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

ETH zürich

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

3DV 2022

https://virtualhumans.mpi-inf.mpg.de/gin/

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
  - MarrNet [3] and GenRe [5] are limited by resolution due to voxel repre.
  - ONet [1] has no explicit regularization -> overfit to training classes
  - SDFNet [2] fails to predict details due to global shape encoding

Input Image
ONet
SDFNet
Our

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

References

[1] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger. **Occupancy networks: Learning 3d reconstruction in function space**. *CVPR*, 2019.

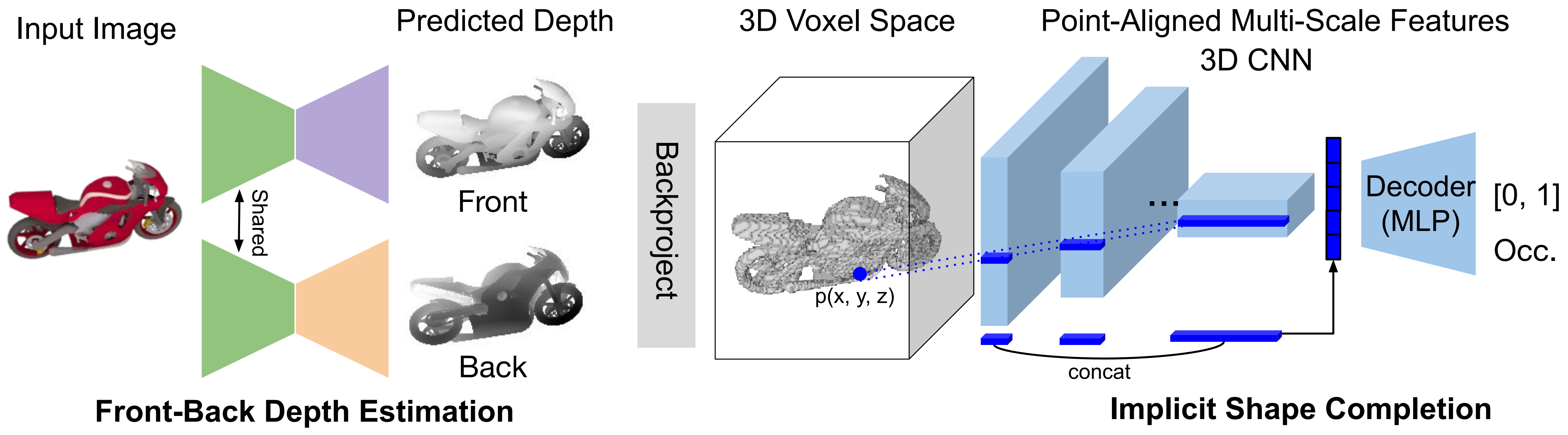
[2] A. Thai, S. Stojanov, V. Upadhya, and J. Rehg. **3d reconstruction of novel object shapes from single images**. *3DV*, 2021.

[3] J. Wu, Y. Wang, T. Xue, X. Sun, W. Freeman and J. Tenenbaum. **Marrnet: 3d shape reconstruction via 2.5 d sketches**. *NIPS*, 2017.

[4] Q. Xu, W. Wang, D. Ceylan, R. Mech, and U. Neumann. **Disn: Deep implicit surface network for high-quality single-view 3d reconstruction**. *NeurIPS*, 2019.

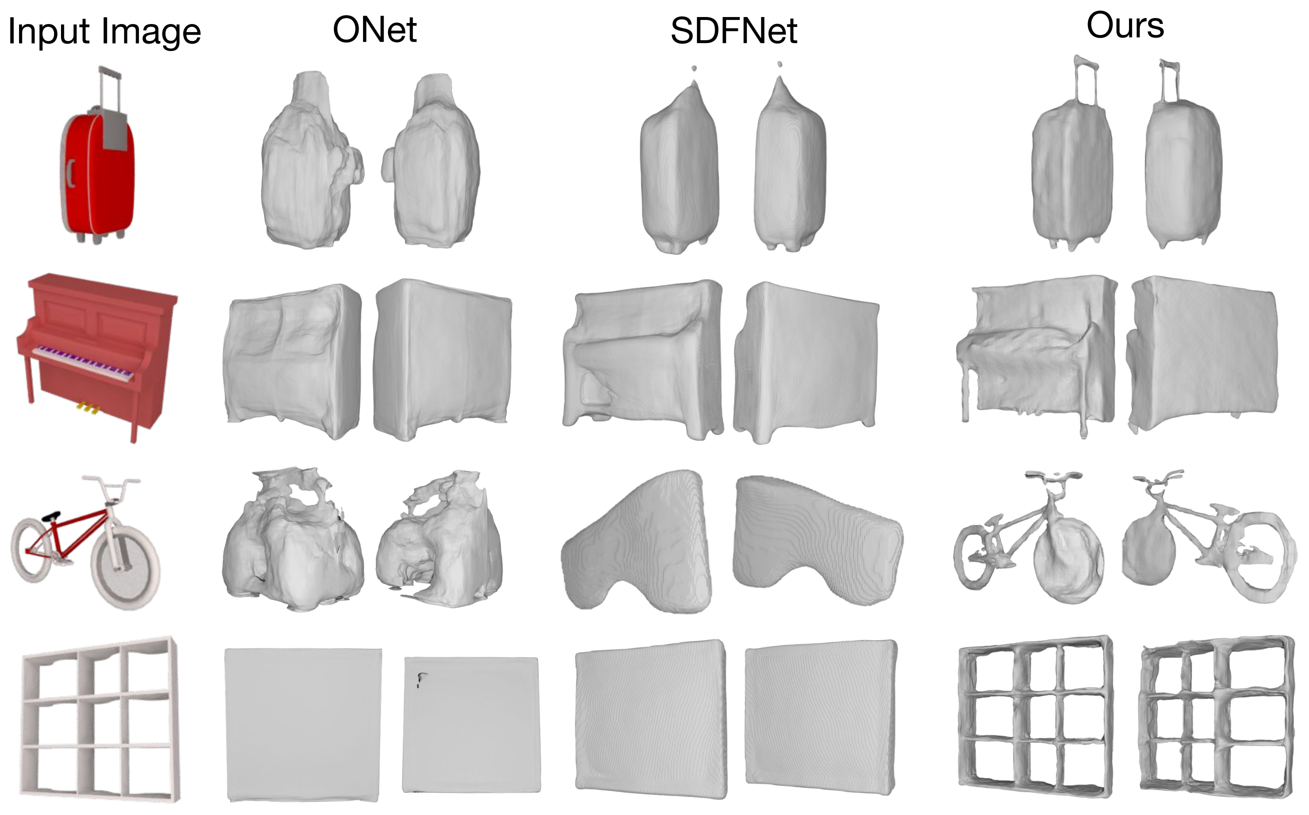
[5] X. Zhang, Z. Zhang, C. Zhang, J. Tenenbaum, W. Freeman and J. Wu. **Learning to Reconstruct Shapes From Unseen Classes**. *NIPS*, 2018.

Method: Generalizing Implicit Networks (GIN)

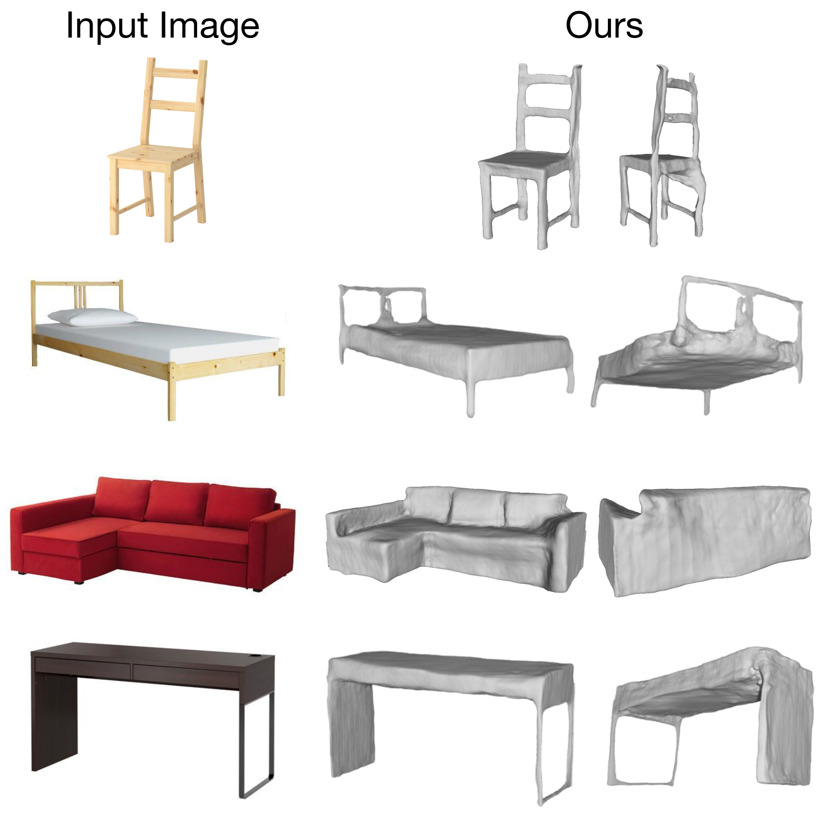


- Training: first, train the depth estimation network and shape completion network separately using the ground truth depth maps and shapes in viewer-center coordinate, afterwards, fine-tune both networks end-to-end.
- Inference: predict front-back depth -> back-project them to 3D voxel space -> estimate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

Qualitative Results



Synthetic Images of Novel Classes



Generalization to Real Images

Quantitative Results

	Seen Classes			Novel Classes		
Method	CD ↓	NC ↑	FS ↑	CD ↓	NC ↑	FS ↑
GenRe [5]	0.153	0.60	0.12	0.172	0.61	0.11
MarrNet [3]	0.116	0.68	0.15	0.127	0.69	0.13
ONet [1]	0.081	0.78	0.25	0.145	0.72	0.15
DISN [4]	0.070	0.77	0.33	0.124	0.72	0.20
SDFNet [2]	0.050	**0.79**	0.42	0.080	0.76	0.31
GIN (ours)	**0.042**	**0.79**	**0.47**	**0.056**	**0.79**	**0.40**

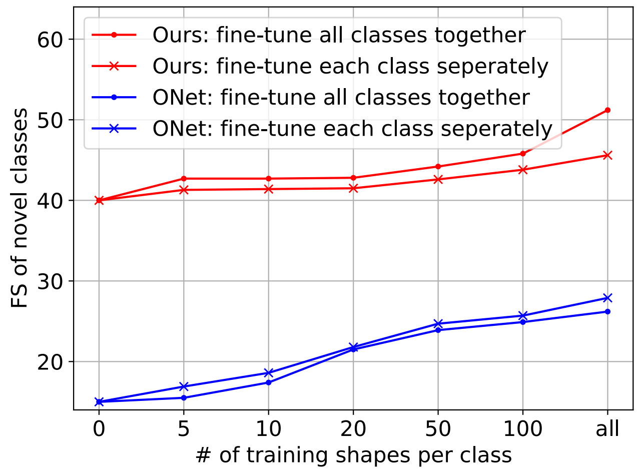
Table 1: Comparisons on ShapeNet. We report Chamfer distance (CD), normal consistency (NC) and F-score (FS). All methods are trained on 13 seen classes and evaluated on both seen and novel classes.

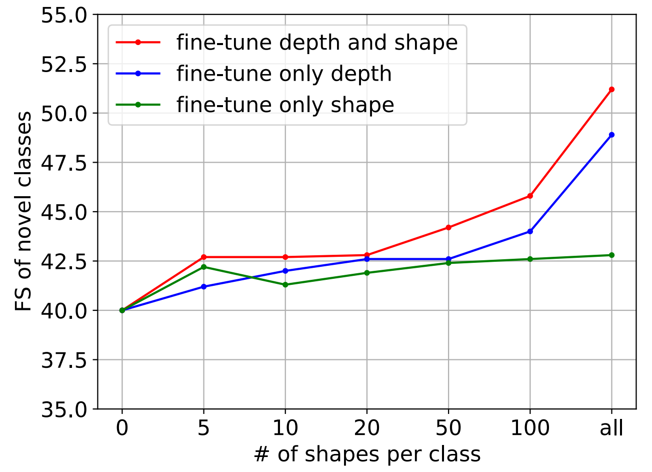
		Seen Classes		Novel Classes	
Method	Coordinate	CD ↓	FS ↑	CD ↓	FS ↑
Ours	VC	**0.042**	0.47	**0.056**	**0.40**
Ours w/o BD	VC	0.043	**0.48**	0.059	**0.40**
Ours	OC	0.042	**0.48**	0.059	0.38
Ours w/o Depth	VC	0.060	0.35	0.086	0.25
Ours w/o PAMSF	VC	0.056	0.36	0.073	0.28

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