MORE EXPERIMENT DETAILS ON EFFACE

Boxu Xiao[1] *Sijia Liu*[2] *Qing Ling*[1]

[1] School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China
[2] Department of Computer Science and Engineering, Michigan State University, East Lansing, USA

1. EXPERIMENT SETUP DETAILS

Datasets, Models, and Federated Frameworks. We test on 5 datasets in image classification: MNIST, Fashion-MNIST, SVHN, CIFAR-10 and CIFAR-100 [1–4]. We set 50 clients to smaller datasets (MNIST and Fashion-MNIST), and 50 clients to the others for scalability evaluation across client count and dataset complexity. We use CNN with 2 convolutional and 2 fully connected layers for MNIST and Fashion-MNIST, ResNet-18 for SVHN and CIFAR-10, and ResNet-34 for CIFAR-100 [5]. We adpot non-i.i.d. partition following the Dirichlet distribution with $\alpha = 0.5$.
Baselines. We select 8 baselines, including:

- FedRetrain: retrain from scratch on $\mathcal{D}^r$ using FedAvg [6];
- FedGD [7]: run federated gradient descent on $\mathcal{D}^r$;
- FedGA [8]: run federated gradient ascent on $\mathcal{D}^u$;
- FedPSGA [9]: It is originally designed for unlearning all data samples of one client. The target client i uses the average of the other clients' local models as the reference model, and runs a projected stochastic gradient ascent algorithm on $\mathcal{D}_i^u$ onto an ℓ_2-norm ball of the reference model. To fix it into a partial sample removel task, we correct it that the unlearning clients use the original model θ^0 as the reference model and run a federated projected stochastic gradient ascent algorithm on $\mathcal{D}^u$;
- FedGB and FedUOSC [10]: FedGB uses federated gradient balancing with ascent on the unlearning dataset and descent on the remaining dataset. FedUOSC based on it applies oriented saliency compression, allowing the clients to upload half of the salient parameters by default.
- NoT [11]: negate the weights of the first few layers of the model and then apply standard federated training on $\mathcal{D}^r$.
- FedRL [12]: set the labels of the unlearning dataset as random labels and train on the $\mathcal{D}^{u\prime}$ and $\mathcal{D}^r$.

Some unlearning methods also need several rounds of refinement that trains on the remaining dataset for fine-tuning, such as FedGA, FedPSGA, FedGB and FedUOSC.

2. MORE EXPERIMENT RESULTS

More results on Fashion-MNIST and SVHN. Tables 1 and 2 show the details of results on Fashion-MNIST and SVHN. Our method, based on FedRL, matches their unlearning quality and utility and achieving superior communication efficiency under Top-1 compression.

Table 1. Unlearning performance on Fashion-MNIST.

Methods	CA (↑)	BA (↓)	ET (↓)	CC (↓)
FedRetrain	91.72(0.12)	1.65(0.16)	2062.94	7893.45
FedGD	91.60(0.17)	96.50(0.05)	2052.01	7893.45
FedGA	77.50(0.70)	0.57(0.84)	**110.25**	1184.02
FedPSGA	86.73(0.61)	6.36(1.28)	125.82	1223.39
FedGB	89.89(0.13)	2.31(1.69)	208.50	1933.89
FedUOSC	89.94(0.30)	2.50(1.99)	209.39	966.95
NoT	88.61(0.52)	17.85(5.20)	2063.69	7893.45
FedRL	**91.47**(0.20)	**0.13**(0.02)	810.05	2682.79
EFFACE	90.39(0.14)	2.07(1.82)	807.74	**79.02**

Table 2. Unlearning performance on SVHN.

Methods	CA (↑)	BA (↓)	ET (↓)	CC (↓)
FedRetrain	96.25(0.08)	0.41(0.06)	6920.31	85250.56
FedGD	96.40(0.08)	99.00(0.15)	6837.16	34100.23
FedGA	94.52(0.68)	1.63(0.33)	324.36	12787.58
FedPSGA	95.78(0.31)	1.88(0.58)	345.72	10656.33
FedGB	96.26(0.10)	0.10(0.08)	271.89	9235.48
FedUOSC	96.08(0.41)	0.76(0.90)	**270.88**	4617.74
NoT	95.92(0.05)	2.31(3.14)	6853.64	34100.23
FedRL	95.22(0.35)	**0.01**(0.01)	953.34	9362.75
EFFACE	**96.36**(0.04)	0.13(0.02)	585.78	**425.92**

3. REFERENCES

[1] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, 1998.

[2] Han Xiao, Kashif Rasul, and Roland Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.

[3] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Baolin Wu, Andrew Y Ng, et al., "Reading digits in natural images with unsupervised feature learning," in *NIPS workshop on deep learning and unsupervised feature learning*, 2011.

[4] A Krizhevsky, "Learning multiple layers of features from tiny images," *Master's Thesis, Department of Computer Science, University of Toronto*, 2009.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Computer Vision and Pattern Recognition*, 2016.

[6] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas, "Communica-tion-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*, 2017.

[7] Alessio Mora, Carlo Mazzocca, Rebecca Montanari, and Paolo Bellavista, "Federated unlearning in healthcare: Why it matters," in *International Joint Conference on Neural Networks*, 2025.

[8] Leijie Wu, Song Guo, Junxiao Wang, Zicong Hong, Jie Zhang, and Yaohong Ding, "Federated unlearning: Guarantee the right of clients to forget," *IEEE Network*, 2022.

[9] Anisa Halimi, Swanand Kadhe, Ambrish Rawat, and Nathalie Baracaldo, "Federated unlearning: How to efficiently erase a client in FL?," *arXiv preprint arXiv: 2207.05521*, 2022.

[10] Boxu Xiao, Sijia Liu, and Qing Ling, "Federated unlearning with oriented saliency compression," in *International Joint Conference on Neural Networks*, 2025.

[11] Yasser H Khalil, Leo Brunswic, Soufiane Lamghari, Xu Li, Mahdi Beitollahi, and Xi Chen, "NoT: Federated unlearning via weight negation," in *Computer Vision and Pattern Recognition Conference*, 2025.

[12] Chongyu Fan, Jiancheng Liu, Yihua Zhang, Dennis Wei, Eric Wong, and Sijia Liu, "Salun: Empowering machine unlearning via gradient-based weight saliency in both image classification and generation," in *International Conference on Learning Representations*, 2024.