Image to Icosahedral Projection for SO(3) Object Reasoning from Single-View Images

Northeastern
Khoury College of Computer and Information Sciences

David Klee, Ondrej Biza, Robert Platt, Robin Walters

dmklee.github.io/image2icosahedral

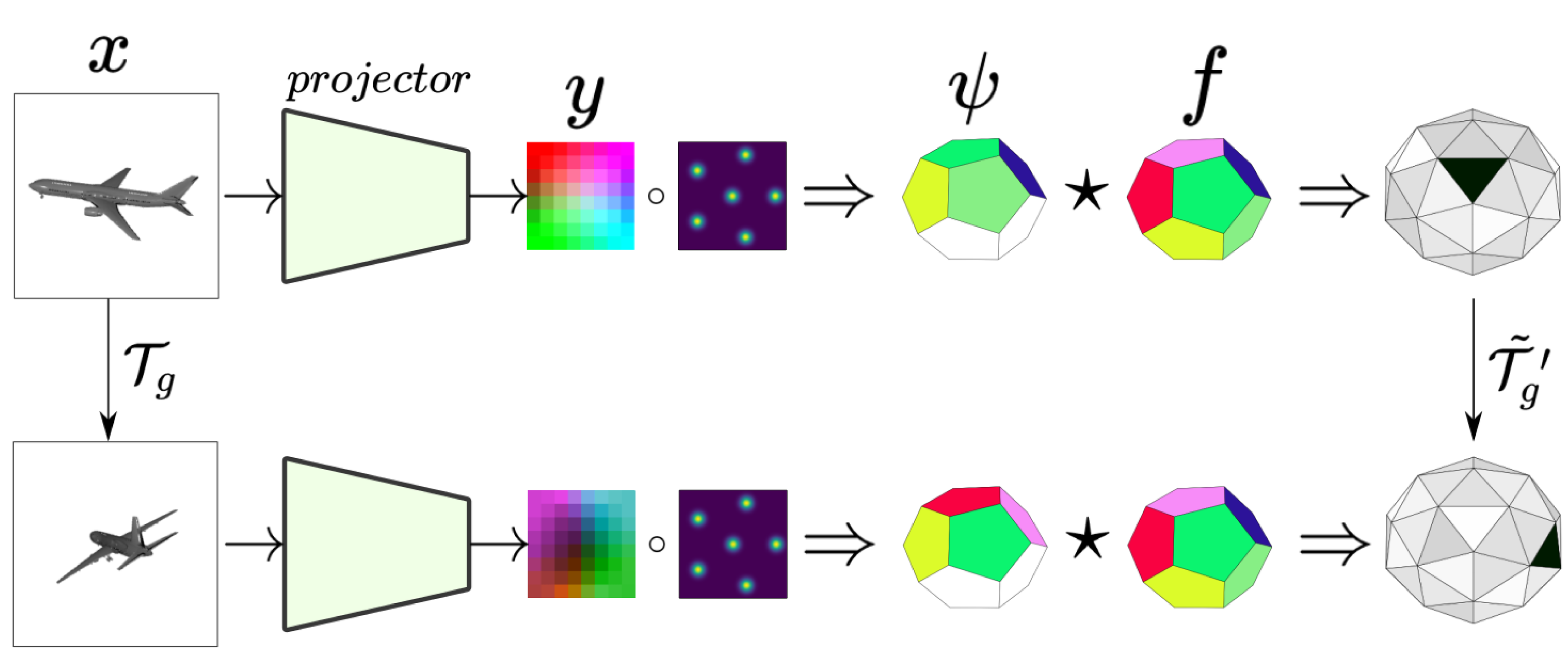


Figure 1: Our method, Image2Icosahedral (I2I), approximates equivariance to SO(3), i.e. rotations of the object correspond to rotations of output representation.

Motivation

- Many computer vision tasks require reasoning about 3D rotations of the scene.
- SO(3) equivariant networks generalize across transformations of the input.
- Existing SO(3)-equivariant models cannot be trained on single-view images since the 2D input is not SO(3)-transformable.

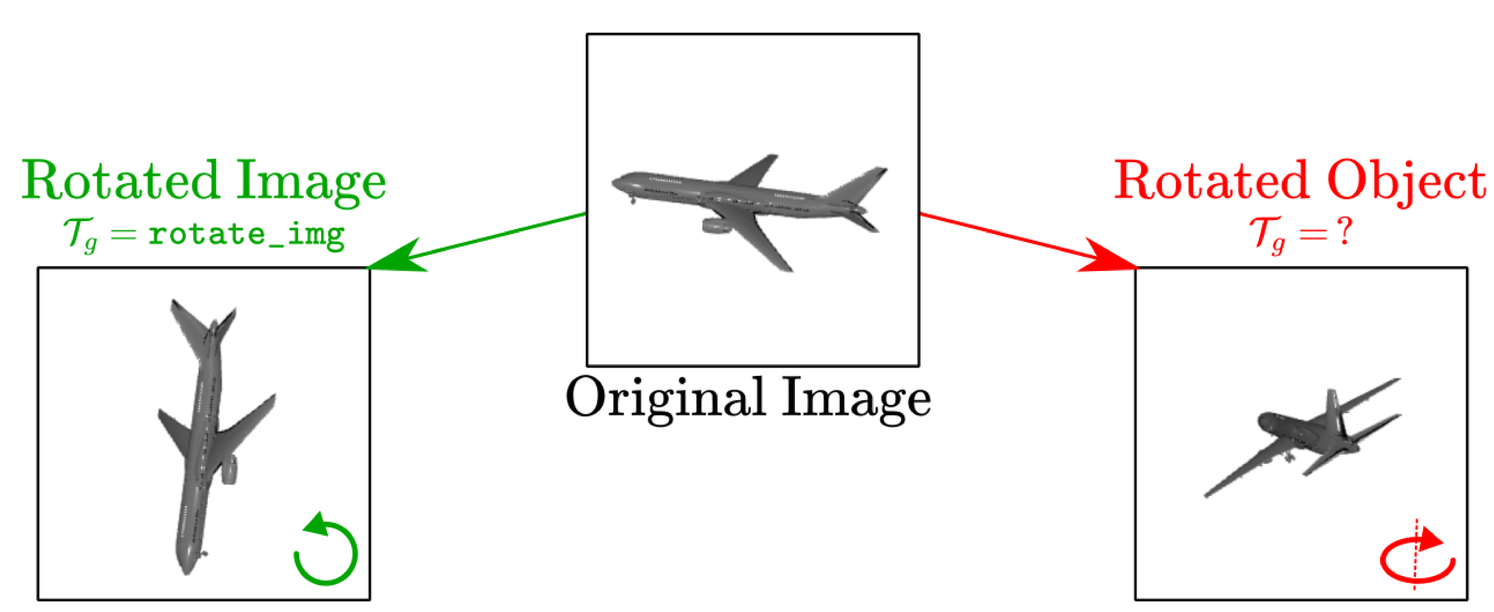


Figure 2: Existing SO(3) equivariant networks cannot be applied to image inputs since the group action is not defined. Unlike SO(2) rotations in the camera plane, arbitary SO(3) rotations cannot be described as a transformation of the image.

Background

- An equivariant function commutes with the action of a group: $f(\mathcal{T}_g x) = \mathcal{T}'_g f(x)$
- Equivariant neural networks can be built with group convolution layers [1]:

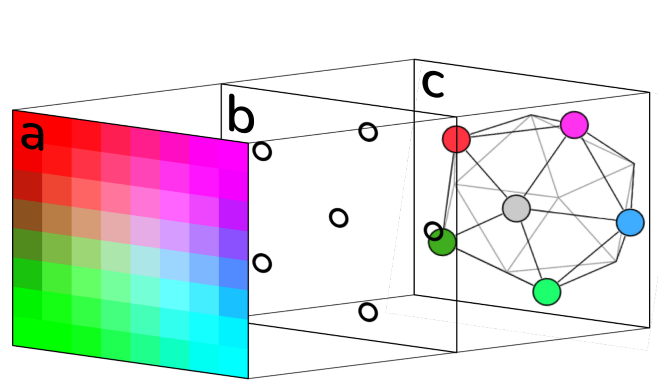
$$[f \star \psi](g) = \sum_{x \in \mathcal{X}} f(x) \cdot \psi(\mathcal{T}_g^{-1} x)$$

- To reduce compute, equivariance can be enforced for a discrete subgroup.
- The largest* subgroup of SO(3) is the Icosahedral group, I_{60}. Group convolution over homogenous spaces of I_{60} was introduced by [2]

Method

Our method combines SO(2) and SO(3) equivariant processing to solve problems that require 3D reasoning from 2D images. It consists of three parts:

1. ResNet-style encoder with SO(2) equivariant convolutional layers [3] processes image to produce dense feature map.
2. Features are orthographically projected from image plane (a) onto vertices of icosahedron (c), forming filter over homogeneous space of Icosahedral group.



3. Icosahedral group convolution is performed between projected filter and trainable signal on icosahedron, generating features over full Icosahedral group.

Experiments

Orientation Prediction

- We generate continuous rotations by predicting an offset rotation from the nearest group element. It is trained with classification and regression loss terms.

input	method	bottle	sofa	car	chair	plane
Grayscale	CNN+GS	5.8	44.7	50.4	22.1	10.6
	CNN+Proc.	5.0	34.8	40.8	19.5	8.2
	CNN+S^3_{exp}	7.8	31.6	57.4	23.7	12.3
	CNN+IER	6.8	47.3	61.5	20.4	10.6
	E2CNN-Eq	4.1	19.7	99.5	17.1	5.7
	I2I (ours)	**2.3**	**5.2**	**5.4**	**7.9**	**2.9**
Point Cloud	KPConv	**2.0**	16.9	113.4	14.6	**1.54**
	EPN	15.4	**3.01**	**93.2**	**3.2**	4.4

Table 1: Median rotation error on ModelNet40 objects.

- Our method significantly outperforms other CNN-based methods.
- I2I is **competitive with point cloud methods** with end-to-end SO(3) equivariance, even when trained without depth information.
- On ambiguous objects like car, our method outperforms point cloud methods; we hypothesize CNN processing is better suited to integrate global context.

Shape Classification

Input	Method	Acc. (%)	mAP
Single Depth Image	CNN	76.5	65.5
	E2CNN-Inv	80.4	70.7
	I2I (ours)	**81.5**	**74.5**
Full Point Cloud	PointNet++	85.0	70.3
	KPConv	86.7	77.5
	EPN	**88.3**	**79.7**

Table 2: Performance on ModelNet40 object set observed at random orientations.

- Our method captures the **rotation invariance** of shape classification using a group pool operation over the Icosahedral group.

Ablation Study

	Average Median Error (°)	
	60 views	15 views
I2I	13.5	16.7
w/o E2CNN	15.2	17.5
w/o GroupConv	18.3	45.5

Table 3: Ablations of I2I on ModelNet40 orientation prediction.

- The SO(3) equivariant layer is more **beneficial** in **low-data regime**.
- End-to-end SO(2) equivariance is not essential to our method.

Conclusions

- We present a novel method for learning 3D representations of objects from 2D images that can be trained end-to-end on challenging computer vision tasks.
- *Limitations*: Our method cannot handle ambiguities in pose and the output is limited to the 60 group elements of the Icosahedral group.
- *Future work*: Extend to continuous SO(3) group and learn object symmetries.

References

[1] T. Cohen and M. Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016.

[2] C. Esteves, Y. Xu, C. Allen-Blanchette, and K. Daniilidis. Equivariant multi-view networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1568–1577, 2019.

[3] M. Weiler and G. Cesa. General E(2)-Equivariant Steerable CNNs. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2019.