Federated Class-Incremental Learning: Hybrid Rehearsal with Latent Exemplars and Data-Free Techniques

Addressing Local and Global Forgetting

Milad Khademi Nori, Il-Min Kim, and Guanghui Wang

Department of Computer Science, Toronto Metropolitan University
Department of Electrical and Computer Engineering, Queen's University

March 24, 2025

Federated Class-Incremental Learning (FCIL)

FCIL enables collaborative learning on decentralized data without direct data sharing.

Crucial for privacy-sensitive applications.

Key Challenge: Local and Global Forgetting due to class imbalance.

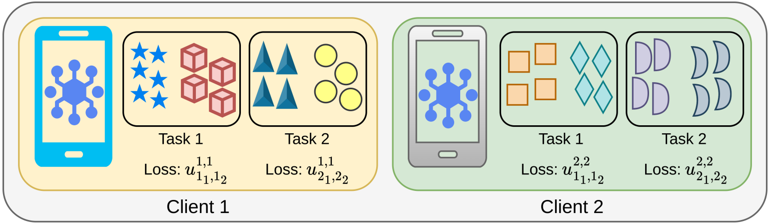


Figure: Local and global forgetting occur due to class imbalance at local (client-level) and global (system-wide) scales.

Local and Global Forgetting

Local Forgetting

- Occurs due to class imbalance within individual clients.
- Clients tend to forget previously learned classes as they learn new ones.
- Results in a biased model that performs poorly on older tasks.

Global Forgetting

- Occurs due to class imbalance across different clients in the federated network.
- Leads to inconsistent knowledge across the system; the global model doesn't represent all classes well.
- Hinders the overall performance and convergence of the federated learning process.

HR: Key Contributions

Theoretical Foundation: Mathematical framework proving HR addresses local & global forgetting.

Unified Approach: Combines latent exemplar replay and a data-free generative method.

Novel Autoencoder: Customized autoencoder & Lennard-Jones potential for efficient latent space management.

State-of-the-Art Results: Outperforms baselines; preserves privacy; minimizes memory.

Mathematical Framework

Formalizes loss update in FCIL:

$$l_{\boldsymbol{\theta}}^{(k+1)} = \underbrace{\sum_{m=1}^{M}\sum_{n=1}^{M}\left(l_{\boldsymbol{\theta}}^{(k),m,n} + \Delta l_{\boldsymbol{\theta}}^{(k+1),m,n}\right)}_{\text{Global Loss Update}}$$

Decomposes loss change:

$$\Delta l_{\boldsymbol{\theta}}^{(k+1),m,n} = \frac{1}{N^2}\left(\underbrace{\sum_{i=1}^{N}\sum_{j=1}^{N} u_{i_{k+1},j_{k+1}}^{m,n}}_{\text{New Task Loss}} + 2\underbrace{\sum_{l=1}^{k}\sum_{i=1}^{N}\sum_{j=1}^{N} u_{i_l,j_{k+1}}^{m,n}}_{\text{Intra-Client Confusion}}\right).$$

HR minimizes these terms to prevent forgetting (see paper for details).

Key Components:

Customized Autoencoder:

Encoder: $f(\boldsymbol{x}) \rightarrow \boldsymbol{z}$
Decoder: $g(\boldsymbol{z}) \rightarrow \boldsymbol{x}$

Lennard-Jones Potential:

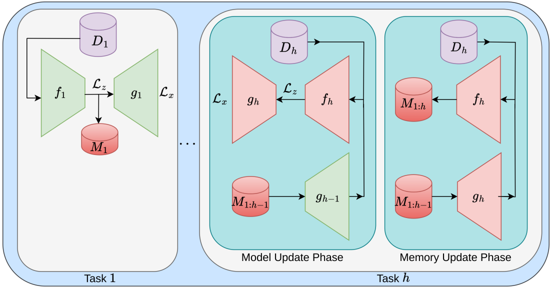
Ensures global consistency.

Latent Exemplar Replay:

Addresses local forgetting.

Data-Free Generation:

Addresses global forgetting.



HR Workflow

HR: Autoencoder and Loss

Customized Autoencoder:

Encoder: $f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}^m$ (maps input to latent space)

Decoder: $g(\boldsymbol{z}) : \mathbb{R}^m \to \mathbb{R}^n$ (reconstructs input from latent space)

Loss Function:

$$\boldsymbol{L} = \underbrace{-\mathbb{E}[\log p(\boldsymbol{x}|\boldsymbol{z})]}_{\text{Reconstruction}} + \underbrace{\mathrm{KL}(q(\boldsymbol{z}|\boldsymbol{x})||p(\boldsymbol{z}))}_{\text{Regularization}} + \underbrace{\lambda \sum ||\boldsymbol{z}_{i_j} - \boldsymbol{p}_{i_j}||^2}_{\text{Clustering}}$$

This loss combines reconstruction, regularization, and a clustering term to ensure a well-structured latent space.

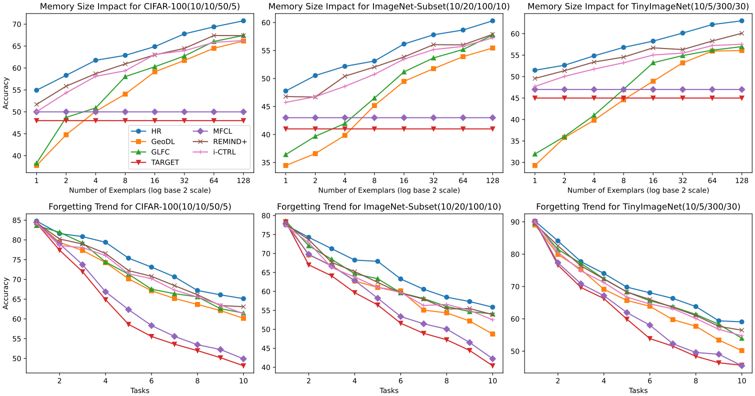
Comparison with Hybrid Approaches

Table: HR vs. Other Hybrid Approaches

Method	**Classif.**	**Learn.**	**Lat. Space**	**Arch.**
REMIND	After Dec.	Online	Unstruct.	Complex
REMIND+	After Dec.	Online	Unstruct.	Complex
i-CTRL	Latent	Offline	Struct.	Simple
HR	**Latent**	**Offline**	**Struct.**	**Simple**

HR combines latent space classification, structured latent space, and simple architecture.

Memory Impact and Forgetting Trend



Impact of memory size on final accuracy (top row of figure). Forgetting trends for HR and baselines (bottom row). Hybrid replay consistently outperforms model-based and data-based approaches.

Conclusion

HR provides a practical and effective solution for FCIL.

HR is proven to address local and global forgetting.

Combines the strengths of exemplar-based and data-free approaches.

Achieves state-of-the-art performance while preserving privacy and minimizing memory overhead.

Future Work: Explore alternative decoder architectures, optimize balance between efficiency, performance, and privacy.