JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Federated Learning with Data-Agnostic Distribution Fusion

Jian-hui Duan, Wenzhong Li*, Derun Zou, Ruichen Li, Sanglu Lu

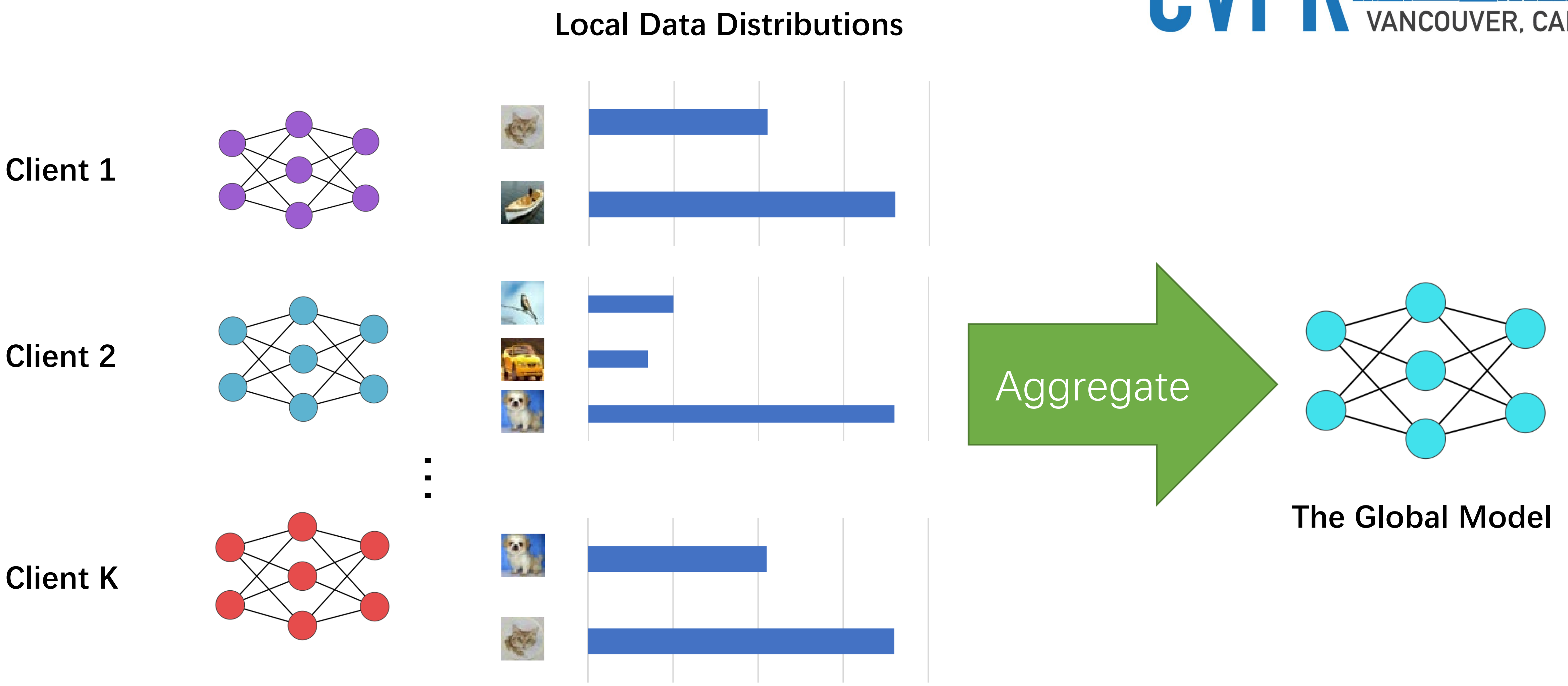
State Key Laboratory for Novel Software Technology, Nanjing University

Nanjing, China

TUE-PM-378

Introduction

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

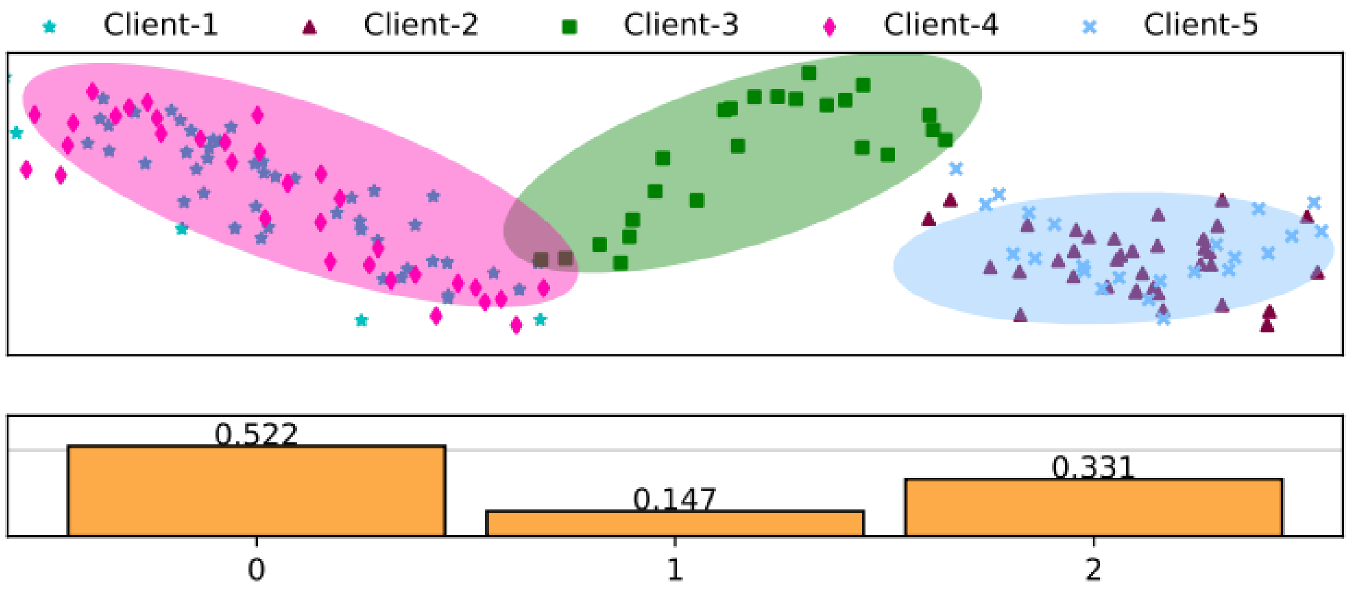


Data distribution heterogeneity in federated learning usually cause accuracy drop in global model

Introduction

- Data distribution of five clients can be represented by a distribution fusion model with three virtual components.

- Client models can be aggregated based on components to better approach centralized training.



Problem Definition

First step is to modify the optimizing target:

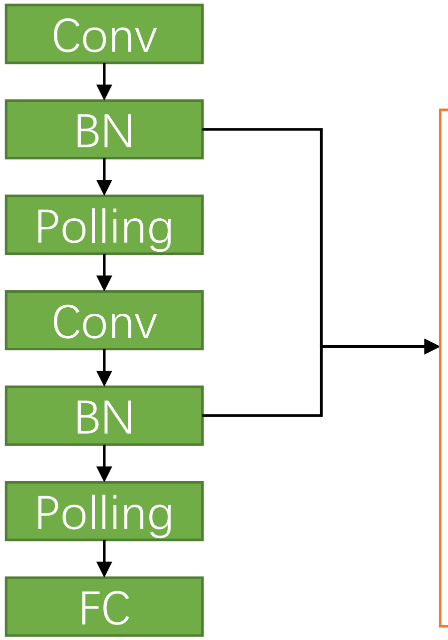
reallocate local models with weight b_{km} to m component, then aggregate component with weight π_m.

$$\min_{\mathbf{w}} \mathcal{L}(\mathbf{w}) = \sum_{k=1}^{K} \frac{n_k}{n} \mathcal{L}_k(\mathbf{w})$$

$$\min_{\mathbf{w}} \mathcal{L}(\mathbf{w}) = \underbrace{\sum_{m=1}^{M} \pi_m}_{\text{Virtual Component Fusion Weight}} \underbrace{\sum_{k=1}^{K} b_{km}}_{\text{Virtual Component Allocation Weight}} \mathcal{L}_k(\mathbf{w})$$

Variational AutoEncoder

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA



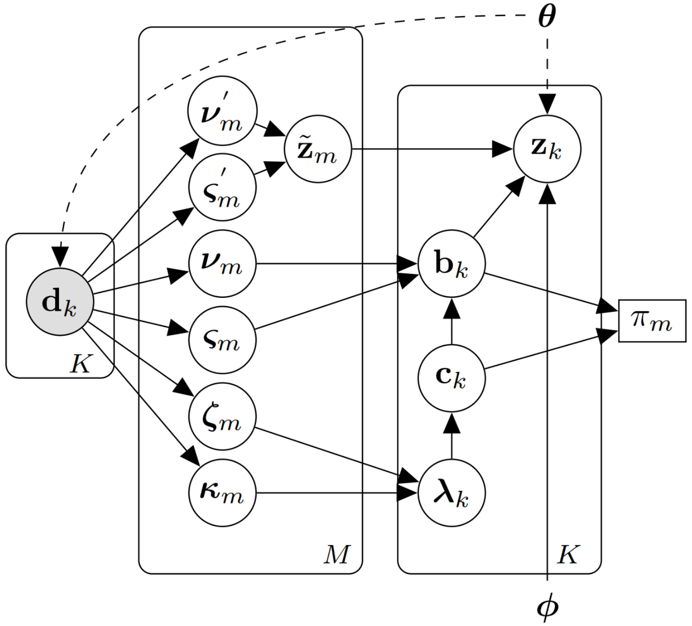
General CNN Architecture

Statistical Information Of Feature Maps On Client k

$$\mathrm{d}_k = \{\hat{\mu}_k, \hat{\sigma}_k, \hat{\beta}_k, \hat{\gamma}_k\}$$

- $\hat{\mu}_k$:means
- $\hat{\sigma}_k$:standard deviations
- $\hat{\beta}_k$:shift means
- $\hat{\gamma}_k$:scaled standard deviations

Input

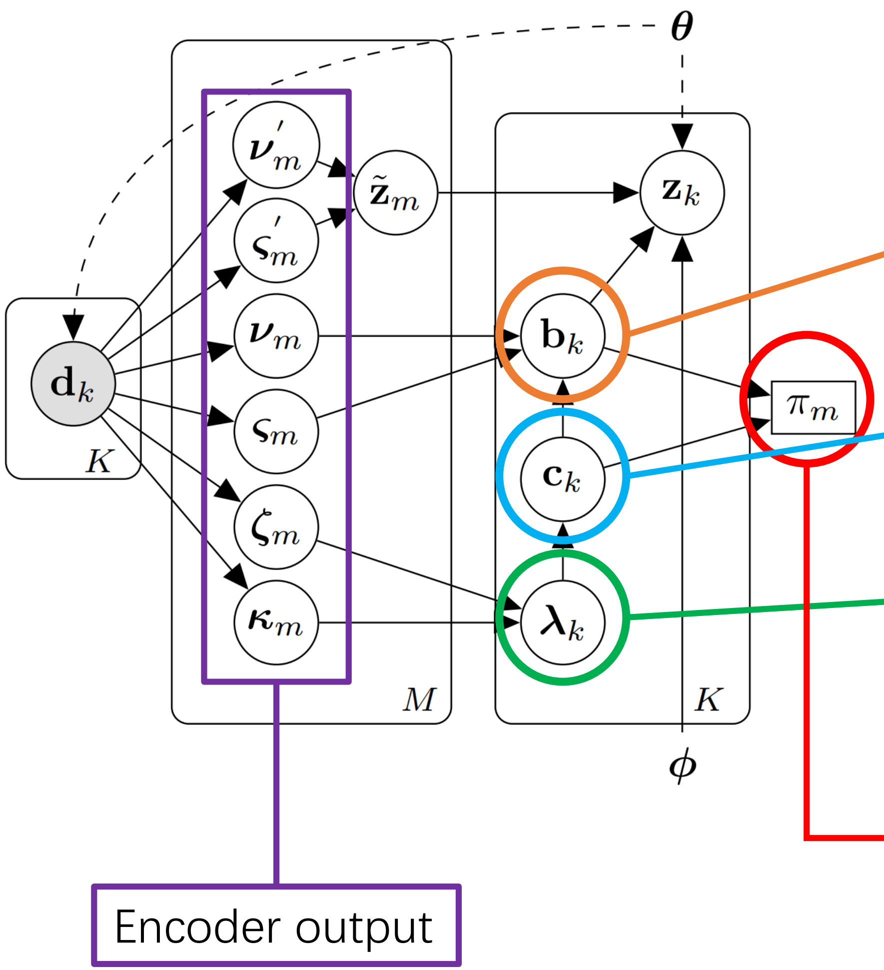


Variational AutoEncoder

Optimal aggregation weights require accurate local data distributions, we gather parameters of normalization layers, and use Variational AutoEncoder to infer local distribution parameters, in order to construct local data distributions.

Variational AutoEncoder

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA



$$\mathbf{b}_k \sim \mathcal{N}(\boldsymbol{\nu}_m, \varsigma_m)$$

$$\mathbf{c}_k \sim Bernoulli(\prod_{m=1}^{M} \lambda_{km})$$

$$\boldsymbol{\lambda}_k \overset{i.i.d.}{\sim} Beta(\boldsymbol{\zeta}_m, \boldsymbol{\kappa}_m)$$

$$\pi_m = \frac{\exp(\frac{1}{K}\sum_{k=1}^{K} q_{\phi}(c_{km}) \cdot b_{km})}{\sum_{m=1}^{M} \exp(\frac{1}{K}\sum_{k=1}^{K} q_{\phi}(c_{km}) \cdot b_{km})}$$

Optimizing Procedure

To better optimize distribution parameters, we design following sampling methods:

$$\boldsymbol{\lambda}_k \overset{i.i.d.}{\sim} Beta(\boldsymbol{\zeta}_m, \boldsymbol{\kappa}_m)$$

Sampling

$$\boldsymbol{\lambda}_k \sim (1 - \xi^{\frac{1}{\boldsymbol{\kappa}_k}})^{\frac{1}{\boldsymbol{\zeta}_k}}, \text{ where } \xi \sim Uniform(0,1)$$

$$\mathbf{c}_k \sim Bernoulli(\prod_{m=1}^{M} \lambda_{km})$$

Sampling

$$\mathbf{c}_{km} = arg\max_i (g_i + \log \prod_{i=1}^{2} \boldsymbol{\lambda}_{ki}) \text{ Where } g_i \sim Gumbel(0,1)$$

Experiment

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Datasets:

Datasets	Data Type	Train	Test	Total
MNIST[25]	1 channel image	60,000	10,000	70,000
Fashion-MNIST[44]	1 channel image	60,000	10,000	70,000
CIFAR-10[22]	3 channel image	50,000	10,000	60,000
Sentiment140[8]	Text data	-	-	1,600,000

BackBone Models：

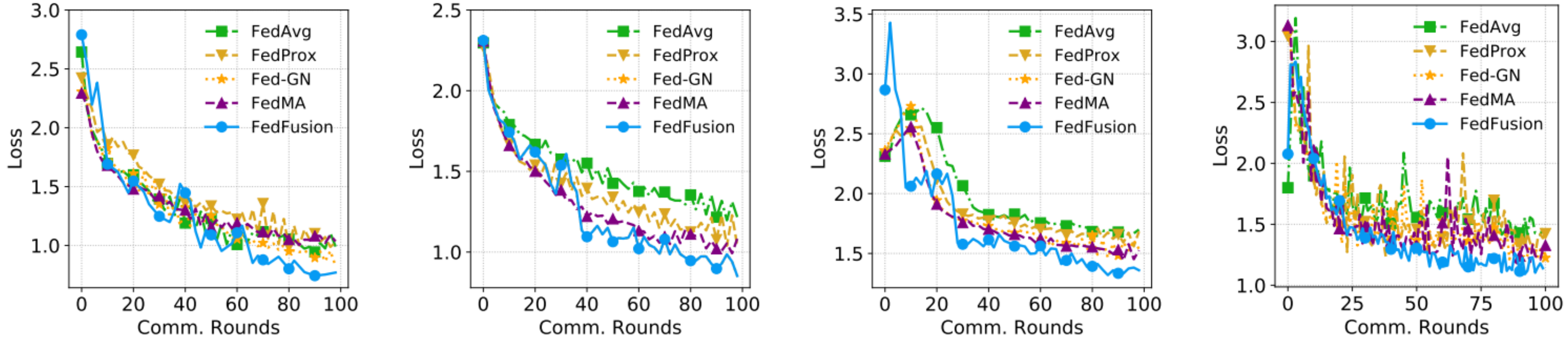
ResNet18[9], DenseNet121[11], MobileNetV2[36], LeNet[24], BiLSTM[]

Benchmarks:

Single-model: FedAvg[30], FedProx[26], Fed-GN[10], FedMA[43]

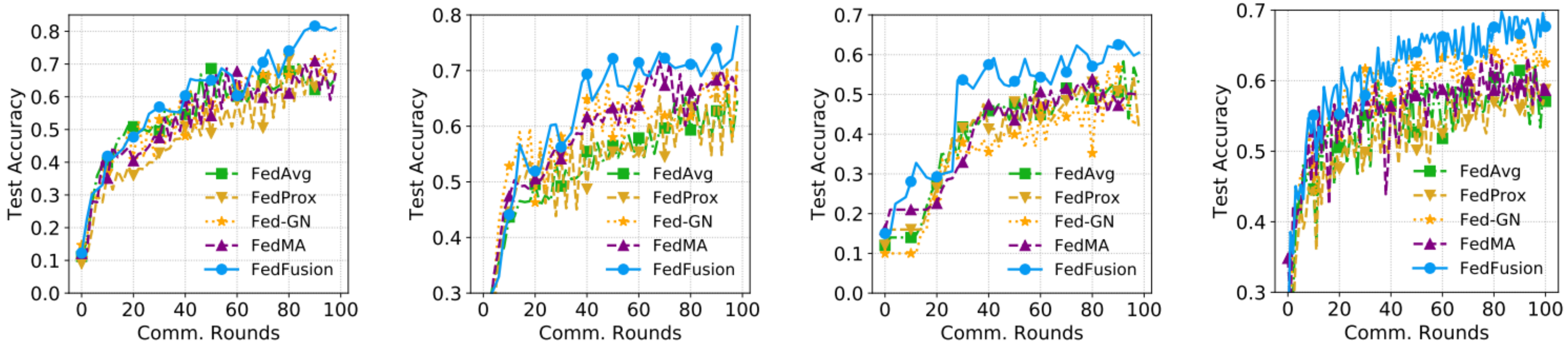
Multi-model: FeSEM[46], IFCA[7], FedCluster[2], FedGroup[6]

Experiment



(a) ResNet18 on CIFAR10 (b) DenseNet121 on CIFAR10 (c) MobileNetV2 on CIFAR10 (d) BiLSTM on Sent140

Figure 3. Training loss of different algorithms.



(a) ResNet18 on CIFAR-10 (b) DenseNet121 on CIFAR-10 (c) MobileNetV2 on CIFAR-10 (d) BiLSTM on Sent140

Figure 4. Training efficiency of different algorithms.

FedFusion(blue line) shows the lowest loss, and converges the fastest among all evaluated algorithms

Experiment

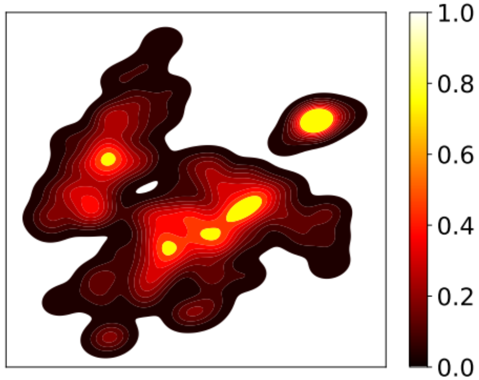
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

	Dataset	CIFAR-10			FMNIST	MNIST	Sent140
	Model	ResNet18	DenseNet121	MobileNetV2	LeNet	LeNet	BiLSTM
Single-model	FedAvg	68.78 (±0.89)	63.33 (±0.67)	54.69 (±3.92)	79.20 (±1.15)	97.32 (±0.04)	58.33 (±2.03)
	FedProx	70.18 (±0.45)	66.85 (±0.93)	55.03 (±2.77)	80.03 (±0.98)	97.55 (±0.02)	59.73 (±1.38)
	Fed-GN	72.57 (±0.78)	70.02 (±1.36)	56.43 (±1.92)	81.11 (±0.74)	97.88 (±0.02)	63.41 (±1.94)
	FedMA	73.43 (±1.03)	70.13 (±1.71)	59.61 (±2.01)	81.02 (±1.35)	98.06 (±0.03)	60.86 (±2.42)
Multi-model	FeSEM	67.78 (±2.58)	62.65 (±0.82)	53.82 (±3.69)	78.18 (±1.45)	96.24 (±0.17)	59.57 (±3.41)
	IFCA	73.04 (±1.45)	70.85 (±2.03)	58.93 (±2.45)	80.82 (±1.29)	97.09 (±0.11)	60.82 (±2.74)
	FedCluster	72.57 (±0.78)	68.77 (±1.38)	58.18 (±1.22)	79.11 (±0.74)	97.88 (±0.02)	63.41 (±1.94)
	FedGroup	74.38 (±1.92)	71.63 (±0.74)	59.86 (±2.09)	81.32 (±2.07)	97.37 (±0.61)	63.61 (±3.26)
	FedFusion	**81.26** (±0.82)	**75.92** (±1.25)	**62.88** (±1.21)	**83.16** (±0.74)	**98.49** (±0.04)	**67.51** (±1.71)

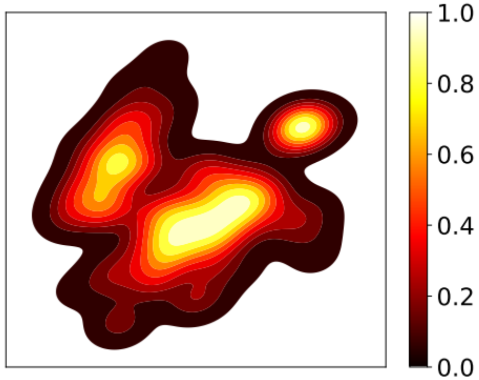
Comparison of average test accuracy on non-IID datasets

Experiment

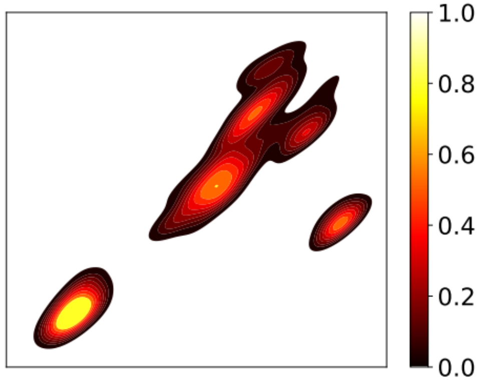
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA



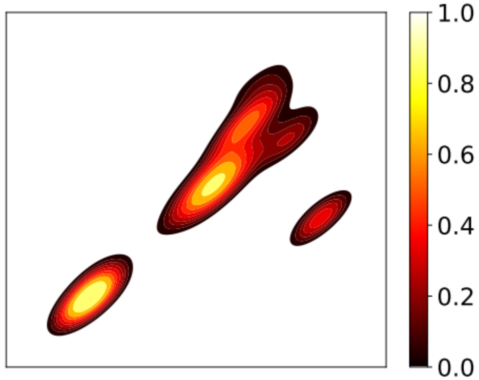
Original Data Distribution Of MNIST Dataset



Data Distribution Of MNIST Dataset Inferred By FedFusion



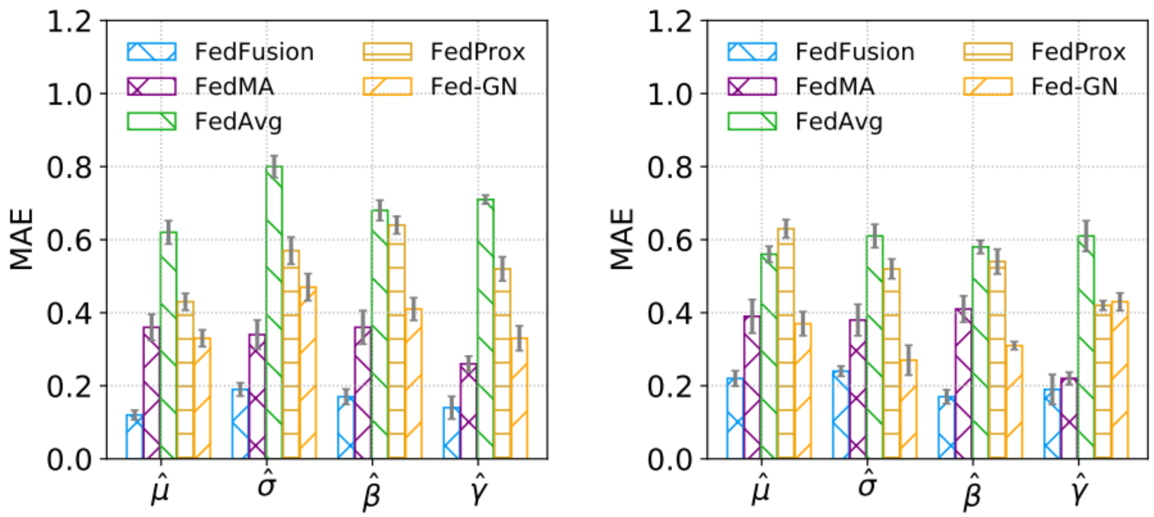
Original Data Distribution Of CIFAR-10 Dataset



Data Distribution Of CIFAR-10 Dataset Inferred by FedFusion

FedFusion accurately infer and reconstruct global data distribution, gives Fedfusion ability to approach centralized training.

Experiment



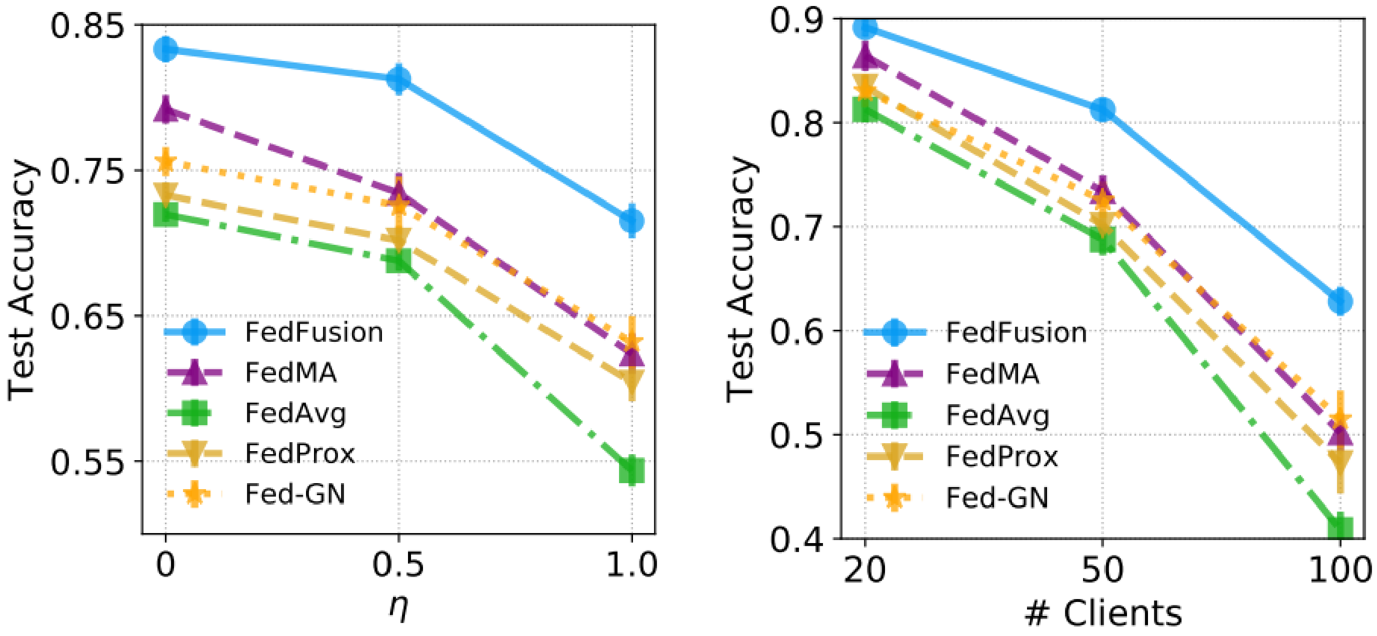
(a) ResNet18 on CIFAR-10 (b) BiLSTM on Sent140

Comparison of feature distribution bias

Normalization layer in models trained by FedFusion has fewer bias compared with centralized trained model, also shows FedFusion approximate centralized training well.

Experiment

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA



With different hyper-parameter settings, FedFusion shows better robustness.

Thank You!