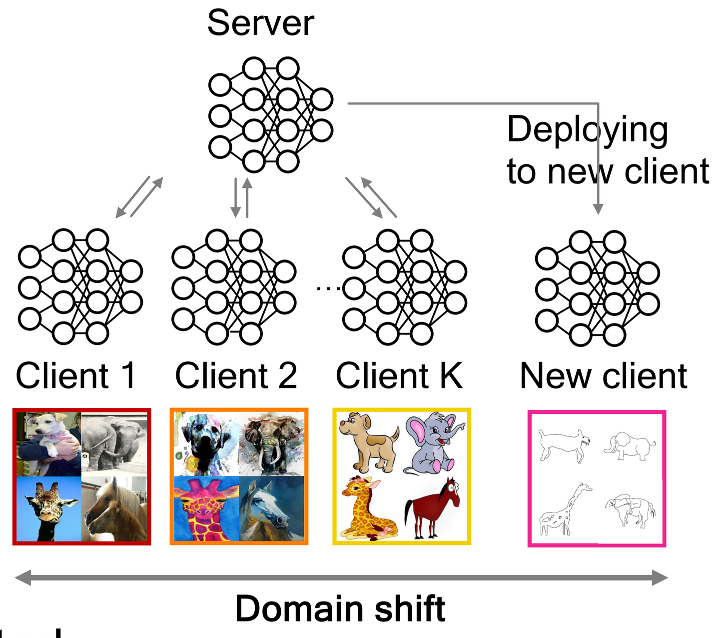
Feature Diversification and Adaptation for Federated Domain Generalization

Seunghan Yang, Seokeon Choi, Hyunsin Park, Sungha Choi, Simyung Chang, Sungrack Yun
Qualcomm AI Research*

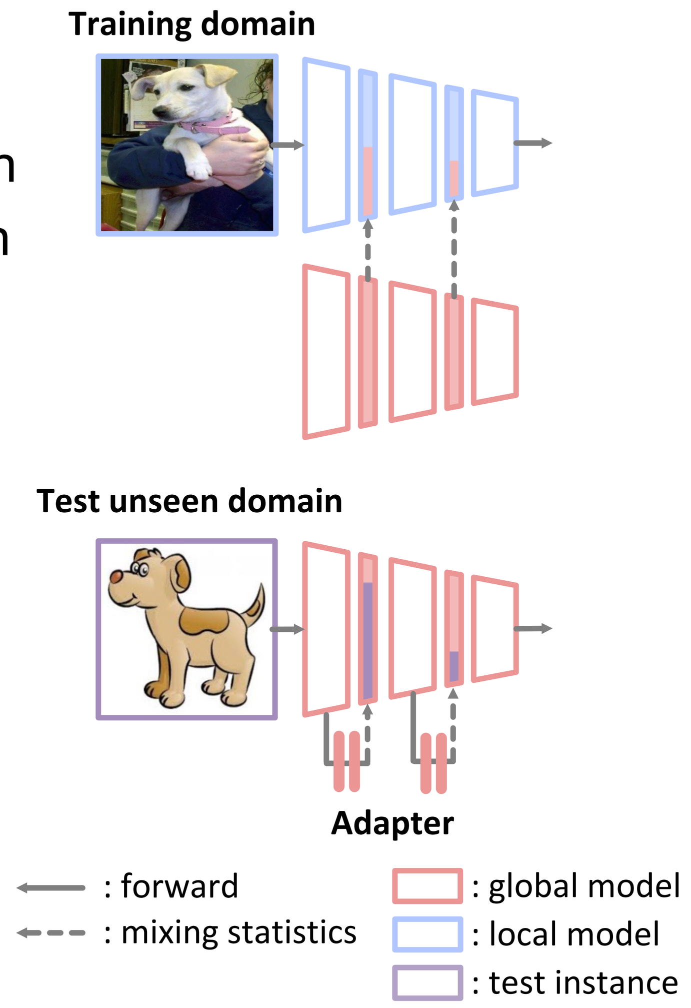
Introduction

- **Federated domain generalization (Federated DG)** aims to improve a federated model's ability to manage various distributed source domains while ensuring reliable generalization to unseen domains
- **Challenges**
  - ***Domain shift:*** local models are trained on limited domains
  - Aggregating local models trained on own limited domain can lead to a significant degradation in a global model performance
- **Limitations of previous research**
  - FL models overfit to local domains, causing aggregation issues.
  - Single-source DG underutilizes the rich server information inherent in FL, leading to suboptimal performance
  - Multi-source DG necessitates sharing elements of local models amongst clients, causing privacy issue



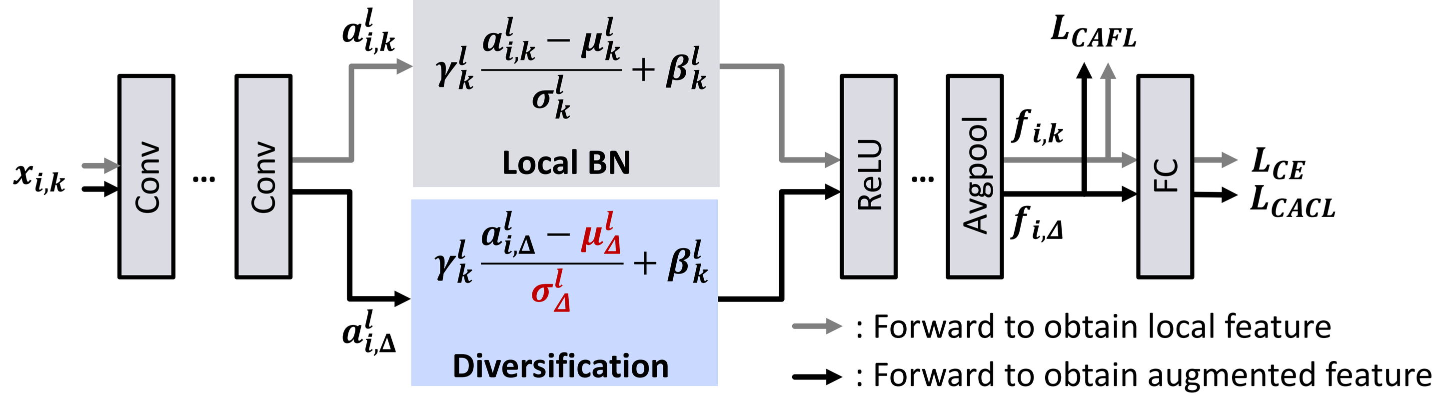
Motivation

- **Feature diversification with global model**
  - Local model relies on CE loss from single domain, which lead to overfitting to single domain
  - The model can learn client invariant features with diverse clients' data, but it poses privacy issue
  - Instead, we augment local data with global BN statistics from the server, which reflects the feature statistics across all clients
- **Feature adaptation with test instance**
  - At test time, it can be difficult to generalize to completely unseen domains where the data distribution has shifted from the training set
  - Using the statistics of the test input can reduce the domain gap between the test and training domains



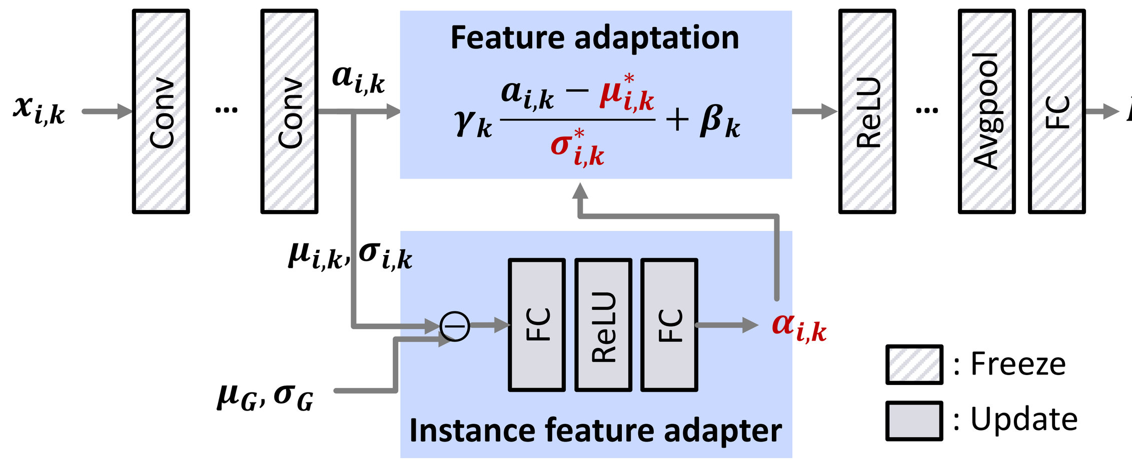
Federated Feature Diversification

- Normalize inputs with a mix of statistics-exploiting global and local BN statistics concurrently
  - Augmented feature $f_{i,\Delta}$ is obtained with $\{\mu_\Delta^l, \sigma_\Delta^l\}_{l=1}^L$
  - $\mu_\Delta^l = u^l \mu_k^l + (1-u^l)\mu_G^l$ and $\sigma_\Delta^l = u^l \sigma_k^l + (1-u^l)\sigma_G^l$
  - $u^l \sim Uniform(0,1)$
- Client-agnostic learning (CAL) with the augmented features
  - $L_{CAFL} = \frac{1}{n_k}\sum \| f_{i,k} - f_{i,\Delta} \|_2^2$, $L_{CACL} = \frac{1}{n_k}\sum \mathrm{CE}\,(C_{\phi_k}(f_{i,\Delta}), y_{i,k})$



Instance Feature Adaptation

- Normalize a test input with a mixed of statistics
  - $\mu_{i,t}^* = \alpha_{i,t}\mu_{i,t} + (1-\alpha_{i,t})\mu_G$ and $\sigma_{i,t}^* = \alpha_{i,t}\sigma_{i,t} + (1-\alpha_{i,t})\sigma_G$
- Instance feature adapter takes the difference between two statistics and generates interpolation value $(\alpha_{i,t})$ as an instance-wise manner
- We employ CE loss to train the adapter on each client, leveraging their differences teaches the adapter how to balance these statistics



Experimental Results

Feature diversification

Mixing components	Acc.
Global only: $u = 0.0$	81.75
Instance only: $u = 1.0$	82.52
Global & instance fixed: $u = 0.5$	83.43
Global & instance randomly: $u \sim U(0,1)$	**84.07**

Computational cost

Methods	#Parms.	Training	Inference	Acc. (%)
FedAvg [35]	11.31M	1.43ms	1.34ms	77.35
SiloBN [1]	11.31M	1.43ms	1.34ms	80.79
FedFD	11.31M	2.25ms	1.34ms	84.07
BIN [37]	11.31M	3.24ms	1.35ms	83.54
MixNorm [21]	11.31M	2.25ms	6.70ms	84.45
IABN [17]	11.31M	2.25ms	1.57ms	84.11
FedFD-A	11.53M	3.50ms	1.37ms	85.53

Ablation study

	Aug.	Loss	Inference	Acc. (%) P	A	C	S	Avg.
(a) Baselines	X	CE	Global statistics	92.74 (0.97)	77.08 (0.59)	75.08 (0.85)	78.28 (1.14)	80.79
		FedProx		92.82 (0.90)	76.43 (0.93)	75.26 (0.99)	78.95 (1.64)	80.86
(b) Effectiveness of FedFD	Mixstyle	CAL	Global statistics	92.04 (0.63)	80.19 (0.28)	77.14 (0.48)	82.53 (0.94)	82.98
	FedFD	**CAL**		**92.99 (0.61)**	**82.17 (1.15)**	**77.71 (1.13)**	**83.40 (0.19)**	**84.07**
(c1) Effectiveness of instance feature adapter	X	CE	Global statistics	92.74 (0.97)	77.08 (0.59)	75.08 (0.85)	78.28 (1.14)	80.79
			Instance statistics (TBN [36])	15.52 (4.89)	10.91 (2.41)	9.36 (7.09)	16.09 (9.44)	12.97
			Instance feature adapter	**93.11 (0.08)**	**80.93 (0.79)**	**77.01 (0.24)**	**82.54 (0.79)**	**83.40**
(c2) Comparison with other inference methods	**FedFD**	**CAL**	BIN [37]	93.86 (0.53)	78.81 (0.45)	79.54 (0.40)	81.94 (0.88)	83.54
			DSON [42]	93.70 (0.39)	83.17 (1.09)	76.08 (1.30)	83.14 (1.02)	84.02
			IABN [17]	93.29 (0.13)	80.18 (1.16)	79.38 (1.33)	83.58 (0.91)	84.11
			MixNorm [21]	93.60 (0.52)	82.77 (0.41)	78.30 (1.13)	83.12 (0.74)	84.45
			Random alpha	90.67 (0.59)	74.82 (2.97)	74.96 (0.76)	79.03 (1.64)	79.87
			Fixed alpha [53]	93.59 (0.25)	79.27 (1.40)	79.64 (1.00)	82.43 (1.12)	83.73
			Instance feature adapter	**94.24 (0.33)**	**84.30 (0.44)**	**79.80 (1.16)**	**83.79 (0.49)**	**85.53**

PACS, VLCS, and OfficeHome

Paradigm	Method	PACS	VLCS	OfficeHome	Avg. of 3 datasets
Decentralized w/o DG	FedAvg [35]	77.35	74.86	63.47	71.89
	FedProx [31]	77.23	74.42	63.02	71.56
	FedBN [32]	79.57	74.85	63.75	72.72
	SiloBN [1]	**80.79**	**75.33**	**64.08**	**73.40**
Decentralized w/ DG	RandAug [8]	77.26	74.32	**64.59**	72.06
	Mixstyle [58]	82.43	**75.54**	63.24	73.74
	SFA [30]	77.98	72.77	59.47	70.06
	RandConv [51]	79.57	73.68	63.25	72.17
	L2D [47]	**82.96**	75.37	63.29	**73.87**
	JiGen [3]	77.69	75.13	64.09	72.30
	RSC [23]	**81.60**	**75.90**	63.01	**73.50**
	SelfReg [26]	80.94	72.79	**64.85**	72.86

Paradigm	Method	PACS	VLCS	OfficeHome	Avg. of 3 datasets
Decentralized w/ federated DG	COPA [48]	85.37	74.51	62.42	74.10
	FedDG [33]	82.52	75.28	64.90	74.23
	CCST [5]	83.56	75.20	63.50	74.09
	CSAC [54]	82.58	75.21	64.35	74.05
	FedSR [38]	79.13	75.36	63.45	72.65
	FedSAM [2]	78.58	74.37	61.93	71.63
	SiloBN w/ GA [55]	80.83	75.41	64.19	73.48
	FedFD	84.07	76.62	64.01	74.90
	FedFD-A	**85.53**	**76.68**	**64.71**	**75.64**

Scalability

Method	PACS P	A	C	S	Avg.
DoPrompt [57]	99.24 (0.16)	90.56 (0.54)	81.06 (0.42)	77.71 (0.56)	87.14
DoPrompt w/ GA [55]	99.33 (0.24)	89.98 (0.53)	81.03 (0.24)	79.01 (1.21)	87.34
FedFD-A	**99.39 (0.06)**	**92.29 (0.42)**	**83.88 (1.07)**	**79.62 (1.23)**	**88.79**

Method	VLCS V	L	C	S	Avg.
DoPrompt [57]	78.32 (1.29)	**65.94 (1.28)**	97.02 (0.49)	77.80 (1.22)	79.82
DoPrompt w/ GA [55]	79.84 (0.49)	65.15 (0.81)	97.14 (0.17)	76.72 (0.59)	79.71
FedFD-A	**81.41 (0.81)**	65.06 (0.62)	**97.46 (0.79)**	**79.64 (0.94)**	**80.89**

*Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc