# A Data-Free Approach to Mitigate Catastrophic Forgetting in Federated Class Incremental Learning for Vision Tasks

NEURAL INFORMATION PROCESSING SYSTEMS

Sara Babakniya[1], Zalan Fabian[1], Chaoyang He[2], Mahdi Soltanolkotabi[1], Salman Avestimehr[1]

[1]University of Southern California [2]FedML

USC University of Southern California

FedML

## Highlights

- In federated learning, clients' data can dynamically change.
- Continual learning targets this problem in centralized setting.
- Instead of relying on user's memory data, we propose using a generative model trained by the server in a data-free manner.
- Our method achieves good performance along with efficiency and privacy.

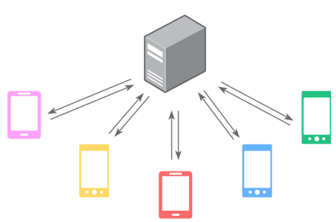
## Background

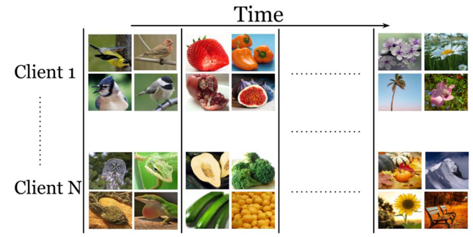
### Federated Learning

**Federated Learning**: Privacy preserving distributed training

**Common Challenges**:

- Privacy
- Efficiency
- Statistical Heterogeneity
- Resource Heterogeneity

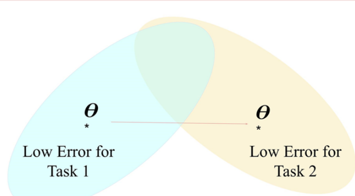





### Continual Learning & Forgetting

**Existing Solutions:**

- Episodic Memory
- GANs
- Regularization
- …

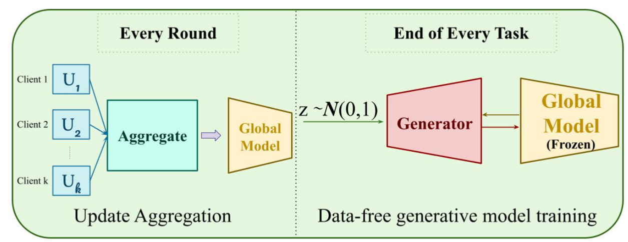


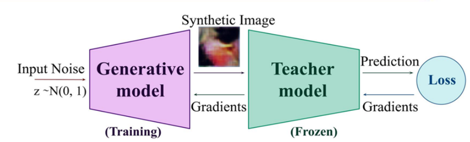

## Challenges

| | |
|---|---|
| Privacy | Cannot share local information |
| Client arrival | Should not only rely on memory |
| Client departure | Should not lose performance |
| Resource constraints | Should not add significant overhead on clients |

## Our Solution: MFCL

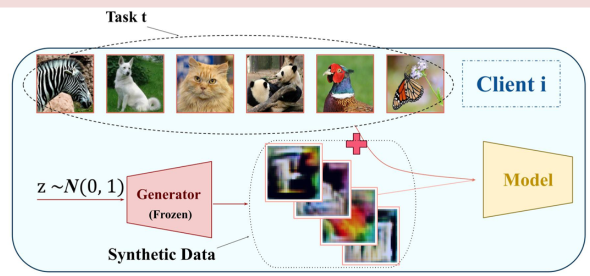
### Server Side




$$\min_{\mathcal{G}} \mathcal{L}_{CE} + w_{div}\mathcal{L}_{div} + w_{BN}\mathcal{L}_{BN} + w_{pr}\mathcal{L}_{pr}$$

### Client Side




$$\min_{\mathcal{F}_t} \mathcal{L}^t_{CE} + w_{FT}\mathcal{L}^t_{FT} + w_{KD}\mathcal{L}^t_{KD}$$

## SuperImageNet Dataset

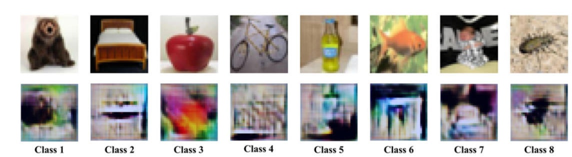
- The #clients in most works is between 5 – 20.
- We propose SuperImageNert; a regrouping of ImageNet tailored for FL with enough data to scale the #clients.

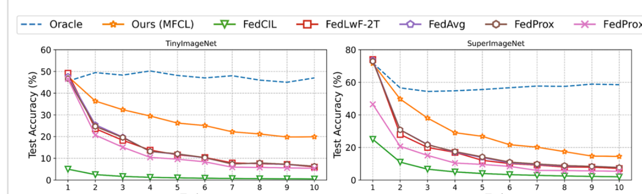

| Dataset | # examples/class | # classes |
|---|---|---|
| *SuperImageNet-S* | 2500 | 100 |
| *SuperImageNet-M* | 5000 | 75 |
| *SuperImageNet-L* | 7500 | 50 |

## Results

We focus on Class Incremental Learning where the tasks do not overlap and the model is evaluated on all the classes observed so far.

| | Average Accuracy $\tilde{A}$ (%) | Average forgetting $\tilde{f}(\%)$ | Training time (s) $(T > 1)$ | Server Runtime (s) |
|---|---|---|---|---|
| **FedAvg** | $22.27 \pm 0.22$ | $78.77 \pm 0.83$ | $\approx 1.2$ | $\approx 1.8$ |
| **FedProx** | $22.00 \pm 0.31$ | $78.17 \pm 0.33$ | $\approx 1.98$ | $\approx 1.8$ |
| **FedCIL** | $26.8 \pm 0.44$ | $38.19 \pm 0.31$ | $\approx 24.5$ | $\approx 2.5$ for $T = 1$, $\approx 4.55$ for $T > 1$ |
| **FedLwF-2T** | $22.17 \pm 0.13$ | $75.08 \pm 0.72$ | $\approx 3.4$ | $\approx 1.8$ |
| **MFCL (Ours)** | $\mathbf{44.98 \pm 0.12}$ | $\mathbf{28.3 \pm 0.78}$ | $\approx 3.7$ | $\approx 330$ (once per task), $\approx 1.8$ O.W. |
| **Oracle** | $67.12 \pm 0.4$ | −− | $\approx 1.2 \times T$ | $\approx 1.8$ |




Scan for the paper →
Contact: babakniy@usc.edu