SCALE: Online Self-supervised Lifelong Learning without Prior Knowledge

Xiaofan Yu[1], Yunhui Guo[2], Sicun Gao[1], Tajana Rosing[1]
[1] University of California San Diego [2] University of Texas at Dallas

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

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

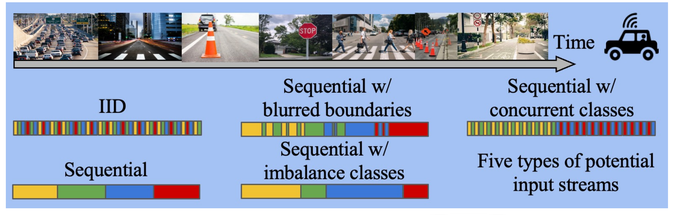
Table 1. Comparison of previous work and SCALE (this paper) on assumed prior knowledge.

Papers	Single-pass	Non-*iid*	No task labels	No class labels
VASE [2], CURL [61], L-VAEGAN [77]	×	✓	✓	✓
He *et al.* [31], CCSL [46], CaSSLe [24], LUMP [49]	✓	✓	×	✓
Tiezzi *et al.* [70], KIERA [57]	✓	✓	✓	×
STAM [68], SCALE (this paper)	✓	✓	✓	✓

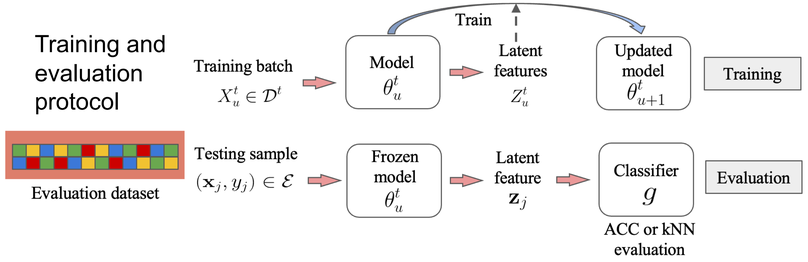
- We aim at closing the gap towards real-world lifelong learning

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



Training and evaluation protocol



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 $\mathcal{L} = \mathcal{L}^{\text{cont}} + \lambda \cdot \mathcal{L}^{\text{forget}}$.

1

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

$$\mathcal{L}^{\text{cont}} = \sum_{i=1}^{2n} \frac{-1}{|\Gamma_i|} \sum_{j \in \Gamma_i} \log \frac{\exp(\tilde{\mathbf{z}}_i \cdot \tilde{\mathbf{z}}_j / \tau)}{\sum_{k=1, k \neq i}^{2(m+n)} \exp(\tilde{\mathbf{z}}_i \cdot \tilde{\mathbf{z}}_k / \tau)},$$

Pseudo-positive set, defined as samples in the near-by neighborhood:

$$\Gamma_i = \{j \in \{1, ..., 2n\} \mid j \neq i, p_{ij} > \mu\},$$

Similarity threshold

2

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

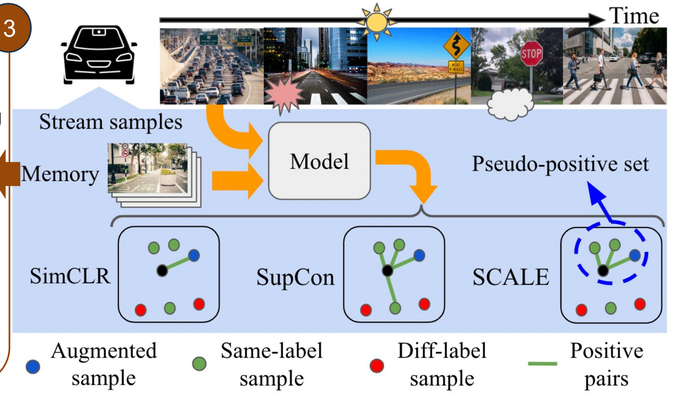
$$\mathcal{L}^{\text{forget}} = \sum_{i=1}^{2(m+n)} \sum_{j=1, j \neq i}^{2(m+n)} -p_{ij} \cdot \log \frac{p_{ij}}{p_{ij}^{past}},$$

p_{ij}, p_{ij}^{past}: pairwise similarity among feature representations, on the current batch, using current model and the frozen model from the previous batch

3

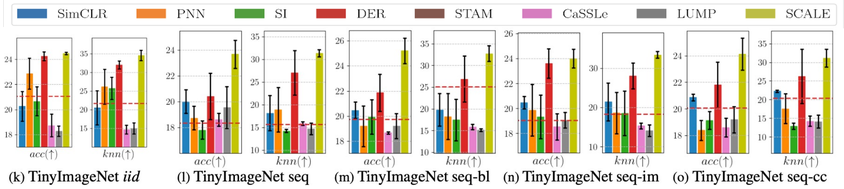
Online memory update is critical for the overall performance, aiming for keeping a balanced subset of historical samples.

We employ the Part and Select Algorithm (PSA) [1] for uniform online subset selection.



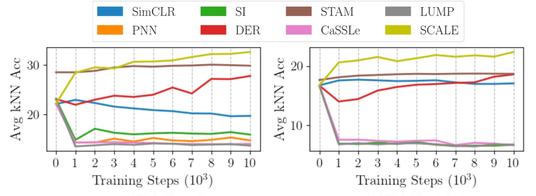
Experiments

- 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



(k) TinyImageNet *iid* (l) TinyImageNet seq (m) TinyImageNet seq-bl (n) TinyImageNet seq-im (o) TinyImageNet seq-cc

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



CIFAR-10 CIFAR-100

- In ablation study, we found that uniform online memory update is important for contrastive-based learning methods like SCALE

Table 3. Average final *k*NN accuracy on the *imbalanced sequential* streams using different memory update policies in SCALE.

Memory update	CIFAR-10	CIFAR-100	TinyImageNet
w/ label	32.41	21.21	27.73
random	29.80	20.10	23.67
KMeans	31.59	22.15	29.07
MinRed [58]	23.66	19.75	25.13
PSA (this paper)	**32.21**	**23.16**	**31.33**

Our code is available at: https://github.com/Orienfish/SCALE (or scan the QR code)

References

[1] Salomon, Shaul, et al. "PSA-A New Scalable Space Partition Based Selection Algorithm for MOEAs." *EVOLVE*. 2012.