Adversarial Teacher-Student Representation Learning for Domain Generalization

NeurIPS 2021 Spotlight Presentation

Fu-En Yang, Yuan-Chia Cheng, Zu-Yun Shiau, Yu-Chiang Frank Wang

Vision and Learning Lab, National Taiwan University

ASUS Intelligent Cloud Services (AICS)

Computer Vision in Real World

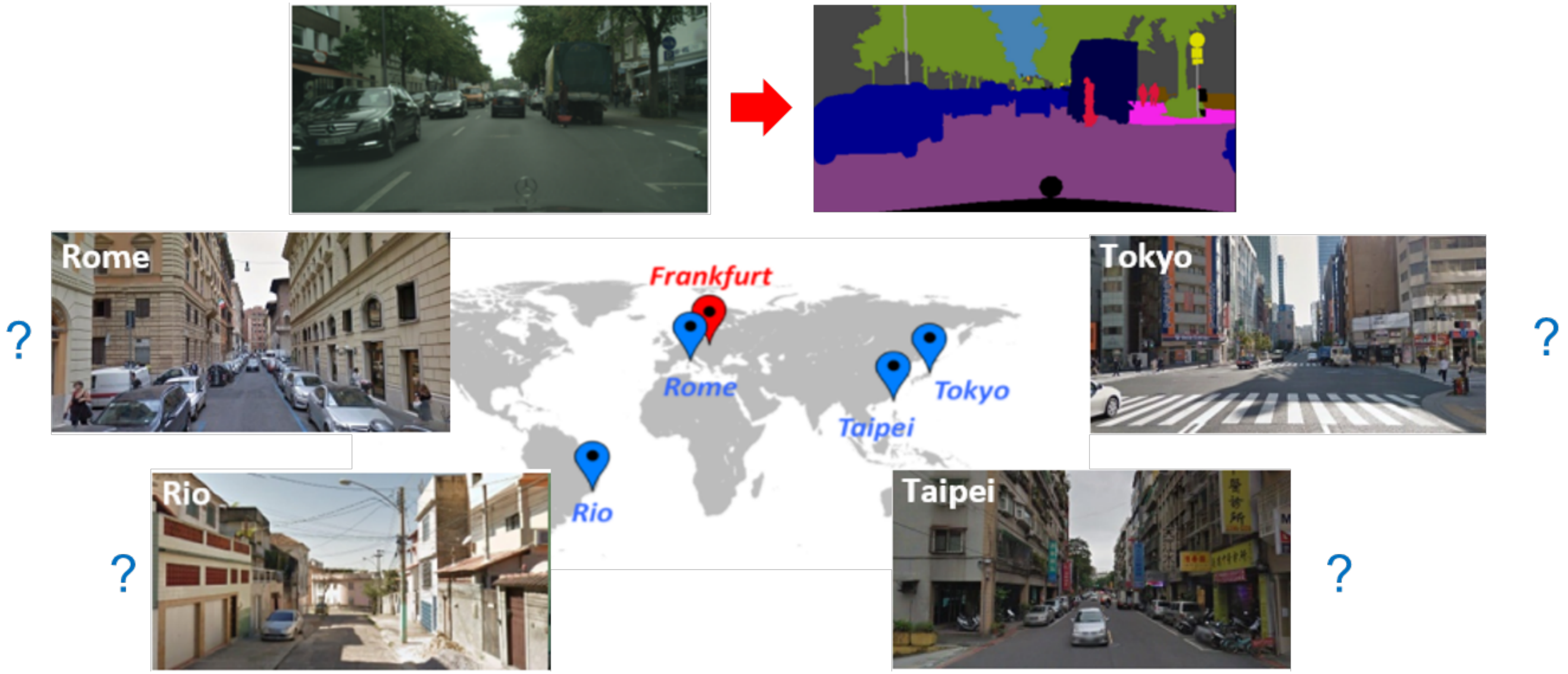
- Image Classification

Computer Vision in Real World

- Person Re-ID

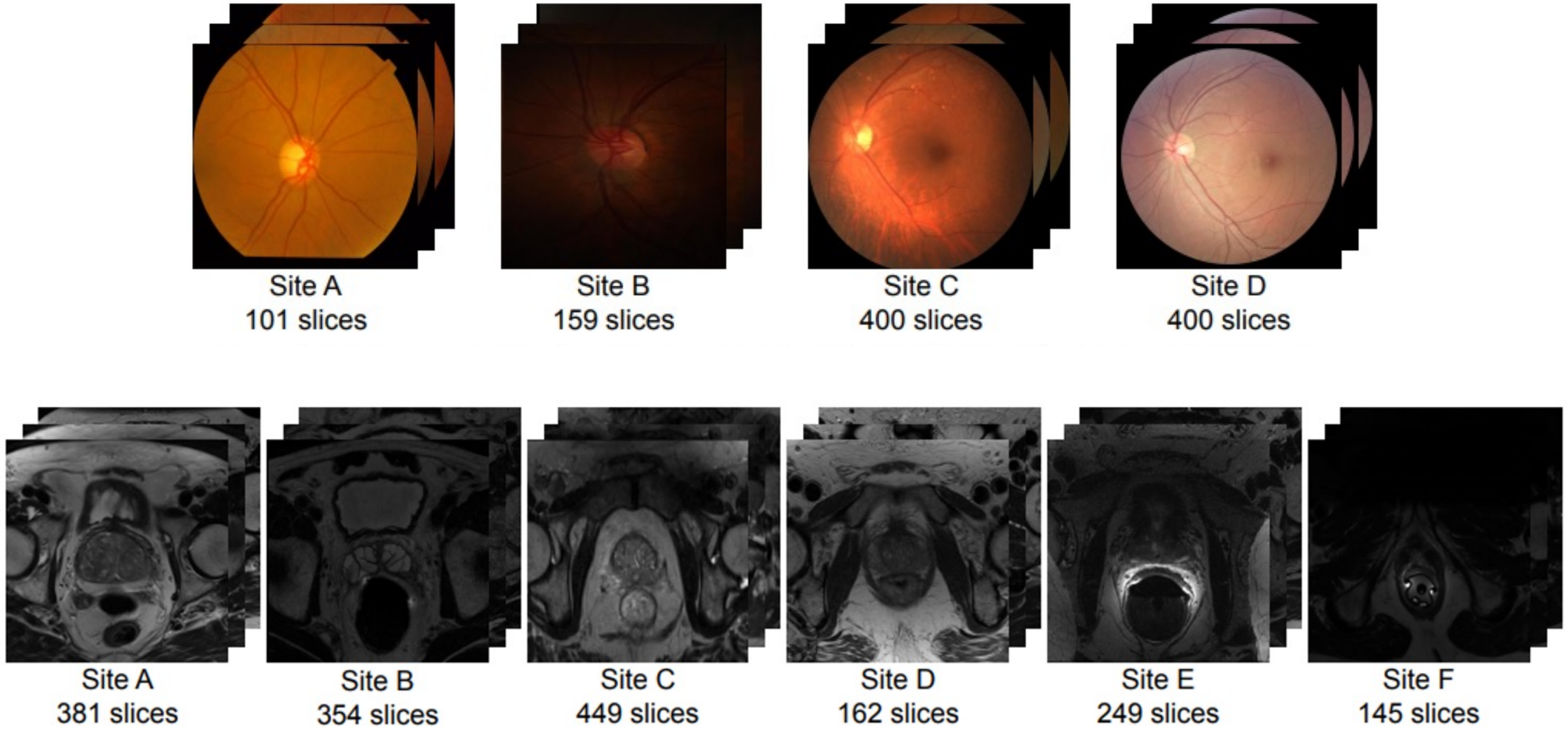
Computer Vision in Real World

- Scene Segmentation



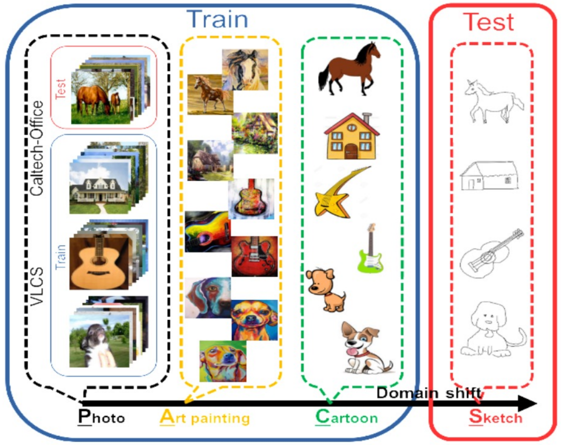
Computer Vision in Real World

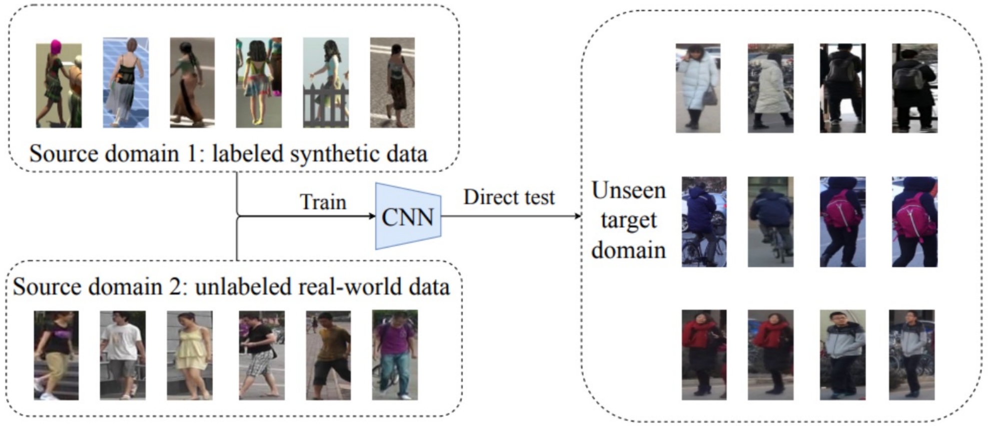
- Medical Image Segmentation



Computer Vision in Real World

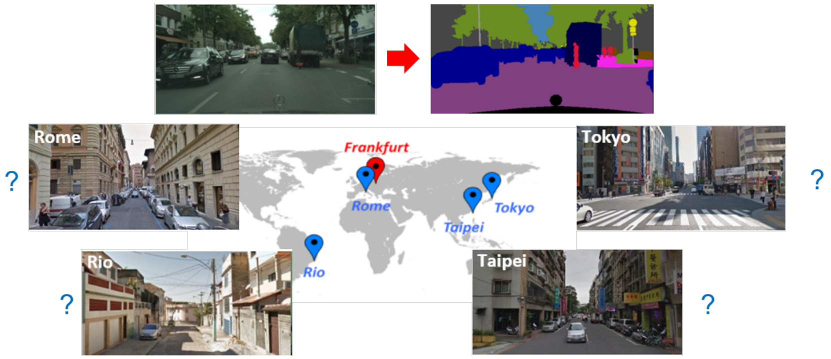
Image Classification (Li *et al.* ICCV'17)

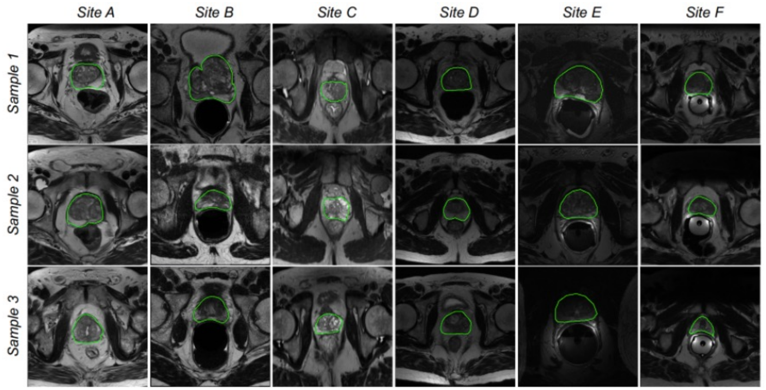




Person Re-ID (Wang *et al.* arXiv'21)

Scene Segmentation (Shiau *et al.* ICIP'21)

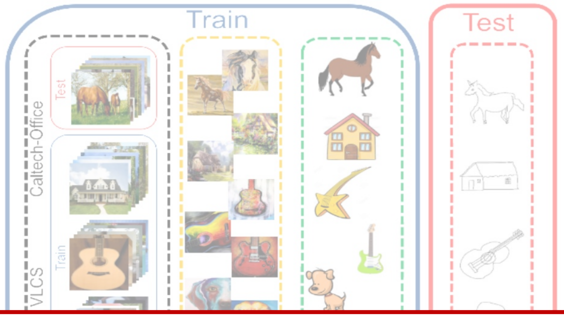


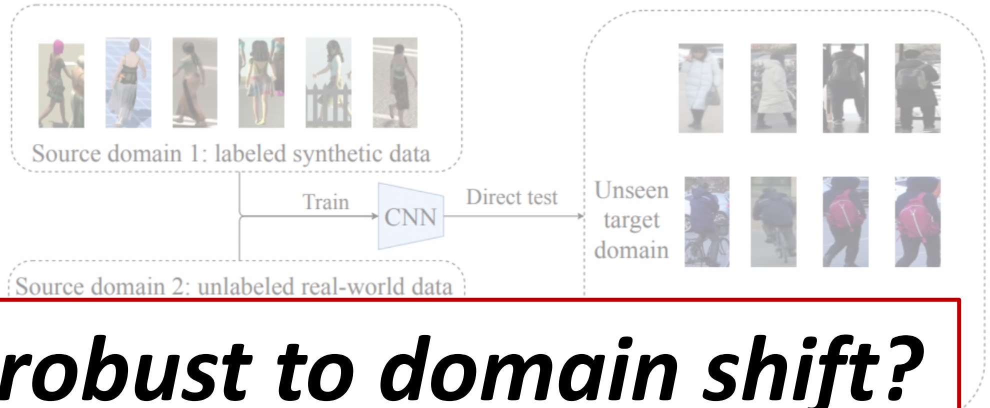


Medical Image Segmentation (Liu *et al.* MICCAI'20)

Computer Vision in Real World

Image Classification (Li *et al.* ICCV'17)

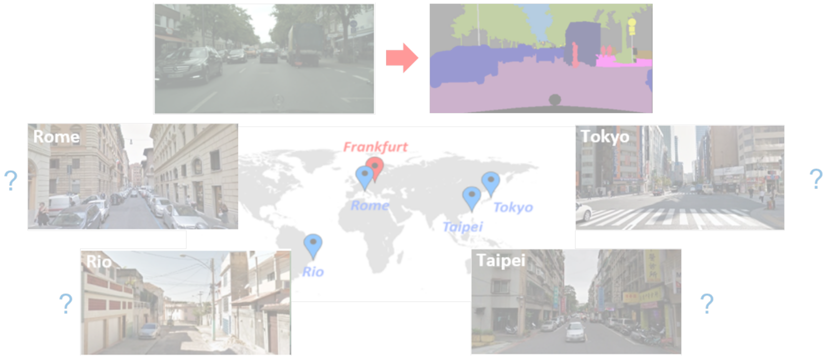


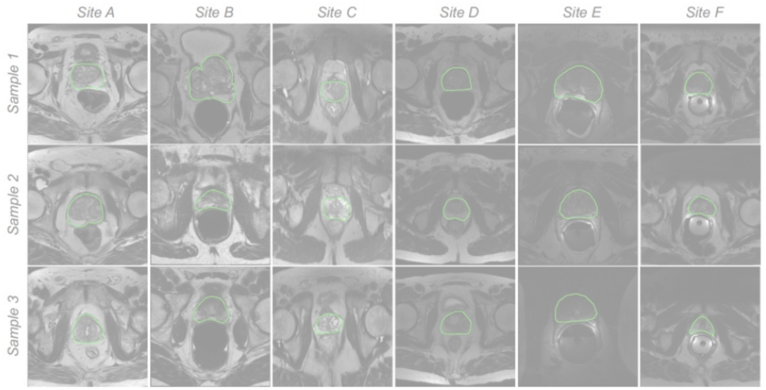


Person Re-ID (Wang *et al.* arXiv'21)

How to learn a model robust to domain shift?

Scene Segmentation (Shiau *et al.* ICIP'21)



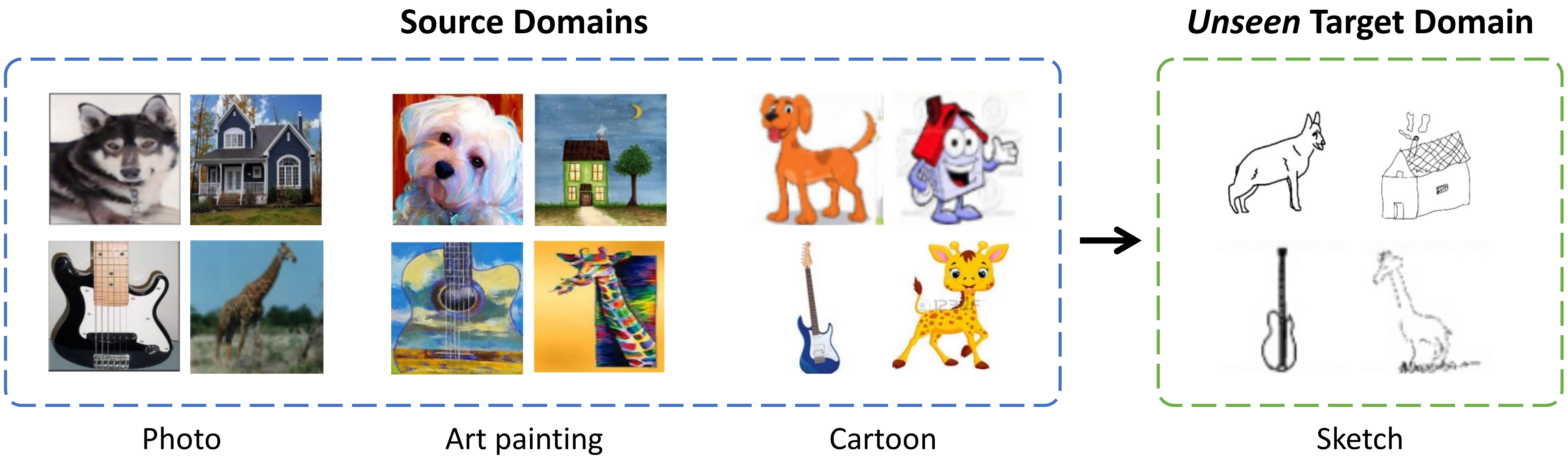


Medical Image Segmentation (Liu *et al.* MICCAI'20)

Domain Generalization (DG)

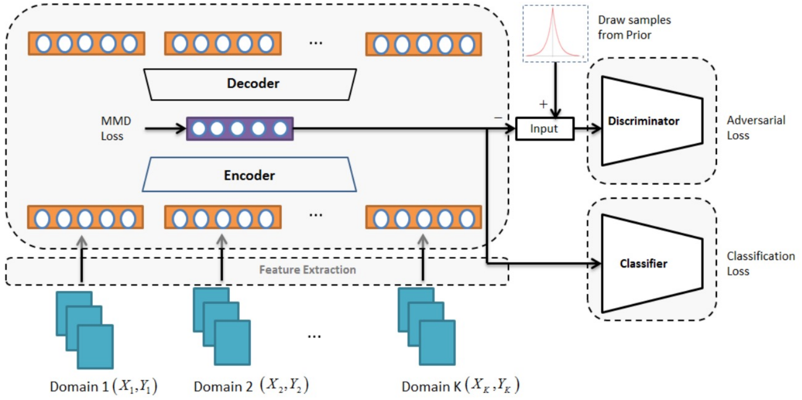
Domain Generalization (DG)

- Train a model on single/multiple source domain(s) and then directly test on *unseen* target domains
- The target data are inaccessible during model training

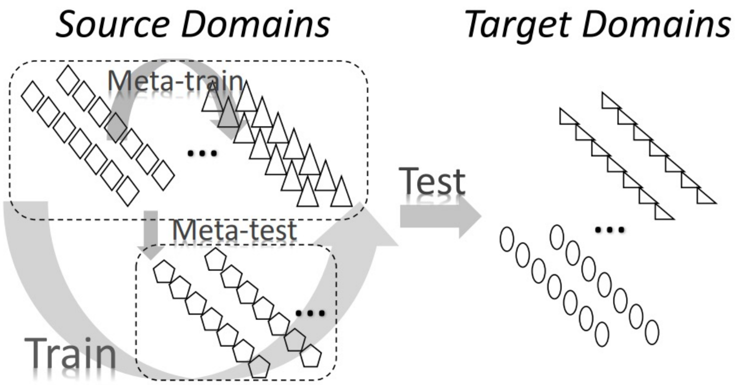


Representation-learning based DG Methods

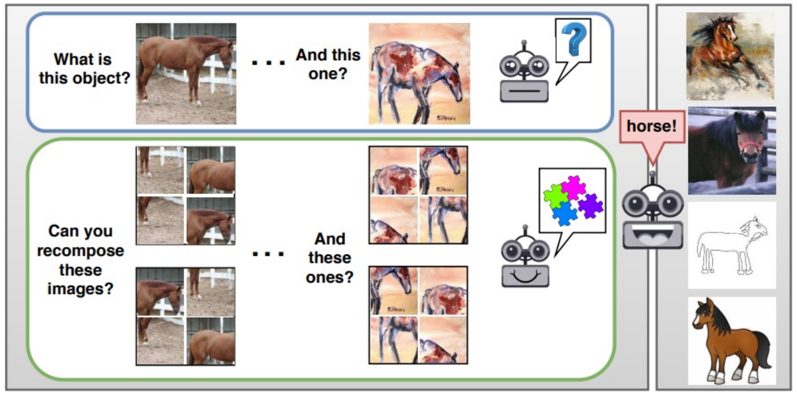
- Domain-invariant feature learning for DG
- Meta-learning for DG
- Self-supervised learning for DG



MMD-AAE
(Li *et al.* CVPR'18)



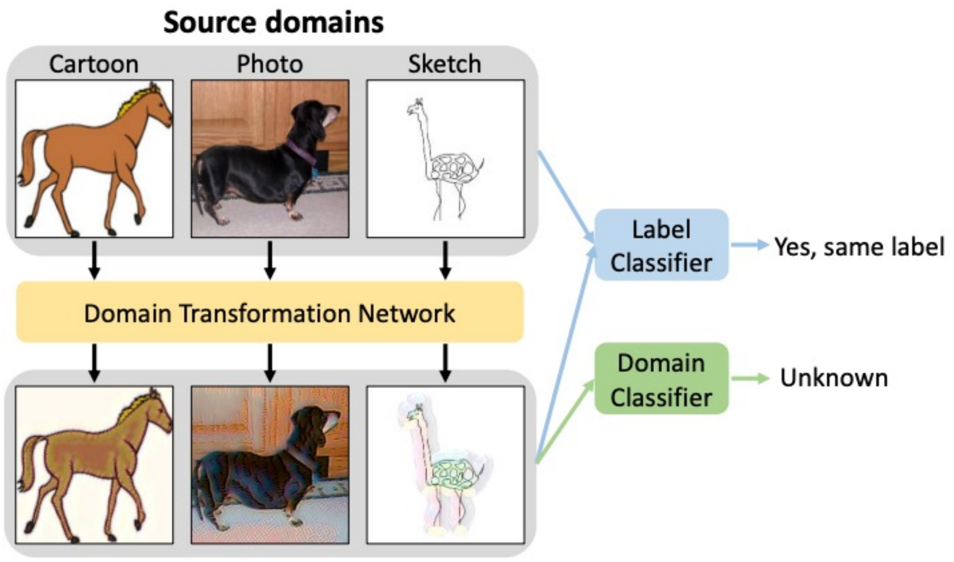
MLDG
(Li *et al.* AAAI'18)



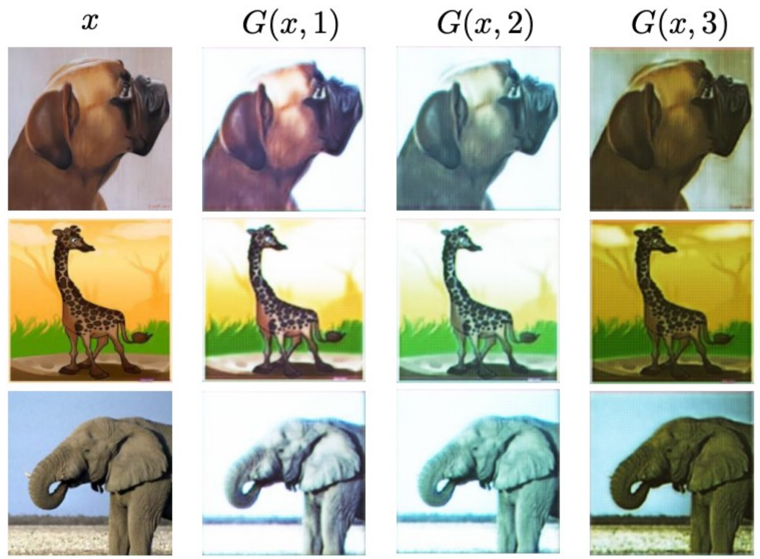
JiGen
(Carlucci *et al.* CVPR'19)

Data-generation based DG Methods

- Generate novel-domain images/features to expand the training domain and increase the diversity of training data distribution



DDAIG
(Zhou *et al.* AAAI'20)

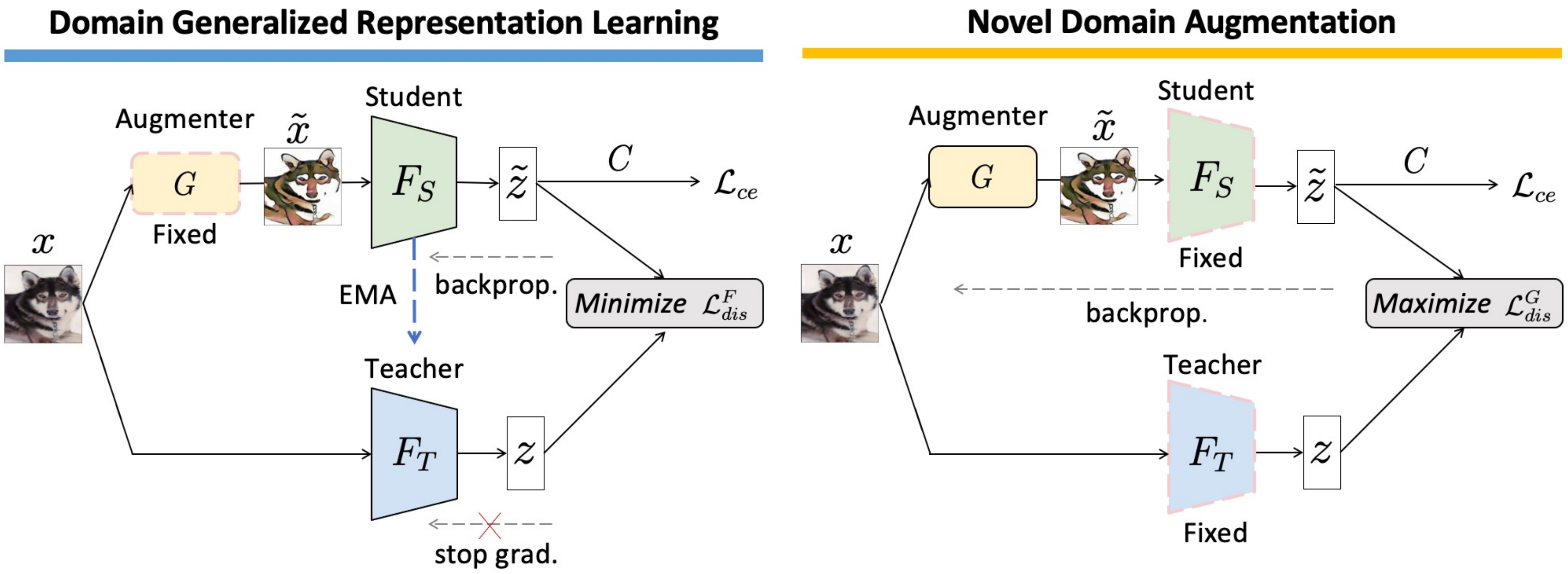


L2A-OT
(Zhou *et al.* ECCV'20)

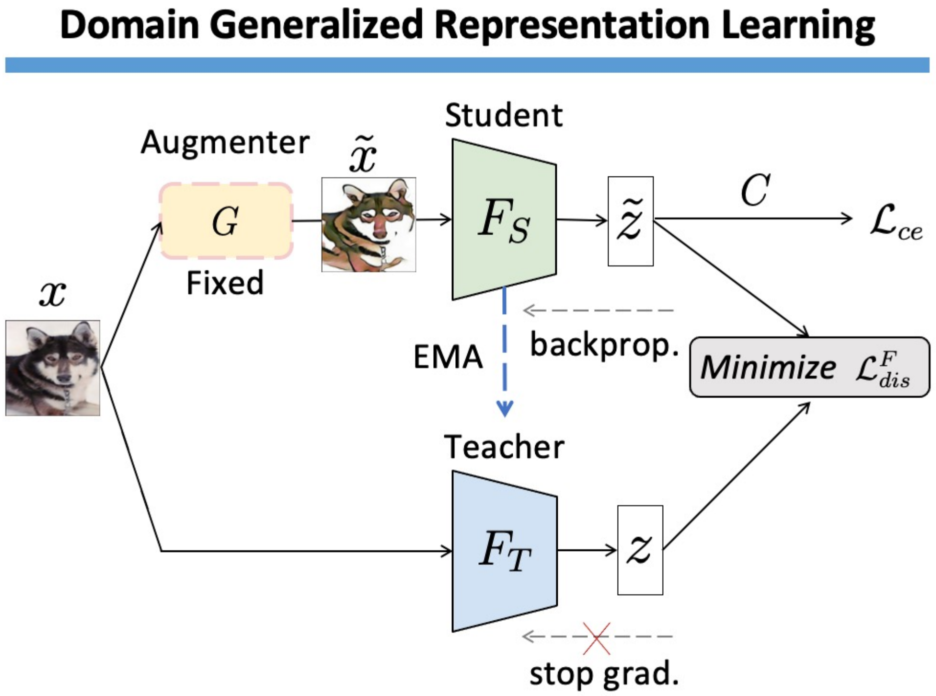
Can we perform representation learning together with novel-domain augmentation in a mutually beneficial manner?

Method – Adversarial Teacher-Student Representation Learning

- Integrate the two stages in an ***adversarial learning*** framework



Method
– Teacher-Student Domain Generalized Representation Learning



- **Minimize** the discrepancy between *Teacher* and *Student*

$$\min_{F_S} \mathcal{L}^F_{dis}(z, \tilde{z}) = \left\| \frac{z}{\|z\|_2} - \frac{\tilde{z}}{\|\tilde{z}\|_2} \right\|_2^2$$

- Distill the knowledge from *Student* to progressively update *Teacher* via exponential moving average (EMA)

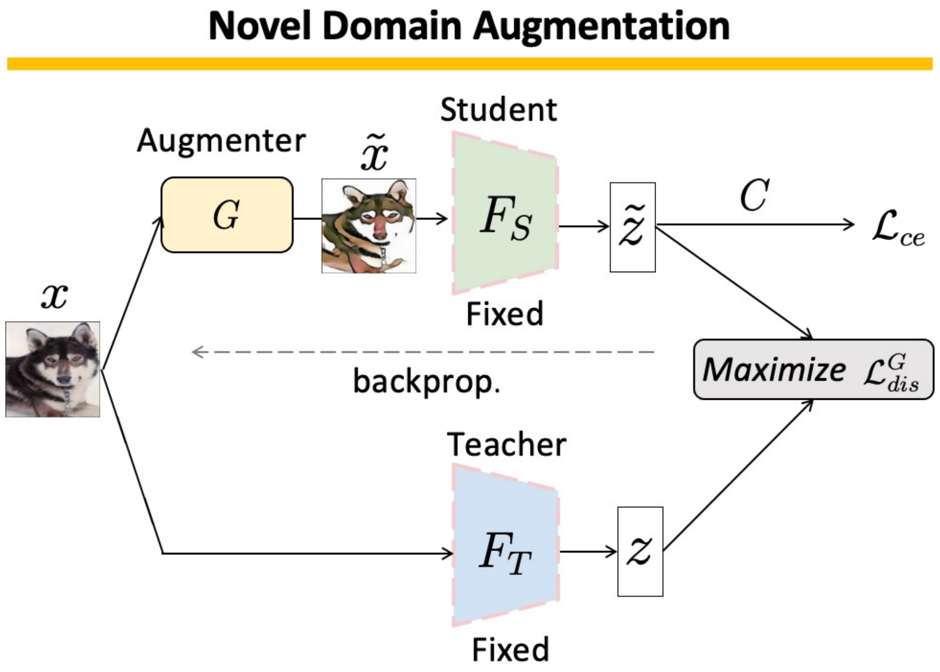
$$\theta_T \leftarrow \tau\theta_T + (1-\tau)\theta_S, \quad \text{where}\ \tau \in [0,1)$$

Method
– Novel Domain Augmentation

- **Maximize** the discrepancy between augmented and existing domains

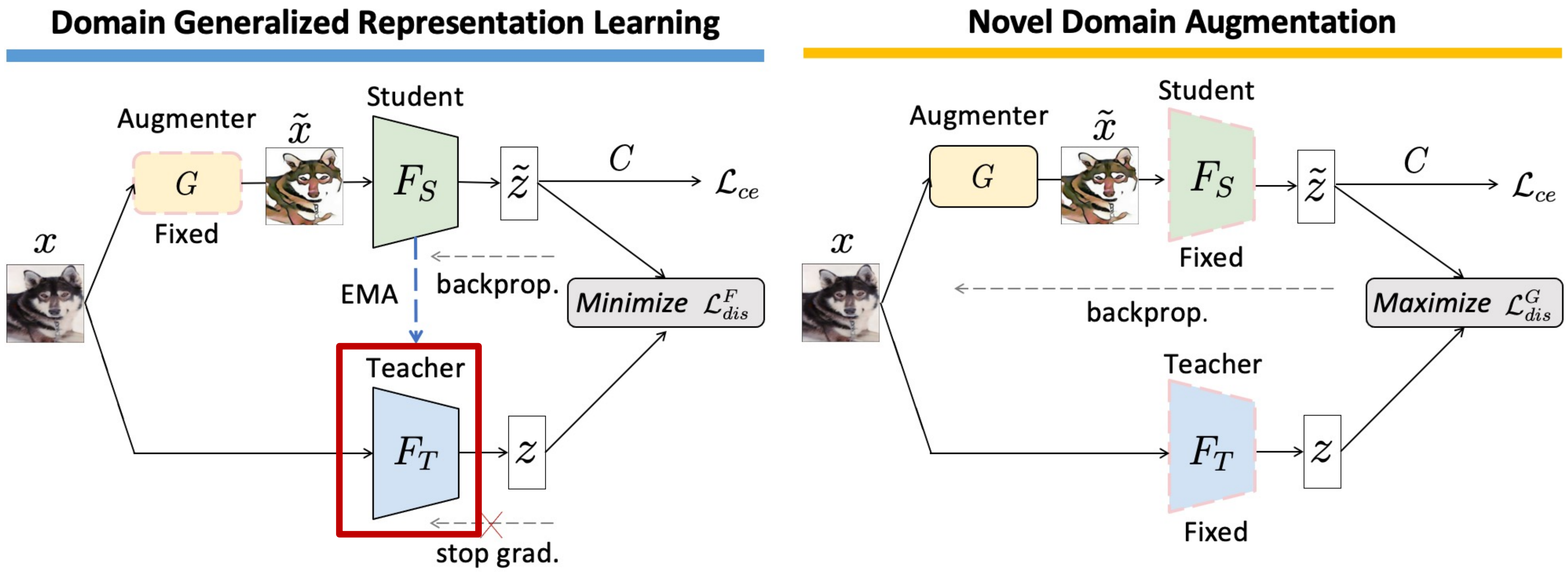
$$\max_G \mathcal{L}^G_{dis}(z, \tilde{z}) = \left[\left\|\frac{z}{\|z\|_2} - \frac{\tilde{z}}{\|\tilde{z}\|_2}\right\|_2^2 - m\right]_-$$

- The semantic information is preserved via CE loss

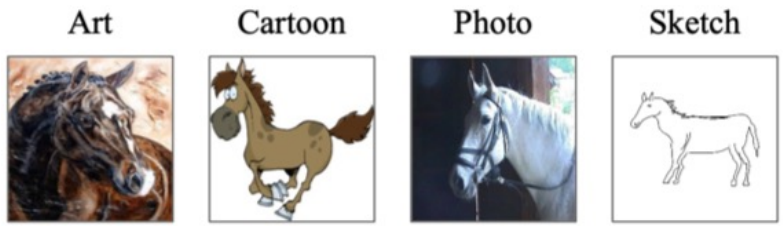


Method – Adversarial Teacher-Student Representation Learning

- During inference, we utilize the teacher network F_T to derive domain generalized representations on target domains



Result
– Quantitative Evaluation

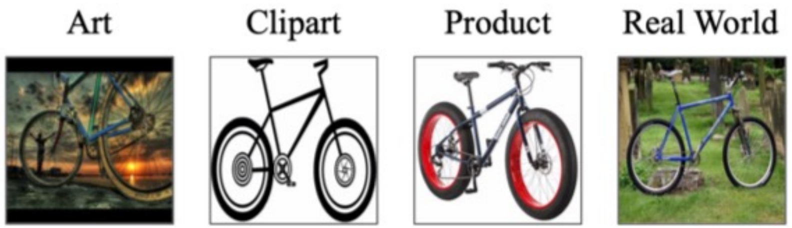


- PACS dataset (*Photo, Art painting, Cartoon, Sketch*)
- leave-one-domain-out comparisons

Target	DeepAll (baseline)	MMD-AAE [1]	MLDG [2]	JiGen [11]	MetaReg [3]	Epi-FCR [4]	MASF [5]	EISNet [12]	DMG [37]	Borlino *et al.* [43]	DSON [44]	RSC [28]	Ours
Photo	95.6	96.0	96.1	96.0	95.5	93.9	95.0	95.9	93.4	95.0	95.9	96.0	**97.3** $\pm$ 0.3
Art painting	75.1	75.2	81.3	79.4	83.7	82.1	80.3	81.9	76.9	82.7	84.7	83.4	**85.8** $\pm$ 0.6
Cartoon	74.2	72.7	77.2	75.3	77.2	77.0	77.2	76.4	80.4	78.0	77.7	80.3	**80.7** $\pm$ 0.5
Sketch	68.4	64.2	72.3	71.4	70.3	73.0	71.7	74.3	75.2	81.6	**82.2**	80.9	77.3 $\pm$ 0.5
Average	78.3	77.0	81.8	80.5	81.7	81.5	81.1	82.2	81.5	84.3	85.1	85.2	**85.3**

Target	ResNet-18						ResNet-50				
	DeepAll (baseline)	CrossGrad [6]	DDAIG [7]	L2A-OT [8]	MixStyle [9]	Ours	DeepAll (baseline)	CrossGrad [6]	DDAIG [7]	MixStyle [9]	Ours
Photo	95.6	96.0	95.3	96.2	96.1	**97.3** $\pm$ 0.3	94.8	97.8	95.7	98.0	**98.9** $\pm$ 0.3
Art painting	75.1	79.8	84.2	83.3	84.1	**85.8** $\pm$ 0.6	81.5	87.5	85.4	87.4	**90.0** $\pm$ 0.3
Cartoon	74.2	76.8	78.1	78.2	78.8	**80.7** $\pm$ 0.5	78.6	80.7	78.5	83.3	**83.5** $\pm$ 0.5
Sketch	68.4	70.2	74.7	73.6	75.9	**77.3** $\pm$ 0.5	69.7	73.9	80.0	78.5	**80.0** $\pm$ 0.6
Average	78.3	80.7	83.1	82.8	83.7	**85.3**	81.2	85.7	84.9	86.8	**88.1**

Result
– Quantitative Evaluation

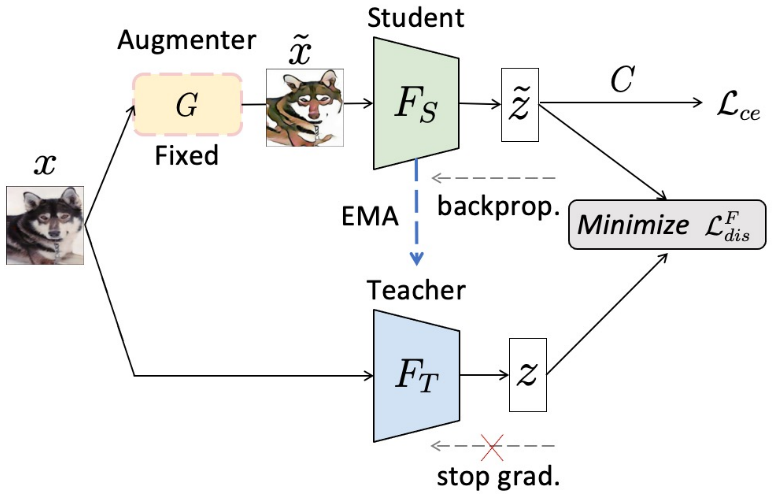


- Office-Home dataset (*Art, Clipart, Product, Real World*)
- leave-one-domain-out comparisons

Target	DeepAll (baseline)	CCSA [45]	MMD-AAE [1]	MLDG [2]	D-SAM [46]	JiGen [11]	Borlino *et al.* [43]	DSON [44]	RSC [28]	Ours
Art	59.0	59.9	56.5	58.1	58.0	53.0	58.7	59.4	58.4	**60.7** ± 0.5
Clipart	48.4	49.9	47.3	49.3	44.4	47.5	52.3	45.7	47.9	**52.9** ± 0.3
Product	72.5	74.1	72.1	72.9	69.2	71.5	73.0	71.8	71.6	**75.8** ± 0.1
Real world	75.5	75.7	74.8	74.7	71.5	72.8	75.0	74.7	74.5	**77.2** ± 0.2
Average	63.9	64.9	62.7	63.8	60.8	61.2	64.8	62.9	63.1	**66.7**

Target	ResNet-18						ResNet-50				
	DeepAll (baseline)	CrossGrad [6]	DDAIG [7]	L2A-OT [8]	MixStyle [9]	Ours	DeepAll (baseline)	CrossGrad [6]	DDAIG [7]	MixStyle [9]	Ours
Art	59.0	58.4	59.2	60.6	58.7	**60.7** ± 0.5	64.7	67.7	65.2	64.9	**69.3** ± 0.2
Clipart	48.4	49.4	52.3	50.1	**53.4**	52.9 ± 0.3	58.8	57.7	59.2	58.8	**60.1** ± 0.6
Product	72.5	73.9	74.6	74.8	74.2	**75.8** ± 0.1	77.9	79.1	77.7	78.3	**81.5** ± 0.4
Real world	75.5	75.8	76.0	73.0	75.9	**77.2** ± 0.2	79.0	80.4	76.7	78.7	**82.1** ± 0.2
Average	63.9	64.4	65.5	65.6	65.5	**66.7**	70.1	71.2	69.7	70.2	**73.3**

Result
– Ablation Study

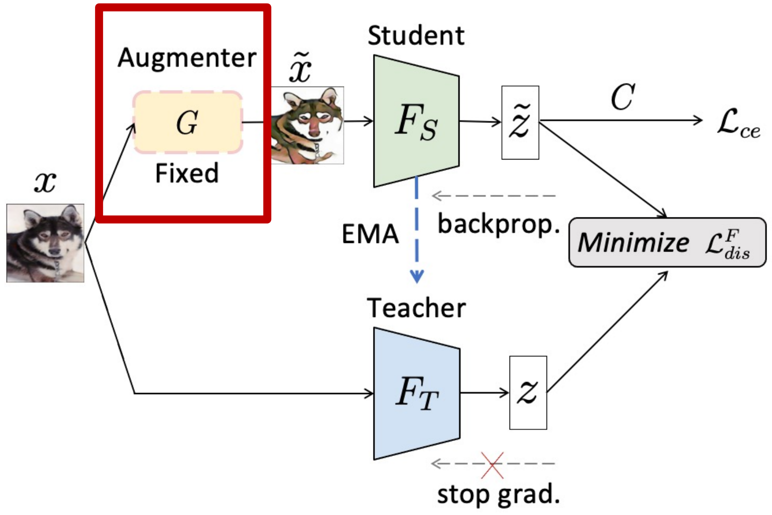


- Ablation studies on PACS using ResNet-50 as the backbone

Module	Method	Photo	Art painting	Cartoon	Sketch	Average
Augmentation	DeepAll	94.8	81.5	78.6	69.7	81.2
	Random Aug.	96.4	83.2	75.9	75.5	82.8
	Jigsaw puzzle	97.1	85.3	79.0	80.5	85.5
Representation	Siamese archi.	98.3	87.5	79.0	80.5	85.8
	F_S w/o EMA	98.2	86.4	80.1	74.7	84.9
	F_S w/ EMA	97.9	88.9	82.0	75.1	86.0
	Ours ($G + F_T$)	**98.9**	**90.0**	**83.5**	**80.0**	**88.1**

Result

– Ablation Study

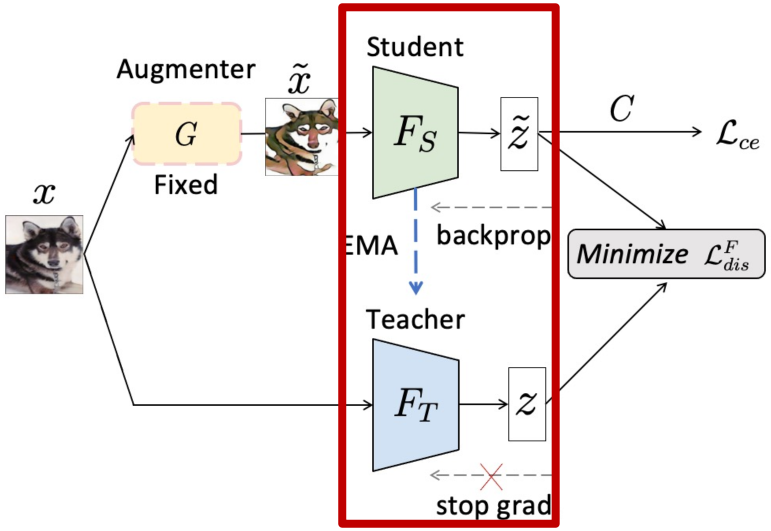


- Ablation studies on PACS using ResNet-50 as the backbone
  - Change Augmenter G to Random Aug. and Jigsaw puzzle

Module	Method	Photo	Art painting	Cartoon	Sketch	Average
	DeepAll	94.8	81.5	78.6	69.7	81.2
Augmentation	Random Aug.	96.4	83.2	75.9	75.5	82.8
	Jigsaw puzzle	97.1	85.3	79.0	80.5	85.5
Representation	Siamese archi.	98.3	87.5	79.0	80.5	85.8
	F_S w/o EMA	98.2	86.4	80.1	74.7	84.9
	F_S w/ EMA	97.9	88.9	82.0	75.1	86.0
	Ours ($G + F_T$)	**98.9**	**90.0**	**83.5**	**80.0**	**88.1**

Result

– Ablation Study



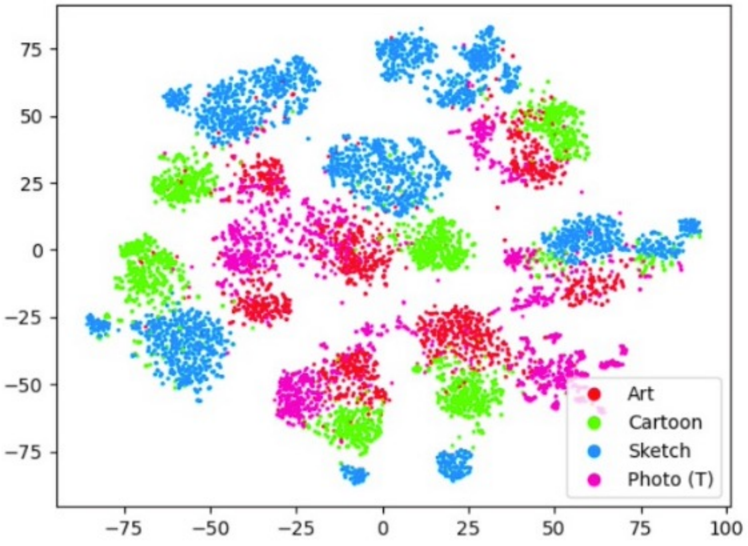
- Ablation studies on PACS using ResNet-50 as the backbone
  - Change Augmenter G to Random Aug. and Jigsaw puzzle
  - Use Siamese archi., F_S w/o EMA, and F_S w/ EMA to extract representations

Module	Method	Photo	Art painting	Cartoon	Sketch	Average
	DeepAll	94.8	81.5	78.6	69.7	81.2
Augmentation	Random Aug.	96.4	83.2	75.9	75.5	82.8
	Jigsaw puzzle	97.1	85.3	79.0	80.5	85.5
	Siamese archi.	98.3	87.5	79.0	80.5	85.8
Representation	F_S w/o EMA	98.2	86.4	80.1	74.7	84.9
	F_S w/ EMA	97.9	88.9	82.0	75.1	86.0
	Ours $(G + F_T)$	**98.9**	**90.0**	**83.5**	**80.0**	**88.1**

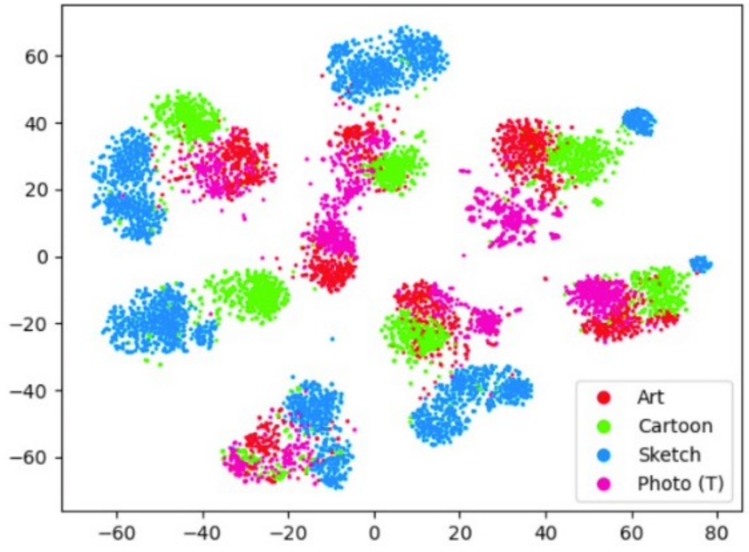
Result

– t-SNE visualization

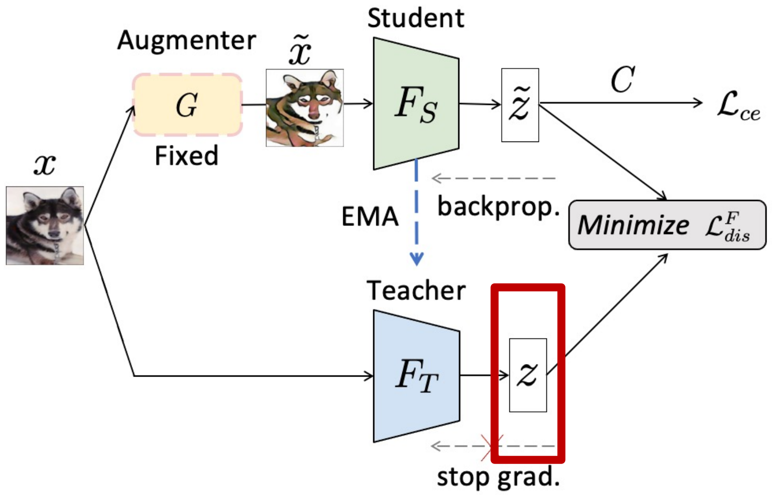
- t-SNE visualization on PACS with *Photo* as the unseen target domain
- The learned representations can be better semantically categorized by our method



(a) DeepAll

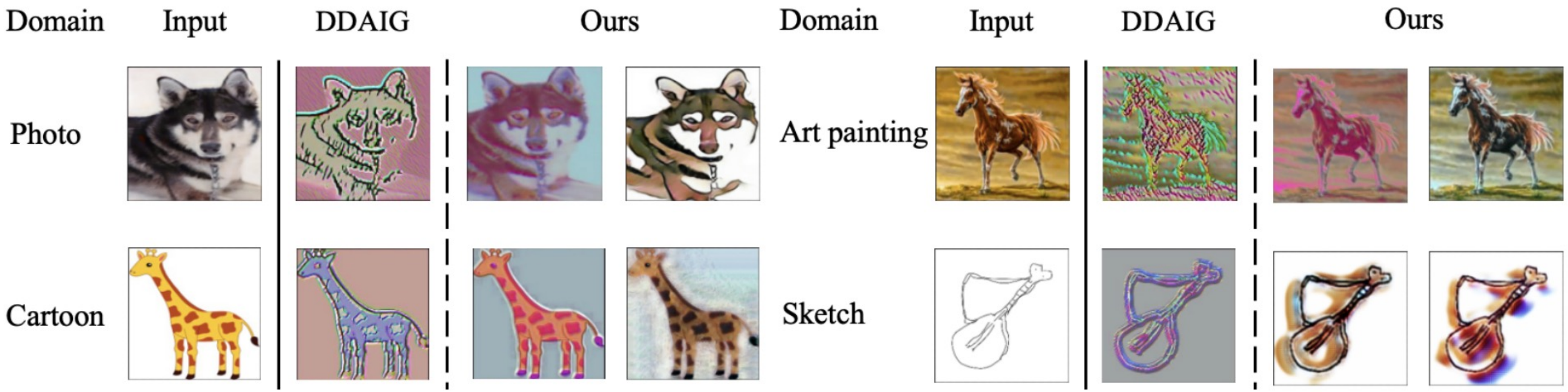


(b) Our method



Result
– Visualization

- PACS dataset
- Qualitative visualization & comparison with DDAIG (AAAI’20)



Result

– Generalization from A Single Source Domain

- Quantitative comparisons on PACS & DomainNet datasets
- PACS: *Photo* as source domain; DomainNet: *Real* as source domain

Method	PACS				DomainNet					
	Art painting	Cartoon	Sketch	Average	Clipart	Infograph	Painting	Quickdraw	Sketch	Average
DeepAll	60.7	23.5	29.0	37.7	34.5	15.7	40.7	3.6	25.9	24.1
JiGen [11]	63.6	28.5	30.2	40.8	50.0	19.0	46.3	7.2	35.5	31.6
CrossGrad [6]	64.2	29.4	32.1	41.9	49.4	19.3	47.3	5.8	35.6	31.5
DDAIG [7]	64.1	32.5	29.6	42.1	41.4	16.5	40.9	3.2	26.7	25.7
M-ADA [41]	64.6	34.6	26.6	41.9	50.3	19.5	48.1	7.1	36.0	32.2
Ours	**68.2** $\pm$ 0.9	**36.3** $\pm$ 0.9	**33.5** $\pm$ 0.3	**46.0**	**52.2** $\pm$ 0.3	**21.6** $\pm$ 0.2	**50.1** $\pm$ 0.2	**8.1** $\pm$ 0.3	**38.3** $\pm$ 0.4	**34.1**

Conclusion

- To directly deploy on target domains *without* the need of target data
- Not only derive the **domain-invariant features** across multiple source domains
- The **novel-domain augmentation** is designed to expand the training domain and diversify the training data distribution
- Our proposed approach *does not* require domain labels, thus can be applied on both multi-source and single-source DG settings

Adversarial Teacher-Student Representation Learning for Domain Generalization

Spotlight Presentation, NeurIPS 2021

Paper

Fu-En Yang
f07942077@ntu.edu.tw

Yuan-Chia Cheng
r08942154@ntu.edu.tw

Zu-Yun Shiau
r09942069@ntu.edu.tw

Yu-Chiang Frank Wang
ycwang@ntu.edu.tw

國立臺灣大學
National Taiwan University

AICS

Thanks for listening!