Equivariant Multi-View Networks

Carlos Esteves* Yinshuang Xu* Christine Allen-Blanchette Kostas Daniilidis

{machc,xuyin,allec,kostas}@seas.upenn.edu

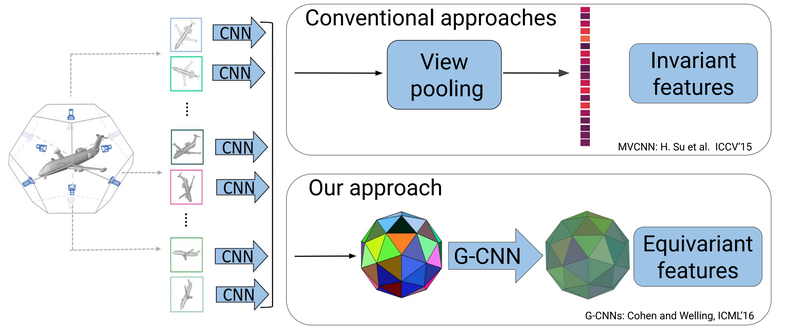
GRASP Laboratory, University of Pennsylvania

Penn Engineering GRASP LABORATORY
General Robotics, Automation, Sensing & Perception Lab

Introduction

- Equivariant representations reduce sample and model complexity.
- In 3D vision, we seek equivariance to the group of 3D rotations, SO(3).
  - Currently, this requires specialized architecture and feature topology.
  - State-of-the-art methods use multi-view 2D CNNs and are not equivariant.
- We propose a group convolutional approach to multi-view aggregation, enabling joint equivariant reasoning over all views.
- Our model can also operate on homogeneous spaces of the rotation group.
- Applications to 3D shape analysis and panoramic scene classification.

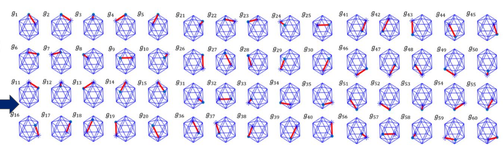
Multi-View to Equivariant Multi-View Networks



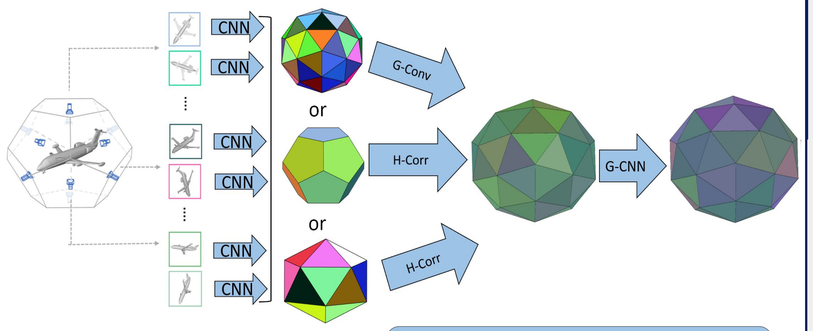
- Convolution on discrete groups:

$$(f * h)(y) = \sum_{g \in G} f(g) h(g^{-1} y)$$

$\mathcal{G}$: rotational icosahedral group



EMVN on group or homogeneous spaces



- Convolution on homogeneous spaces:

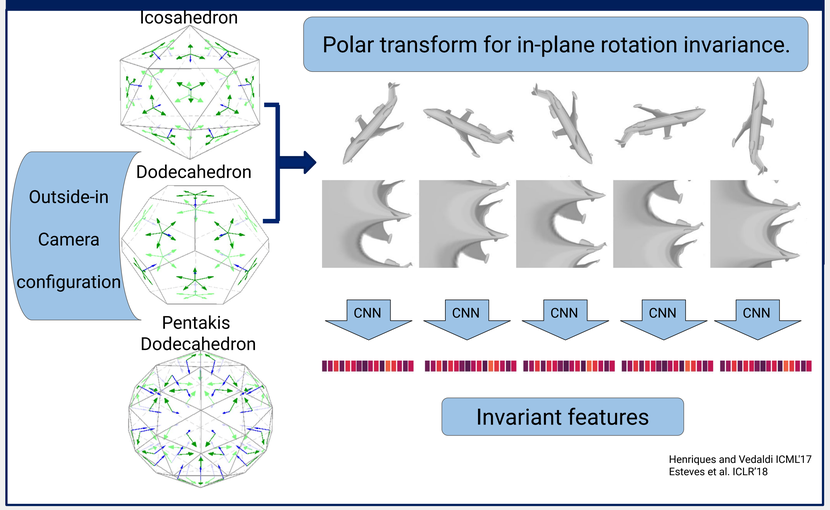
$$(f * h)(x) = \int_{g \in G} f(g\eta) h(g^{-1} x) dg$$

- Correlation on homogeneous spaces:

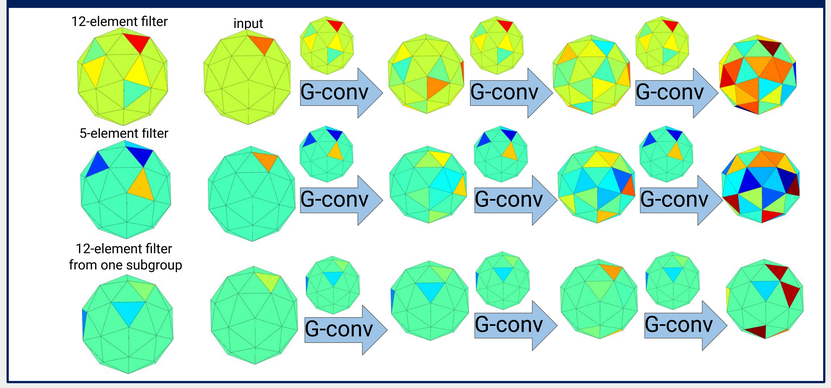
$$(f \star h)(g) = \int_{x \in \mathcal{X}} f(gx) h(x) dx$$

We lift homogeneous space features to the group via correlation

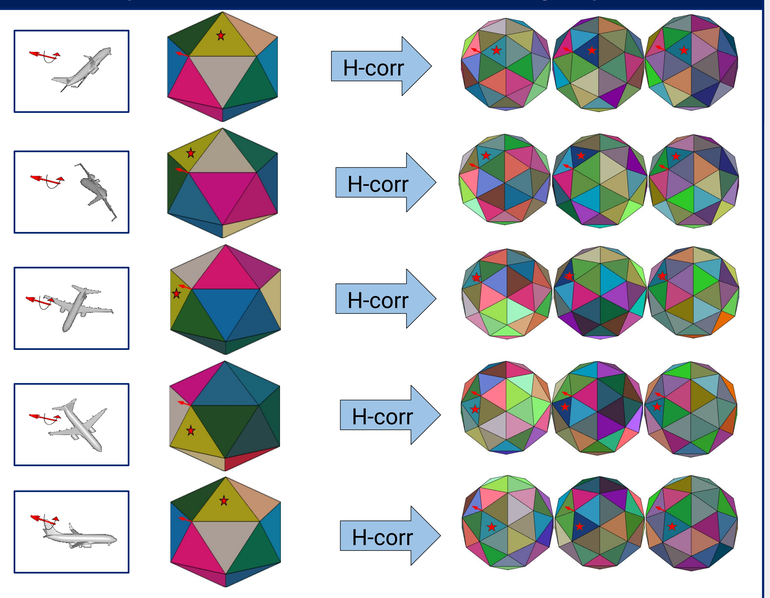
Variable number of input views



Localizing filters on the discrete rotation group

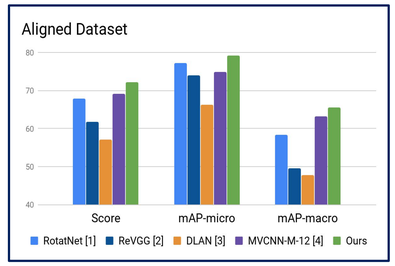


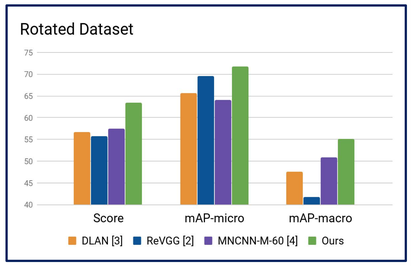
Equivariance to the discrete rotation group



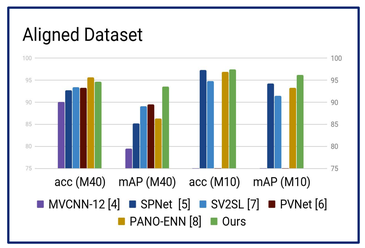
3D shape analysis benchmarks

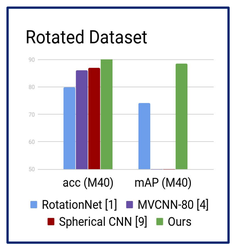
SHREC'17 3D shape retrieval challenge:





Modelnet classification and and retrieval:

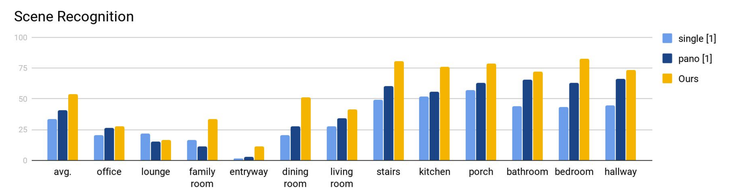




[1] Kanezaki et al. CVPR'18
[2] M. Savva et al. 3DOR'17
[3] T. Furuya et al. BMVC'16
[4] H. Su et al. ICCV'15
[5] M. Yavartanoo et al. arXiv'18
[6] Haoxuan You et al. ACM'18
[7] Z. Han et al. TIP'19
[8] Kostantinos Sfikas C&G'18 [9] C. Esteves ECCV'18

Panoramic scene recognition (Matterport3D)

Outperforms baselines by a large margin!



[1] A.Chang et al. Matterport3d

Conclusion

- We combine the power of conventional CNNs with the robustness of equivariant CNNs, enabling joint equivariant reasoning over multiple views.
- We surpass the state of the art on several 3D shape analysis benchmarks.
- Our code is available at https://github.com/daniilidis-group/emvn