NeuSE: Neural SE(3)-Equivariant Embedding for Consistent Spatial Understanding with Objects

Jiahui Fu, Yilun Du, Kurran Singh, Joshua B. Tenenbaum, and John J. Leonard
MIT CSAIL, Marine Robotics Group

Scan for Project Page

Goals

Leverage neural implicit object representations for consistent long-term scene understanding.

What is NeuSE?

- A novel Neural SE(3)-Equivariant Embedding for objects
- A compact point cloud surrogate for complete object models, encoding full shape information while transforming SE(3) equivariantly with the object in the physical world.

Contributions

- NeuSE, a neural SE(3)-equivariant embedding for objects, encoding the full object shape and transforms SE(3)-equivariantly with the real-world object.
- A NeuSE-based object SLAM paradigm targeting long-term scene inconsistencies, enabling NeuSE-predicted object-level localization and change-aware mapping.

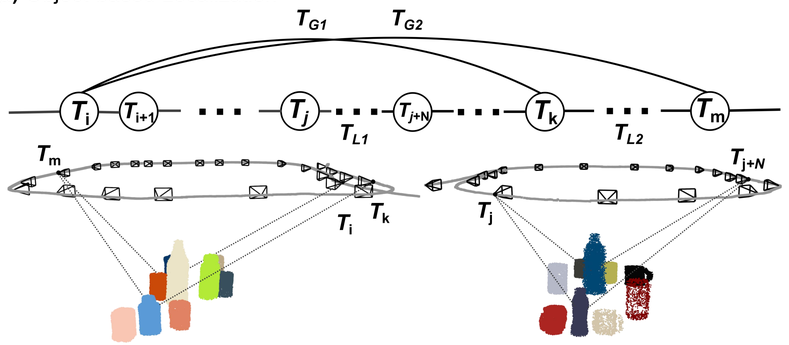
Experiments show:

- Synthetic and Real-world sequences with object changes
- Improved localization performance and flexible mapping capability when working standalone or jointly with a common SLAM pipeline.

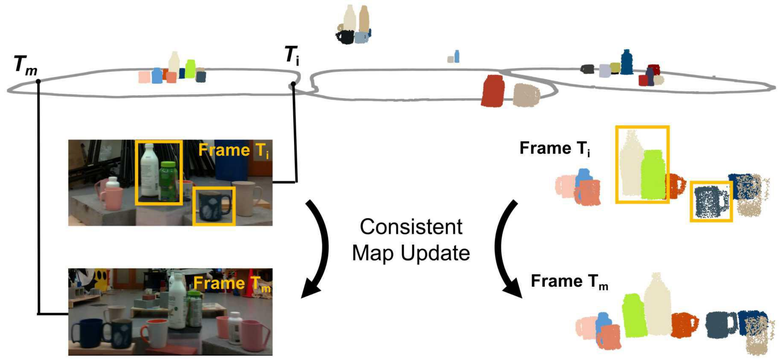
Contact: Jiahui Fu jiahuifu@mit.edu

Methodology

(a) Object-based Localization



(b) Object-centric Change Detection and Mapping



(c) Category-level SE(3)-Equivariant Embeddings for Objects

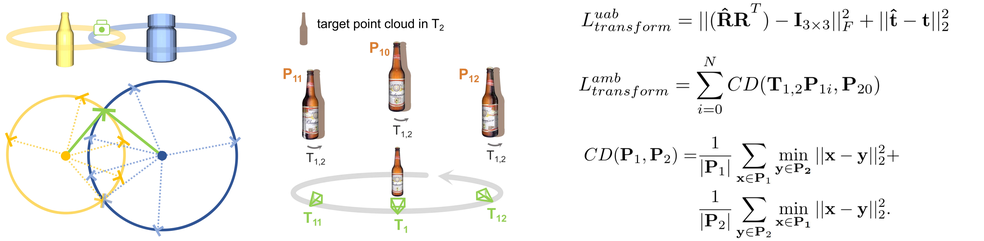
$$f_\theta(\mathbf{RP}) = \mathbf{Rz}, \mathbf{R} \in \mathbb{SO}(3)$$

$$f_\theta(\mathbf{P} - \overline{\mathbf{P}}) = \mathbf{z}_0 : \mathbb{R}^{n\times 3} \rightarrow \mathbb{R}^{k\times 3}$$

$$\mathbf{z} = \mathbf{z}_0 + \overline{\mathbf{P}}, \mathbf{z}' = \mathbf{z} - \overline{\mathbf{z}}$$

$$\Phi(\mathbf{x}, \mathbf{f}_\theta(\mathbf{P})) = \Phi(\mathbf{x} - \overline{\mathbf{z}}, \mathbf{z}') : \mathbb{R}^3 \times \mathbb{R}^{k\times 3} \rightarrow [0, 1]$$

(d) Training Objectives for Ambiguous and Unambigous Object Categories



$$L^{uab}_{transform} = ||(\hat{\mathbf{R}}\mathbf{R}^T) - \mathbf{I}_{3\times 3}||^2_F + ||\hat{\mathbf{t}} - \mathbf{t}||^2_2$$

$$L^{amb}_{transform} = \sum_{i=0}^{N} CD(\mathbf{T}_{1,2}\mathbf{P}_{1i}, \mathbf{P}_{20})$$

$$CD(\mathbf{P}_1, \mathbf{P}_2) = \frac{1}{|\mathbf{P}_1|}\sum_{\mathbf{x}\in\mathbf{P}_1}\min_{\mathbf{y}\in\mathbf{P}_2}||\mathbf{x}-\mathbf{y}||^2_2 + \frac{1}{|\mathbf{P}_2|}\sum_{\mathbf{y}\in\mathbf{P}_2}\min_{\mathbf{x}\in\mathbf{P}_1}||\mathbf{x}-\mathbf{y}||^2_2.$$

Results