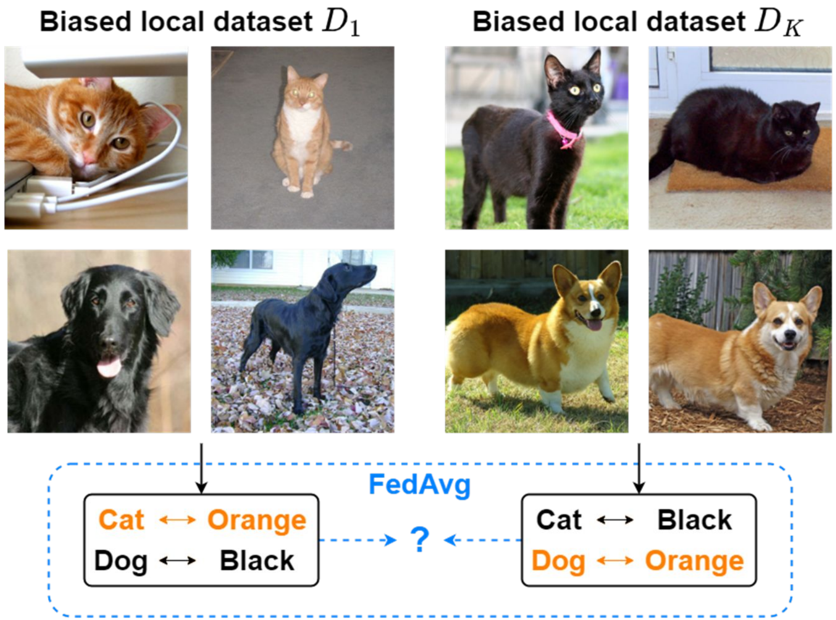
Bias-Eliminating Augmentation Learning for Debiased Federated Learning

Yuan-Yi Xu[1] Ci-Siang Lin[1] Yu-Chiang Frank Wang[1, 2]
[1]National Taiwan University [2]NVIDIA, Taiwan

THU-AM-377

Motivation

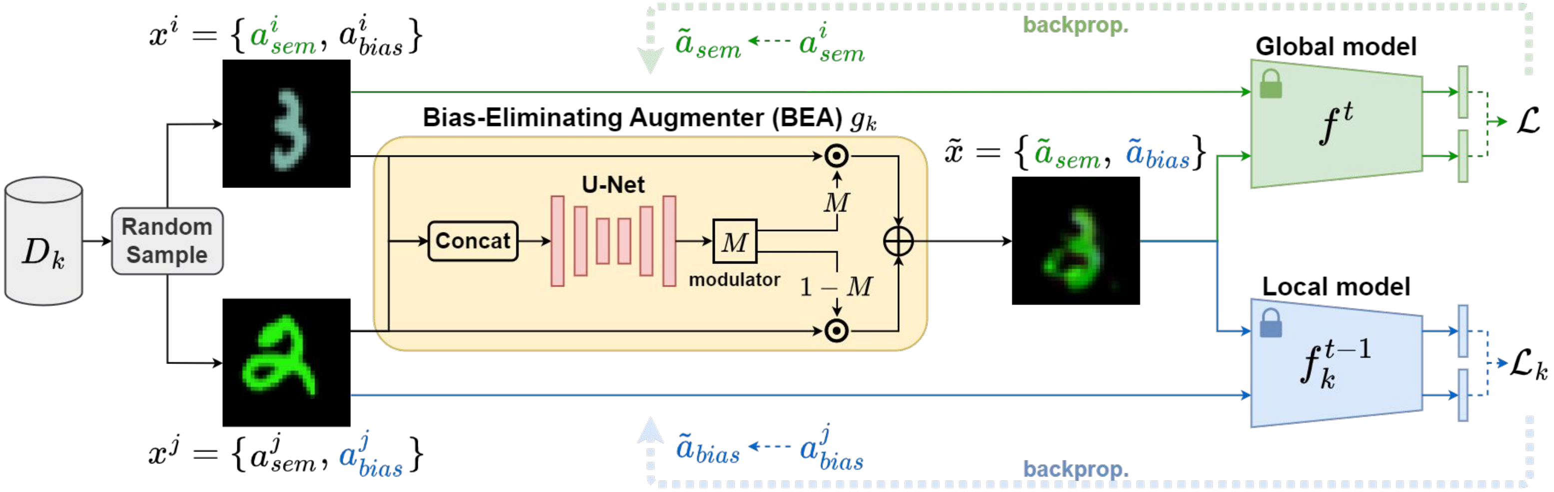
- Local data bias is likely to happen in real-world FL applications
- Debiased federated learning aims to learn unbiased models from biased local datasets



Method Overview

- Our proposed FedBEAL enables each client to train a **B**ias-**E**liminating **A**ugmenter (**BEA**) for generating bias-conflicting samples to debias local training

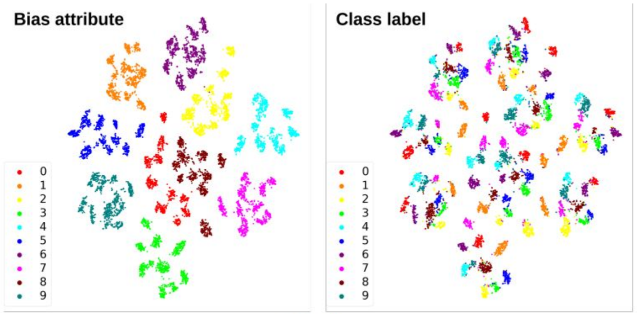
Semantic attribute: Digits
Bias attribute: Color



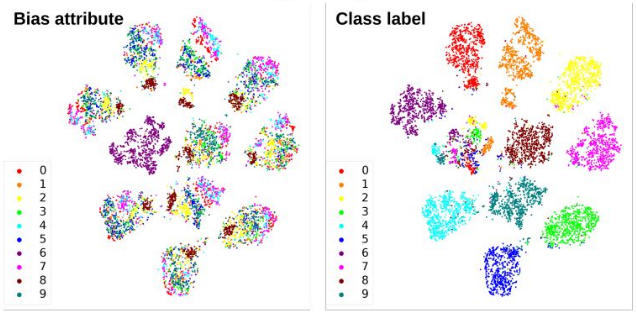
Results

- Extensive experiments confirm the effectiveness of our method

Dataset	Colored MNIST		Corrupted CIFAR-10		Collage CIFAR-10	
Bias ratio β	0.99	0.999	0.99	0.999	0.99	0.999
Baselines						
SOLO	46.90	14.46	16.80	13.19	12.28	10.58
FedAvg [35]	93.90	72.67	49.03	40.28	52.93	36.91
Centralized Debiasing Methods						
LfF [36]	87.64	55.27	53.47	42.25	46.53	26.96
SoftCon [18]	96.75	86.39	55.38	47.61	54.19	42.98
Lee *et al.* [27]	90.28	61.35	54.86	45.90	41.02	22.58
Data Heterogeneous Federated Learning						
FedProx [30]	94.51	73.07	44.06	34.01	41.87	25.94
SCAFFOLD [21]	95.01	68.41	41.73	34.35	38.37	33.85
MOON [29]	93.33	69.37	36.79	26.06	34.71	19.97
FedBN [31]	N/A	N/A	48.46	36.52	46.51	32.53
Ours	**98.58**	**91.99**	**59.18**	**49.09**	**69.53**	**64.53**



(a) FedAvg



(b) FedBEAL

More Details

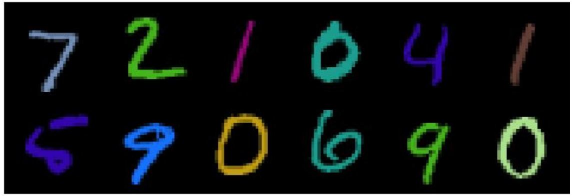
Problem Definition

- Training data (biased): Each client has disparate bias-label correlations
- Test data (unbiased)
- Colored MNIST dataset:
  - Label: digits
  - Bias: color

Client 1 training data (biased)

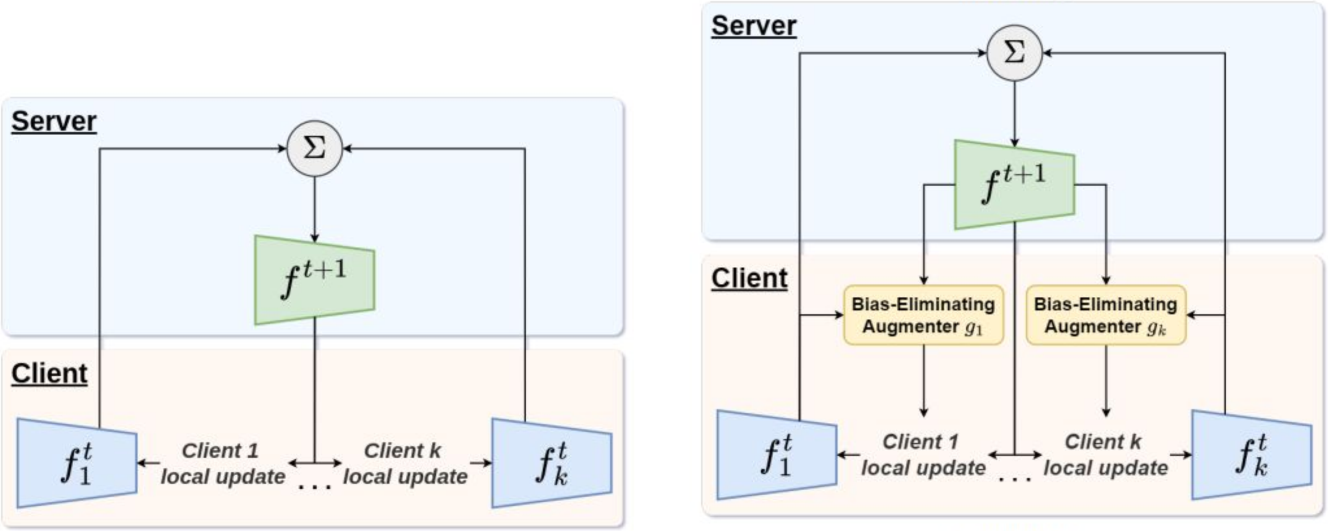
Client 2 training data (biased)

Test data (unbiased)



Method Overview

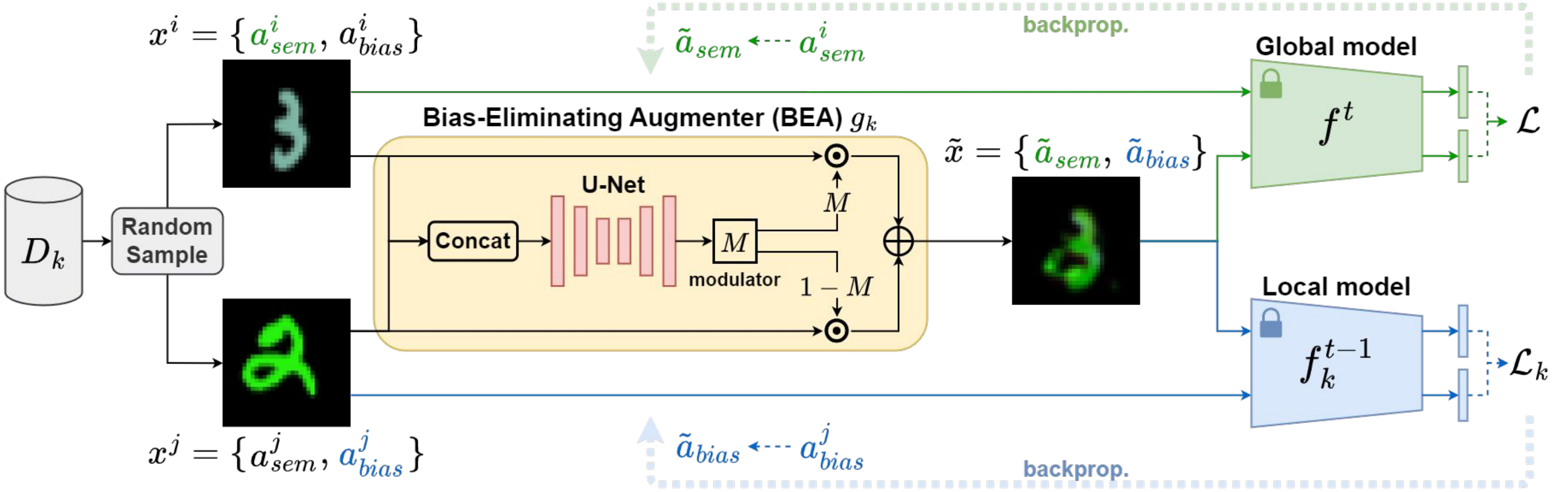
- Our FedBEAL learns Bias-Eliminating Augmenters (BEA) to produce bias-conflicting samples at each client



(a) FedAvg (b) FedBEAL

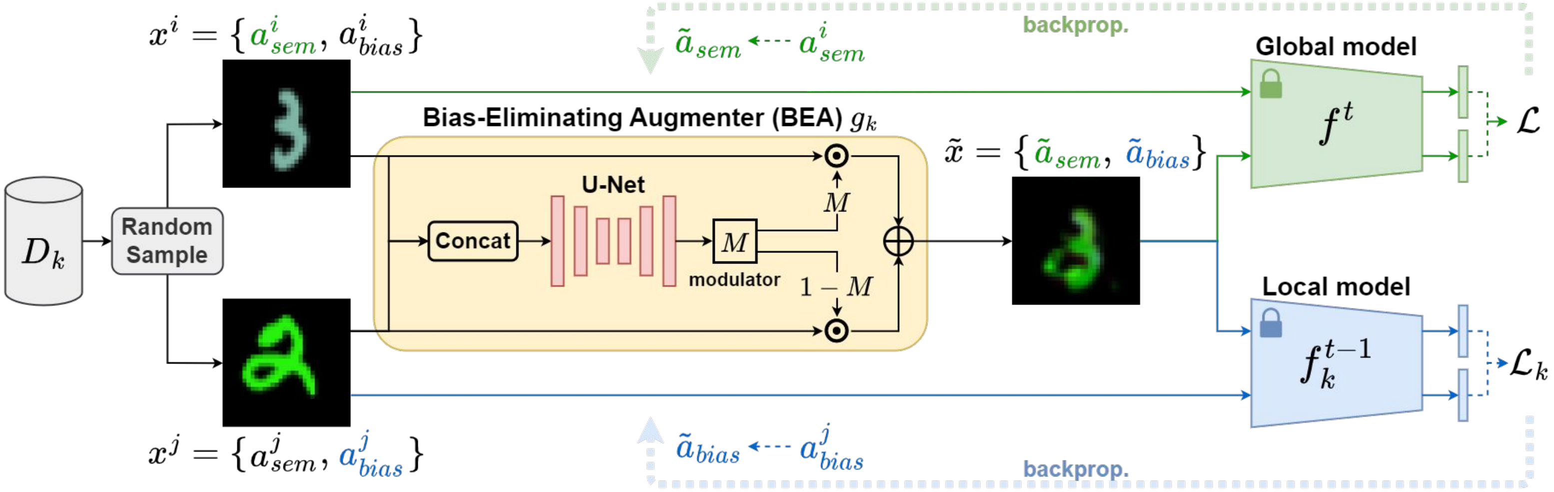
Design and Architecture of BEA

- Mixing two biased data to produce bias-conflicting sample
  - $\tilde{x} = M \odot x^i + (1 - M) \odot x^j$
- Utilizing U-Net as the backbone to produce the modulator M ∈ $[0, 1]^{H \times W \times 3}$

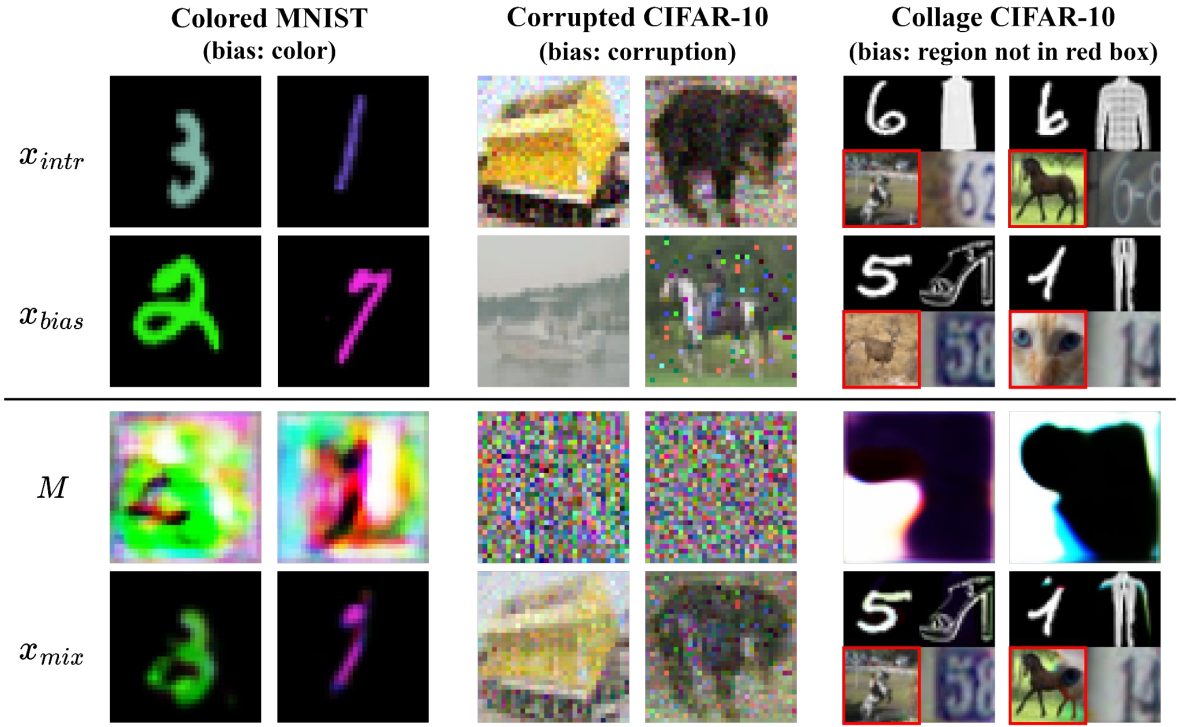


Learning of BEA

- Extracting semantic attributes via unbiased global prediction:
  - $\mathcal{L} = d_{KL}(f^t(\tilde{x}), f^t(x^i))$
- Producing bias attributes via biased local prediction:
  - $\mathcal{L}_k = d_{KL}(f_k^{t-1}(\tilde{x}), f_k^{t-1}(x^j))$



Visualization of Images Produced by BEA



Quantitative Evaluation

Dataset	Colored MNIST		Corrupted CIFAR-10		Collage CIFAR-10	
Bias ratio β	0.99	0.999	0.99	0.999	0.99	0.999
Baselines						
SOLO	46.90	14.46	16.80	13.19	12.28	10.58
FedAvg [35]	93.90	72.67	49.03	40.28	52.93	36.91
Centralized Debiasing Methods						
LfF [36]	87.64	55.27	53.47	42.25	46.53	26.96
SoftCon [18]	<u>96.75</u>	<u>86.39</u>	<u>55.38</u>	<u>47.61</u>	<u>54.19</u>	<u>42.98</u>
Lee *et al*. [27]	90.28	61.35	54.86	45.90	41.02	22.58
Data Heterogeneous Federated Learning						
FedProx [30]	94.51	73.07	44.06	34.01	41.87	25.94
SCAFFOLD [21]	95.01	68.41	41.73	34.35	38.37	33.85
MOON [29]	93.33	69.37	36.79	26.06	34.71	19.97
FedBN [31]	N/A	N/A	48.46	36.52	46.51	32.53
Ours	**98.58**	**91.99**	**59.18**	**49.09**	**69.53**	**64.53**

Conclusion

- We propose FedBEAL to address the challenging task of debiased federated learning
- BEAs generate bias-conflicting samples that automatically mitigate bias in federated learning
- Extensive experiments confirm the effectiveness and robustness of FedBEAL