Method	Test error (%)	#params
SVM [1]	11.11	-
Transformation RBM [2]	4.2	-
Conv-RBM [3]	3.98	-
CNN [4]	5.03	22k
CNN + data aug* [4]	3.50	22k
P4CNN rotation pooling [4]	3.21	25k
P4CNN [4]	2.28	25k
Harmonic Networks (Ours)	1.69	33k

Boundary segmentation: BSD500

Feature magnitudes: w/o transfer learning

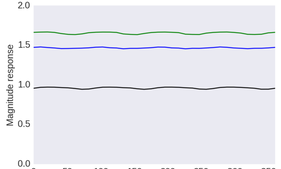
Method	ODS	OIS	#params
Holistically-nested edge detection (HED) [5]	0.640	0.650	2346k
HED (low # params) [5]	0.697	0.709	115k
Kivinen et al. [6]	0.702	0.715	-
Harmonic Networks (Ours)	**0.726**	**0.742**	116k

Data efficiency



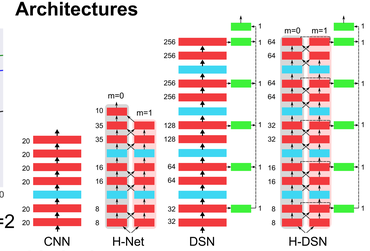
Feature orientation

Feature Stability



black: m=0, blue: m=1, green: m=2

Architectures



Pros

- Good generalization for low data
- Few parameters
- Interpretable features for rotation
- No rotational data augmentation

Limitations

- Only small--medium networks
- No noise rejection yet
- 4x computational cost
- Still need to learn other transformations

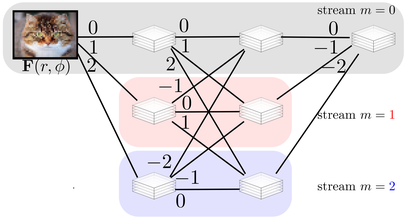
Future work:

- Include explicit noise rejection into convolution
- Consider other continuous groups (Lie Groups)

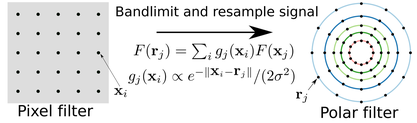
4) Details

The Equivariance Condition

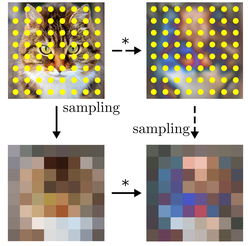
Chained convolution of rotation orders m_1 and m_2 leasd to response order m_1+m_2. At feature map, all incoming features must have same rotation order.



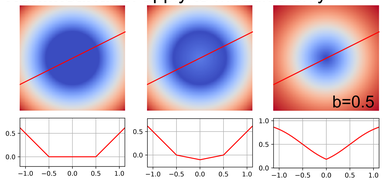
Circular harmonics are defined in polar coordinates, so **resample** with Gaussian **anti-aliasing filter**.



Resampling per-patch leads to **rotational equivariance about each pixel** (yellow dots).



Nonlinearities: apply to modulus only



$$g_b(|\mathbf{F}|e^{i\angle\mathbf{F}}) = g(|\mathbf{F}| + b)e^{i\angle\mathbf{F}}$$

real-valued bias

Computational cost

1 complex convolution

$$\mathbf{W}_m * \mathbf{F} = \mathbf{W}_m^{\mathrm{Re}} * \mathbf{F}^{\mathrm{Re}} - \mathbf{W}_m^{\mathrm{Im}} * \mathbf{F}^{\mathrm{Im}} + i(\mathbf{W}_m^{\mathrm{Re}} * \mathbf{F}^{\mathrm{Im}} + \mathbf{W}_m^{\mathrm{Im}} * \mathbf{F}^{\mathrm{Re}})$$

4 real convolutions

Harmonic convolution is bandlimited **Fourier transform** on circle. Rotation is translation on circle, so phase shift in the (Fourier) response.

$$\mathbf{F}(\theta - \phi) \overset{\mathrm{FT}}{\Longleftrightarrow} e^{i\omega\phi}\tilde{\mathbf{F}}(\omega)$$

This complexity is function of feature bandwidth

Our method: 360° equivariant convolutions:

- Computational complexity: O(#bandwidth)

Naïve rotational weight-tying:

- Computational complexity: O(#rotations)

Acknowledgements: Fight for Sight UK, Microsoft Research PhD Scholarship, EPSRC Nature Smart Cities EP/K503745/1, ENGAGE EP/K015664/1

[1] H. Larochelle, D. Erhan, A. C. Courville, J. Bergstra, and Y. Bengio. An empirical evaluation of deep architectures on problems with many factors of variation. In Machine Learning, Proceedings of the Twenty-Fourth International Conference (ICML 2007), Corvallis, Oregon, USA, June 20-24, 2007, pages 473–480
[2] K. Sohn and H. Lee. Learning invariant representations with local transformations. In Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012
[3] U. Schmidt and S. Roth. Learning rotation-aware features: From invariant priors to equivariant descriptors. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012, pages 2050–2057, 2012
[4] T. Cohen and M. Welling. Group equivariant convolutional networks. In Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016, pages 2990–2999, 2016. 2, 6, 7
[5] S. Xie and Z. Tu. Holistically-nested edge detection. In 2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015 [33], pages 1395–1403
[6] J. J. Kivinen, C. K. I. Williams, and N. Heess. Visual boundary prediction: A deep neural prediction network and quality dissection. In Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics, AISTATS 2014, Reykjavik, Iceland, April 22-25, 2014 [14], pages 512–521