Sony AI — FedWon: Triumphing Multi-domain Federated Learning Without Normalization
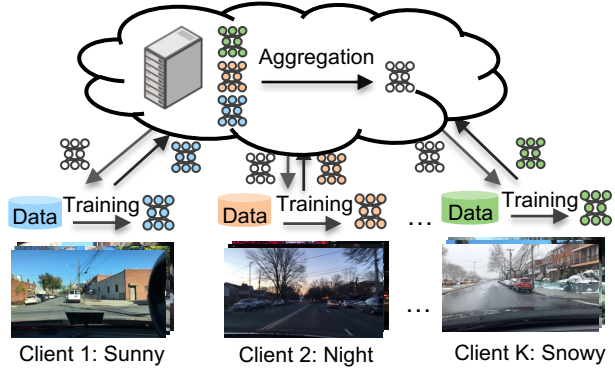
Weiming Zhuang, Lingjuan Lyu

ICLR

Motivation

Multi-domain Federated Learning (FL): each FL client contains data of one domain.

Applications: e.g., autonomous driving cars in different locations capture images in varying weather conditions.

Client 1: Sunny   Client 2: Night   Client K: Snowy

Challenge: clients trained on data with different domains capture different feature statistics in Batch Normalization (BN) layers.

4-th BN Layer
Client 0   Client 1
Running Variance
Running Mean
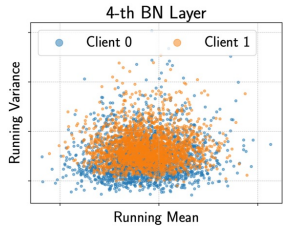
Contribution

The 1st FL method without normalizations.

Significance: (1) Superior performance in multi-domain FL; (2) Versatile for both cross-silo and cross-device FL; (3) Strong performance even with batch size as small as 1; (4) Adaptable to skewed label distribution challenge.

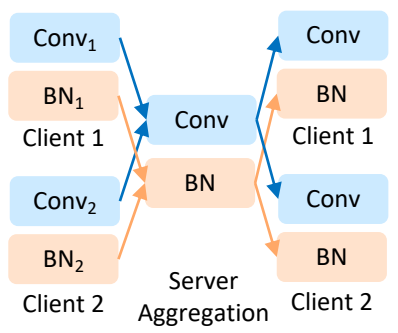## Federated Learning Without Normalization (FedWon)

Objective

$$\min_{\theta} f(\theta) := \sum_{k=1}^{K} p_k \mathrm{E}_{\xi_k \sim \mathcal{D}_k}[f_k(\theta; \xi_k)]$$
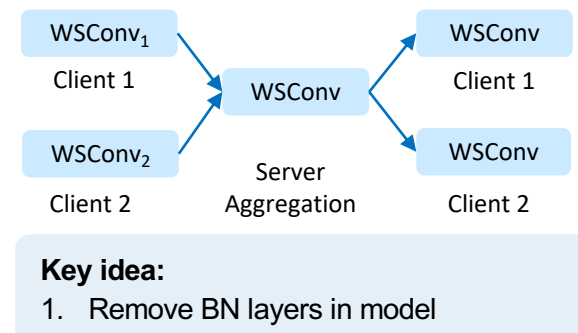
Problem Setup

Clients possess data from various domains and each client only owns data samples from a single domain:

- $\mathcal{P}_k(x) \not\sim \mathcal{P}_{k'}(x)$ for different clients k and k'.
- $\mathcal{P}_k(x_i) \sim \mathcal{P}_k(x_j)$ for $\forall x_i, x_j \in \left\{(x_i^k, y_i^k)\right\}_{\{i=1\}}^{n_k}$ all data samples of a client k.

FedAvg

FedWon (Ours)

Key idea:
1. Remove BN layers in model
2. Reparametrize Conv layers with scaled weight standardization (WSConv)

Scaled Weight Standardization (WSConv)

Goal: approximate zero mean and unit variance for outputs

Weight matrix of a layer

$$\widehat{W}_{i,j} = \gamma \cdot \frac{W_{i,j} - \mu_i}{\sigma_i \sqrt{N}}$$

Output channel   Input channel   Constant   Weight normalization   → mean   → Input channel size

Evaluation

Cross-silo FL: DomainNet dataset with 6 clients

Methods	C	I	P	Q	R	S	Avg.
Standalone	42.7	24.0	34.2	**71.6**	51.2	33.5	42.9
FedAvg	48.9	26.5	37.7	44.5	46.8	35.7	40.0
FedProx	51.1	24.1	37.3	46.1	45.5	37.5	40.2
FedAvg+GN	45.4	21.1	35.4	57.2	50.7	36.5	41.1
FedAvg+LN	42.7	23.6	35.3	46.0	43.9	28.9	36.7
SiloBN	51.8	25.0	36.4	45.9	47.7	38.0	40.8
FixBN	49.2	24.5	38.2	46.3	46.2	37.4	40.3
FedBN	49.9	28.1	40.4	69.0	55.2	38.2	46.8
FedWon	**57.2**	**28.1**	**43.7**	69.2	**56.5**	**51.9**	**51.1**

Cross-device FL: Digits-Five dataset with 100 clients, a fraction C clients participate in training each round.

C	Method	M	S	U	Syn	M-M	Avg.
10%	FedAvg	98.2	81.0	97.2	91.6	89.3	91.5
	FedWon	**98.6**	**85.4**	**98.3**	**93.6**	**90.5**	**93.3**
40%	FedAvg	98.1	80.5	97.0	91.4	89.4	91.3
	FedWon	**98.8**	**86.4**	**98.4**	**94.2**	**91.0**	**93.7**

Small Batch Size B: Office-Caltech-10 dataset, 4 clients

B	Methods	A	C	D	W
1	FedAvg+GN	60.4	52.0	87.5	84.8
	FedAvg+LN	55.7	43.1	84.4	88.1
	FedWon	**66.7**	**55.1**	**96.9**	**89.8**
2	FedAvg	64.1	49.3	87.5	89.8
	FedAvg+GN	63.5	52.0	81.3	84.8
	FedAvg+LN	58.3	44.9	87.5	86.4
	FixBN	66.2	50.7	87.5	88.1
	SiloBN	61.5	47.1	87.5	86.4
	FedBN	59.4	48.0	96.9	86.4
	FedWon	**66.2**	**54.7**	**93.8**	**89.8**

Code & Paper
weiming.me