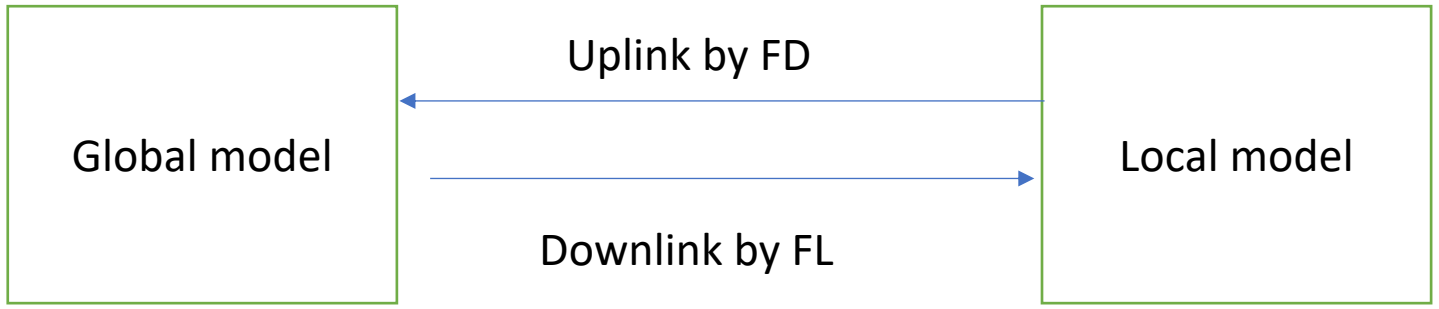
Mix2FLD: Downlink Federated Learning After Uplink Federated Distillation With Two-Way Mixup

Problem Addressed

- communication efficiency of FL is problematic in deep neural network models (DNNs).



Solution

- Communication issue occurs due to:

Size of DNN increases ->play load increases->uplink latency ->asymmetric uplink downlink channels

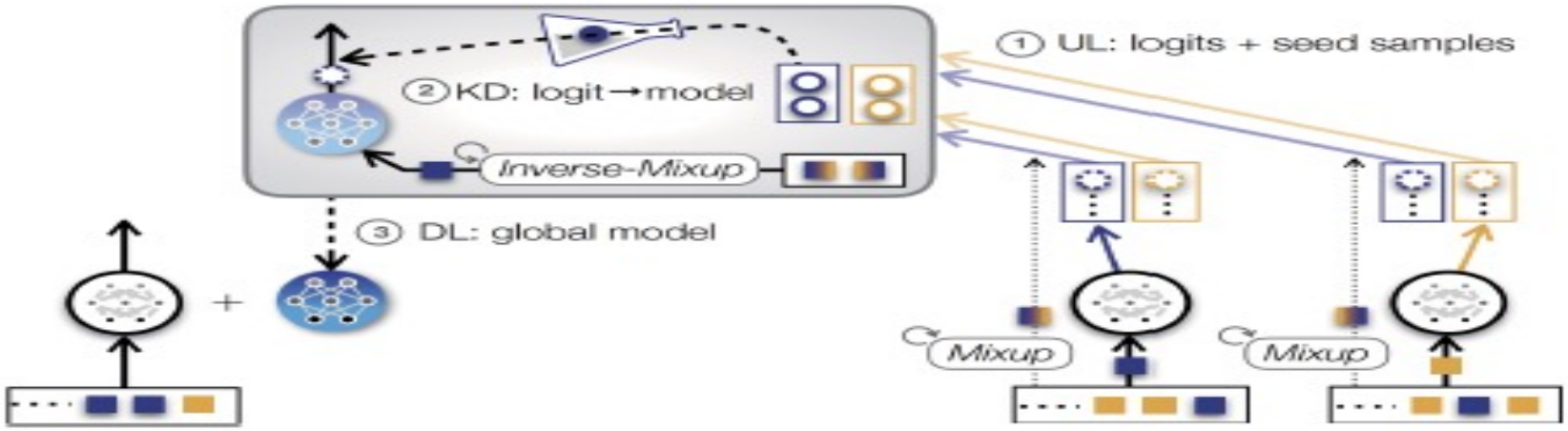
- Solution : Federated Distillation (FD)

It is independent of the model size ,communication payload sizes of FD are fixed as the model output dimension (e.g., 10 labels in MNIST)

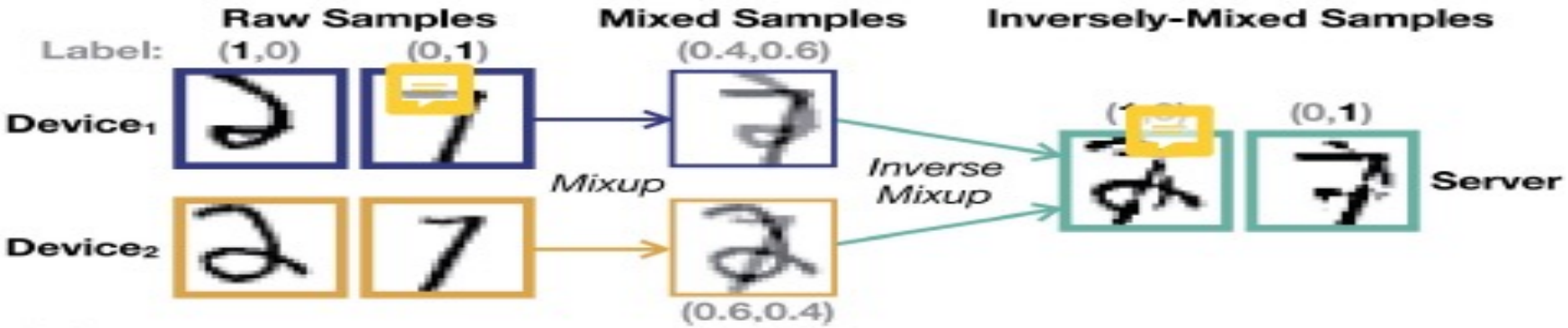
- Drawback of FD: Accuracy low

Solution To Improve Accuracy Of FD

- Addition of synthetic data so that privacy is also preserved ->**Mix2FLD**



(a) **Mix2FLD**: downlink federated learning (FL) & uplink federated distillation (FD) with two-way Mixup (Mix2up) seed sample collection.



(b) **Mix2up**: mixing raw samples at devices & inversely mixing them across different devices at the server (mixing ratio $\lambda = 0.4$).

System Settings

- D_1 , D_2 , $D_{3...}$ D_n devices
- each device in range (1 to n) has a local model M_i
- each device in range (1 to n) has a private dataset S_i
- Each sample is has unlabeled X_i and the ground truth label l_i.
- Classification problem
- Loss function is cross entropy

FLOWCHART

- FD converts the local output vector(obtain using SGD for K iterations) to global average output vectors

- KD convert the Global output vector to Global model parameter

use

- Mix2FLD creates synthetic data

- FL downloads the global model

Federated Distillation (FD)

INPUT

Local average output vectors for each ground truth label .

PROCESS

- Create GlobalaverageOutputVector()

(The size of the output of is equal to the number of labels in the classification problem.)

-DownloadGlobalAverageOutputVector() by each device .

- UpdateLocalweight using KD

-DistillationRegularization()

Measures the gap between average output vector and global output vector
(If this knowledge gap is negligible, the device's)

- weight is updated based on its own prediction
- Otherwise perturbed proportionally to the gap.

OUTPUT

Global model output.

Knowledge Distillation (KD)

Input

- global output vector
- for each device upload seed samples (from its own local dataset)

Process

-Run SGD

-UpdateGlobalModelWeight

Output

Global model parameter

Federated Learning (FL)

Input

local data of device Di

Process

- updateLocalWeights (through K iterations of SGD)
    - CalculateEntrophyLoss
    - GDtoUpdateWeight

Output

Normalised averaged local output vector for each label

- any device can upload the weight vectors at n <K

Check List

- Data heterogeneity : true
- System heterogeneity : no
- Model accuracy : measure using test accuracy on test data .
- Use of synthetic data : yes , for maintaining privacy .
- Data transfer : yes for improving accuracy