ETH zürich University of Cyprus V4RL VISION FOR ROBOTICS LAB

BMVC 2023 The 34th British Machine Vision Conference
20th - 24th November 2023, Aberdeen, UK

Domain-Adaptive Semantic Segmentation with Memory-Efficient Cross-Domain Transformers

Ruben Mascaro
Lucas Teixeira
Margarita Chli

Vision for Robotics Lab
ETH Zurich & University of Cyprus

1. Motivation

Transformer-based architectures have demonstrated to greatly outperform CNNs when applied to UDA tasks.

In semantic segmentation, current approaches still struggle to effectively learn context dependencies in the target domain.

This typically leads to the confusion of classes that have similar appearance, such as *road* and *sidewalk* in these examples from DAFormer [1].

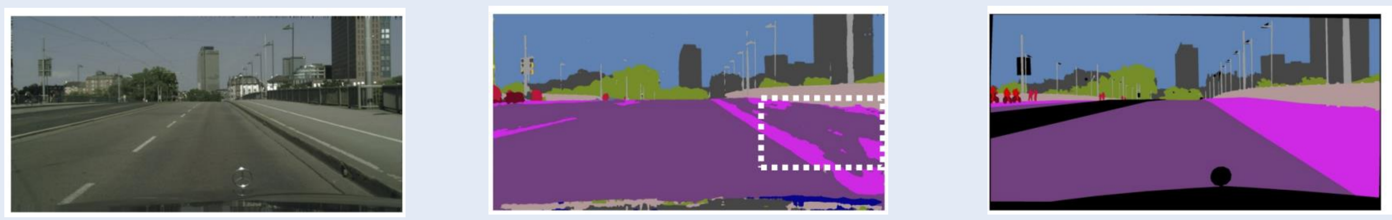
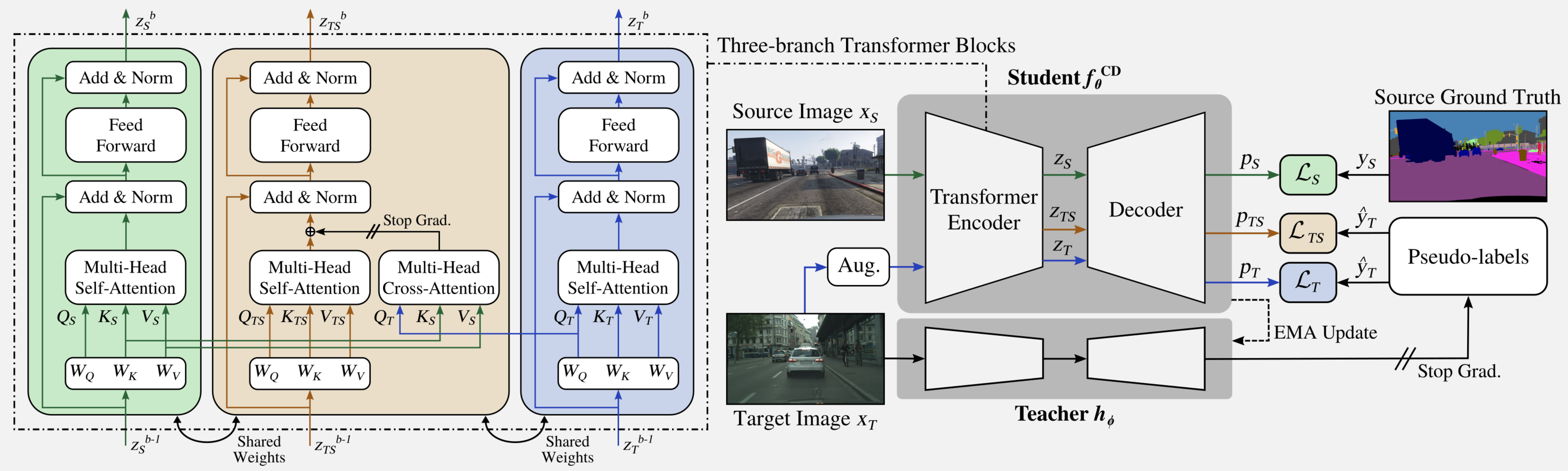


Image DAFormer Ground Truth

2. UDA Self-Training with Memory-Efficient Cross-Domain Transformers

We present a new Transformer block combining intra- and cross-domain attention for better source-target feature alignment.

It can be easily incorporated into state-of-the-art self-training UDA frameworks to enhance knowledge transfer.



Training loss:

$$\mathcal{L} = \mathcal{L}_S + \mathcal{L}_T + \mathcal{L}_{TS}$$

$$\mathcal{L}_S = -\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{C} y_S^{(h,w,c)} \log p_S^{(h,w,c)}$$

$$\mathcal{L}_T = -\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{C} q_T \hat{y}_T^{(h,w,c)} \log p_T^{(h,w,c)}$$

$$\mathcal{L}_{TS} = -\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{C} q_T \hat{y}_T^{(h,w,c)} \log p_{TS}^{(h,w,c)}$$

3. Comparison with the State of the Art

We evaluate our approach on synthetic-to-real and clear-to-adverse-weather UDA tasks using benchmarking datasets.

A comparison against SOTA UDA approaches that leverage Transformer architectures is provided.

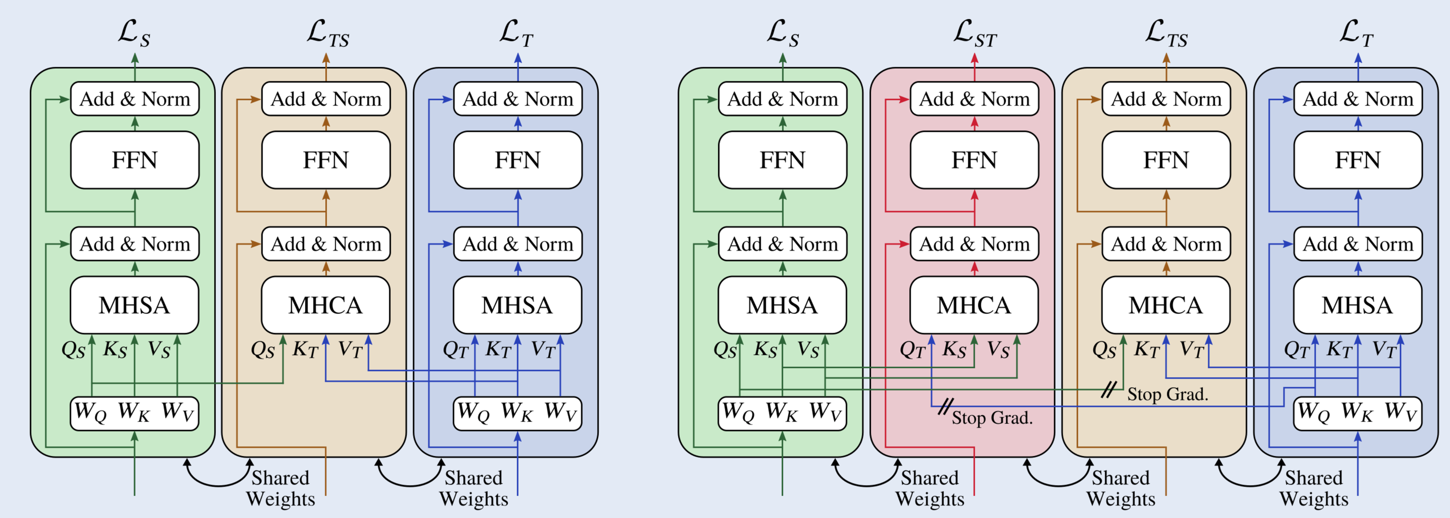
Method	road	sidew.	build.	wall	fence	pole	t.light	t.sign	veget.	terrain	sky	person	rider	car	truck	bus	train	m.bike	bike	mIoU
Synthetic-to-Real: GTA → Cityscapes (Val.)																				
DAFormer	95.7	70.2	89.4	53.5	48.1	49.6	55.8	59.4	89.9	47.9	**92.5**	**72.2**	44.7	92.3	74.5	78.2	65.1	55.9	61.8	68.3
CDTDA	**96.5**	**73.9**	89.5	**56.8**	48.9	50.7	55.8	**63.3**	89.9	**49.1**	91.2	**72.2**	**45.4**	**92.7**	78.3	**82.9**	67.5	55.2	**63.4**	69.6
Ours	96.3	73.7	**89.9**	56.2	**49.7**	**52.0**	**56.8**	62.7	**90.0**	**49.1**	91.5	71.5	44.6	92.5	**79.4**	77.8	**71.6**	**56.8**	63.2	**69.7**
Synthetic-to-Real: SYNTHIA → Cityscapes (Val.)																				
DAFormer	84.5	40.7	88.4	41.5	6.5	50.0	55.0	**54.6**	**86.0**	-	89.8	73.2	**48.2**	87.2	-	53.2	-	53.9	**61.7**	60.9
CDTDA	83.7	42.9	87.4	39.8	7.5	**50.7**	55.7	53.5	85.9	-	**90.9**	**74.5**	47.2	86.0	-	60.2	-	**57.8**	60.8	61.5
Ours	**86.0**	**44.9**	**88.7**	**44.0**	**7.9**	50.3	**56.0**	54.0	85.6	-	88.4	73.8	46.2	**87.7**	-	**61.5**	-	55.8	60.3	**62.0**
Clear-to-Adverse Weather: Cityscapes → ACDC (Test)																				
DAFormer	58.4	51.3	84.0	42.7	35.1	**50.7**	30.0	**57.0**	**74.8**	52.8	51.3	58.3	32.6	82.7	58.3	54.9	82.4	**44.1**	50.7	55.4
CDTDA	57.6	43.7	**85.1**	43.5	33.9	50.1	42.9	53.9	72.8	52.9	52.2	59.4	34.7	**83.6**	60.4	**68.7**	**84.3**	41.4	**53.0**	56.5
Ours	**69.0**	**53.1**	84.7	**45.8**	**36.0**	50.1	**43.2**	**57.0**	73.4	**54.2**	**65.9**	**59.9**	**37.0**	83.0	**65.8**	62.3	83.9	42.3	51.5	**58.8**

➔ Our method leads to **more effective learning of context relationships** in the target domain, resulting in better distinction of visually similar classes (road/sidewalk, road/sky, wall/fence/building, etc.).

Image DAFormer Ours Ground Truth

4. Architecture Evaluation

We compare our design with other cross-domain Transformers.



CDTrans [2] CDTDA [3]

Architecture	mIoU	Throughput	GPU Memory
DAFormer	68.1 ± 0.7	**0.70 it/s**	**9.81 GB**
CDTrans	68.8 ± 0.4	0.44 it/s	17.51 GB
CDTDA	68.9 ± 0.6	0.37 it/s	13.35 GB
Ours	**69.7 ± 0.4**	0.52 it/s	**9.81 GB**

➔ Our method requires **less GPU memory** for training while offering **better adaptation** capabilities.

References

[1] L. Hoyer et al., "DAFormer: Improving Network Architectures and Training Strategies for Domain-Adaptive Semantic Segmentation," CVPR 2022.

[2] T. Xu et al., "CDTrans: Cross-Domain Transformer for Unsupervised Domain Adaptation," ICLR 2022.

[3] K. Wang et al., "Exploring Consistency in Cross-Domain Transformer for Domain Adaptive Semantic Segmentation," ICCV 2023.

Code