ReDSDF: Regularized Deep Signed Distance Fields for Robotics

Puze Liu, Kuo Zhang, Davide Tateo, Snehal Jauhri, Jan Peters and Georgia Chalvatzaki

***Abstract*— The safe operation of autonomous robots requires avoiding unintentional contact between the robot itself and the environment, such as obstacles and humans. Distance-based constraints are fundamental for enabling robots to plan and act safely. Moreover, different applications require different distance resolutions, leading to various heuristic approaches for measuring distance fields w.r.t. obstacles. We propose Regularized Deep Signed Distance Fields (ReDSDF), a single neural implicit function that can compute smooth distance fields at any scale, with fine-grained resolution over high-dimensional manifolds and articulated bodies like humans, thanks to our effective data generation and a simple *inductive bias* during training. We demonstrate the effectiveness of our approach in representative simulated tasks for whole-body control (WBC) and safe Human-Robot Interaction (HRI) in shared workspaces.**

I. Introduction

Distance-based constraints restrict robot operation space to avoid unintentional contacts with the environment. Defining distances between query points and objects with complex shapes or dynamic articulations, such as shelves or humans, is a challenging problem. A Signed Distance Function (SDF) constructs an implicit function that inputs the query point and the surface encoding and outputs the signed distance between the query point and the surface. The constraint surface is defined as the 0-level curve. Furthermore, SDFs can be differentiable, enabling the exploitation of gradient-based methods to deal with the constraints.

SDFs generally can be categorized into sensor-centric SDF and object-centric SDF. Sensor-centric SDF tries to reconstruct the partially observed surface of the environment based on the distance information obtained from the sensor. Peer research focuses on the improvement of surface reconstruction, such as robustness [1], refinement [2]. Sensor-centric SDF are widely used for mobile robot mapping, planning, and navigation [3]–[9]. However, sensor-centric SDF are not applicable when the query point differs from the sensor. Object-centric SDF approximates the object surface as a level-0 curve. We can query the distances of any points in the space w.r.t the object reference frame. Object-centric SDF can be approximated by simple geometry primitives, such as spheres and capsules [10] or deep neural networks [11], [12]. Object-centric approaches majorly focus on object surface reconstruction and are beneficial in robotic tasks, such as manipulation [13]–[16]. However, object-centric approaches majorly focus on object surface reconstruction but are lacking accuracy outside of the object boundary. The incorrect distance information may provide a misleading sense of safety and cause hazardous behavior. In addition, existing approaches do not consider the articulated objects, while our proposed approach tries to fill this gap.

In this paper, we present ReDSDF, an approach that extends the concept of SDFs for arbitrary **articulated objects** such as robotic manipulators and humans. Our method provides a single deep model that learns distance fields which can be used **at any scale**. Notably, we define a simple yet effective **inductive bias** that regularizes implicit function to approximate the distance by the L2 norm w.r.t. the object's center. This regularization allows us to obtain expressive deep SDFs that can provide precise distance at the proximity of the object and meaningful level curves when the robot is far from it. We encourage the reader to read *full paper* for more technical details [17].

Technische Universität Darmstadt, Karolinenpl. 5, 64289 Darmstadt, Germany puze@robot-learning.de, kuo.zhang@stud.tu-darmstadt.de, {davide.tateo, snehal.jauhri, jan.peters, georgia.chalvatzaki} @tu-darmstadt.de

This research received funding by the DFG Emmy Noether Programme (#448644653), the RoboTrust project of the Centre Responsible Digitality Hessen, Germany and the China Scholarship Council (No. 201908080039).

II. Learning ReDSDFnd Robot Control

We define ReDSDF w.r.t. a given (articulated) object with configuration $\boldsymbol{q}$ at any query point $\boldsymbol{x}$ as the distance

$$d_{\boldsymbol{q}}(\boldsymbol{x}) = (1 - \sigma_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{q}))\, f_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{q}) + \sigma_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{q}) \|\boldsymbol{x}\|_2, \qquad (1)$$

with the fully connected network $f_{\boldsymbol{\theta}}$ which takes the articulated target configuration $\boldsymbol{q}$ and the query point $\boldsymbol{x} \in \mathbb{R}^3$ as input. $\sigma_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{q}) = \mathrm{sigmoid}\left(\alpha_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{q})\left(\|\boldsymbol{x}\|_2 - \rho_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{q})\right)\right)$ is a mode switching function with shaping parameter $\alpha_{\boldsymbol{\theta}}$ and $\rho_{\boldsymbol{\theta}}$. Here, the $\rho_{\boldsymbol{\theta}}$ defines a (soft) threshold for switching from the network approximated distance $f_{\boldsymbol{\theta}}$ to the regularized distance $\|\boldsymbol{x}\|_2$. The $\alpha_{\boldsymbol{\theta}}$ regulates the sharpness of the change between the two regimes. We omit the network details due to the page limit.

To generate the data, we first obtain a point cloud of the target of interest from different viewpoints and at different configurations $\boldsymbol{q}$. We estimate the normal direction $\bar{\boldsymbol{n}}_{\boldsymbol{q}}(\boldsymbol{x})$ for each point $\boldsymbol{x}$ in the point cloud. We then prune the points whose normal are not consistent with the neighbors. Using the estimated normals, we augment the data following the same procedure of [12], i.e., performing forward tracing along the normal direction of each point and rejecting the points that do not re-project on the original point based on KD-tree. Finally, we assign to every generated point $\boldsymbol{x}$ the distance from the object $\bar{d}_{\boldsymbol{q}}(\boldsymbol{x})$ and a weight $\omega_{\boldsymbol{q}}(\boldsymbol{x})$ using the inverse of the success rate of the augmentation. To obtain the distance field of human model of various configurations, we leverage the realistic 3D model of the human body, SMPL [18], and the AMASS [19] dataset.

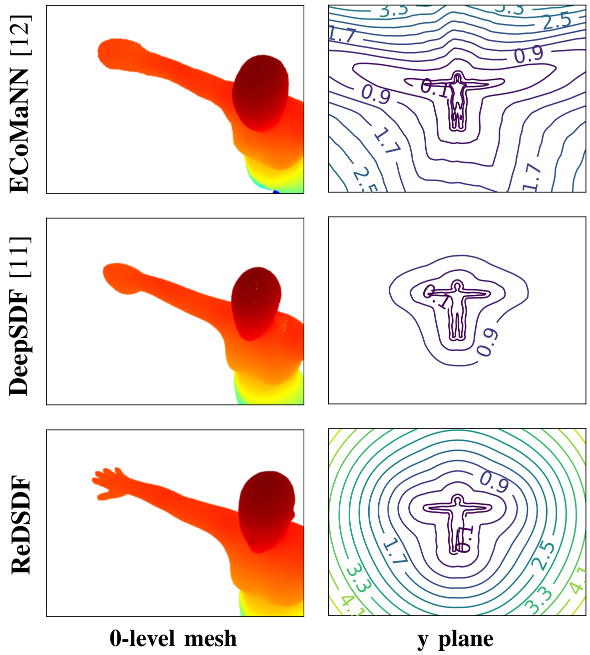

Fig. 1: Reconstruction of the human distance function

We train the model by optimizing the following loss function over the collected dataset $\mathcal{D}$

$$\mathcal{L}(\mathcal{D}) = \sum_{\mathcal{D}} \omega_{\boldsymbol{q}}(\boldsymbol{x}) \left(\bar{d}_{\boldsymbol{q}}(\boldsymbol{x}) - d_{\boldsymbol{q}}(\boldsymbol{x})\right)^2 + \left(\|D_{\boldsymbol{q}}(\boldsymbol{x})\bar{\boldsymbol{n}}_{\boldsymbol{q}}(\boldsymbol{x})\|_2^2 + \|N_{\boldsymbol{q}}(\boldsymbol{x})\nabla_{\boldsymbol{x}} d_{\boldsymbol{q}}(\boldsymbol{x})\|_2^2\right) + \gamma \rho_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{q})^2, \quad (2)$$

where $D_{\boldsymbol{q}}(\boldsymbol{x}) = \text{null}\left(\nabla_{\boldsymbol{x}} d_{\boldsymbol{q}}(\boldsymbol{x})\right)$ and $N_{\boldsymbol{q}}(\boldsymbol{x}) = \text{null}\left(\bar{\boldsymbol{n}}_{\boldsymbol{q}}(\boldsymbol{x})\right)$ are, respectively, the null-space of the gradient of the distance field, and the null space of the normal of the points. The first term in (2) is the weight squared error of the distance approximation. The second and the third term is similar to the one proposed by [12] to align the estimated normals of the points. The last component of the loss is a regularization term, where γ is a regularization coefficient, set in our experiments to 0.02. Reducing the output of the learnable function $\rho_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{q})$ has the effect of switching the distance regime of the network as soon as possible. This regularization is important in particular where the dataset is sparse.

III. Experiments

We compare ReDSDF against the state-of-the-art approaches Equality Constraint Manifold Neural Network (ECoMaNN) [12] and DeepSDF [11] for a fixed human configuration. All methods originate from the same mesh file but use individual data-generation techniques to train the SDF. As illustrated in Figure 1, all methods reconstruct roughly the human shape. However, ReDSDF provides more precise details like human fingers and ears. Our approach exploits the inductive bias to enforce a well-behaved, even if approximate, distance field.

To evaluate the applicability of ReDSDF for robot control, we devise a WBC for the bimanual mobile manipulator TIAGo++ robot (Fig. 2a). Note that the distance field is learned by excluding the controlled arm from the robot itself; otherwise, every point is considered a self-collision. Fig. 2a. demonstrates the learned ReDSDF and Points of Interest (PoI)s for the distance query. The distance field for the other arm is computed by mirroring through the symmetry plane. We use a PID controller to provide the control velocity of the end-effector to reach the target. The repulsive field generates joint velocities to avoid collisions based on the distance and the gradient obtain from ReDSDF. We compare our WBC to one without obstacle avoidance and to another one that uses Artificial Potential Fields (APF)s using a sphere-based distance field. We keep the PID controller the same but tune the repulsive controllers to achieve a task-oriented controller or a cautious controller.

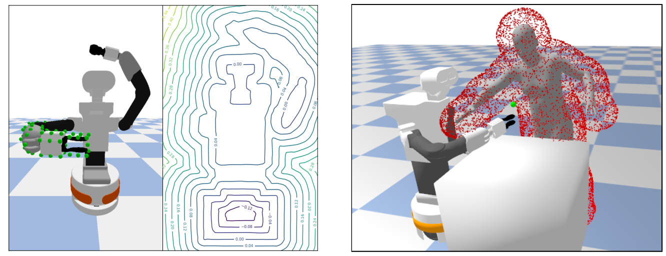

Fig. 2: a): ReDSDF for WBC. *Left:* PoI of the controlled arm that we check to avoid collisions. *Right:* The distance field of the robot, excluding the controlled arm. b): Human-robot shared-workspace simulated task.

To test the applicability of ReDSDF for reactive motion generation for challenging HRI tasks, we built a simulation task where a human and a robot perform sequential pick-and-place-like actions in a shared workspace. To simulate realistic movements, we record a set of trajectories from an actual human performing the task, and we infer the SMPL trajectories with the help of VIBE [20]. We replay these trajectories in our PyBullet simulator, constituting a novel way of simulating HRI tasks with the robot dynamically interacting with the human (Fig. 2b). However, how to obtain reliable human pose in real time remains an open question.

Our experiment shows that the batch queries of the distance field enable the controller to achieve exquisite collision avoidance. The smooth distance function with a well-defined gradient avoids oscillating movements while sphere based approach introduces discontinuity when closest sphere changes. Please find our videos: `https://irosalab.com/2022/02/28/redsdf`.

IV. Conclusions

In this work, we presented the Regularized Deep Signed Distance Fields (ReDSDF) framework, which generalizes the concept of SDF to arbitrary articulated objects. We use induced bias to normalize the training of the distance function to ensure that the distance field is smooth and of high quality near the target and reasonable at a distance. Deep neural networks can also perform distance queries in batches, which allows the robot controller to perform delicate collision avoidance. Our method can be easily deployed to generate reactive motions both in the context of a whole-body control (WBC) and in a human-robot shared workspace scenario, using real human data.

References

[1] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in *SIGGRAPH*, 1996.

[2] M. Zollhöfer, A. Dai, M. Innmann, C. Wu, M. Stamminger, C. Theobalt, and M. Nießner, "Shading-based refinement on volumetric signed distance functions," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 4, pp. 1–14, 2015.

[3] H. Oleynikova, A. Millane, and Z. Taylor et al., "Signed distance fields: A natural representation for both mapping and planning," in *RSS Workshops*, 2016.

[4] H. Oleynikova, Z. Taylor, M. Fehr, R. Siegwart, and J. Nieto, "Voxblox: Incremental 3d euclidean signed distance fields for on-board mav planning," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 1366–1373.

[5] D. R. Canelhas, "Truncated signed distance fields applied to robotics," Ph.D. dissertation, Örebro University, 2017.

[6] K. Saulnier, N. Atanasov, G. J. Pappas, and V. Kumar, "Information theoretic active exploration in signed distance fields," in *ICRA*, 2020.

[7] M. N. Finean, W. Merkt, and I. Havoutis, "Simultaneous scene reconstruction and whole-body motion planning for safe operation in dynamic environments," in *IROS*, 2021.

[8] L. Han, F. Gao, B. Zhou, and S. Shen, "Fiesta: Fast incremental euclidean distance fields for online motion planning of aerial robots," in *IROS*, 2019.

[9] S. Izadi, D. Kim, O. Hilliges, and D. Molyneaux et al., "Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera," in *ACM Symp. on User interface software and tech.*, 2011.

[10] N. Ratliff, M. Zucker, J. A. Bagnell, and S. Srinivasa, "Chomp: Gradient optimization techniques for efficient motion planning," in *ICRA*, 2009.

[11] J. J. Park and P. Florence et al., "Deepsdf: Learning continuous signed distance functions for shape representation," in *CVPR*, 2019.

[12] G. Sutanto, I. M. R. Fernández, and P. Englert et al., "Learning equality constraints for motion planning on manifolds," in *CoRL*, 2020.

[13] B. Xu, W. Li, and D. Tzoumanikas et al., "Mid-fusion: Octree-based object-level multi-instance dynamic slam," in *ICRA*, 2019.

[14] M. Grinvald, F. Furrer, and T. Novkovic et al., "Volumetric instance-aware semantic mapping and 3d object discovery," *IEEE RA-L*, 2019.

[15] J. McCormac, R. Clark, M. Bloesch, A. Davison, and S. Leutenegger, "Fusion++: Volumetric object-level slam," in *3DV*, 2018.

[16] S. Peng, M. Niemeyer, L. Mescheder, M. Pollefeys, and A. Geiger, "Convolutional occupancy networks," in *ECCV*, 2020.

[17] P. Liu, K. Zhang, D. Tateo, S. Jauhri, J. Peters, and G. Chalvatzaki, "Regularized deep signed distance fields for reactive motion generation," *arXiv preprint arXiv:2203.04739*, 2022.

[18] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "Smpl: A skinned multi-person linear model," *ACM TOG*, 2015.

[19] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black, "AMASS: Archive of motion capture as surface shapes," in *ICCV*, 2019.

[20] M. Kocabas, N. Athanasiou, and M. J. Black, "Vibe: Video inference for human body pose and shape estimation," in *CVPR*, 2020.