Learning Prior Mean Function for Gaussian Process Implicit Surfaces

Tarik Kelestemur[1], Taskin Padir[1], Robert Platt[2] and David Rosen[1]

I. INTRODUCTION

Having full knowledge of 3D geometries of objects is an important skill for robotic manipulation. However, this knowledge might not be available for robots operating in unstructured environments with a limited number of sensors. The shape completion methods deal with this problem by estimating the full shape of the object from a partial observation. With the rise of deep learning, a series of methods [1], [2], [3], [4] addressed the shape completion problem and achieved impressive results. Following the success of these methods, the robotics community incorporated these methods into manipulation [5], [6] to improve the performance of tasks such as grasping and pick-and-place. While these methods work well with the in-class objects, they lack the capability of generalizing to novel shapes. Ideally, a robot should be able to make sense of an unknown object's geometries and reason about the reconstruction quality.

On the other hand, another line of research addresses the uncertainty estimation in shape completion methods by leveraging probabilistic models such as Gaussian processes (GP). Work by [7], [8], [9], [10] has shown that the Gaussian process can be used to estimate the complete shapes by approximating the signed distance functions (SDF) of objects. Furthermore, they are able to produce meaningful uncertainty estimation of unobserved points. The posterior uncertainty is then used to guide the exploration of the objects to gather more knowledge via either visual or tactile feedback. These methods usually use simple geometric primitives priors (e.g. constant or primitive shape mean functions) which are not expressive enough to represent a large class of shapes. Therefore, they require a large number of exploration steps to successfully produce the complete shape.

To bring the best of the two worlds, we propose a simple procedure (which we call *DeepGPIS*) to dramatically improve the fidelity of Gaussian process implicit surface estimation by learning informative prior mean functions parameterizing Gaussian process models over SDF shape representations. Specifically, we learn an implicit neural network that represents the SDF of a multiple class of objects. Later, this network is used as the mean function of a GP along with a standard radial basis kernel (RBF). Our approach thus combines the flexibility of learning-based shape completion methods with the principled uncertainty quantification and test-time non-parametric adaptability that Gaussian process regression affords. In particular, we show that this method is able to generalize well at test-time to objects that it was not trained to recognize a priori. This is important for real-world applications, where robots may encounter many novel objects in the wild. We validate DeepGPIS via a series of shape completion experiments and show that we can achieve improved shape reconstructions even with unseen classes of shapes.

1. The College of Engineering, 2. The Khoury College of Computer Sciences, 1., 2. The Institute for Experiential Robotics at Northeastern University, Boston, Massauchusetts 02115, USA. kelestemur.t@northeastern.edu

II. METHODS

Background: *Signed-distance function* (SDF) is an implicit shape representation commonly used in robotics. It is defined as a continuous function that takes spatial points $\boldsymbol{x} \in \mathcal{R}^3$ and outputs their signed closest distance $s \in \mathcal{R}$ to object boundary: $SDF(\boldsymbol{x}) = s$. The distance is positive for points outside the object and negative for inside the object. The zero-level set of this function represents the underlying surface of the object which in practice can be extracted using e.g. Marching Cubes. [11]. To be able to represent arbitrarily shaped objects, Park et al. [3] introduced the DeepSDF where the SDF is represented by a neural network. Let $f_\theta(\boldsymbol{z}_i, \boldsymbol{x})$ be a network that represents the SDF and parameterized by θ. It takes 3D points $\boldsymbol{x}$ and a latent code $\boldsymbol{z}_i$ representing the shape. During training, the parameters of the network and the latent codes are optimized jointly using L_1 loss and latent code regularization.

$$\underset{\theta, \{\boldsymbol{z}_i\}_{i=1}^{N}}{\arg\min} \sum_{i=1}^{N} \left(\sum_{j=1}^{M} |f_\theta(\boldsymbol{x}) - s| + \frac{1}{\sigma_r^2} \|\boldsymbol{z}_i\|_2^2 \right) \tag{1}$$

where N is the number of objects, M is the number of points per object, $s_{i,j}$ is the ground truth SDF value for the i-th object and j-th point of point and σ_r is the latent code regularization term. At inference time, the latent codes for test objects are found via Maximum-a-Posterior (MAP) estimation from the observed points of the objects.

$$\hat{\boldsymbol{z}} = \underset{\boldsymbol{z}}{\arg\min} \sum_{(\boldsymbol{x}_j, s_j) \in \boldsymbol{X}} |f_\theta(\boldsymbol{x}) - s| + \frac{1}{\sigma_r^2} \|\boldsymbol{z}\|_2^2 \tag{2}$$

However, we found that the MAP estimation does not work well in the case of a multi classes of shapes. To overcome this problem, we trained an encoder (using pre-trained ResNet [12]) that takes the depth image and predicts a latent code: $h_\phi(I_i) = \hat{\boldsymbol{z}}_i$ which is then used to initialize the MAP estimation (Eq. 2). We train this encoder using L_2 loss with respect to the learned latent codes from DeepSDF training.

The *Gaussian process* (GP) is a non-parametric and probabilistic modelling [13] method that is well suited for continuous functions. A GP can be interpreted as distrubtion

TABLE I
MEAN CHAMFER DISTANCE AND INFERENCE TIME - SINGLE VIEW

Models	Chamfer distance	Time (s)
DeepSDF - Validation Objects	0.128	6.65
DeepGPIS - Validation Objects	0.050	7.15
DeepSDF - Test Objects	0.099	6.80
DeepGPIS - Test Objects	0.059	7.22

over functions and it is specified by its mean and kernel function $GP(m(x), k(x, x'))$. It can produce closed-form posterior distributions to describe uncertainty over shape by conditioning on the observed points. Naturally, they are good candidates for modelling SDFs with uncertainty estimation. The Gaussian Process Implicit Surfaces (GPIS) [13] represents the SDF as a $GP(\boldsymbol{x}, k(\boldsymbol{x}, \boldsymbol{x}'))$ where $\boldsymbol{x}$ is the observed points and $\boldsymbol{x}'$ are the test points.

Proposed Method: We propose to use a learned mean function to improve the fidelity and generalization capabilities of GPIS estimation while leveraging the closed-form uncertainty estimates and non-parametric test-time flexibility that GP models afford. Formally, we will use a DeepSDF as the mean function of a GP model along with an RBF kernel.

$$m(\boldsymbol{x}) = f_\theta(\boldsymbol{z}, \boldsymbol{x}) \tag{3}$$

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \exp(-(\boldsymbol{x} - \boldsymbol{x}')^2)/l^2) \tag{4}$$

where σ_f is the output-scale and l is the length-scale of the kernel function. Compared to most prior work, DeepGPIS will have better inductive biases than what is achieved using simple primitive mean functions [8]. Moreover, it is able to fix the bad predictions of vanilla DeepSDF which is a crucial capability when encountering previously-unknown objects.

III. EXPERIMENTS

Dataset: Our dataset is generated using randomly selected meshes from the ShapeNet [14]. For training and validation, we use 10 classes where each class has 45 training and 5 validation objects. For the test set, we select 5 different classes where each class has 5 objects. Similar to [3], we normalize the meshes to the unit sphere and sample 500000 points inside the sphere. Then, we calculate points' SDF values using KDTree. For shape completion experiments, 8 equidistant cameras are placed on the unit-sphere. The depth images and partial point clouds are saved from these cameras.

Experiments: We evaluate our proposed method on shape completion experiments where the models see a partial point cloud and reconstruct the whole shape. Given a depth image, I, we first predict the initial latent code using the depth encoder which is followed by the MAP estimation using Eq. 2. With the estimated latent code, we evaluate the DeepSDF and DeepGPIS at a 3D grid with 128^3 resolution. Finally, we reconstruct the shape using the Marching Cubes method. The Chamfer distance is used as the evaluation metric which measures how close two point sets are to each other. After reconstructions of the shapes, we sample 100000 points on the surface of the ground truth and estimated meshes: X_1 and X_2. Given these two point sets, the Chamfer distance is calculated as follows:

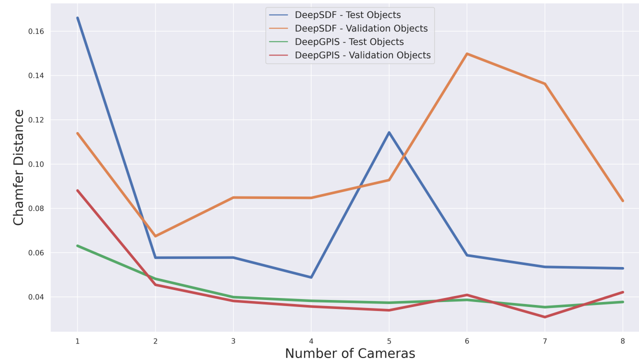

Fig. 1. Mean Chamfer Distance

$$d = \frac{1}{2K} \sum_{x \in X_1} \min_{y \in X_2} ||x - y||_2^2 + \sum_{y \in X_2} \min_{x \in X_1} ||x - y||_2^2 \tag{5}$$

where K is the number of sampled points. The first experiment considers only a single view of the object. We calculate the Chamfer distance for 50 validation objects and 25 test objects. For each object, we reconstruct the shapes from 8 camera viewpoints resulting in 400 and 200 reconstructions for validation and test sets, respectively. The mean root square Chamfer distance $\sqrt{d}$ and the mean inference time are given in Table I . As can be seen from the results, our proposed method outperforms the DeepSDF baseline by a large margin in both validation and test sets. This shows that our method not only generalizes the in-class object but also can work with out-of-the distribution classes.

In the second experiment, we evaluate our method on sequential point cloud streams. In this case, the models see consecutive partial point clouds which are merged with every new data. Note that we rerun the latent code inference with every new data since it will estimate a better code with more coverage of the object. We also condition the DeepGPIS on the merged data. The Chamfer distance over all the objects with respect to the number of cameras is shown in Fig. 1. The results show that our method is able to improve the reconstruction quality with more data while DeepSDF struggles to leverage the new data. In both validation and test object, DeepGPIS produces better results.

IV. CONCLUSIONS

In this work, we combined a powerful deep learning-based shape representation (DeepSDF) with a probabilistic model for implicit surfaces (GPIS) which allows us to generate complete shapes of novel objects and estimate their predictive uncertainties. Through our experiments, we showed that DeepGPIS outperforms one of the state-of-the-art shape completion methods. For future work, we would like to explore an active perception setting where the uncertainty estimation of shape predictions will be used to find the next-best-view or next-best-touch to generate complete shapes in a minimum number of steps.

REFERENCES

[1] W. Yuan, T. Khot, D. Held, C. Mertz, and M. Hebert, "Pcn: Point completion network," in *2018 International Conference on 3D Vision (3DV)*. IEEE, 2018, pp. 728–737.

[2] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger, "Occupancy networks: Learning 3d reconstruction in function space," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4460–4470.

[3] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "Deepsdf: Learning continuous signed distance functions for shape representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 165–174.

[4] H. Huang, Z. Yang, and R. Platt, "Gascn: Graph attention shape completion network," in *2021 International Conference on 3D Vision (3DV)*. IEEE, 2021, pp. 1269–1278.

[5] J. Lundell, F. Verdoja, and V. Kyrki, "Beyond top-grasps through scene completion," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 545–551.

[6] W. Gao and R. Tedrake, "kpam-sc: Generalizable manipulation planning using keypoint affordance and shape completion," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 6527–6533.

[7] S. Dragiev, M. Toussaint, and M. Gienger, "Gaussian process implicit surfaces for shape estimation and grasping," in *2011 IEEE International Conference on Robotics and Automation*. IEEE, 2011, pp. 2845–2850.

[8] W. Martens, Y. Poffet, P. R. Soria, R. Fitch, and S. Sukkarieh, "Geometric priors for gaussian process implicit surfaces," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 373–380, 2016.

[9] S. Suresh, Z. Si, J. G. Mangelson, W. Yuan, and M. Kaess, "Efficient shape mapping through dense touch and vision," *arXiv preprint arXiv:2109.09884*, 2021.

[10] C. de Farias, N. Marturi, R. Stolkin, and Y. Bekiroglu, "Simultaneous tactile exploration and grasp refinement for unknown objects," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3349–3356, 2021.

[11] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," *ACM siggraph computer graphics*, vol. 21, no. 4, pp. 163–169, 1987.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[13] O. Williams and A. Fitzgibbon, "Gaussian process implicit surfaces," in *Gaussian Processes in Practice*, 2006.

[14] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, *et al.*, "Shapenet: An information-rich 3d model repository," *arXiv preprint arXiv:1512.03012*, 2015.