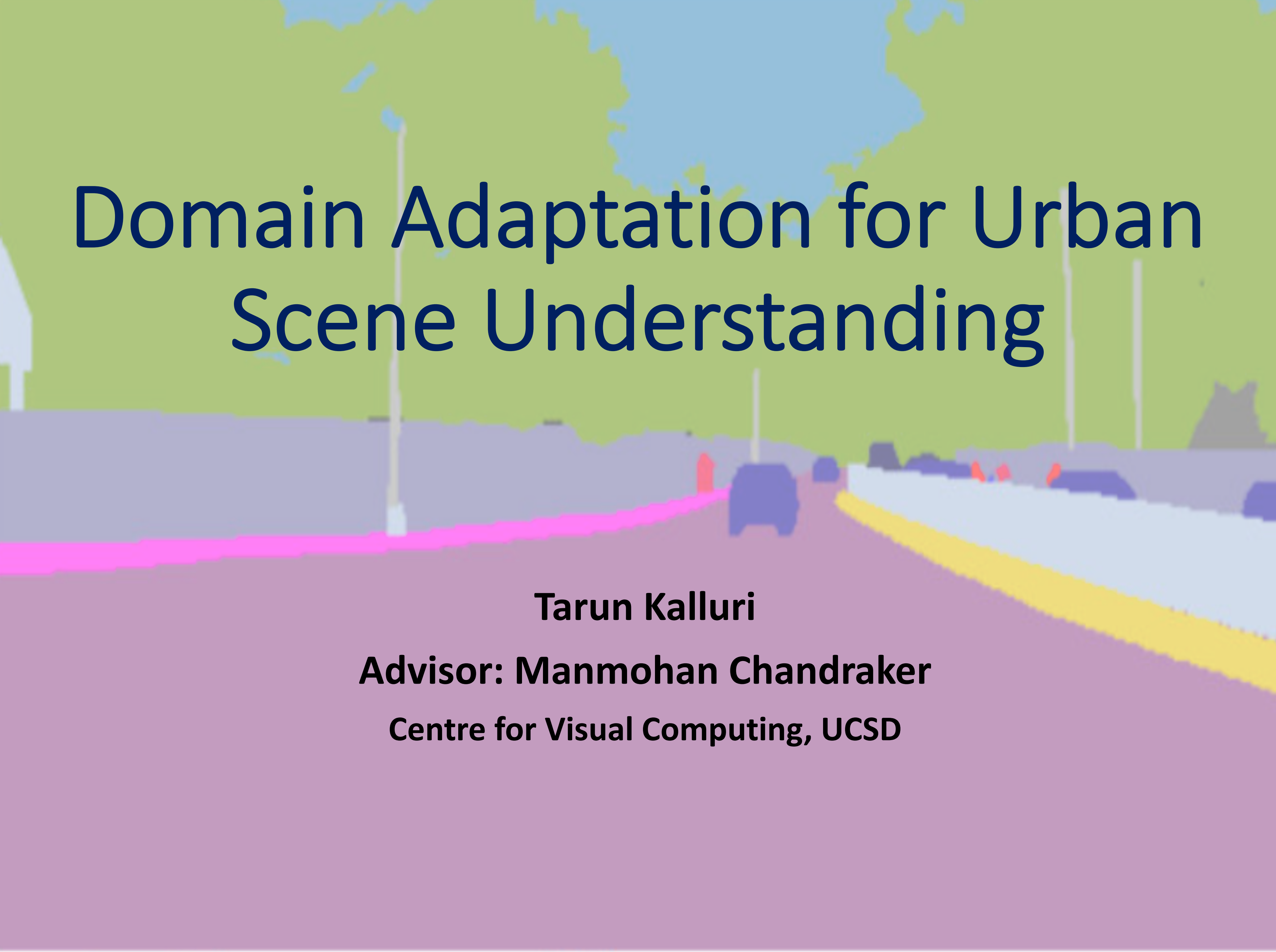


Domain Adaptation for Urban Scene Understanding

Tarun Kalluri

Advisor: Manmohan Chandraker

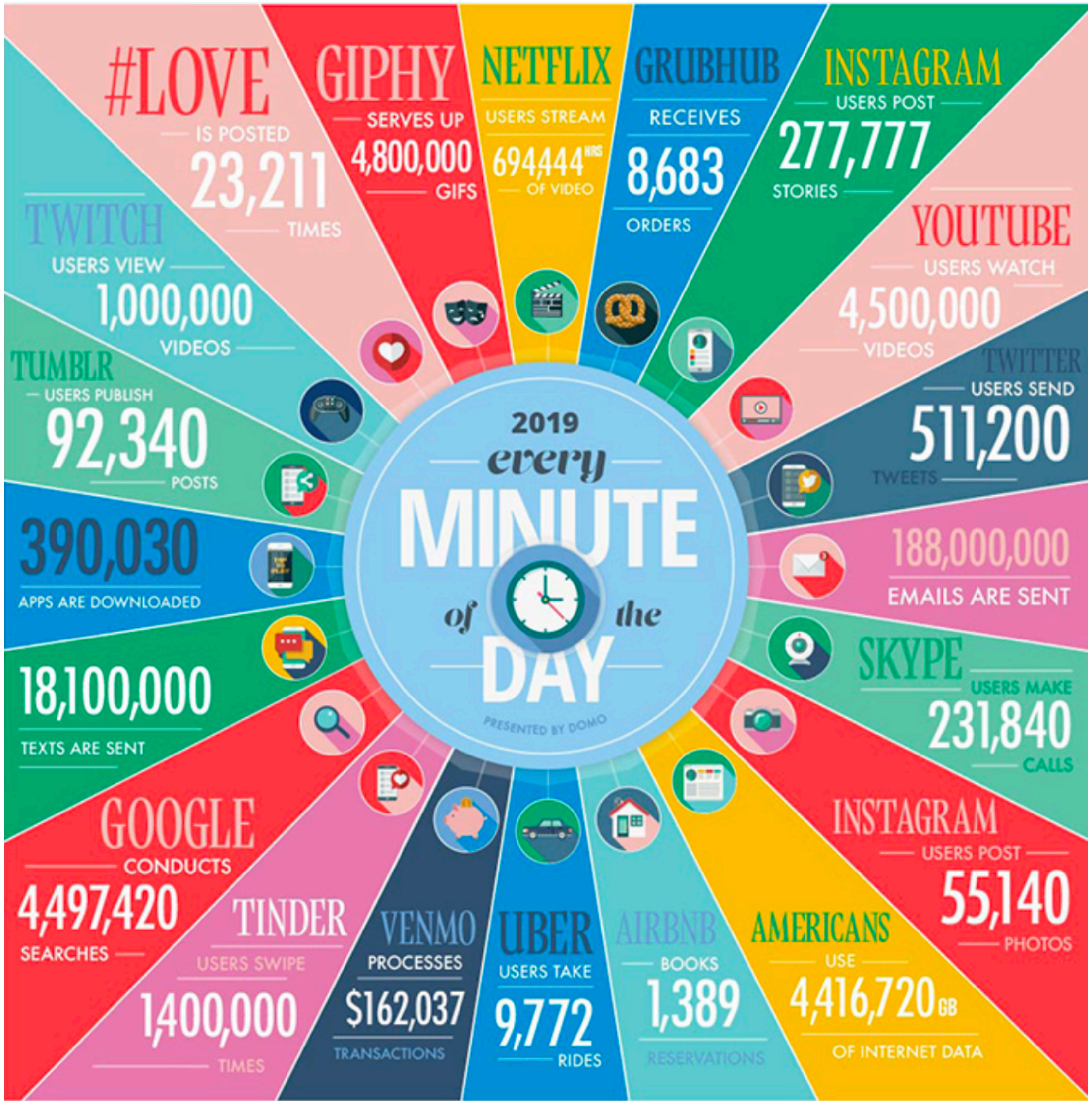
Centre for Visual Computing, UCSD

Today's talk

1. Domain Adaptation for Driving Scenes

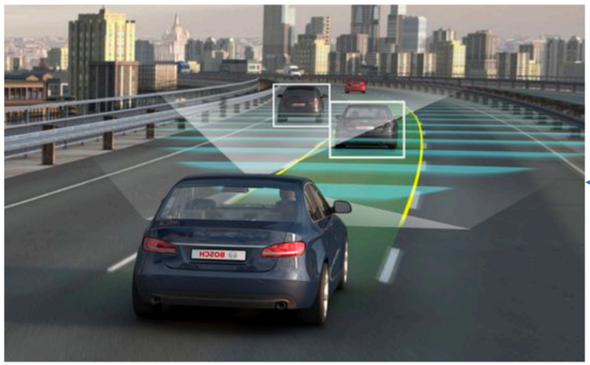
2. Universal Semantic Segmentation

2020: Data is driving research

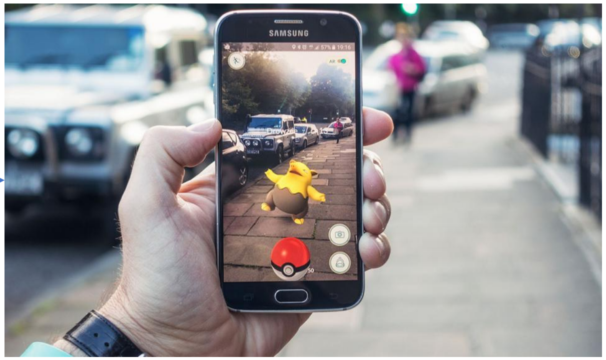


https://www.domo.com/blog/data-never-sleeps-4-0/

Computer vision

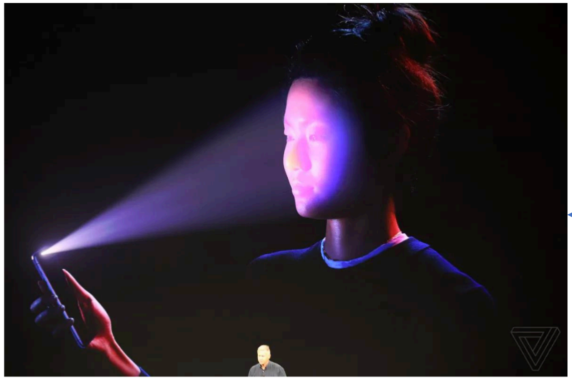


Autonomous Driving



Mobile AR / VR

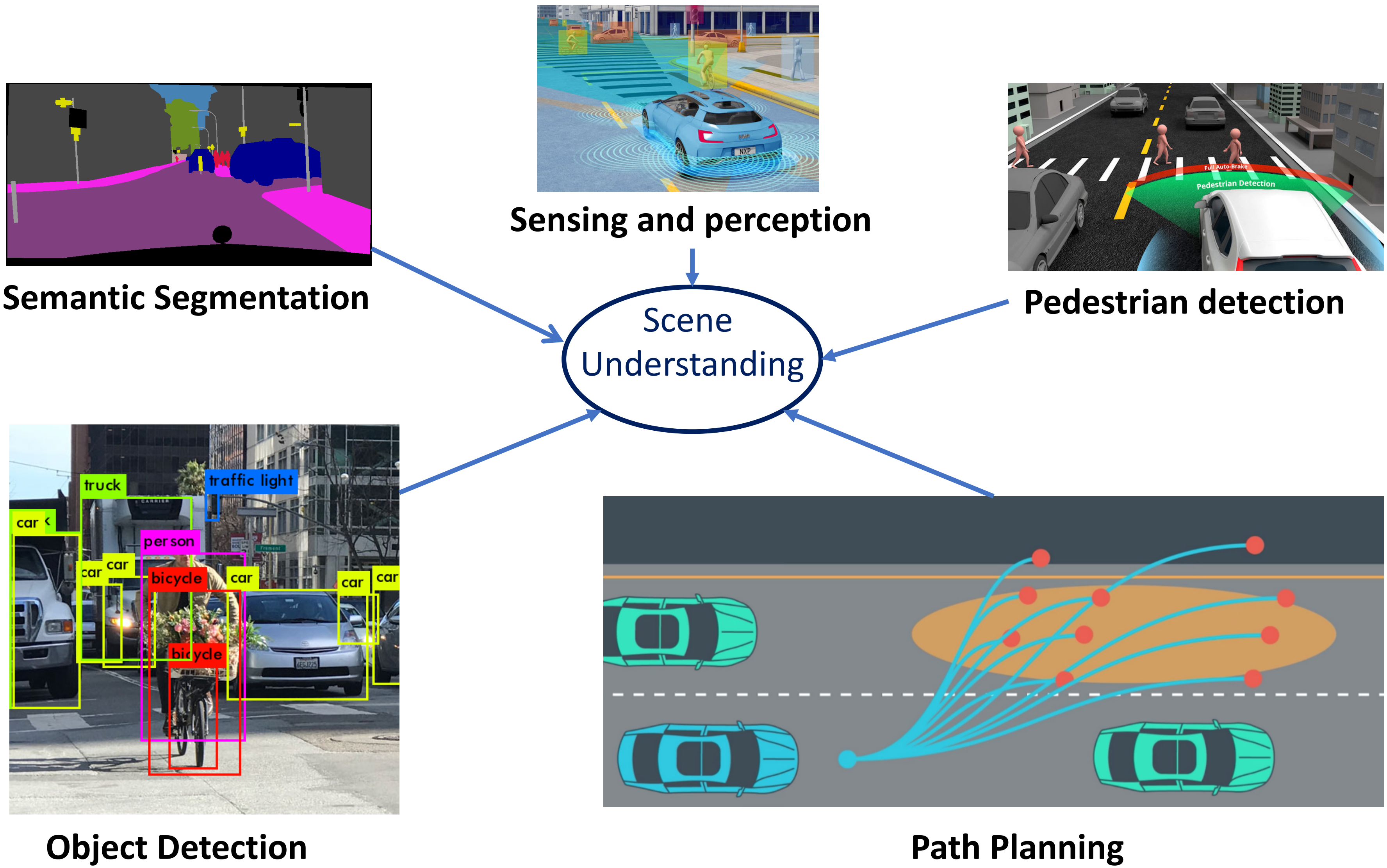
Computer Vision



Security / Surveillance

Human Assisting Robots

Holistic urban scene understanding



Transformation due to deep learning

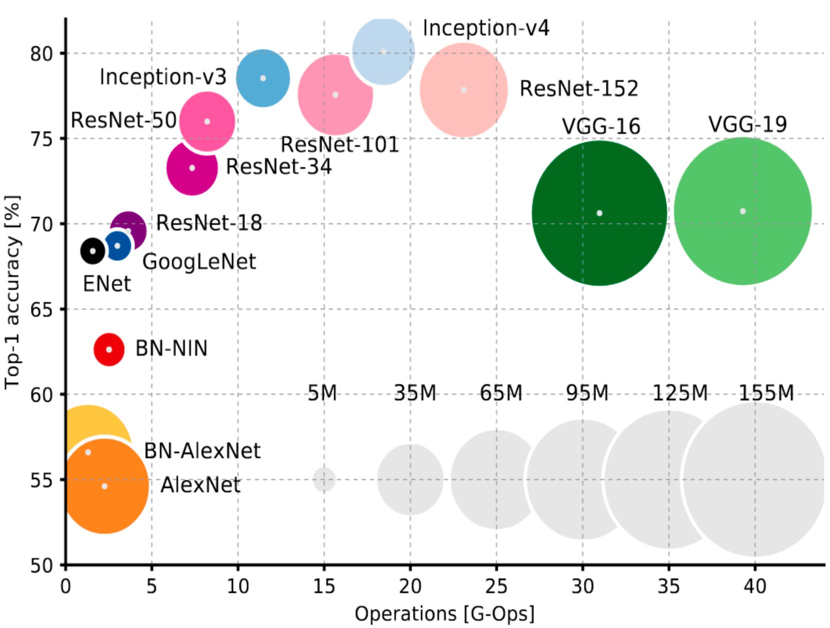
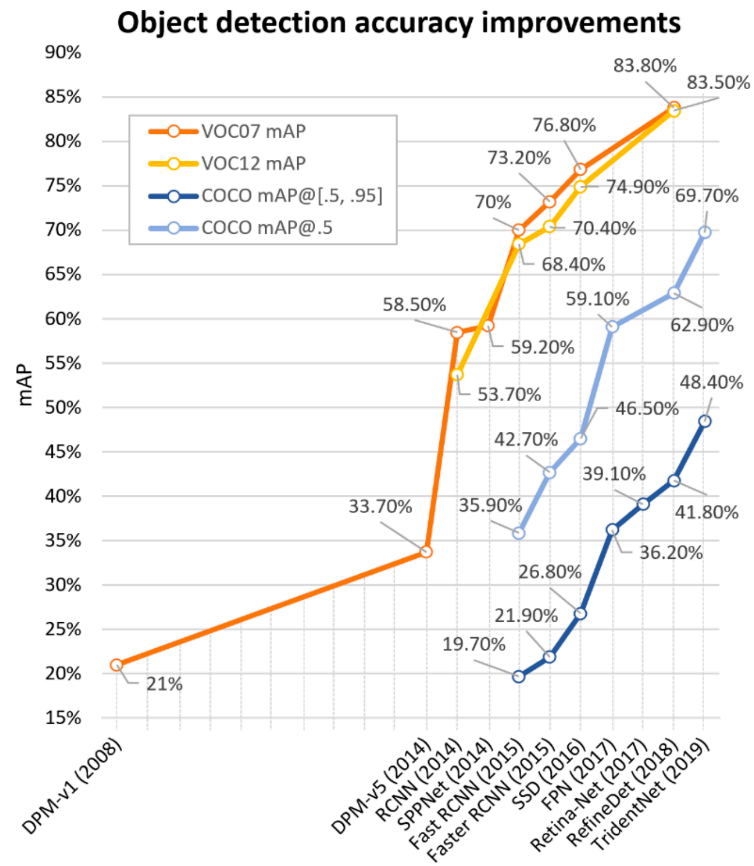


Image Classification on ImageNet



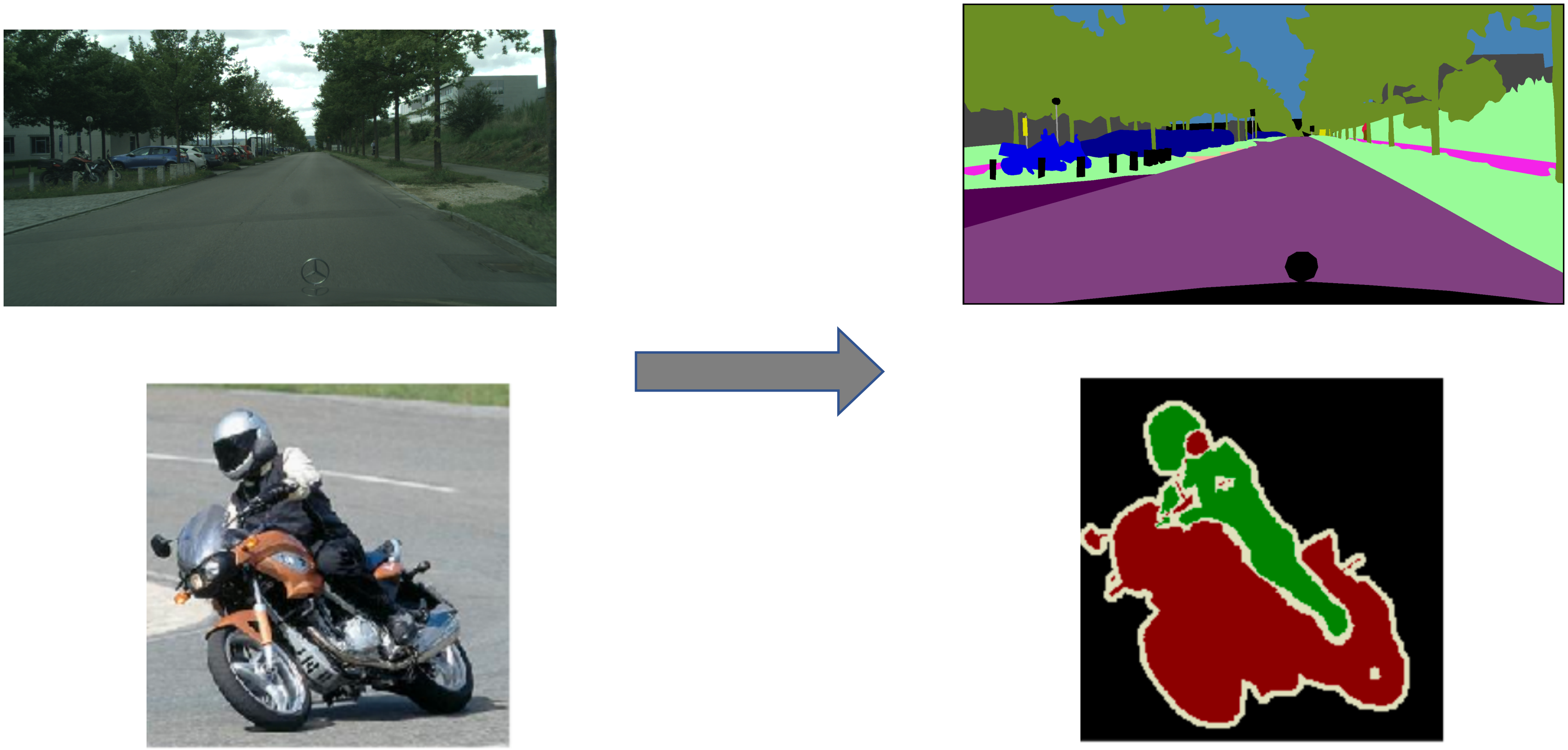
Object Detection

Object Detection in 20 Years: A Survey[2019]

Crucial factor: Large labeled data

Places: A 10 million Image Database for Scene Recognition [2017]

Semantic Segmentation



Data annotation: A serious challenge

- Car
- Road
- Sidewalk
- Person
- Sky
- Vegetation
- Street Sign
- Building

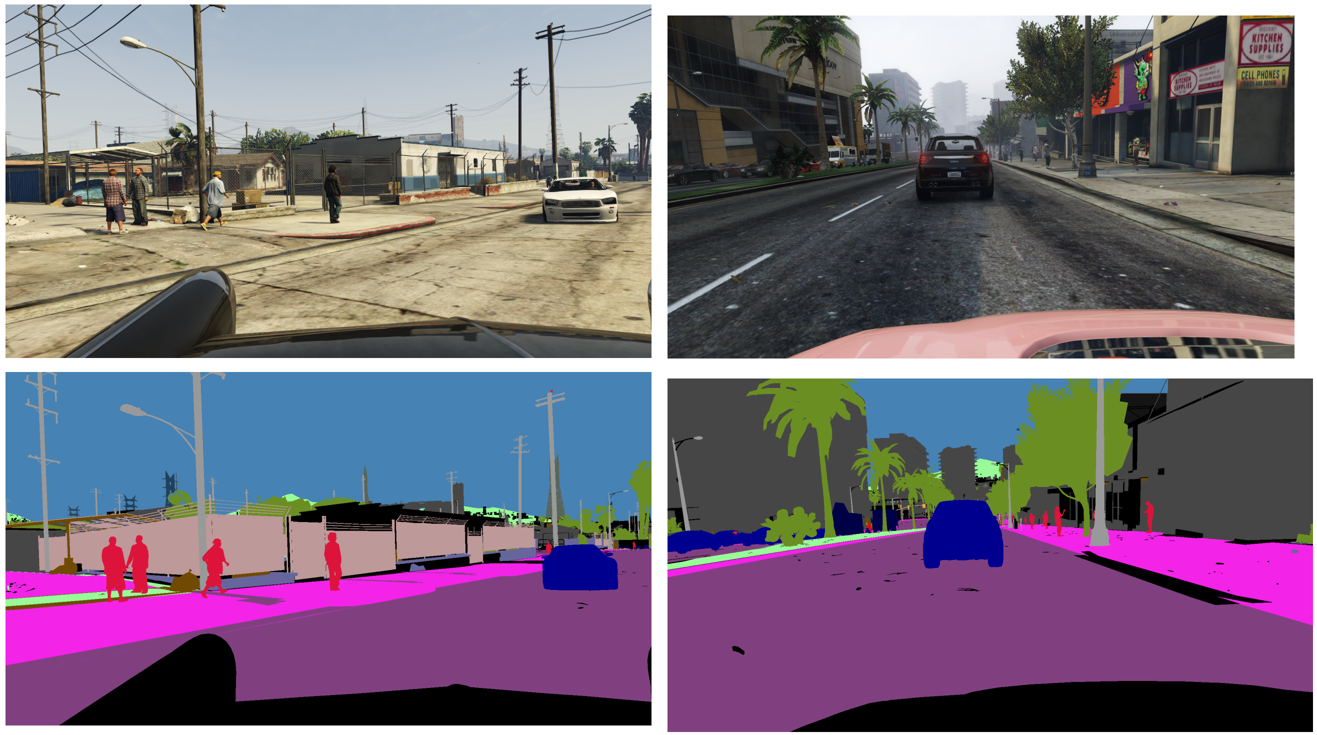
<u>EXPENSIVE</u>
10-12$ per image

<u>TIME CONSUMING</u>
90-96 min/Img

<u>NOISE</u>
Inter annotator agreement

Learning from a different source

Synthetic Images

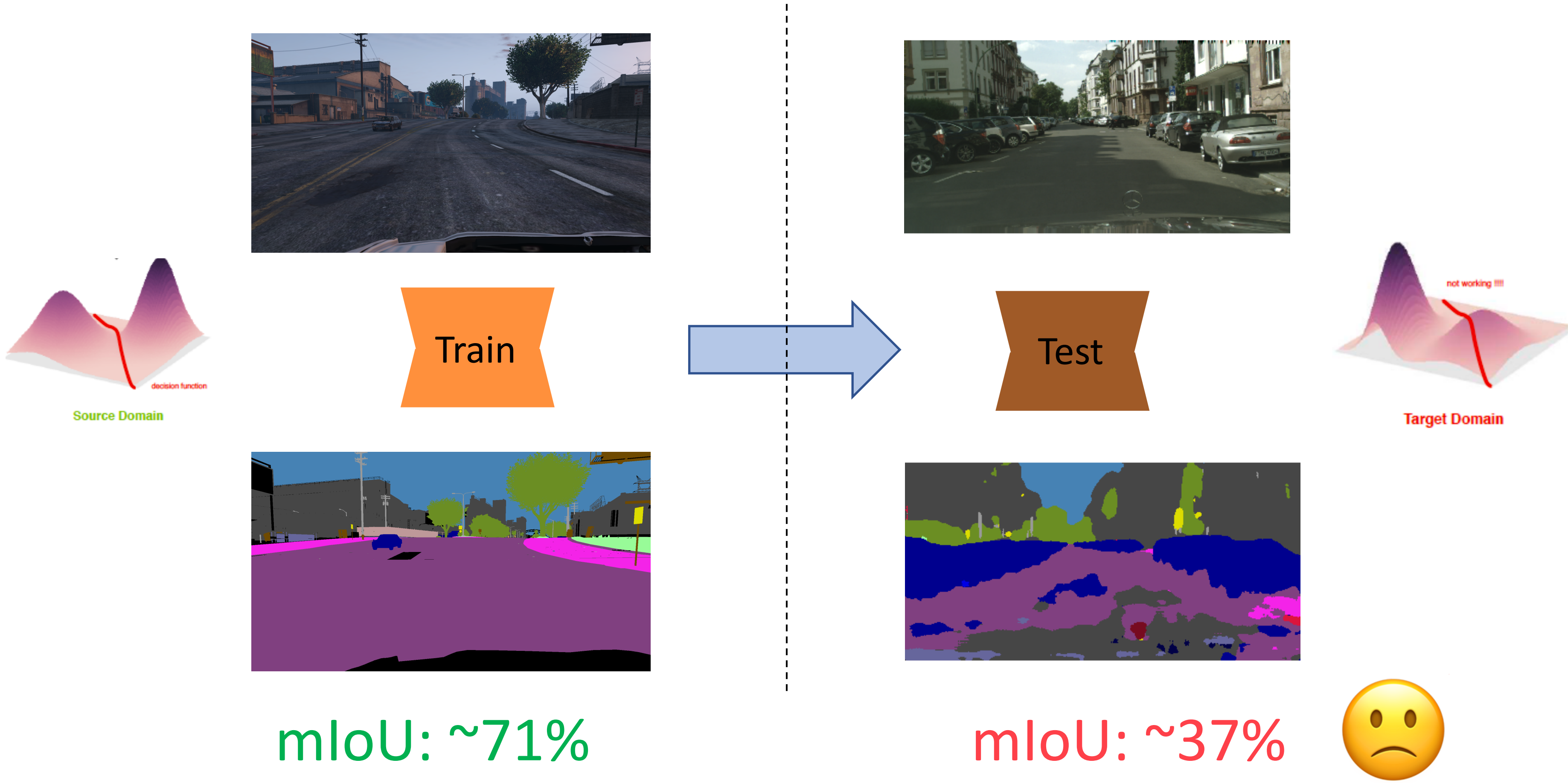


✓ **Easy to acquire** ✓ **Quicker** ✓ **Accurate**

GTA , Synthia

Domain discrepancy hurts transfer learning

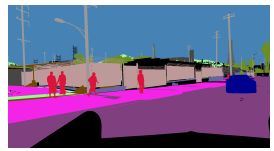
Train: Synthetic Domain | **Test: Real Domain**



Models trained on one source dataset **do not generalize** to other target images.

Use Unsupervised Domain Adaptation (UDA)

Source Domain



labels !!

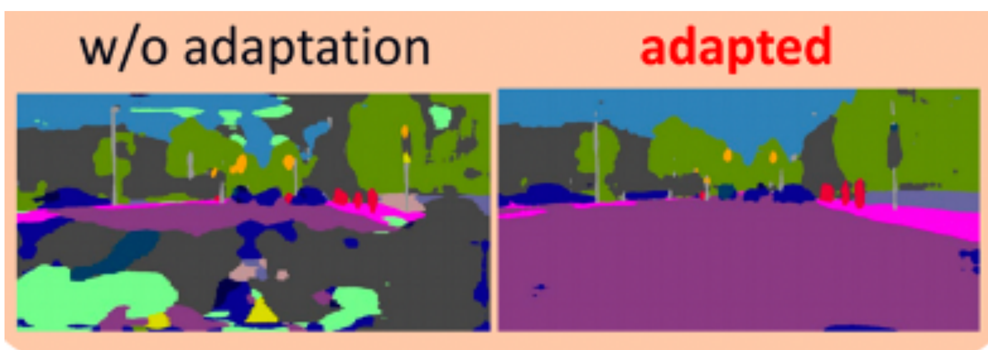
Target Domain

no labels

Adapt

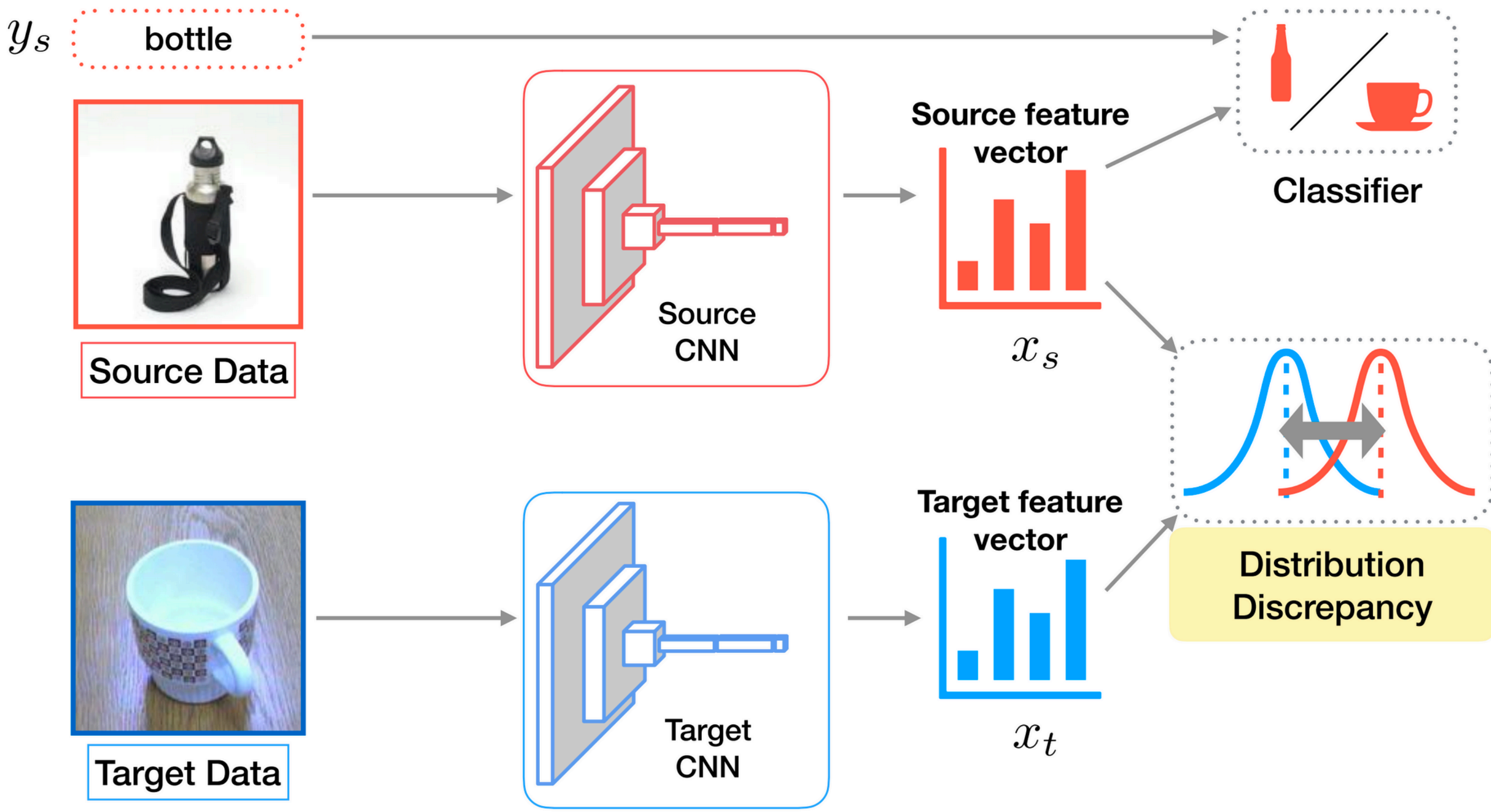
Target Segmentation

mIoU: ~37%



mIoU: ~47%

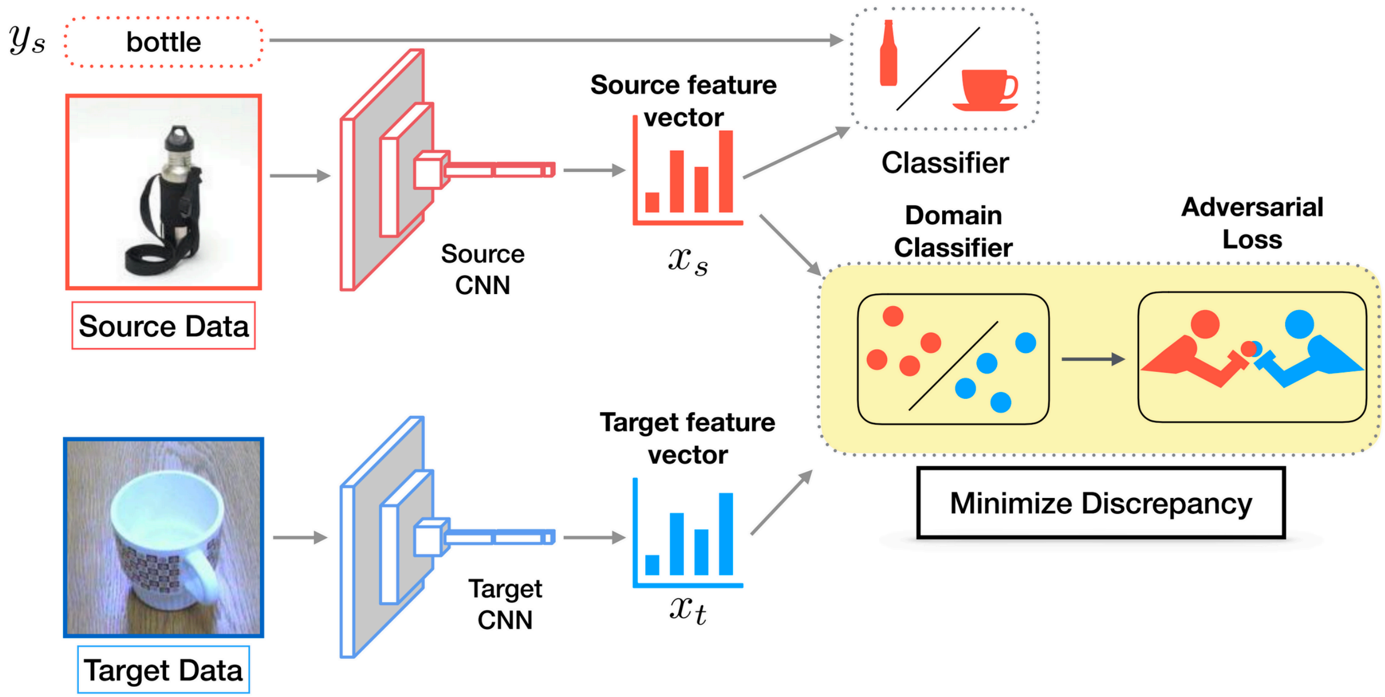
Deep Domain Adaptation



Learning with limited labels, Kate Saenko, ICCV 2019

Adversarial Domain Adaptation

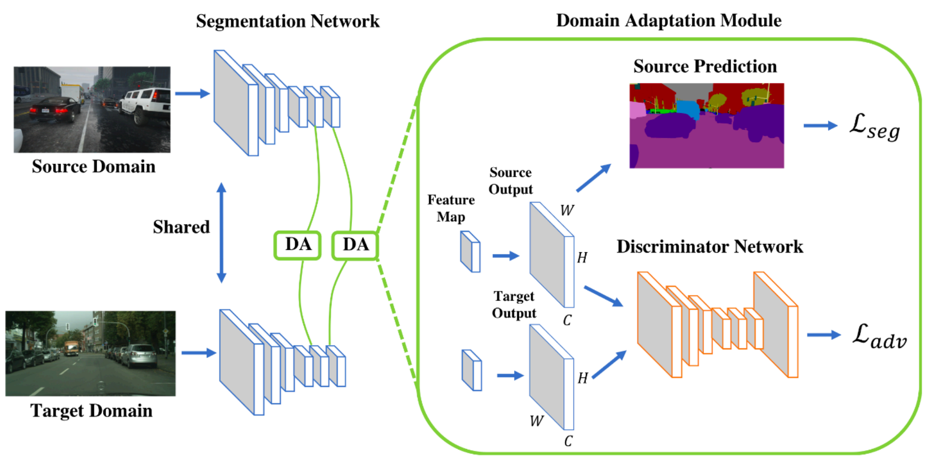
- Discriminator Based Learning
  - Optimize simultaneously for
    - Feature extractor θ_f, Classifier θ_C, Domain Discriminator θ_D.
  - Learn **domain agnostic features** through **adversarial training.**



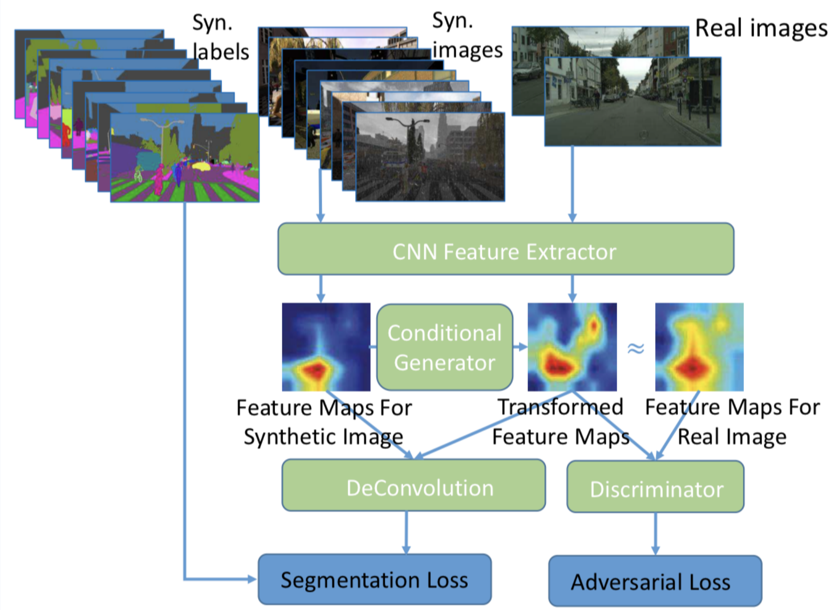
Learning with limited labels, Kate Saenko, ICCV 2019

Adversarial training for semantic segmentation

Output space adaptation



Input space adaptation



[Tsai'18 , Hong'18]

Universal Learning: Adaptation across very distinct datasets

Day Scenes

Unconstrained Scenes

Rainy scenes

Night scenes

A tale of two cities: Cityscapes vs. IDD

Cityscapes: German Roads

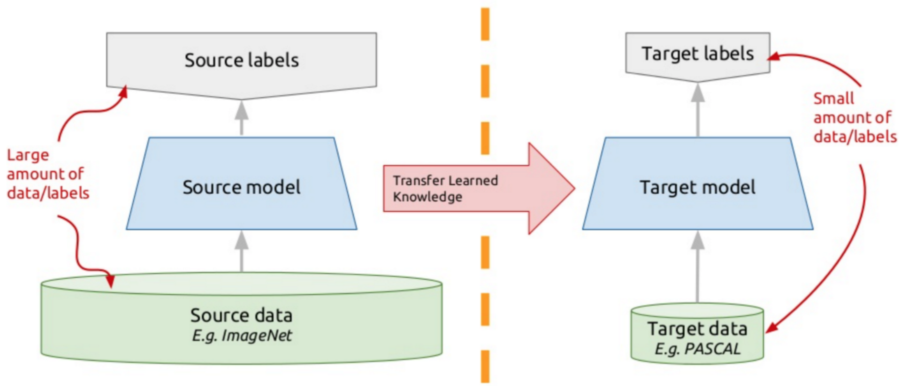
IDD: Indian Roads

- Unconstrained
- Weather
- New labels: auto-rickshaw, local traffic signs
- Semantic Shift: Sidewalk (US) vs. Footpath (UK) vs. Pavement (Ind)

Varma, Girish, et al. "IDD: A dataset for exploring problems of autonomous navigation in unconstrained environments, WACV 2019.

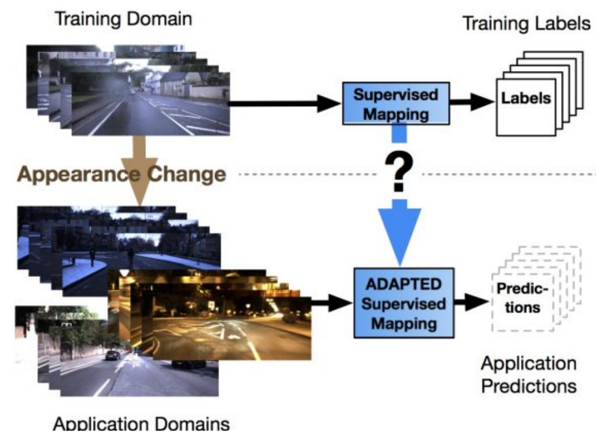
Existing alternatives not suitable for universal learning

Transfer Learning



- Low accuracy on source
  [Catastrophic Forgetting]
- Does not use unlabeled data
- Requires large amount target data

Domain Adaptation

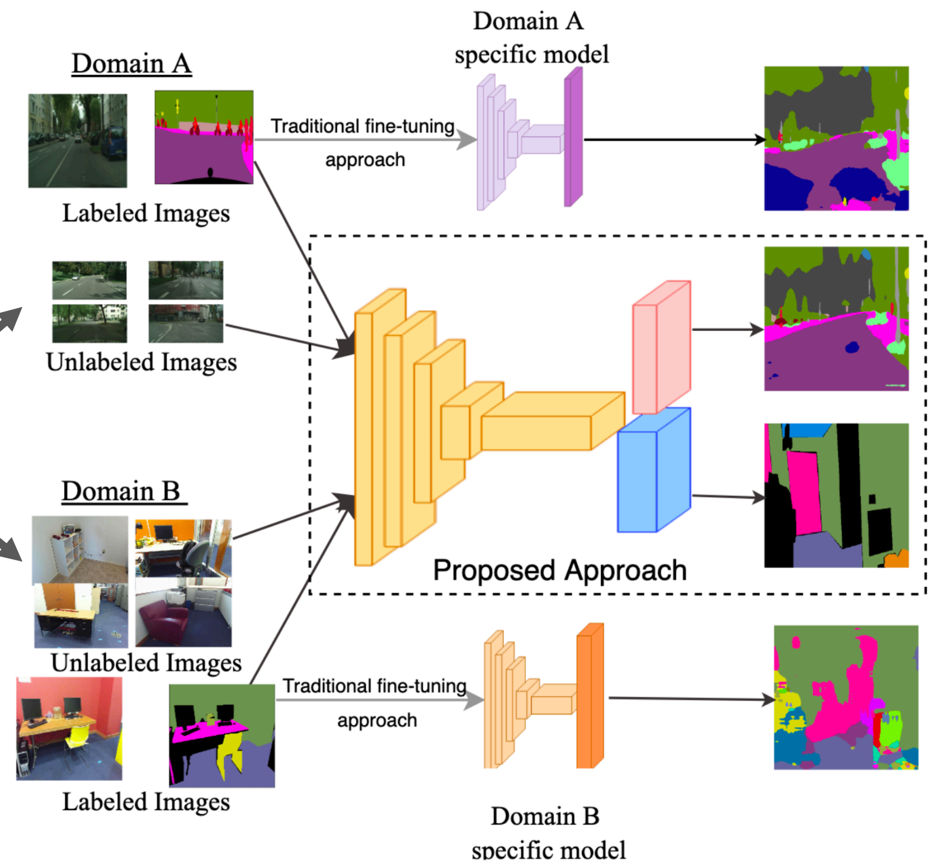


- Needs same categories across source and target
- Requires large labeled data in source domain.

Markus,. "Addressing appearance change in outdoor robotics.", IROS 2017

Use Universal Semi-Supervised Segmentation

Few labeled data + lots of unlabeled data!



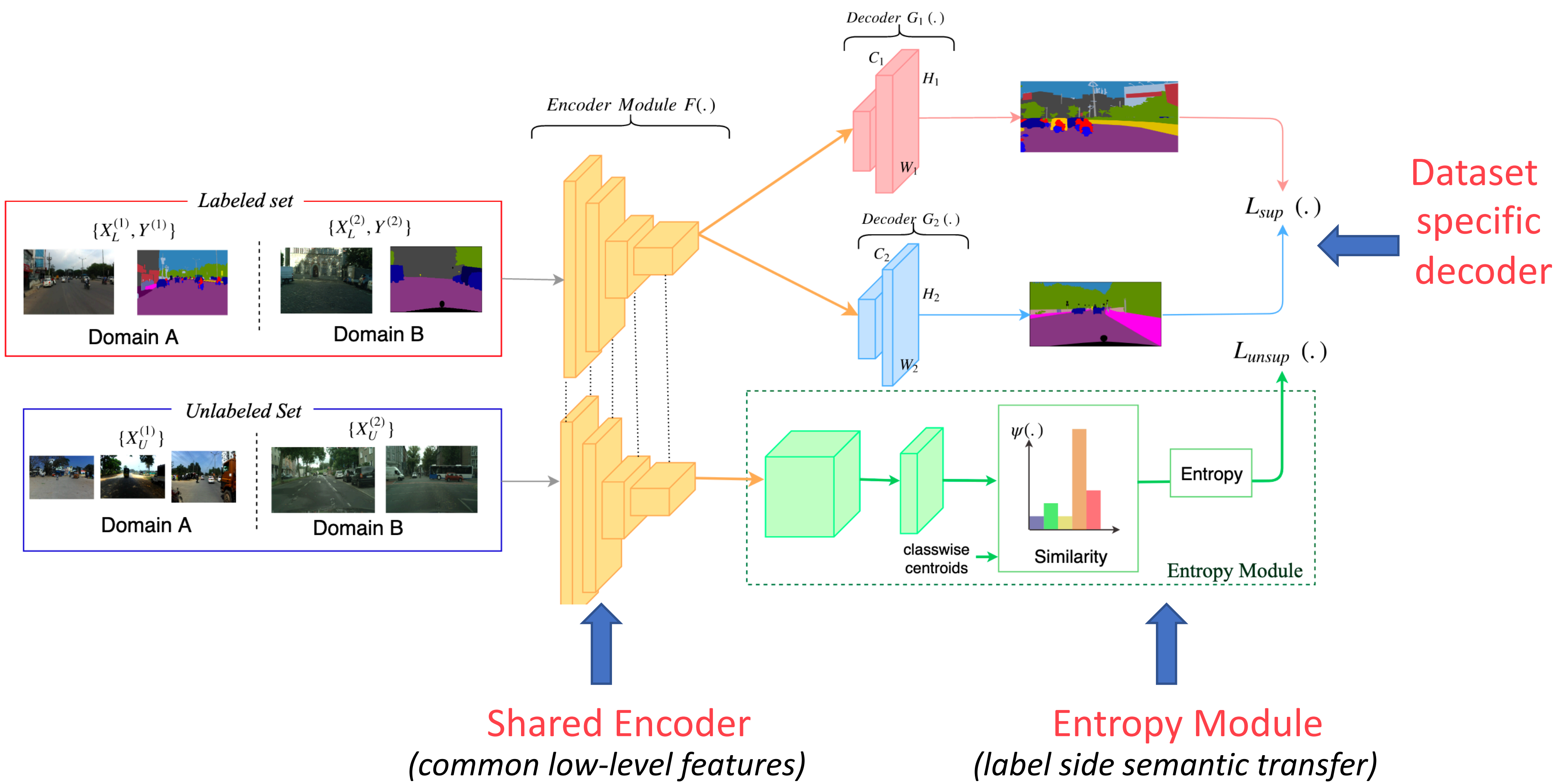
knowledge transfer → better segmentation

Kalluri,. "Universal Semi-supervised Segmentation", ICCV 2019

Address semantic shift using unlabeled data

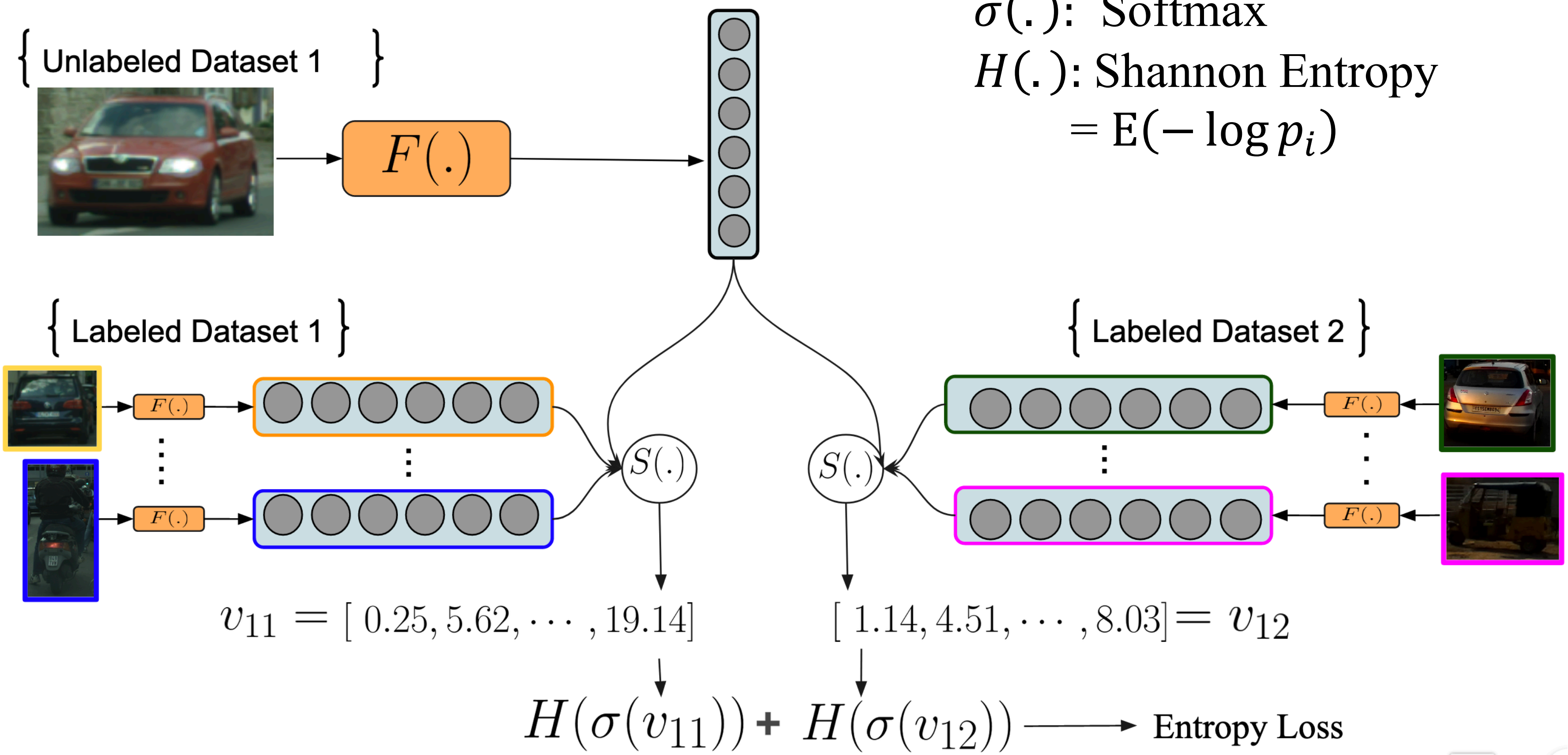
	Source Unlabeled Data	Target Unlabeled Data	Joint Model	Mixed Labels Support
Fine Tuning	✗	✗	✗	✓
Semi-supervised [Hung 2018]	✓	✗	✗	NA
CyCADA [Hoffman 2018]	✗	✓	✓	✗
Joint Training	✗	✗	✓	✓
Our Approach	✓	✓	✓	✓

CNN architecture for universal training



Using entropy regularization for feature alignment

$S(.)$: Dot product similarity
$\sigma(.)$: Softmax
$H(.)$: Shannon Entropy
$= \mathrm{E}(-\log p_i)$



Datasets for evaluating universal models

Cityscapes

+

CamVid

IDD (Indian Roads)

SUN RGB-D (Indoor Scenes)

Easy → Hard

Universal models performs better than individual models

Take N=100 labeled examples from each dataset

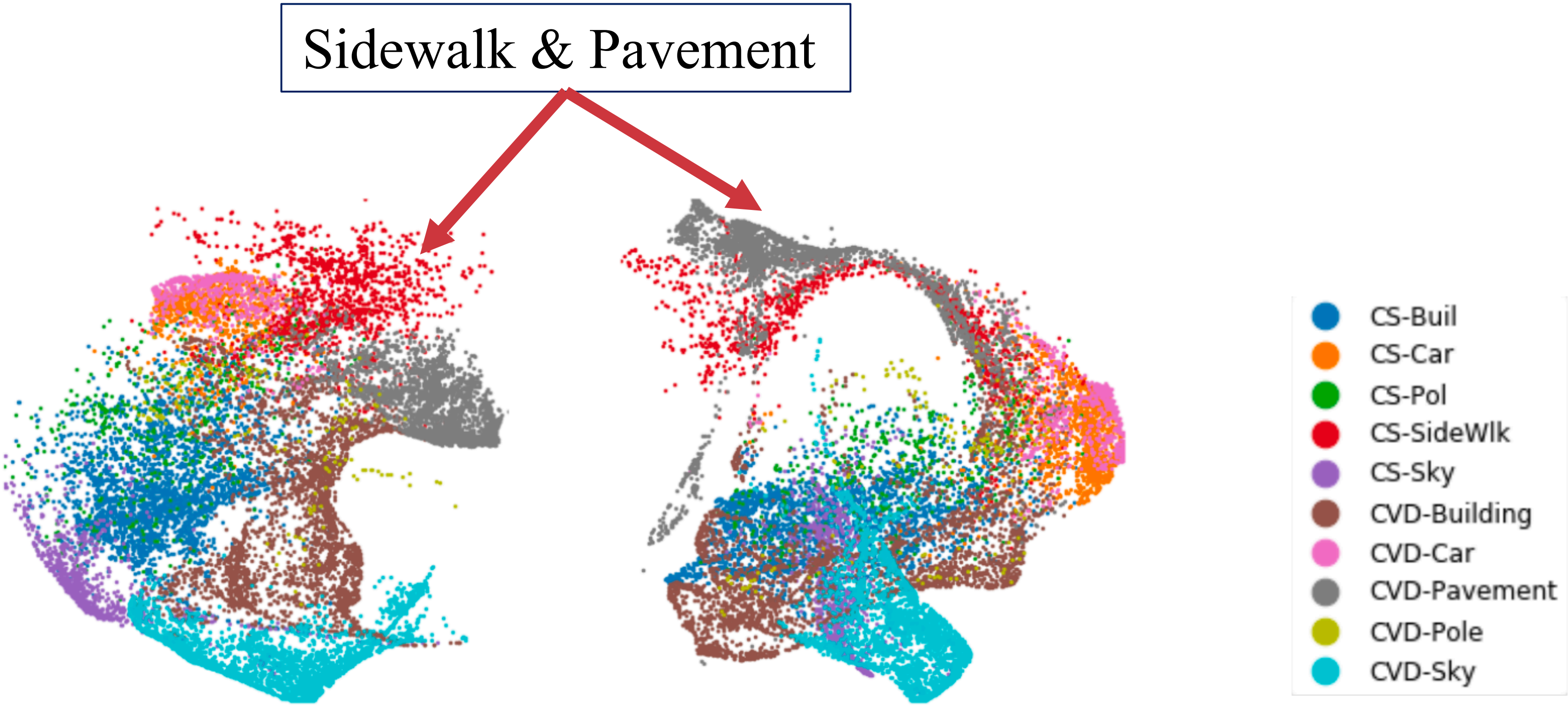
Method	Evaluate on CS	Evaluate on CVD	Average
Train on CS	40.9	36.5 (↓ 14%)	38.7
Train on CVD	22.2 (↓ 18%)	50.1	36.1
Ours Universal	41.1 (↑ 0.2%)	54.6 (↑ 4%)	**47.8**

Universal model on Cityscapes + CamVid

Method	Evaluate on CS	Evaluate on IDD	Average
Train on CS	64.2	32.5(↓ 18%)	48.4
Train on IDD	46.3(↓ 18%)	55.0	50.7
Ours Universal	64.1 (↓ 1%)	55.1(↑ 5%)	**59.6**

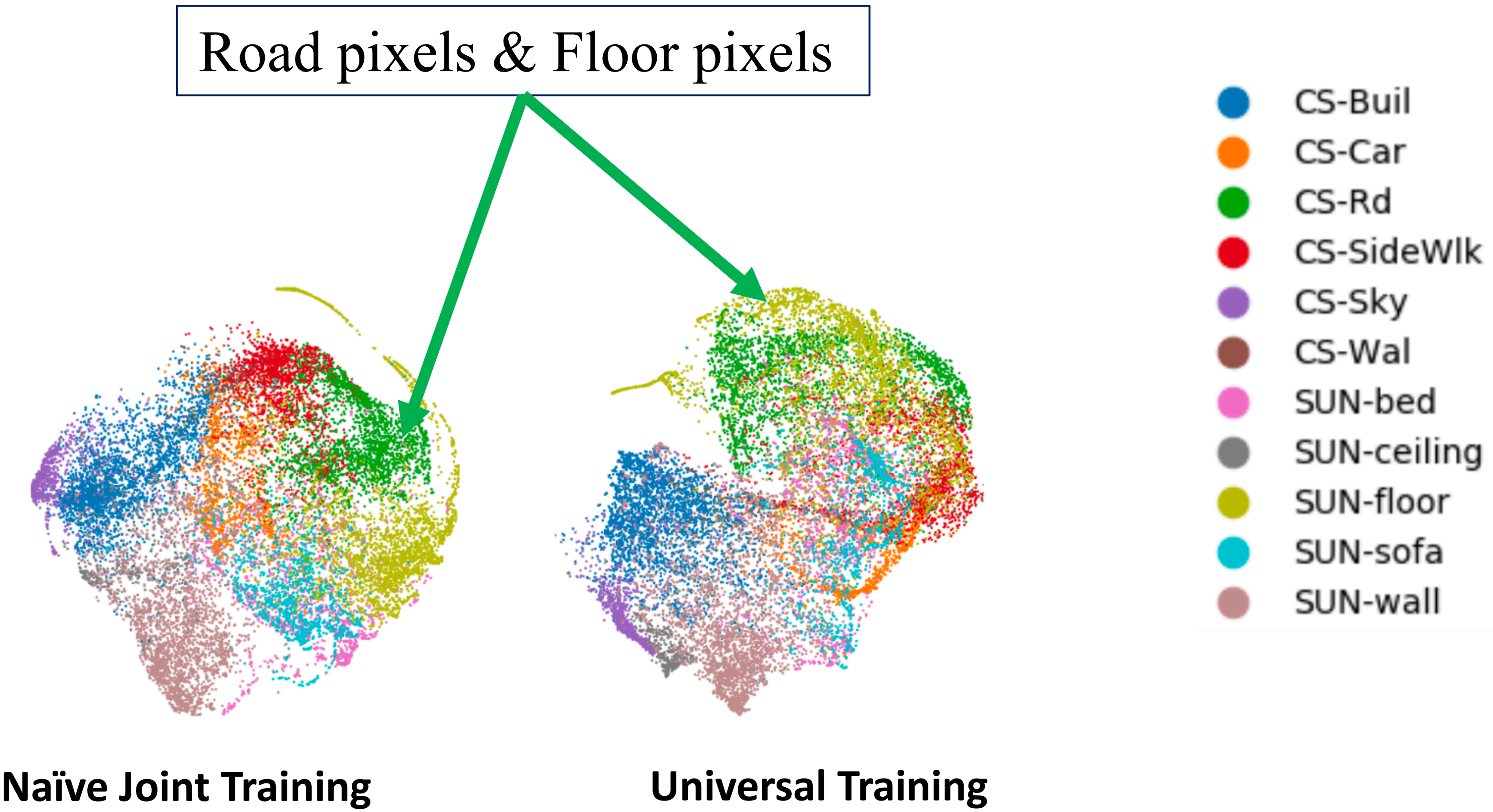
Universal model on Cityscapes + IDD

Semantic Alignment for Cityscapes vs CamVid

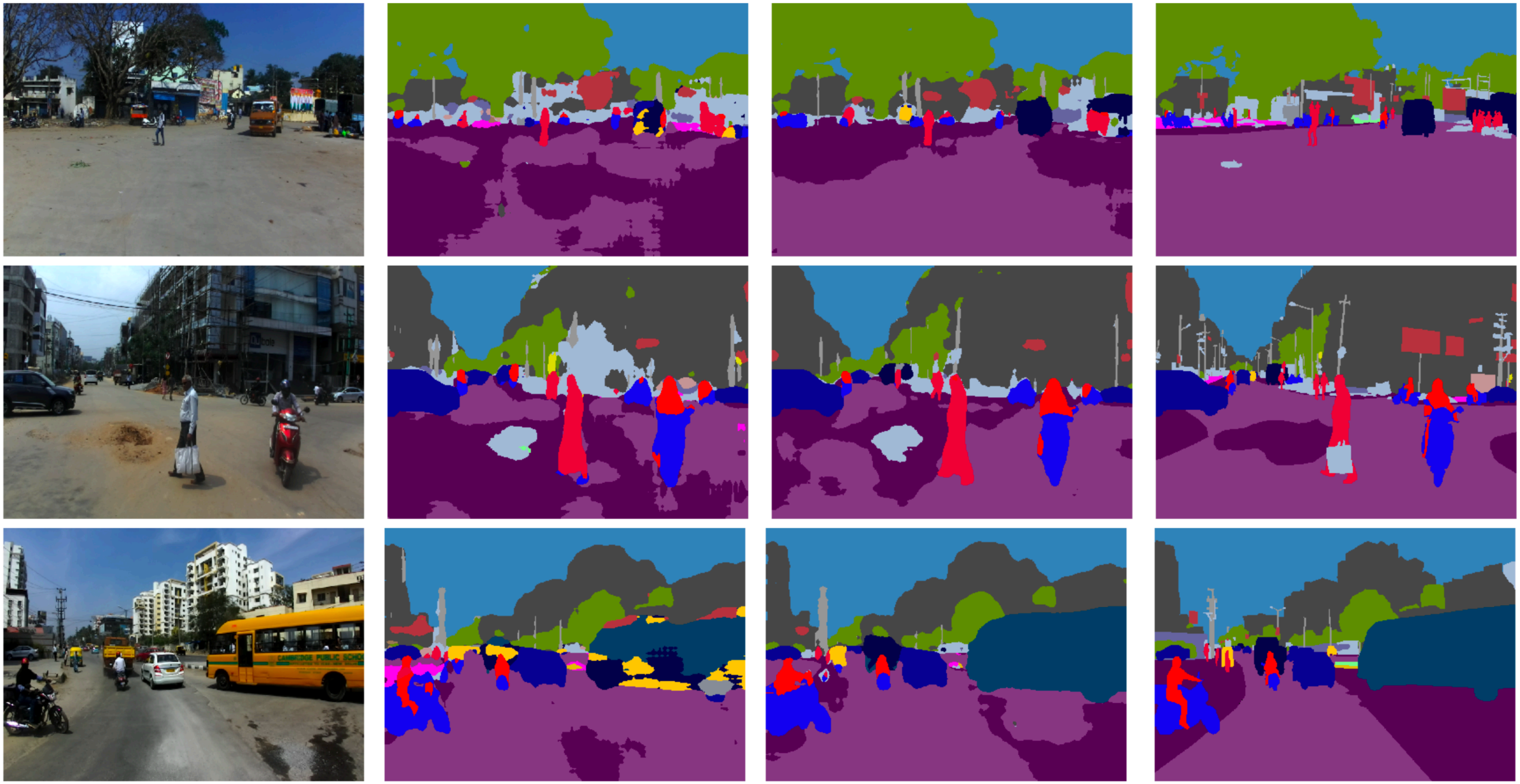


Naïve Joint Training **Universal Training**

Geometrical Alignment for Cityscapes vs SunRGB



Refined segmentation outputs on IDD using universal training



(a) Original Image (b) Without Entropy Module (c) With Entropy Module (d) Ground Truth Segmentation

Summary

1. Domain adaptation needed to overcome annotation overhead.

2. Universal Segmentation enables effective knowledge transfer and improves performance.