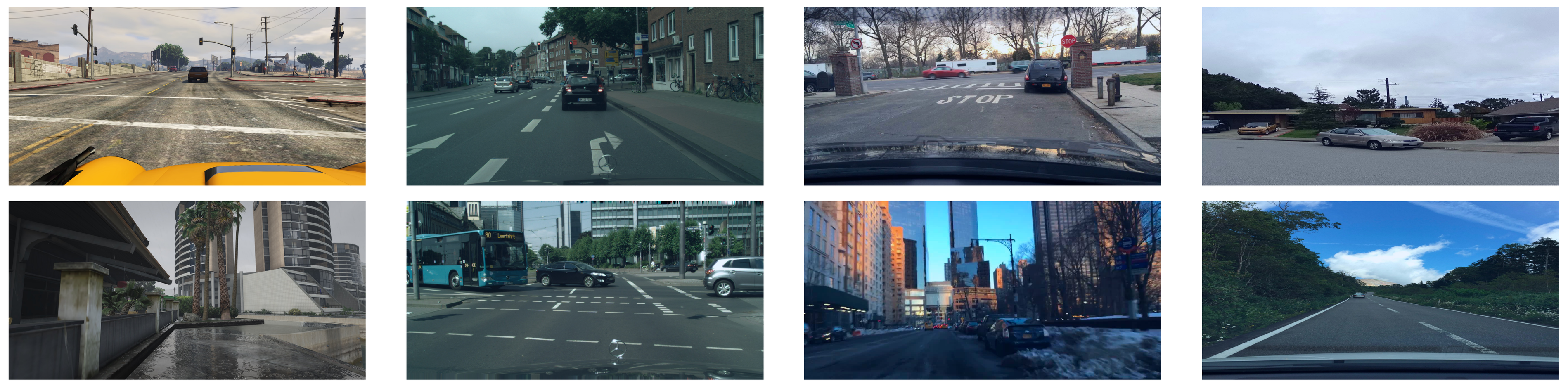
Style-Hallucinated Dual Consistency Learning for Domain Generalized Semantic Segmentation

Yuyang Zhao[1] Zhun Zhong[2] Na Zhao[1] Nicu Sebe[2] Gim Hee Lee[1]
[1] National University of Singapore [2] University of Trento

ECCV TEL AVIV 2022

Paper and code are available:
https://github.com/HeliosZhao/SHADE

Domain Generalization

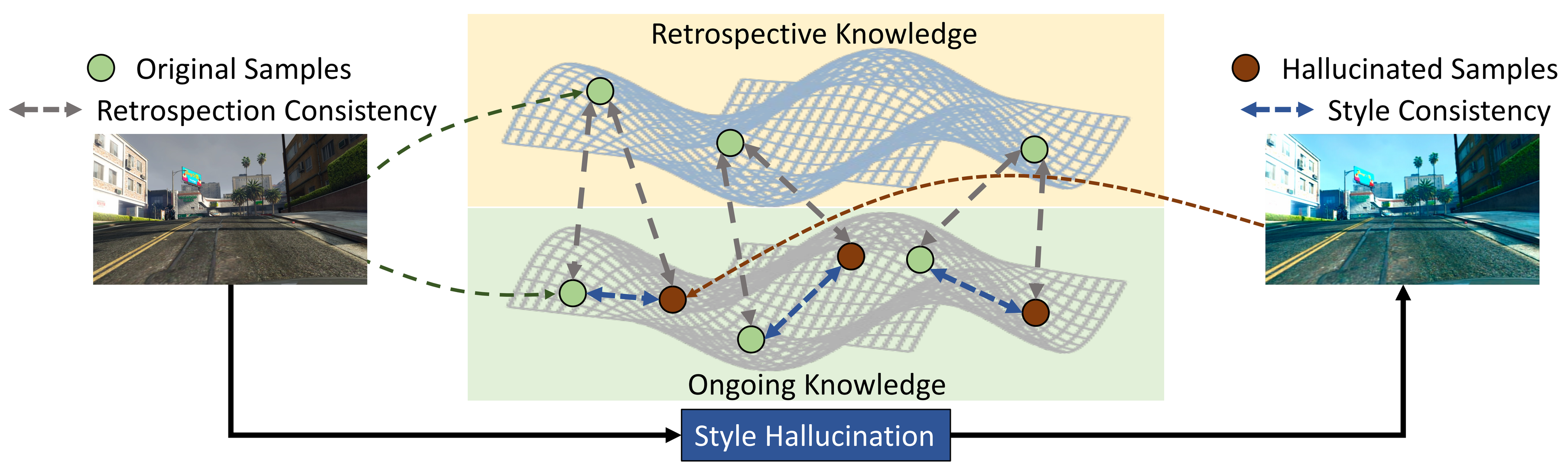


GTAV CityScapes BDD100K Mapillary

Domain Shift:
- Limited source variations.
- Distribution gap between source and target.

Contributions:
- Dual consistency learning constraints for learning style invariant representation and narrowing domain gap.
- Style hallucination module to generate new and diverse styles with the representative basis styles.

Dual Consistency Learning



Style Consistency:
- Learning style invariant representation

$$\mathcal{L}_{SC}(x_S, \tilde{x}_S) = JSD\left(p(x_S); p(\tilde{x}_S)\right) = \frac{1}{2}\left(D[p(x_S)||Q] + D[p(\tilde{x}_S)||Q]\right).$$

Retrospective Consistency:
- Narrowing the domain gap between synthetic and real data

$$\mathcal{L}_{RC}(x_S, \tilde{x}_S) = \frac{1}{\sum_m M_{things}^{(m)}} \sum_m M_{things}^{(m)} \cdot \left(\left(f(x_S; \theta_S)^{(m)} - f(x_S; \theta_{IN})^{(m)} \right)^2 + \left(f(\tilde{x}_S; \theta_S)^{(m)} - f(x_S; \theta_{IN})^{(m)} \right)^2 \right).$$

Style Hallucination Module

Style hallucination module aims to generate new styles:
- Generate diverse styles by linearly combining basis styles.
- Alleviate dominant styles by selecting basis styles with FPS.

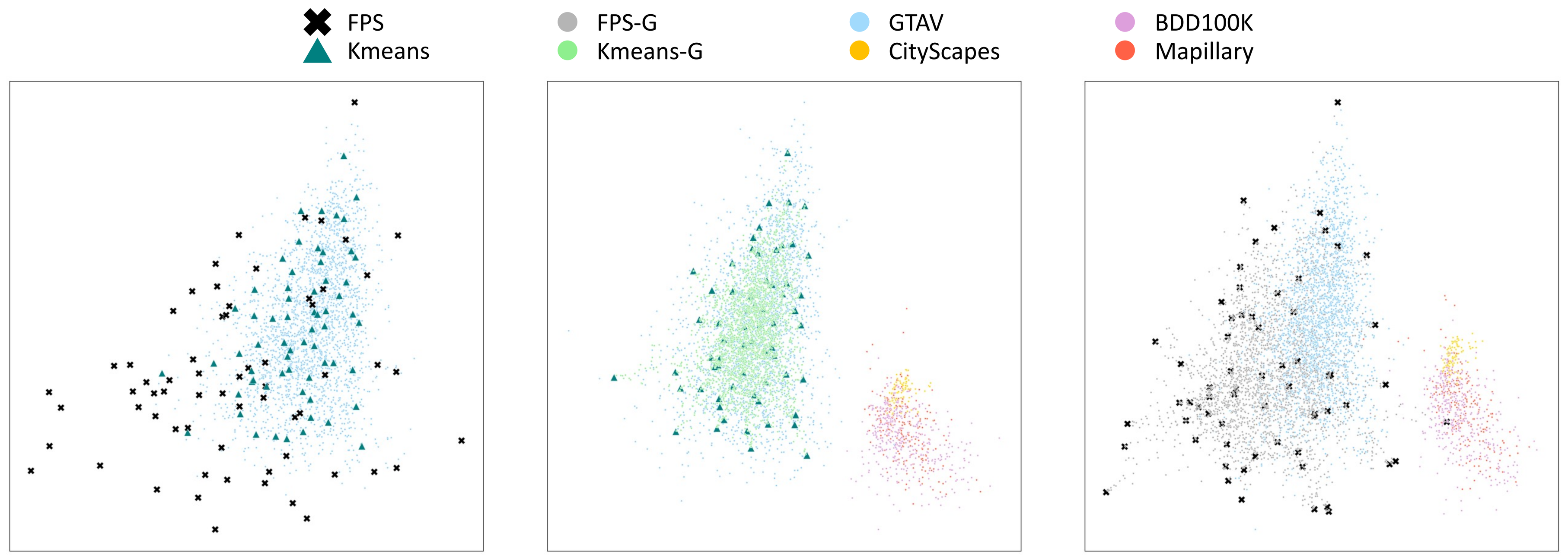
Linear combination for a new style:

$$\mu_{HS} = W \cdot \mu_{base}, \qquad \sigma_{HS} = W \cdot \sigma_{base}.$$

Generate a new sample with AdaIN:

$$\tilde{x}_S = \sigma_{HS}\left(\frac{x_S - \mu(x_S)}{\sigma(x_S)}\right) + \mu_{HS}.$$

Selection of basis style:
- Farthest point sampling (FPS): selecting new points that are the most distant points with respect to the remaining points.
- Kmeans: clustering points based on the relative distance.



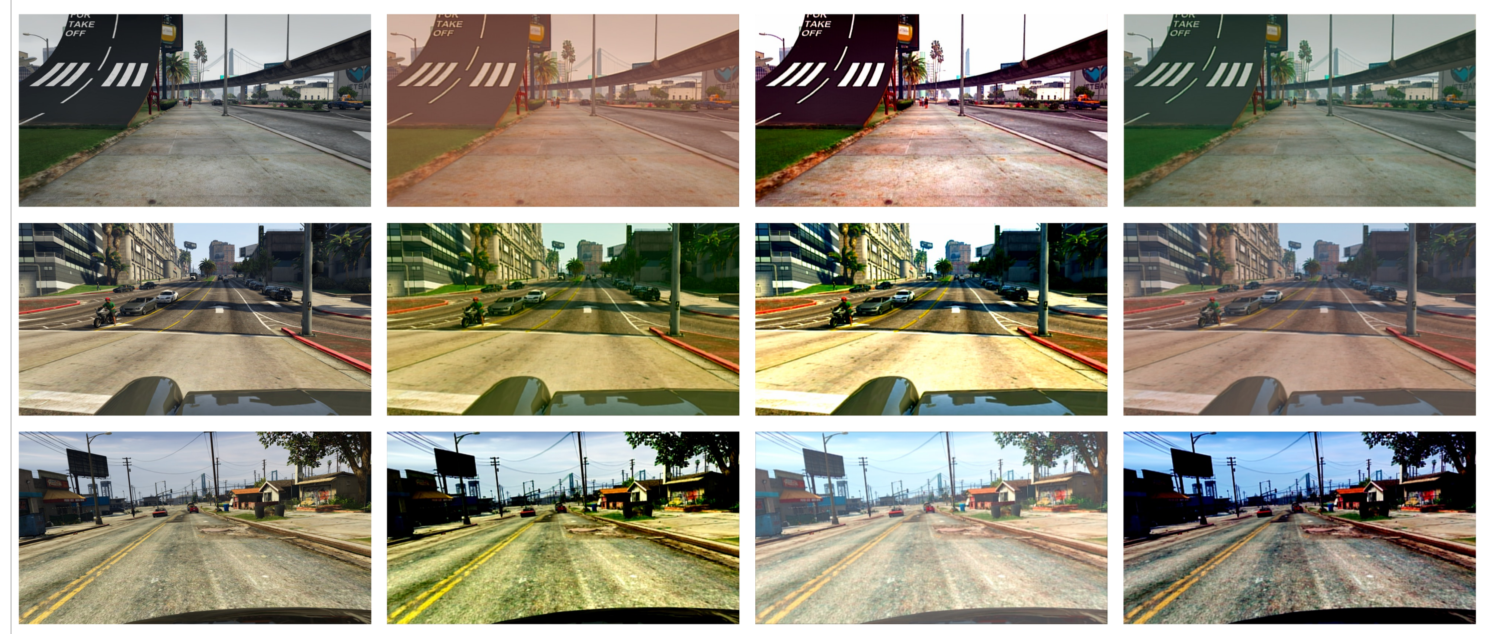
(a) Basis Style (b) Kmeans Generation (c) FPS Generation

Advantages of FPS over Kmeans:
- Rare styles are commonly far from the dominant styles, and thus can be chosen.
- Represent the style space and enlarge the source distribution.

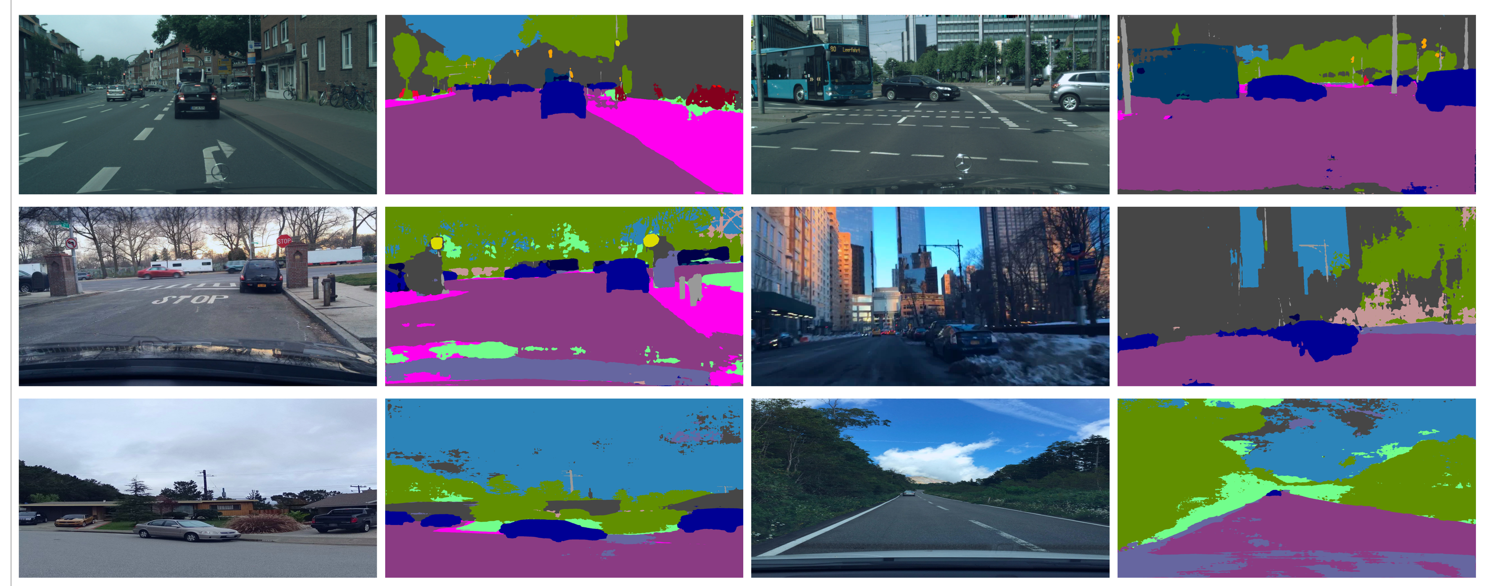
Comparison of different style variation methods

Methods (GTAV)	CityScapes	BDD100K	Mapillary	Mean
Baseline	28.95	25.14	28.18	27.42
Random Style	37.99	37.63	38.06	37.89
MixStyle [44]	43.14	37.94	42.22	41.10
CrossNorm [33]	43.13	37.20	41.83	40.72
Kmeans Basis	40.50	37.62	39.46	39.19
Ours	**44.65**	**39.28**	**43.34**	**42.42**

Hallucinated Samples



Segmentation Results



Quantitative Results

Performance on single-source DG benchmark

Net	Methods (GTAV)	Extra Data	CityScapes	BDD100K	Mapillary	Mean
ResNet-50	Baseline	✗	28.95	25.14	28.18	27.42
	SW [27]	✗	29.91	27.48	29.71	29.03
	IterNorm [15]	✗	31.81	32.70	33.88	32.79
	IBN-Net [26]	✗	33.85	32.30	37.75	34.63
	DRPC [41]§	✓	37.42	32.14	34.12	34.56
	ISW [4]	✗	36.58	35.20	40.33	37.37
	Ours	✗	**44.65**	**39.28**	**43.34**	**42.42**
ResNet-101	Baseline	✗	32.97	30.77	30.68	31.47
	IBN-Net [26]	✗	37.37	34.21	36.81	36.13
	ISW [4]	✗	37.20	33.36	35.57	35.38
	DRPC [41]§	✓	42.53	38.72	38.05	39.77
	FSDR [14]§	✓	44.80	41.20	43.40	43.13
	Ours	✗	**46.66**	**43.66**	**45.50**	**45.27**

Performance on multi-source DG benchmark

Methods (G+S)	CityScapes	BDD100K	Mapillary	Mean
Baseline	35.46	25.09	31.94	30.83
IBN-Net [26]	35.55	32.18	38.09	35.27
ISW [4]	37.69	34.09	38.49	36.76
Ours	**47.43**	**40.30**	**47.60**	**45.11**

Ablation studies

SHM	$\mathcal{L}_{SC}$	$\mathcal{L}_{RC}$	EMA	CityScapes	BDD100K	Mapillary	Mean
✗	✗	✗	✗	28.95	25.14	28.18	27.42
✓	✗	✗	✗	38.68	32.40	35.96	35.68
✓	✓	✗	✗	42.66	35.92	40.42	39.67
✓	✗	✓	✗	41.43	37.65	41.77	40.29
✓	✓	✗	✓	42.38	38.04	42.34	40.92
✓	✓	✓	✗	**44.65**	**39.28**	**43.34**	**42.42**