

TACO: Temporal Consensus Optimization for Continual Neural Mapping

Xunlan Zhou*

School of Intelligence Science
and Technology
Nanjing University

Email: wyattzhouxl@mail.nju.edu.cn

Hongrui Zhao*

Department of Aerospace Engineering
University of Illinois Urbana-Champaign
Email: hongrui5@illinois.edu

Negar Mehr

Department of Mechanical Engineering
University of California Berkeley
Email: negar@berkeley.edu

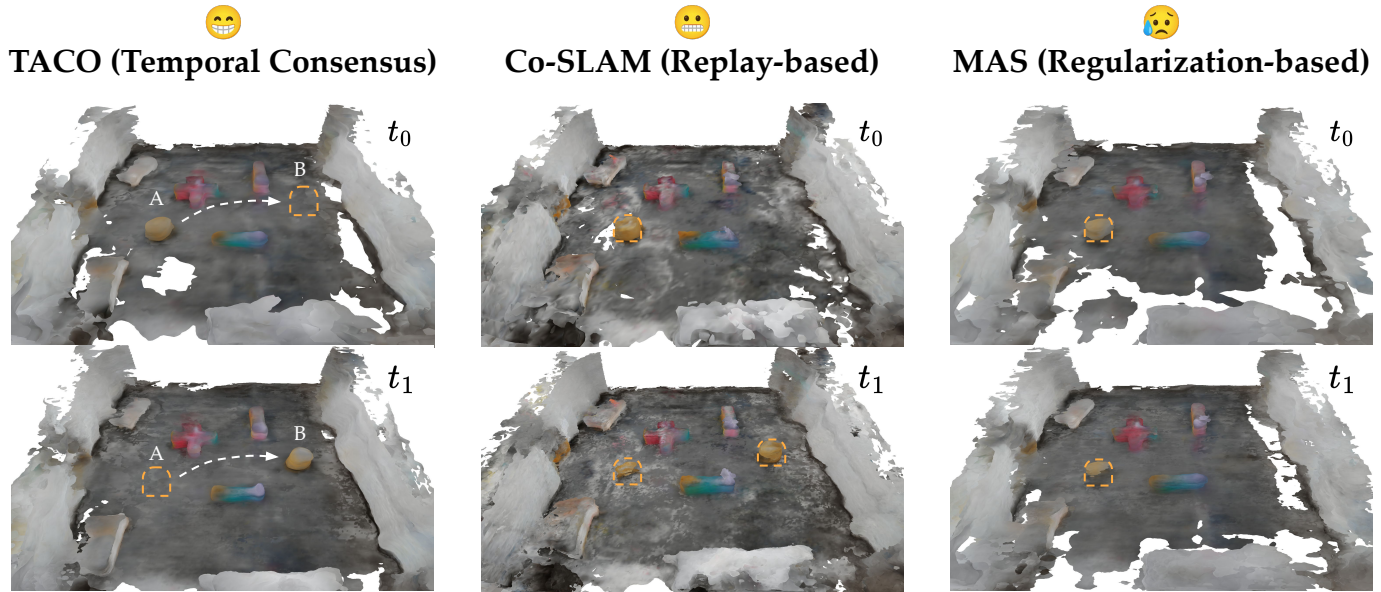


Fig. 1: **Comparison of continual neural mapping under environment changes.** The yellow stool was moved from *A* to *B*. **Left:** TACO (Temporal Consensus Optimization) revises outdated regions while preserving consistent geometry and yields an accurate and up-to-date map. **Middle:** Replay-based methods (e.g., Co-SLAM) preserve high-quality reconstructions overall, but their reliance on replaying past observations can lead to persistent “ghost” artifacts when scene geometry changes. **Right:** Regularization-based methods (e.g., MAS) over-constrain the optimization, unable to adapt to the changing environment.

Abstract—Neural implicit mapping has emerged as a powerful paradigm for robotic navigation and scene understanding. However, real-world robotic deployment requires continual adaptation to changing environments under strict memory and computation constraints, which existing mapping systems fail to support. Most prior methods rely on replaying historical observations to preserve consistency and assume static scenes. As a result, they cannot enable continual learning in dynamic robotic tasks. To address these challenges, we propose TACO (Temporal Consensus Optimization), a replay-free framework for continual neural mapping. We reformulate mapping as a temporal consensus optimization problem, where we treat past model snapshots as temporal neighbors. Intuitively, our approach resembles a model consulting its own past knowledge. We update the current map by enforcing weighted consensus with historical representations. Our method allows reliable past geometry to constrain optimization while permitting unreliable or outdated regions to be revised in response to new observations. TACO achieves a balance between memory efficiency and adaptability

without storing or replaying previous data. Through extensive simulated and real-world experiments, we show that TACO robustly adapts to scene changes, and consistently outperforms other continual learning baselines. The code can be found at <https://iconlab.negarmehr.com/TACO>

I. INTRODUCTION

Neural implicit representation has become an important paradigm for robotic dense scene mapping [23, 47, 53, 55]. Unlike traditional discrete map representations [27, 28, 48], they encode scene geometry and appearance as continuous functions, which yields compact and expressive scene models with high reconstruction fidelity and low memory cost. Recent work demonstrates that neural implicit maps support direct optimization from streaming RGB-D observations and enable online dense mapping at real-time rates [42, 45, 59]. Despite their success, most existing online neural mapping systems rely on storing and replaying past observations or keyframes

*Equal contribution.

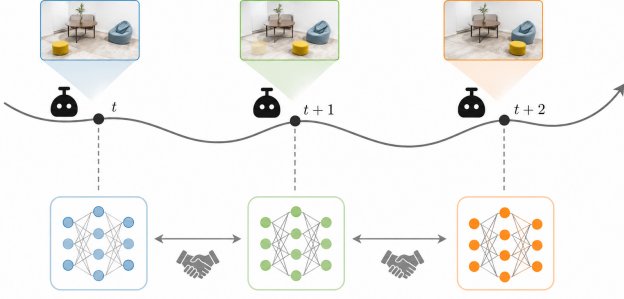


Fig. 2: **Intuition of temporal consensus.** A robot receives sequential visual observations over time, while neural map snapshots are encouraged to reach consensus across adjacent time steps. This allows the map to preserve reliable past knowledge while adapting to newly observed scene changes.

to stabilize training and prevent the forgetting of the previously visited regions. However, in dynamic real-world environments, past and new observations often conflict. Training on such inconsistent data degrades map quality and negatively impacts downstream tasks which require up-to-date environmental information. In addition to keyframe replay, methods such as Continual Neural Mapping (CNM) [51] and UNIKD [15] draw inspiration from continual learning literature [9, 15, 29, 33, 39] to prevent forgetting. However, these approaches often struggle to adapt to dynamic scenes, as they overly constrain updates to prevent deviation from the prior model.

To address these challenges, we introduce TACO, a Temporal Consensus Optimization framework for continual neural mapping. We preserve previous model parameters as “historical snapshots”. Upon receiving new observations, a *temporal consensus* process enforces consistency between the current model parameters and these snapshots. We also incorporate an importance-aware weighting scheme to enable a more adaptive consensus enforcement. Intuitively, our weighting allows the current model to align with reliable aspects of its past states while remaining free to revise regions that conflict with new observations. As illustrated in Fig. 2, a robot receives sequential observations over time, while the corresponding neural map snapshots are encouraged to reach consensus across adjacent time steps.

Our approach avoids catastrophic forgetting and allows adaptation to the changing environment, as illustrated in Fig. 1. Our formulation enables continual adaptation under streaming observations *without relying on data replay* and remains well-suited for long-term online mapping under realistic resource constraints. Through extensive experiments, we demonstrate that we consistently produce more accurate and up-to-date reconstructions under environment changes compared to existing baselines.

In summary, our key contributions in this paper are:

- 1) We propose TACO, a Temporal Consensus Optimization framework that reformulates continual neural mapping

as a consensus problem over time.

- 2) We introduce an importance-aware temporal consensus mechanism that dynamically balances stability and adaptation: we preserve reliable past knowledge, while updating regions that conflict with new observations.
- 3) We validate TACO through extensive simulation and real-world robotic experiments. We show that TACO adapts to dynamic scene changes while state-of-the-art methods fail. Our approach mitigates forgetting under realistic online and memory-constrained conditions.

II. RELATED WORKS

Continual Learning. Continual learning studies how a model acquires knowledge over time from sequential data streams while preserving useful past information and adapting to new data [9, 29, 35, 46]. A widely studied subset is incremental learning which considers settings where training data arrive sequentially and only a subset of observations is available at each step [1, 9]. Most incremental learning methods address catastrophic forgetting under the assumption of a stationary task or data distribution and typically fall into three categories [9]: data replay [7, 19, 21, 32, 34, 39], parameter regularization [3, 18, 20, 31], and parameter isolation [2, 13, 22, 50]. Representative examples include replay-based strategies such as PTAM with keyframe replaying [19], regularization-based approaches like MAS [3] and EWC [18], parameter isolation methods such as PackNet [22], and knowledge distillation techniques such as POD [12], AFC [16], and UNIKD [15]. Some of these strategies have also been applied to neural implicit representations. Examples include Continual Neural Mapping (CNM) [51] for neural surface fields, CLNeRF [6] for NeRF-based novel view synthesis, and UNIKD [15], which leverages uncertainty-filtered distillation to mitigate forgetting in incremental neural implicit reconstruction.

While effective at preserving previously reconstructed content, these approaches primarily focus on consistency with past model states. In neural mapping systems, this emphasis often manifests as aligning with earlier reconstructions, even when new sensor observations indicate geometric changes. As a result, the map can retain outdated structure and fail to reflect scene evolution over time. Our work instead considers a broader continual mapping setting [9, 29], where the objective extends beyond preventing forgetting to updating historical geometry in response to new observations without requiring access to past data.

Consensus Optimization for Neural Mapping. Consensus optimization provides an approach to reconcile conflicting information across multiple learners by combining local updates with consensus constraints [26]. In neural mapping, this paradigm has been explored primarily in multi-robot settings, where independent agents observe different parts of the scene and must agree on a shared representation. Prior work applies consensus-based optimization to neural implicit mapping in order to maintain geometric consistency across agents under

limited or unreliable communication [4, 10, 11, 54, 56–58]. Although originally developed for spatially distributed systems, these methods reveal a broader insight: *consensus mechanisms, especially when weighted by importance, provide an effective means of resolving conflicting information without requiring access to raw data*. In this work, we build on this insight by shifting the notion of consensus from the spatial domain to the temporal domain. Rather than coordinating multiple agents, we propose to coordinate multiple model states over time, which enables continual neural mapping that updates outdated geometry while preserving reliable structure.

Finding the Important Parameters in Neural Implicit Maps. Quantifying the importance of model parameters in neural implicit maps is relevant to tasks such as view selection, refinement, and artifact suppression, as well as to continual and incremental optimization. Much of the existing literature approaches this problem from an uncertainty quantification perspective. Classical techniques such as deep ensembles [43] and Bayesian neural networks [37] provide principled uncertainty quantification approaches. Alternative methods estimate spatial uncertainty by perturbing inputs or parameters [14, 52]; these approaches primarily capture output variability and do not explicitly characterize the relative importance of individual learnable parameters. More recent works utilize computationally efficient uncertainty proxies to accelerate real-time mapping. RAMEN [57] estimates parameter relevance using frequency-based statistics over multi-resolution hash grids to modulate consensus in distributed neural mapping. UNIKD [15] incorporates uncertainty estimates into a distillation framework to filter unreliable knowledge when transferring information across incremental training stages. In contrast, our work adopts an output-sensitivity-based importance formulation inspired by Memory Aware Synapses (MAS) [3]. We quantify parameter importance through the sensitivity of network outputs to parameter perturbations. This measure reflects the temporal reliability of learned representations, and we use it to regulate continual optimization without introducing additional network components or auxiliary inference procedures.

III. PRELIMINARIES

In this section, we introduce the core building blocks and notations used throughout the paper. We first review Co-SLAM [45], a real-time neural implicit mapping framework that serves as the optimization backbone of our approach. We then provide an overview of the method of multipliers [38, 40], which provides the algorithmic foundation of our work for later incorporating temporal consensus into neural mapping.

A. Neural Implicit Mapping with Co-SLAM

Co-SLAM represents a scene using a neural implicit map composed of three components: a multi-resolution hash-based feature grid [25], denoted as $V_\alpha = \{V_\alpha^l\}_{l=1}^L$, a geometry decoder F_τ , and a color decoder F_ϕ . The feature grid contains L resolution levels with geometrically increasing resolutions, where higher levels encode finer geometric details. We denote the learnable parameters of the neural implicit map as

$\Theta = \{\alpha, \tau, \phi\}$, where α parameterizes the hash grid, and τ and ϕ parameterize the geometry and color decoders, respectively.

Given a 3D point with world coordinate x , the neural implicit map predicts its truncated signed distance function (SDF) value s and RGB color c . The SDF value encodes the signed distance to the closest surface, with the sign indicating whether the point lies inside or outside the surface. To improve geometric smoothness, one-blob positional encoding [24] $\gamma(\cdot)$ is applied to x . Using the encoded coordinate $\gamma(x)$ and interpolated feature vectors $V_\alpha(x)$, the geometry decoder outputs a latent feature h and the SDF value:

$$F_\tau(\gamma(x), V_\alpha(x)) \rightarrow (h, s). \quad (1)$$

The color decoder then predicts the RGB value:

$$F_\phi(\gamma(x), h) \rightarrow c. \quad (2)$$

Co-SLAM renders RGB and depth images by volume rendering along camera rays. Given camera intrinsics K and pose P , each image pixel with coordinates (u, v) corresponds to a ray with origin o and direction $r = PK^{-1}(u, v, 1)^\top$. Along each ray, M sample points $x_i = o + d_i r$ are evaluated, where d_i denotes the distance from the camera origin. For each x_i , the neural implicit map predicts (s_i, c_i) using (1) and (2). The rendered pixel color $\hat{c}$ and depth $\hat{d}$ are computed as weighted averages:

$$\hat{c} = \frac{1}{\sum_{i=1}^M w_i} \sum_{i=1}^M w_i c_i, \quad (3a)$$

$$\hat{d} = \frac{1}{\sum_{i=1}^M w_i} \sum_{i=1}^M w_i d_i, \quad (3b)$$

where w_i is the weight quantifying the contribution of point i to the final rendering [5]. We truncate the SDF value s to tr to exclude the points far from the surface, as these regions provide little information for reconstruction. Then, we compute the weights as

$$w_i = \sigma\left(\frac{s_i}{tr}\right) \sigma\left(-\frac{s_i}{tr}\right), \quad (4)$$

with $\sigma(\cdot)$ denoting the Sigmoid function. This formulation assigns high weights to samples whose signed distance s_i is close to zero, which corresponds to points near the estimated surface intersection along the ray, while suppressing contributions from samples in free space. The sigmoid functions provide a smooth truncation controlled by tr , avoiding hard thresholds and ensuring stable gradients during optimization by softly limiting the influence of samples far from the surface.

The neural implicit map is optimized by minimizing the discrepancy between rendered RGB-D images and observations collected by the robot. Let R denote the incoming RGB-D frames. The training loss is defined as

$$L^{obj}(\Theta, R) = L_{rgb} + L_d + L_{sdf} + L_{fs} + L_{smooth}, \quad (5)$$

where L_{rgb} and L_d measure the photometric and depth reconstruction errors, L_{sdf} supervises SDF predictions, L_{fs} constrains points far from surfaces, and L_{smooth} encourages

geometric smoothness. We refer readers to [45] for detailed definitions of each loss term.

B. Constrained Optimization with the Method of Multipliers

In this section, we review the method of multipliers [40] for solving constrained optimization problems, which serves as a key algorithmic building block in our approach. Consider the following optimization problem

$$\min_{\Theta} L^{obj}(\Theta, R), \text{ s.t. } A\Theta = z, \quad (6)$$

where the linear equality constraint $A\Theta = z$ defines the feasible parameter space. Here, A is the constraint matrix and z is the target vector. The neural implicit map parameters are flattened into a vector, denoted as Θ , which serves as the optimization variable.

To solve (6), we can construct the augmented Lagrangian of the problem:

$$\mathcal{L}^a = L^{obj}(\Theta, R) + (A\Theta - z)^\top p + \frac{\rho}{2} \|A\Theta - z\|_2^2, \quad (7)$$

where p is the vector of dual variables, and $\rho > 0$ is a penalty scalar parameter controlling the strength of enforcing constraint. We apply dual ascent to minimize the augmented Lagrangian (7), which iteratively updates the dual and the primal variables. At iteration k , the updates take the form:

$$\begin{aligned} \Theta^k &= \arg \min_{\Theta} L^{obj}(\Theta, R) + (A\Theta - z)^\top p^{k-1} \\ &\quad + \frac{\rho}{2} \|A\Theta - z\|_2^2, \end{aligned} \quad (8a)$$

$$p^k = p^{k-1} + \rho (A\Theta^k - z). \quad (8b)$$

The update in (8a) is referred to as the *primal update*, while (8b) corresponds to the *dual update*. Intuitively, the dual variable acts as accumulated disagreement signals that record how much our primal variable Θ violates the constraint over past iterations. The dual variable drives subsequent updates to correct these discrepancies. Since Θ is the neural network parameters, we only approximate the primal update using a small number of stochastic gradient descent steps. In the next section, we will discuss how we develop our temporal consensus formulation based on this framework.

IV. TEMPORAL CONSENSUS OPTIMIZATION FOR CONTINUAL NEURAL MAPPING

Now, we explain the main idea behind TACO. We propose to view consensus as enforcing agreement between the current model and its historical snapshots (model parameters from the past). From this perspective, continual neural mapping can be viewed as optimizing a current neural implicit map from streaming observations while retaining a set of frozen historical snapshots that encode past knowledge.

A. Problem Formulation

Let Θ_t denote the neural implicit map parameters obtained by optimizing on the incoming observations R_t at time step t . We consider an online continual mapping setting in which observations arrive sequentially, and the model is updated incrementally using only the current data stream, without storing or revisiting past RGB-D frames.

This temporal setting gives rise to two competing requirements that a continual neural mapping system must balance:

- 1) *Adaptation to scene changes*: Historical models may encode outdated or incorrect geometry when the environment evolves, and indiscriminately enforcing consistency with such information can impede the incorporation of new observations.
- 2) *Retention of valid historical knowledge*: At the same time, historical models often contain accurate information about regions of the environment that remain unchanged. An effective continual learning strategy should preserve such reliable knowledge to avoid unnecessary relearning.

To address these competing requirements, we formulate continual neural mapping as a temporal consensus optimization problem. We force the current model toward agreement with a set of frozen historical snapshots $\{\Theta_{t-k}\}$, where Θ_{t-k} denotes parameters obtained at earlier time steps. We also introduce importance-aware weighting to modulate the strength of consensus to preserve reliable historical knowledge while enabling adaptation to newly observed regions.

B. Importance-Weighted Temporal Consensus

For simplicity, we consider only the current model Θ_t and its immediate predecessor Θ_{t-1} . We aim to update the neural map to incorporate the new observation R_t while ensuring that it does not deviate significantly from the previous state Θ_{t-1} . To enforce this, we define a consensus target $z_{t,t-1}(\Theta_t)$

$$\begin{aligned} z_{t,t-1}(\Theta_t) &= (W_t(\Theta_t) + W_{t-1}(\Theta_t))^{-1} \\ &\quad (W_t(\Theta_t)\Theta_t + W_{t-1}(\Theta_t)\Theta_{t-1}), \end{aligned} \quad (9)$$

where $W_t(\cdot)$ and $W_{t-1}(\cdot)$ are square weight matrices encoding parameter-wise importance of Θ_t and Θ_{t-1} . We note that W_t, W_{t-1} and $z_{t,t-1}$ are functions of Θ_t . We depend W_{t-1} on Θ_t due to the normalization process. We compute W_t, W_{t-1} using the gradient information of Θ_t . We will discuss the computations of W_t, W_{t-1} in detail in the next section. Constraint (9) computes a parameter-wise weighted average of the current and historical models, where parameters with higher weights exert a stronger influence on the consensus target. As a result, important historical parameters are preserved, while parameters associated with lower weights are allowed to adapt to new observations.

In this setting, our optimization objective is:

$$\min_{\Theta_t} L^{obj}(\Theta_t, R_t), \text{ s.t. } \Theta_t = z_{t,t-1}(\Theta_t). \quad (10)$$

Here, $L^{obj}(\Theta_t, R_t)$ denotes the reconstruction loss with the new observations. Following the process described in the preliminaries section, we next obtain the augmented Lagrangian of (10) as

$$\mathcal{L}^a = L^{obj}(\Theta_t, R_t) + (\Theta_t - z_{t,t-1}(\Theta_t))^\top p + \frac{\rho}{2} \|\Theta_t - z_{t,t-1}(\Theta_t)\|_2^2, \quad (11)$$

where ρ , a user-picked hyperparameter, controls the strength of temporal consensus, and p is the dual variable.

We now solve this optimization problem (11) iteratively by applying the method of multipliers. We update the primal and dual variables for K iterations. As mentioned in (9), W_t, W_{t-1} and $z_{t,t-1}$ are functions of Θ_t . Thus, we also need to update them when the value of Θ_t is changing at each iteration. After the dual update of the previous iteration $k-1$, we utilize the information of Θ_t^{k-1} to compute W_t^{k-1} and W_{t-1}^{k-1} . Then, with W_t^{k-1}, W_{t-1}^{k-1} and Θ_t^{k-1} , we can get $z_{t,t-1}^{k-1}$:

$$z_{t,t-1}^{k-1}(\Theta_t^{k-1}) = (W_t^{k-1}(\Theta_t^{k-1}) + W_{t-1}^{k-1}(\Theta_t^{k-1}))^{-1} (W_t^{k-1}(\Theta_t^{k-1})\Theta_t^{k-1} + W_{t-1}^{k-1}(\Theta_t^{k-1})\Theta_{t-1}), \quad (12)$$

At the current iteration k , the state of the neural map, Θ_t^k , is obtained by the primal update:

$$\Theta_t^k = \arg \min_{\Theta_t} L^{obj}(\Theta_t, R_t) + (\Theta_t - z_{t,t-1}^{k-1}(\Theta_t^{k-1}))^\top p^{k-1} + \frac{\rho}{2} \|\Theta_t - z_{t,t-1}^{k-1}(\Theta_t^{k-1})\|_2^2. \quad (13)$$

We then update the dual variable:

$$p^k = p^{k-1} + \rho(\Theta_t^k - z_{t,t-1}^{k-1}(\Theta_t^{k-1})). \quad (14)$$

This dual update penalizes deviations from the importance-weighted temporal consensus target while allowing reliable historical parameters to exert stronger influence on the optimization.

After the method of multipliers iterative updates, we set the next state of the neural map to be $\Theta_t = \Theta_t^K$, where K is the last iteration. In the next section, we will discuss the importance weight matrix in detail.

C. Importance Estimation for Temporal Consensus

To construct importance-aware weights for temporal consensus optimization, we estimate a parameter-wise importance that controls how strongly each parameter is constrained over time. In neural implicit mapping, parameters contribute unequally to representing past scene geometry, and uniform temporal constraints can either impede adaptation or preserve outdated structure. Our importance measure captures each parameter's contribution to the learned geometry, protecting reliable regions while allowing flexibility where the scene has changed.

To measure how strongly model predictions depend on individual parameters, we first define a proxy objective on the rendered outputs:

$$\mathcal{L}_s^{\text{proxy}} = \frac{1}{|\mathcal{S}_s|} \sum_{r \in \mathcal{S}_s} (\|\hat{c}_{s,r}\|_2^2 + \|\hat{d}_{s,r}\|_2^2), \quad (15)$$

where $\mathcal{S}_s$ denotes the set of rays or pixels sampled at method of multiplier iteration s , and $\hat{c}_{s,r}$ and $\hat{d}_{s,r}$ are the predicted color and depth defined in Eqs. (3a) and (3b). This proxy objective captures the overall magnitude of the model's predictions and serves as a task-agnostic signal for probing output sensitivity without requiring ground-truth supervision.

We then estimate parameter importance by accumulating the sensitivity of this proxy objective with respect to the model parameters over iterations:

$$u_t^k(\Theta_t^k) = \sum_{s=1}^k \left| \frac{\partial \mathcal{L}_s^{\text{proxy}}}{\partial \Theta_t^s} \right| = \left| \frac{\partial \mathcal{L}_k^{\text{proxy}}}{\partial \Theta_t^k} \right| + u_t^{k-1}. \quad (16)$$

Here, s indexes iterations (up to current iteration k), and Θ_t^s denotes the optimization variable at the iteration s . We compute $\nabla_{\Theta_t^s} \mathcal{L}_s^{\text{proxy}}$ via backpropagation and accumulate its element-wise absolute value.

Intuitively, u_t^k measures how sensitive the model's predictions are to perturbations of each parameter across past iterations. Parameters with larger u_t^k have consistently exerted stronger influence on the rendered outputs and are therefore treated as more important, while parameters with smaller values are allowed to adapt more freely.

Since accumulated importance can grow unboundedly over time, it cannot be directly used as a consensus weight in our optimization. Moreover, independently normalizing the importance vectors of the current and historical models (e.g., to $[0, 1]$) destroys the relative magnitude information across parameters.

To preserve relative importance while maintaining numerical stability, we adopt a relative linear scaling strategy to construct temporal consensus weights. Using (16), we obtain the importance vector u_t^k . We also have u_{t-1} saved alongside the frozen historical snapshot Θ_{t-1} . We then compute their joint importance $u_{\text{sum}}^k := u_t^k + u_{t-1}$. Rather than imposing a fixed range, we normalize the overall scale of the weights by aligning their average magnitude with the penalty parameter ρ . Specifically, we define a global scaling factor

$$\epsilon^k(\Theta_t^k) = \frac{\rho}{\text{avg}(u_{\text{sum}}^k)}. \quad (17)$$

By normalizing with the average over all parameters of u_{sum}^k , we preserve the relative importance structure across parameters while preventing the absolute weight magnitudes from becoming arbitrarily large or small.

Then, we can construct our weight matrices as the following diagonal matrices

$$W_t^k(\Theta_t^k) = \text{diag}(\epsilon^k \cdot u_t^k), \quad W_{t-1}^k(\Theta_t^k) = \text{diag}(\epsilon^k \cdot u_{t-1}). \quad (18)$$

where $\text{diag}(\cdot)$ denotes the diagonal matrix formed by placing the entries of a vector on the main diagonal.

Note that during early mapping stages, some parameters may exhibit non-zero importance values due to noise, initialization effects, or indirect parameter coupling, even though they do not yet encode stable scene structure. Applying temporal consensus uniformly to such parameters can introduce

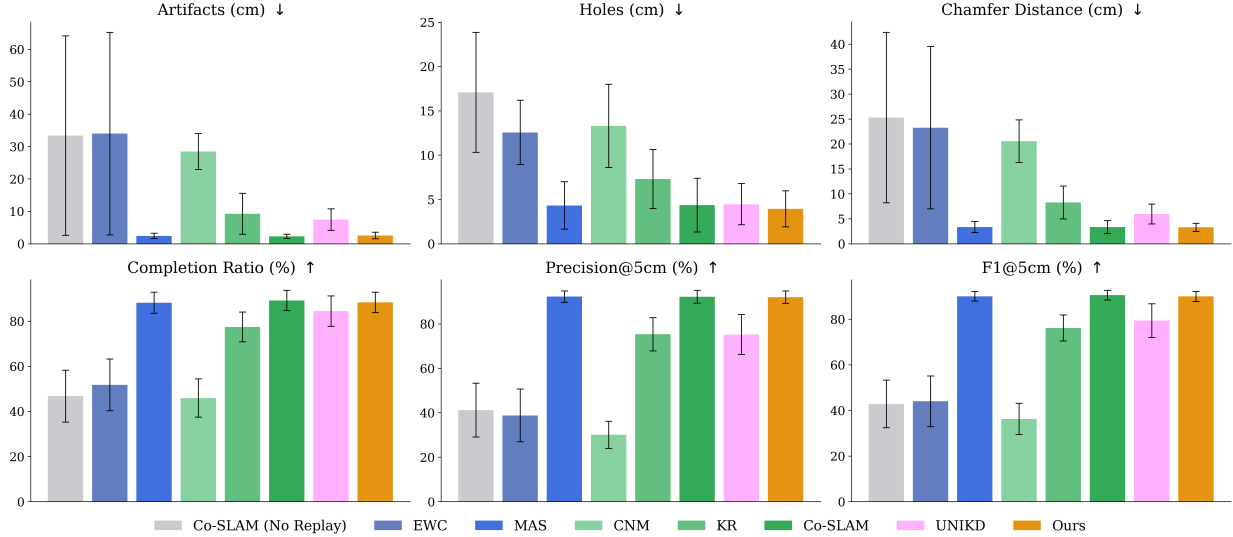


Fig. 3: Quantitative results on eight Replica scenes under static settings. We report reconstruction accuracy and completeness metrics, including Artifacts, Holes, Chamfer Distance, Completion Ratio, Precision@5cm, and F1@5cm. TACO achieves performance comparable to Co-SLAM with replay and matches or outperforms all continual learning baselines across all metrics.

unnecessary constraints in the early stages of the training. To make the importance-weighted consensus more robust in these cases, we incorporate a simple masking mechanism on the temporal consensus weights. Specifically, given a threshold $\beta > 0$, we suppress temporal constraints on parameters whose associated weights fall below this threshold:

$$W_{t-1}^{(i,i)} = \begin{cases} W_{t-1}^{(i,i)}, & W_{t-1}^{(i,i)} \geq \beta, \\ 0, & W_{t-1}^{(i,i)} < \beta. \end{cases} \quad (19)$$

This masking operation effectively disables temporal consensus for parameters with extremely low importance, allowing unconstrained adaptation in poorly observed regions, while preserving strong temporal constraints where sufficient evidence has been accumulated.

V. EXPERIMENTS AND RESULTS

We evaluate **TACO** in both simulation and real-world settings considering static and dynamic scenes. Evaluating both regimes serves two complementary goals: static scenes allow us to measure how well we preserve previously learned geometry over time without forgetting; dynamic scenes introduce controlled scene changes to evaluate whether we can adapt to new observations while retaining reliable historical structure.

We conduct static-scene experiments on two standard RGB-D benchmarks: **Replica** [41] which provides high-fidelity synthetic indoor scenes with accurate ground-truth geometry, and **ScanNet** [8] which captures large-scale real-world indoor environments with sensor noise and partial observations.

To study continual adaptation under dynamic conditions, we construct dynamic benchmarks in both simulation and the real world. In simulation, we base our dynamic experiments on

the **Habitat Synthetic Scenes Dataset** [17] and use **Habitat-Sim** [30, 36, 44] to introduce controlled object-level scene changes. We further collect a **real-world dynamic dataset** using a turtlebot. We deliberately modify the scene to validate that TACO remains effective when the environment changes.

We compare TACO against a representative set of continual learning baselines, all implemented within the Co-SLAM framework for a controlled comparison:

- **Parameter regularization methods:** **EWC** [18] and **MAS** [3], which mitigate forgetting by penalizing changes to parameters deemed important for past observations.
- **Replay-based methods:** **CNM** [51] and **KR** [19], which preserve historical information via explicit data replay. For KR, we use a replay buffer of 10 keyframes [42].
- **Distillation-based method:** **UNIKD** [15], which leverages uncertainty-filtered knowledge distillation to reduce forgetting during incremental updates.

We also report results for **Co-SLAM with replay** and **Co-SLAM without replay** as reference points. Note that Co-SLAM with replay represents an upper bound as it optimizes over both past and current observations via a replay buffer, while Co-SLAM without replay serves as a lower bound, updating only on recent observations and exposing the severity of catastrophic forgetting.

All our continual learning baselines operate in a strict streaming setting: observations are processed sequentially, past data are not revisited unless explicitly allowed by the method (e.g., replay-based baselines).

To quantitatively evaluate reconstruction accuracy, we adopt a set of standard 3D geometric metrics widely used in prior neural implicit mapping works [15, 42, 57, 59]. These metrics

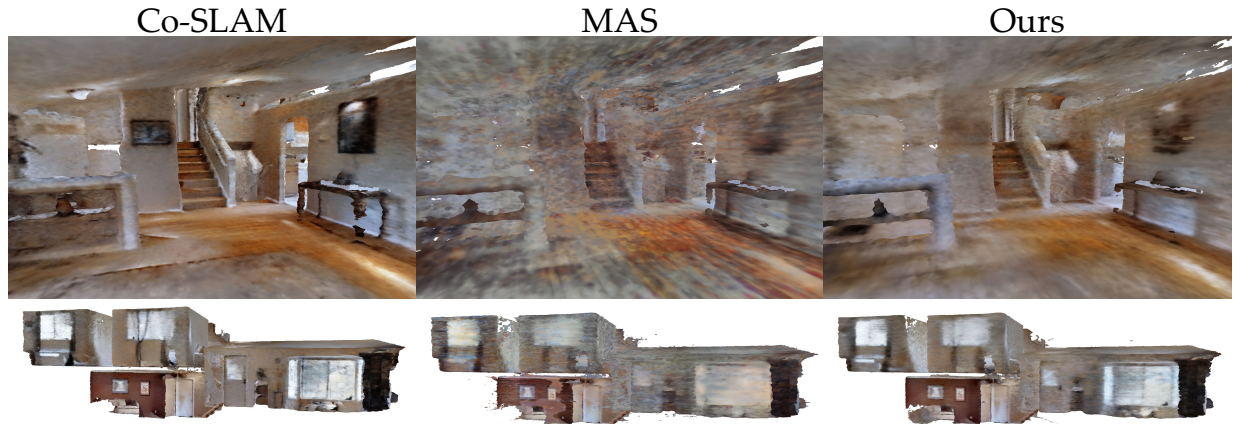


Fig. 4: Qualitative reconstruction results on a representative Gibson scene. TACO achieves geometry quality comparable to the replay-based baseline while producing less textural artifacts than MAS.

jointly assess surface accuracy, completeness, and artifact suppression. We have provided their detailed definitions and evaluation protocols in Appendix A.

A. Results on Static Simulated Scenes

We evaluate all our methods on static scenes to assess their ability to mitigate catastrophic forgetting when the environment remains unchanged. As shown in Fig. 3, TACO achieves top-tier performance on the Replica benchmarks, closely matching the upper bound given by Co-SLAM with replay and performing on par with or better than the strongest continual learning baseline, MAS. More details, including results on ScanNet, can be found in the Appendix B. These results indicate that importance-weighted temporal consensus effectively preserves global geometry over time without revisiting past observations.

To further test scalability under static conditions, we evaluate the top performing methods, TACO, MAS, and Co-SLAM, on a larger and more structurally complex environment from the Gibson dataset [49]. Compared to Replica and ScanNet, Gibson scenes cover substantially larger spatial extents and exhibit increased geometric complexity. As shown in Fig. 4, TACO produces more complete and smoother geometric reconstructions than MAS, with fewer discontinuities and less mesh fragmentation. Although both TACO and MAS exhibit color inconsistency relative to replay-based Co-SLAM due to the absence of explicit appearance modeling, TACO maintains coherent geometry without relying on replay. These results demonstrate that temporal consensus scales effectively to larger static environments.

For the subsequent evaluations on dynamic scenes, we only compare TACO to Co-SLAM with replay and MAS, which are the strongest baseline in the static settings.

B. Results on Dynamic Simulated Scenes

We further evaluate continual neural mapping under dynamic environment changes using simulated scenes with staged object motion. Unlike static reconstruction settings, dynamic scenes require the model to adapt to newly observed

geometry while preserving previously reliable map content, making them a direct test of continual adaptation.

To construct this evaluation, we build dynamic datasets based on the Habitat Synthetic Scenes Dataset. Using Habitat-Sim, we generate simulated environments in which scene geometry is modified in a staged manner, allowing us to isolate and analyze the behavior of different continual mapping strategies under conflicting observations. Specifically, the dynamic sequence consists of two stages, as illustrated in Fig. 5. In the first stage, a beanbag sofa and a potted tree are relocated to new positions, while the remaining scene structure remains unchanged. In the second stage, the table is moved to a different location, and the tableware originally stored on the built-in lower shelf is disturbed and scattered onto the carpet, which introduces additional fine-grained geometric changes. Throughout both stages, large portions of the scene remain static.

Fig. 5 illustrates qualitative reconstruction results across stages. TACO accurately captures the staged scene modifications while maintaining consistent reconstruction of unchanged regions. After each change, outdated geometry is correctly revised, and the updated map reflects the current object configuration without introducing artifacts in previously stable areas. This demonstrates that temporal consensus enables selective adaptation rather than uniform preservation of past states. In contrast, MAS, as a regularization-based method, enforces strong parameter constraints to preserve prior knowledge. When confronted with conflicting observations caused by object motion, this over-constraining behavior prevents proper geometric revision. As a result, mesh tearing and fragmented surfaces emerge near both the original and updated object locations. Replay-based methods such as Co-SLAM exhibit a different failure mode. By explicitly replaying past observations, Co-SLAM enforces consistency with both historical and current geometry. When objects are moved, this leads to simultaneous reconstruction at both the old and new positions, producing persistent ghost artifacts as a result. While replay mitigates catastrophic forgetting in static settings, it

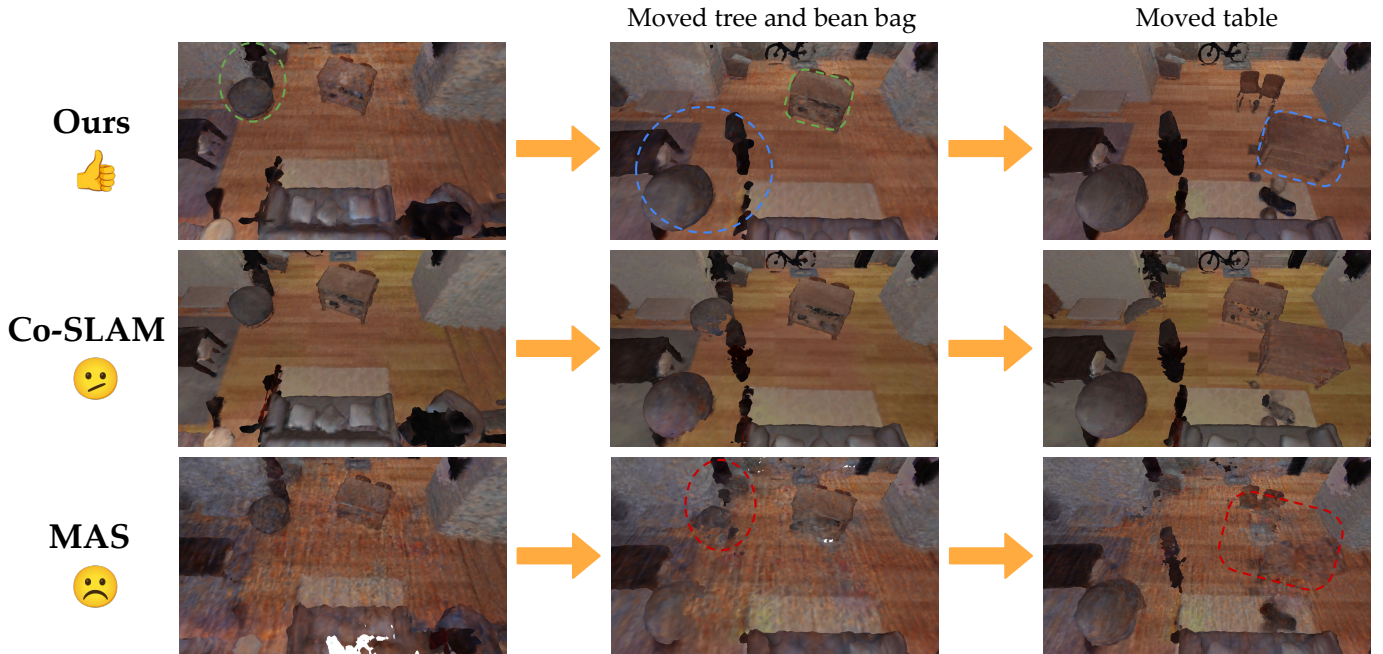


Fig. 5: **Dynamic scene reconstruction under staged object motion.** TACO adapts to scene changes without introducing artifacts in stable regions, whereas replay- and regularization-based methods produce ghosting and mesh tearing, respectively. Dashed boxes highlight objects affected by staged motion and the resulting reconstruction behavior: **green boxes** mark objects that will move in the next stage, **blue boxes** mark objects whose positions have changed in the current stage, and **red boxes** highlight failure regions with severe artifacts.

becomes detrimental under environment changes where past observations are no longer valid.

Table I reports quantitative results on the dynamic simulated scenes. TACO consistently outperforms MAS across all metrics. It produces fewer reconstruction artifacts and achieves higher completion ratio and F1@5cm. These results demonstrate that the proposed temporal consensus mechanism is effective not only for preserving past geometry but also for adapting to dynamic scene changes.

TABLE I: Quantitative results on dynamic simulated scenes. TACO outperforms the regularization-based MAS baseline, producing fewer reconstruction artifacts and achieving higher completion ratio and F1 score.

Method	Artifacts ↓	Completion Ratio ↑	F1@5cm ↑
MAS	8.53	71.41	73.43
TACO	6.38	74.37	78.05

Overall, these qualitative and quantitative results highlight the limitations of replay- and regularization-based continual mapping under dynamic scenes. By treating past representations as temporally weighted constraints rather than immutable supervision, TACO achieves robust adaptation to scene changes while preserving consistent geometry.

C. Continual Mapping in Dynamic Real-World Environments

We further evaluate continual neural mapping performance in a real-world setting with controlled scene changes.

We first consider a simple scenario where we move a single object—the yellow stool highlighted in Fig. 1—from position A at time t_0 to position B at time t_1 while keeping the rest of the scene unchanged. As shown in Fig. 1, different methods exhibit distinct behaviors after the relocation. Replay-based methods such as Co-SLAM reconstruct the stool at both its original and new locations. It produces persistent artifacts because replay enforces consistency with historical observations even when they no longer reflect the current scene. Regularization-based methods such as MAS strongly penalize parameter updates associated with past observations, which over-constrains optimization and leads to distorted or fragmented geometry in regions affected by the conflict. In contrast, our method selectively relaxes constraints on parameters linked to outdated geometry while preserving consistency in unchanged regions. As a result, we reconstruct the stool only at its current location, without ghost artifacts or geometric breakage.

We further test continual mapping under more challenging real-world dynamics by introducing multiple sequential scene changes. As illustrated in Fig. 6, this experiment proceeds through four stages. In the first stage, we move a row of cubes to new locations. In the second stage, we reshuffle a larger set of cubes into a different configuration. In the third stage, we move two cubes back to their original positions and relocate the yellow stool to the positions previously occupied by those cubes. Across these stages, the environment undergoes repeated and heterogeneous changes that induce accumulating

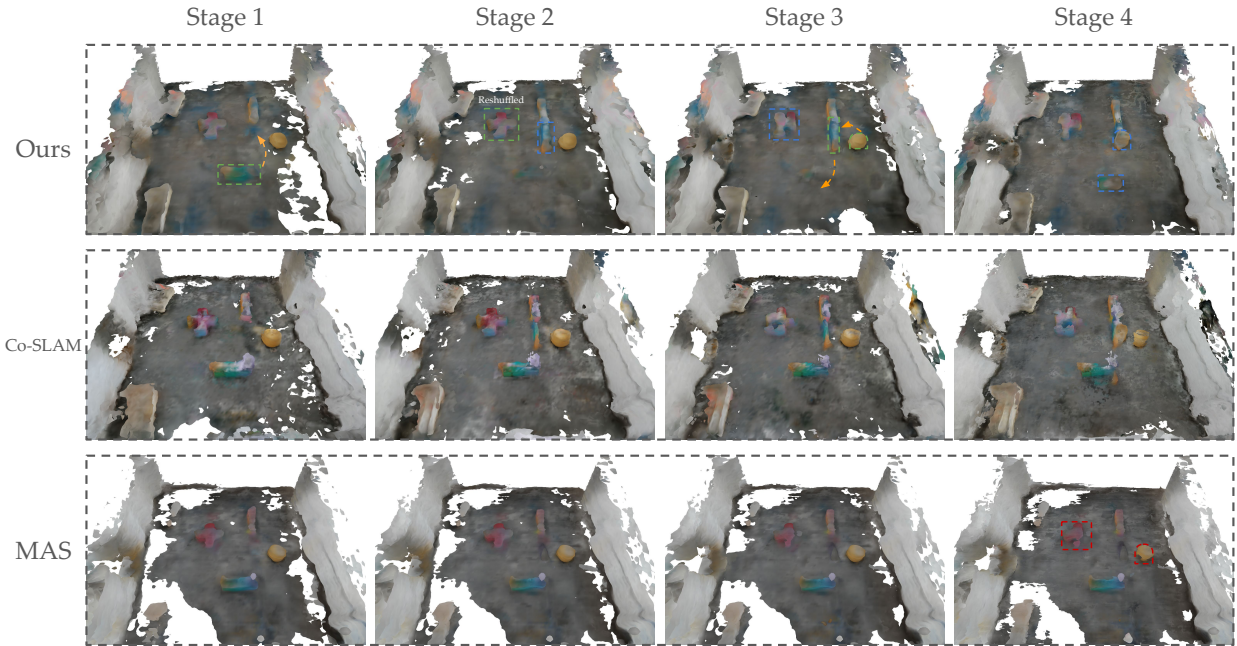


Fig. 6: Additional hardware experiment with multiple sequential scene changes. TACO consistently adapts to changing environments, outperforming the baseline methods. Dashed boxes follow the same annotation scheme as in Fig. 5: **green boxes** mark objects that will move in the next stage, **blue boxes** mark objects whose positions have changed in the current stage, and **red boxes** highlight failure regions with severe artifacts.

conflicts between historical and current observations.

Across all stages, we adapt consistently to the evolving scene by revising outdated geometry while preserving stable regions and producing coherent reconstructions throughout the sequence. In contrast, Co-SLAM accumulates severe ghost artifacts as changes progress, and MAS fails to adapt meaningfully under repeated changes as it produces torn and fragmented geometry in regions affected by object motion.

Overall, these real-world experiments demonstrate that we robustly balance stability and adaptivity across real-world dynamics, and validate temporal consensus as an effective mechanism for replay-free continual neural mapping in dynamic environments.

D. Analysis of Memory Usage

We analyze the CPU memory footprint of representative methods as CPU memory is a critical constraint for real-world robotic deployment. We report *resident set size* (RSS), which measures the amount of physical CPU memory actively occupied by a process during execution, excluding disk-backed virtual memory. Unlike GPU memory, CPU RSS directly reflects the on-board memory pressure faced by robotic systems and thus provides a practical indicator of deployability.

As shown in Table II, full replay (Co-SLAM) incurs substantial memory overhead due to storing historical observations. In contrast, TACO achieves a CPU memory footprint comparable to replay-based, distillation-based, and regularization-based baselines, while reducing memory usage by more than 50% relative to full replay.

TABLE II: CPU memory footprint (RSS) comparison in MB.

Method	CPU RSS (MB) ↓
Co-SLAM	4165.1
KR (Replay-based)	1912.4
UNIKD (Knowledge Distillation)	1945.1
Online MAS (Regularization)	1800.6
Ours	1864.5

VI. CONCLUSION

We introduced TACO, a replay-free framework for continual neural implicit mapping that formulates online mapping as an importance-aware temporal consensus optimization problem. By treating historical models as temporally weighted constraints rather than immutable supervision, TACO enables selective revision of outdated geometry while preserving consistent structure in unchanged regions. Our experiments in both simulated and real-world environments demonstrate that TACO robustly adapts to scene changes while avoiding ghost artifacts and over-constrained updates common to replay- and regularization-based methods. This work highlights temporal consensus as a principled mechanism for scalable long-term neural mapping in dynamic robotic environments.

ACKNOWLEDGMENTS

This work was supported in part by NSF grants ECCS-2438314 (CAREER Award) and CNS-2529645. This work was also supported by the UC Berkeley Center for Digital Assets (CDA). This research was made possible by GPU resources provided via the NVIDIA Academic Grant.

REFERENCES

- [1] RR Ade and PR Deshmukh. Methods for incremental learning: a survey. *International Journal of Data Mining & Knowledge Management Process*, 3(4):119, 2013.
- [2] Rahaf Aljundi, Punarjay Chakravarty, and Tinne Tuytelaars. Expert gate: Lifelong learning with a network of experts. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3366–3375, 2017.
- [3] Rahaf Aljundi, Francesca Babiloni, Mohamed Elhoseiny, Marcus Rohrbach, and Tinne Tuytelaars. Memory aware synapses: Learning what (not) to forget. In *Proceedings of the European conference on computer vision (ECCV)*, pages 139–154, 2018.
- [4] Mahboubeh Asadi, Kourosh Zareinia, and Sajad Saeedi. Di-nerf: Distributed nerf for collaborative learning with relative pose refinement. *IEEE Robotics and Automation Letters*, 2024.
- [5] Dejan Azinović, Ricardo Martin-Brualla, Dan B Goldman, Matthias Nießner, and Justus Thies. Neural rgb-d surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6290–6301, June 2022.
- [6] Zhipeng Cai and Matthias Müller. Clnrf: Continual learning meets nerf. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 23185–23194, 2023.
- [7] Arslan Chaudhry, Marc’Aurelio Ranzato, Marcus Rohrbach, and Mohamed Elhoseiny. Efficient lifelong learning with a-gem. *arXiv preprint arXiv:1812.00420*, 2018.
- [8] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2017.
- [9] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3366–3385, 2021.
- [10] Yinan Deng, Yujie Tang, Yi Yang, Danwei Wang, and Yufeng Yue. Macim: Multi-agent collaborative implicit mapping. *IEEE Robotics and Automation Letters*, 9(5): 4369–4376, 2024.
- [11] Jianyu Dou, Yinan Deng, Jiahui Wang, Xingsi Tang, Yi Yang, and Yufeng Yue. Openmulti: Open-vocabulary instance-level multi-agent distributed implicit mapping. *IEEE Robotics and Automation Letters*, 2025.
- [12] Arthur Douillard, Matthieu Cord, Charles Ollion, Thomas Robert, and Eduardo Valle. Podnet: Pooled outputs distillation for small-tasks incremental learning. In *European Conference on Computer Vision*, pages 86–102. Springer, 2020.
- [13] Chrisantha Fernando, Dylan Banarse, Charles Blundell, Yori Zwols, David Ha, Andrei A Rusu, Alexander Pritzel, and Daan Wierstra. Pathnet: Evolution channels gradient descent in super neural networks. *arXiv preprint arXiv:1701.08734*, 2017.
- [14] Lily Goli, Cody Reading, Silvia Sellán, Alec Jacobson, and Andrea Tagliasacchi. Bayes’ rays: Uncertainty quantification for neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20061–20070, 2024.
- [15] Mengqi Guo, Chen Li, Hanlin Chen, and Gim Hee Lee. Unikd: Uncertainty-filtered incremental knowledge distillation for neural implicit representation. In *European Conference on Computer Vision*, pages 237–254. Springer, 2024.
- [16] Minsoo Kang, Jaeyoo Park, and Bohyung Han. Class-incremental learning by knowledge distillation with adaptive feature consolidation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16071–16080, 2022.
- [17] Mukul Khanna, Yongsan Mao, Hanxiao Jiang, Sanjay Haresh, Brennan Shacklett, Dhruv Batra, Alexander Clegg, Eric Undersander, Angel X Chang, and Manolis Savva. Habitat synthetic scenes dataset (hssd-200): An analysis of 3d scene scale and realism tradeoffs for objectgoal navigation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16384–16393, 2024.
- [18] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.
- [19] Georg Klein and David Murray. Parallel tracking and mapping for small ar workspaces. In *2007 6th IEEE and ACM international symposium on mixed and augmented reality*, pages 225–234. IEEE, 2007.
- [20] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. Barf: Bundle-adjusting neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5741–5751, 2021.
- [21] David Lopez-Paz and Marc’Aurelio Ranzato. Gradient episodic memory for continual learning. *Advances in neural information processing systems*, 30, 2017.
- [22] Arun Mallya and Svetlana Lazebnik. Packnet: Adding multiple tasks to a single network by iterative pruning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 7765–7773, 2018.
- [23] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.
- [24] Thomas Müller, Brian McWilliams, Fabrice Rousselle, Markus Gross, and Jan Novák. Neural importance

- sampling. *ACM Trans. Graph.*, 38(5), October 2019. ISSN 0730-0301. doi: 10.1145/3341156. URL <https://doi.org/10.1145/3341156>.
- [25] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4), July 2022. ISSN 0730-0301. doi: 10.1145/3528223.3530127. URL <https://doi.org/10.1145/3528223.3530127>.
- [26] Angelia Nedic and Asuman Ozdaglar. Distributed sub-gradient methods for multi-agent optimization. *IEEE Transactions on automatic control*, 54(1):48–61, 2009.
- [27] Richard A. Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J. Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. Kinectfusion: Real-time dense surface mapping and tracking. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pages 127–136, 2011. doi: 10.1109/ISMAR.2011.6092378.
- [28] Matthias Nießner, Michael Zollhöfer, Shahram Izadi, and Marc Stamminger. Real-time 3d reconstruction at scale using voxel hashing. *ACM Transactions on Graphics (ToG)*, 32(6):1–11, 2013.
- [29] German I Parisi, Ronald Kemker, Jose L Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural networks*, 113:54–71, 2019.
- [30] Xavi Puig, Eric Undersander, Andrew Szot, Mikael Dal-laire Cote, Ruslan Partsey, Jimmy Yang, Ruta Desai, Alexander William Clegg, Michal Hlavac, Tiffany Min, Theo Gervet, Vladimír Vondruš, Vincent-Pierre Berges, John Turner, Oleksandr Maksymets, Zsolt Kira, Mrinal Kalakrishnan, Jitendra Malik, Devendra Singh Chaplot, Unnat Jain, Dhruv Batra, Akshara Rai, and Roozbeh Mottaghi. Habitat 3.0: A co-habitat for humans, avatars and robots, 2023.
- [31] Amal Rannen, Rahaf Aljundi, Matthew B Blaschko, and Tinne Tuytelaars. Encoder based lifelong learning. In *Proceedings of the IEEE international conference on computer vision*, pages 1320–1328, 2017.
- [32] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2001–2010, 2017.
- [33] Anthony Robins. Catastrophic forgetting, rehearsal and pseudorehearsal. *Connection Science*, 7(2):123–146, 1995.
- [34] David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy Lillicrap, and Gregory Wayne. Experience replay for continual learning. *Advances in neural information processing systems*, 32, 2019.
- [35] Kanghyun Ryu, Qiayuan Liao, Zhongyu Li, Payam Del-gosha, Koushil Sreenath, and Negar Mehr. Curricullm: Automatic task curricula design for learning complex robot skills using large language models. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4470–4477. IEEE, 2025.
- [36] Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, Devi Parikh, and Dhruv Batra. Habitat: A Platform for Embodied AI Research. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.
- [37] Jianxiong Shen, Adria Ruiz, Antonio Agudo, and Francesc Moreno-Noguer. Stochastic neural radiance fields: Quantifying uncertainty in implicit 3d representations. In *2021 International Conference on 3D Vision (3DV)*, pages 972–981. IEEE, 2021.
- [38] Wei Shi, Qing Ling, Kun Yuan, Gang Wu, and Wotao Yin. On the linear convergence of the admm in decentralized consensus optimization. *IEEE Transactions on Signal Processing*, 62(7):1750–1761, 2014. doi: 10.1109/TSP.2014.2304432.
- [39] Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. Continual learning with deep generative replay. *Advances in neural information processing systems*, 30, 2017.
- [40] Boyd Stephen, Parikh Neal, Chu Eric, Peleato Borja, and Eckstein Jonathan. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Information Retrieval*, 3(1):1–122, 07 2011. ISSN 1554-0669. doi: 10.1561/22000000016. URL <https://doi.org/10.1561/22000000016>.
- [41] Julian Straub, Thomas Whelan, Lingni Ma, Yufan Chen, Erik Wijmans, Simon Green, Jakob J. Engel, Raul Mur-Artal, Carl Ren, Shobhit Verma, Anton Clarkson, Mingfei Yan, Brian Budge, Yajie Yan, Xiaqing Pan, June Yon, Yuyang Zou, Kimberly Leon, Nigel Carter, Jesus Briales, Tyler Gillingham, Elias Mueggler, Luis Pesqueira, Manolis Savva, Dhruv Batra, Hauke M. Strasdat, Renzo De Nardi, Michael Goesele, Steven Lovegrove, and Richard Newcombe. The Replica dataset: A digital replica of indoor spaces. *arXiv preprint arXiv:1906.05797*, 2019.
- [42] Edgar Sucar, Shikun Liu, Joseph Ortiz, and Andrew J Davison. imap: Implicit mapping and positioning in real-time. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6229–6238, 2021.
- [43] Niko Sünderhauf, Jad Abou-Chakra, and Dimity Miller. Density-aware nerf ensembles: Quantifying predictive uncertainty in neural radiance fields. *arXiv preprint arXiv:2209.08718*, 2022.
- [44] Andrew Szot, Alex Clegg, Eric Undersander, Erik Wijmans, Yili Zhao, John Turner, Noah Maestre, Mustafa Mukadam, Devendra Chaplot, Oleksandr Maksymets, Aaron Gokaslan, Vladimir Vondrus, Sameer Dharur, Franziska Meier, Wojciech Galuba, Angel Chang, Zsolt Kira, Vladlen Koltun, Jitendra Malik, Manolis Savva, and Dhruv Batra. Habitat 2.0: Training home assistants to rearrange their habitat. In *Advances in Neural Information*

Processing Systems (NeurIPS), 2021.

- [45] Hengyi Wang, Jingwen Wang, and Lourdes Agapito. Co-slam: Joint coordinate and sparse parametric encodings for neural real-time slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13293–13302, 2023.
- [46] Liyuan Wang, Xingxing Zhang, Hang Su, and Jun Zhu. A comprehensive survey of continual learning: Theory, method and application. *IEEE transactions on pattern analysis and machine intelligence*, 46(8):5362–5383, 2024.
- [47] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021.
- [48] Thomas Whelan, Stefan Leutenegger, Renato F Salas-Moreno, Ben Glocker, and Andrew J Davison. Elastic-fusion: Dense slam without a pose graph. In *Robotics: science and systems*, volume 11. Rome, 2015.
- [49] Fei Xia, Amir R. Zamir, Zhi-Yang He, Alexander Sax, Jitendra Malik, and Silvio Savarese. Gibson Env: real-world perception for embodied agents. In *Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on*. IEEE, 2018.
- [50] Ju Xu and Zhanxing Zhu. Reinforced continual learning. *Advances in neural information processing systems*, 31, 2018.
- [51] Zike Yan, Yuxin Tian, Xuesong Shi, Ping Guo, Peng Wang, and Hongbin Zha. Continual neural mapping: Learning an implicit scene representation from sequential observations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15782–15792, 2021.
- [52] Zike Yan, Haoxiang Yang, and Hongbin Zha. Active neural mapping. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10981–10992, October 2023.
- [53] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *Advances in neural information processing systems*, 34:4805–4815, 2021.
- [54] Javier Yu, Joseph A Vincent, and Mac Schwager. Dinno: Distributed neural network optimization for multi-robot collaborative learning. *IEEE Robotics and Automation Letters*, 7(2):1896–1903, 2022.
- [55] Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction. *Advances in neural information processing systems*, 35:25018–25032, 2022.
- [56] Hongrui Zhao, Boris Ivanovic, and Negar Mehr. Distributed nerf learning for collaborative multi-robot perception. *arXiv preprint arXiv:2409.20289*, 2024.
- [57] Hongrui Zhao, Boris Ivanovic, and Negar Mehr. Ramen: Real-time asynchronous multi-agent neural implicit mapping. *arXiv preprint arXiv:2502.19592*, 2025.
- [58] Hongrui Zhao, Xunlan Zhou, Boris Ivanovic, and Negar Mehr. Udon: Uncertainty-weighted distributed optimization for multi-robot neural implicit mapping under extreme communication constraints. *arXiv preprint arXiv:2509.12702*, 2025.
- [59] Zihan Zhu, Songyou Peng, Viktor Larsson, Weiwei Xu, Hujun Bao, Zhaopeng Cui, Martin R Oswald, and Marc Pollefeys. Nice-slam: Neural implicit scalable encoding for slam. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12786–12796, 2022.

APPENDIX

A. Quantitative Metrics for Geometry Evaluation

To assess the accuracy and completeness of reconstructed geometry, we employ a set of 3D quantitative metrics commonly used in prior neural implicit mapping works [15, 42, 57, 59], with minor renaming for clarity. Together, these metrics capture complementary aspects of geometric fidelity, surface completeness, and artifact suppression.

- **Artifacts (cm)** measure the average distance from reconstructed surface points to the ground-truth geometry, reflecting the presence of spurious or floating artifacts. Lower values indicate fewer artifacts.
- **Holes (cm)** compute the average distance from ground-truth surface points to the reconstructed mesh, quantifying missing or unreconstructed regions. Lower values indicate better coverage.
- **Completion Ratio (%)** reports the percentage of ground-truth points whose nearest reconstructed point lies within a 5 cm threshold. This metric serves as our primary measure of surface completeness.
- **Chamfer Distance (CD)** is defined as the symmetric average distance between reconstructed and ground-truth point sets, jointly reflecting geometric accuracy and surface completeness. Lower values are better.
- **Precision@5 cm** measures the fraction of reconstructed surface points whose nearest ground-truth point lies within a 5 cm threshold, penalizing spurious geometry and floating artifacts. Higher values are better.
- **F1@5 cm** is the harmonic mean of Precision@5 cm and Recall@5 cm at a 5 cm threshold, providing a balanced metric that captures both geometric fidelity and scene coverage.

B. Results on Static Simulated Scenes

We evaluate all methods on static scenes to establish a controlled setting for assessing reconstruction quality and catastrophic forgetting in the absence of scene changes. Fig. 3 reports results on eight Replica scenes across all metrics. Among parameter-regularization baselines, EWC collapses severely, exhibiting large artifacts and Chamfer errors and substantially reduced completion, precision, and F1 scores, indicating that naive stability constraints are insufficient for continual neural mapping. In contrast, MAS performs strongly and remains competitive with the best methods across most metrics. Replay-based methods (CNM, KR) partially recover performance—KR consistently outperforms CNM—but both fall short of the top tier in terms of accuracy and coverage. The distillation-based baseline UNIKD improves over replay and regularization failures and achieves solid reconstruction quality, yet remains below the best-performing methods overall.

Across all Replica scenes, we achieve top-tier performance on every metric, closely matching Co-SLAM with replay and exhibiting performance comparable to or better than MAS. In particular, TACO maintains low artifact and Chamfer

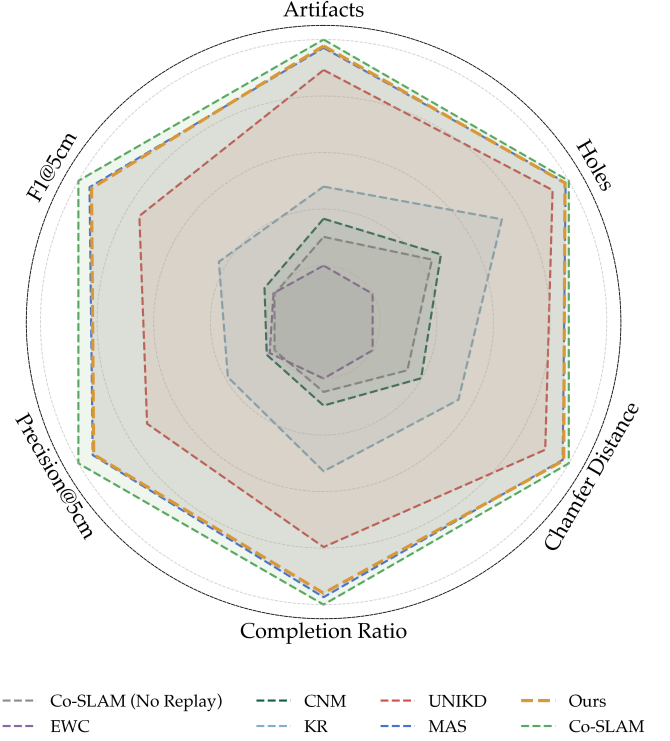


Fig. 7: Quantitative Results on ScanNet

distance while preserving high completion, precision@5cm, and F1@5cm, demonstrating that importance-weighted temporal consensus preserves global geometry without revisiting past observations. Compared to explicit replay (CNM, KR) and knowledge distillation (UNIKD), TACO achieves a more favorable accuracy-completeness trade-off, reducing spurious geometry without sacrificing surface coverage.

We further provide quantitative evaluation on two real-world ScanNet scenes (scene0000_00 and scene0673_05), summarized by the radar plots in Fig. 7. The overall performance trends closely mirror those observed on Replica. Parameter-regularization baselines degrade substantially across all metrics, reflecting limited robustness to real-world sensor noise and partial observations. Replay-based and distillation-based methods recover partial performance but remain uneven across accuracy and completeness dimensions. In contrast, TACO consistently lies near the upper envelope of the radar plots, closely aligning with Co-SLAM and matching the strongest continual learning baseline across all metrics.

We next visualize reconstruction quality on ScanNet. As shown in Fig. 8, results on scene0000_00 reveal consistent qualitative behaviors under real-world sensing noise and incomplete coverage. Co-SLAM without replay fails to preserve coherent geometry and produces fragmented reconstructions, while UNIKD largely maintains scene structure but exhibits severe color inconsistency. Regularization-based methods such as MAS preserve the overall layout but show noticeable geometric breakage in several regions. In contrast, TACO produces stable and coherent reconstructions and more



Fig. 8: Visualization on ScanNet scene0000_00.

completely recovers the table region highlighted in the figure.

C. Ablation on Mask Mechanism

We conduct an ablation study to analyze the effect of the mask mechanism used in temporal consensus weighting. The mask is designed to suppress unreliable parameter updates during early stages of mapping, where non-informative gradients may arise due to noise, initialization effects, or indirect coupling within the network. Setting the mask to zero disables this mechanism, corresponding to importance-based temporal consensus without masking.

TABLE III: Ablation study on the mask mechanism for temporal consensus weighting on room1.

Method	Chamfer Distance (CD) ↓	F1@5 cm ↑
Ours ($\tau = 10^{-2}$)	2.87	89.96
Ours ($\tau = 10^{-3}$)	2.21	92.69
Ours ($\tau = 10^{-4}$)	2.05	93.65
Ours ($\tau = 10^{-5}$)	2.12	93.20
Ours ($\tau = 0$)	2.17	93.06

Table III reports results on room1 using Chamfer Distance (CD) and F1@5 cm. Introducing the mask consistently improves reconstruction quality compared to the no-mask baseline. In particular, moderate masking strengths (e.g., $1e-4$) achieve the best trade-off between geometric accuracy and surface completeness, yielding the lowest CD and highest F1 score. When the mask is too strong, performance degrades due to overly aggressive suppression of parameter updates, while disabling the mask entirely leads to slightly inferior results.

These results indicate that the proposed mask mechanism plays a critical role in stabilizing importance-aware temporal consensus, enabling robust continual adaptation while avoiding the influence of spurious early-stage gradients.

We further observe that the choice of the masking threshold τ is coupled with the penalty parameter ρ , which controls the overall strength of temporal consensus. Larger values of ρ amplify the influence of consensus terms, requiring a correspondingly higher threshold τ to avoid over-constraining early-stage or unreliable parameters, while smaller ρ values allow for weaker masking without destabilizing optimization.

D. Ablation on the Number of Temporal Snapshots

We conduct an ablation study to examine the effect of using multiple historical snapshots for temporal consensus optimization. Specifically, we vary the number of snapshots K jointly participating in temporal consensus while keeping all other components fixed.

TABLE IV: Ablation on the number of temporal snapshots K used for temporal consensus on room1.

K (number of snapshots)	Chamfer Distance (CD) ↓	F1@5 cm ↑
$K = 1$	2.50	93.65
$K = 2$	2.62	91.53
$K = 3$	2.85	90.29
$K = 4$	3.51	87.60

Table IV reports results on Replica room1. Using a single historical snapshot ($K = 1$) yields the best overall performance, achieving the lowest Chamfer Distance and highest F1 score. As K increases, performance degrades consistently across both metrics.

This behavior reflects the trade-off between stability and adaptability in temporal consensus. In dynamic mapping scenarios, older snapshots are more likely to encode outdated geometry. Including multiple historical models introduces conflicting constraints that bias optimization toward obsolete scene states, making it harder for the current model to fully adapt to recent changes. As a result, enforcing consensus with too many snapshots over-regularizes the model and degrades reconstruction quality.

These results suggest that temporal consensus should prioritize recency rather than aggregating long histories, and that a small number of carefully weighted snapshots is sufficient for stabilizing continual neural mapping without impeding adaptation.