Equivariance to the discrete rotation group

We train another network to invert the embeddings with a loss to reconstruct the input architecture similar to a flipped embedding network, with $L_2$ loss.

Low sample complexity: training with a single image, not pairs

Pre-trained Spherical-CNN

GRASP Laboratory, University of Pennsylvania

Equivariant Multi-View Networks

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

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

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

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:
  \[ (f * h)(g) = \sum_{g' \in G} f(g')h(g^{-1}g') \]
  G rotational icosahedral group

- Convolution on homogeneous spaces:
  \[ (f * h)(x) = \int_{g \in G} f(g)h(g^{-1}x)dg \]

Invariant features

Localizing filters on the discrete rotation group

Equivariance to the discrete rotation group

We lift homogeneous space features to the group via correlation

Panoramic scene recognition (Matterport3D)

Modelnet classification and retrieval:

- Outperforms baselines by a large margin!