Test-Time Style Shifting: Handling Arbitrary Styles in Domain Generalization

Jungwuk Park*, Dong-Jun Han*, Soyeong Kim and Jaekyun Moon

July 2023

ICML
International Conference On Machine Learning
40 Years

KAIST
School of Electrical Engineering

Background:
Domain Generalization

- Goal: Perform well on the unseen domain

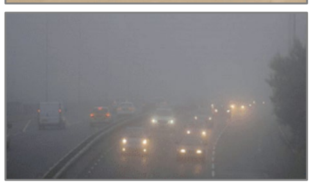
Domain 1
(rain)

D_1

Domain 2
(sand)

D_2

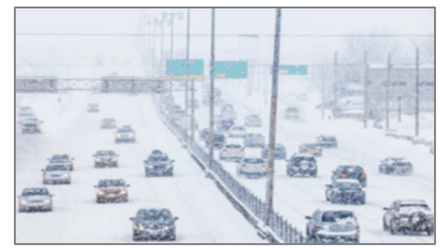
Domain 3
(fog)



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:
Domain Generalization

- Goal: Perform well on the unseen domain

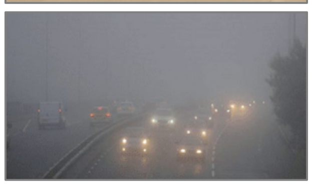
Domain 1
(rain)

D_1

Domain 2
(sand)

D_2

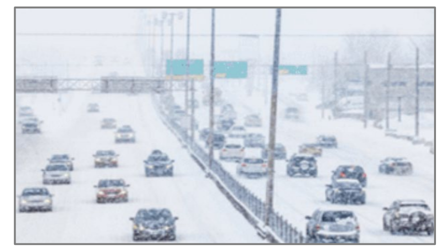
Domain 3
(fog)



D_3

Source domains

Training



Testing on unseen target domain
(snow)

→ The target domain is unknown during training

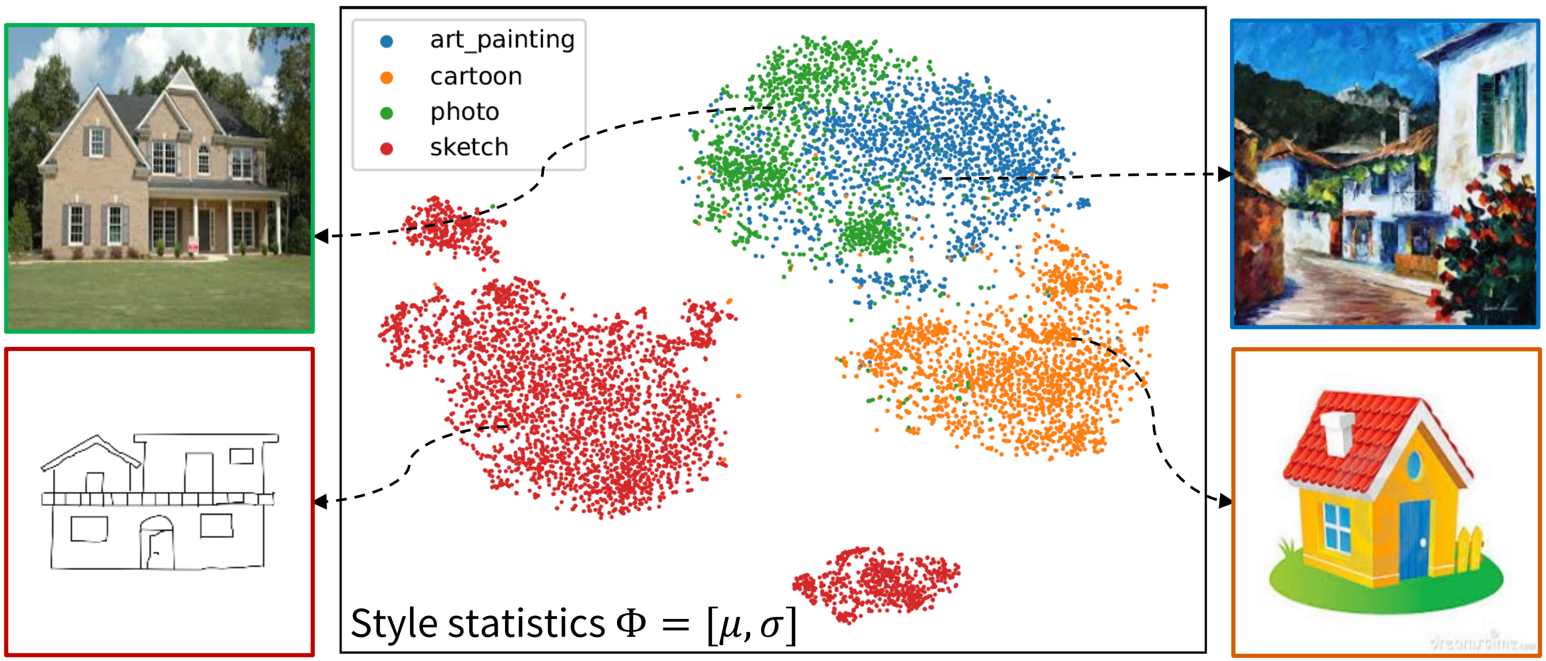
→ The model should have generalization capability on the unseen domain

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

our focus

Background:
Domain Generalization via Feature Augmentation

- t-SNE visualization of concatenated feature-level style statistics



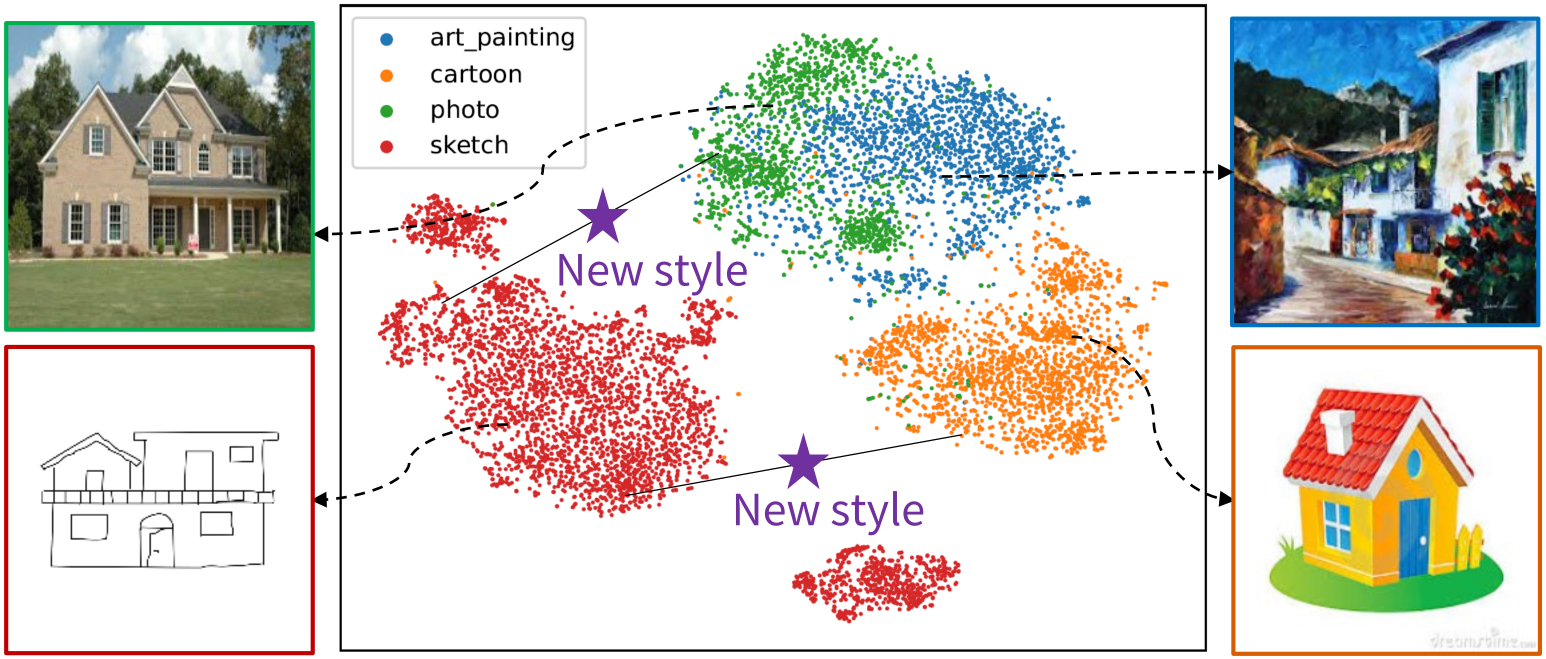
μ: channel-wise mean of the feature
σ: channel-wise standard deviation of the feature

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

→ Samples are clustered based on domain/style characteristics

Background:
Domain Generalization via Feature Augmentation

- MixStyle[ICLR'21]



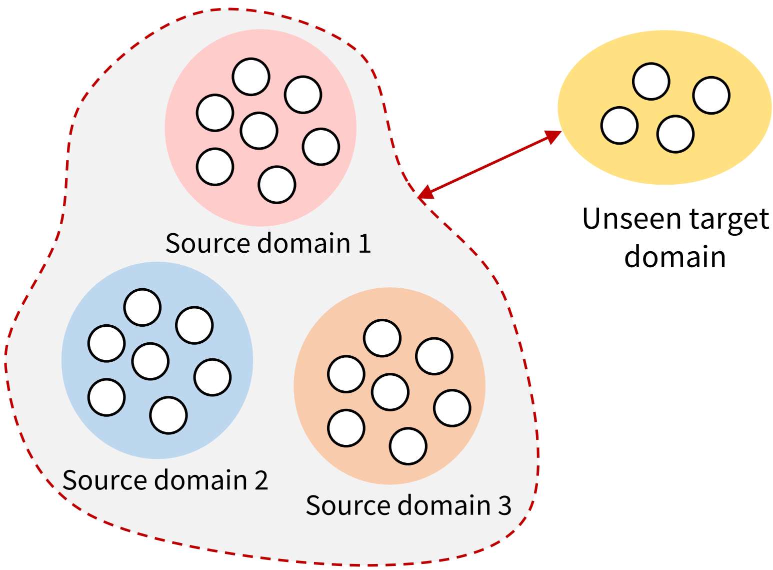
→ Generate new styles while keeping the content via AdaIN[ICCV'17]

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

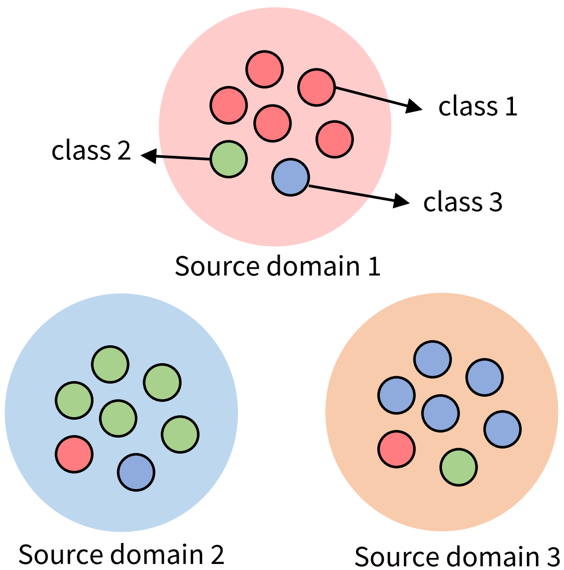
→ Improves generalization capability via style augmentation

Motivation:
Issues in Style Augmentation based Ideas

- DG is still regarded as a challenging problem due to:



Large style gap between source and target domains

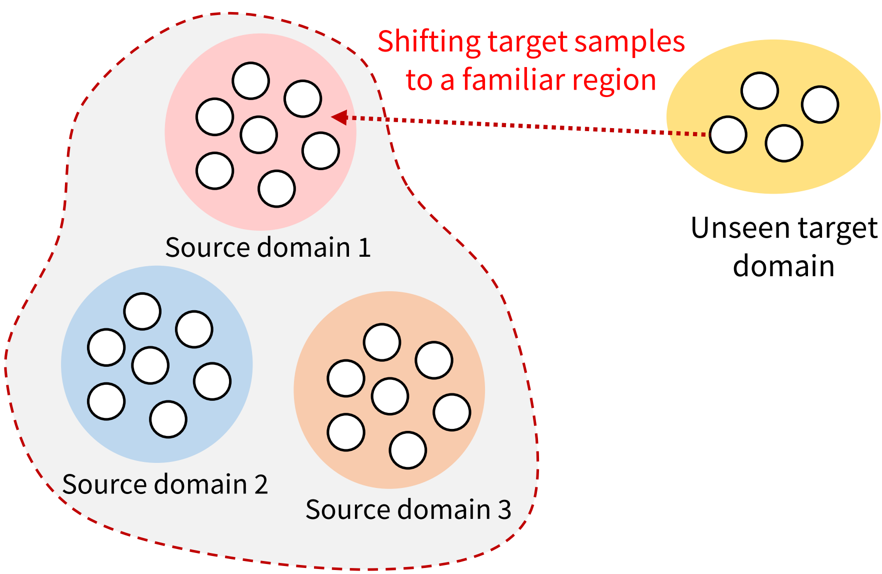


Cross-domain imbalance issue

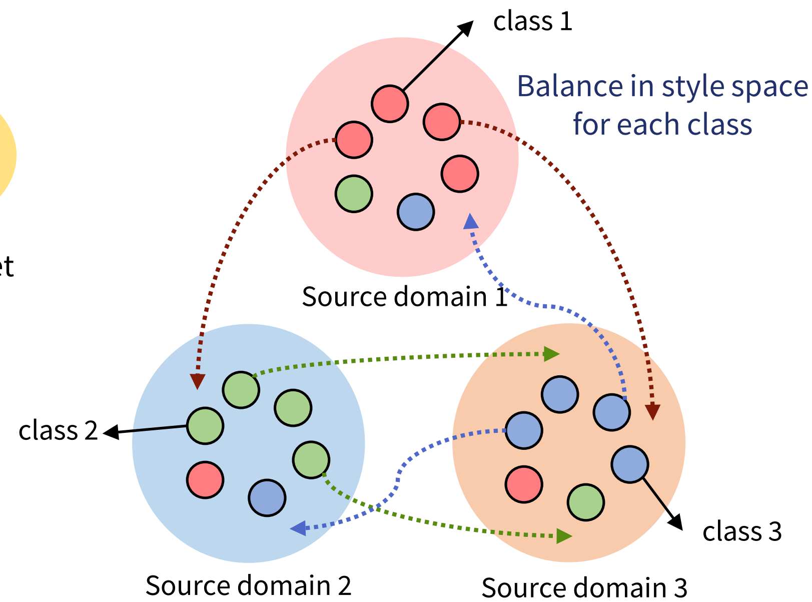
→ Lead to performance degradation

Motivation:
Contributions

- Contributions of our work



Large style gap between source and target domains
-> Test-time style shifting at testing phase

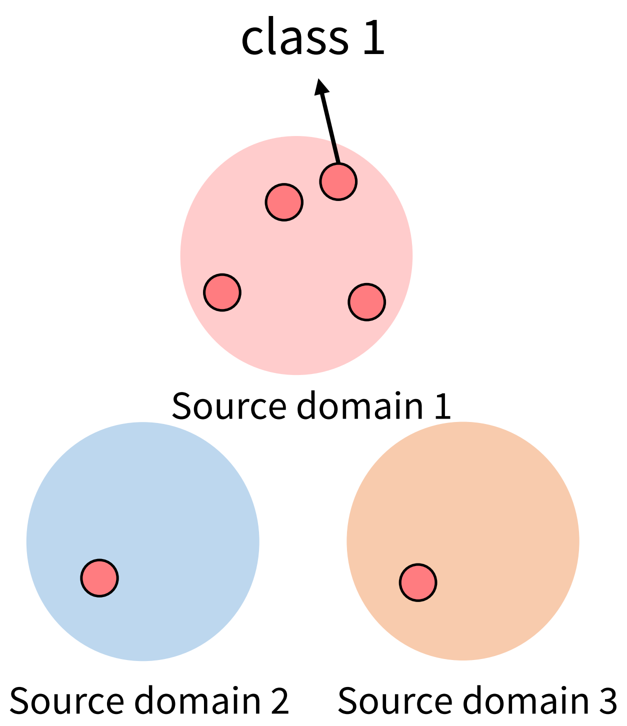


Cross-domain imbalance issue
-> Style balancing strategy at training

→ Can handle arbitrary domains while mitigating imbalance issue in DG

Proposed Idea:
Style balancing at training

➢ Process of style balancing for each class

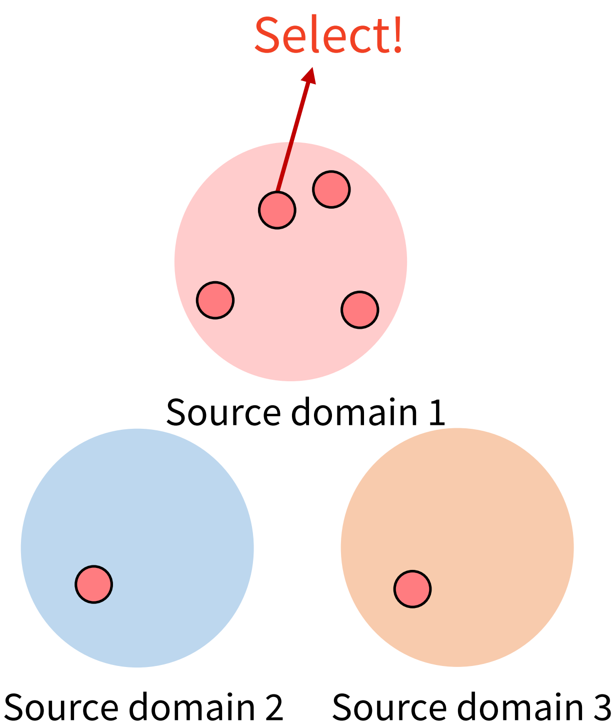


✓**Step 1:** Determining the number of samples to shift in each domain by class

❖ Two samples from domain 1 should be shifted to domain 2 and 3.

Proposed Idea:
Style balancing at training

➢ Process of style balancing for each class

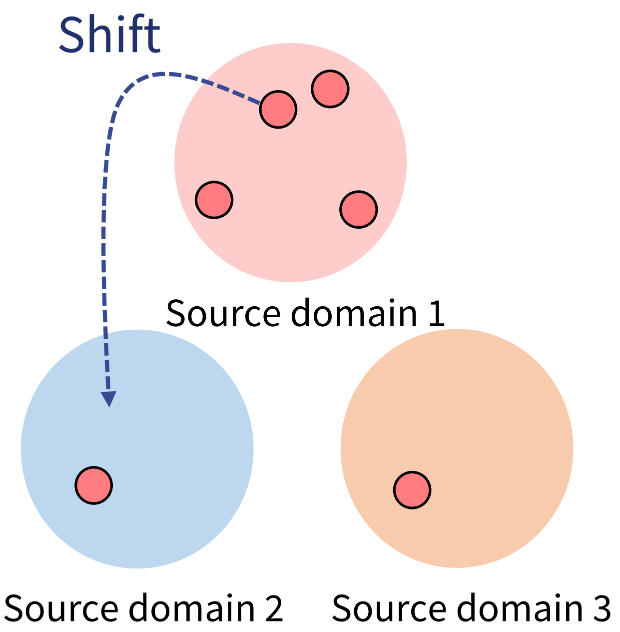


✓**Step 2:** Sample selection

❖ Based on the distance measured between two samples (i.e., $d_{i,j} = \|\Phi(f(s_i)) - \Phi(f(s_j))\|$) for all sample pairs, choose one pair that has a minimum distance and select one sample closer to all other samples.

Proposed Idea:
Style balancing at training

➢ Process of style balancing for each class



✓**Step 3:** Balancing

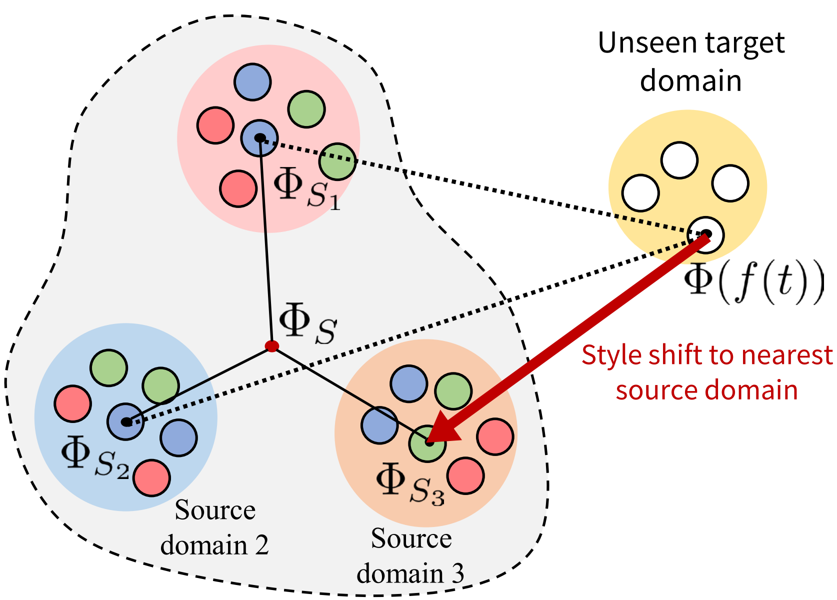
❖ Shift the selected sample to the other domain via EFDMix [CVPR'22].

→ Through this balancing process, all classes can explore more diverse styles.

Proposed Idea:
Test-time style shifting at testing

➢ Test-time style shifting

Key Idea: Shifting the styles of each target sample to the nearest source domain that the model is already familiar (well trained) with.



Test-Time Style Shifting

- Given a target sample t, new style statistic $\Phi(f(t))_{\text{new}}$ of the sample are generated by

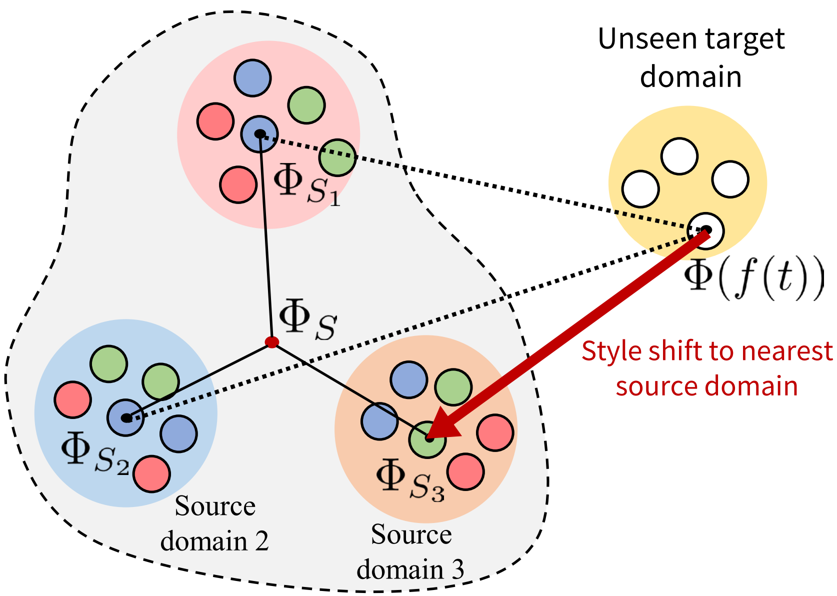
$$\begin{cases} \Phi_{S_{n'}} & \text{if } \frac{1}{N}\sum_{n=1}^{N}\|\Phi(f(t)) - \Phi_{S_n}\| > \alpha\left(\frac{1}{N}\sum_{n=1}^{N}\|\Phi_S - \Phi_{S_n}\|\right) \\ \Phi(f(t)) & \text{otherwise,} \end{cases}$$

- Where $n' = \text{argmin}_n \|\Phi(f(t)) - \Phi_{S_n}\|$, Φ_{S_i} indicates the mean of feature statistics in i-th source domain, Φ_S is the mean of feature statistics over all source domains

Proposed Idea:
Test-time style shifting at testing

➢ Test-time style shifting

Key Idea: Shifting the styles of each target sample to the nearest source domain that the model is already familiar (well trained) with.

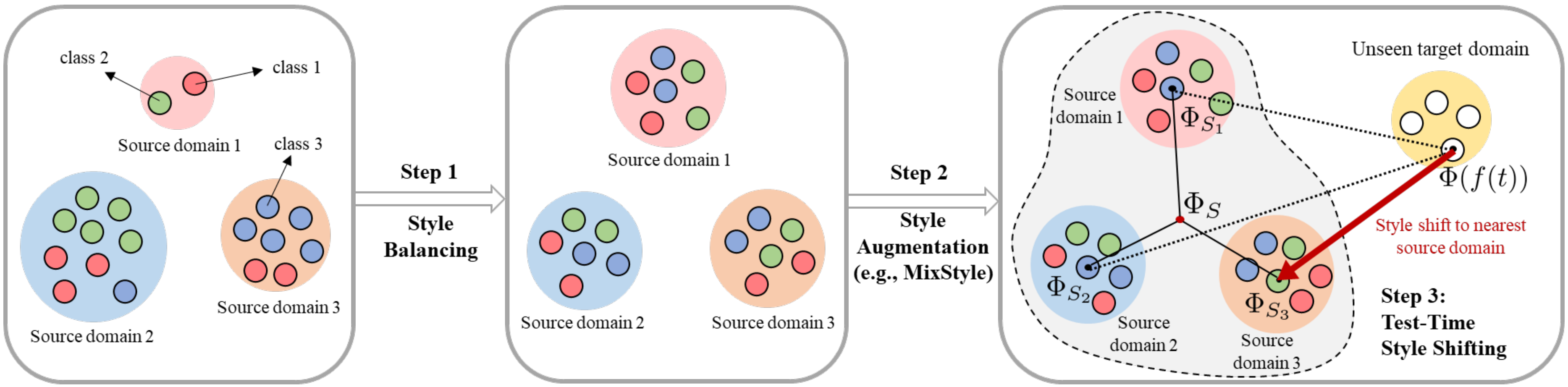


Test-Time Style Shifting

→ This enables the model to always make reliable predictions in well-trained domains at test time without any additional model update.

Proposed Idea:
Style Balancing and Test-Time Style Shifting

- Proposed method



- **Step 1:** Style Balancing

  → Balance the number of samples for each class.

- **Step 2:** DG schemes

  → Existing DG schemes such as MixStyle can be used.

- **Step 3:** Test-Time Style Shifting

  → Shift the style of each test sample to the nearest source domain.

Experimental Results

- Effect of style balancing (SB) and test-time style shifting (TS) on PACS dataset

Methods	Art	Cartoon	Photo	Sketch	Avg.
L2A-OT* (Zhou et al., 2020)	83.3	78.2	96.2	73.6	82.8
pAdaIN* (Nuriel et al., 2021)	81.74	76.91	96.29	75.13	82.51
SagNet* (Nam et al., 2021)	83.58	77.66	95.47	76.3	83.25
Tent* (Wang et al., 2020)	81.55	77.67	95.49	77.64	83.09
T3A* (Iwasawa & Matsuo, 2021)	80.4	75.2	94.7	76.5	81.7
SSG* (Xiao et al., 2022)	82.02	79.73	95.87	78.96	84.15
Baseline - ResNet18	73.97	74.71	96.07	65.71	77.62
SB (Baseline)	80.55	77.16	96.39	71.68	81.44
TS (Baseline)	73.89	75.14	95.87	72.00	79.23
TSB (Baseline)	80.60	77.58	96.35	74.37	**82.22**
MixStyle (Zhou et al., 2021)	82.54	79.42	95.88	74.06	82.98
SB (+ MixStyle)	83.48	79.07	96.15	73.74	83.11
TS (+ MixStyle)	82.59	79.99	95.88	78.66	84.28
TSB (+ MixStyle)	83.62	80.07	96.15	78.66	**84.63**
DSU (Li et al., 2022)	81.78	78.66	95.91	76.75	83.27
SB (+ DSU)	80.98	79.61	95.95	78.66	83.80
TS (+ DSU)	81.12	80.31	95.82	79.19	84.11
TSB (+ DSU)	80.73	80.69	95.83	79.47	**84.18**
EFDMix (Zhang et al., 2022)	83.12	79.76	96.43	75.08	83.60
SB (+ EFDMix)	83.98	79.75	96.47	75.12	83.83
TS (+ EFDMix)	83.05	81.31	96.40	78.93	84.92
TSB (+ EFDMix)	84.00	80.72	96.46	78.85	**85.00**

Experimental Results

- Results on imbalanced PACS dataset
  - Notably, in domain and class imbalance scenarios, our style balancing module effectively plays an important role to resolve the imbalance issues.

Methods	Reference	Cross-domain data imbalance					Cross-domain class imbalance				
		Art	Cartoon	Photo	Sketch	Avg.	Art	Cartoon	Photo	Sketch	Avg.
MixStyle	ICLR'21	71.73	73.80	90.60	66.48	75.65	39.91	54.08	56.45	44.82	48.82
SB (+ MixStyle)	Ours	76.53	75.61	93.33	68.34	78.45	44.49	55.57	56.28	44.93	50.32
TS (+ MixStyle)	Ours	72.04	74.01	90.60	75.12	77.94	39.98	54.01	56.45	44.44	48.74
TSB (+ MixStyle)	Ours	76.97	76.62	93.29	75.88	**80.69**	44.50	55.84	56.28	46.68	**50.83**
DSU	ICLR'22	75.76	75.26	91.90	72.45	78.84	29.61	45.24	46.90	39.37	40.28
SB (+ DSU)	Ours	76.04	76.15	92.87	73.47	79.64	45.09	53.93	60.25	47.74	51.75
TS (+ DSU)	Ours	75.49	76.69	91.92	76.36	80.12	29.78	44.54	46.90	36.65	39.47
TSB (+ DSU)	Ours	75.93	77.39	92.85	75.90	**80.52**	45.03	54.42	60.24	49.20	**52.22**
EFDMix	CVPR'22	75.33	75.67	90.59	71.07	78.16	44.68	54.87	58.15	44.64	50.59
SB (+ EFDMix)	Ours	77.91	76.38	92.79	70.99	79.52	46.63	54.84	57.89	44.47	50.96
TS (+ EFDMix)	Ours	75.39	75.92	90.56	74.97	79.21	44.56	55.05	58.15	45.96	50.93
TSB (+ EFDMix)	Ours	77.90	76.54	92.71	76.37	**80.88**	46.03	55.29	57.87	49.99	**52.30**

Experimental Results

- Results on imbalanced VLCS (left) and results in a single-domain generalization setup on PACS dataset (right).
  - Our TS significantly boosts up the performance of existing methods in a single-DG setup (right).

Methods	Caltech	LabelMe	Pascal	Sun	Avg.
MixStyle	68.87	53.32	55.12	39.09	54.10
SB (+ MixStyle)	69.97	53.87	55.51	38.51	54.47
TS (+ MixStyle)	73.51	53.20	55.15	38.98	<u>**55.21**</u>
TSB (+ MixStyle)	73.27	53.78	55.02	38.58	55.16
DSU	63.07	54.13	56.01	39.90	53.28
SB (+ DSU)	74.02	53.40	55.91	40.22	55.89
TS (+ DSU)	65.99	53.90	55.93	40.02	53.96
TSB (+ DSU)	75.99	53.50	55.46	40.28	<u>**56.31**</u>

Methods	Art	Cartoon	Photo	Sketch	Avg.
MixStyle	64.32	71.77	42.98	32.18	52.81
TS (+ MixStyle)	72.19	77.25	48.50	43.62	<u>**60.39**</u>
DSU	64.85	74.53	39.48	36.20	53.77
TS (+ DSU)	70.99	73.95	51.18	49.03	<u>**61.28**</u>
EFDMix	66.56	73.93	44.74	36.36	55.40
TS (+ EFDMix)	73.87	76.79	53.04	49.41	<u>**63.28**</u>

- Results on person re-ID task using Market1501 and GRID datasets.
  - Our TSB can be applied to various tasks and bring performance improvements.

Methods	Reference	Market → GRID				GRID → Market			
		mAP	R1	R5	R10	mAP	R1	R5	R10
MixStyle (Zhou et al., 2021)	ICLR'21	35.30	26.67	**44.53**	53.07	5.25	16.40	30.05	37.05
TSB (+ MixStyle)	Ours	**36.30**	**28.27**	42.93	**55.47**	**5.70**	**17.75**	**31.90**	**39.65**
DSU (Li et al., 2022)	ICLR'22	38.57	30.40	46.40	53.07	4.45	14.90	27.65	34.60
TSB (+ DSU)	Ours	**40.10**	**30.67**	**48.00**	**58.13**	**5.25**	**16.70**	**31.60**	**38.85**
EFDMix (Zhang et al., 2022)	CVPR'22	36.33	**27.47**	45.87	52.27	6.07	19.27	33.70	41.30
TSB (+ EFDMix)	Ours	**36.67**	26.93	**46.67**	**55.57**	**6.53**	**20.23**	**35.37**	**43.13**

Conclusion

- We propose two effective strategies to handle the issues in domain generalization
- Test-time style shifting: handles any target domains with arbitrary styles.
- Style balancing: increases the potential of test-time style shifting while handling the DG-specific imbalance issues.
- We believe that our solution provides a new guideline for DG in practice with imbalance and domain shift issues.

Thank you

Any questions?

Contact:

Jungwuk Park

Email: savertm@kaist.ac.kr