Geometric Knowledge-Guided Localized Global Distribution Alignment for Federated Learning

Yanbiao Ma*,[1], Wei Dai*,[1], Wenke Huang[2], Jiayi Chen[1]

CVPR Oral

(wdai139@stu.xidian.edu.cn)

[1]Xidian University

[2]Wuhan University

Presenter: Wei Dai 戴威 3rd-year undergraduate student at Xidian University

Background

Algorithm 1 `FederatedAveraging`. The K clients are indexed by k; B is the local minibatch size, E is the number of local epochs, and η is the learning rate.

Server executes:
initialize w_0
for each round $t = 1, 2, \dots$ **do**
  $m \leftarrow \max(C \cdot K, 1)$
  $S_t \leftarrow$ (random set of m clients)
  for each client $k \in S_t$ **in parallel do**
    $w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$
  $m_t \leftarrow \sum_{k \in S_t} n_k$
  $w_{t+1} \leftarrow \sum_{k \in S_t} \frac{n_k}{m_t} w_{t+1}^k$ *// Erratum*[4]

ClientUpdate(k, w): *// Run on client k*
$\mathcal{B} \leftarrow$ (split $\mathcal{P}_k$ into batches of size B)
for each local epoch i from 1 to E **do**
  for batch $b \in \mathcal{B}$ **do**
    $w \leftarrow w - \eta \nabla \ell(w; b)$
return w to server

FedAvg

Parameter average aggregation

Algorithm 2 `FedProx` (Proposed Framework)

Input: $K, T, \mu, \gamma, w^0, N, p_k, k = 1, \cdots, N$
for $t = 0, \cdots, T-1$ **do**
  Server selects a subset S_t of K devices at random (each device k is chosen with probability p_k)
  Server sends w^t to all chosen devices
  Each chosen device $k \in S_t$ finds a w_k^{t+1} which is a γ_k^t-inexact minimizer of: $w_k^{t+1} \approx \arg\min_w h_k(w; w^t) = F_k(w) + \frac{\mu}{2}\|w - w^t\|^2$
  Each device $k \in S_t$ sends w_k^{t+1} back to the server
  Server aggregates the w's as $w^{t+1} = \frac{1}{K}\sum_{k \in S_t} w_k^{t+1}$
end for

FedProx

Model consistency regularization term

Algorithm 1: The MOON framework

Input: number of communication rounds T, number of parties N, number of local epochs E, temperature τ, learning rate η, hyper-parameter μ
Output: The final model w^T

Server executes:
initialize w^0
for $t = 0, 1, ..., T-1$ **do**
  for $i = 1, 2, ..., N$ **in parallel do**
    send the global model w^t to P_i
    $w_i^t \leftarrow$ **PartyLocalTraining**(i, w^t)
  $w^{t+1} \leftarrow \sum_{k=1}^{N} \frac{|\mathcal{D}^i|}{|\mathcal{D}|} w_k^t$
return w^T

PartyLocalTraining(i, w^t):
$w_i^t \leftarrow w^t$
for epoch $i = 1, 2, ..., E$ **do**
  for each batch $\mathbf{b} = \{x, y\}$ of $\mathcal{D}^i$ **do**
    $\ell_{sup} \leftarrow CrossEntropyLoss(F_{w_i^t}(x), y)$
    $z \leftarrow R_{w_i^t}(x)$
    $z_{glob} \leftarrow R_{w^t}(x)$
    $z_{prev} \leftarrow R_{w_i^{t-1}}(x)$
    $\ell_{con} \leftarrow -\log \frac{\exp(\text{sim}(z, z_{glob})/\tau)}{\exp(\text{sim}(z, z_{glob})/\tau) + \exp(\text{sim}(z, z_{prev})/\tau)}$
    $\ell \leftarrow \ell_{sup} + \mu \ell_{con}$
    $w_i^t \leftarrow w_i^t - \eta \nabla \ell$
return w_i^t to server

MOON

Model comparison loss term

Algorithm 1: `FedProto`

Input: $D_i, \omega_i, i = 1, \cdots, m$
Server executes:
Initialize global prototype set $\{\bar{C}^{(j)}\}$ for all classes.
for each round $T = 1, 2, \dots$ **do**
  for each client i **in parallel do**
    $C_i \leftarrow$ LocalUpdate $(i, \bar{C}_i)$
  end for
  Update global prototype by Eq. 6.
  Update local prototype set C_i with prototypes in $\{\bar{C}^{(j)}\}$
end for

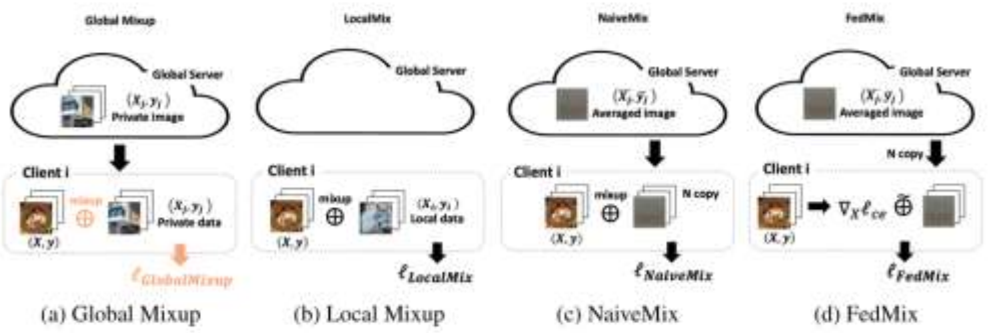
LocalUpdate$(i, \bar{C}_i)$:
for each local epoch **do**
  for batch $(x_i, y_i) \in D_i$ **do**
    Compute local prototype by Eq. 3.
    Compute loss by Eq. 7 using local prototypes.
    Update local model according to the loss.
  end for
end for
return $C^{(i)}$

FedProto

Prototype comparison supervision

Federal Data Augmentation

Previous Methods



$$\mu_m^k = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} \boldsymbol{X}_m^{k,(h,w)} \in \mathbb{R}^{B\times C}, \quad \sigma_m^k = \sqrt{\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}(\boldsymbol{X}_m^{k,(h,w)} - \mu_m^k)^2} \in \mathbb{R}^{B\times C},$$

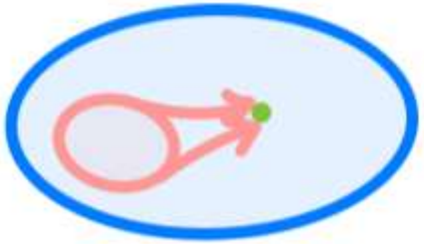
$$\Sigma^2_{\mu_m^k} = \frac{1}{B}\sum_{b=1}^{B}(\mu_m^k - \mathbb{E}[\mu_m^k])^2 \in \mathbb{R}^C, \quad \Sigma^2_{\sigma_m^k} = \frac{1}{B}\sum_{b=1}^{B}(\sigma_m^k - \mathbb{E}[\sigma_m^k])^2 \in \mathbb{R}^C,$$

$$\hat{\boldsymbol{X}}_m^k = \hat{\sigma}_m^k \frac{\boldsymbol{X}_m^k - \mu_m^k}{\sigma_m^k} + \hat{\mu}_m^k, \quad \text{where} \quad \hat{\mu}_m^k \sim \mathcal{N}(\mu_m^k, \hat{\Sigma}^2_{\mu_m^k}), \quad \hat{\sigma}_m^k \sim \mathcal{N}(\sigma_m^k, \hat{\Sigma}^2_{\sigma_m^k}).$$

FedMix

1: Aggregate the average class features of each client on the server-side to form a global class prototype

2: The client receives the class prototype and enhances it by mixing it with local samples in the feature space



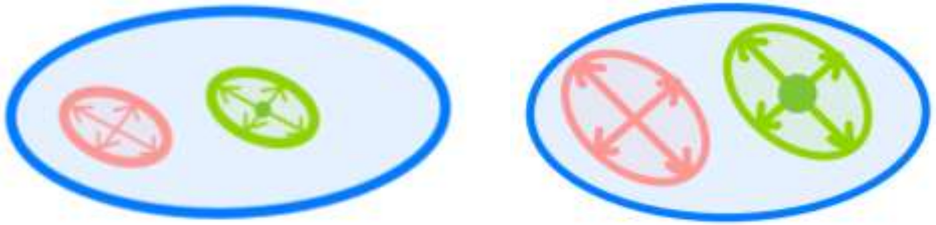
Observing the interpolated distribution towards the center of the true distribution

FedFA

1: Aggregate the category mean features of each client on the server-side to form a global class prototype

2: The client calculates the feature variance based on the observed distribution and records the distribution disturbance

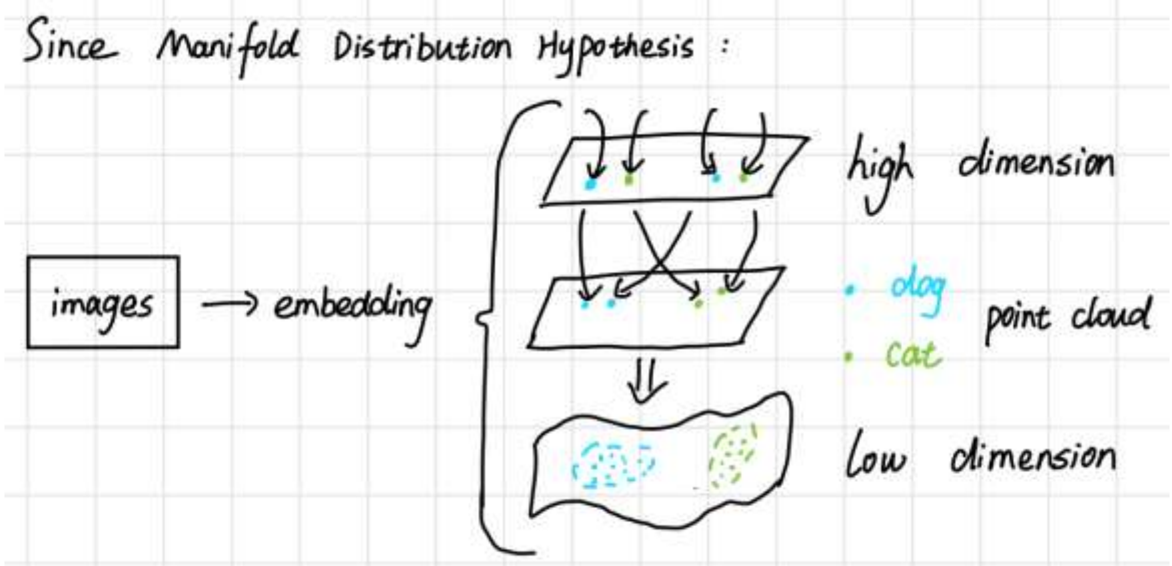
3: Each client locally perturbs the global class prototype



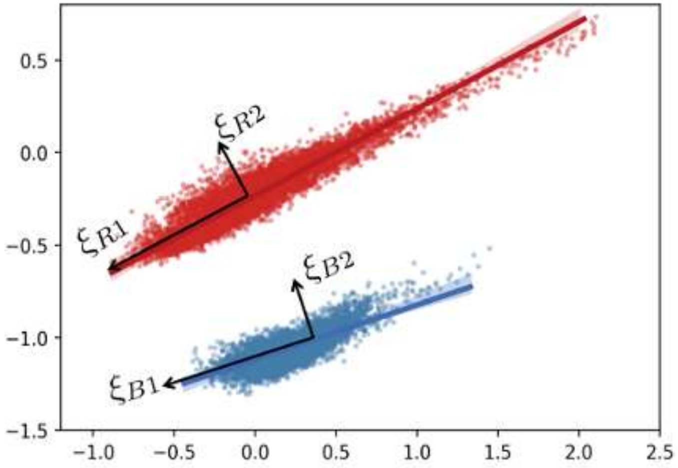
Simulate and observe the distribution from the disturbance of the distribution center

Ref: ICLR "FedFA: Federated feature augmentation", ICLR "Fedmix: Approximation of mixup under mean augmented federated learning"

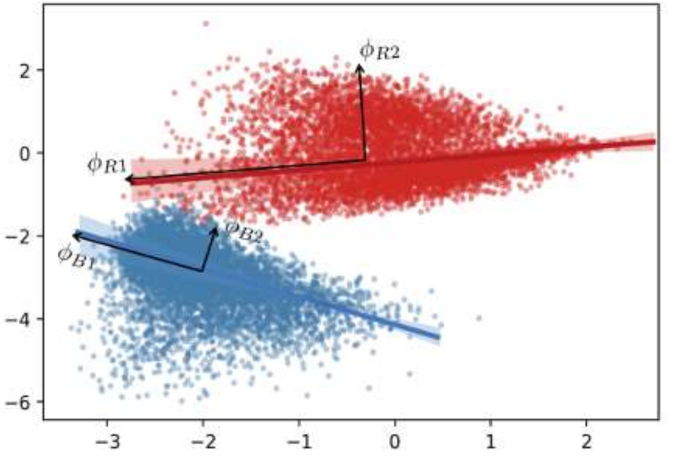
Inspiration



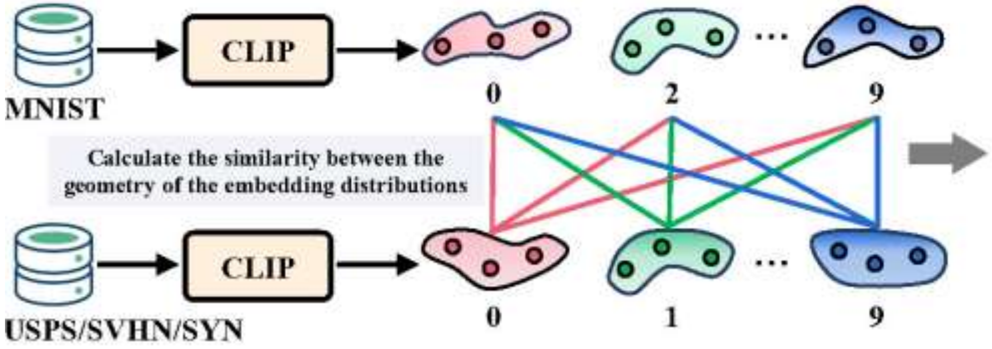
Manifold distribution Hypothesis

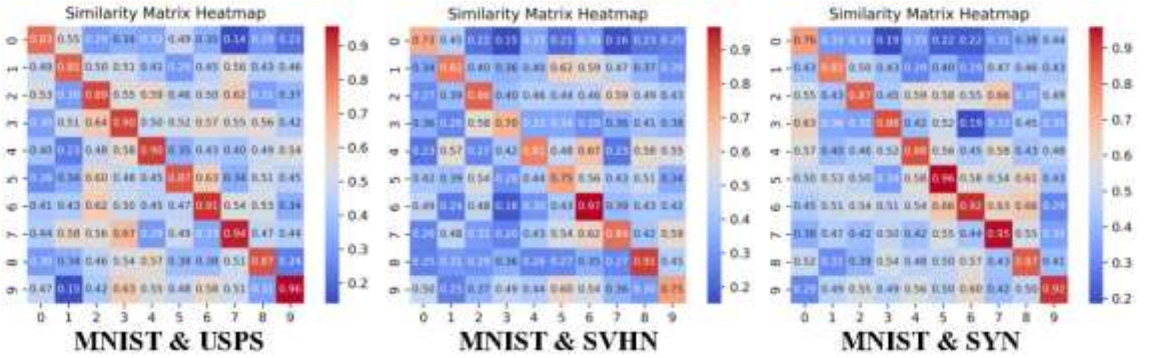


Similar geometric directions



Different geometric directions

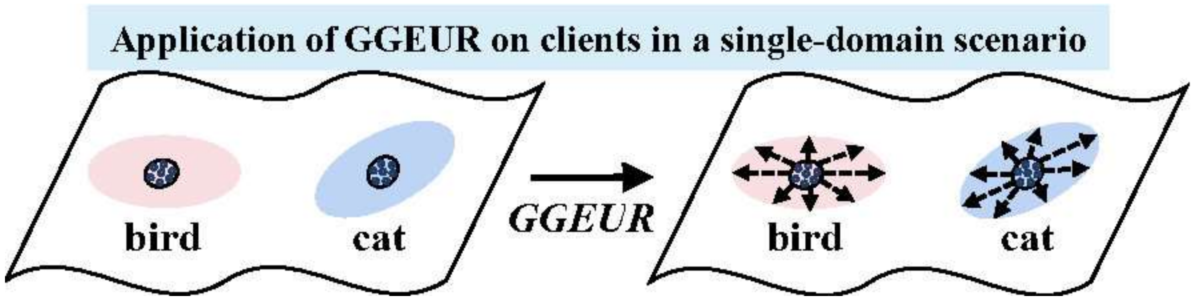




General Feature of VLM

Ref: IJCV "Geometric Prior Guided Feature Representation Learning for Long-Tailed Classification"

Methodology-Single domain



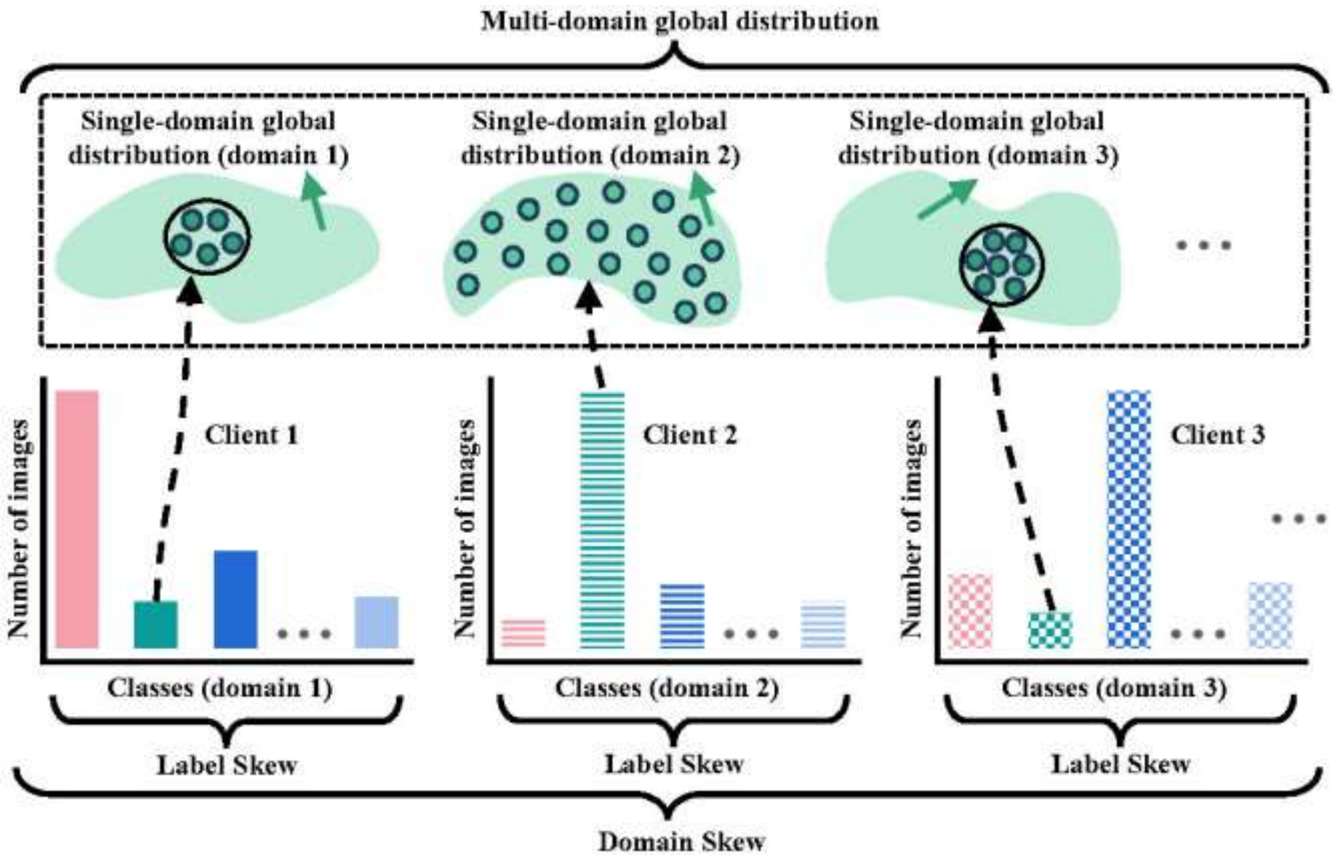
Single domain GGEUR

$$\Sigma_X = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_i x_i^T\right] = \frac{1}{n} X X^T \in \mathbb{R}^{P\times P}.$$

Local covariance matrix

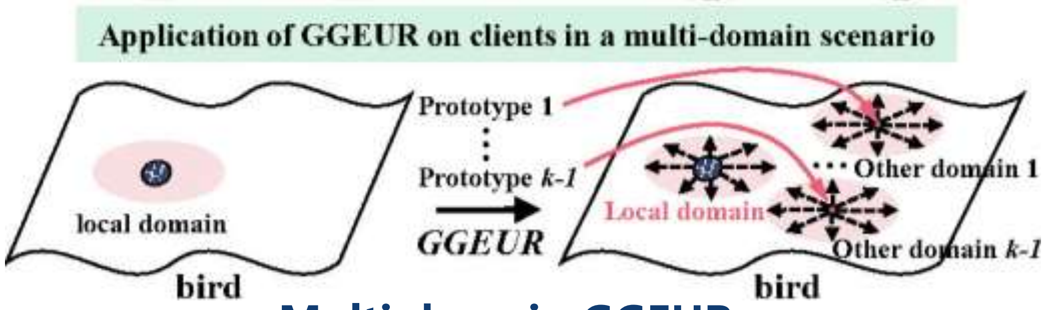
$$\Sigma_i = \frac{1}{N_i}\sum_{k=1}^{K}\sum_{j=1}^{n_k^i}(x_k^{i,j} - \mu_i)(x_k^{i,j} - \mu_i)^T,$$

Global covariance matrix



Single domain Geometric augmentation

Methodology-Multi domain



Multi domain GGEUR

$$\mu_i = \frac{1}{N_i}\sum_{k=1}^{K}\sum_{j=1}^{n_k^i} x_k^{i,j} = \frac{1}{N_i}\sum_{k=1}^{K} n_k^i \mu_k^i,$$

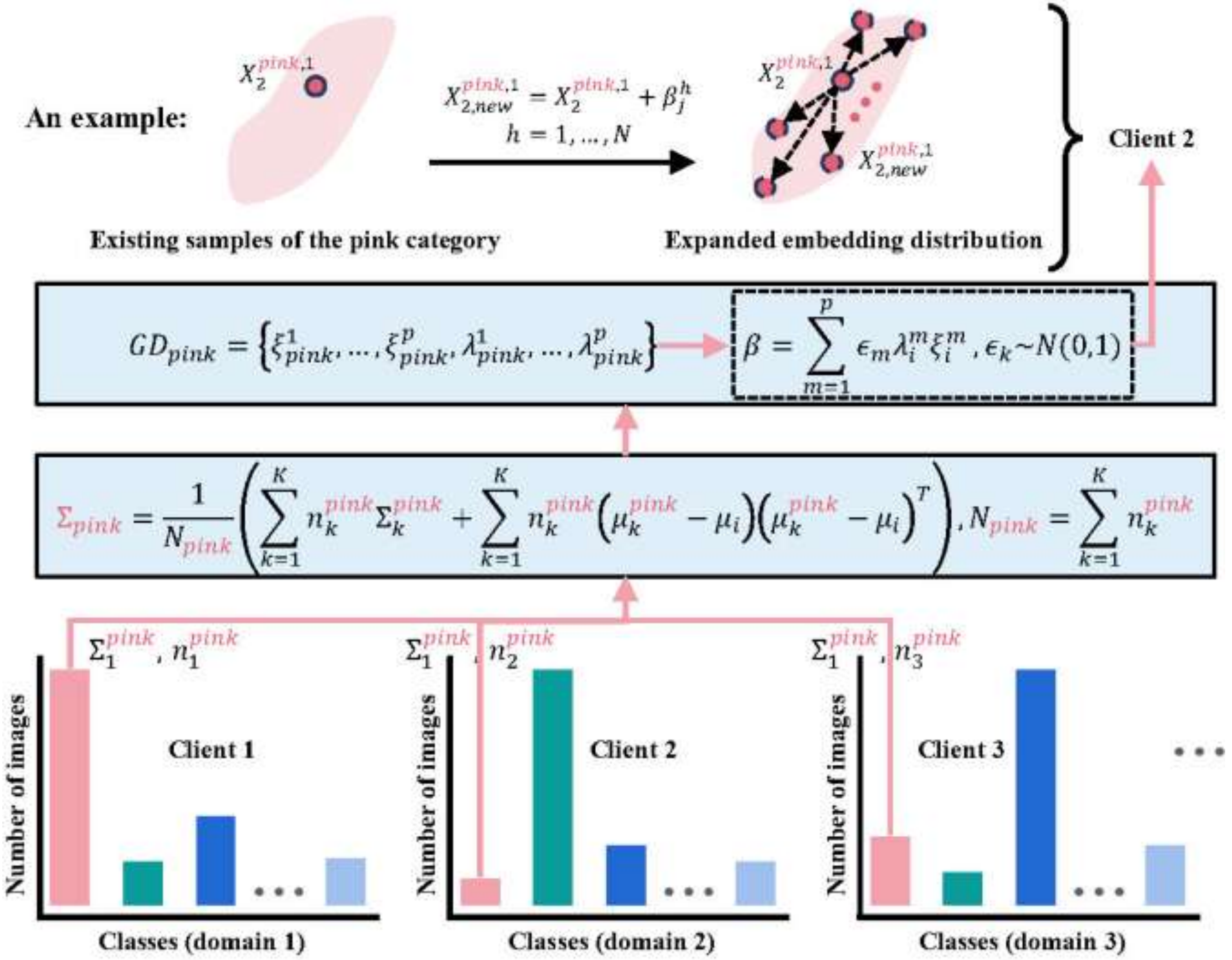
Global class prototype

$$\Sigma_X = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_i x_i^T\right] = \frac{1}{n} X X^T \in \mathbb{R}^{P\times P}.$$

Local covariance matrix

$$\Sigma_i = \frac{1}{N_i}\sum_{k=1}^{K}\sum_{j=1}^{n_k^i}(x_k^{i,j} - \mu_i)(x_k^{i,j} - \mu_i)^T,$$

Global covariance matrix



Multi domain Geometric augmentation

Experiment

Methods	CIFAR-100 0.5	CIFAR-100 0.3	CIFAR-100 0.1	Tiny-ImageNet 0.5	Tiny-ImageNet 0.3	Tiny-ImageNet 0.1
Zero-Shot CLIP	64.87			63.67		
FedTPG	71.40	70.95	68.63	67.63	66.72	64.71
FedCLIP	72.03	71.20	70.64	70.41	70.37	69.50
FedMix (CLIP+MLP)	81.31	79.62	73.85	70.89	68.57	63.43
FEDGEN (CLIP+MLP)	81.24	78.97	73.15	72.37	70.35	64.16
FedFA (CLIP+MLP)	81.98	79.31	74.68	70.41	70.68	64.62
FedAvg (CLIP+MLP)	81.41	77.68	68.22	70.08	67.65	60.10
+ GGEUR	**83.31**	**81.65**	**77.70**	**73.89**	**72.19**	**66.86**

Methods	CIFAR-10 0.01	0.03	0.05	0.07	0.09
FedAvg (CLIP+MLP)	90.87	90.13	91.96	92.05	91.82
+ GGEUR	**94.39**	**94.25**	**95.07**	**95.21**	**95.38**

Methods	CIFAR-100 0.01	0.03	0.05	0.07	0.09
FedAvg (CLIP+MLP)	58.71	60.77	62.32	61.69	66.51
+ GGEUR	**75.72**	**75.40**	**75.96**	**76.72**	**78.00**

Methods	Tiny-ImageNet 0.01	0.03	0.05	0.07	0.09
FedAvg (CLIP+MLP)	53.03	54.57	58.91	58.77	59.13
+ GGEUR	**64.27**	**65.79**	**66.49**	**66.34**	**66.85**

Single domain(label shift)

Methods	Digits MNIST	USPS	SVHN	SYN	AVG ↑	STD ↓
FedMix (CLIP+MLP)	95.03	90.25	57.50	72.60	78.85	14.89
FEDGEN (CLIP+MLP)	95.85	92.52	58.77	73.62	80.19	14.99
FedFA (CLIP+MLP)	96.68	92.97	57.87	75.53	80.76	15.44
FedAvg [39]	90.40	60.30	34.68	46.99	58.09	20.74
FedAvg (CLIP+MLP)	95.12	89.74	56.36	65.17	76.60	16.25
+ GGEUR (Step 1)	96.02	93.02	58.55	73.13	80.18	15.28
+ GGEUR (Step 1 & 2)	97.05	94.12	63.54	74.73	**82.36**	**13.84**
SCAFFOLD [16]	97.79	94.45	26.64	90.69	77.39	29.41
SCAFFOLD (CLIP+MLP)	94.62	90.08	54.33	68.71	76.93	16.31
+ GGEUR	95.91	92.08	63.25	71.54	**80.70**	**13.69**
MOON [25]	92.78	68.11	33.36	39.28	58.36	23.82
MOON (CLIP+MLP)	75.64	73.09	38.83	52.74	60.07	**15.14**
+ GGEUR (Step 1)	84.64	81.96	43.04	60.35	67.50	16.97
+ GGEUR (Step 1 & 2)	95.16	91.13	55.23	71.00	**78.13**	16.08
FedDyn [1]	88.91	60.34	34.57	50.72	58.65	19.76
FedDyn (CLIP+MLP)	95.46	92.13	58.89	70.30	79.19	15.19
+ GGEUR	97.07	94.02	63.34	74.83	**82.31**	**13.88**
FedOPT [48]	92.71	87.62	31.32	87.92	74.89	25.38
FedOPT (CLIP+MLP)	94.57	88.79	58.65	66.47	77.12	14.96
+ GGEUR	96.43	93.47	62.35	70.75	**80.75**	**14.54**
FedProto [55]	90.54	89.54	34.61	58.00	68.18	23.38
FedProto (CLIP+MLP)	94.86	92.63	54.29	65.52	76.83	17.40
+ GGEUR	97.19	94.12	63.70	73.83	**82.21**	**13.96**
FedNTD [22]	52.31	58.07	18.03	97.29	56.43	28.12
FedNTD (CLIP+MLP)	95.82	91.43	58.26	69.95	78.86	15.41
+ GGEUR	97.08	94.32	63.57	73.53	**82.13**	**14.06**

Methods	Office-Caltech Am	Ca	D	W	AVG ↑	STD ↓
FedAvg [39]	81.99	73.21	79.37	67.93	75.62	6.31
FedAvg (CLIP+MLP)	98.26	96.74	100	100	**98.75**	**1.57**
SCAFFOLD [16]	39.77	42.50	78.02	70.69	57.75	19.44
SCAFFOLD (CLIP+MLP)	96.18	92.88	95.83	93.26	**94.54**	**1.70**
MOON [25]	84.42	75.98	84.67	68.97	78.51	7.53
MOON (CLIP+MLP)	98.61	97.33	100	98.88	**98.70**	**1.09**
FedDyn [1]	84.02	72.59	77.34	68.97	75.72	6.50
FedDyn (CLIP+MLP)	98.61	96.74	100	100	**98.84**	**1.54**
FedOPT [48]	79.05	71.96	89.34	74.48	78.71	7.67
FedOPT (CLIP+MLP)	98.26	97.33	100	100	**98.90**	**1.33**
FedProto [55]	87.79	75.98	90.00	79.31	83.27	6.70
FedProto (CLIP+MLP)	98.26	96.74	100	100	**98.75**	**1.57**
FedNTD [22]	10.95	10.89	14.67	10.34	11.71	1.99
FedNTD (CLIP+MLP)	97.92	96.14	100	100	**98.51**	**1.86**

Methods	PACS P	AP	Ct	Sk	AVG ↑	STD ↓
FedAvg [39]	76.09	64.19	83.50	89.40	78.30	9.41
FedAvg (CLIP+MLP)	99.40	98.37	99.01	93.64	**97.60**	**2.32**
SCAFFOLD [16]	61.95	45.44	58.87	54.64	55.25	6.22
SCAFFOLD (CLIP+MLP)	92.42	81.63	80.68	87.28	**85.50**	**4.72**
MOON [25]	74.44	64.19	83.92	89.17	77.93	9.53
MOON (CLIP+MLP)	99.60	99.02	99.43	93.89	**97.99**	**2.37**
FedDyn [1]	78.17	63.29	82.27	89.93	78.66	9.70
FedDyn (CLIP+MLP)	99.40	98.37	99.01	93.55	**97.58**	**2.36**
FedOPT [48]	78.66	67.66	82.41	83.68	78.12	6.31
FedOPT (CLIP+MLP)	99.40	98.37	99.01	93.64	**97.60**	**2.32**
FedProto [55]	85.63	73.69	83.57	91.14	83.51	6.31
FedProto (CLIP+MLP)	99.40	98.21	99.01	94.23	**97.71**	**2.06**
FedNTD [22]	16.77	18.23	28.47	93.18	39.16	31.51
FedNTD (CLIP+MLP)	99.40	98.54	99.29	92.28	**97.38**	**2.96**

Multi domain(domain shift)

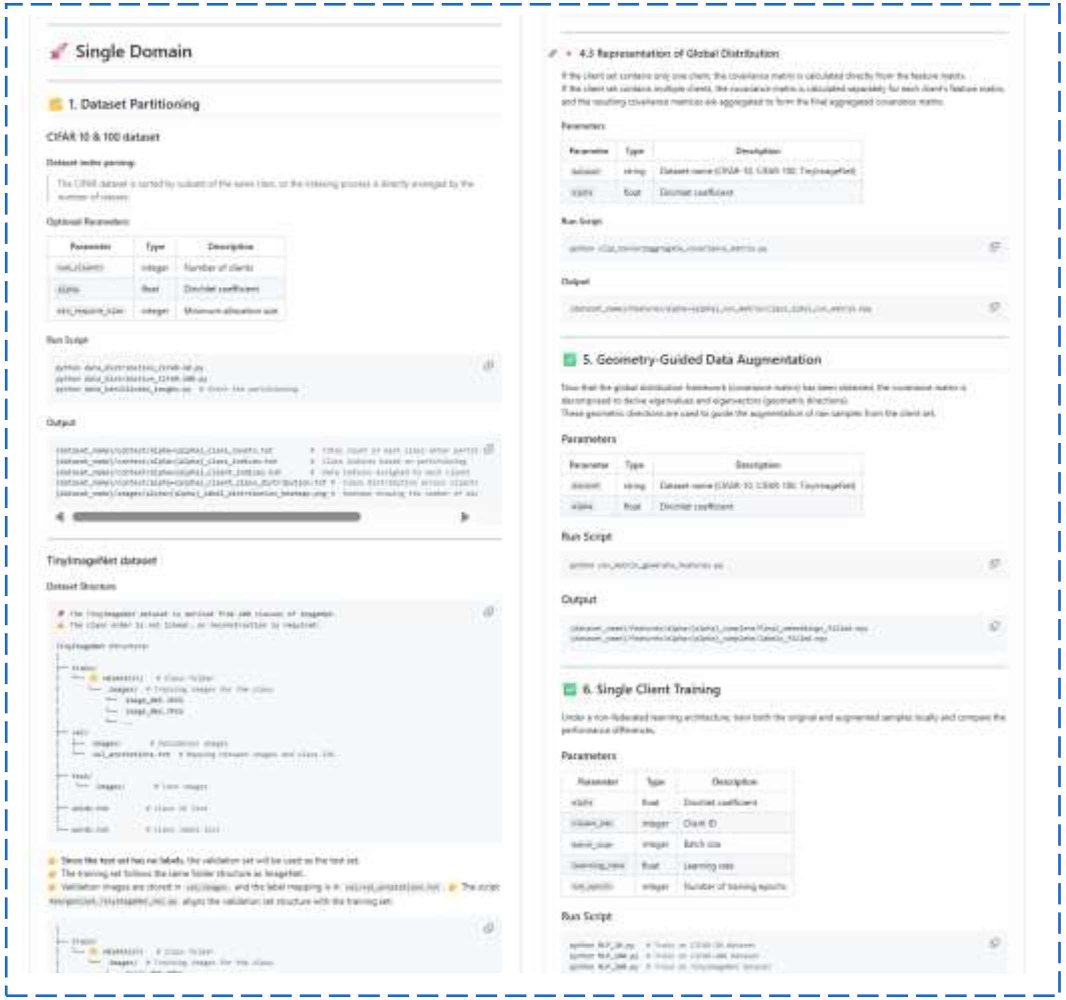
Benchmark: Tpami “Federated Learning for Generalization, Robustness, Fairness: A Survey and Benchmark”

Experiment

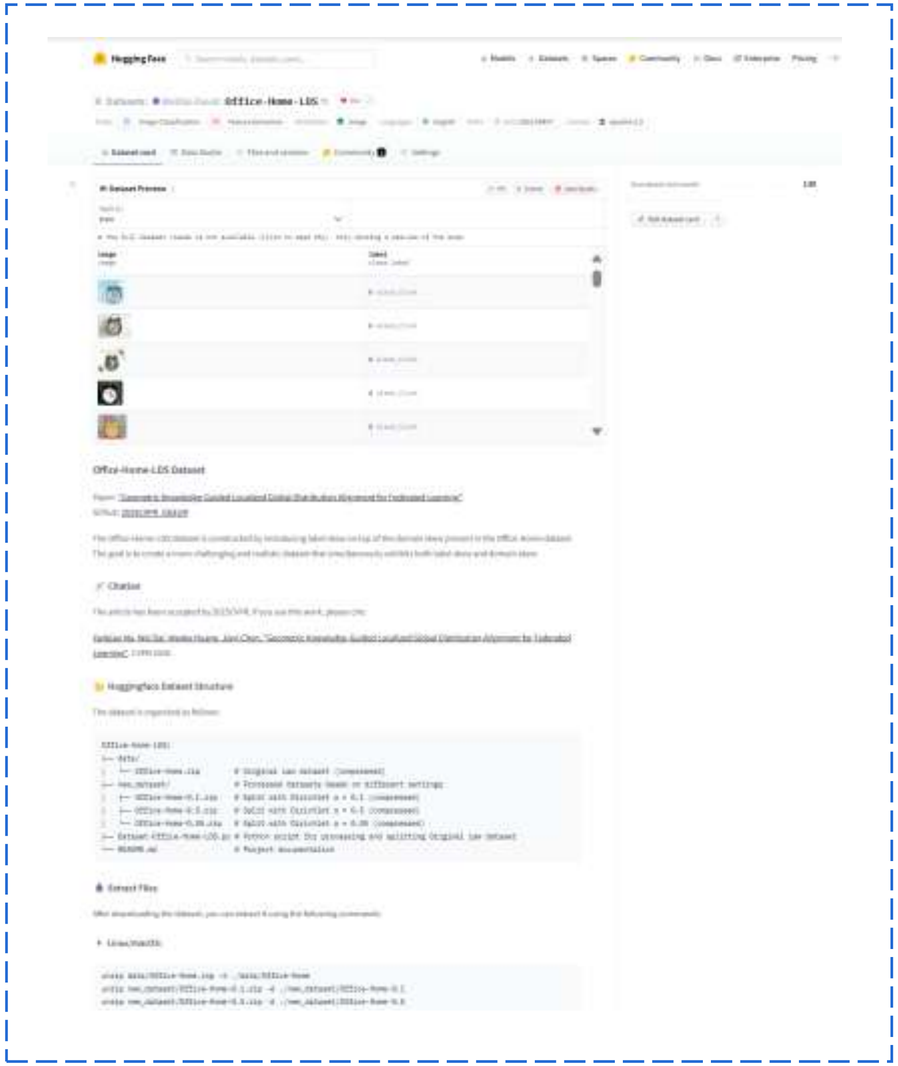
Methods	Office-Home-LDS A	C	P	R	AVG ↑	STD ↓
FedAvg (CLIP+MLP) [39]	65.29	58.17	80.56	76.53	70.14	8.89
+ GGEUR	78.33	79.01	90.17	88.46	**83.99**	**5.36**
SCAFFOLD (CLIP+MLP) [16]	68.72	66.79	83.63	80.12	74.82	7.20
+ GGEUR	78.60	78.32	89.86	89.07	**83.96**	**5.51**
MOON (CLIP+MLP) [25]	69.27	68.63	86.56	82.87	76.83	7.99
+ GGEUR	72.02	70.31	86.11	83.87	**78.08**	**6.98**
FedDyn (CLIP+MLP) [1]	58.30	55.19	77.63	72.86	65.99	9.47
+ GGEUR	78.88	78.55	90.47	88.46	**84.09**	**5.42**
FedOPT (CLIP+MLP) [48]	58.44	54.89	76.80	72.25	65.59	9.16
+ GGEUR	79.01	78.32	90.84	88.61	**84.20**	**5.59**
FedProto (CLIP+MLP) [55]	65.84	56.49	80.41	74.85	69.40	9.09
+ GGEUR	78.05	77.71	89.79	87.84	**83.35**	**5.51**
FedNTD (CLIP+MLP) [22]	69.68	66.64	84.53	80.96	75.46	7.48
+ GGEUR	78.19	74.66	90.24	86.77	**82.46**	**6.29**

Benchmark: Tpami “Federated Learning for Generalization, Robustness, Fairness: A Survey and Benchmark”

Open Resource



Code



Dataset

Thanks for your time and attention.

If you have any questions, don’t hesitate to drop me an email!

wdai139@stu.xidian.edu.cn

Arxiv：https://arxiv.org/pdf/2503.06457

Code：https://github.com/WeiDai-David/2025CVPR_GGEUR

Dataset：https://huggingface.co/datasets/WeiDai-David/Office-Home-LDS