SCARP: 3D Shape Completion in ARbitrary Poses for Improved Grasping

Bipasha Sen*†, Aditya Agarwal*†, Gaurav Singh*†, Brojeshwar B.‡, Srinath Sridhar §and Madhava Krishna†
*Equal contribution
†Robotics Research Center, IIIT-Hyderabad
‡TCS Research, India
§Brown University

Abstract—**Recovering full 3D shapes from partial observations is a challenging task that has been extensively addressed in the computer vision community. Many deep learning methods tackle this problem by training 3D shape generation networks to learn a prior over the full 3D shapes. In this training regime, the methods expect the inputs to be in a fixed canonical form, without which they fail to learn a valid prior over the 3D shapes. We propose SCARP, a model that performs Shape Completion in ARbitrary Poses. Given a partial pointcloud of an object, SCARP learns a disentangled feature representation of pose and shape by relying on rotationally equivariant pose features and geometric shape features trained using a multi-tasking objective. Unlike existing methods that depend on an external canonicalization method, SCARP performs canonicalization, pose estimation, and shape completion in a single network, improving the performance by 45% over the existing baselines. In this work, we use SCARP for improving grasp proposals on tabletop objects. By completing partial tabletop objects directly in their observed poses, SCARP enables a SOTA grasp proposal network improve their proposals by 71.2% on partial shapes.**

I. Introduction

Given a partial observation of an object, 3D shape completion aims to recover the full 3D shape of the object. This has been widely addressed in computer vision [26, 11, 8, 13, 18, 31, 27, 2] and has many diverse downstream applications in robotics including visual servoing [6], manipulation [5, 1, 9, 21], visual inspection [12], autonomous driving [3, 29, 22].

Many existing methods tackle shape completion by incorporating a training scheme that learns a prior over the full 3D shapes. This is done by training an autoencoder [24, 28, 26, 31] or a GAN [30] over many different instances of full shapes. At inference, this learned prior space is conditionally queried on the partial observations. These methods however, suffer from a major limitation: they expect the partial input to be in a fixed canonical frame–a common frame of reference that is shared between instances in that category [23, 16]. A particular shape X in two different poses $\{R_1, T_1\}$ and $\{R_2, T_2\}$ will have very different geometry. As a result, X in different poses appear as novel instances for these methods inhibiting them from learning a valid prior over shapes.

Existing datasets like ShapeNet [4] have shapes that are manually aligned to a canonical frame, but real shape observations (e.g., depth maps) do not contain this information.

A naive approach to tackling this challenge is to *canonicalize*, i.e., map a 3D (full or partial) shape to a category-level canonical frame with [23] or without supervision [16, 20, 19].

A multi-stage pipeline can be built involving the sequential steps of (1) canonicalization, (2) shape completion, and (3) de-canonicalization (bringing the object back in the original pose). In such a pipeline however, the performance of a shape completion network directly depends on the output quality of the canonicalization module. This can lead to errors propagating between these modules leading to a sub-optimal completion.

We propose SCARP, a method that performs **S**hape **C**ompletion in **AR**bitrary **P**oses. Unlike existing methods that have to directly learn a prior over all possible poses and shapes, we first disentangle the pose from the shape of a partial pointcloud. We build a multi-task objective that: (1) generates a disentangled feature representation of pose and shape by canonicalizing an object to a fixed frame of reference, (2) estimates the exact pose of the object, and (3) completes the shape of the object using the disentangled representation. This multi-task objective allows our network to jointly understand the pose and shape of the input.

It does so by learning rotationally-equivariant and translationally-invariant pose features using Tensor Field Networks [14], and global geometric shape features using PointNet++ [15].

Application: Robotic grasp pose estimation [21, 9, 7, 10] is a challenging area of research that often expects a faithful reconstruction of the scene in 3D.

Under a partial observation, [21] generates grasp proposals that directly collide with the actual object in the scene (shown in red). As a result, the manipulator is likely to collide with the object as it attempts to grasp the objects using one of these predicted grasp poses. We use SCARP to complete these partial shapes directly in their observed poses and estimate grasp proposals on these completed shapes. We show that SCARP reduces such invalid grasps by 71.2% over predicting grasp poses directly on the partial observations.

To summarize, our contributions are:

1) We propose SCARP, a novel architecture to perform shape completion from partial pointclouds in arbitrary poses.
2) We show for the first time how a multi-task objective can support: (1) canonicalization, (2) 6D pose estimation, and (3) shape completion on partial pointclouds.

3) We demonstrate that SCARP outperforms the existing shape completion baselines (with pre-canonicalization) by 45% and improves grasp pose estimation by reducing invalid grasp poses by 71%.

II. Method

Given a partial object pointcloud $\hat{X}_p$ at an unknown pose $\{R, T\}$, we want to estimate this pose and the corresponding full object pointcloud $\hat{X}$ in the same pose.

This is a challenging task as for a neural network, a pointcloud X in two different poses $\{R_1, T_1\}$ and $\{R_2, T_2\}$ are two completely different pointclouds. Thus, we adopt a multi-tasking objective that disentangles the pose and the shape of the input partial pointcloud $\hat{X}_p$. The shape component allows us to understand that $\hat{X}_p$ is a partial observation of X which is $\hat{X}$ in its canonical form. The pose component is then used to estimate the pose transform $\{R, T\}$ between $\hat{X}$ and X.

A. Multi-tasking Pipeline for disentangling Shape and Pose

Let X_p and X be a partial and its corresponding full pointcloud in a fixed canonical frame. Then $\hat{X}_p$ and $\hat{X}$ are X_p and X in an unknown arbitrary pose $\{R, T\}$ such that $\hat{X}_p = R(X_p) + T$ and $\hat{X} = R(X) + T$. The input to our network is $\hat{X}_p$ which is mean centered at the origin and normalized to a unit bounding box. Our aim is to predict $\{R, T\}$ and the full pointcloud $\hat{X}$ which is posed as:

$$\{R, T, \hat{X}\} = \Phi(\hat{X}_p) \tag{1}$$

where Φ denotes our proposed network, SCARP.

Our multi-tasking objective is formulated to (1) complete the partial pointcloud in a fixed canonical frame given by X and (2) estimate the pose transformation from the canonical frame to the original pose $\{R, T\}$. In this pipeline, the two components (1) pose and (2) shape are predicted separately using two different output heads as shown in Fig. 1.

1) Feature Extraction: To estimate the input's shape, we compute global geometric shape features, $p \in \mathbb{R}^E$, using Pointnet++ [15] To estimate the pose of the input, we adapt TFN [14]. Our TFN computes a global equivariant feature, $F \in \mathbb{R}^{N \times E}$ by max pooling over the types $\{\ell\}_{\ell=0}^{\ell=\ell_{max}}$, where E is the dimension of the equivariant embeddings, N and ℓ_{max} are user-defined.

The input to our shape completion network is a non-linear combination of p and a global invariant embedding, $F_{\mathcal{X}} \in \mathbb{R}^E$, computed by max pooling F over the channel dimension, N. Additionally, F is used to estimate an equivariant frame of reference, $\{R' \in \mathbb{R}^{3\times3}, T' \in \mathbb{R}^3\}$ that transforms the invariant embeddings to X's original pose.

2) Task I: Shape Completion: Completing the shape of a partial input at any arbitrary orientation is difficult. Therefore, we aim to first complete the shape at a fixed canonical frame. To learn this canonical frame, the model needs to build an understanding of the full shape of the partial input. To achieve this, we train our model to predict a full canonicalized pointcloud $X' \in \mathbb{R}^{1024\times3}$ directly from $\hat{X}_p \in \mathbb{R}^{512\times3}$. Shape completion enables our model to learn a prior over the global shape of a category (a typical chair would have four legs and a backrest) enabling our network to directly canonicalize the partial inputs accurately.

We adopt our generator G, from [17], as our shape completion network where (1) the input to G is a semantically meaningful embedding generated from a partial input $\hat{X}_p$ and (2) is trained using a distance loss against the full pointcloud X to learn a relationship between the partial input $\hat{X}_p$ and the predicted full canonical pointcloud X'. As shown in Fig. 1 (right), the input to G is a globally invariant feature vector $f \in \mathbb{R}^E$ computed by combining $p = P(\hat{X}_p)$ and $F_{\mathcal{X}} = \mathcal{X}(\hat{X}_p)$ non-linearly using a neural network ϕ_S given as:

$$X' = G(f) \quad \text{and} \quad f = \phi_S(\mathcal{X}(\hat{X}_p) \oplus P(\hat{X}_p))) \tag{2}$$

3) Task II: Pose Estimation: Once G predicts the full pointcloud X' in a canonical pose, it is important to estimate the correct rotation $SO(3)$ matrix $R \in \mathbb{R}^{3\times3}$ and translation $T \in \mathbb{R}^3$ to register X' back on $\hat{X}_p$. We predict R' and T' on the second head of our model using the rotationally equivariant TFN features F given as:

$$R' = \phi_R(F) \quad \text{and} \quad T' = \phi_T(F) \tag{3}$$

where ϕ_R and ϕ_T are multi-layered perceptrons.

4) Subsubsection Heading Here: Subsubsection text here.

B. Loss Functions for Multitask Training

1) Shape Completion in a fixed Canonical Frame: In the first task, we estimate the completed pointcloud in a fixed canonical frame given by X'. We use DCD [25] to minimize the distance between the predicted pointcloud X' and the ground truth canonical pointcloud X given by:

$$\mathcal{L}_{shape} = d_{DCD}(X', X) \tag{4}$$

2) Estimating the pose of the object: To estimate the pose given by $\{R, T\}$, we use rotationally equivariant pose features F and pass it through $\{\phi_R, \phi_T\}$. We constrain this prediction against the canonical frame. To do so, we rotate the canonical output X' to obtain $R'(X')$ and compare it against the rotated ground truth $\hat{X}$. At this point however, the pointwise correspondences between X and X' are lost. Thus, a hard distance loss such as Euclidean distance cannot be directly used. To tackle this, we minimize permutation invariant CD [1] objective between X and X'. However, CD only minimizes the distance between the nearest neighbors of the points in the pointcloud. This results in local minimas where the loss is minimal even when the actual correspondences are far. As a result, the predicted pointcloud is often flipped about one of the axes. To tackle this issue, we rotate the canonical ground truth X using the predicted R' and compare against $\hat{X}$ using $L2$ loss. The overall loss is:

$$\mathcal{L}_{rot} = \delta d_{CD}(\hat{X}, R'(X')) + \gamma ||\hat{X}, R'(X)||_2 \tag{5}$$

[1] https://pdal.io/en/latest/apps/chamfer.html

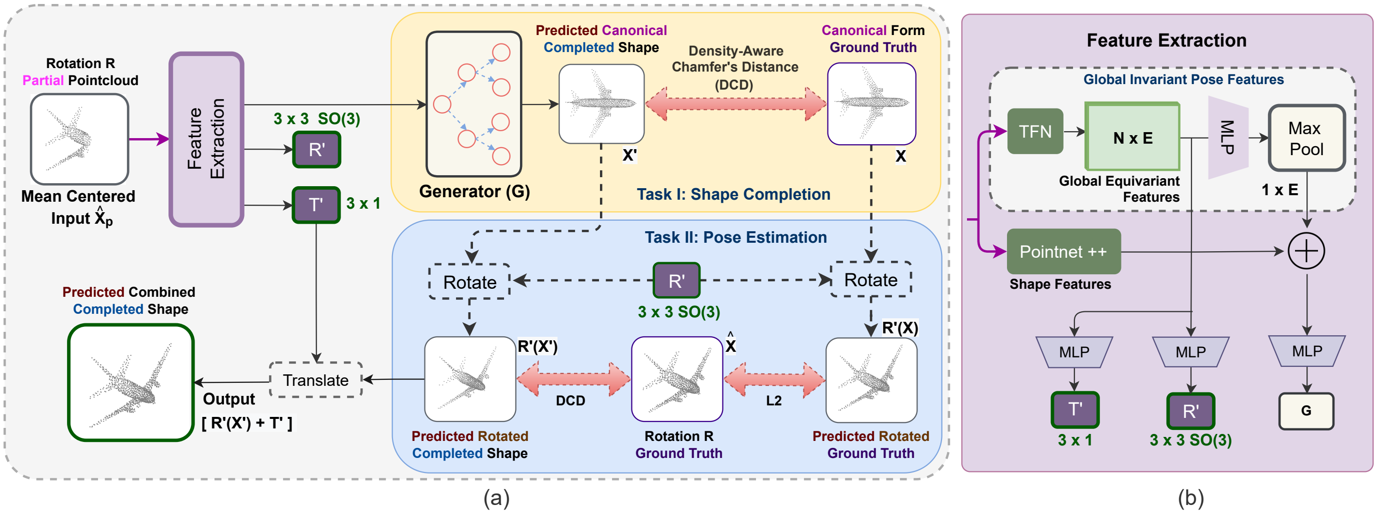

Fig. 1. **Overview of our proposed approach:** The input to SCARP is a mean-centered partial pointcloud $\hat{X}_p$ in an arbitrary orientation R. Our feature extraction module **(b)** disentangles the partial pointcloud's pose and shape and is trained in a multi-tasking objective **(a)**. In the first task, SCARP combines Pointnet++ [15] and TFN [14] features to generate a shape feature that is used by a pointcloud completion network, G, to generate X'. In the second task, the TFN pose feature is used to generate an equivariant frame $\{R', T'\}$. Our loss functions enable the overall network to learn a prior over the shape while understanding the pose of the partial input.

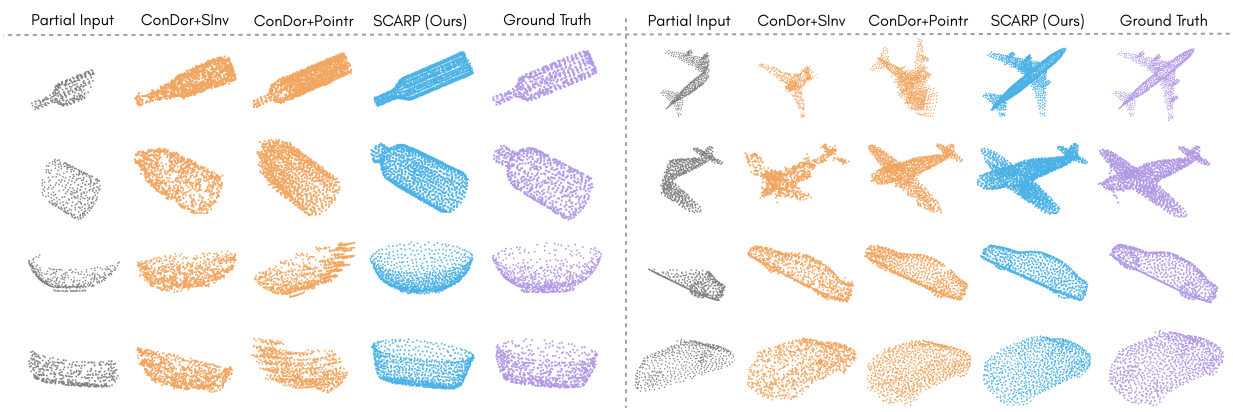

Fig. 2. Qualitative comparison of shape completion in arbitrary poses on SCARP and the existing multi-stage baselines: Canonicalization using ConDor, Shape Completion, and De-canonicalization. Pointr[27] is a SOTA pointcloud completion network that generates high-resolution completed pointclouds. Shape Inversion (SInv.) [30] is based on tree-GAN [17] that shares our generator G.

$R'(X')$ is computed by detaching the forward computation graph at the output of G. The gradients from the loss does not backpropogate through G at the first head.

For symmetrical objects such as bowls and glasses, multiple R' predictions can be correct. A hard $L2$ loss penalizes the network for correct predictions even for correct R' if the correspondences do not exactly match. Thus, for symmetrical objects, we keep $\delta \sim 1.0$ and $\gamma \sim 0.0$ and for non-symmetrical objects we keep $\delta \sim 1.0$ and $\gamma \sim 1.0$.

The input to our network is a mean-centered partial pointcloud $\hat{X}_p$. At this point, we train our network to regress to $\hat{X}_p$'s centroid in the full pointcloud X given by T'. We directly supervise T' against the ground truth T given as:

$$\mathcal{L}_{trans} = ||T' - T||_2 \tag{6}$$

The final output is obtained by rotating and translating our predicted pointcloud X' by R' and T' respectively as:

$$X_o = R'(X') + T' \tag{7}$$

Orthonormality Loss: The rotation R' predicted by our network is a 3×3 matrix in the $SO(3)$ space. However, the matrix predicted by Eqn. 3 is not guaranteed to be a valid $SO(3)$ matrix. We therefore, enforce orthonormality on R' by minimizing its difference to its closest orthonormal matrix. To do so, we compute the SVD decomposition of $R = U\Sigma V^T$ and enforce unit eigenvalues as:

$$\mathcal{L}_{orth} = ||UV^T - R||_2 \tag{8}$$

3) Combined Loss: We train our network end-to-end by combining all the losses as:

$$\mathcal{L} = \mathcal{L}_{shape} + \mathcal{L}_{rot} + \mathcal{L}_{trans} + \mathcal{L}_{orth} \tag{9}$$

		Tabletop					Off-Table				
		Bowl	Bottle	Can	Mug	Basket	Plane	Car	Chair	Watercraft	Average
CD↓	ConDor+SInv.	82.7	27.4	45.4	41.5	85.3	34.2	14.7	59.4	39.9	47.8
	ConDor+Pointr	30.8	20.9	29.9	14.2	40.9	22.1	6.4	19.8	8.5	21.5
	SCARP (Ours)	**21.8**	**7.9**	**11.8**	**12.1**	**34.2**	**6.9**	**5.6**	**19.1**	**7.1**	**14.0**
		Bowl	Bottle	Can	Mug	Basket	Plane	Car	Chair	Watercraft	Average
MMD-EMD↓	ConDor+SInv.	27.3	17.2	20.1	19.9	29.2	19.6	11.3	22.2	18.9	20.6
	ConDor+Pointr	21.6	13.6	14.8	12.6	18.8	14.4	8.1	13.5	9.1	14.1
	SCARP (Ours)	**9.6**	**6.3**	**8.8**	**8.4**	**10.6**	**5.0**	**5.6**	**8.4**	**6.0**	**7.6**

TABLE I

QUANTITATIVE COMPARISON OF SHAPE COMPLETION IN ARBITRARY POSES FOR TABLETOP AND OFF-TABLETOP OBJECTS. MOST TABLETOP OBJECTS ARE SYMMETRICAL WHEREAS OFF-TABLE OBJECTS HAVE MORE VARIATIONS IN STRUCTURE. CHAMFER'S DISTANCE (CD) AND EARTH MOVERS DISTANCE-MAXIMUM MEAN DISCREPANCY (MMD-EMD) ARE SCALED BY 10^3 AND 10^2.

III. RESULTS

Dataset: Our dataset is a subset of [4] derived from [16] and [27] made of 5 tabletop (Bowl, Bottle, Can, Mug, Basket) and 4 non-tabletop (Plane, Car, Chair, Watercraft) categories.

Results: As shown in Table I, SCARP outperforms the existing multi-stage baselines on all the categories on an average by 45%. The existing shape completion methods rely on the output of an external canonicalization model that suffer from their own inconsistencies as reported in their paper [16]. This results in an error propagation as the input to the shape completion networks are not always in the exact canonical forms. The errors in the input map to a larger error in the output of the networks. This is followed by an error in the transform from the canonical form to the original pose. The resulting output of the multi-stage pipeline suffer from high inconsitensies and sub-optimal outputs.

Unlike these networks, our model is trained jointly on both tasks (canonicalization and shape-completion) using a multi-tasking objective. As we show in the ablations, this objective plays a crucial role in achieving a disentangled representation of shape and pose.

Qualitative results are shown in Fig. 2 that vividly show the closeness of SCARP's output to the ground truth when compared with others.

Improvement in Grasp Proposals: Generating grasp proposals for partial pointclouds is a challenging task as a network may mistake a missing portion of an object as a potential area to grasp (see Fig. 3). We apply SCARP to complete these partial observations directly in the observed poses and predict grasp poses on these completed pointclouds using a SOTA grasp generation network Contact-Graspnet [21]. Qualitative results are shown in Fig. 3.

As can be seen, the grasps proposed on partial observations collide with the actual object (ground truth), whereas, the grasp proposals made on the completed object by SCARP are valid.

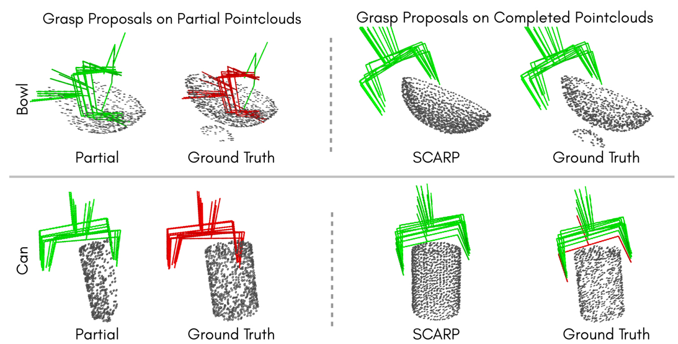

Fig. 3. **(left)**: Grasp proposals made by a SOTA grasp proposal network, [21], on partial observations lead to collisions with the actual object. *Partial* is a partial observation and *Ground truth* denotes the actual object. The proposals are made on *Partial* (shown in green) but collide with the actual object (shown in red). **(right)**: We use SCARP to complete the partial observations. Grasp proposals made on the completed objects align well with the actual object on the table reducing such collisions by a large margin.

IV. CONCLUSION

Existing shape completion works assume the partial inputs to be in a fixed canonical frame. This is difficult to achieve in a robotics setting where the objects are observed in arbitrary poses thus needing pre-canonicalization. This leads to an error propagation resulting in a sub-optimal shape completion. We propose SCARP, a novel architecture that performs Shape Completion in ARbitrary Poses. SCARP is trained using a multi-task objective to perform (1) canonicalization, (2) 6D pose estimation, and (3) shape completion. SCARP outperforms the existing multi-stage baselines by 45% and showcases its potential in improving grasp proposals on tabletop objects, reducing colliding grasps by more than 70%. SCARP has a huge potential in many more robotics applications like collision avoidance in trajectory planning or differential simulators for model-based RL planners.

ACKNOWLEDGMENTS

We would like to thank Karthik Desingh, who is an Assistant Professor at the University of Minnesota, and Adrien Poulenard, who is a Postdoctoral Fellow at the Stanford University, for their valuable feedback.

REFERENCES

[1] Aditya Agarwal, Bipasha Sen, Shankara Narayanan, Vishal Reddy Mandadi, Brojeshwar Bhowmick, and K Madhava Krishna. Approaches and challenges in robotic perception for table-top rearrangement and planning, 2022. URL https://arxiv.org/abs/2205.04090.

[2] Armen Avetisyan, Angela Dai, and Matthias Nießner. End-to-end cad model retrieval and 9dof alignment in 3d scans. In *Proceedings of the IEEE/CVF International Conference on computer vision*, pages 2551–2560, 2019.

[3] Mark Campbell, Magnus Egerstedt, Jonathan How, and Richard Murray. Autonomous driving in urban environments: Approaches, lessons and challenges. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 368:4649–72, 10 2010. doi: 10.1098/rsta.2010.0110.

[4] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

[5] Charles C. Kemp, Aaron Edsinger, and Eduardo Torres-Jara. Challenges for robot manipulation in human environments [grand challenges of robotics]. *IEEE Robotics & Automation Magazine*, 14(1):20–29, 2007. doi: 10.1109/MRA.2007.339604.

[6] Gourav Kumar, Harit Pandya, Ayush Gaud, and K. Madhava Krishna. Pose induction for visual servoing to a novel object instance. page 2953–2959. IEEE Press, 2017. doi: 10.1109/IROS.2017.8206130. URL https://doi.org/10.1109/IROS.2017.8206130.

[7] Jeffrey Mahler, Jacky Liang, Sherdil Niyaz, Michael Laskey, Richard Doan, Xinyu Liu, Juan Aparicio Ojea, and Ken Goldberg. Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics. *Robotics: Science and Systems*, 2017.

[8] Paritosh Mittal, Yen-Chi Cheng, Maneesh Singh, and Shubham Tulsiani. Autosdf: Shape priors for 3d completion, reconstruction and generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 306–315, 2022.

[9] Arsalan Mousavian, Clemens Eppner, and Dieter Fox. 6-dof graspnet: Variational grasp generation for object manipulation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2901–2910, 2019.

[10] Adithyavairavan Murali, Weiyu Liu, Kenneth Marino, Sonia Chernova, and Abhinav Gupta. Same object, different grasps: Data and semantic knowledge for task-oriented grasping. *Conference on Robot Learning*, 2020.

[11] Liang Pan, Tong Wu, Zhongang Cai, Ziwei Liu, Xumin Yu, Yongming Rao, Jiwen Lu, Jie Zhou, Mingye Xu, Xiaoyuan Luo, et al. Multi-view partial (mvp) point cloud challenge 2021 on completion and registration: Methods and results. *arXiv preprint arXiv:2112.12053*, 2021.

[12] Harit Pandya, Ayush Gaud, Gourav Kumar, and Madhava Krishna. Instance invariant visual servoing framework for part-aware autonomous vehicle inspection using mavs: Pandya et al. *Journal of Field Robotics*, 36, 01 2019. doi: 10.1002/rob.21859.

[13] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019.

[14] Adrien Poulenard and Leonidas J. Guibas. A functional approach to rotation equivariant non-linearities for tensor field networks. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13169–13178, 2021. doi: 10.1109/CVPR46437.2021.01297.

[15] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017.

[16] Rahul Sajnani, Adrien Poulenard, Jivitesh Jain, Radhika Dua, Leonidas J Guibas, and Srinath Sridhar. Condor: Self-supervised canonicalization of 3d pose for partial shapes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16969–16979, 2022.

[17] Dong Wook Shu, Sung Woo Park, and Junseok Kwon. 3d point cloud generative adversarial network based on tree structured graph convolutions. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3859–3868, 2019.

[18] Vincent Sitzmann, Eric Chan, Richard Tucker, Noah Snavely, and Gordon Wetzstein. Metasdf: Meta-learning signed distance functions. *Advances in Neural Information Processing Systems*, 33:10136–10147, 2020.

[19] Riccardo Spezialetti, Federico Stella, Marlon Marcon, Luciano Silva, Samuele Salti, and Luigi Di Stefano. Learning to orient surfaces by self-supervised spherical cnns. *Advances in Neural Information Processing Systems*, 33:5381–5392, 2020.

[20] Weiwei Sun, Andrea Tagliasacchi, Boyang Deng, Sara Sabour, Soroosh Yazdani, Geoffrey E Hinton, and Kwang Moo Yi. Canonical capsules: Self-supervised capsules in canonical pose. *Advances in Neural Information Processing Systems*, 34:24993–25005, 2021.

[21] Martin Sundermeyer, Arsalan Mousavian, Rudolph Triebel, and Dieter Fox. Contact-graspnet: Efficient 6-dof grasp generation in cluttered scenes. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13438–13444. IEEE, 2021.

[22] Sebastian Thrun, Michael Montemerlo, Hendrik Dahlkamp, David Stavens, Andrei Aron, James Diebel, Philip Fong, John Gale, Morgan Halpenny, Gabriel Hoffmann, Kenny Lau, Celia Oakley, Mark Palatucci, Vaughan Pratt, Pascal Stang, Sven Strohband, Cedric Dupont, Lars-Erik Jendrossek, Christian Koelen, and Pamela Mahoney. Stanley: The robot that won the darpa grand challenge. *J. Field Robotics*, 23:661–692, 01 2006.

[23] He Wang, Srinath Sridhar, Jingwei Huang, Julien Valentin, Shuran Song, and Leonidas J Guibas. Nor-

malized object coordinate space for category-level 6d object pose and size estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2642–2651, 2019.

[24] Xiaogang Wang, Marcelo H Ang Jr, and Gim Hee Lee. Cascaded refinement network for point cloud completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 790–799, 2020.

[25] Tong Wu, Liang Pan, Junzhe Zhang, Tai Wang, Ziwei Liu, and Dahua Lin. Density-aware chamfer distance as a comprehensive metric for point cloud completion. *Advances in Neural Information Processing Systems*, 2021.

[26] Xingguang Yan, Liqiang Lin, Niloy J Mitra, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. Shapeformer: Transformer-based shape completion via sparse representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6239–6249, 2022.

[27] Xumin Yu, Yongming Rao, Ziyi Wang, Zuyan Liu, Jiwen Lu, and Jie Zhou. Pointr: Diverse point cloud completion with geometry-aware transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12498–12507, 2021.

[28] Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. Pcn: Point completion network. In *2018 International Conference on 3D Vision (3DV)*, pages 728–737. IEEE, 2018.

[29] Wenyuan Zeng, Shenlong Wang, Renjie Liao, Yun Chen, Bin Yang, and Raquel Urtasun. Dsdnet: Deep structured self-driving network. In *European conference on computer vision*, pages 156–172. Springer, 2020.

[30] Junzhe Zhang, Xinyi Chen, Zhongang Cai, Liang Pan, Haiyu Zhao, Shuai Yi, Chai Kiat Yeo, Bo Dai, and Chen Change Loy. Unsupervised 3d shape completion through gan inversion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1768–1777, 2021.

[31] Haoran Zhou, Yun Cao, Wenqing Chu, Junwei Zhu, Tong Lu, Ying Tai, and Chengjie Wang. Seedformer: Patch seeds based point cloud completion with upsample transformer, 2022.