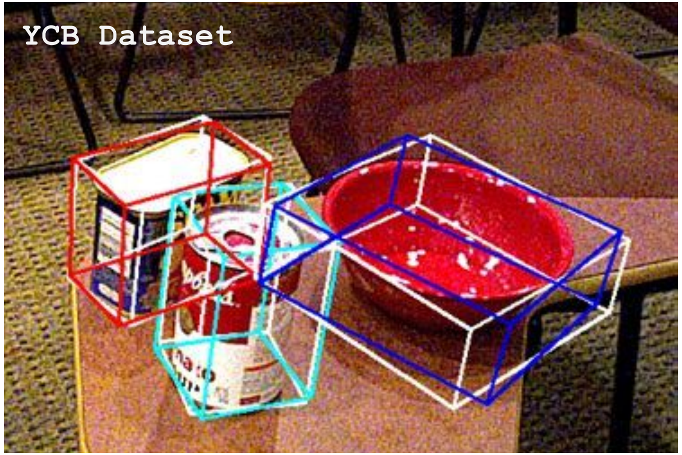
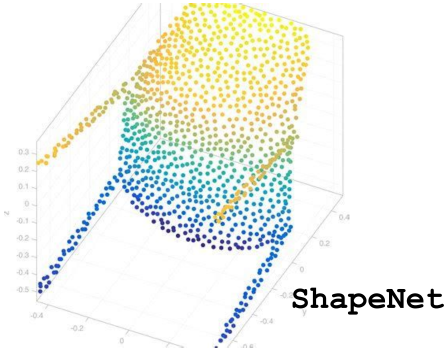
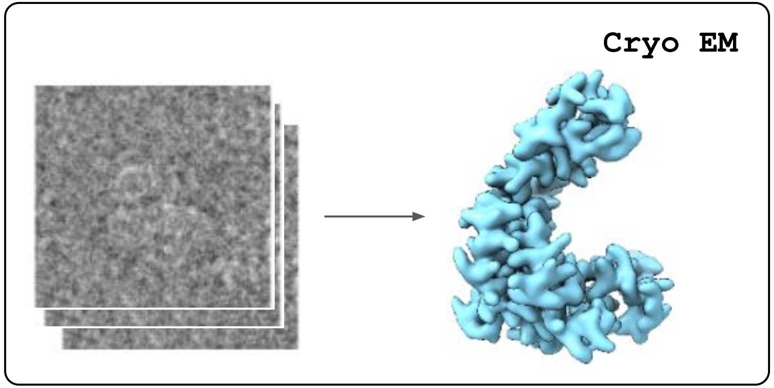
Leveraging SO(3) symmetry for Object Pose Estimation from 2D Images

David Klee

Object pose prediction is an important problem in robotics, AR, and medicine.



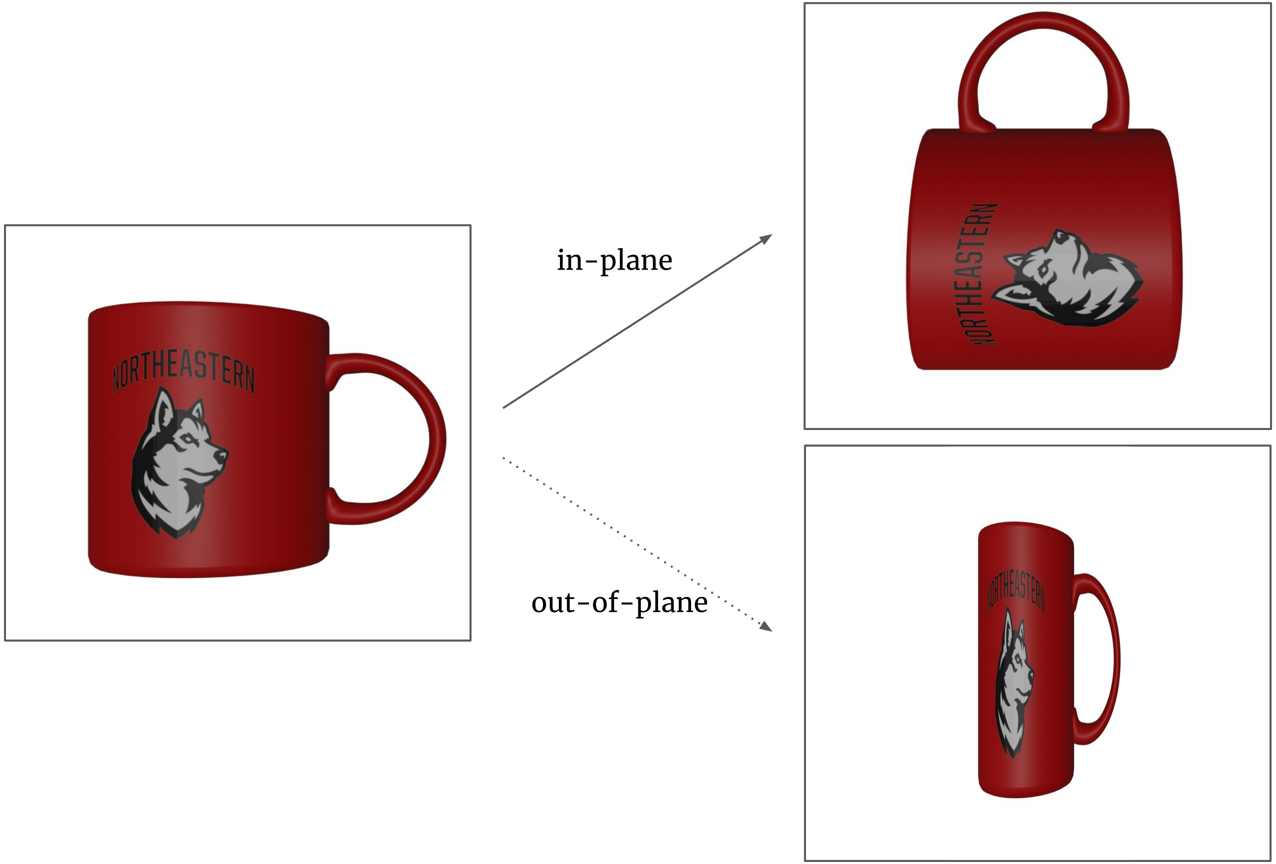




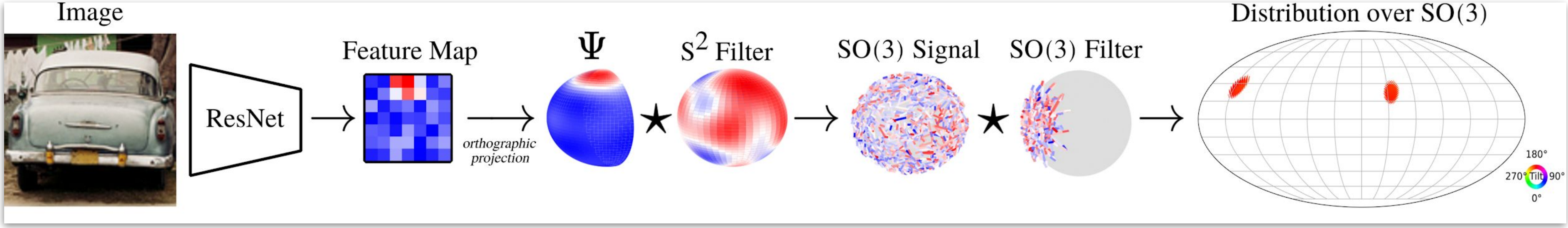
Pose prediction is an equivariant mapping.

$$f_\theta(\mathcal{T}_g x) = \mathcal{T}_g f_\theta(x)$$

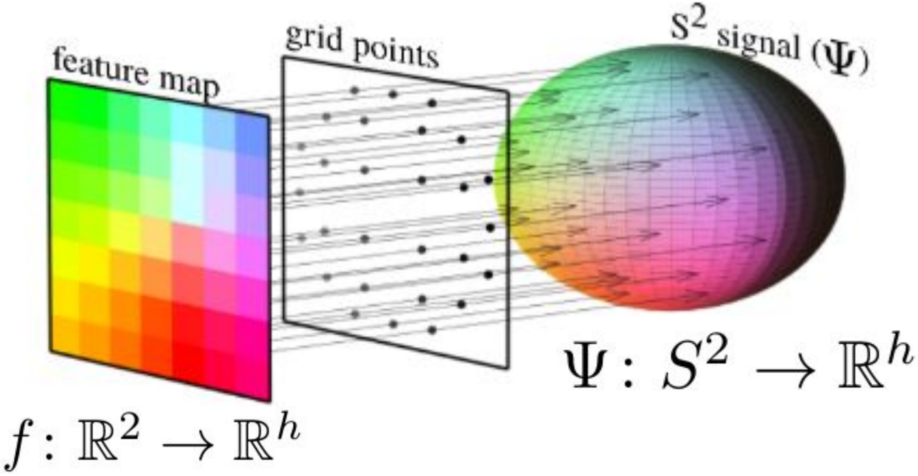
Image inputs destroy end-to-end SO(3) equivariance.



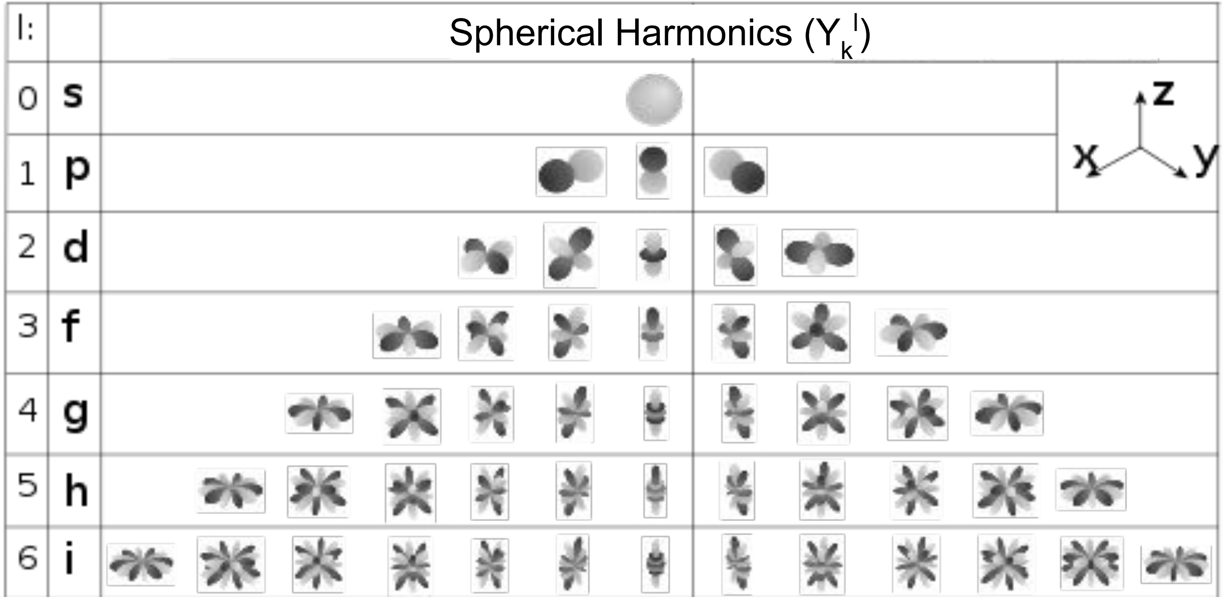
Idea: map learned features to SO(3) transformable space to restore desired equivariance.



1. Extract Image features
2. Map to Features to Sphere
3. Convert to Fourier Basis
4. Perform SO(3) equivariant convolutions
5. Normalize distribution in spatial domain

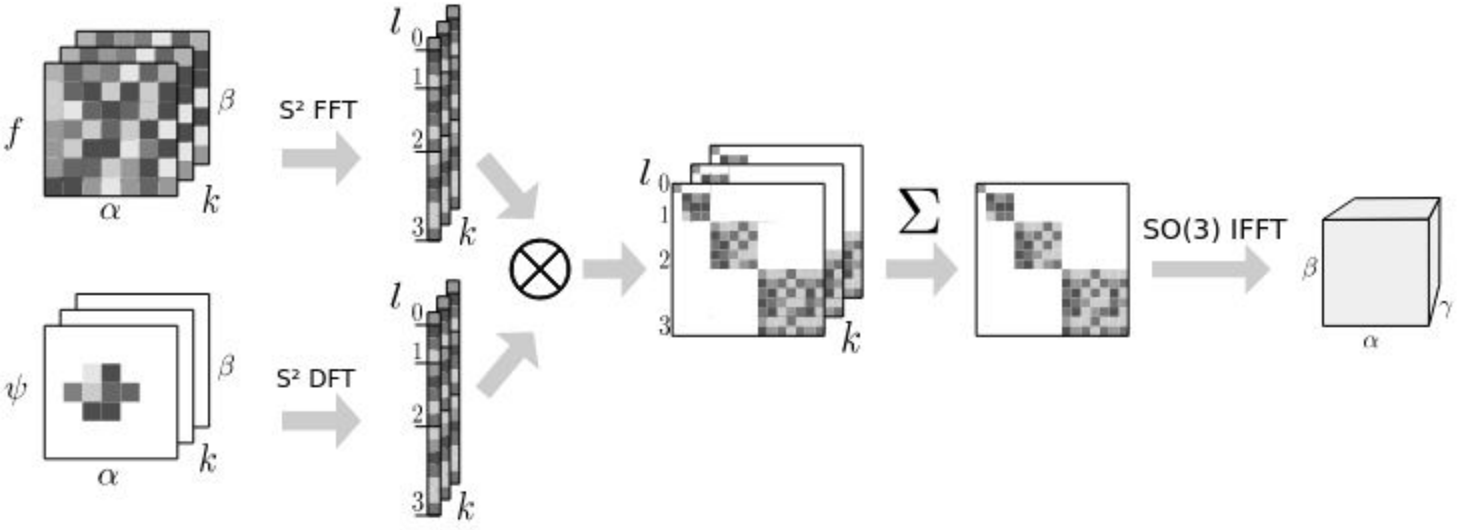


$$P : S^2 \to \mathbb{R}^2$$
$$P(x, y, z) = (x, y)$$
$$\Psi(x) = f(P(x))$$



$$\Psi(x) \approx \sum_{l=0}^{L} \sum_{k=0}^{2l+1} c_k^l Y_k^l(x)$$

$$f(g) \approx \sum_{l=0}^{L} \sum_{m=0}^{2l+1} \sum_{n=0}^{2l+1} c_{mn}^l D_{mn}^l(g)$$



Cohen, Taco S., et al. "Spherical cnns." *arXiv preprint arXiv:1801.10130* (2018).

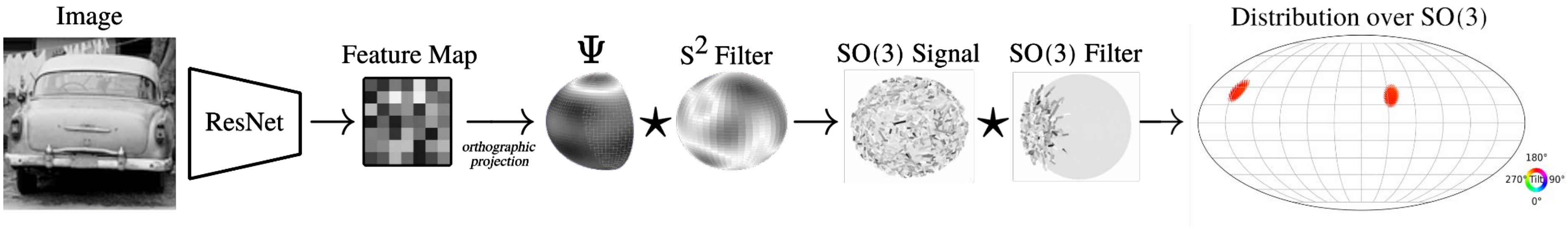


Image
ResNet
Feature Map
orthographic projection
Ψ
S2 Filter
SO(3) Signal
SO(3) Filter
Distribution over SO(3)
180°
270°
Tilt
90°
0°

Training with cross entropy loss

1. Calculate signal over equi-volumetric grid

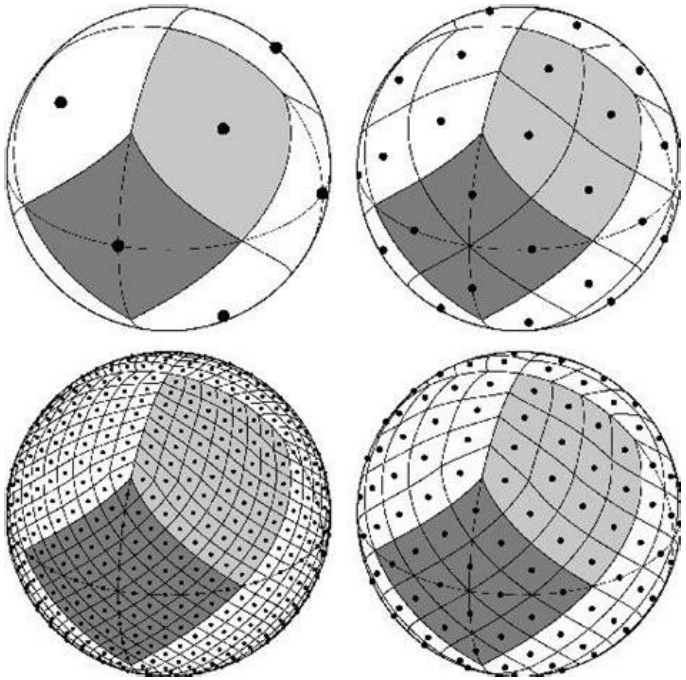
$$f(g) \approx \sum_{l=0}^{L} \sum_{m=0}^{2l+1} \sum_{n=0}^{2l+1} c_{mn}^{l} D_{mn}^{l}(g)$$

2. Perform softmax operation to normalize distribution

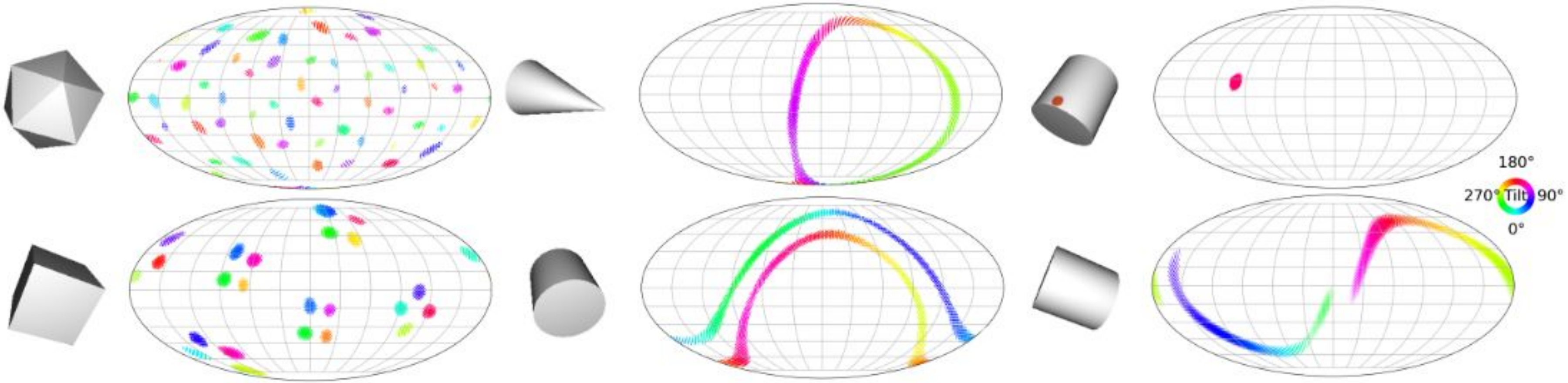
$$p(g) = \frac{\exp f(g)}{\sum_h \exp f(h)}$$

3. Minimize negative log likelihood

$$\mathcal{L}(p, g^*) = -\log p(g^*)$$



The SO(3) HEALPix grid is generated recursively to achieve different resolutions


180°
270°
Tilt
90°
0°

Colab Guide

Full paper