NEURAL INFORMATION PROCESSING SYSTEMS

StableFDG: Style and Attention Based Learning for Federated Domain Generalization

Jungwuk Park*, Dong-Jun Han*, Jinho Kim, Shiqiang Wang, Christopher Brinton and Jaekyun Moon

Dec. 2023

KAIST PURDUE UNIVERSITY IBM

Background:
Domain Generalization

- Goal: Perform well on the unseen domain

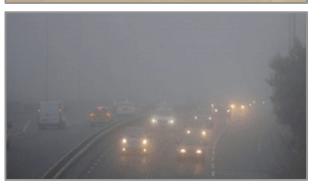
Domain 1
(rain)

D_1

Domain 2
(sand)

D_2

Domain 3
(fog)



D_3

Source domains

Training

Testing on unseen target domain
(snow)

→ The target domain is unknown during training.

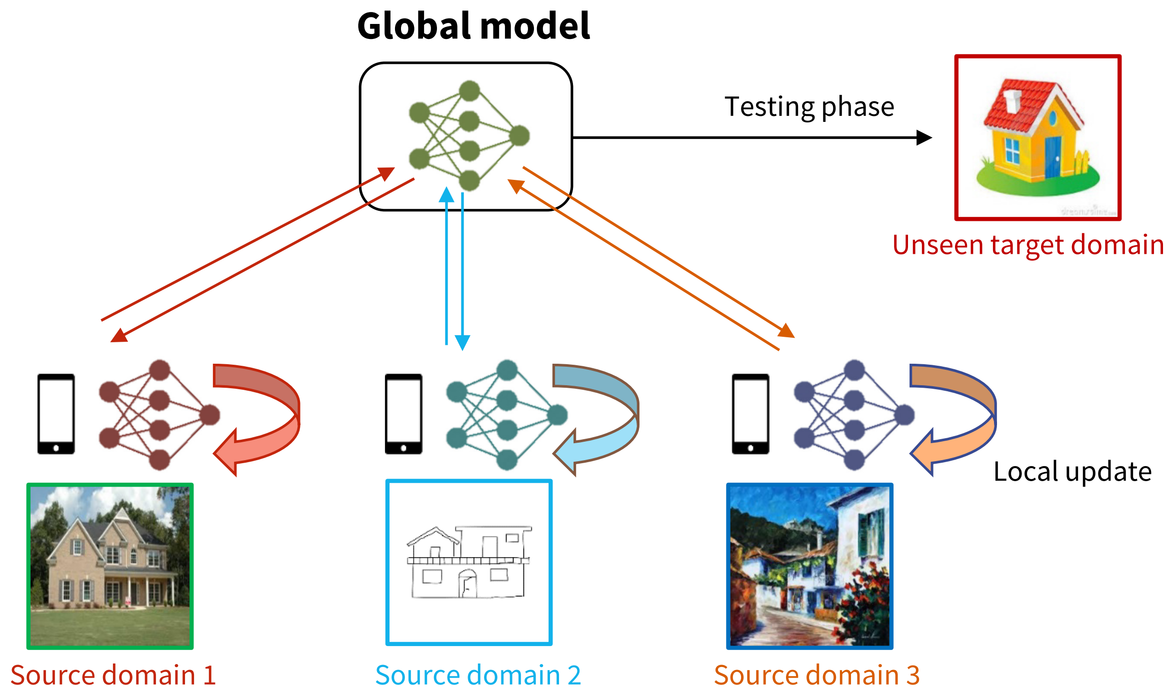
→ The model should have generalization capability on the unseen domain.

→ Solved via meta-learning, data augmentation, style augmentation ..

Background:
Federated Domain Generalization

- In real-world FL scenarios, clients have different domains and the trained global model should predict well on any unseen domains.

→ Raising an important problem: federated domain generalization.

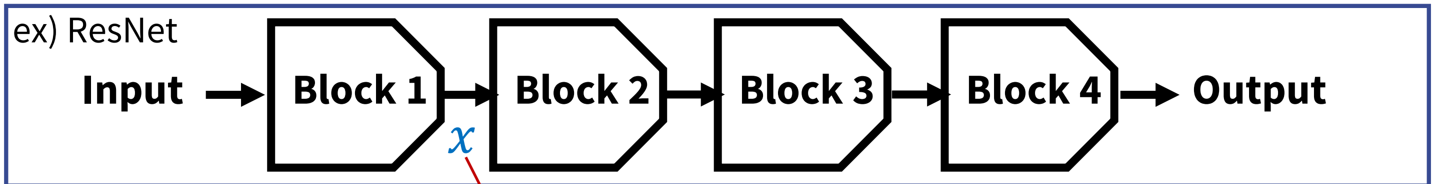


- However, federated domain generalization is challenging due to:

→ Limited data/domain in each user.

Background:
Domain Generalization via Feature Augmentation

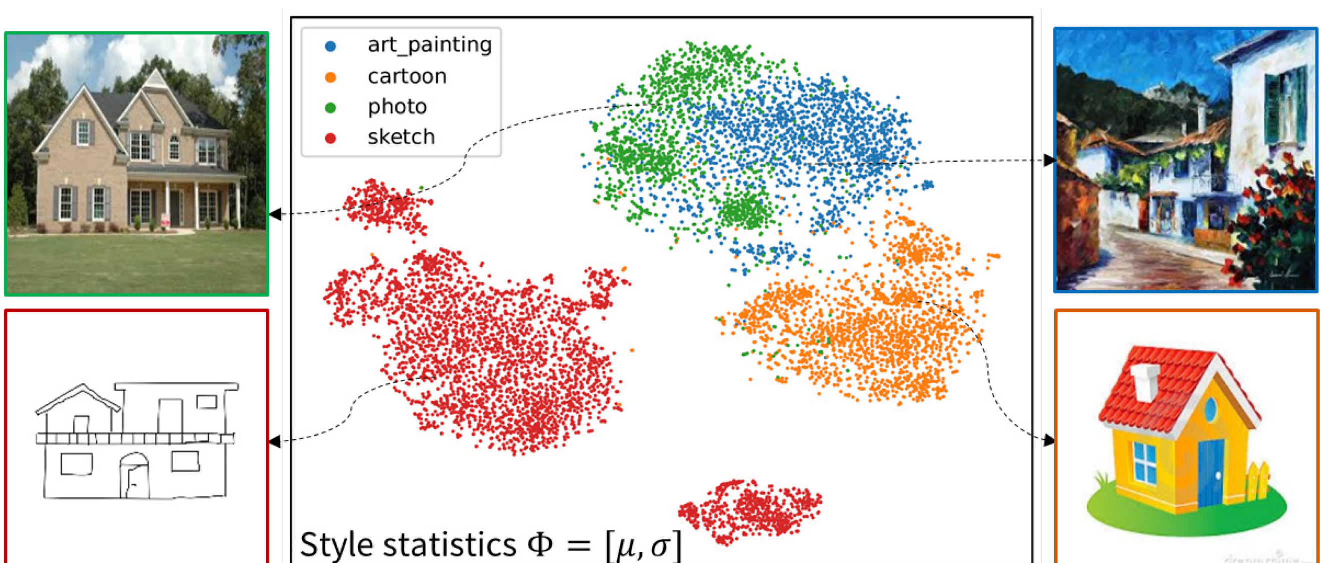
- **Key observations:** feature statistics of CNN layers capture domain information
  - Define x as the encoded feature of a specific sample at an early layer.



Channel-wise mean & variance $[\mu(x), \sigma(x)]$

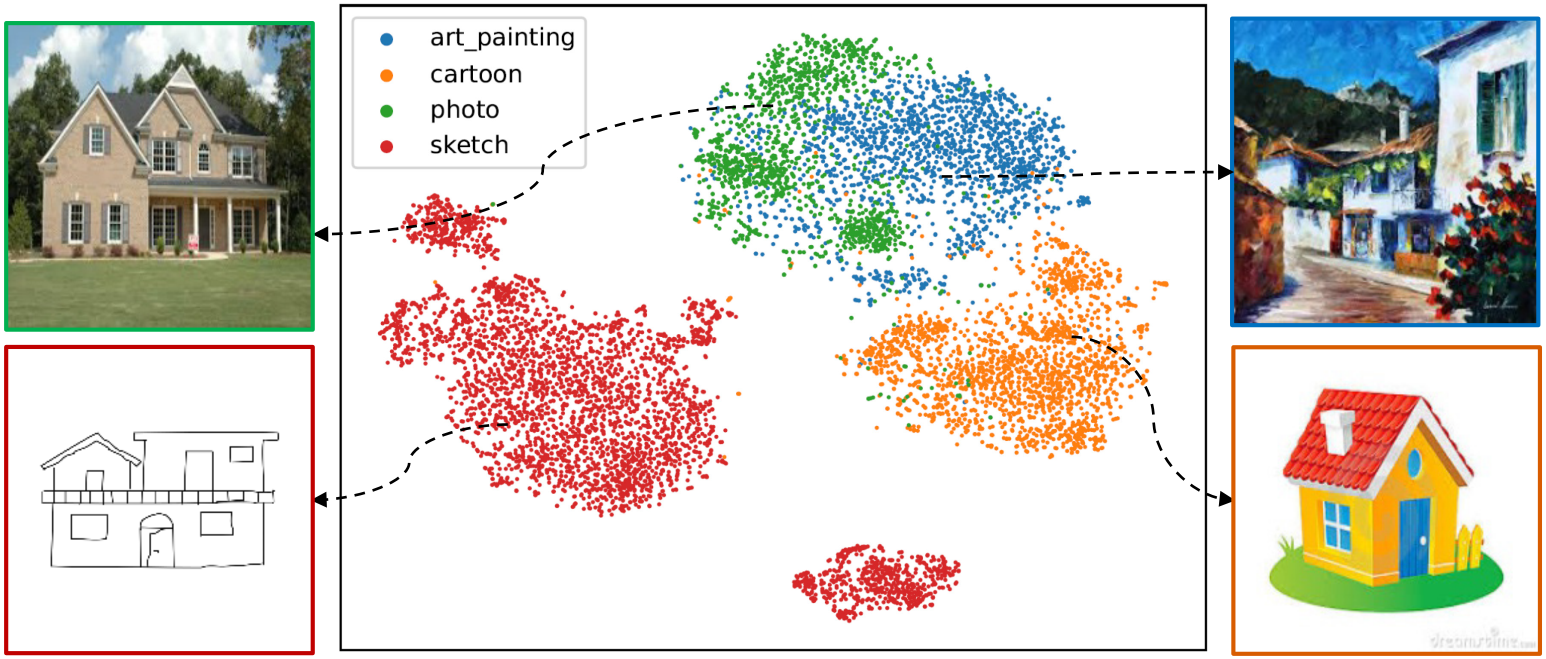
$$\mu(x)_{b,c} = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} x_{b,c,h,w}$$

$$\sigma^2(x)_{b,c} = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\left(x_{b,c,h,w}-\mu_{b,c}(x)\right)^2$$



Background:
Domain Generalization via Feature Augmentation

- AdaIN [ICCV'17]



→ Based on this observation, AdaIN proposed a new style transfer method.

Content information of x is preserved

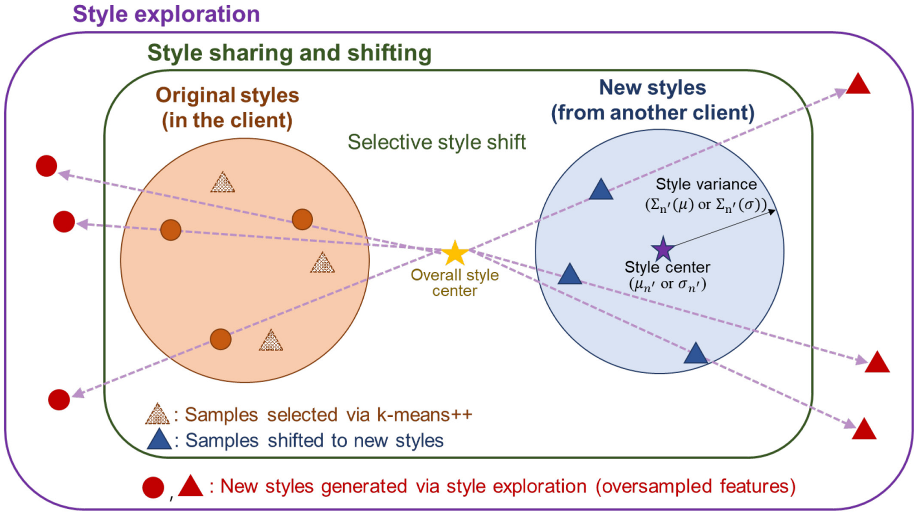
$$\mathrm{AdaIN}(x) = \underbrace{\sigma(y)}\boxed{\frac{x-\mu(x)}{\sigma(x)}}\underbrace{+\,\mu(y)}.$$

Transfer to the style of y

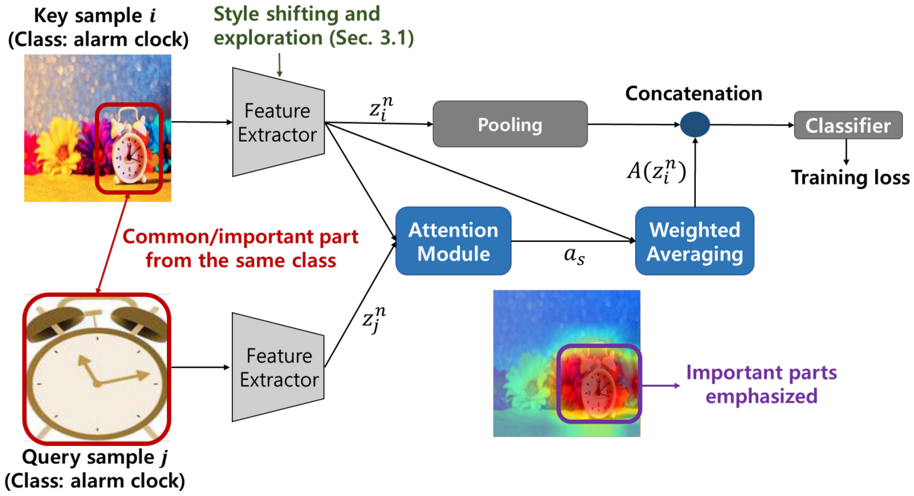
Contribution:
Overview of Approach

- Two key components to tackle federated DG

1. Style-based learning



2. Attention-based learning

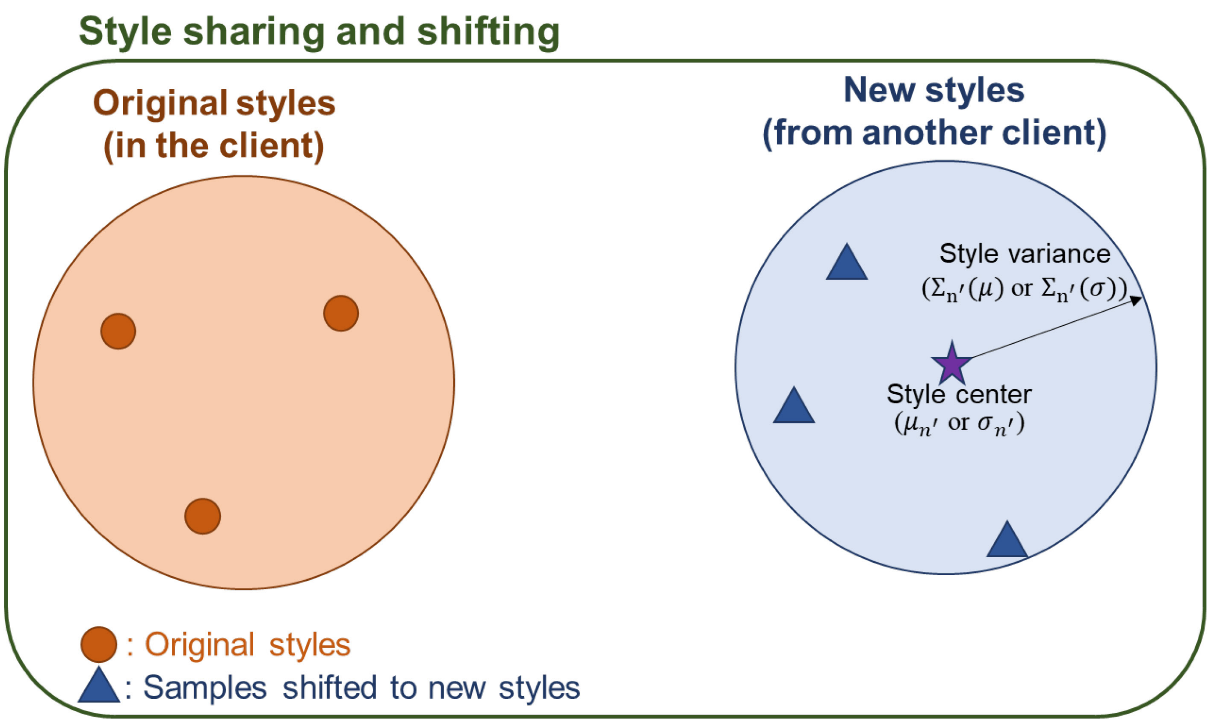


1) **Style-based learning:** Improving domain diversity (to tackle the lack of domains in each client).
2) **Attention-based learning:** Extracting common/important feature information within each class and emphasize them (to tackle the lack of data in each client).

Proposed Method:
Style-Based Learning

- Step 1: Style information sharing

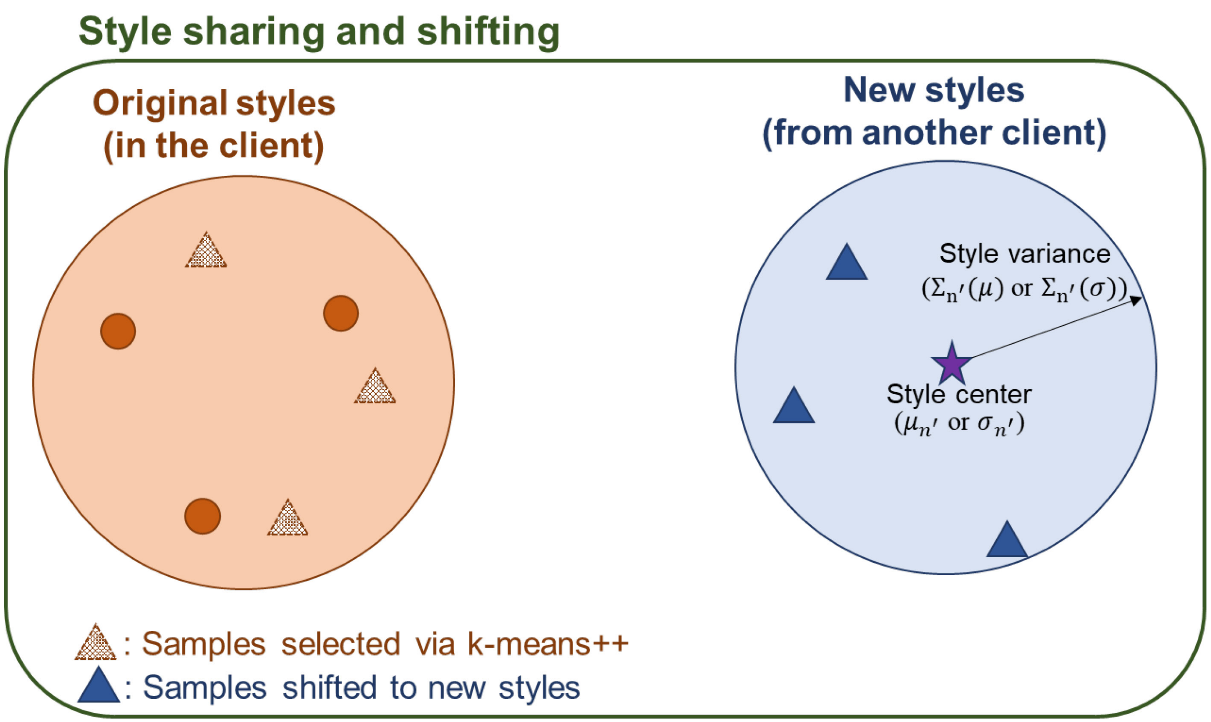
→ Share the statistic information of each client's style



Proposed Method:
Style-Based Learning

- Step 2: Selective style shifting

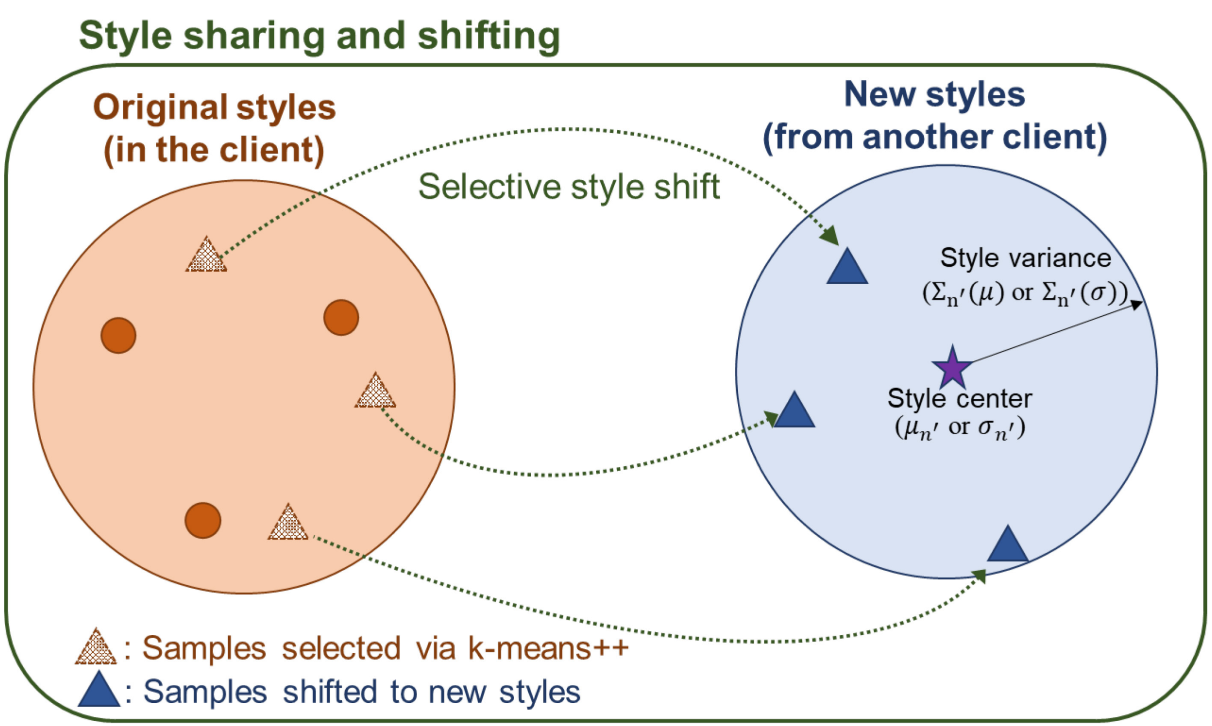
→ Choose B/2 cluster centers via k-means++



Proposed Method:
Style-Based Learning

- Step 2: Selective style shifting

→ Shift the remaining B/2 samples (not cluster centers) to the new style

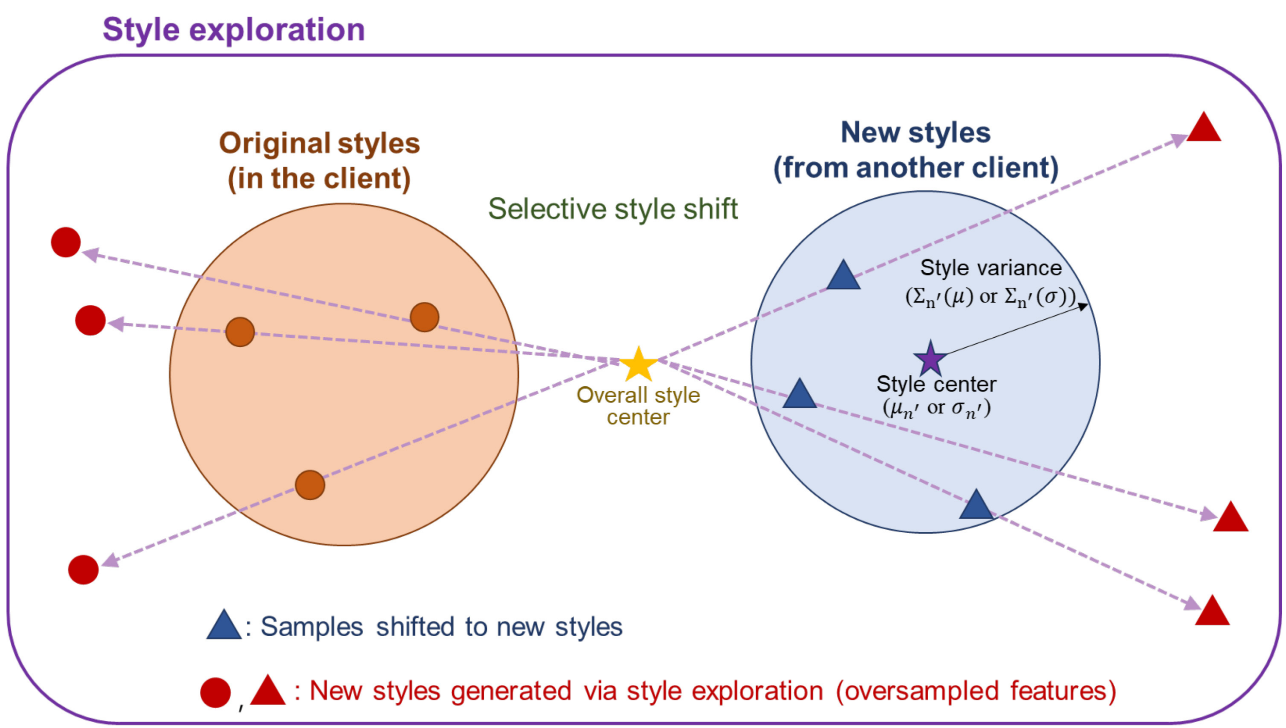


Proposed Method:
Style-Based Learning

- Step 3 & 4: Feature-level oversampling & Style exploration
  1) Perform feature-level oversampling
  2) Extend the styles of oversampled features according to the below equations

$$\mu_{new}(\tilde{s}_i^n) = \mu(\tilde{s}_i^n) + \alpha \cdot (\mu(\tilde{s}_i^n) - \mu_n(s^n)),$$
$$\sigma_{new}(\tilde{s}_i^n) = \sigma(\tilde{s}_i^n) + \alpha \cdot (\sigma(\tilde{s}_i^n) - \sigma_n(s^n)),$$

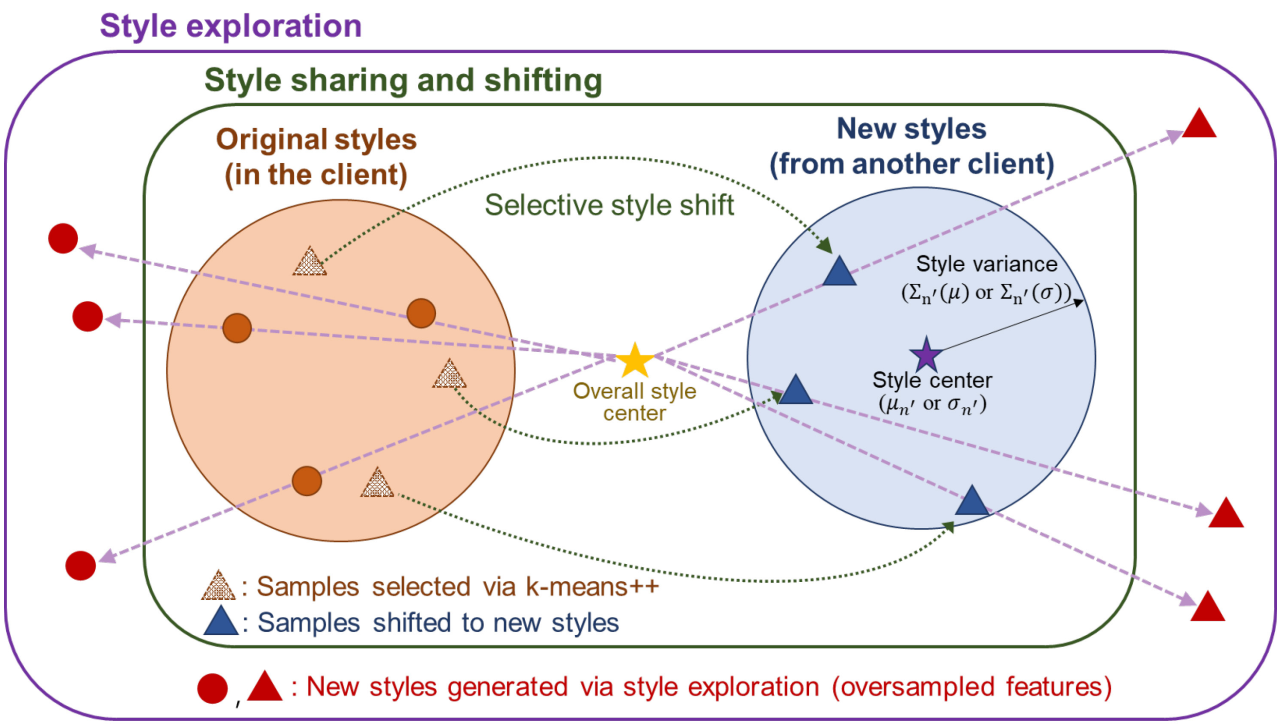
α: exploration level



Proposed Method:
Style-Based Learning

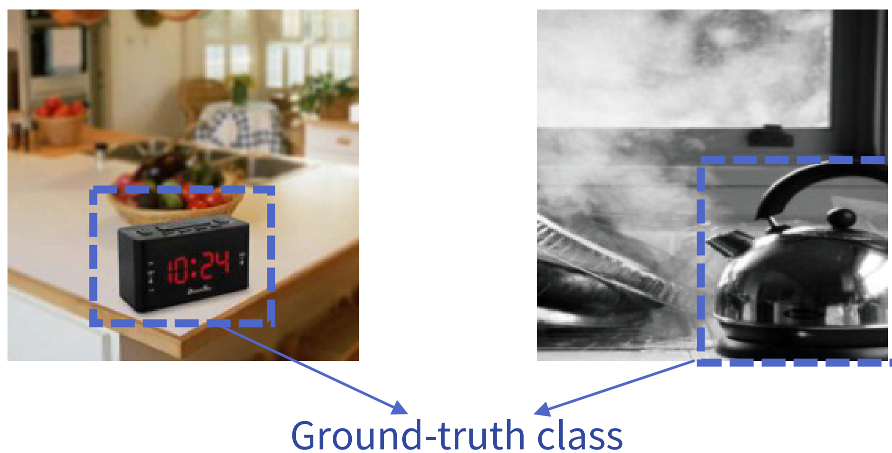
- Advantage of style-based learning

→ Each client can expose the model to diverse styles, handling the issue of the lack of styles in each FL client.



Motivation:
Attention-Based Learning

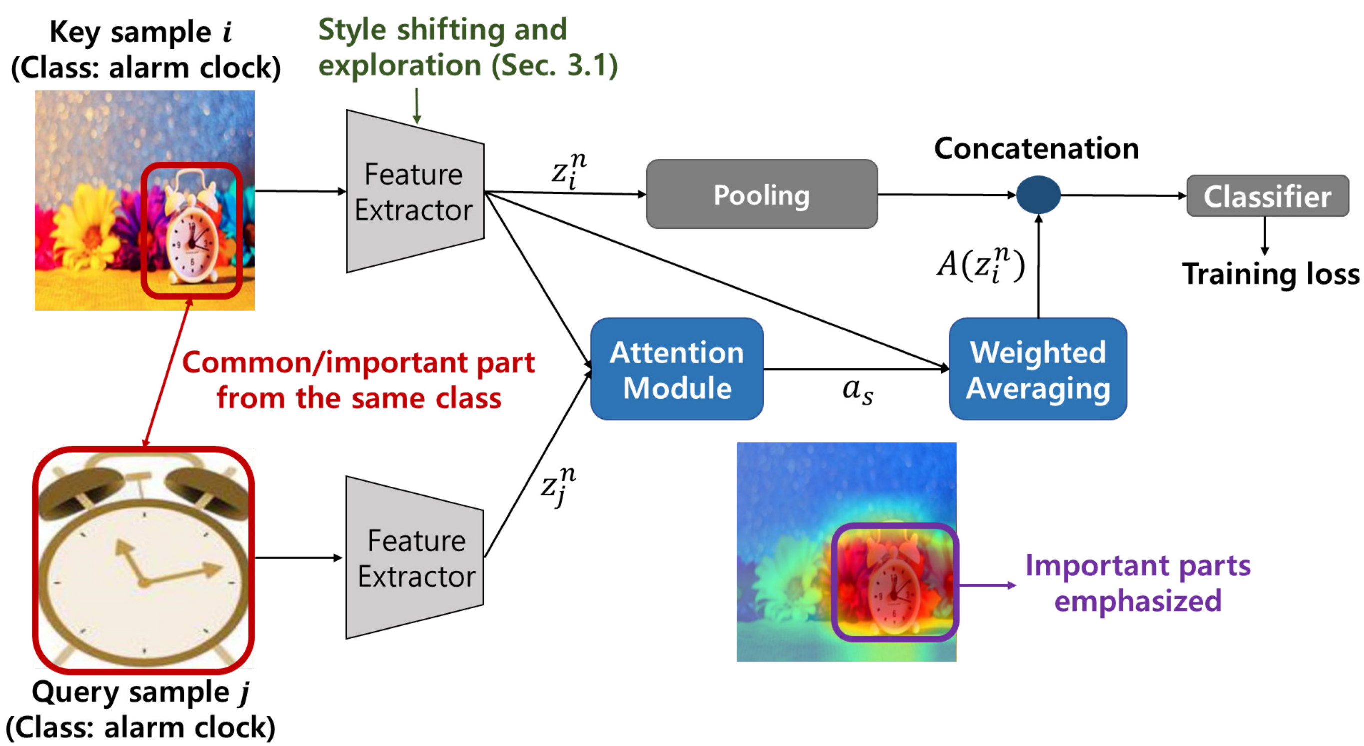
- In data-poor FL, each client has a limited number of data samples.
- In practice, each real-world image usually includes background noises.
    - E.g., images in Office-Home dataset



→ Leading to overfitting to small local datasets

Motivation:
Attention-Based Learning

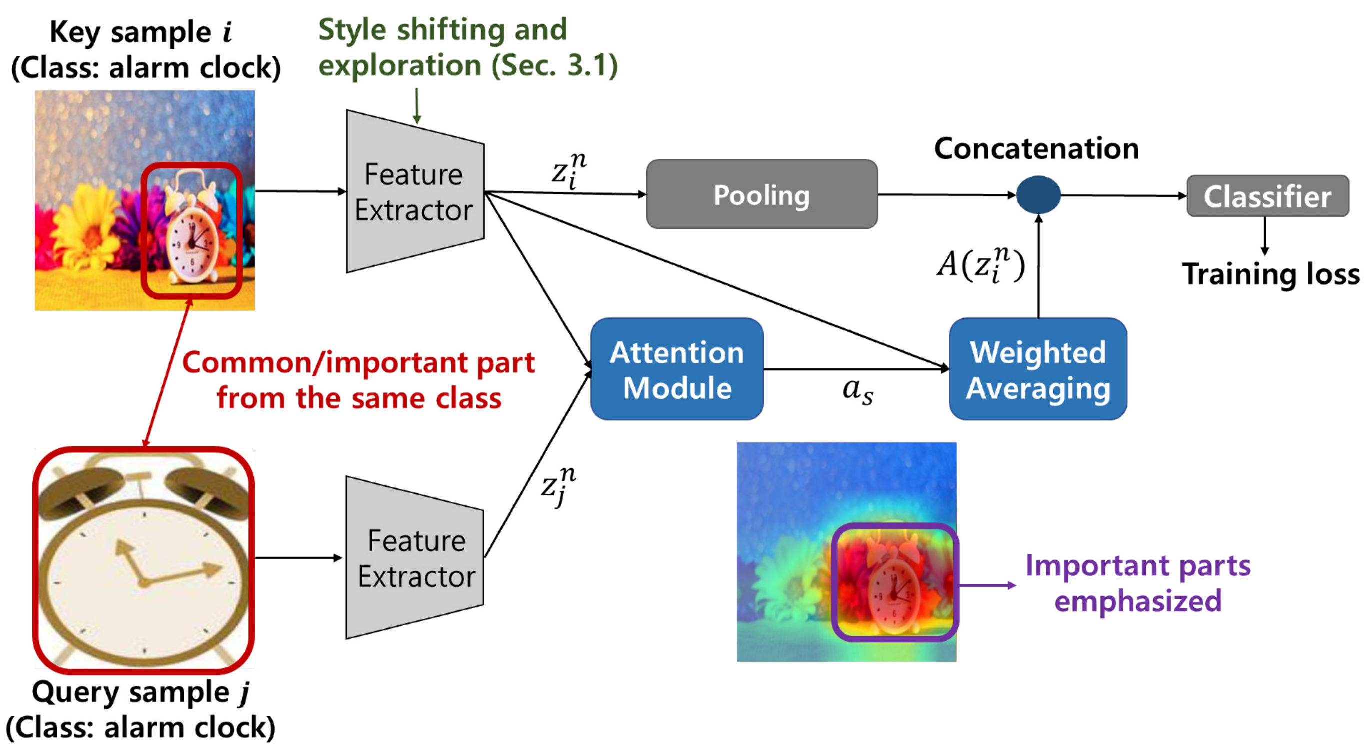
- Training: Apply both cross/self-attention to capture the common information between images from the same class.
  - Similarity function: $\mathrm{Sim}_{\mathrm{mix}}(X_i, X_j) = \left(\frac{\theta_q X_j + \theta_q X_i}{2}\right)^T (\theta_k X_i),$
- Inference
  - Similarity function: $\mathrm{Sim}(X_i, X_i)$



Motivation:
Attention-Based Learning

- Advantage of attention-based learning

→ Focus on more important features while the effect of unimportant factors such as backgrounds is effectively reduced, improving the performance.



Experimental Results

- Achievable accuracy of different schemes

Methods	PACS					VLCS				
	Art	Cartoon	Photo	Sketch	Avg.	Caltech	LabelMe	Pascal	Sun	Avg.
FedAvg [26]	73.67	70.87	90.27	55.70	72.63	93.75	59.30	70.05	69.90	73.25
FedBN [22]	78.42	70.9	90.96	54.07	73.59	94.81	58.59	72.06	70.36	73.96
MixStyle [39]	79.10	76.30	90.10	60.63	76.53	95.20	60.40	72.10	69.93	74.41
DSU [21]	80.43	75.70	92.60	69.87	79.65	96.13	58.77	71.80	71.87	74.64
CCST [2]	71.35	72.40	88.65	64.10	74.13	92.50	61.20	68.20	66.50	72.10
FedDG [25]	71.20	71.40	90.70	59.20	73.13	95.3	57.5	72.8	69.8	73.85
FedSR [28]	76.40	71.25	93.25	60.55	75.36	92.10	60.50	70.75	71.65	73.75
StableFDG (ours)	84.10	78.57	95.40	72.73	**82.70**	98.13	59.20	73.60	70.27	**75.30**

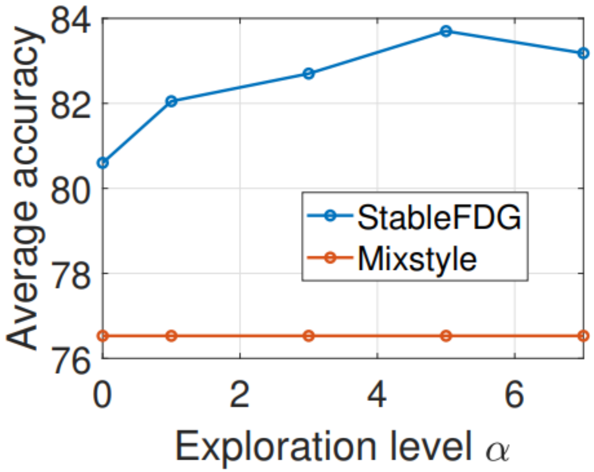
(a) PACS and VLCS datasets.

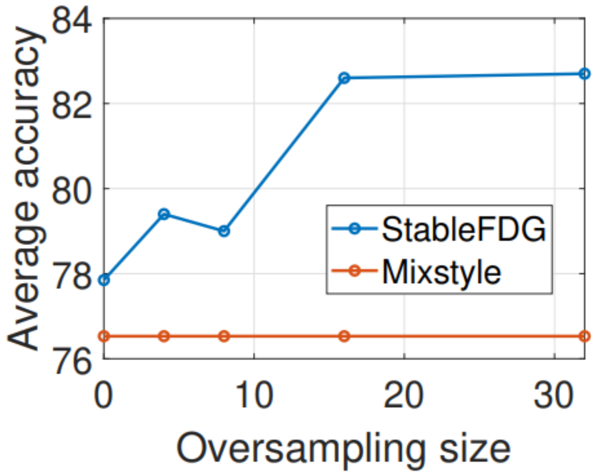
Methods	Office-Home					Digits-DG				
	Art	Clipart	Product	Real	Avg.	MNIST	MNIST-M	SVHN	SYN	Avg.
FedAvg [26]	57.27	48.23	72.77	74.60	63.22	98.05	70.95	68.95	86.40	81.09
FedBN [22]	57.56	48.13	72.65	74.57	63.23	97.33	72.68	71.77	85.36	81.79
MixStyle [39]	56.05	51.55	70.95	73.25	62.95	97.75	74.25	70.85	85.50	82.09
DSU [21]	58.55	52.60	71.60	73.15	63.98	98.10	75.60	70.47	85.80	82.49
CCST [2]	51.3	51.75	70.2	70.3	60.89	95.10	62.80	56.60	74.90	72.35
FedDG [25]	57.6	48.1	72.55	74.33	63.15	97.97	72.13	71.03	87.87	82.25
FedSR [28]	57.8	48.1	72.1	74.2	63.05	98.00	73.00	68.50	86.70	81.55
StableFDG (ours)	57.57	54.30	72.33	74.97	**64.79**	97.23	74.53	72.95	85.85	**82.64**

(b) Office-Home and Digits-DG datasets.

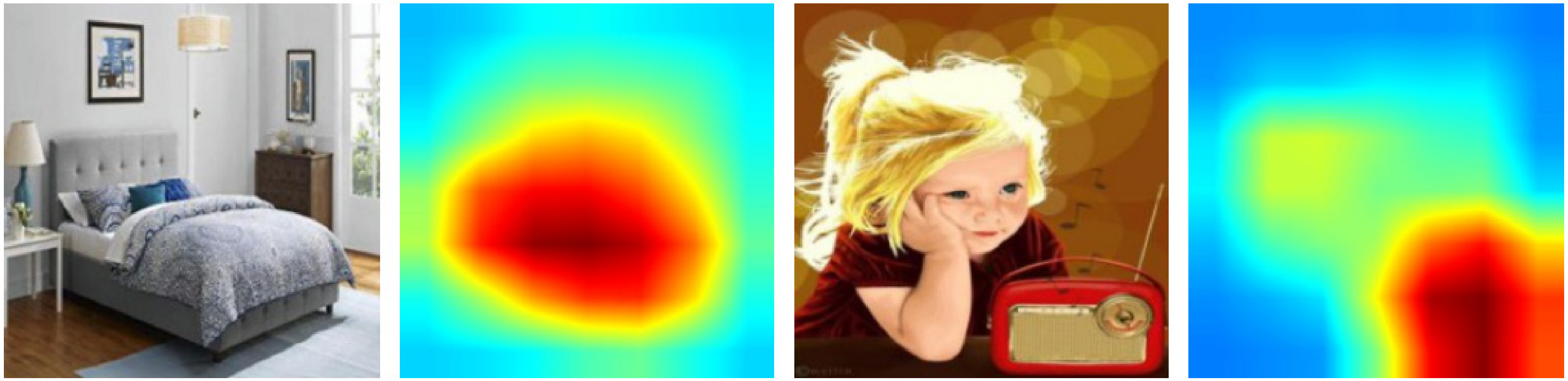
Experimental Results

- Ablation studies





- Visualization of attention score maps



Conclusion

- We proposed StableFDG, a new training strategy tailored to this unexplored area.
- Our style-based strategy enables the model to get exposed to various novel styles beyond each client's source domains.
- Our attention-based method captures and emphasizes the important / common characteristics of each class.
- We believe that our solution provides an interesting direction for DG and FedDG community in practice.

Reference

- [ICCV'17] Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization
- [ICLR'21] Domain Generalization with MixStyle
- [ICLR'22] Modeling Uncertain Feature Representation for Domain Generalization

Thank you

Any questions?

Contact:

Jungwuk Park

Email: savertm@kaist.ac.kr