MIC: Masked Image Consistency for Context-Enhanced Domain Adaptation

Lukas Hoyer[1], Dengxin Dai[2], Haoran Wang[2], Luc Van Gool[1,3]

[1] ETH Zürich, Switzerland
[2] Max Planck Institute for Informatics, Germany
[3] KU Leuven, Belgium

github.com/lhoyer/MIC

CVPR’23 Poster WED-AM-333

ETH zürich CVL Computer Vision Lab

max planck institut informatik

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

MIC: Overview

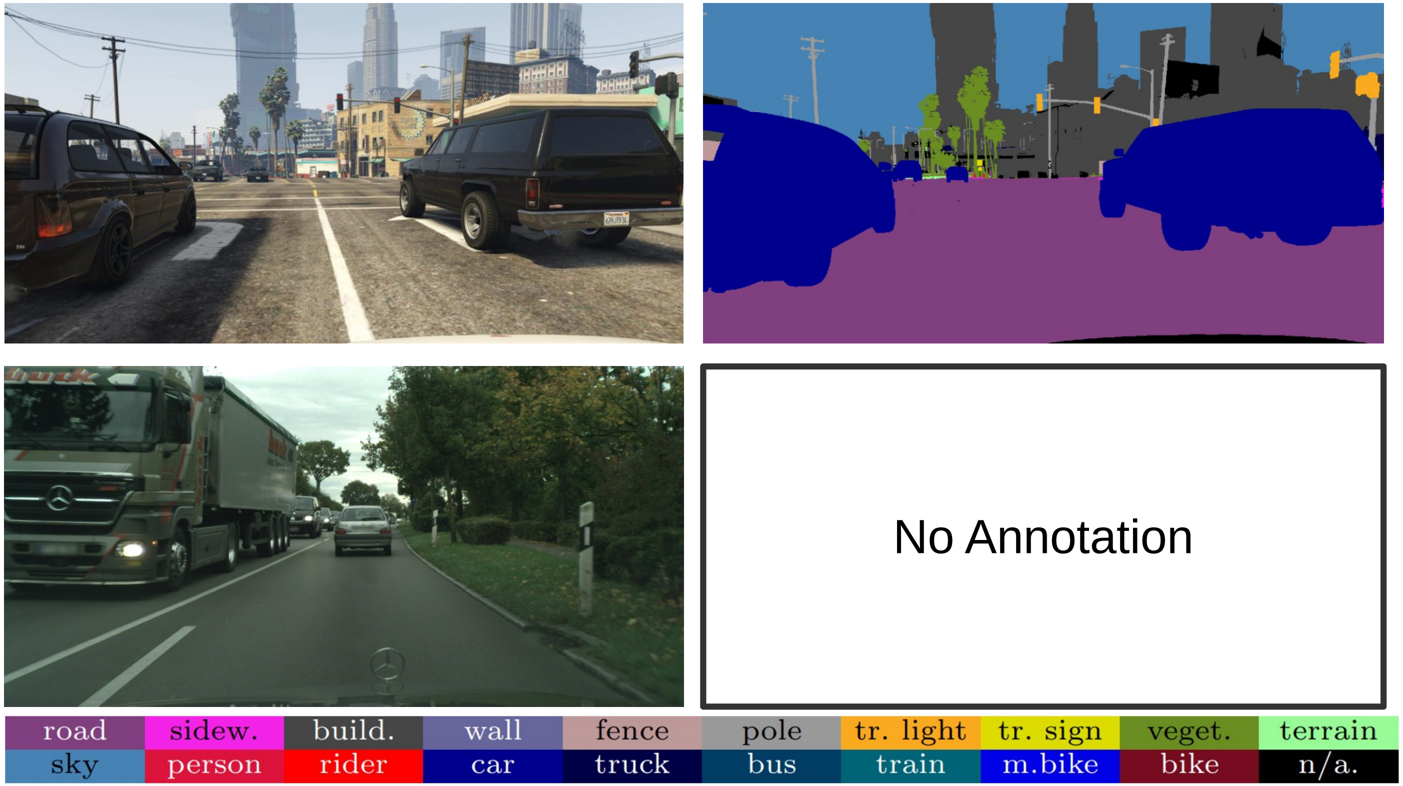
Unsupervised Domain Adaptation (UDA)

Image

Ground Truth

Source Domain

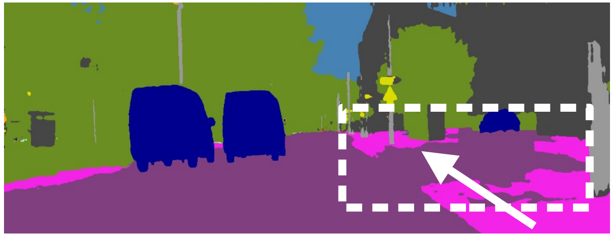
Target Domain


No Annotation

MIC: Overview

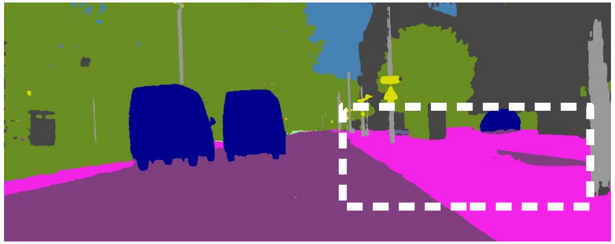
Target Domain Image

SotA UDA Prediction

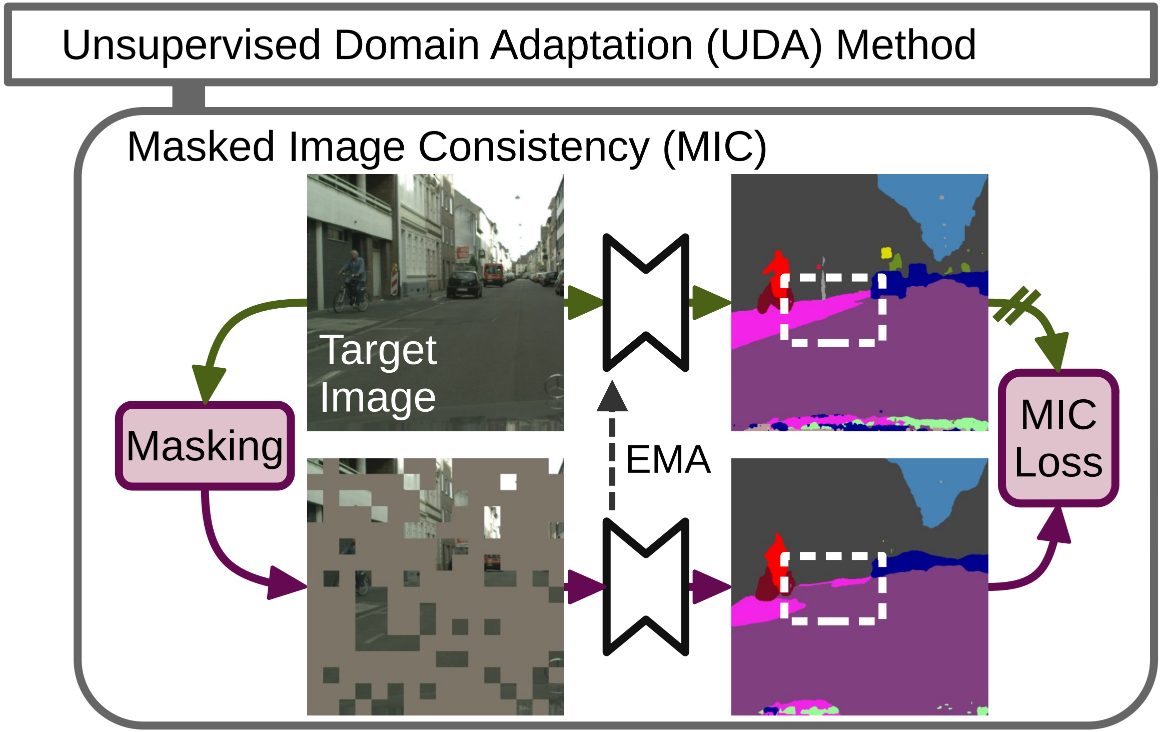


➔ Classes with a similar local appearance are confused

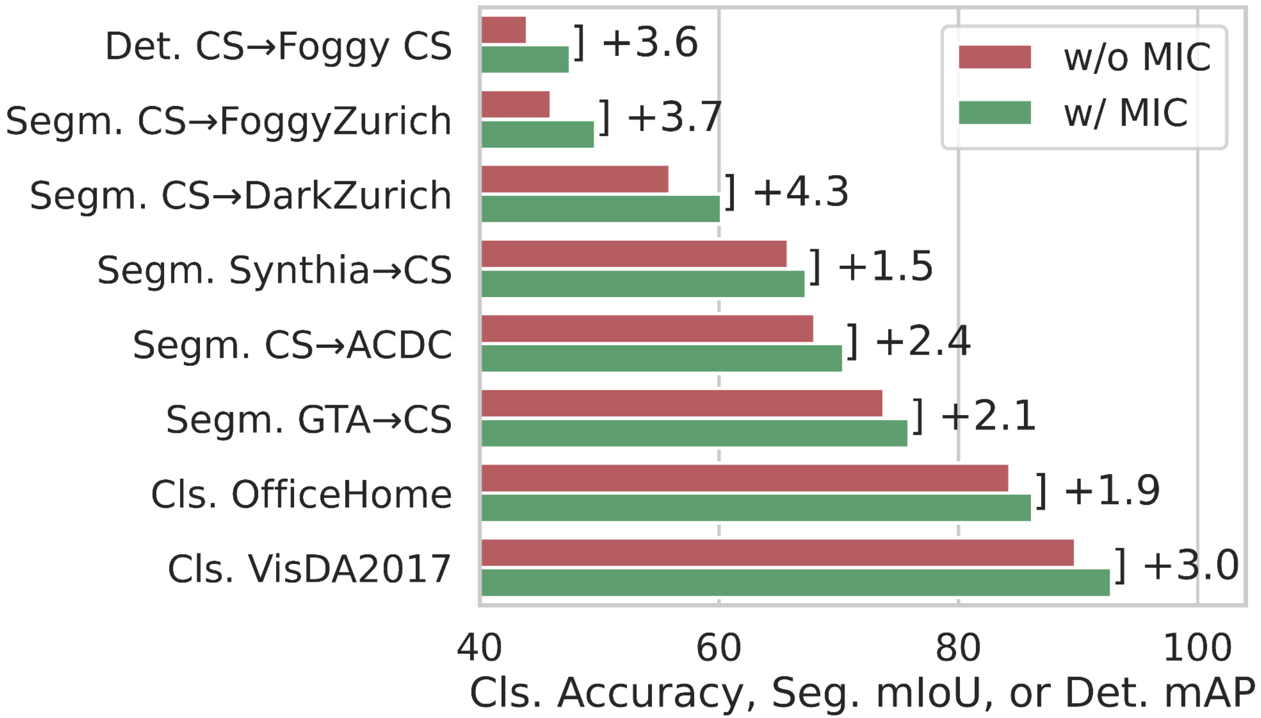
MIC Prediction



Idea: Enhance learned context relations on target domain


Unsupervised Domain Adaptation (UDA) Method
Masked Image Consistency (MIC)
Target Image
Masking
EMA
MIC Loss

MIC: Overview



➔ MIC improves the State of the Art on various UDA benchmarks

Unsupervised Domain Adaptation (UDA)

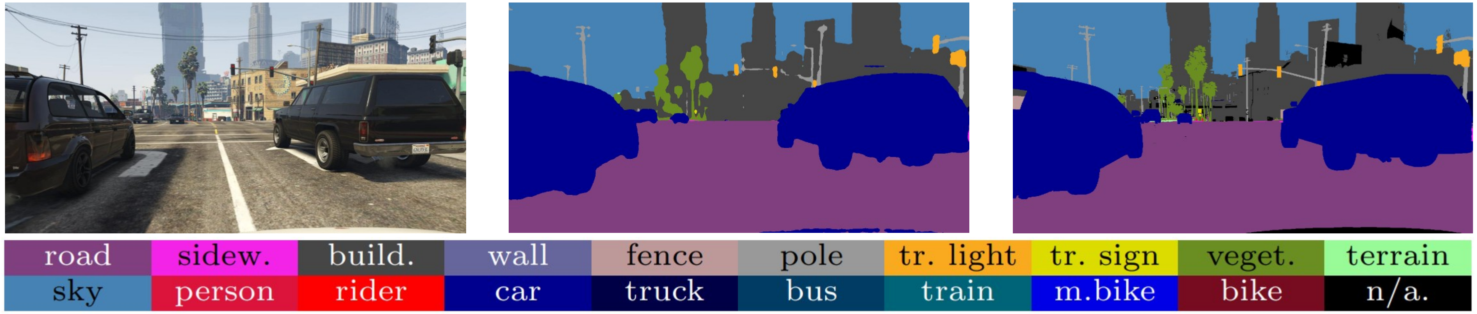
Motivation: Reduce annotation effort with synthetic data

Image

Prediction

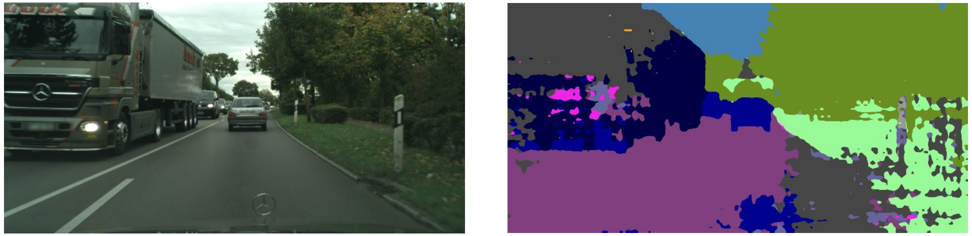
Ground Truth

Training on
Source Domain
(Synthetic)



Problem: Performance drop on target domain

Inference on
Target Domain
(Real)

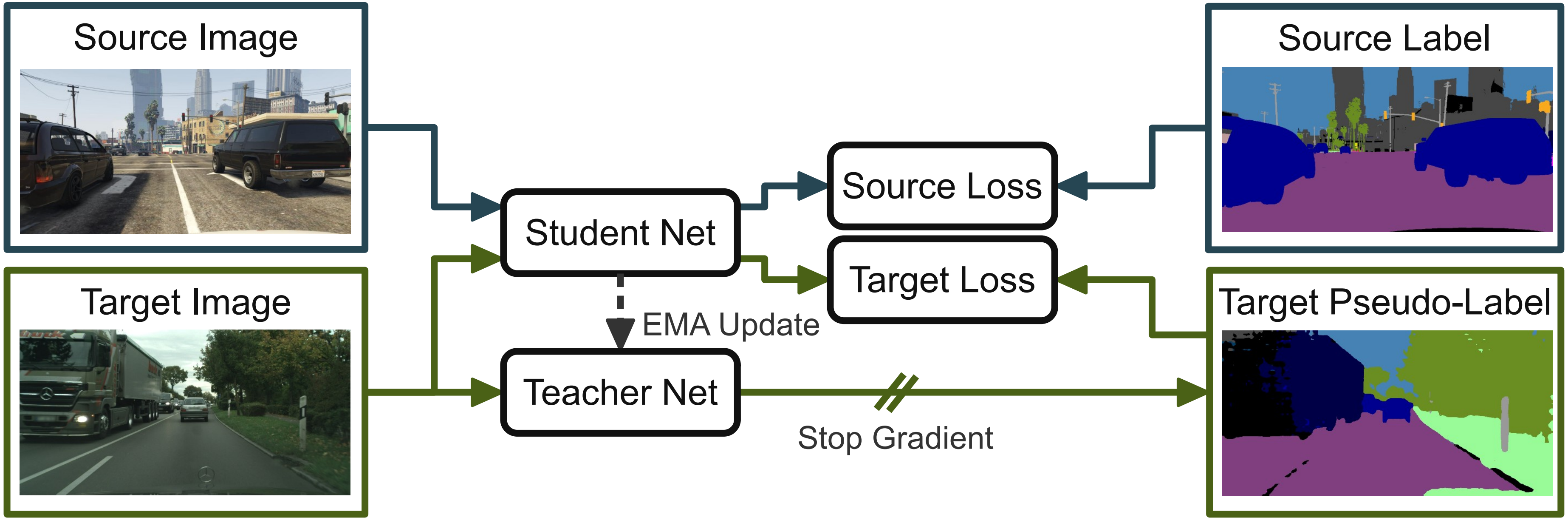


Without Annotation

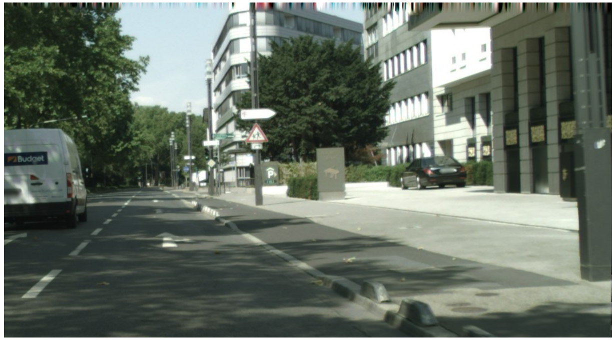
➔ Adapt network to unlabeled target images (UDA)

Preliminary: Self-Training for UDA

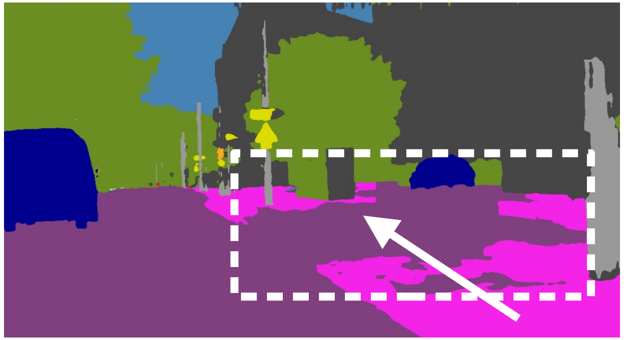
Idea: Use confident target predictions as pseudo-labels



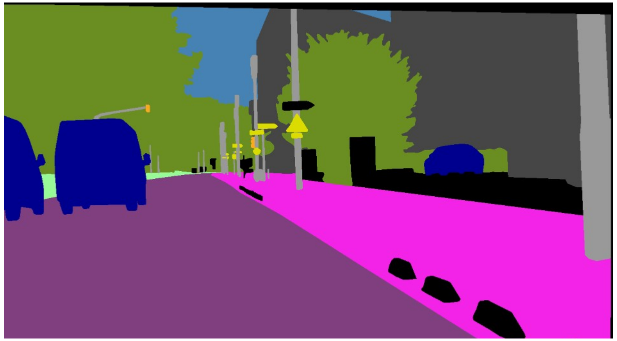
MIC: Motivation



Target Image



SotA UDA Prediction [1]



Ground Truth

Problem: Classes with similar local appearance are confused such as road/sidewalk

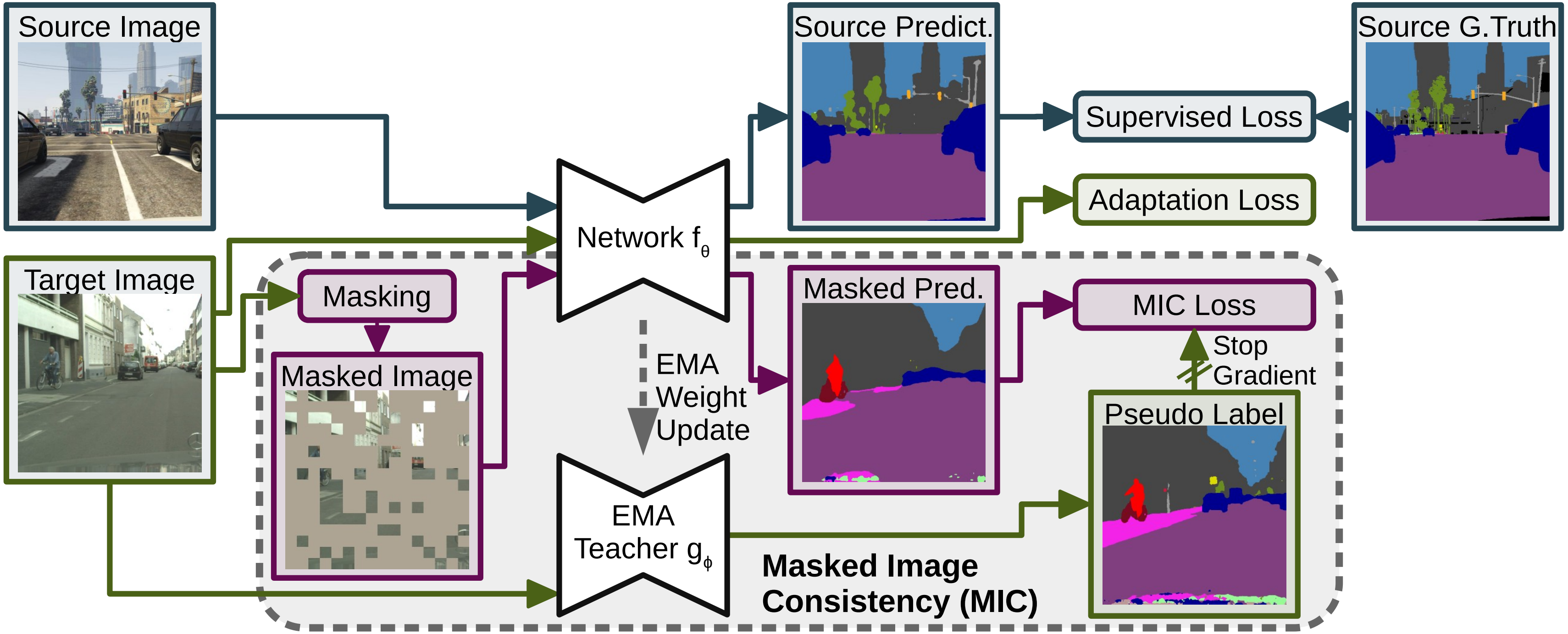
Idea: Enhance learning of spatial context relations (e.g. curb in foreground)

[1] Hoyer et al. “HRDA: Context-aware high-resolution domain-adaptive semantic segmentation”, ECCV 2022.

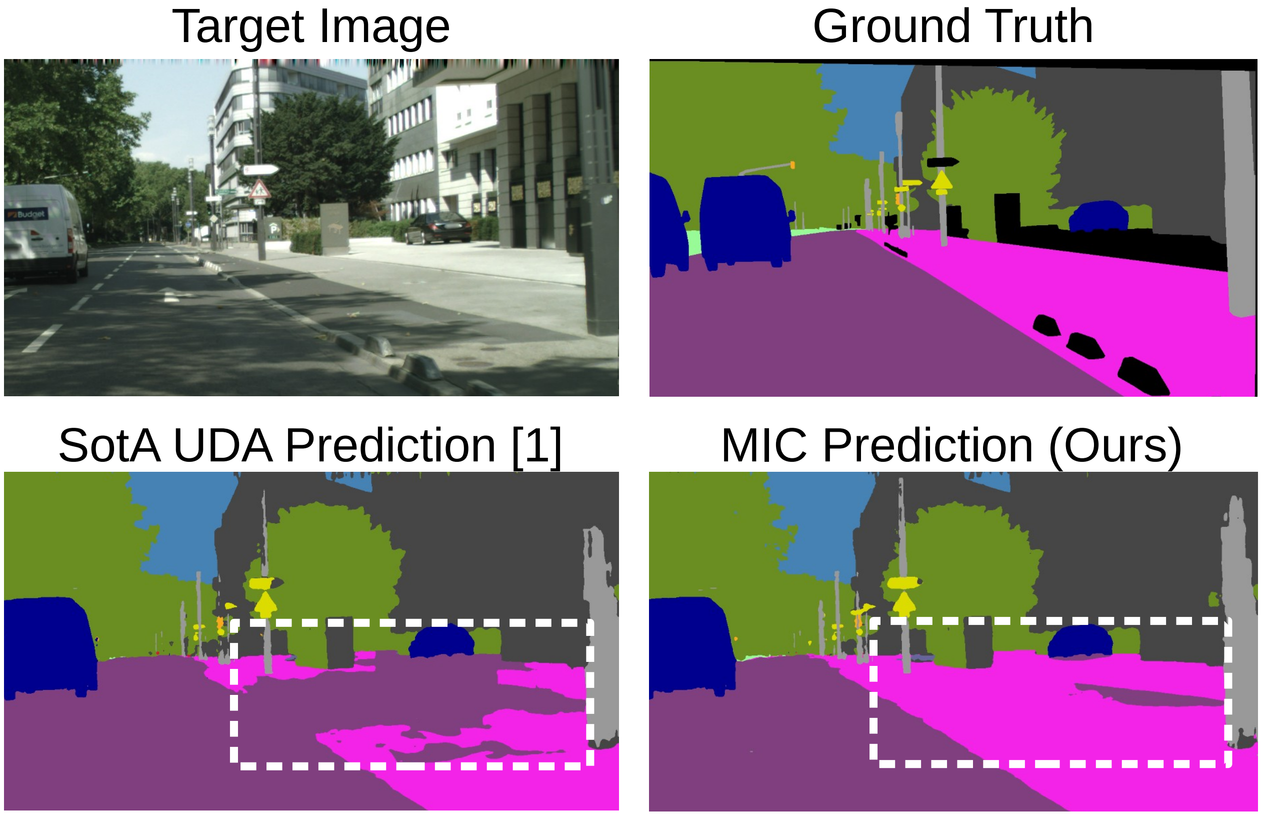
MIC: Method

Masked Image Consistency (MIC) plug-in for UDA

- Randomly mask out target image patches
- Predict semantics of entire image
- ➔ Network learns to utilize context



MIC: Example Prediction



Target Image

Ground Truth

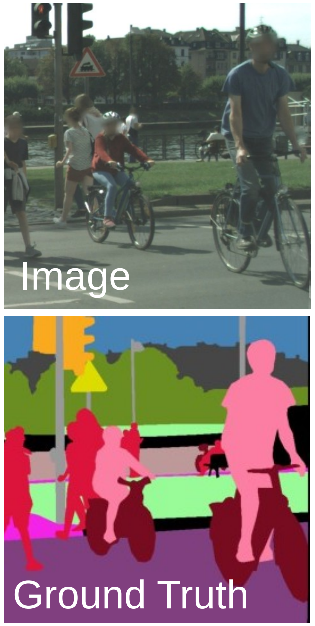
SotA UDA Prediction [1]

MIC Prediction (Ours)

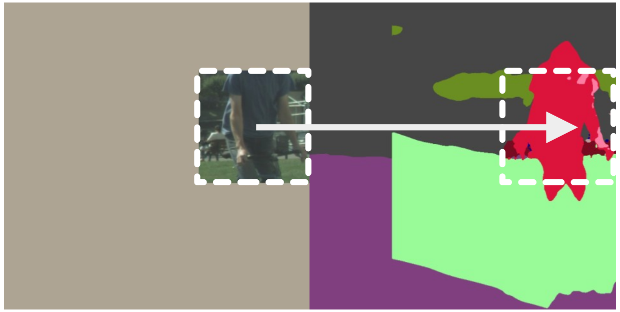
➔ MIC better distinguishes visually similar classes such as road/sidewalk

[1] Hoyer et al. “HRDA: Context-aware high-resolution domain-adaptive semantic segmentation”, ECCV 2022.

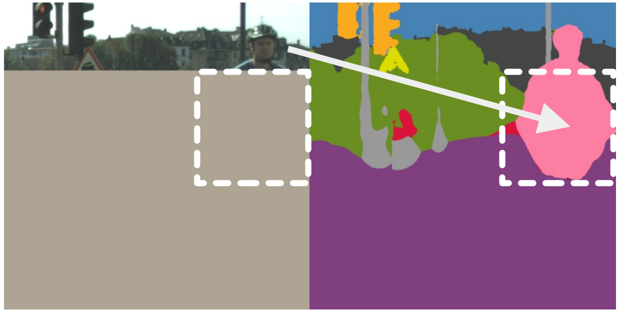
MIC: Predictions from Different Context



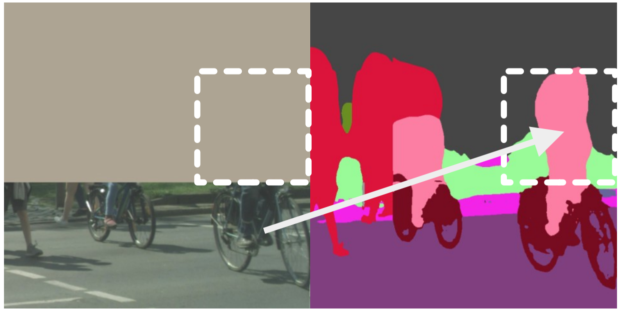
Masked Variants



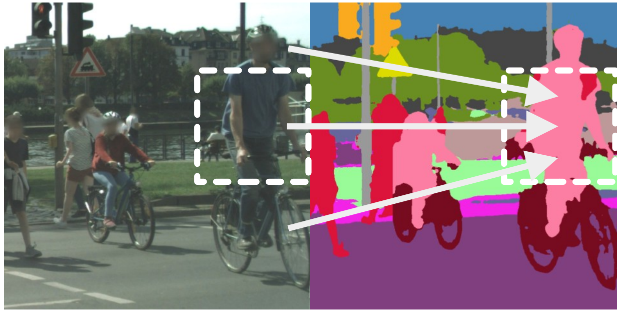
Only local patch
→ Rider is confused with pedestrian



Only context above
→ Rider's body is predicted from helmet

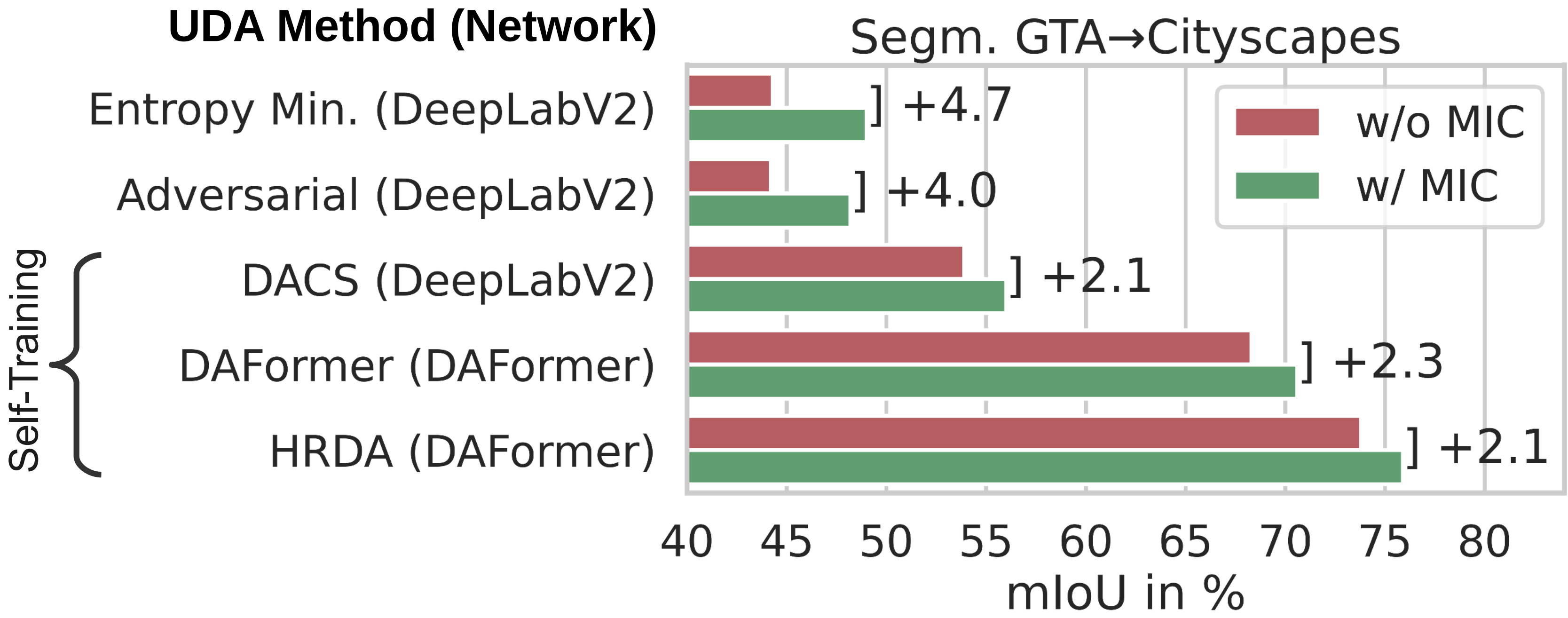


Only context bow
→ Rider's body is predicted from bicycle



Entire image
→ All local and context clues can be used

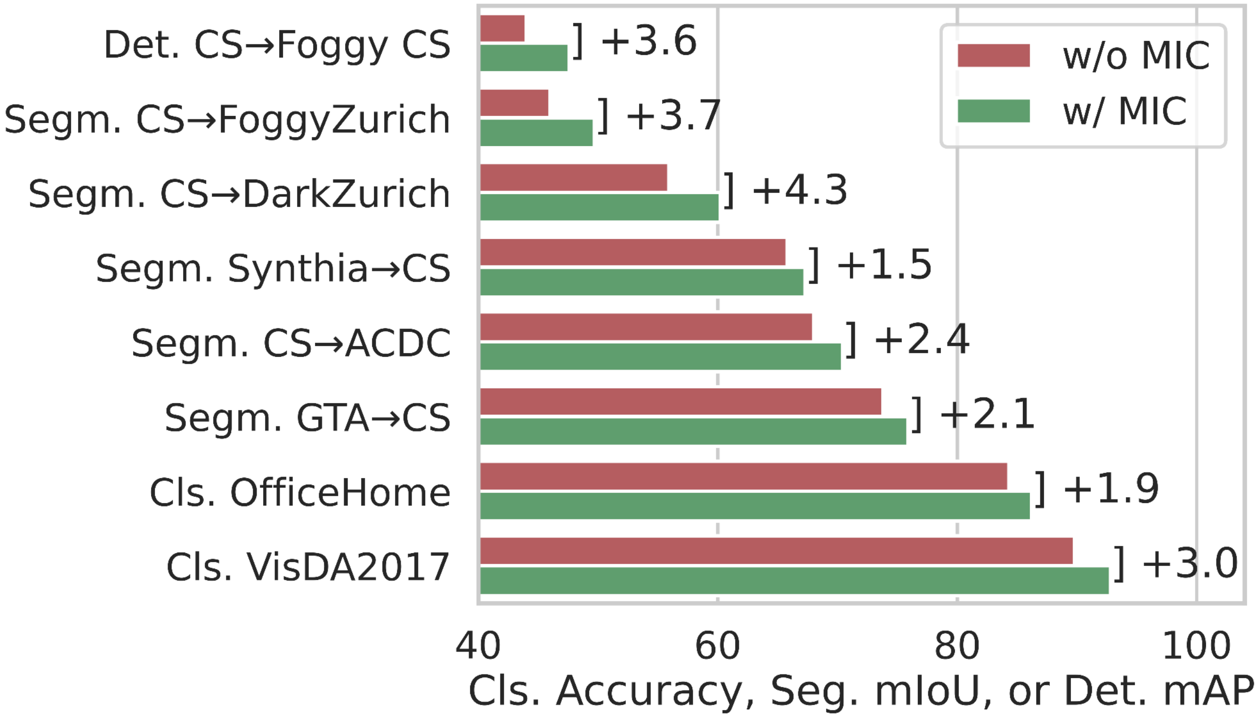
MIC: Evaluation



➔ MIC improves performance across different

- UDA methods
- Network architectures

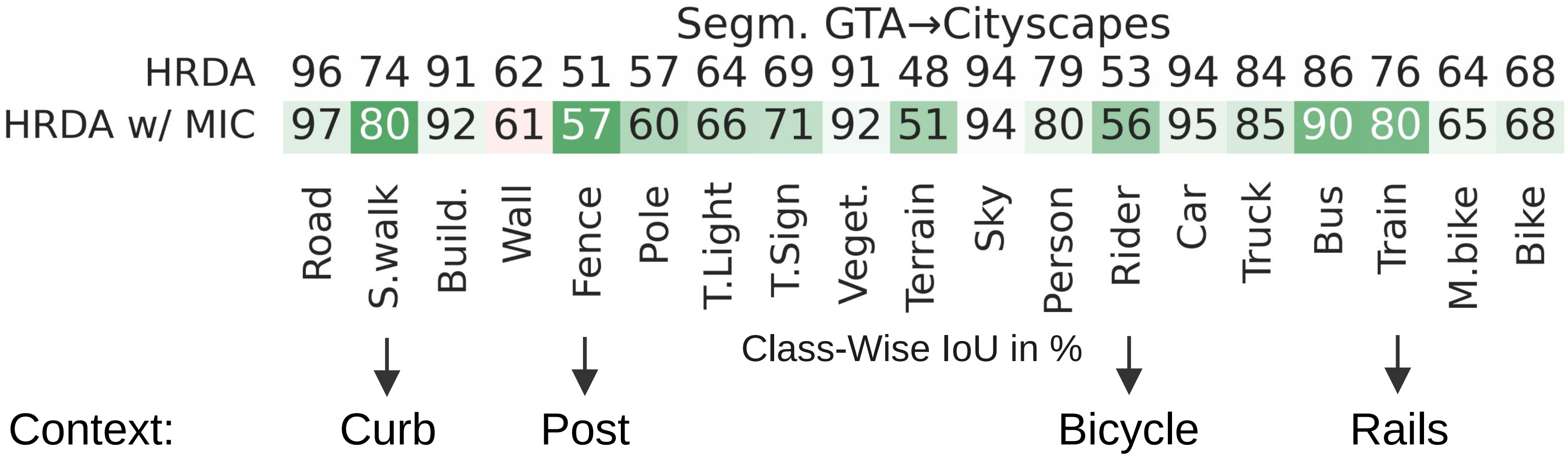
MIC: Evaluation



➔ MIC improves the State of the Art across different

- Vision tasks: classification, segmentation, detection
- Domain gaps: synthetic/real, day/night, clear/adverse-weather

MIC: Evaluation



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
HRDA	96	74	91	62	51	57	64	69	91	48	94	79	53	94	84	86	76	64	68
HRDA w/ MIC	97	80	92	61	57	60	66	71	92	51	94	80	56	95	85	90	80	65	68

Class-Wise IoU in %

Context: Curb Post Bicycle Rails

➔ MIC most improves classes with relevant context clues

MIC: Example Predictions

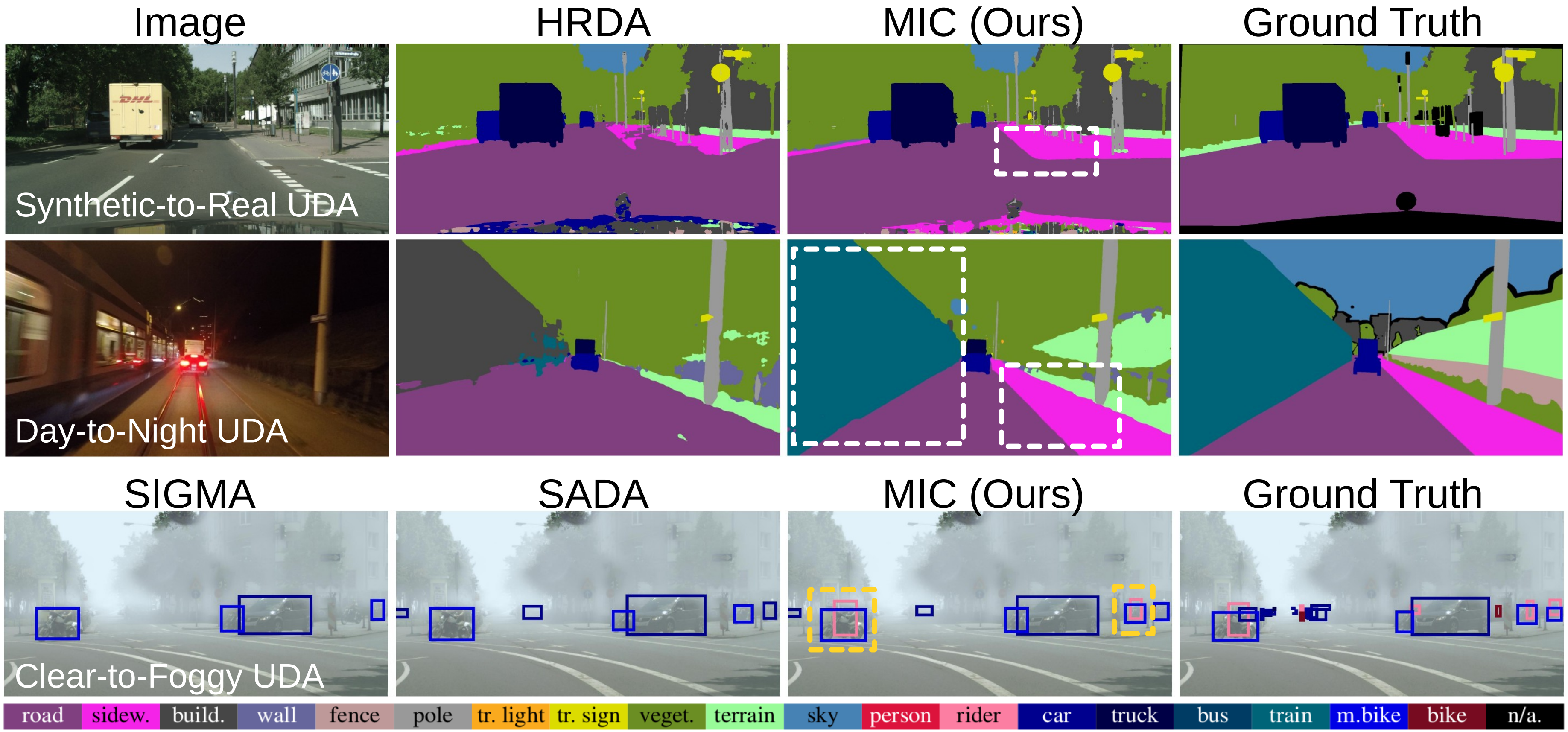


Image HRDA MIC (Ours) Ground Truth

SIGMA SADA MIC (Ours) Ground Truth

The implementation is available at:
github.com/lhoyer/MIC