Continual Multi-Robot Learning from Black-Box Visual Place Recognition Models

Kenta Tsukahara, Kanji Tanaka, Daiki Iwata, Jonathan Tay Yu Liang and Wuhao Xie
University of Fukui, 3-9-1 Bunkyo, Fukui City, Fukui 910-0017, Japan

Keywords: Visual Place Recognition, Continual Multi-Robot Learning, Black-Box Teachers.

Abstract: In the context of visual place recognition (VPR), continual learning (CL) techniques offer significant potential for avoiding catastrophic forgetting when learning new places. However, existing CL methods often focus on knowledge transfer from a known model to a new one, overlooking the existence of unknown black-box models. This study explores a novel multi-robot CL approach that enables knowledge transfer from black-box VPR models (teachers), such as those of local robots encountered by traveler robots (students) in unknown environments. Specifically, we introduce Membership Inference Attack (MIA), a privacy attack applicable to black-box models, and leverage it to reconstruct pseudo training sets, which serve as the transferable knowledge between robots. Furthermore, we address the low sampling efficiency of MIA by leveraging prior insights from the literature on place class prediction distributions and unseen-class detection. Finally, we analyze both the individual and combined effects of these techniques.

1 INTRODUCTION

Visual place recognition (VPR) enables autonomous robots and self-driving vehicles to classify their place from visual input (Weyand et al., 2016)(Morita et al., 2005)(Seo et al., 2018). While conventional VPR systems rely on supervised learning from direct visual experiences, they face two fundamental limitations: the high cost of collecting training data in new environments and catastrophic forgetting when learning new places. These challenges become particularly acute in long-term autonomous operations where robots must continuously adapt to environmental changes. Recent advancements in continual learning techniques (Lange et al., 2022)—such as regularization methods, experience replay, and dynamic network expansion—have yielded promising results in mitigating catastrophic forgetting and improving continual adaptation performance (Gao et al., 2022; Vödisch et al., 2023; Vödisch et al., 2022).

This study explores multi-robot Continual Learning (CL) for Visual Place Recognition (VPR), framing CL as knowledge transfer from an old model to a new one. We extend conventional knowledge transfer between different models of the same robot to knowledge transfer between different models of different robots. This approach offers key advantages, including scalability and fault tolerance. Specifically,

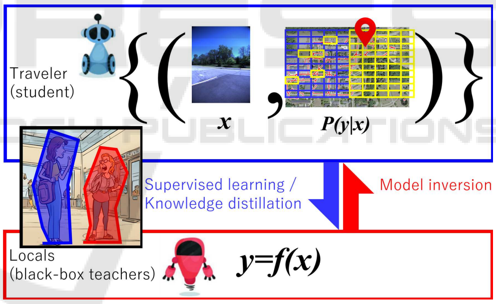


Figure 1: Illustration of human-inspired knowledge transfer (KT) via local-to-traveler interaction, motivating robot-to-robot KT using VPR models.

a traveling robot can update its knowledge by acquiring information from local robots, avoiding the cost and risk of collecting training data independently. Additionally, sharing VPR knowledge ensures critical information is preserved even if some robots fail. Despite these advantages, multi-robot CL introduces new challenges, such as variations in robot capabilities, communication overhead, and concerns related to privacy and security.

One of the key research challenges in multi-robot systems is the question: ”Can a pseudo-training set be reconstructed from a black-box teacher?” For example, when a traveling robot explores an unfamiliar city and encounters a local robot for the first time, it cannot access the internal parameters or training

Tsukahara, K., Tanaka, K., Iwata, D., Liang, J. T. Y. and Xie, W.
Continual Multi-Robot Learning from Black-Box Visual Place Recognition Models.
DOI: 10.5220/0013771900003982

In *Proceedings of the 22nd International Conference on Informatics in Control, Automation and Robotics (ICINCO 2025) - Volume 1*, pages 126-135
ISBN: 978-989-758-770-2; ISSN: 2184-2809

data of the local robot's VPR model, which thus acts as a black-box teacher. Generating a pseudo-training set from such a black box is essential for knowledge transfer, yet remains a difficult and unresolved problem. This scenario represents a typical case where black-box teachers naturally arise in multi-robot continual learning, and reconstructing pseudo-training sets is crucial for integrating new data and retraining models.

Many existing approaches rely on the assumption that the existing model is a white box, allowing access to its internal structure or training set. Indeed, in single-robot systems, this assumption often holds, as the existing model belongs to the robot itself. However, in multi-robot systems, especially in open-world scenarios involving heterogeneous or unknown robots, this assumption is overly optimistic. In fact, reconstructing a training set from a black-box model remains an ongoing challenge in machine learning and is largely an open problem.

In this work, we address the key challenge of reconstructing (pseudo) training samples from black-box teachers using membership inference attacks (MIAs) (Hu et al., 2022) (Fig. 1). Unlike traditional approaches that analyze the model to reconstruct the training data, MIA generates samples and predicts the likelihood that they were part of the model's training data. It is the only major privacy attack method effective with black-box models. However, this flexibility comes with a downside: low sampling efficiency, particularly when dealing with high-dimensional input data, such as images used in VPR. The naive approach of randomly generating training sample candidates is highly inefficient. To overcome this limitation, we propose improving the sampling efficiency of MIA by leveraging prior knowledge from the VPR literature, such as place class prediction distributions, and unlearned class detection techniques. We also evaluate the effects of these methods both individually and in combination.

Through extensive experiments, we demonstrate that the proposed approach significantly enhances performance in continual learning under a challenging setup where all robots are black boxes and perform poorly.

Our key contributions include:

1. We extend the continual learning (CL) framework from a single-robot to a multi-robot paradigm, enabling knowledge transfer between distinct robots.
2. We introduce a method for reconstructing pseudo-training samples from a black-box teacher model using membership inference attacks (MIA), targeting the teacher-student VPR knowledge transfer problem.
3. We empirically demonstrate that even in a near-worst-case scenario involving only low-performing black-box robots, collaborative learning can significantly boost VPR performance.

This work generalizes earlier studies on VPR, continual learning, and knowledge transfer into a novel data-free setting applicable to black-box teachers. This study significantly extends our late-breaking paper in (Tsukahara et al., 2024) and generalizes our previous research on VPR, including continual learning (Tanaka, 2015), knowledge transfer from teacher to student (Takeda and Tanaka, 2021), and knowledge distillation (Hiroki and Tanaka, 2019), into a novel data-free variant applicable to black-box teachers. As autonomous robots continue to proliferate, universal knowledge transfer protocols and continual domain adaptation will become increasingly important. This research lays the foundation for optimizing robot-to-robot interactions in various VPR scenarios and aims to enhance robotic autonomy in open-world environments.

2 RELATED WORK

Visual Place Recognition (VPR). VPR is a specialized self-localization task that has garnered considerable research attention (Weyand et al., 2016; Morita et al., 2005; Seo et al., 2018). Compared to other self-localization tasks, including image retrieval (Garcia-Fidalgo and Ortiz, 2018), map matching (Sobreira et al., 2019), sequence matching (Milford and Wyeth, 2012), and multiple-hypothesis tracking (Rozsypalek et al., 2023), VPR is particularly scalable. This scalability is demonstrated through the development of compact VPR models for large-scale self-localization at a planetary scale (Weyand et al., 2016), tuning-free VPR models for diverse weather and seasonal conditions (Morita et al., 2005), and fine-grained models with combinatorial spatial partitioning (Seo et al., 2018).

Continual Learning (CL). Continual learning in robotics addresses the challenge of maintaining and updating knowledge over time. Recent advances span loop closure detection (Gao et al., 2022), visual odometry (Vödisch et al., 2023), SLAM (Vödisch et al., 2022), and localization (Cabrera-Ponce et al., 2023). However, most existing approaches focus on single-robot scenarios, with limited attention to multi-robot knowledge transfer. For an overview of CL taxonomies and techniques, see (Lange et al., 2022). CL methods have been increasingly applied in both computer vision (Lomonaco et al., 2022) and robotics

(Lesort et al., 2020). Benefits of CL in localization include online adaptation (Wang et al., 2021) and real-time learning (Cabrera-Ponce et al., 2023). While some studies address CL in multi-agent systems (Casado et al., 2020), research on data-free multi-robot CL remains sparse.

According to the CL taxonomy (Lange et al., 2022), the scenario targeted in this study spans domain-incremental, class-incremental, and vocabulary-incremental settings, wherein a robot learns across seasonal domains (e.g., inter-season), accumulates new place classes, and adapts to new teacher models (e.g., vocabulary changes).

Privacy Attacks. In machine learning applications such as medical image analysis, privacy concerns about reconstructing training samples—e.g., portraits or sensitive images—during model transfer have led to a growing body of research on privacy attacks and defenses (Liu et al., 2021). Among these, model inversion methods (Fredrikson et al., 2015) are of particular relevance. These aim to reconstruct pseudo-training samples (or “impressions”) from teacher models, which the student model can then learn from.

For example, early zero-shot KT approaches modeled the softmax space with Dirichlet distributions and reconstructed data from the teacher accordingly (Nayak et al., 2019). Other works such as DAFL (Chen et al., 2019) introduced regularization based on assumed access to activations and outputs. While recent studies have succeeded in reconstructing pseudo-datasets (Buzaglo et al., 2023), they still assume access to the internal architecture of teacher models. In contrast, membership inference attacks (MIAs) are the only major method applicable to black-box teachers.

The MIA method proposed in this study draws on insights from unlearned place detection in VPR (Kim et al., 2019), and aligns with recent metric-based MIA techniques (Ko et al., 2023). However, most prior MIA research has focused on one-shot attacks, and due to poor sampling efficiency, dataset reconstruction for VPR knowledge transfer remains an open challenge in black-box settings.

3 FORMULATION OF MULTI-ROBOT CONTINUAL LEARNING

This section first reviews the conventional tasks, including visual place recognition (VPR), supervised learning, and single-robot continual learning (CL). Building on this, we formulate the multi-robot CL problem and describe membership inference attacks (MIAs), which are central to our approach. Furthermore, as preparation for evaluation, we introduce a metric for assessing knowledge transfer cost.

A VPR model is defined as a function M that takes an input image x and outputs a probability distribution over place classes C:

$$M : x \to P(y \mid x), \quad y \in C, \tag{1}$$

where, C denotes a predefined set of place classes, each corresponding to a real-world spatial region. For example, in the NCLT public dataset (Carlevaris-Bianco et al., 2016) used in our experiments, grid-based partitioning (Kim et al., 2019) is applied, dividing the workspace into 10×10 grid cells in a bird's-eye coordinate system, each cell representing a distinct place class (Fig. 1). This approach provides a standardized place definition; however, it increases intra-class variation, thereby complicating classification[1] (Fig. 2).

In a traditional supervised learning setting, a robot experiences a set of place classes C_0, converts these experiences into a training set T_0, and trains a model M_0 via supervised learning L:

$$M_0 = L(T_0),$$

where

$$T = \{(x_i, y_i)\}_{i=1}^N.$$

In single-robot CL, when a new class set C_0^+ is observed with corresponding training data T_0^+, the goal is to update the existing model M_0 to a new model M_1. This is typically done by first reconstructing a pseudo training set $\bar{T}_0$ using a model inversion function I (e.g., MIA):

$$\bar{T}_0 = I(M_0),$$

where

$$\bar{T} = \{(x_i, P(y \mid x_i))\}_{i=1}^N.$$

The reconstructed pseudo set is combined with the new training data:

$$\tilde{T}_0^+ = \bar{T}_0 \cup T_0^+,$$

and a distilled model M_1 is learned using distillation $\bar{L}$ (Hinton et al., 2015):

$$M_1 = \bar{L}(\tilde{T}_0^+).$$

The overall update process is summarized as

$$M_1 = \bar{L}(I(M_0) \cup T_0^+). \tag{2}$$

[1] Optimizing place definitions to reduce intra-class variation is a fundamental, ongoing challenge in the VPR community (Seo et al., 2018). This topic lies beyond the scope of the present work but is considered complementary.

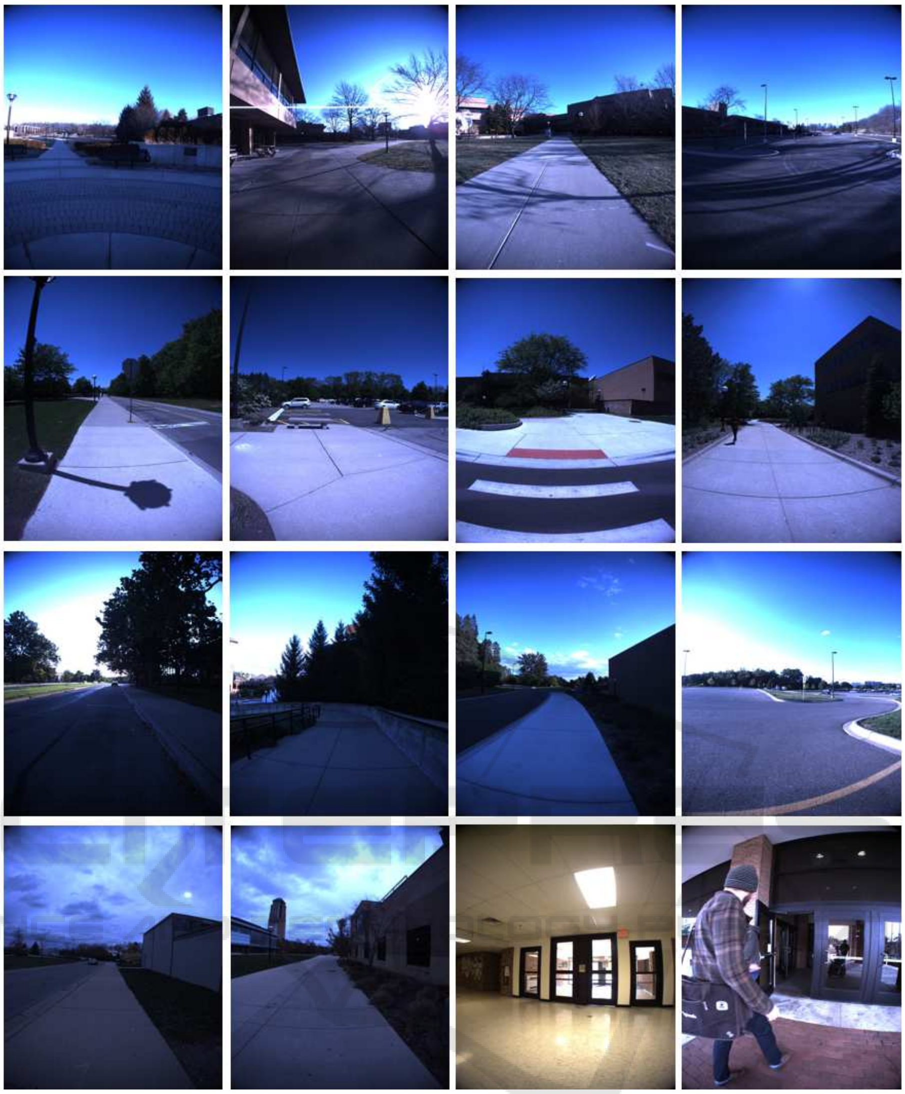


Figure 2: Sample input images from independent sessions. Each row corresponds to distinct place classes. Grid partitioning introduces intra-class variation, challenging the VPR task.

In multi-robot CL, instead of receiving new training data T_0^+, a new black-box teacher model M_0^+ is provided. In this scenario, a specialized inversion function I^* (i.e., MIA) applicable to black-box models is employed to reconstruct the pseudo training set:

$$\bar{T}_0^+ = I^*(M_0^+),$$

where, MIAs aim to approximate the inaccessible training data of a black-box teacher model via queries and responses. The overall model update can be written as:

$$M_1 = \bar{L}(I(M_0) \cup I^*(M_0^+)). \tag{3}$$

Finally, we consider the knowledge transfer cost in multi-robot CL. Interactions between the student and black-box teacher occur via queries (from student) and responses (from teacher). The queries are automatically generated programmatic requests, typically consisting of short code snippets, incurring negligible communication overhead. Responses, however, consist of pseudo training samples represented as high-dimensional real-valued vectors or tensors, imposing significant communication cost proportional to the number of samples transferred. Therefore, this study evaluates the knowledge transfer cost primarily based on the number of pseudo samples sent from the teacher to the student, assuming the cost of queries is negligible.

4 DATASET RECONSTRUCTION USING BLACK-BOX MEMBERSHIP INFERENCE ATTACKS (BB-MIA)

A critical challenge in applying Membership Inference Attacks (MIA) to black-box models (BB-MIA) lies in the high dimensionality of the sample space. Since the attacker (student) has no access to the teacher model's internal parameters or its training data distribution, generating informative samples to effectively probe the model's decision boundaries is highly non-trivial.

To address this, we approximate the teacher model with a cascade pipeline comprising two modules: a pre-trained embedding module and a trainable MIA module. The embedding module projects high-dimensional input images into lower-dimensional embedding vectors via a scene graph classifier, a method validated for generalizability and effectiveness in various VPR tasks (Takeda and Tanaka, 2021). The MIA module then employs multiple sampling strategies within the embedding space to enable efficient MIA sample generation. Details of each module follow.

4.1 Generic Embedding Model

Figure 3 illustrates the approximated pipeline. It consists of three key stages:

1. **Semantic Segmentation:** Input images are segmented into regions by DeepLab v3+ (Chen et al., 2018).
2. **Scene Graph Generation:** Segmented regions form nodes in a scene graph. Each node is represented by a 189-dimensional one-hot vector encoding semantic labels, orientation, and range (Lowry et al., 2016). Edges connect spatially adjacent nodes, capturing multimodal features including appearance, semantics, and spatial relations (Yoshida et al., 2024).
3. **Graph Neural Network:** The scene graph is fed into a pre-trained graph convolutional network (GCN) classifier, producing a class-specific probability map (CPM) of dimension $|C|$ (Ohta et al., 2023). This CPM serves as the final embedding vector.

4.2 Black-Box Membership Inference Attacks

The problem is to reconstruct a pseudo-sample set $\{(x, P(y|x))\}$ approximating the training data of a black-box teacher model, which takes embedding vectors as input and outputs CPMs. This is an inverse supervised learning problem.

Uniform Sampling (US) Strategy The simplest approach samples vectors $x \in \mathbb{R}^{|C|}$ from a uniform distribution, normalizes them with the L1 norm, and queries the teacher to obtain $P(y|x)$:

$$x_i = \frac{u_i}{\|u_i\|_1}, \quad u_i \sim U, \quad i = 1, \ldots, N. \tag{4}$$

Reciprocal Rank (RR) Strategy. The Uniform Sampling (US) strategy generates samples randomly and does not take into account the teacher model's output distribution. The RR strategy improves on this by converting uniform samples into reciprocal rank features (RRF), which reflect the relative strength of each class in the model's predictions (Takeda and Tanaka, 2021):

$$x_i = f^{RR}\left(\frac{u_i}{\|u_i\|_1}\right), \quad u_i \sim U, \quad i = 1, \ldots, N, \tag{5}$$

where

$$f^{RR}(x) = \left[\frac{1}{\text{rank}_1}, \frac{1}{\text{rank}_2}, \ldots, \frac{1}{\text{rank}_{|C|}}\right], \tag{6}$$

and rank_j is the position of the j-th element in descending order. In simple terms, the largest value gets 1/1 = 1, the second largest 1/2 = 0.5, the third largest $1/3 \simeq 0.33$, and so on. This transforms random values into features that encode the relative order of classes, making the samples more informative for probing the teacher model.

Entropy Strategy. While the RR strategy approximates general predictive distributions, it may not fit the target teacher model's specifics. The Entropy strategy selects samples with low predictive entropy, hypothesizing these correspond to training data members (Kim et al., 2019). From many RR-generated samples, the N with the lowest entropy

$$H(x) = -\sum_{i=1}^{|C|} P(y_i|x) \log P(y_i|x) \tag{7}$$

are chosen.

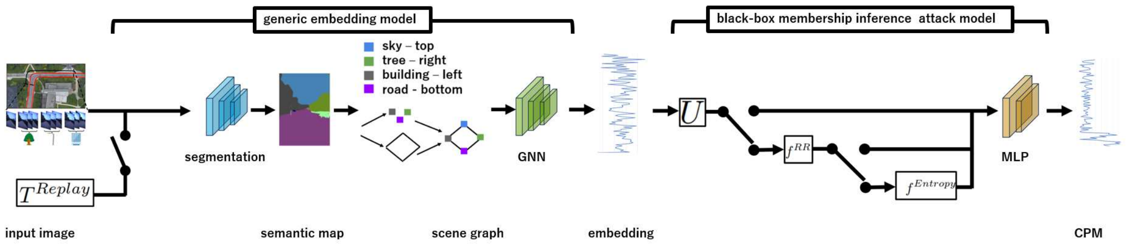


Figure 3: Overview of our VPR pipeline. The architecture includes a generic embedding module to reduce dimensionality and multiple sampling switches for model inversion analysis (MIA). Switch states shown correspond to the proposed ”Entropy” method.

Replay/Prior Strategies. Beyond BB-MIA, replay-based sampling assumes partial access to the teacher's training samples (Isele and Cosgun, 2018), contradicting the black-box assumption. Such settings are plausible with high communication and memory capacity in multi-robot systems. The Prior strategy assumes the student has access to its own training samples for querying the teacher, applicable when only the student cooperates.

Mixup Strategy. A hybrid approach combines a small subset R of retained training samples with $N-R$ samples from RR or Entropy strategies:

$$x_i \sim \begin{cases} T^{Replay}, & i = 1, \ldots, R \\ T^{RR/Entropy}, & i = R+1, \ldots, N \end{cases}, \tag{8}$$

where T^{Replay} is the teacher outputs for replayed (retained) training samples, and $T^{RR/Entropy}$ is the teacher outputs for samples generated by Reciprocal Rank or Entropy strategies. This strategy serves as an oracle baseline, not applicable to black-box teachers.

5 EXPERIMENTAL EVALUATION

This section evaluates the performance of the proposed method in a typical multi-robot continual learning (CL) scenario, where a traveling robot (student) sequentially interacts with three teacher robots via wireless communication. A key focus is the trade-off between acquiring new knowledge and avoiding forgetting previously learned place classes.

5.1 Experimental Setup

Experiments utilized the NCLT dataset (Carlevaris-Bianco et al., 2016), which contains sensor data from a Segway robot navigating multiple sessions across different seasons on the University of Michigan North Campus. Visual Place Recognition (VPR) tasks employ onboard camera images with ground-truth viewpoint GPS annotations. The dataset, originally collected by a single robot, is adapted to multi-robot scenarios by pairing distinct sessions with different robots (Mangelson et al., 2018).

The evaluation protocol comprises:

1. Sequential interactions with up to three teachers.
2. Utilization of 27 NCLT sessions:
   - One test session: “2012/08/04”
   - One session for embedding model training: “2012/04/29”
   - Twenty-five sessions for student/teacher VPR model training
3. Teacher and student VPR models implemented as multi-layer perceptrons (MLPs) with a 4,096-dimensional hidden layer.
4. Evaluation of six distinct knowledge transfer scenarios.
5. Performance metrics including:
   - Top-1 VPR accuracy measured after:
     – Student's supervised learning
     – Knowledge transfer from each teacher sequentially
   - Knowledge transfer (KT) cost quantified by the number of samples N used.
   - Knowledge retention, assessing avoidance of catastrophic forgetting.
   - Computational efficiency.

Additional details include:

1. Each student and teacher robot is trained on K place classes from its assigned session, using fully supervised learning with the corresponding training data. Following standard continual learning protocols, training sets are discarded after training, except for Replay, Prior, and Mixup strategies where some samples are retained.
2. Sessions are indexed from 0 to 24 chronologically according to navigation dates.

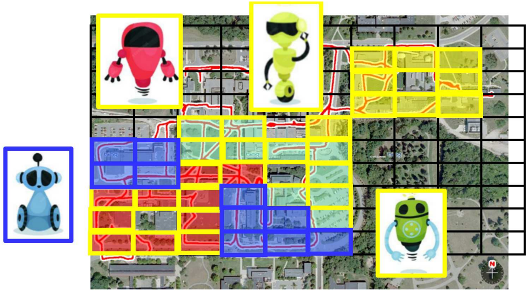


Figure 4: Experiment design overview. At stage $i = 0$, VPR model is trained on the student's prior experience. At subsequent stages $i = 1, 2, 3$, it is updated via knowledge transfer from distinct teachers.

3. Six scenarios ($j = 0, \ldots, 5$) evaluate different student-teacher combinations. For the j-th scenario, models with ID $i \in \{0, 1, 2, 3\}$ correspond to the student ($i = 0$) and teachers ($i = 1, 2, 3$), each trained on $K = 10$ random place classes from session $((6i + j) \bmod 25)$.
4. Sample counts are reported per place class for simplicity.
5. When encountering a teacher, pseudo-training samples for classes known to the teacher are reconstructed from the teacher model, while samples for classes exclusive to the student are reconstructed from the student model.
6. For Replay, Prior, and Mixup strategies, the number of replay samples per class R is set to 1 by default.
7. Reciprocal Rank Feature (RRF) vectors are approximated using sparse k-hot RRF with $k = 10$ as in (Ohta et al., 2023).

Computational costs were modest: MLP training required tens of seconds, and query sample generation took approximately 25 milliseconds per sample. Knowledge transfer cost for a 100-dimensional k-hot RRF sample was under 128 bits.

5.2 Results and Discussions

We first evaluated the basic performance of the proposed method. Performance was measured initially after supervised training of the student, and subsequently after sequential knowledge transfers from new teachers.

At the initial stage, the student's performance was predictably low, as the test set included samples from previously unseen place classes, highlighting the inherent difficulty of the VPR task under such conditions.

Figure 5 shows that performance improved with an increasing number of samples across all five

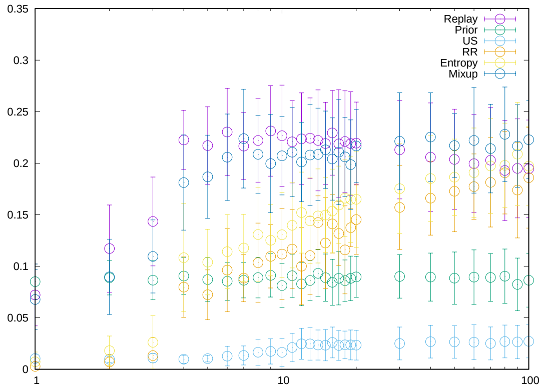


Figure 5: Top-1 accuracy vs. KT cost (N) under different replay conditions. $K = 10$.

strategies, though notable differences were observed among them:

Replay Strategy: When provided with sufficiently many samples, this strategy effectively mitigated catastrophic forgetting (Kang et al., 2023) and consistently achieved the highest accuracy. However, it does not conform to strict continual learning principles. The Prior strategy, a variant which queries the teacher with the student's own training samples, exhibited inferior performance, likely due to a low probability of overlap between the student's and teacher's training samples.

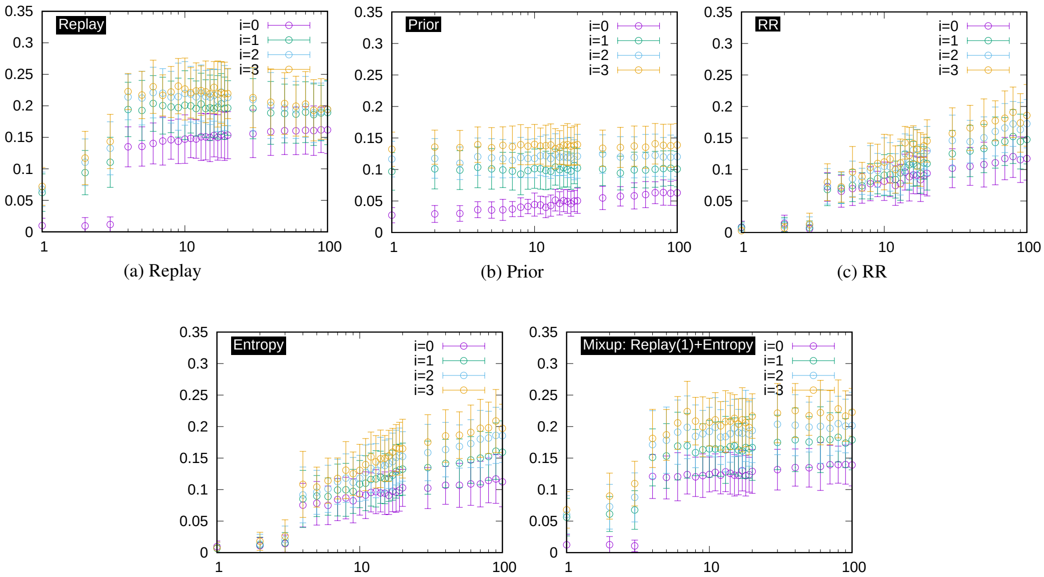
US Strategy: Uniform Sampling (US) showed the lowest performance in all experiments, presumably due to the uniform distribution not matching the teacher model's predictive distribution in the embedding space.

RR Strategy: Despite its simplicity, the Reciprocal Rank (RR) strategy surprisingly achieved high performance, indicating that the reciprocal rank feature distribution approximates the teacher's predictive distribution well.

Entropy Strategy: The Entropy strategy performed comparably or better than RR, especially excelling when the number of samples N was small by effectively selecting high-quality samples.

Mixup Strategy: Mixup balances generalization and communication cost effectively. Though it requires retaining a small number R of replay samples, its knowledge transfer cost is substantially lower than Replay while still mitigating forgetting and achieving comparable performance.

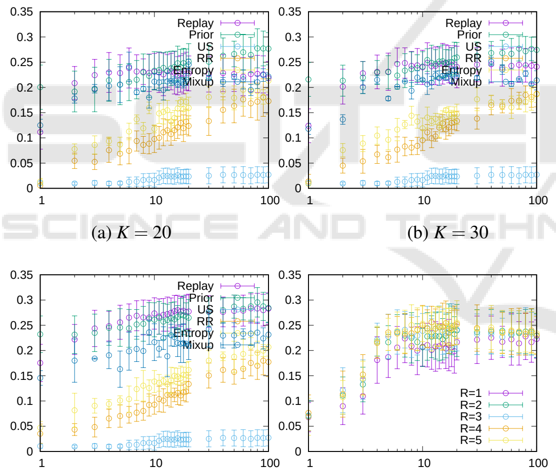
Figure 6 illustrates the student's VPR performance evolution through continual learning. The Replay strategy demonstrated the greatest stability, followed by Mixup, Entropy, RR, and US. Notably, Mixup closely matched Replay's stability by leveraging replay samples. Even strictly black-box



(a) Replay (b) Prior (c) RR

(d) Entropy (e) Mixup

Figure 6: Performance evolution of VPR models across continual learning stages.



(a) $K = 20$ (b) $K = 30$

(c) $K = 40$ (d) R=1,2,3,4,5

Figure 7: Parameter analysis on (a-c) known class count K and (d) replay sample size R.

teacher-compliant strategies such as RR and Entropy approached the performance of non-black-box approaches (Replay) when more than 20 samples per class were used. This demonstrates the promise of the proposed methods in challenging black-box MIA scenarios.

Figure 7 (a–c) depicts performance variation as the number of place classes K experienced by each robot changes. While RR and Entropy strategies exhibited a slight performance decline with increasing K, their accuracy remained comparable to Replay, evidencing robustness throughout continual learning.

Finally, Figure 7 (d) shows the impact of varying the number of replay samples R in the Mixup strategy. Even with a small R (e.g., $R = 1$), increasing the number of non-replay samples sufficiently improved VPR accuracy, suggesting that Mixup's practical knowledge transfer cost can be minimized without sacrificing performance.

In summary, the proposed approach substantially improved performance under challenging black-box MIA settings. In particular, the RR strategy achieved surprisingly strong results despite its simplicity. The Entropy strategy further enhanced performance with minimal additional cost. These findings build upon insights from non-black-box MIA tasks in VPR (Takeda and Tanaka, 2021)(Kim et al., 2019) (Section 4.2) and lay a foundation for future multi-robot continual learning research.

6 CONCLUSION AND FUTURE WORK

This study formulated continual multi-robot learning for visual place recognition as a teacher-student knowledge transfer problem, leveraging membership inference attacks (MIA)— the only major privacy attack applicable to black-box teachers. To address MIA's critical bottleneck of low sampling efficiency, we proposed utilizing the student model's prior knowledge to improve practical sampling efficiency. Extensive experiments demonstrated significant effectiveness of the approach, revealing the interplay between VPR accuracy, sampling efficiency, and computational cost.

Our investigation focused on a near-worst-case scenario where all teachers are black boxes and all robots have limited initial performance. For practical deployment, future work should explore leveraging heterogeneous teacher models, including white-box and high-performing teachers, to further enhance learning. Moreover, improving VPR performance may benefit from several well-known extensions: (1) advancing from grid-based to adaptive workspace partitioning for place definitions, (2) extending from single-view to multi-view VPR, and (3) shifting from passive to active VPR with robot control.

Immediate future directions include applying these advancements to challenging real-world tasks and addressing the lost robot problem to realize robust long-term autonomy in open-world environments.

REFERENCES

Buzaglo, G., Haim, N., Yehudai, G., Vardi, G., Oz, Y., Nikankin, Y., and Irani, M. (2023). Deconstructing data reconstruction: Multiclass, weight decay and general losses. In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*.

Cabrera-Ponce, A. A., Martin-Ortiz, M., and Martinez-Carranza, J. (2023). Continual learning for topological geo-localisation. *Journal of Intelligent & Fuzzy Systems*, 44(6):10369–10381.

Carlevaris-Bianco, N., Ushani, A. K., and Eustice, R. M. (2016). University of michigan north campus long-term vision and lidar dataset. *Int. J. Robotics Res.*, 35(9):1023–1035.

Casado, F. E., Lema, D., Iglesias, R., Regueiro, C. V., and Barro, S. (2020). Collaborative and continual learning for classification tasks in a society of devices. *arXiv preprint arXiv:2006.07129*.

Chen, H., Wang, Y., Xu, C., Yang, Z., Liu, C., Shi, B., Xu, C., Xu, C., and Tian, Q. (2019). Data-free learning of student networks. In *2019 IEEE/CVF International Conference on Computer Vision*, pages 3513–3521. IEEE.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *ECCV*.

Fredrikson, M., Jha, S., and Ristenpart, T. (2015). Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security*, pages 1322–1333.

Gao, D., Wang, C., and Scherer, S. (2022). Airloop: Lifelong loop closure detection. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 10664–10671. IEEE.

Garcia-Fidalgo, E. and Ortiz, A. (2018). ibow-lcd: An appearance-based loop-closure detection approach using incremental bags of binary words. *IEEE Robotics Autom. Lett.*, 3(4):3051–3057.

Hinton, G. E., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. *CoRR*, abs/1503.02531.

Hiroki, T. and Tanaka, K. (2019). Long-term knowledge distillation of visual place classifiers. In *2019 IEEE Intelligent Transportation Systems Conference, ITSC*, pages 541–546. IEEE.

Hu, H., Salcic, Z., Sun, L., Dobbie, G., Yu, P. S., and Zhang, X. (2022). Membership inference attacks on machine learning: A survey. *ACM Computing Surveys (CSUR)*, 54(11s):1–37.

Isele, D. and Cosgun, A. (2018). Selective experience replay for lifelong learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, pages 3302–3309. AAAI Press.

Kang, M., Zhang, J., Zhang, J., Wang, X., Chen, Y., Ma, Z., and Huang, X. (2023). Alleviating catastrophic forgetting of incremental object detection via within-class and between-class knowledge distillation. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 18848–18858.

Kim, G., Park, B., and Kim, A. (2019). 1-day learning, 1-year localization: Long-term lidar localization using scan context image. *IEEE Robotics Autom. Lett.*, 4(2):1948–1955.

Ko, M., Jin, M., Wang, C., and Jia, R. (2023). Practical membership inference attacks against large-scale multi-modal models: A pilot study. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4871–4881.

Lange, M. D., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., Slabaugh, G. G., and Tuytelaars, T. (2022). A continual learning survey: Defying forgetting in classification tasks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(7):3366–3385.

Lesort, T., Lomonaco, V., Stoian, A., Maltoni, D., Filliat, D., and Díaz-Rodríguez, N. (2020). Continual learning for robotics: Definition, framework, learning strategies, opportunities and challenges. *Information fusion*, 58:52–68.

Liu, Y., Zhang, W., Wang, J., and Wang, J. (2021). Data-free knowledge transfer: A survey. *CoRR*, abs/2112.15278.

Lomonaco, V., Pellegrini, L., Rodriguez, P., Caccia, M., She, Q., Chen, Y., Jodelet, Q., Wang, R., Mai, Z., Vazquez, D., et al. (2022). Cvpr 2020 continual learning in computer vision competition: Approaches, results, current challenges and future directions. *Artificial Intelligence*, 303:103635.

Lowry, S. M., Sünderhauf, N., Newman, P., Leonard, J. J., Cox, D. D., Corke, P. I., and Milford, M. J. (2016). Visual place recognition: A survey. *IEEE Trans. Robotics*, 32(1):1–19.

Mangelson, J. G., Dominic, D., Eustice, R. M., and Vasudevan, R. (2018). Pairwise consistent measurement set maximization for robust multi-robot map merging. In *2018 IEEE international conference on robotics and automation (ICRA)*, pages 2916–2923. IEEE.

Milford, M. J. and Wyeth, G. F. (2012). Seqslam: Visual route-based navigation for sunny summer days and stormy winter nights. In *2012 IEEE international conference on robotics and automation*, pages 1643–1649. IEEE.

Morita, H., Hild, M., Miura, J., and Shirai, Y. (2005). View-based localization in outdoor environments based on support vector learning. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2965–2970. IEEE.

Nayak, G. K., Mopuri, K. R., Shaj, V., Radhakrishnan, V. B., and Chakraborty, A. (2019). Zero-shot knowledge distillation in deep networks. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 4743–4751. PMLR.

Ohta, T., Tanaka, K., and Yamamoto, R. (2023). Scene graph descriptors for visual place classification from noisy scene data. *ICT Express*, 9(6):995–1000.

Rozsypalek, Z., Rou?ek, T., Vintr, T., and Krajnik, T. (2023). Multidimensional particle filter for long-term visual teach and repeat in changing environments. *IEEE Robotics and Automation Letters*, 8(4):1951–1958.

Seo, P. H., Weyand, T., Sim, J., and Han, B. (2018). Cplanet: Enhancing image geolocalization by combinatorial partitioning of maps. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part X*, volume 11214 of *Lecture Notes in Computer Science*, pages 544–560. Springer.

Sobreira, H., Costa, C. M., Sousa, I., Rocha, L., Lima, J., Farias, P., Costa, P., and Moreira, A. P. (2019). Map-matching algorithms for robot self-localization: a comparison between perfect match, iterative closest point and normal distributions transform. *Journal of Intelligent & Robotic Systems*, 93:533–546.

Takeda, K. and Tanaka, K. (2021). Dark reciprocal-rank: Teacher-to-student knowledge transfer from self-localization model to graph-convolutional neural network. In *IEEE International Conference on Robotics and Automation, ICRA 2021, Xi'an, China, May 30 - June 5, 2021*, pages 1846–1853. IEEE.

Tanaka, K. (2015). Cross-season place recognition using NBNN scene descriptor. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2015, Hamburg, Germany, September 28 - October 2, 2015*, pages 729–735. IEEE.

Tsukahara, K., Tanaka, K., and Iwata, D. (2024). Training self-localization models for unseen unfamiliar places via teacher-to-student data-free knowledge transfer. In *IEEE Conference on Intelligent Robots and Systems (IROS 2024 Late Breaking Paper)*.

Vödisch, N., Cattaneo, D., Burgard, W., and Valada, A. (2022). Continual slam: Beyond lifelong simultaneous localization and mapping through continual learning. In *The International Symposium of Robotics Research*, pages 19–35. Springer.

Vödisch, N., Cattaneo, D., Burgard, W., and Valada, A. (2023). Covio: Online continual learning for visual-inertial odometry. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2464–2473.

Wang, S., Laskar, Z., Melekhov, I., Li, X., and Kannala, J. (2021). Continual learning for image-based camera localization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3252–3262.

Weyand, T., Kostrikov, I., and Philbin, J. (2016). Planet - photo geolocation with convolutional neural networks. In Leibe, B., Matas, J., Sebe, N., and Welling, M., editors, *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII*, volume 9912 of *Lecture Notes in Computer Science*, pages 37–55. Springer.

Yoshida, M., Yamamoto, R., Iwata, D., and Tanaka, K. (2024). Open-world distributed robot self-localization with transferable visual vocabulary and both absolute and relative features. *CoRR*, abs/2109.04569.