EVOLVE: Enhancing Unsupervised Continual Learning with Multiple Experts

Xiaofan Yu[1], Tajana Rosing[1], Yunhui Guo[2]

[1] University of California San Diego
[2] University of Texas at Dallas

https://github.com/Orienfish/EVOLVE

WACV 2024

Continual Learning

- Continual Learning
  - To continually learn over time by acquiring new knowledge as well as consolidating past experiences
  - Key assumption: continuously changing environments
  - Key challenge: catastrophic forgetting

- While great progress has been made in continual learning, there is still a gap between existing continual learning algorithms and real-world deployments, due to
  - Unpredictable streaming input
  - Lack of supervision and prior knowledge

We consider the **unsupervised continual learning (UCL)** problem

Related Work

- Most existing works in UCL rely on various prior knowledge to produce good results
  - Cons: these prior knowledge may not be available in real-world applications

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]	✓	✓	✓	✓

- Self-supervised learning methods with a replay buffer can alleviate catastrophic forgetting on non-iid data streams [ECCV'22]
  - Cons: it is unclear how to effectively learn new patterns in an unpredictable world

- In this work, EVOLVE aims at closing the gap between continual learning algorithms and real-world applications by eliminating the prior assumptions.

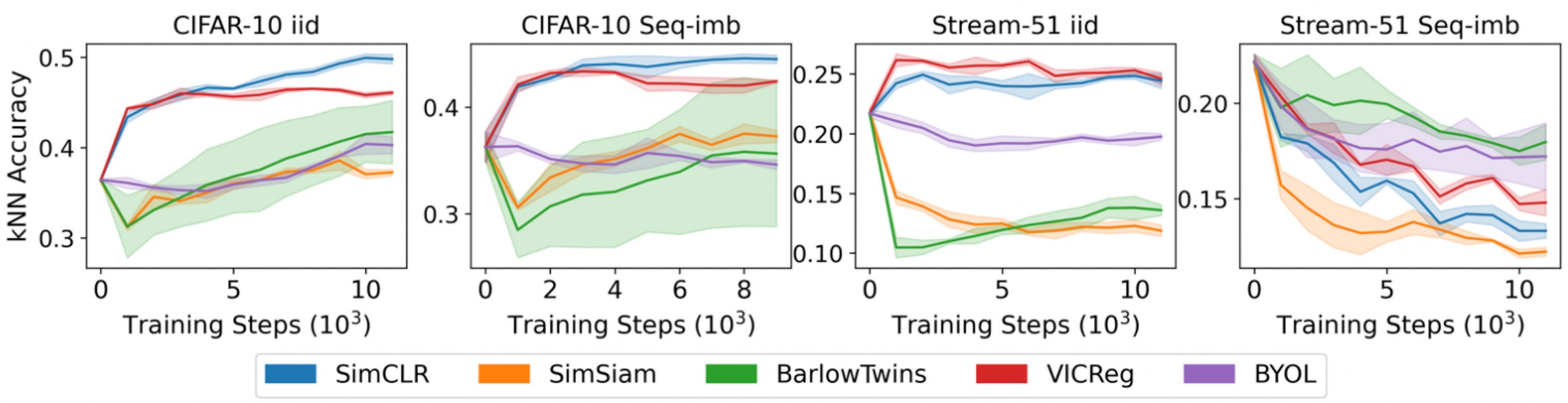
Purushwalkam, Senthil, et al. "The challenges of continuous self-supervised learning." *ECCV'22*

An Empirical Study of Existing Self-Supervised Learning (SSL) Baselines

- To study whether SSL w/ replay buffer is sufficiently practical, we conduct empirical study
  - Dataset: CIFAR-10 (image), Stream-51 (video) [CVPRW'20]
  - Data streams: iid vs. temporally correlated

Table 1. Overview of state-of-the-art SSL methods and losses.

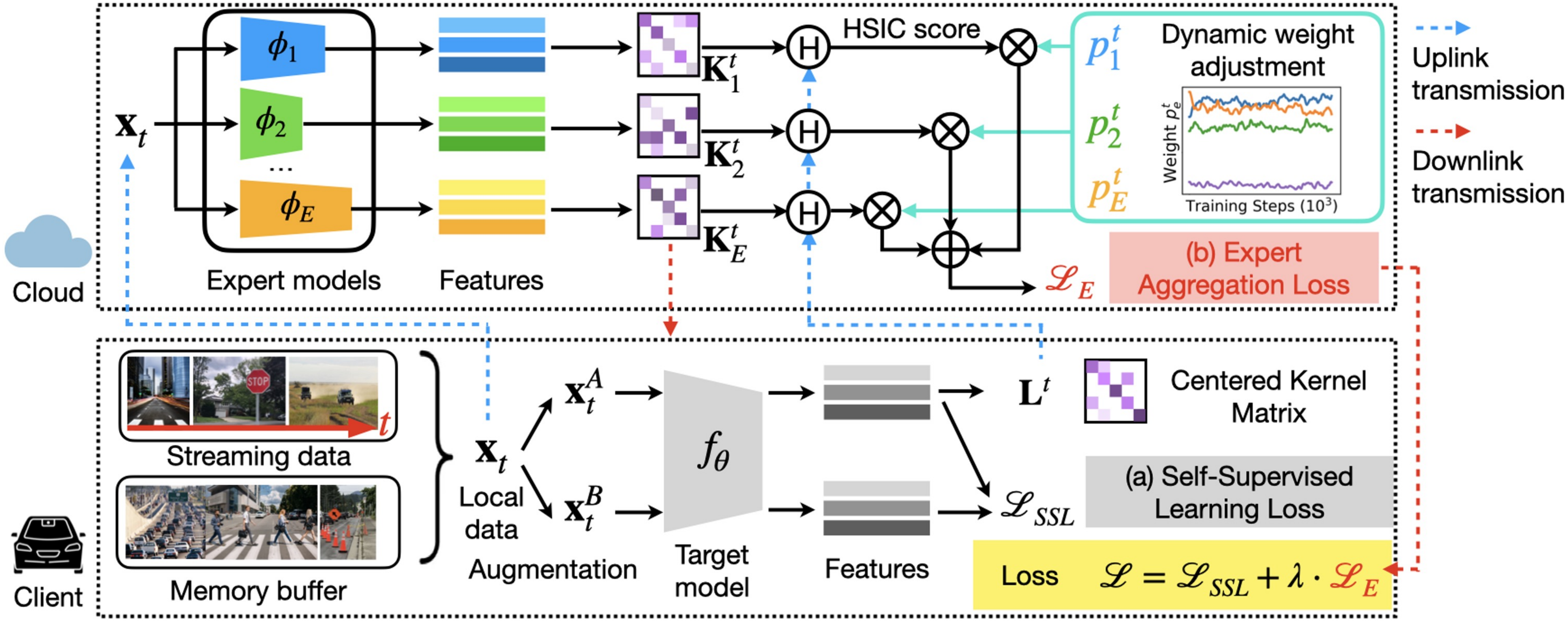
Methods	Loss	Loss Function $\mathcal{L}_{SSL}$
SimCLR [20]	InfoNCE	$-\log \frac{\exp(\mathbf{z}_i^A \cdot \mathbf{z}_j^B/\tau)}{\sum_{k\neq i}\exp(\mathbf{z}_i^A \cdot \mathbf{z}_k/\tau)}$
BYOL [35]	MSE	$\left\Vert \mathbf{q}_t^A - \mathbf{z}_t^B \right\Vert_2^2$
SimSiam [23]	MSE	$\mathcal{D}(\mathbf{q}_t^A, \mathbf{z}_t^B)$
Barlow Twins [92]	Cross-Correlation	$\sum_u (1-\mathcal{C}_{uu})^2 + \psi \sum_u \sum_{v\neq u} \mathcal{C}_{uv}^2$
VICReg [7]	MSE + Variance + Cross-Correlation	$\psi s(\mathbf{z}_t^A, \mathbf{z}_t^B) + \mu v(\mathbf{z}_t^A) + \nu c(\mathbf{z}_t^A)$



SSL baselines experience a **significant accuracy drop** when applied on video datasets with temporally correlated data streams

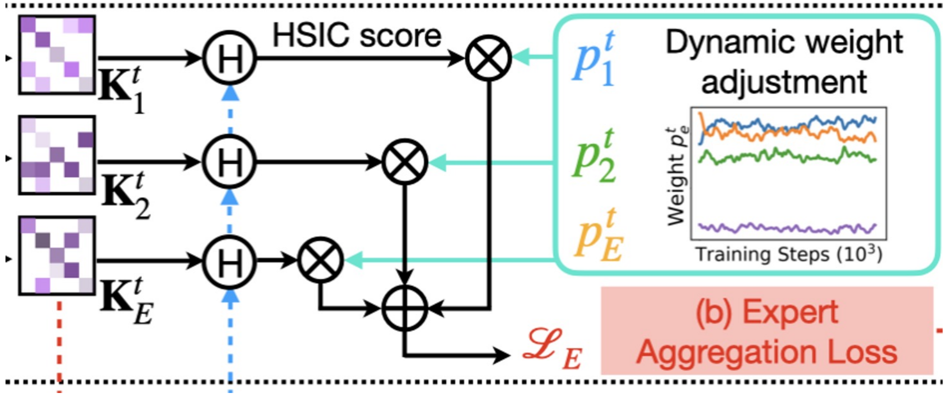
Roady, Ryne, et al. "Stream-51: Streaming classification and novelty detection from videos." CVPRW'20

Our Method: EVOLVE



- We propose EVOLVE, a hybrid UCL framework with (1) SSL training on the local client and (2) expert-guided training on the cloud, transmitting a small set of data and intermediate features to the cloud

Two Key Designs in EVOLVE



- The design of EVOLVE has two key parts
- We use the Hilbert-Schmidt independence criterion (HSIC) for assessing a model's ability in representation learning

$$HSIC(X,Y) = HSIC(K,L) = \frac{1}{(n-1)^2} Tr(KHLH),$$

$$\text{where } H = I - \frac{1}{N} 11^T$$

1 Expert Aggregation Loss based on HSIC

$$\mathcal{L}_E = -\sum_{e=1}^{E} p_e^t \cdot \mathrm{HSIC}(\mathbf{K}_e^t, \mathbf{L}^t),$$

Expert weights

2 Dynamic weight adjustment based on a confidence metric

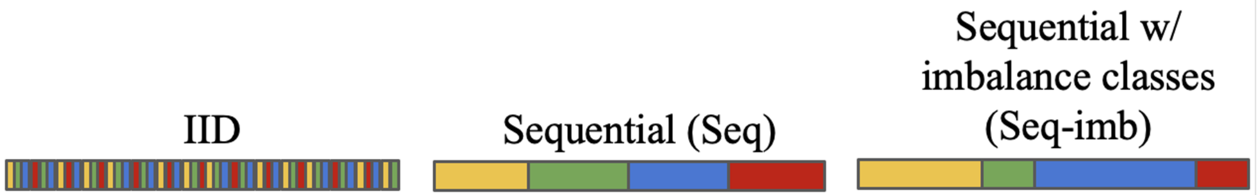
$$q_e^t = \sum_i \frac{\exp(\mathbf{h}_{e,i}^A \cdot \mathbf{h}_{e,i}^B/\tau)}{\sum_{k\neq i} \exp(\mathbf{h}_{e,i}^A \cdot \mathbf{h}_{e,k}^A/\tau) + \sum_k \exp(\mathbf{h}_{e,i}^A \cdot \mathbf{h}_{e,k}^B/\tau)},$$

$$w_e^{t+1} = \alpha w_e^t + (1-\alpha) q_e^t. \qquad p_e^t = w_e^t / \sum_l w_l^t$$

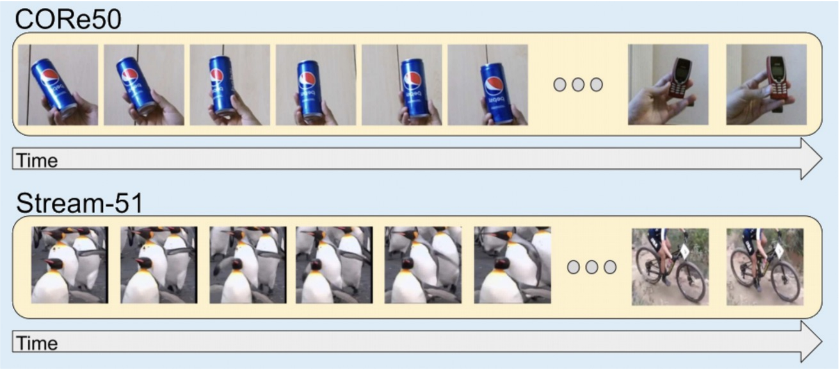
Moving average-based update

Experimental Setup

- **Datasets:** CIFAR-10, TinyImageNet, CORe50 [PMLR'17], Stream-51 [CVPRW'20]
- **Data streams:**



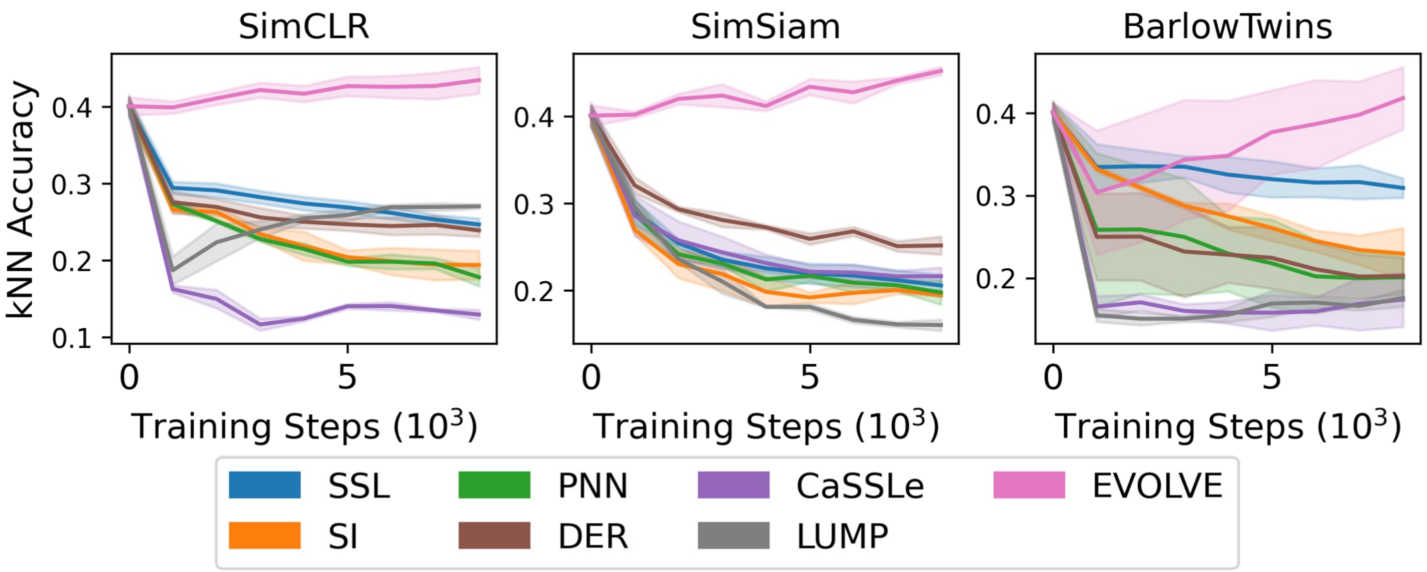
- **Baselines:**
  - Synaptic Intelligence (SI) [ICML'17]
  - Progressive Neural Network (PNN) [arXiv'16]
  - Dark Experience Replay (DER) [NeurIPS'20]
  - CaSSLe [CVPR'22]
  - Lifelong Unsupervised Mixup (LUMP) [ICLR'22]



- **Target model:** ResNet-18
- **Experts:** pretrained ResNet-50, Swin Transformer downloaded from torchvision
- **Metrics:** kNN accuracy, linear evaluation accuracy

Main Accuracy Results

- EVOLVE outperforms the top baseline using the same SSL by **3.6-20.0% in kNN accuracy and 6.1- 53.7% in top-1 linear evaluation accuracy** across diverse data streams



Conclusion

- Existing UCL algorithms cannot generalize to real-world scenarios due to
  - Unpredictable streaming input
  - Lack of supervision and prior knowledge
- We propose a general expert-guided continual learning framework, called EVOLVE, with local SSL training on clients and expert-guided training on the cloud
- EVOLVE has two key designs
  - Expert aggregation loss based on HSIC to distill guidance
  - Dynamic weight update to adjust the “power” of diverse experts based on latest data input
- EVOLVE outperforms existing UCL baselines on both image- and video-based data streams using the same SSL backbone

References

- Purushwalkam, Senthil, et al. "The challenges of continuous self-supervised learning." *ECCV’22*
- Roady, Ryne, et al. "Stream-51: Streaming classification and novelty detection from videos." CVPRW’20
- Lomonaco, Vincenzo, and Davide Maltoni. "Core50: a new dataset and benchmark for continuous object recognition." PMLR’17
- Zenke, Friedemann, Ben Poole, and Surya Ganguli. "Continual learning through synaptic intelligence." ICML’17
- Rusu, Andrei A., et al. "Progressive neural networks." *arXiv’16*
- Buzzega, Pietro, et al. "Dark experience for general continual learning: a strong, simple baseline." *NeurIPS’20*
- Fini, Enrico, et al. "Self-supervised models are continual learners." *CVPR’22*
- Madaan, Divyam, et al. "Representational continuity for unsupervised continual learning." ICLR’22