iSDF: Real-Time Neural Signed Distance Fields for Robot Perception

Joseph Ortiz\textsuperscript{1,2} Alexander Clegg\textsuperscript{2} Jing Dong\textsuperscript{3} Edgar Sucar\textsuperscript{1}
David Novotny\textsuperscript{2} Michael Zollhoefer\textsuperscript{3} Mustafa Mukadam\textsuperscript{2}

Abstract—We present iSDF, a continual learning system for real-time signed distance field (SDF) reconstruction. Given a stream of posed depth images from a moving camera, it trains a randomly initialised neural network to map input 3D coordinate to approximate signed distance. The model is self-supervised by minimising a loss that bounds the predicted signed distance using the distance to the closest sampled point in a batch of query points that are actively sampled. In contrast to prior work based on voxel grids, our neural method is able to provide adaptive levels of detail with plausible filling in of partially observed regions and denoising of observations, all while having a more compact representation. In evaluations against alternative methods on real and synthetic datasets of indoor environments, we find that iSDF produces more accurate reconstructions, and better approximations of collision costs and gradients useful for downstream planners in domains from navigation to manipulation. Code and video results can be found at our project page: https://joeortiz.github.io/iSDF/.

I. INTRODUCTION

Signed distance fields are a common map representations in both robotics and vision that associate each point in space with the signed distance to the closest surface, encoding the surface as the zero level set. In robotics, SDFs are a commonly used for collision avoidance in motion planning [24], [20], [17], [4], [12], however they are usually assumed as given a priori or too expensive to compute in real-time. On the other hand, in real-time vision systems, depth fusion is commonly used to produce truncated signed distance fields (TSDF) [14], [13], but there is little work on reconstructing non-truncated SDFs in room scale environments. In order to close this gap, in this work, we focus on the problem of real-time full SDF reconstruction.

Prior work on real-time reconstruction of non-truncated SDFs [15], [7] (most notably Voxblox [15]) operates in two stages, first reconstructing surface geometry and then transforming it to the SDF using wavefront propagation algorithms. These methods are based on voxel grids and are limited to a resolution of 5cm by real-time constraints due to the cost of wavefront propagation.

Neural fields provide an alternative paradigm to voxel grids for modelling scenes [16], [10], [11] and can be optimised from scratch to accurately fit a specific scene. Building on recent advances, we present iSDF (incremental Signed Distance Fields), a system for real-time SDF reconstruction based on a neural signed distance field that maps a 3D coordinate $x$ to the signed distance value $s$, $f(x; \theta) = s$. Specifically we build two advances; iMAP [18] which demonstrated that neural radiance fields can be trained in real-time as part of a SLAM system and the body of work using neural SDFs for object modelling [6], [23], [8], [22], [21], [9], [1].

II. iSDF - REAL-TIME SDF RECONSTRUCTION

iSDF is a system for real-time SDF reconstruction that takes as input a stream of posed depth images captured by a moving camera and, during online operation, optimises a neural SDF of the environment. We assume an external tracking module and focus on mapping.

The SDF is modelled by a randomly initialised MLP with 4 hidden layer and the “off-axis” positional embedding [2]. At each iteration, we select a subset of frames, i.e. a few posed depth images. Given the continual learning setting, we follow iMAP [18] and select frames via active sampling from a sparse set of representative keyframes to mitigate catastrophic forgetting. For each frame, we sample 200 random pixels and then, along each backprojected ray, generate 20 stratified samples, 8 Gaussian samples around the measured depth and one sample at the depth. These points for all sampled pixels form the batch of points that are fed to the MLP.

To supervise the predicted signed distances, we propose to use the distance to the closest surface point in the batch. This distance $b$ will be larger than the true signed distance and so we use it to bound the prediction via the loss:

$$L_{df}(f(x; \theta), b) = \max(0, e^{-\beta f(x; \theta)} - 1, f(x; \theta) - b) .$$  \(1\)

Close to the surface, we expect the bound to be very tight and so we instead directly supervise the SDF prediction to take the bound value. The closest surface point in the batch can also provide signal to supervise the spatial gradient of the predicted signed distance and we introduce a loss $L_{\text{grad}}$ penalising the cosine distance between the predicted signed distance and the vector from the closest surface point to the query point. Following Gropp et al. [6] we also apply Eikonal regularisation: $L_{\text{eik}}(f(x; \theta)) = ||\nabla f(x; \theta)|| - 1$. The network parameters are optimised to minimise the loss:

$$l(\theta) = L_{df} + 0.02L_{\text{grad}} + 0.25L_{\text{eik}} .$$  \(2\)

III. RESULTS

Experimental setup. We experiment with sequences from the ScanNet dataset [3] and the synthetic ReplicaCAD dataset [19] (in which sequences are generated by simulating the measurements by a camera mounted on a mobile manipulator). To evaluate the SDF reconstruction, we pre-compute a voxel grid (resolution 1cm) of ground truth SDF.

\textsuperscript{1}Imperial College London, \textsuperscript{2}Meta AI, \textsuperscript{3}Reality Labs Research.

joeortiz16@gmail.com
values by voxelizing the mesh into an occupancy grid and then computing the Euclidean distance transform (EDT). We compare iSDF against two voxel-based methods for real-time SDF reconstruction: (1) Voxelblox [15] and (2) KinectFusion+. Both methods operate in two stages, first reconstructing the surface before computing the SDF via wavefront propagation. Voxelblox is a CPU only method and to the best of our knowledge there are no GPU methods for real-time SDF reconstruction. Consequently, KinectFusion+ is our own GPU implementation in C++ with custom CUDA kernels for fusion and the efficient EDT [5].

**Quantitative results.** To visualise the reconstructed SDFs, we compare 2D slices of the SDFs at constant heights in Fig. 2. Unlike Voxelblox and KinectFusion+, which only map the visible region, iSDF is predictive and reconstructs a plausible field in the full environment. For example, in the interiors of the beanbag (highlighted by the orange boxes in Fig. 1) or the interior of a wall, iSDF is the only method that returns predictions. As Voxelblox and KinectFusion+ employ fusion in the first stage, there are holes in the reconstructions in regions that no rays reach, shown in white in the slices. These holes are problematic for downstream planners as they must be marked as un navigable and given a high cost.

**Additional results.** We provide further qualitative and quantitative results for more sequences on our website ([https://joeaortiz.github.io/iSDF/](https://joeaortiz.github.io/iSDF/)). We show that iSDF inherits the positive characteristics of neural fields. In contrast to prior work based on voxel grids, iSDF is: (i) efficient - it can seamlessly allocate memory capacity to model different parts of an environment with different levels of detail, and (ii) predictive - it can denoise and consolidate noisy measurements and sensibly fill in gaps in partially observed regions; all while requiring only 1MB of memory for the network weights. Results demonstrating these properties can be found on our project website.

**IV. LIMITATIONS AND CONCLUSIONS**

There remain many important challenges in training neural fields in real-time, for example: handling dynamic environments, using local models to reduce replay, and designing more flexible positional embeddings. In addition, for downstream planning, it would be useful to have a measure of confidence associated with the iSDF predictions.

To conclude, we have presented iSDF, a continual learning system for real-time signed distance field reconstruction. Key to our approach is a novel loss that bounds the predicted signed distance using the distance to the closest batch surface point to enable learning signed distances away from surfaces. In evaluations against alternative methods on real and synthetic datasets, we show that iSDF produces more accurate reconstructions as well as better collision costs and gradients for downstream planners. We hope that the generality and differentiability of iSDF means it can easily be integrated into downstream robotics applications.
