MIC: Masked Image Consistency for Context-Enhanced Domain Adaptation

Lukas Hoyer, Dengxin Dai, Haoran Wang, Luc Van Gool • github.com/lhoyer/MIC

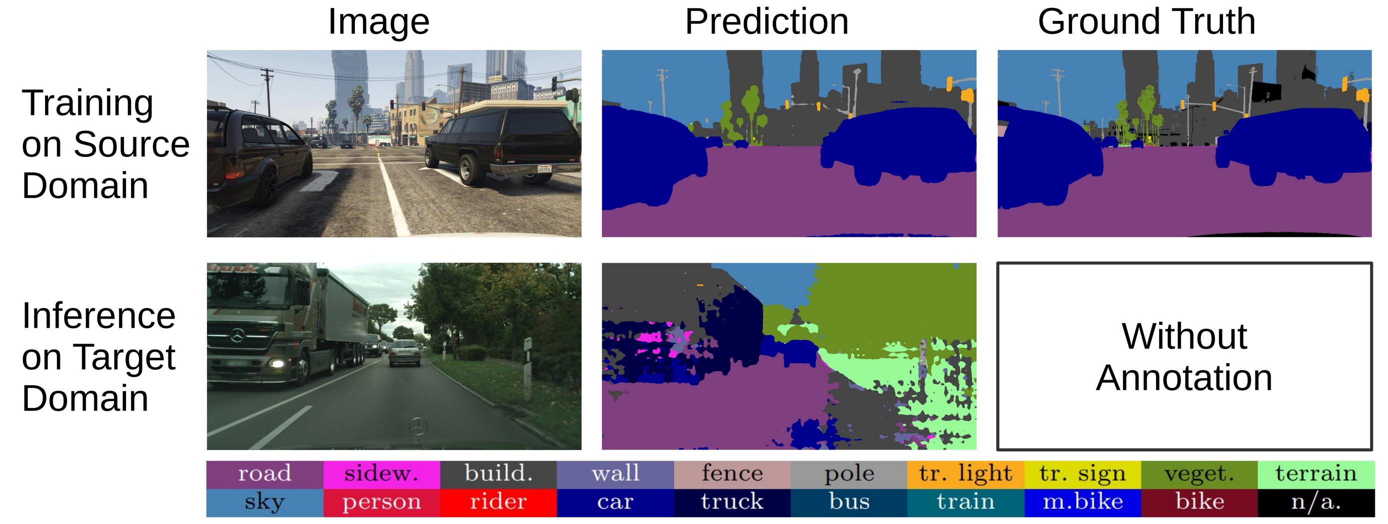
max planck institut informatik

ETH zürich

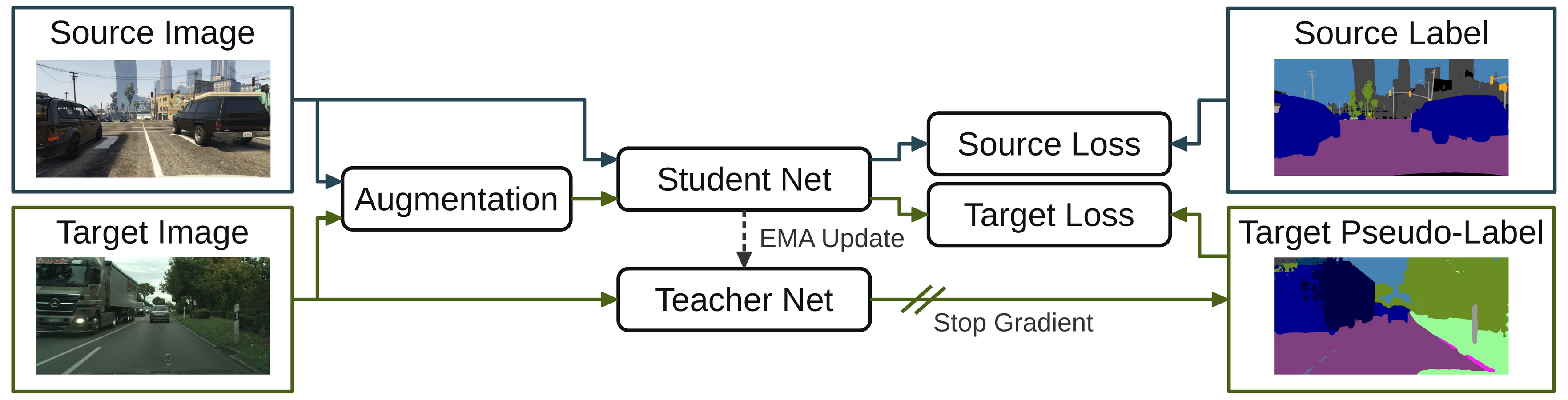
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

1 Unsupervised Domain Adaptation (UDA)

Problem: Networks trained on one domain often experience a performance drop on another domain

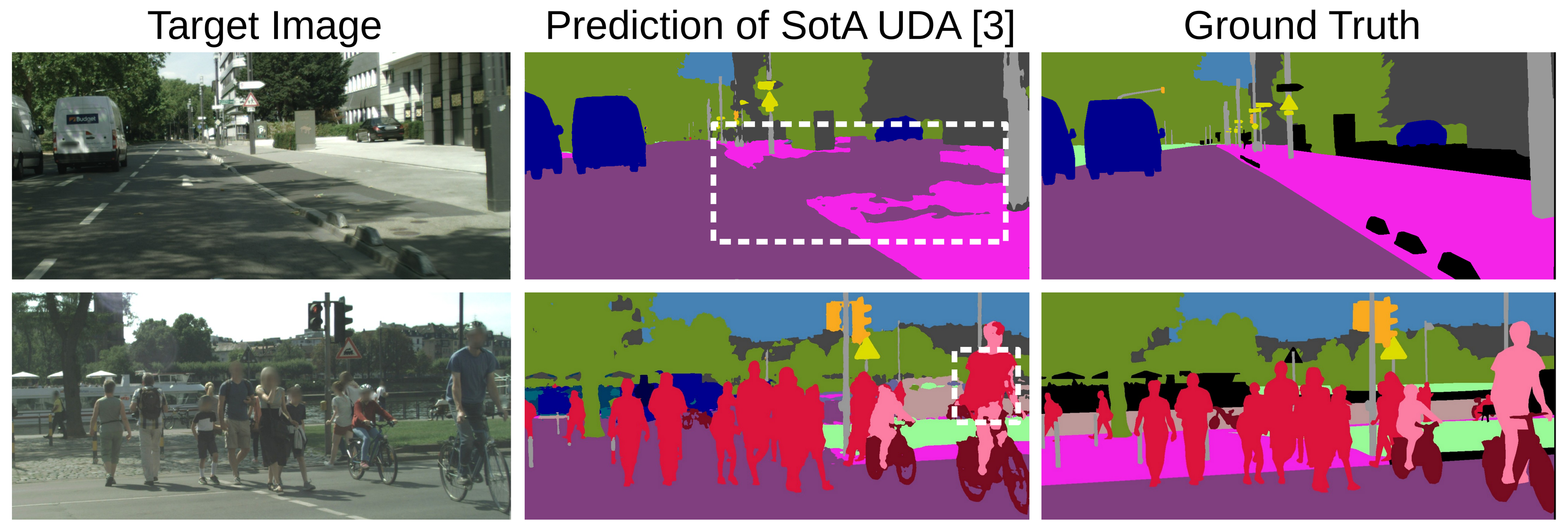


Solution: Adapt network to unlabeled target images, e.g. using self-training



2 MIC Motivation

Problem: Previous UDA methods confuse classes with similar local appearance on the unlabeled target domain



- ➔ Road/sidewalk and pedestrian/rider are confused
- ➔ Parts are correctly recognized (e.g. curb in foreground and legs of rider)
- ➔ Previous UDA methods insufficiently exploit context clues

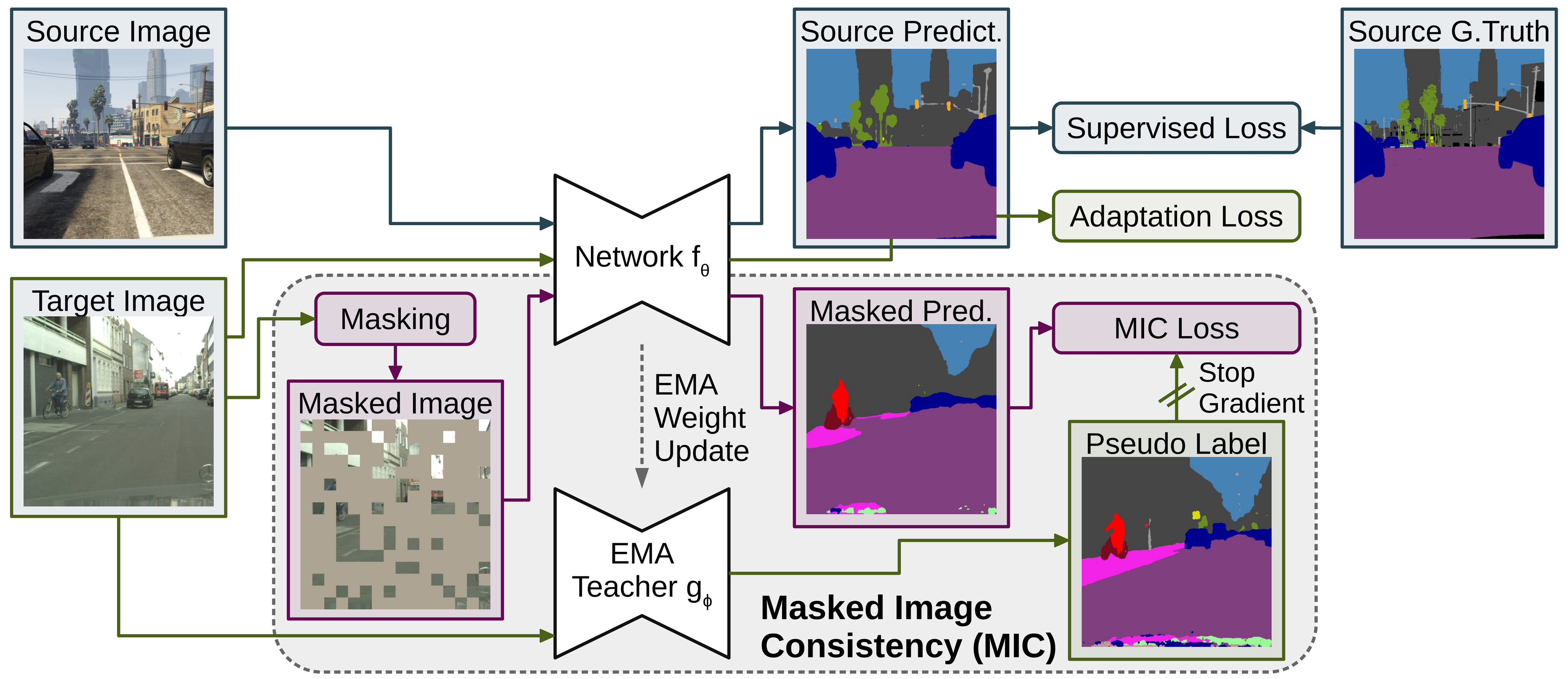
Goal: Enhance the learning of context relations on target domain

3 MIC Method

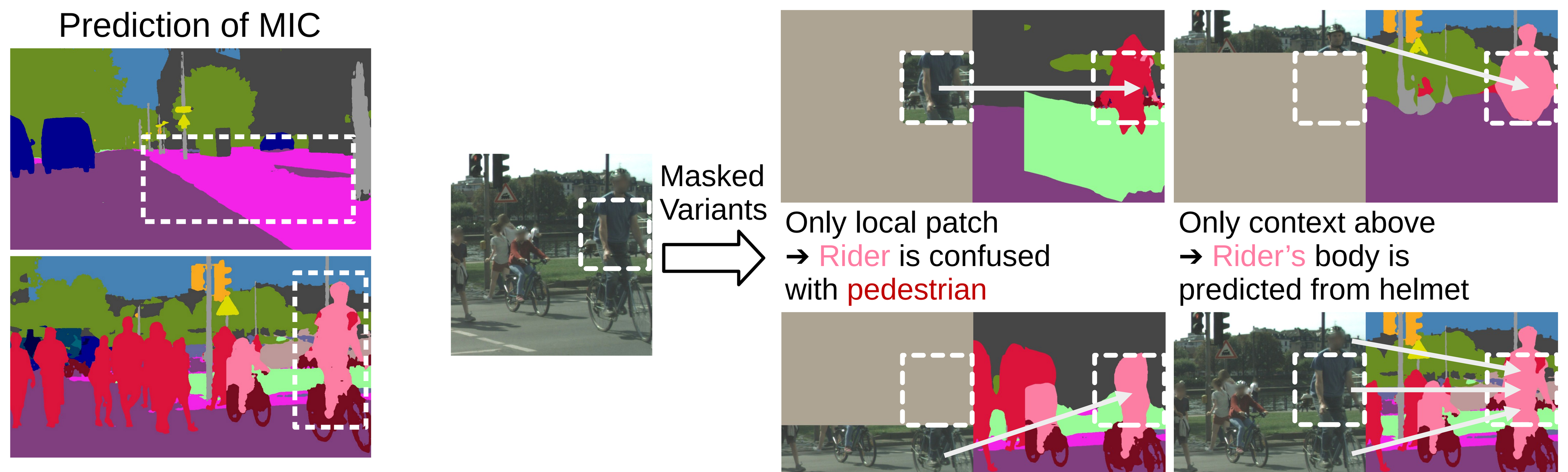
Masked Image Consistency (MIC) Plug-In for UDA

- Randomly mask out target image patches
- Predict semantics for entire image (incl. masked patches)

➔ Network learns to utilize context



Context Utilization of MIC



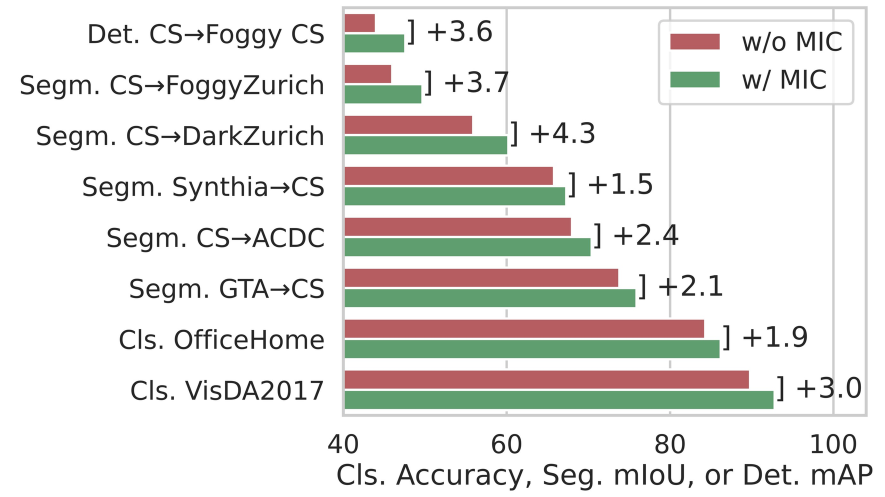
➔ MIC better distinguishes visually similar classes such as road/sidewalk and pedestrian/rider

References

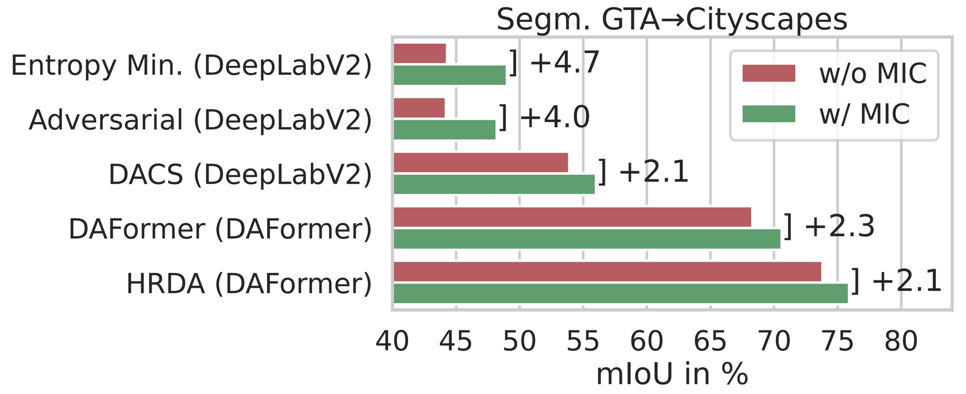
[1] Chen et al. "Scale-aware domain adaptive faster r-cnn", IJCV 2021.
[2] Hoyer et al. "DAFormer: Improving network architectures and training strategies for domain-adaptive semantic segmentation", CVPR 2022.
[3] Hoyer et al. "HRDA: Context-aware high-resolution domain-adaptive semantic segmentation", ECCV 2022.
[4] Li et al. "Sigma: Semantic-complete graph matching for domain adaptive object detection", CVPR 2022.
[5] Tranheden et al. "DACS: Domain Adaptation via Cross-domain Mixed Sampling", WACV 2021.

4 MIC Evaluation

Comparison with the State of the Art



- ➔ MIC improves the State of the Art across
  - Vision tasks: classification, segmentation, detection
  - Domain gaps: synthetic/real, day/night, clear/adverse weather



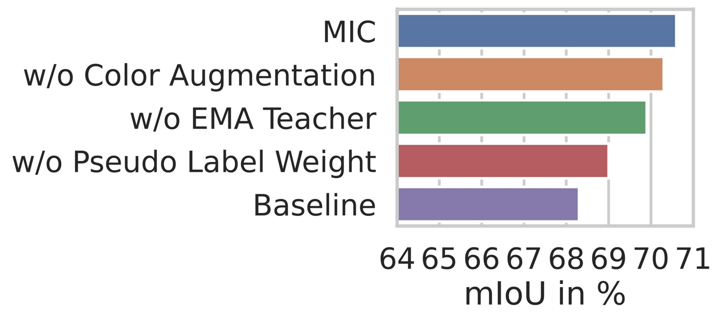
➔ MIC can be combined with various UDA methods and yields significant gains

Segm. GTA→Cityscapes

	Road	S.walk	Build.	Wall	Fence	Pole	T.Light	T.Sign	Veget.	Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	M.bike	Bike
w/o MIC	96	74	91	62	51	57	64	69	91	48	94	79	53	94	84	86	76	64	68
w/ MIC	97	80	92	61	57	60	66	71	92	51	94	80	56	95	85	90	80	65	68
Context:		Curb			Post								Bicycle				Rails		

➔ MIC improves classes with important context

Ablation Study



➔ EMA teacher and pseudo label confidence weighting enhance MIC

Patch Size b \ Mask Ratio r	0.3	0.5	0.7	0.9
32	69.3	69.9	69.7	69.3
64	69.2	70.3	70.6	69.7
128	68.7	70.5	70.4	68.2
256	66.2	69.5	69.8	68.0

➔ High mask ratio with medium-sized patches works best

MIC Domain	GTA→CS	CS→ACDC
---	73.8	65.3
Source	71.1	66.5
Target	75.9	66.9
Src+Trg	74.5	68.0

➔ Domain context gap can require learning target-specific relations

Example Predictions