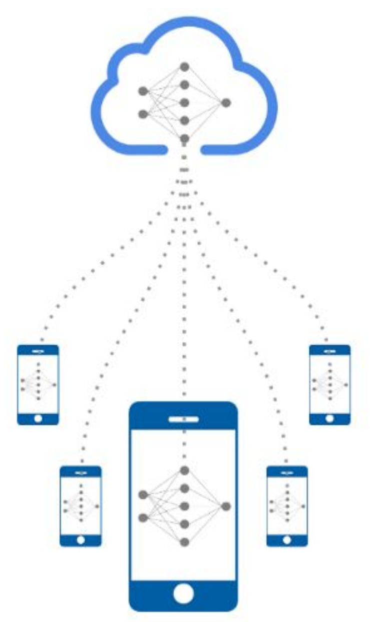
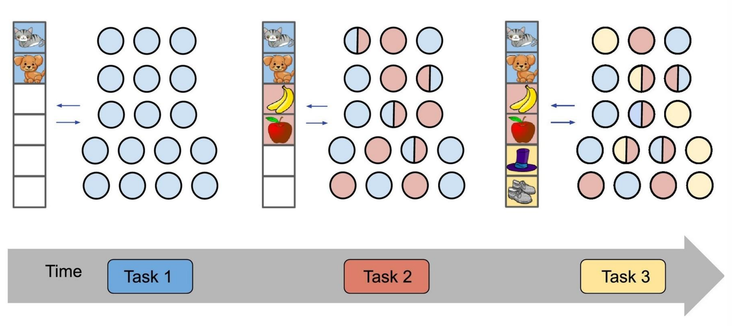
LEarning and Algorithms for People and Systems @ JHU-UCLA

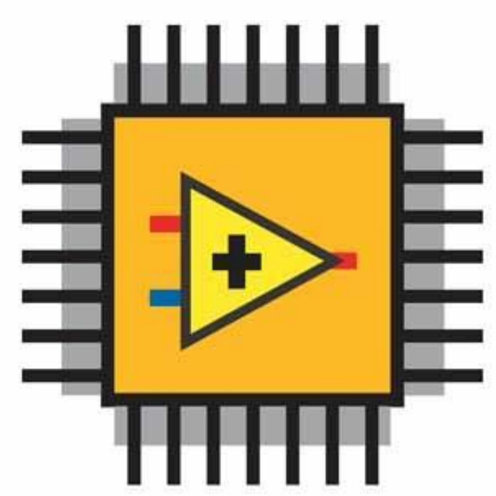
Poster: FLLEdge: Federated Lifelong Learning on Edge Devices

Ziang Song, Zhuolong Yu, Jingfeng Wu, Lin Yang, Vladimir Braverman

Motivation







Federated Learning (FL)

FL learns generalized central model learns generalized knowledge by aggregating collective wisdom from each independent client.

Lifelong Learning (CL)

LL can effectively **resist catastrophic forgetting** when models train on a sequence of unique tasks by retaining knowledge of previous tasks.

Low-power Devices

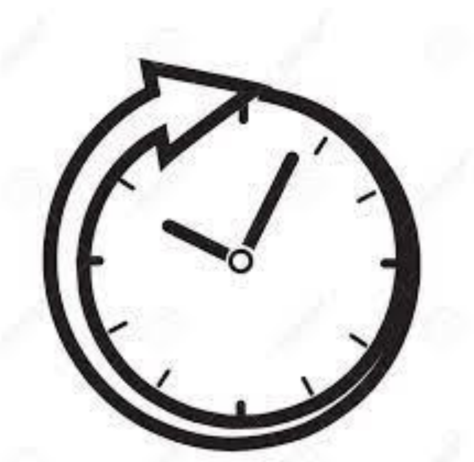
Single low-power devices may be **challenging to train on computationally expensive machine learning tasks**. But they are **cheap** and **numerous**, can be used to overcome hardware constraints.

Question: Can we support **Federated Lifelong Learning on low-power devices**?

Challenges

Challenge 1: Contents

FLLEdge agents must share knowledge useful to other agents' generalization.



Challenge 2: Communications

An **efficient asynchronous** communication scheme is needed for sharing agent knowledge.

Challenge 3: Computation

Lifelong Learning algorithms are **computationally expensive** to run on low power devices.

Challenge 4: Catastrophic Forgetting

Deployed Lifelong Learning algorithms must effectively balance new and old knowledge.

System Overview

Figure 1: Overview of FLLEdge

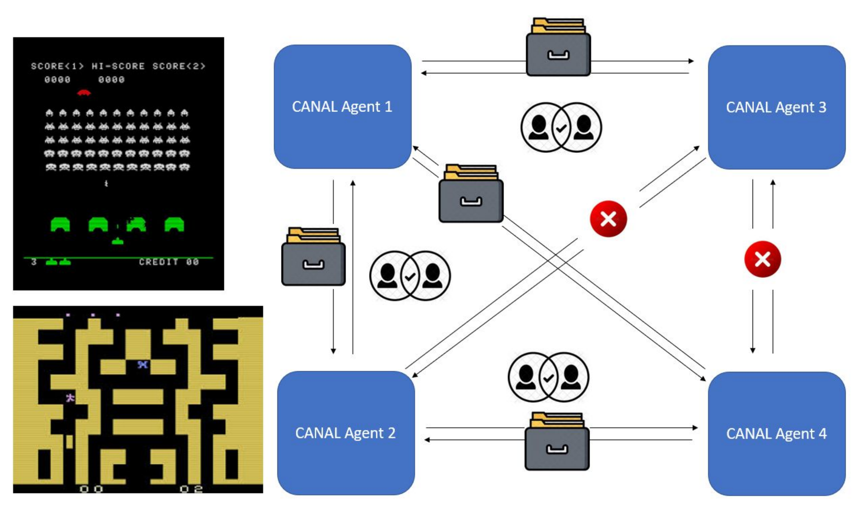


Figure 2: Use Case Demonstration of FLLEdge

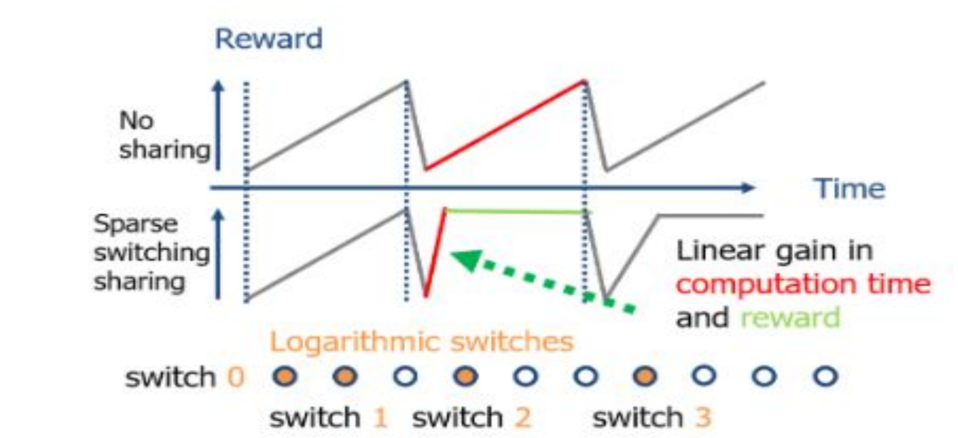


Figure 3: Comparison of Sparse Switching Sharing No Sharing

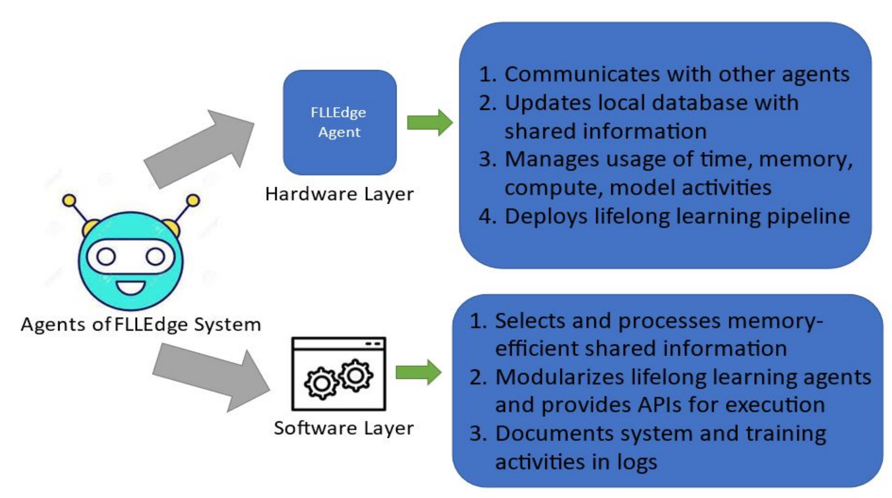
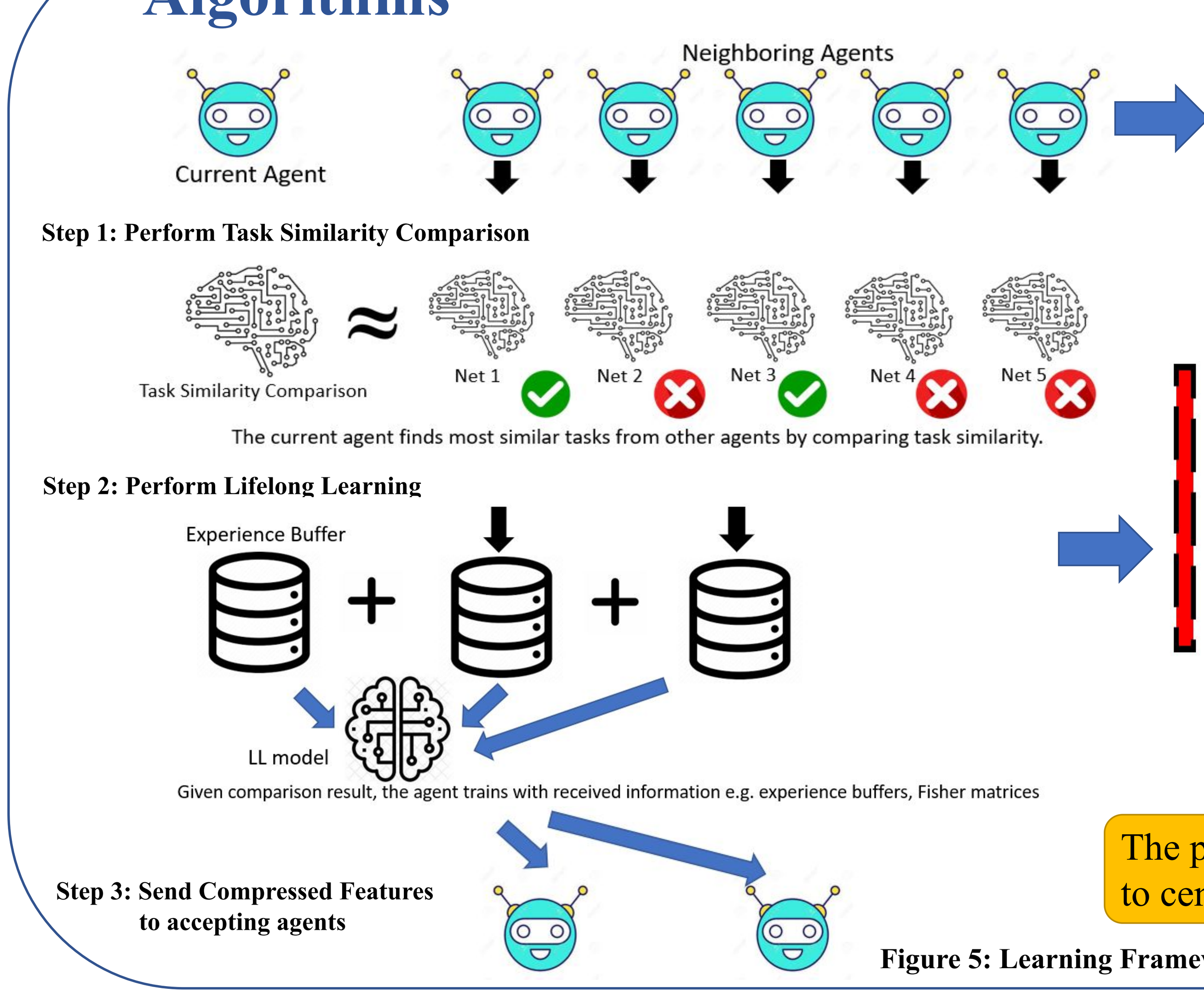


Figure 4: Responsibilities of Software and Hardware Layers

- Each FLLEdge agent is deployed on a **Jetson Nano**.
- Each FLLEdge agent trains on a unique sequence of Atari games.
- All agents maintain a distributed database for knowledge sharing.

Algorithms



Metrics of comparison include:

1. Euclidean distance difference of model parameters.
2. Proximity of model's output distribution given the same inputs.

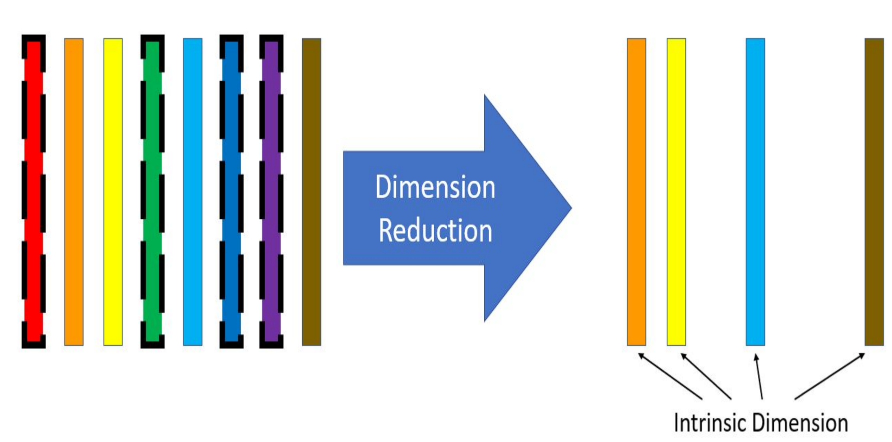


Figure 6: Compressive Linear Operators

The peer-to-peer topology can be easily converted to centralized topology for standard FL setting

Figure 5: Learning Framework

Experimental Results

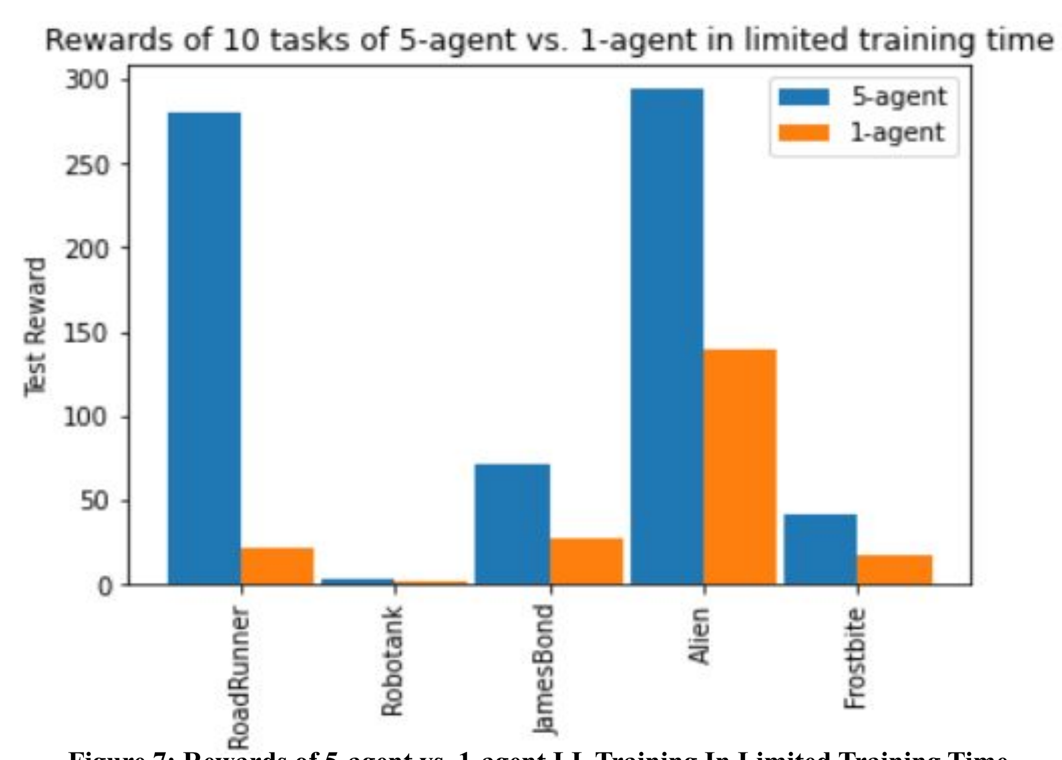


Figure 7: Rewards of 5-agent vs. 1-agent LL Training In Limited Training Time

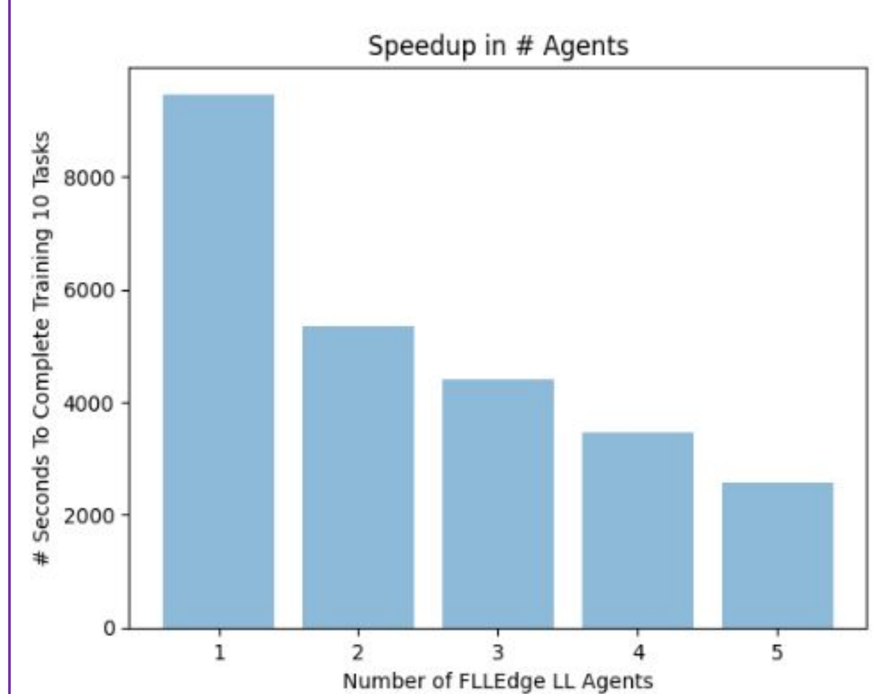


Figure 8: Scalability of Speedup In Number of Agents

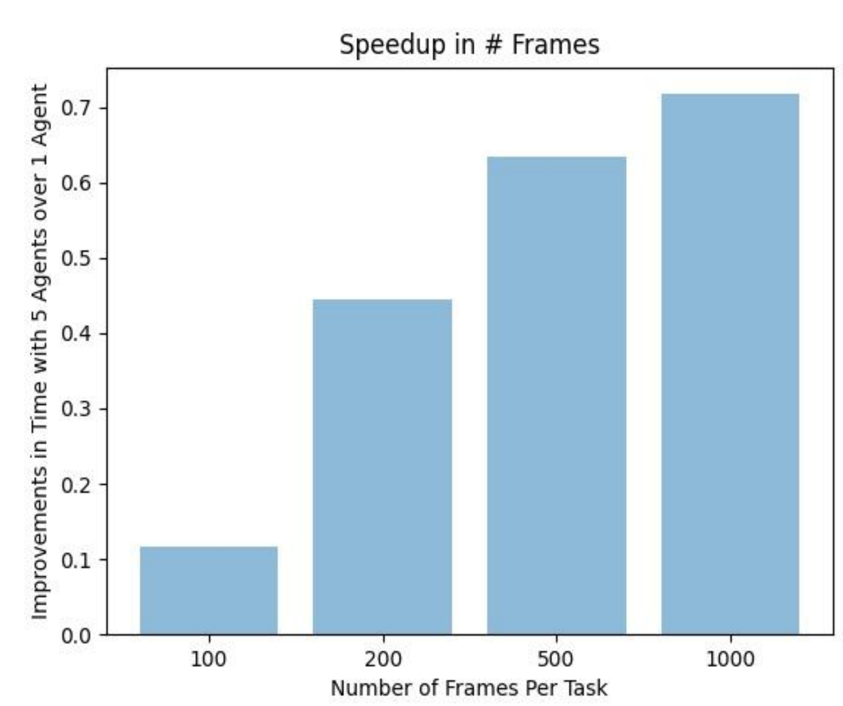


Figure 9: Scalability of Speedup In Number of Frames

	Round 1	Round 2	Rounds 3-10
Agent 1	873	872	<< 0.01 seconds per round
Agent 2	872	870	<< 0.01 seconds per round
Agent 3	871	870	<< 0.01 seconds per round
Agent 4	872	872	<< 0.01 seconds per round
Agent 5	872	872	<< 0.01 seconds per round

Figure 10: # seconds Spent on Training 10 tasks with 5 FLLEdge Agents

Each task is trained for 1000 frames. Until 2^{nd} round, all new tasks are seen. Hence, in all remaining rounds, each agent replays memory shared experienced buffers. Memory replay generally takes about 0.01 seconds to complete, while training from scratch (see rounds 1 and 2) takes about 870 seconds.

FLLEdge is part of ShELL project funded by DARPA. For more, please refer to https://tinyurl.com/4cnx8jnm