**Introduction**

While great progress has been made in lifelong learning, it is still challenging to deploy the existing algorithms in the wild to learn over time in a real-world application, e.g., self-driving vehicles. Existing works rely on prior knowledge to produce good results. We aim at closing the gap towards real-world lifelong learning.

**Problem Definition**

Online unsupervised lifelong learning without prior knowledge:
- Non-iid and single-pass data streams
- No task or class labels
- No prior knowledge, e.g., task/class shift boundaries

We consider four different types of class-incremental streams inspired from real-world applications.

**Method**

We design SCALE around three components which function collaboratively to maximize the performance:
- Pseudo-supervised contrastive loss
- Self-supervised forgetting loss
- Online memory update for uniform subset selection

**Loss Function**

\[ \mathcal{L} = \mathcal{L}_{\text{contrastive}} + \lambda \cdot \mathcal{L}_{\text{forgetting}} \]

Pseudo-supervised contrastive loss aims to enhance similarity between positive pairs over negative pairs.

\[ \mathcal{L}_{\text{contrastive}} = \sum_{i=1}^{2m} \sum_{j=1}^{m} \log \frac{\exp(d_i - d_j)}{\sum_{k=1}^{2m} \exp(d_i - d_k)} \]

Self-supervised forgetting loss aims to preserve the knowledge of pairwise similarity, and mitigate forgetting.

\[ \mathcal{L}_{\text{forgetting}} = \sum_{i=1}^{2m} \sum_{j=1}^{m} -p_{ij} \log \frac{p_{ij}}{p_{ij} + p_{ji}} \]

**Online Memory Update** is critical for the overall performance, aiming for keeping a balanced subset of historical samples. We employ the Part and Select Algorithm (PSA) [1] for uniform online subset selection.

**Experiments**

We experiment on five types of iid and class-incremental streams sampled from CIFAR-10, CIFAR-100 and TinyImageNet.

Key baselines: STAM [IJCAI 2021], CaSSLe [CVPR 2022], LUMP [ICLR 2021]

SCALE outperforms the best state-of-the-art algorithm on all settings with improvements of up to 6.43%, 5.23%, 5.86% kNN accuracy on CIFAR-10, CIFAR-100 and TinyImageNet.

SCALE enjoys gradually increasing kNN accuracy as we introduce new classes, while most baselines are subject to forgetting.

**References**