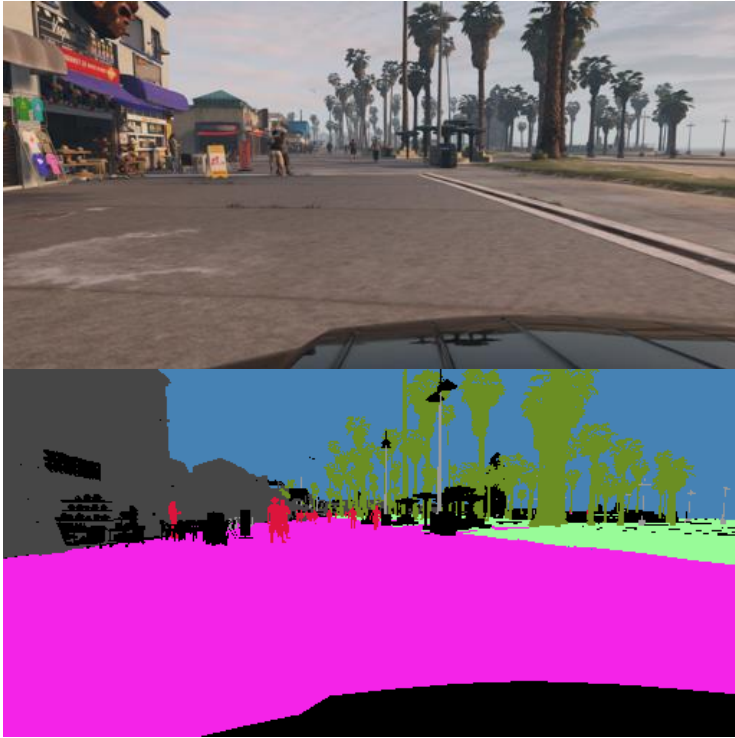


# Visual Domain Adaptation Challenge 2017: Semantic Segmentation

Yiheng Zhang, Zhaofan Qiu, Ting Yao, Tao Mei

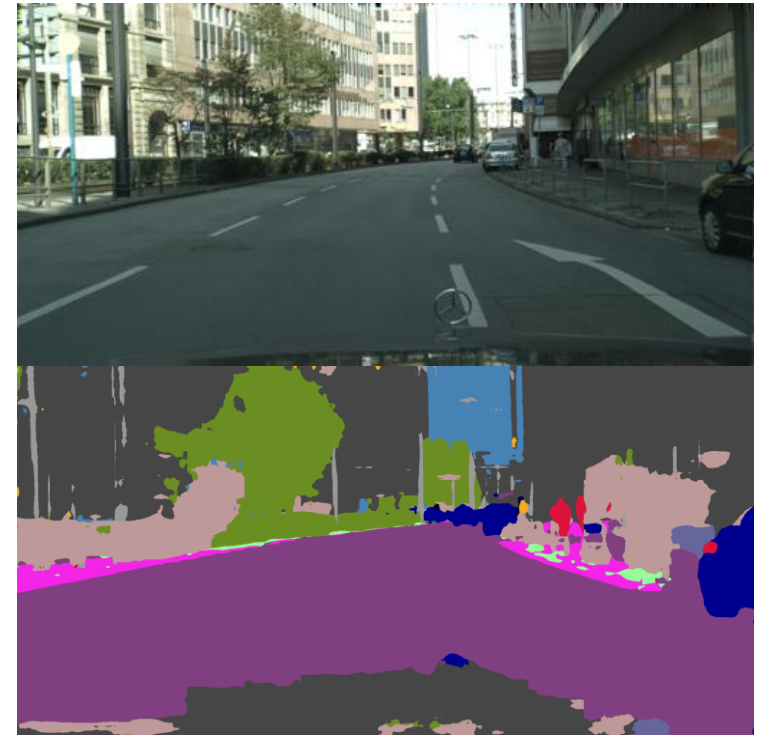
*Multimedia Search and Mining Group @ Microsoft Research Asia*

# Task: Domain Adaptation for Semantic Segmentation



Source Domain: labeled data

Training  
Adaptation



Target Domain: unlabeled data

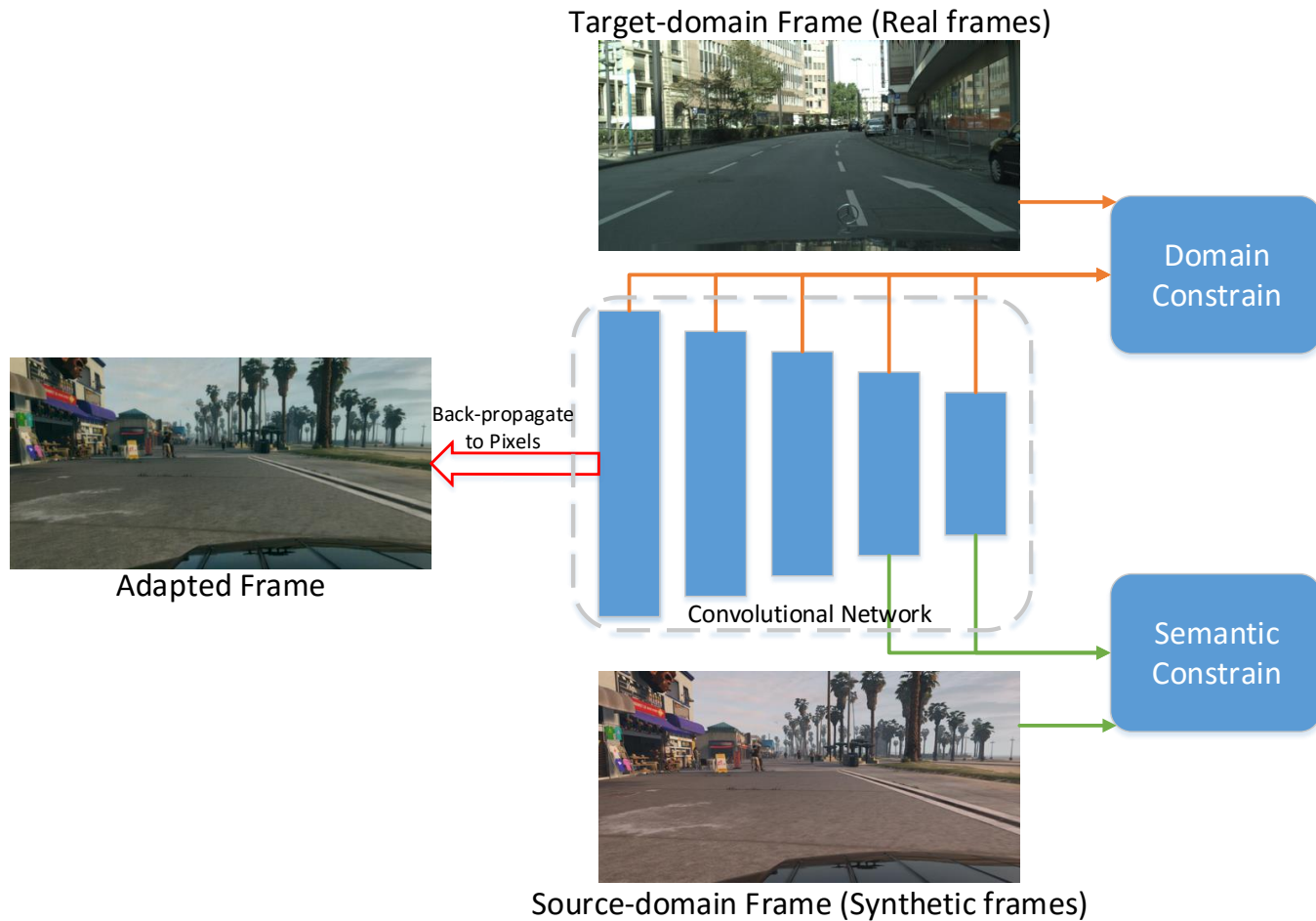
# Outline

- Frame level domain adaptation
  - Explore visual similarity between domains
- Feature level domain adaptation
  - Explore domain invariant representation
- Detailed Implementation for semantic segmentation
  - ResNet + PSP + Multi-scale Test + Ensemble

# Outline

- Frame level domain adaptation
  - Explore visual similarity between domains
- Feature level domain adaptation
  - Explore domain invariant representation
- Detailed Implementation for semantic segmentation
  - ResNet + PSP + Multi-scale Test + Ensemble

# Frame level domain adaptation



- Observation:

- Visually difference between synthetic frames and real frames

- Solution:

- Explicitly adapt frames across domains
- Draw frames with domain and semantic constraints

# Outline

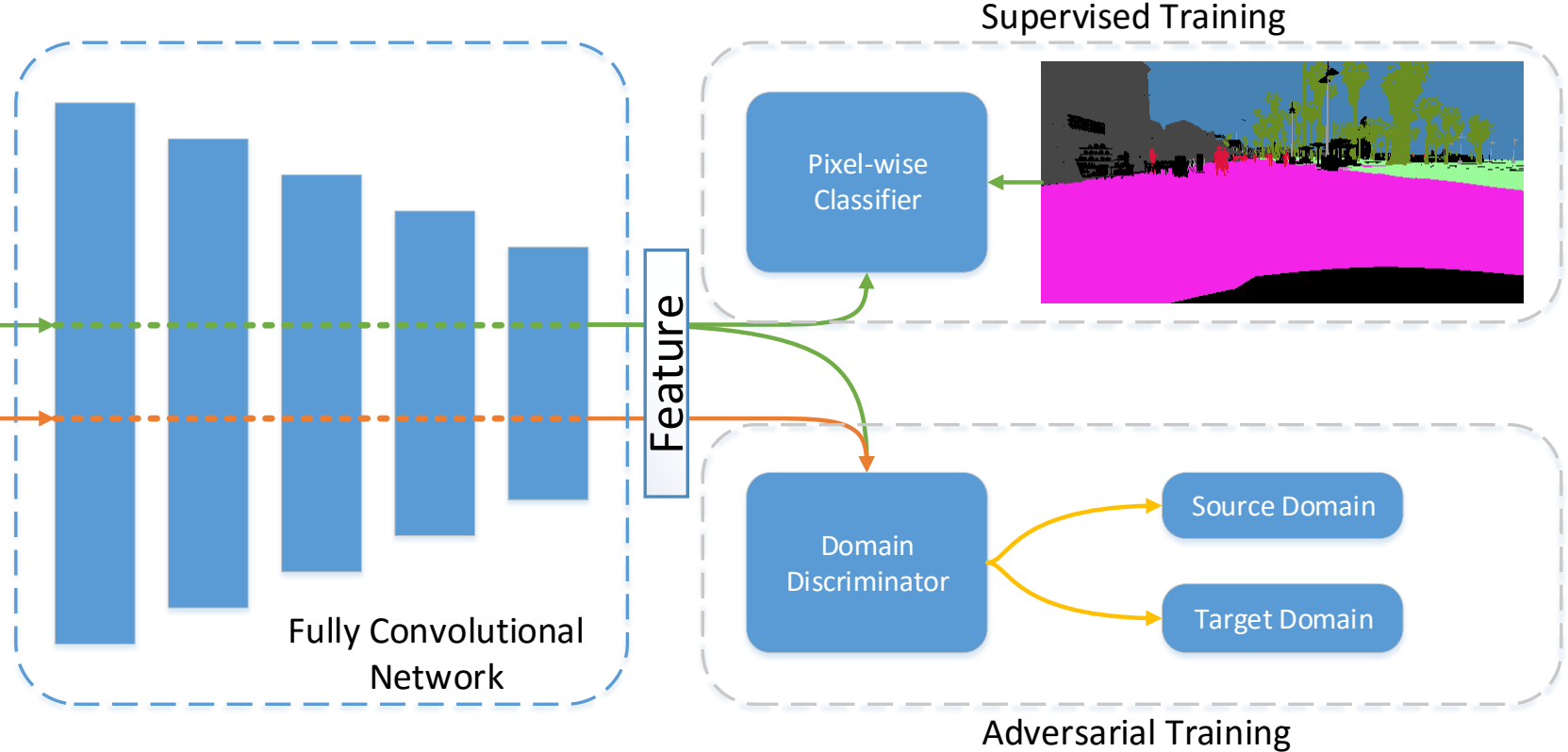
- Frame level domain adaptation
  - Explore visual similarity between domains
- Feature level domain adaptation
  - Explore domain invariant representation
- Detailed Implementation for semantic segmentation
  - ResNet + PSP + Multi-scale Test + Ensemble

# Feature level domain adaptation

Adapted Source-domain Frame



Adapted Target-domain Frame



# Outline

- Frame level domain adaptation
  - Explore visual similarity between domains
- Feature level domain adaptation
  - Explore domain invariant representation
- Detailed Implementation for semantic segmentation
  - ResNet + PSP + Multi-scale Test + Ensemble



# Detailed Implementation

- ImageNet pre-trained ResNet:
  - ResNet-101 [K. He *et al.*, CVPR'16]
  - ResNet-152 [K. He *et al.*, CVPR'16]
  - SE-ResNeXt-101 [J. Hu *et al.*, arXiv'17]
- Pyramid Spatial Pooling (PSP) [H. Zhao *et al.*, CVPR'17]
- Multi-scale Test:
  - Scales: [0.50, 0.75, 1.00, 1.25, 1.50, 1.75]
- Ensemble

# Domain Adaptation Results

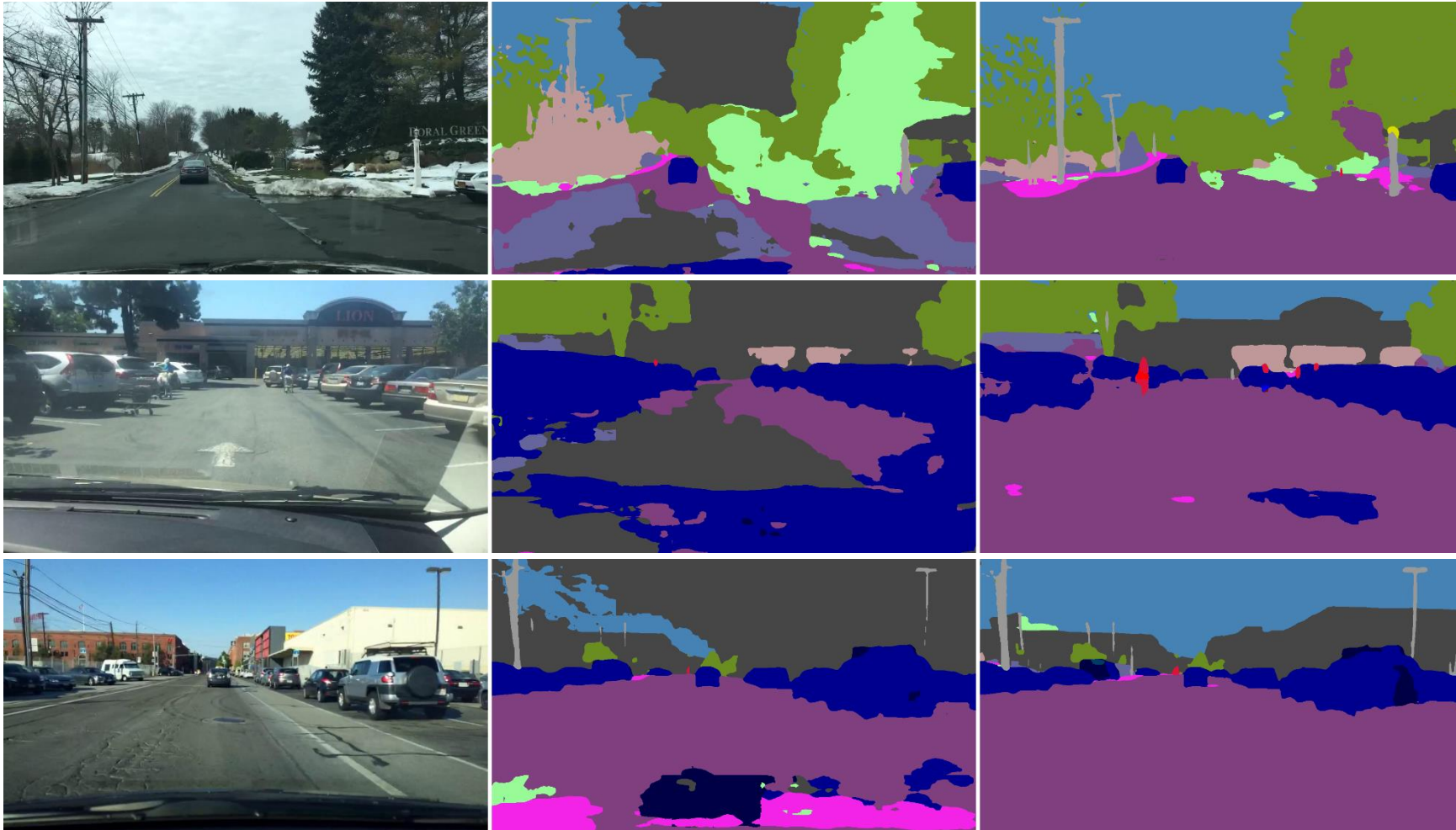


Original Frame

Before Adaptation

After Adaptation

# Domain Adaptation Results



Original Frame

Before Adaptation

After Adaptation

# Evaluations

- Dataset
  - Training Domain: GTA5, 24,966 labeled synthetic frames
  - Validation Domain: Cityscapes, 500 labeled real frames
  - Test Domain: 1,500 unlabeled real frames

Evaluation Results on Test Domain

Model	road	sdwlk	bldng	wall	fence	pole	light	sign	vgttn	trrn	sky	person	rider	car	truck	bus	train	mcycl	bcycl	mIoU
ResNet-101	84.5	36.9	72.9	15.8	23.3	39.4	41.8	36.8	67.1	25.2	89.1	50.5	20.6	77.8	22.1	24.3	22.8	28.5	37.9	43.0
ResNet-152	<b>87.4</b>	38.3	73.4	20.0	29.7	39.9	42.2	<b>36.9</b>	70.9	30.7	89.4	53.3	22.4	79.0	<b>23.5</b>	<b>25.6</b>	11.6	30.2	41.1	44.5
SE-ResNeXt-101	83.4	34.9	72.3	20.8	27.7	39.3	39.3	34.6	67.5	27.1	89.4	49.6	21.7	77.5	21.5	25.1	12.0	25.0	42.1	42.7
Ensemble	87.0	<b>38.5</b>	<b>74.7</b>	<b>23.7</b>	<b>30.5</b>	<b>41.1</b>	<b>45.2</b>	<b>36.9</b>	<b>72.1</b>	<b>32.6</b>	<b>90.4</b>	<b>55.9</b>	<b>26.8</b>	<b>80.0</b>	23.4	25.1	<b>28.7</b>	<b>44.6</b>	<b>46.0</b>	<b>47.5</b>

**Thanks !**