Any-Shot GIN: Generalizing Implicit Networks for Reconstructing Novel Classes

Yongqin Xian, Julian Chibane, Bharat Lal Bhatnagar, Bernt Schiele, Zeynep Akata, and Gerard Pons-Moll

Motivation

- Task: 3D reconstruction from a single RGB image
- Setting: train on 13 ShapeNet classes and evaluate on a large number of novel/unseen classes
- Weakness of previous methods:
  - ONet [1] has no explicit regularization -> overfit to training classes

Contributions

- A new method which sequentially predicts front-back depth, projects depth into 3D and estimates shapes with implicit surfaces which reason in 3D
- A new state-of-the-art on both seen and novel shape classes for single-image 3D reconstruction
- Insights: using depth for learning implicit surfaces enhances generalization; projecting depth into 3D to extract 3D features preserve shape details
- Good few-shot learner: novel classes can be further improved by using only few-shot depth supervision

Method: Generalizing Implicit Networks (GIN)

- Input Image
- Predicted Depth
- 3D Voxel Space
- Point-Aligned Multi-Scale Features 3D CNN
- Front-Back Depth Estimation
- Implicit Shape Completion

Qualitative Results

- Training: first, train the depth estimation network and shape completion network separately using the ground truth depth maps and shapes in viewer-center coordinate, afterwards, fine-tune both networks end-to-end.
- Inference: predict front-back depth -> back-project them to 3D voxel space -> estimate occupancies of all grid points at desired resolution -> obtain a mesh with Marching Cube algorithm

References


Quantitative Results

Table 1: Comparisons on ShapeNet. We report Chamfer distance (CD), normal consistency (NC) and F-score (FS). All methods are trained on 13 seen classes and evaluated on both seen and novel classes.

Table 2: Ablations on ShapeNet. BD: back-view depth, VC: viewer-centered, OC: object-centered coordinates. PAMSF: point-aligned multi-scale features

- Our GIN outperforms SOTA in all metrics, improving SDFNet by 5% on seen classes and 9% on novel classes in terms of F-Score
- Depth and PAMSF are both crucial for achieving the best performance
- Viewer-centered supervision enhances generalization on novel classes

Few-Shot Learning Results

- Left figure: fine-tune each class separately vs fine-tune all classes together, showing that the depth allows our method to benefit more from the geometric knowledge shared across shape categories.
- Right figure: the effect of using different supervision signals, indicating that our method can be further improved using only few-shot depth supervision