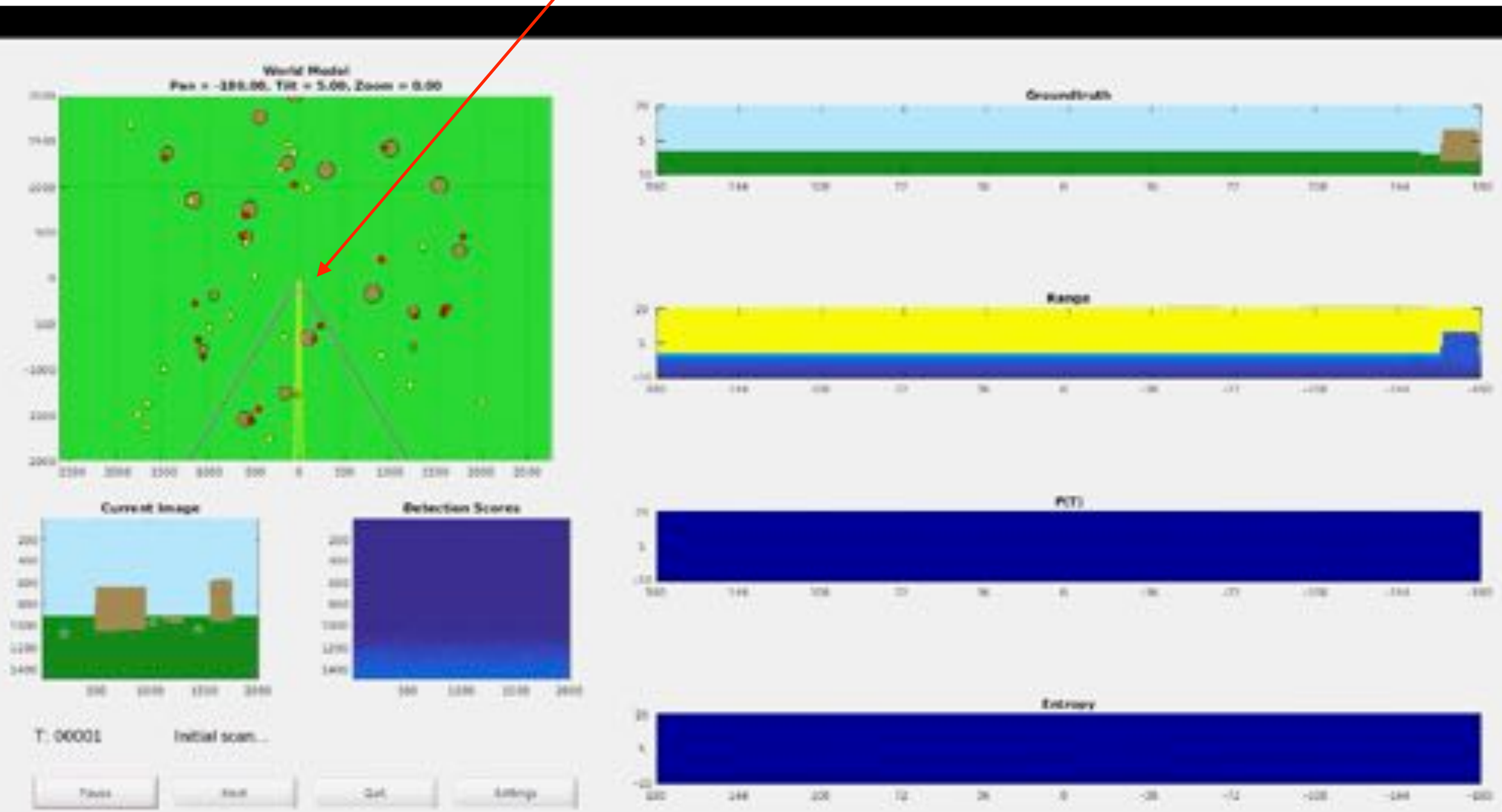
Domain Adaptation & Transfer: All You Need to Use Simulation "for Real"

Boqing Gong

Tencent AI Lab

An intelligent robot



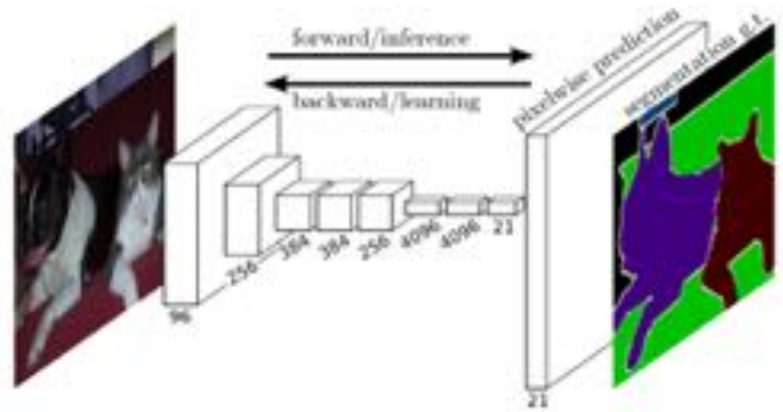
Semantic segmentation of urban scenes

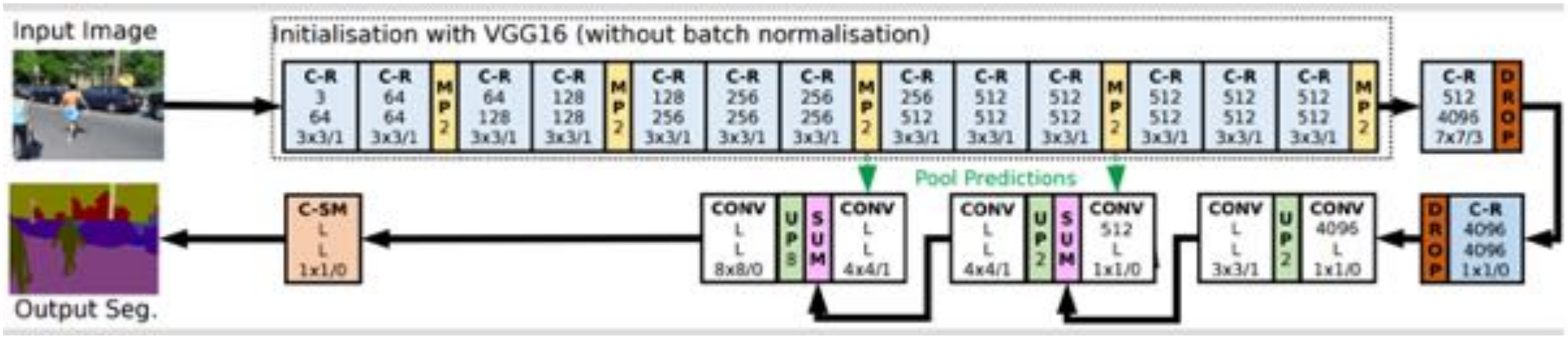
Assign each pixel a semantic label

An appealing application: **self-driving**

Image credit: https://www.cityscapes-dataset.com/

Triumphal approach: CNNs convolutional neural networks





Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

To teach/train CNNs to segment images and videos

About 1.5 hrs to label one such image!

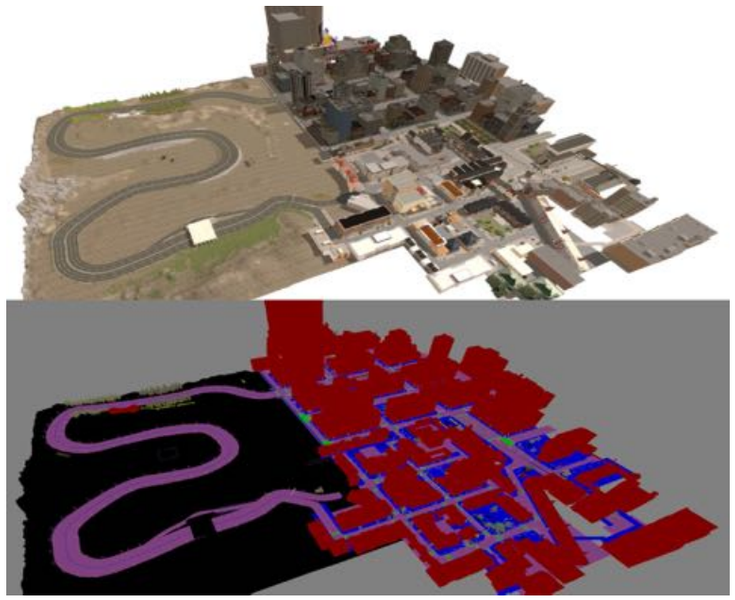
Cityscapes:

30k images captured from 50 cities

Only 5k are well labeled thus far

Image credit: https://www.cityscapes-dataset.com/

Labeling-free training data by simulation



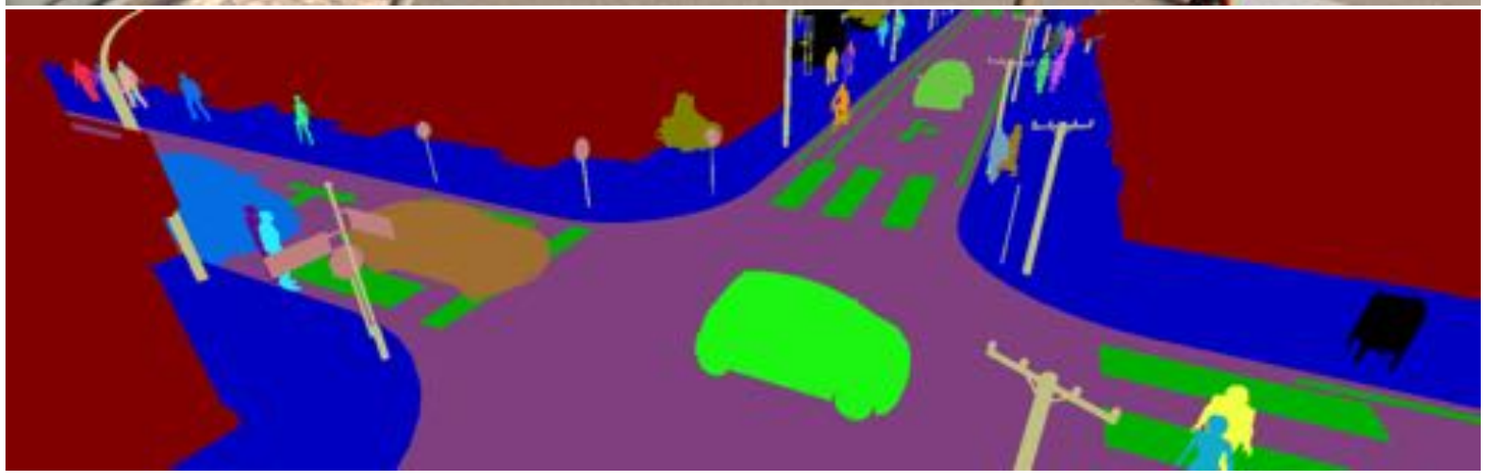
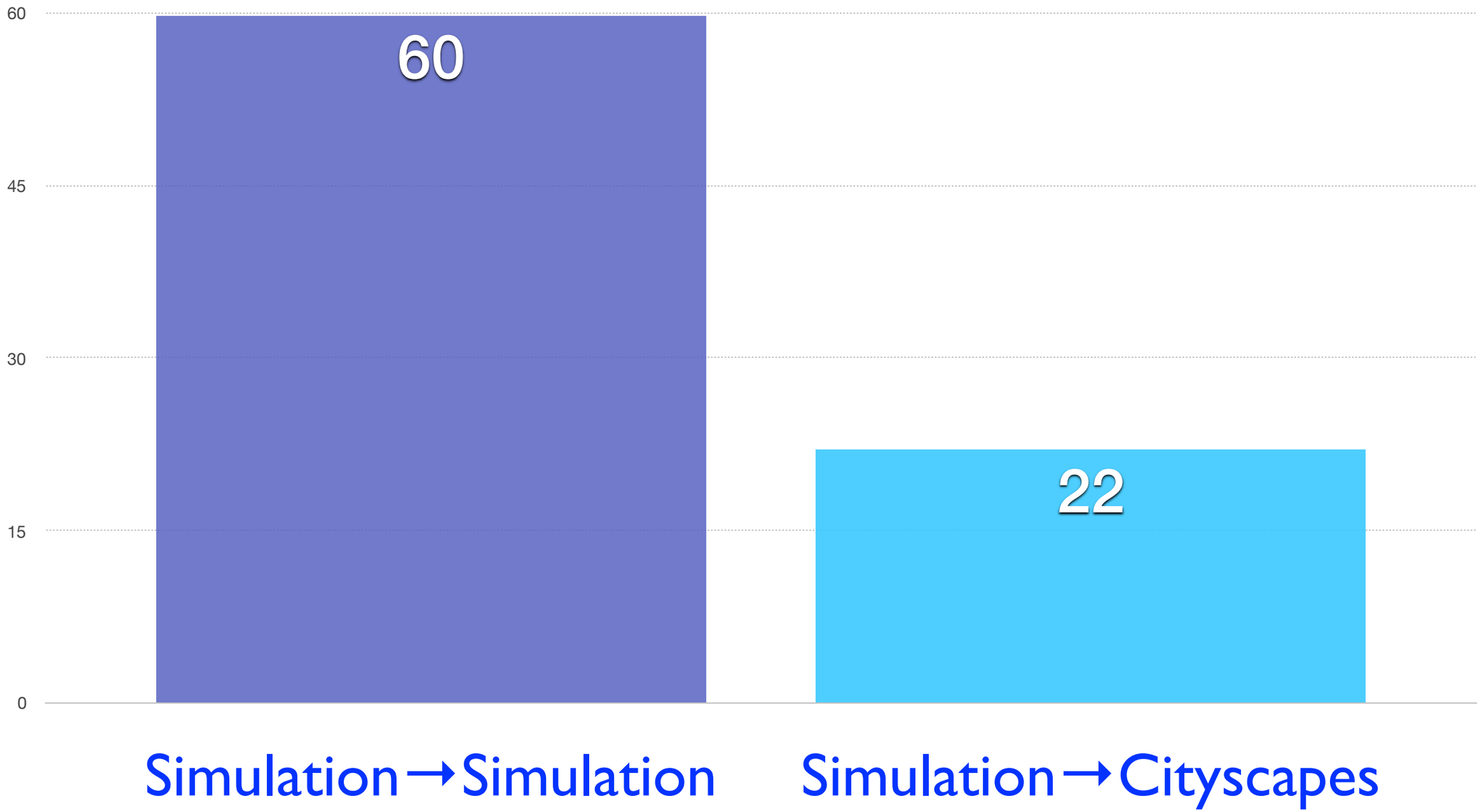


Image credit: http://synthia-dataset.net/

Simulation to real world: catastrophic performance drop



The perils of mismatched domains

Cause: standard assumption in machine learning

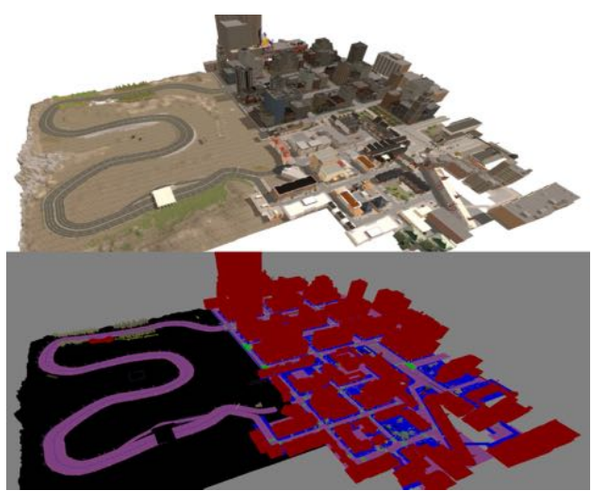
Same underlying distribution for training and testing

Consequence:

Poor cross-domain generalization

Brittle systems in dynamic and changing environment

The perils of mismatched domains



Synthetic imagery → Real photos

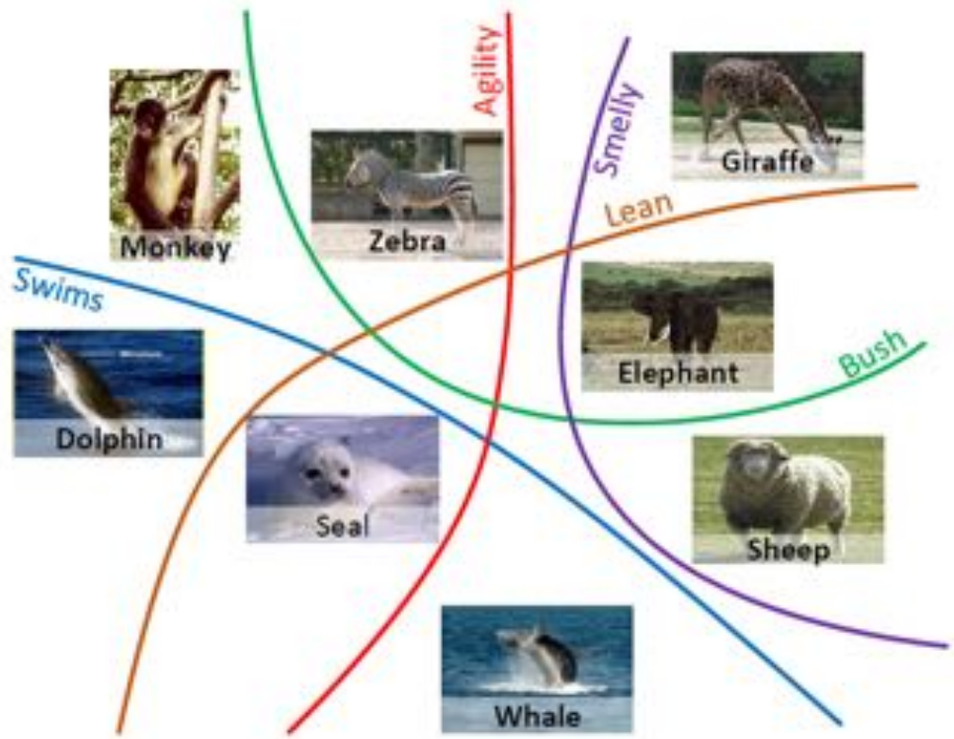
[Zhang et al., ICCV'17]

The perils of mismatched domains

Adapting face detector to a user's album

[Jamal et al., CVPR'18]

The perils of mismatched domains



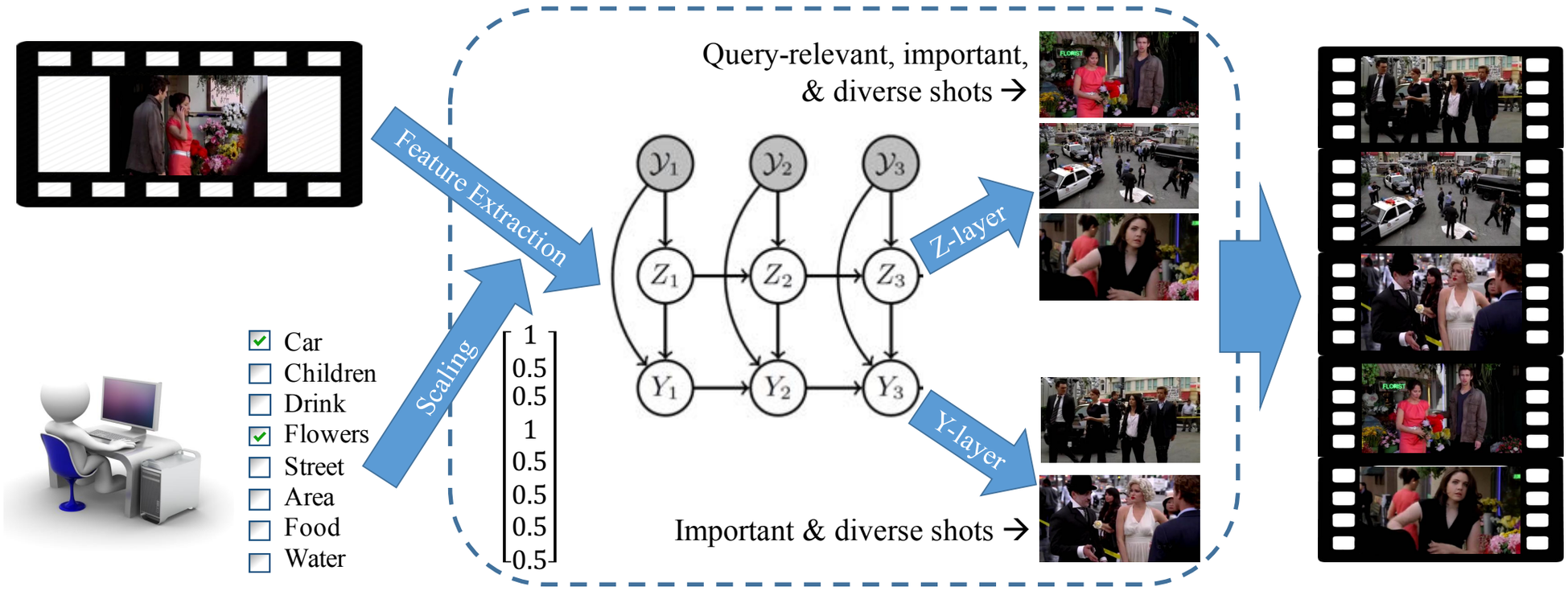
Middle-level concepts describing objects, faces, etc.

Shared by different categories

Attribute detection

[Gan et al., CVPR'17]

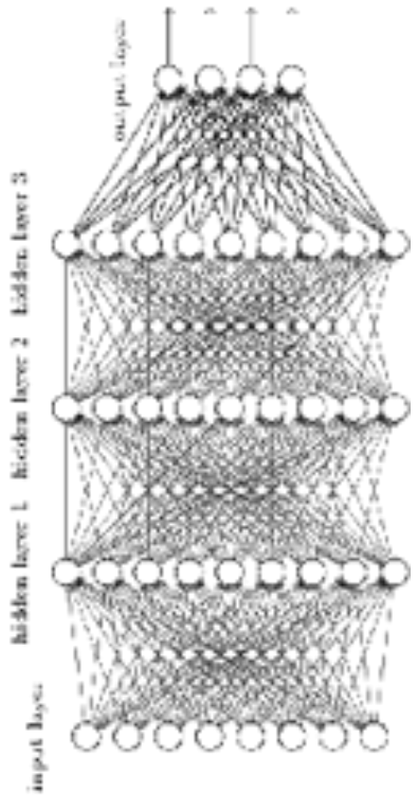
The perils of mismatched domains



Personalization of video summarizers

[Sharghi et al., ECCV'16, CVPR'17, ECCV'18]

The perils of mismatched domains



Webly supervised learning

[Ding et al., WACV'18]

[Gan et al., ECCV'16, CVPR'18]

Abstract form: *unsupervised* domain adaptation (DA)

Setup

Source domain (with labeled data)

$$D_S = \{(x_m, y_m)\}_{m=1}^{M} \sim P_S(X, Y)$$

Target domain (no labels for training)

$$D_T = \{(x_n, \textbf{?})\}_{n=1}^{N} \sim P_T(X, Y)$$

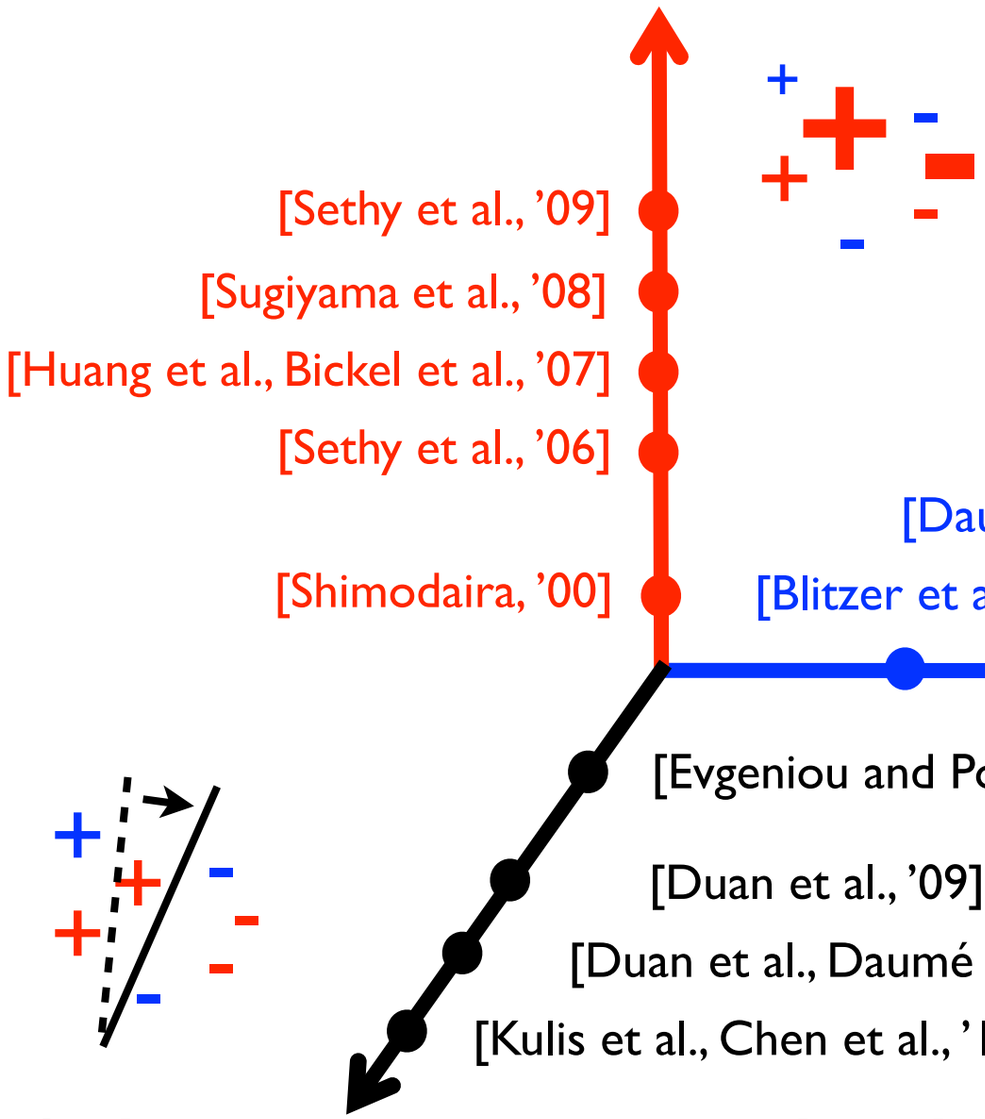
Different distributions

Objective

Learn models to work well on target

Existing methods

Correcting ***sampling*** bias



[Sethy et al., '09]

[Sugiyama et al., '08]

[Huang et al., Bickel et al., '07]

[Sethy et al., '06]

[Shimodaira, '00]

[Pan et al., '09]

[Argyriou et al, '08]

[Daumé III, '07]

[Blitzer et al., '06]

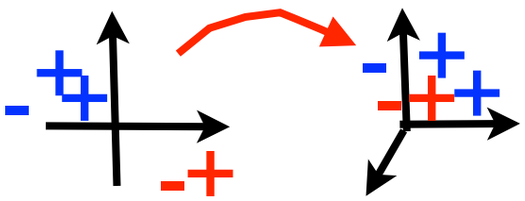
[Muandet et al., '13]

[Gong et al., '12]

[Chen et al., '12]

[Gopalan et al., '11]

Inferring domain-invariant ***features***



[Evgeniou and Pontil, '05]

[Duan et al., '09]

[Duan et al., Daumé III et al., Saenko et al., '10]

[Kulis et al., Chen et al., '11]

Adjusting mismatched ***models***

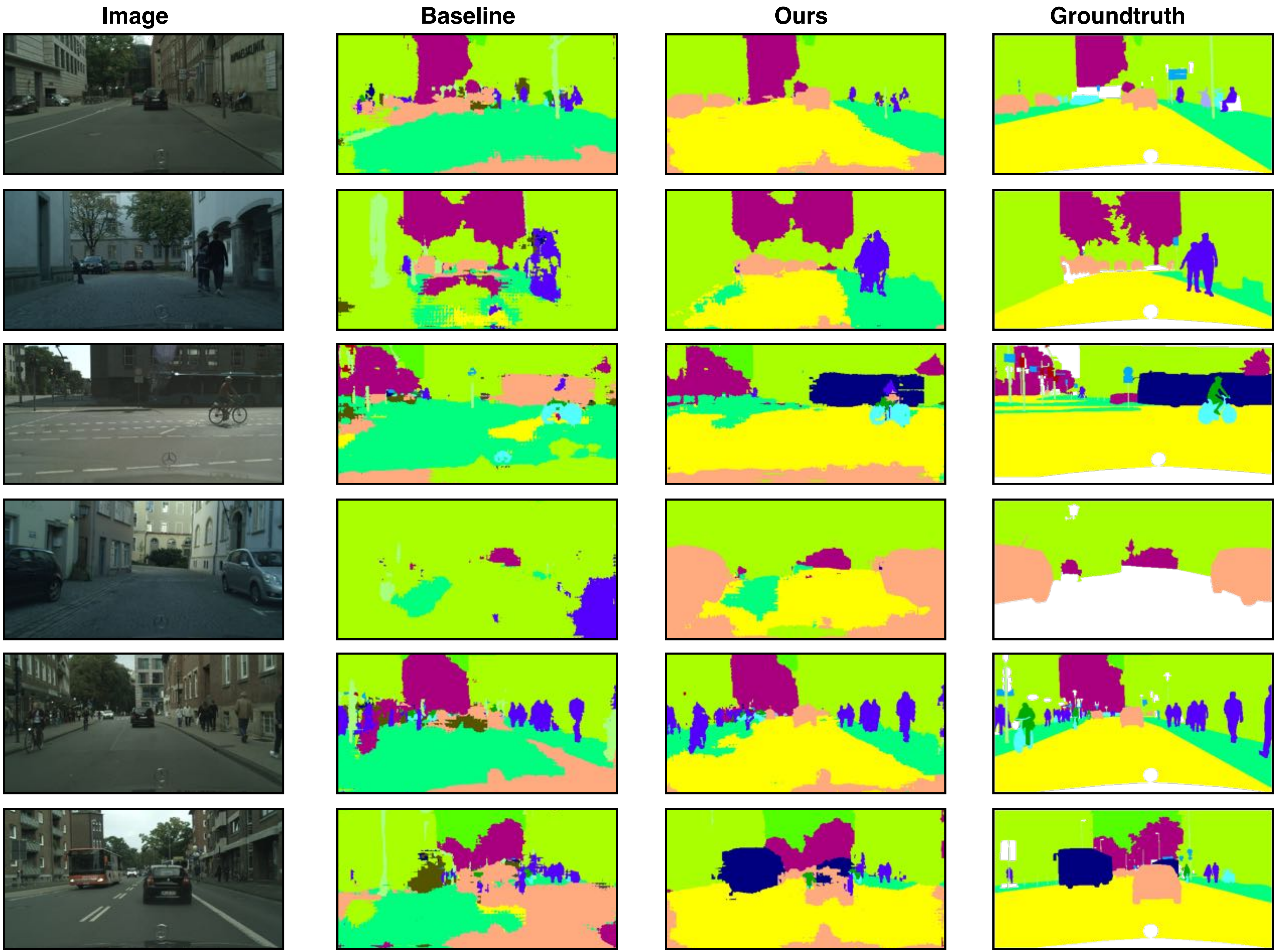
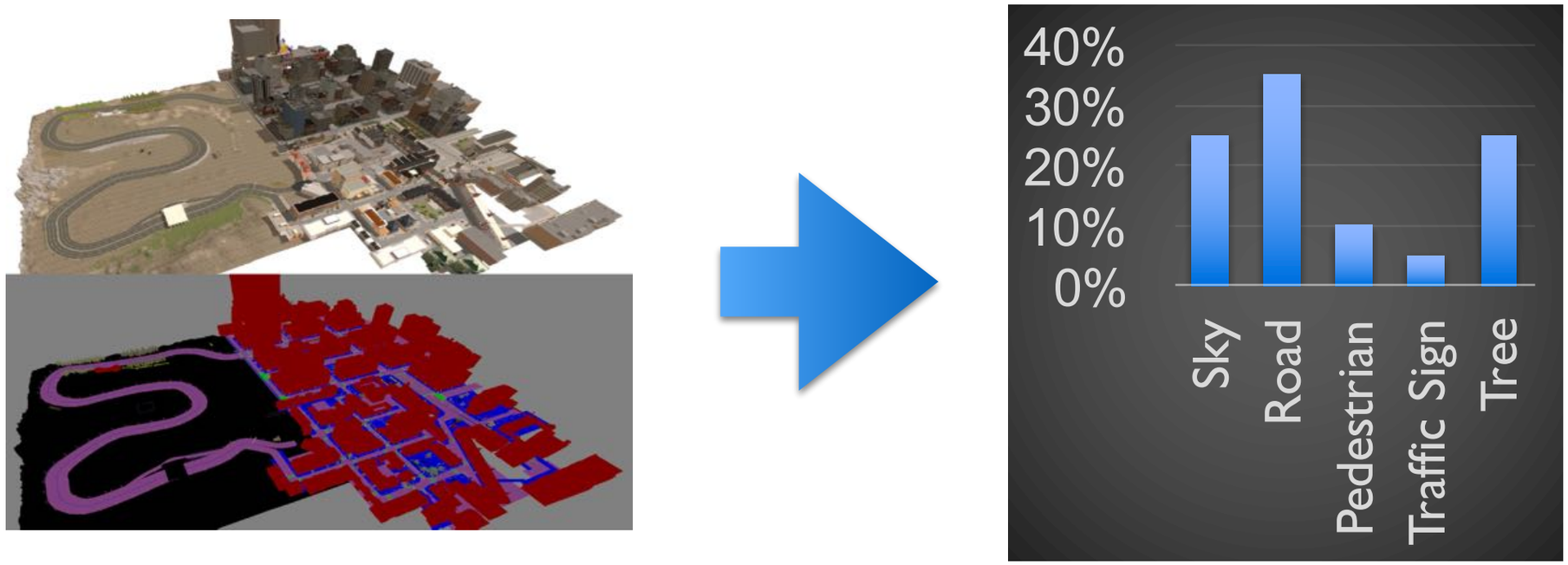


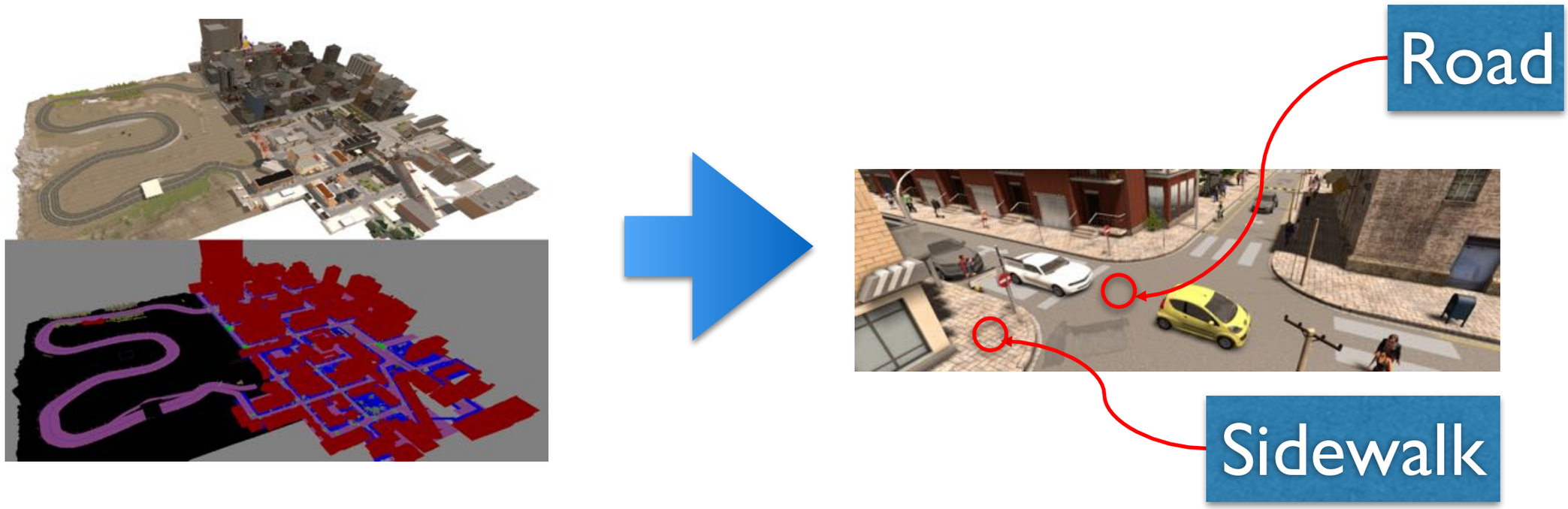
Image
Baseline
Ours
Groundtruth

Let teacher model hint segmentation net (student)



Input: An urban scene image
Algorithm: Logistic regression
Output: Label distributions

Let 2nd teacher model hint segmentation net (student)

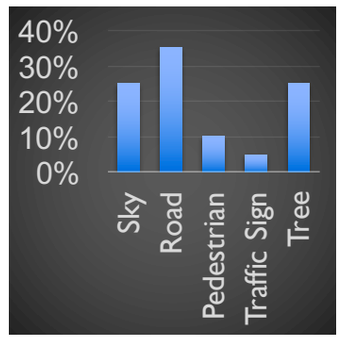


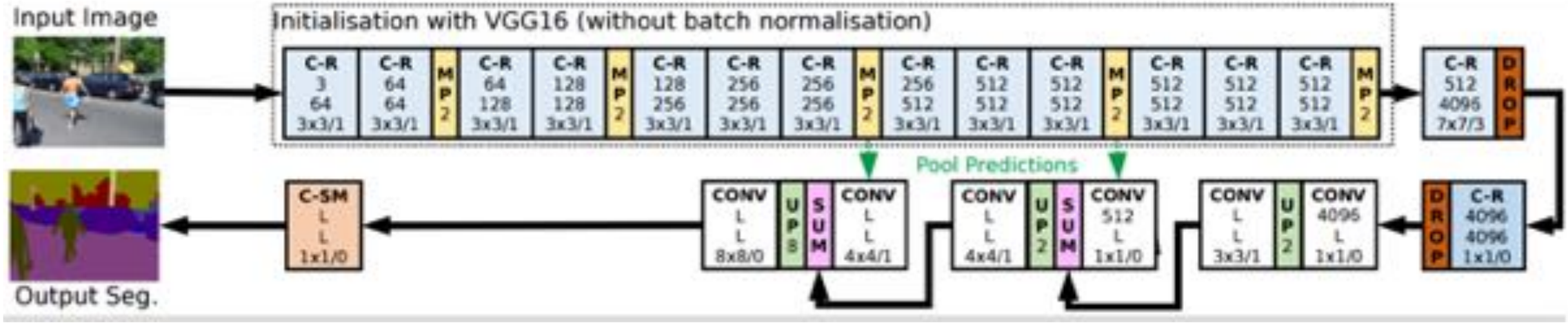
Input: An urban scene image
Algorithm: Super-pixel + Logistic regression
Output: Labels of some super-pixels

Curriculum domain adaptation *for training CNNs*

$$\min_{\Theta} \; \mathcal{L}(Y_s, \widehat{Y}_s) + d(p_t, \widehat{p}_t(\widehat{Y}_t))$$

s : Source, t : Target

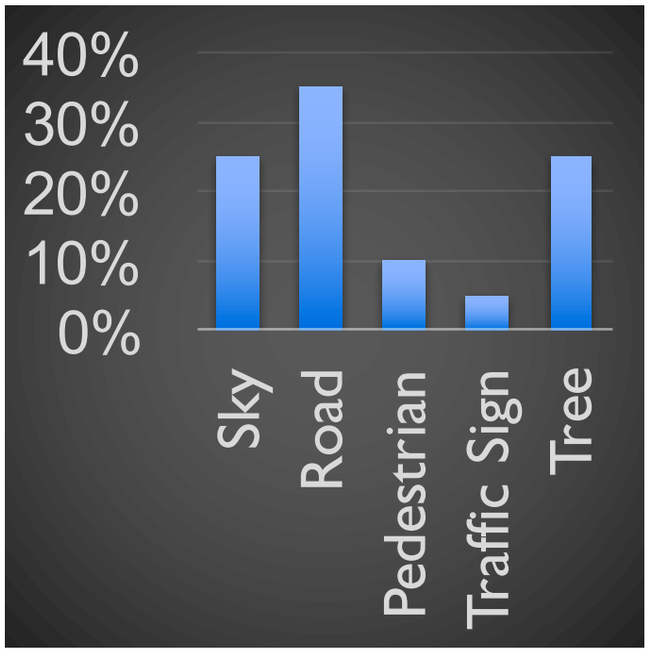




$\widehat{Y}$

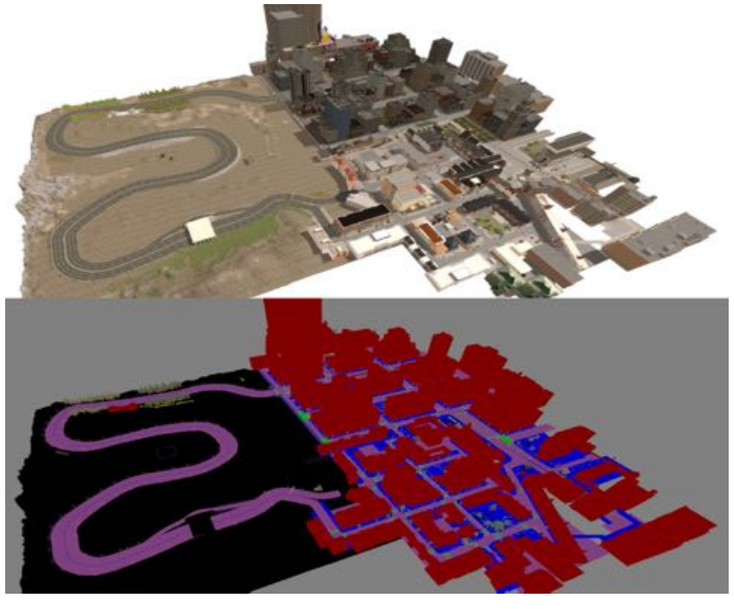
[ICCV'17]

Curriculum domain adaptation



C

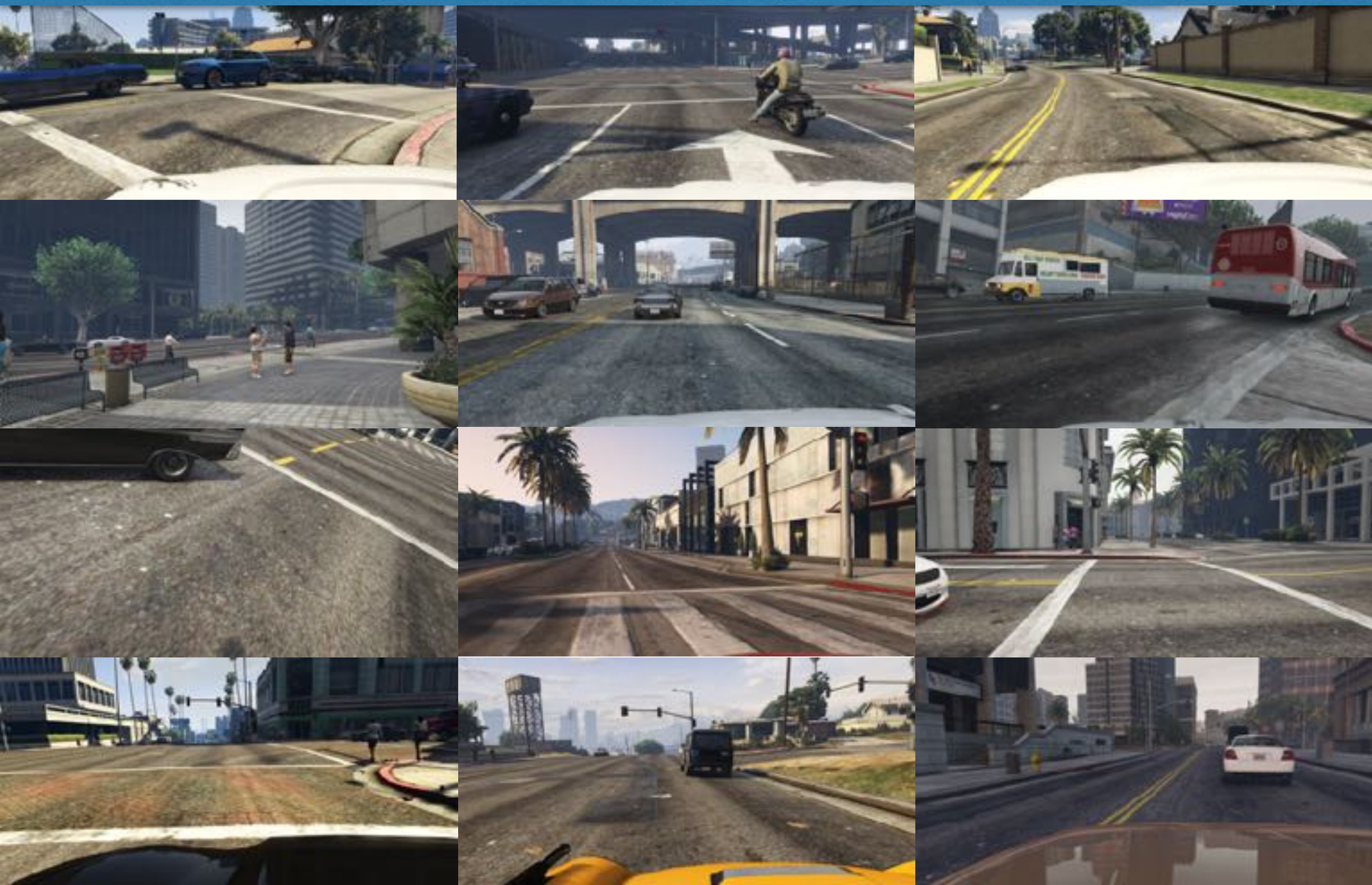
B



A

Cityscapes:
Train/val/test: 2993/503/1531

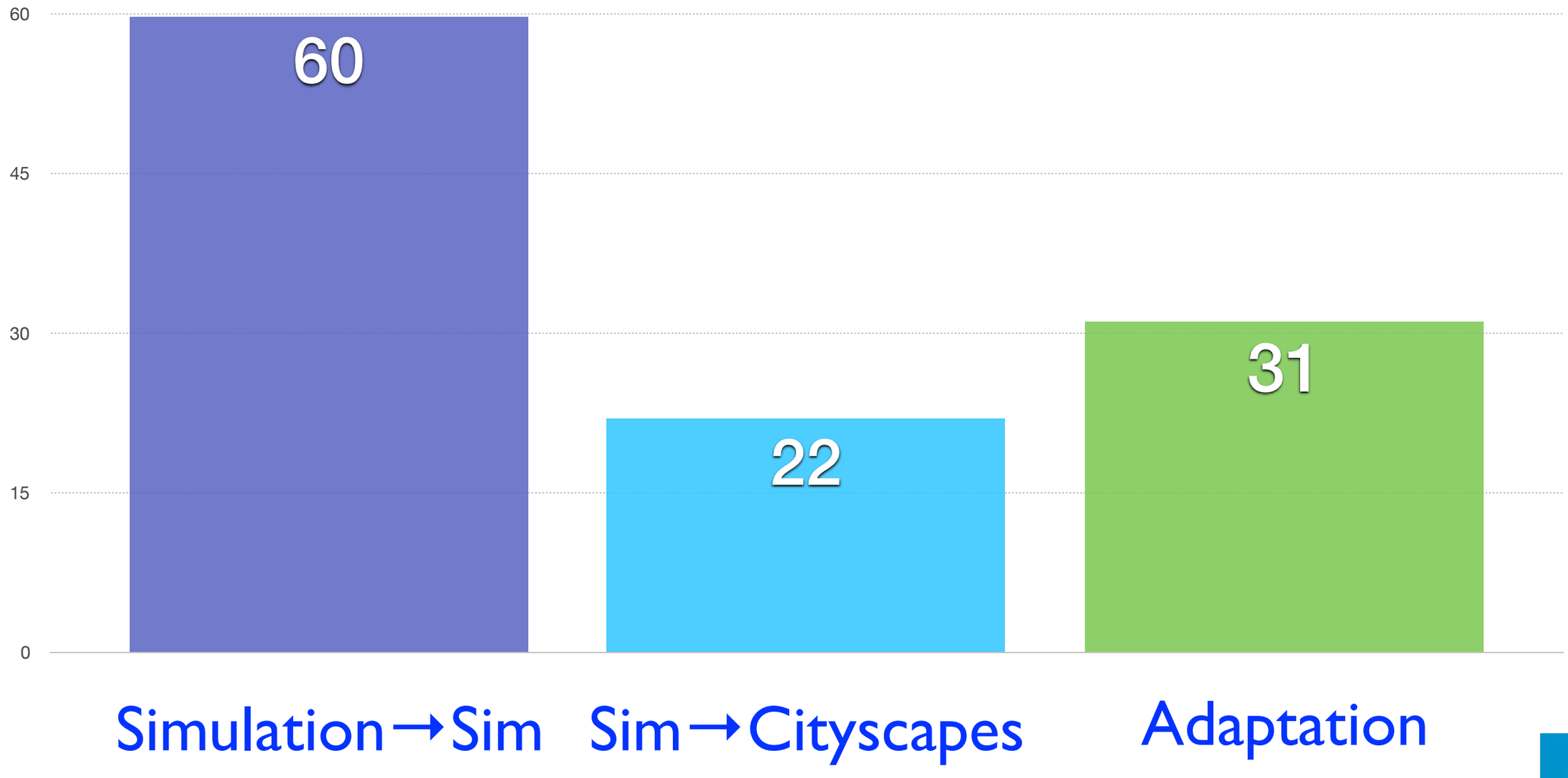
GTA: 24,996 images from the video game



SYNTHIA: 9,400 images

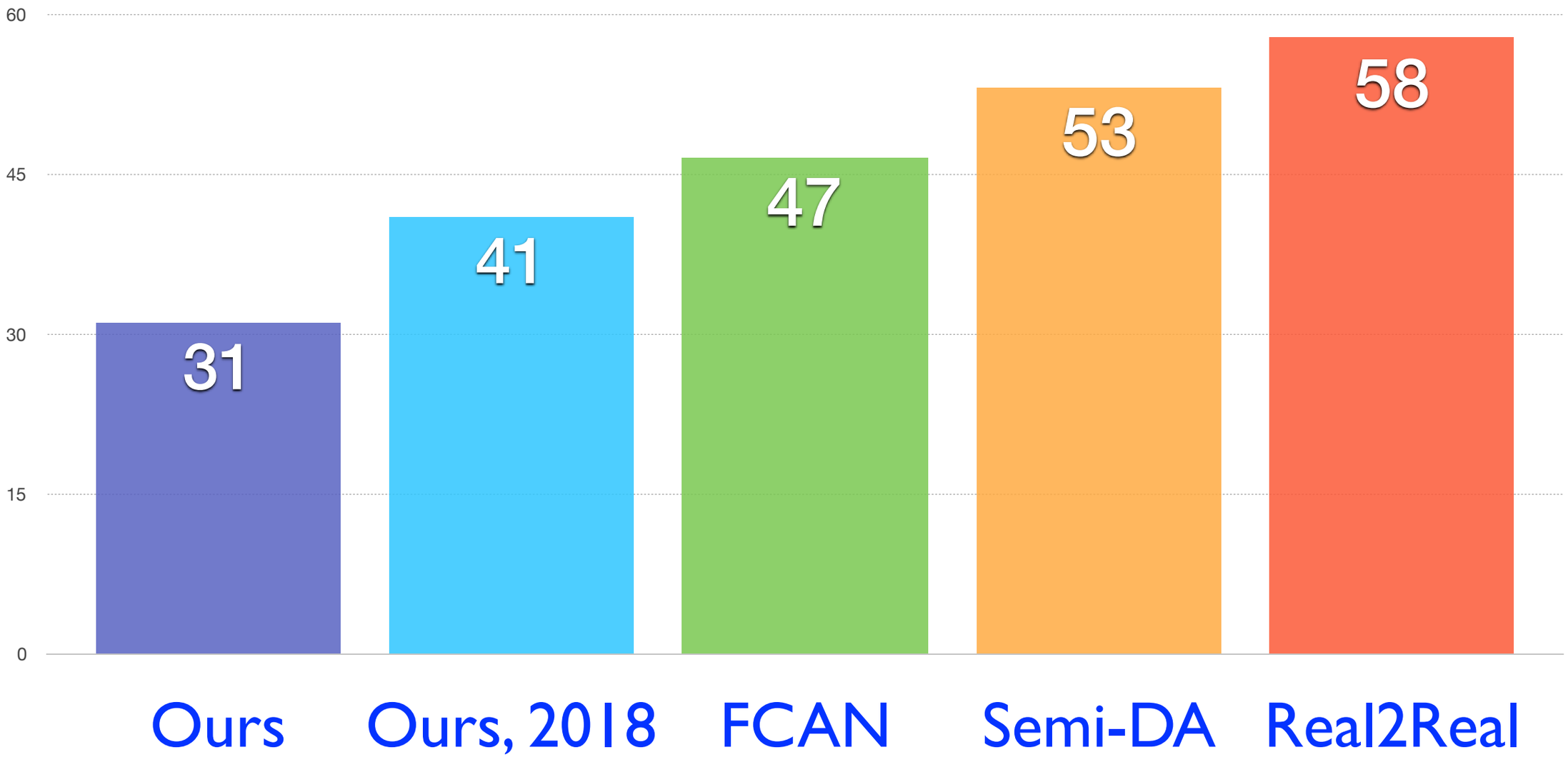
Simulation to real world:

~~catastrophic~~ performance drop

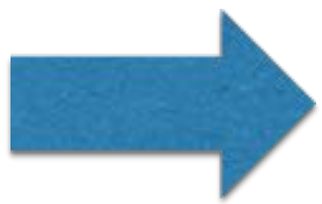


[Zhang et al., ICCV'17]

Recent progress



Domain adaptation: key to use simulation "for real"

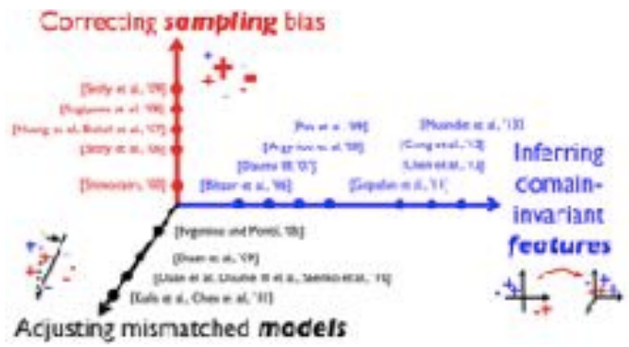


Tencent AI Lab

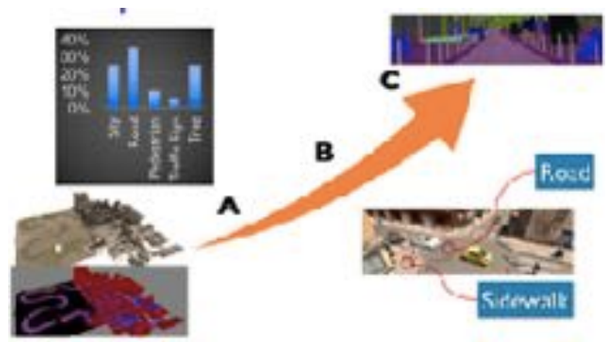
Simulation to reality for segmentation, detection, Dynamics planning & control, etc.

Domain adaptation: key to use simulation "for real"

Domain-invariant **features**
Importance sampling of **data**
Adapt background **models**
etc.

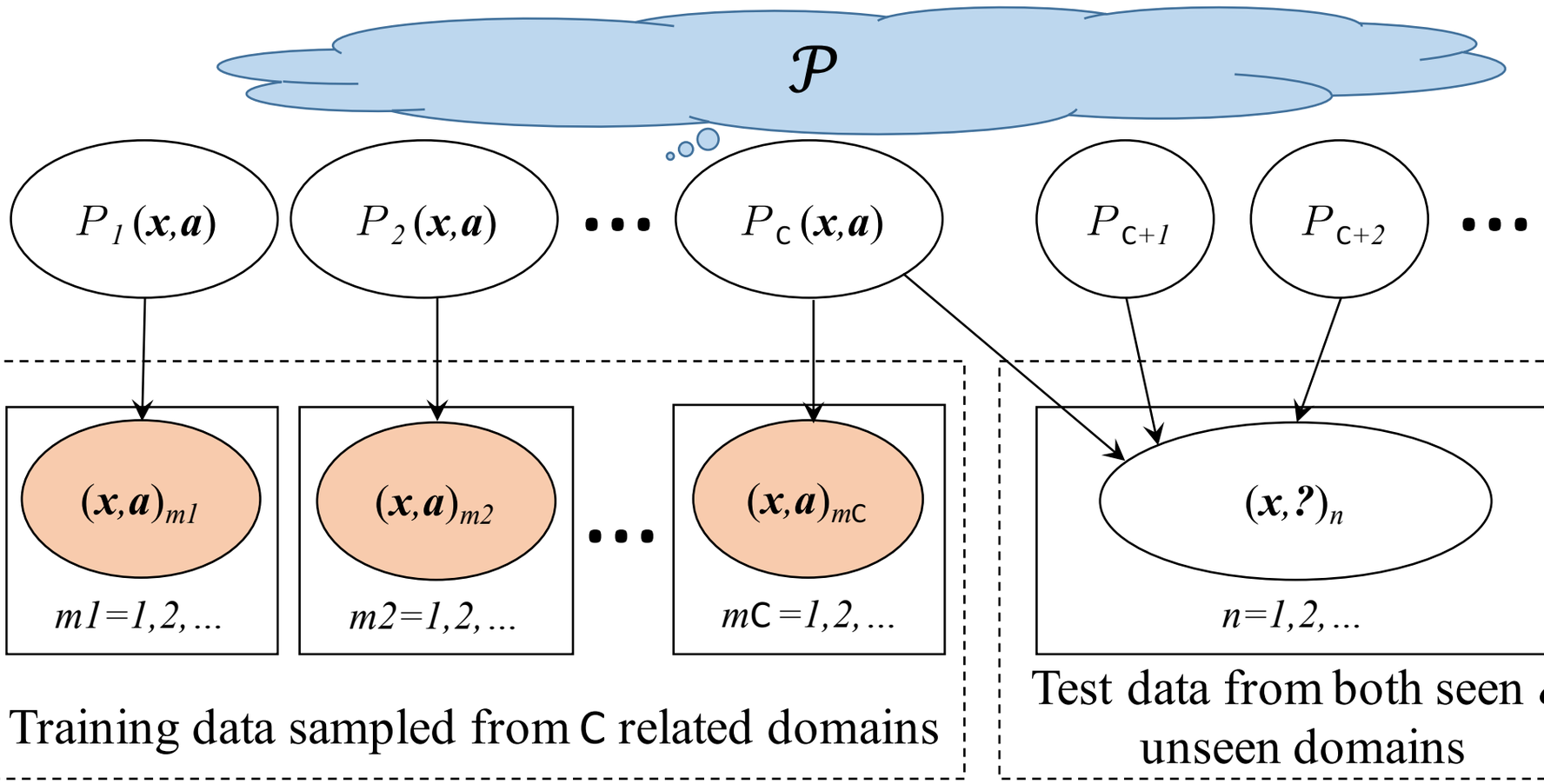


Curriculum domain adaptation
Style transfer, etc.

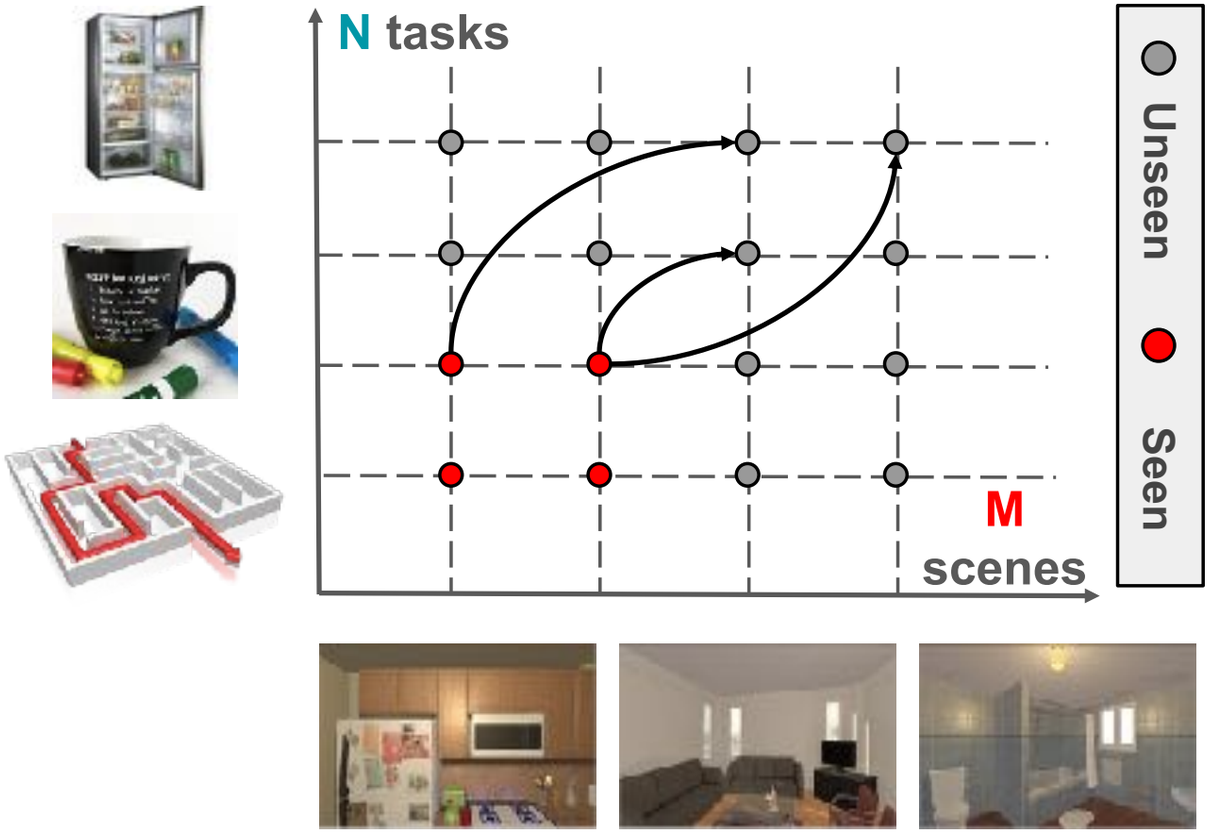


Simulation to reality for segmentation, detection, Dynamics planning & control, etc.

Domain adaptation → domain generalization



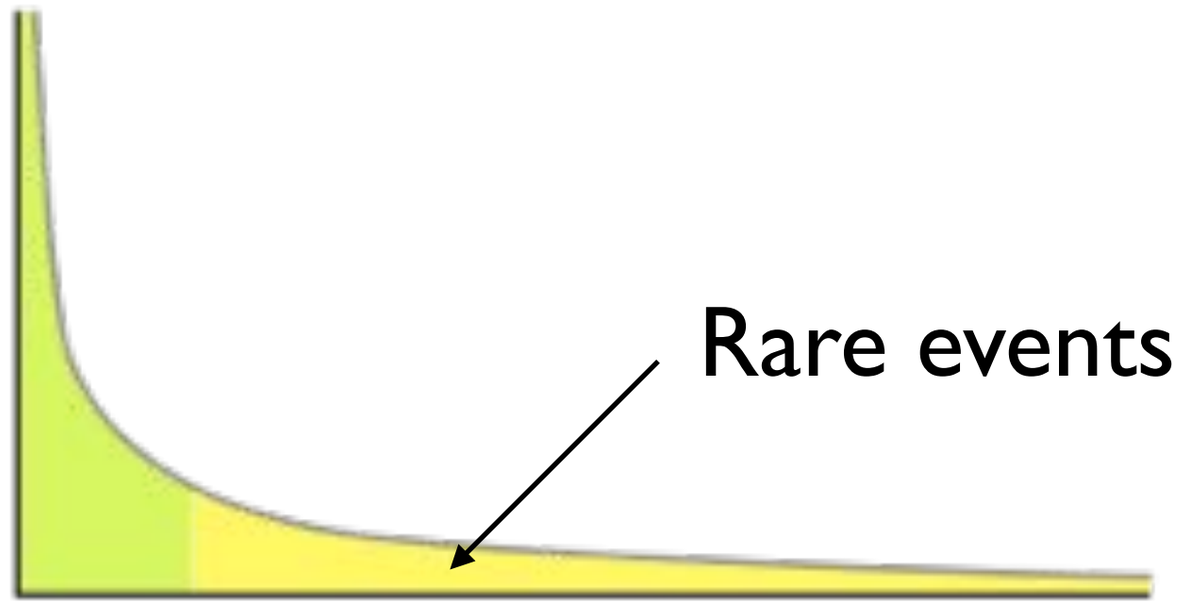
Simulation for domain generalization



Synthesize Policy for Transfer and Adaptation across Environments and Tasks

[NIPS'18, Spotlight]

What to simulate?



What to simulate?

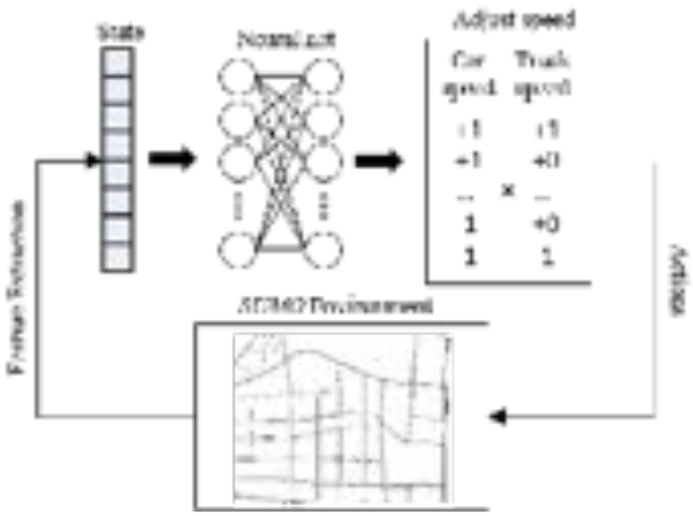
Active Simulation

More data, better model

Simulator

Reality

Actively **tune simulator**



[*Proof-of-concept paper submitted*]

Thank you!

Visual Learning and Embodied Agents in Simulation Environments

ECCV 2018 Workshop, Munich, Germany

Sunday, 9th September, 08:45 AM to 05:30 PM, Room: N1095Z3 at TU München