IDD

A Dataset for Exploring Problems of Autonomous Navigation in Unstructured Environments

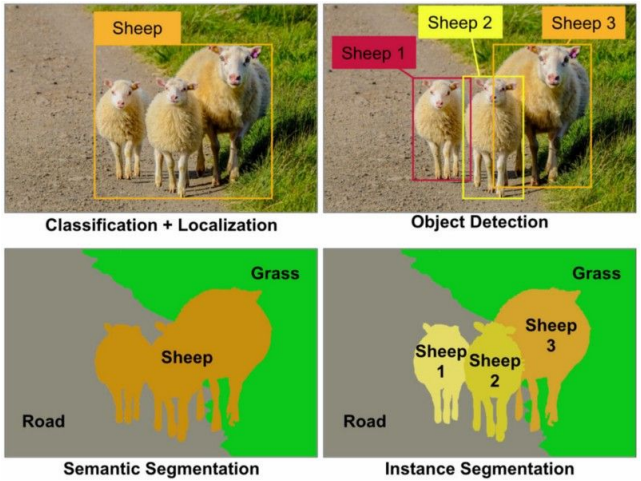
http://idd.insaan.iiit.ac.in/

Girish Varma
(IIIT Hyderabad)


WHAT?

Auto. Nav. Dataset

- Images from road scenes
- Pixel level/Bounding box Annotations
- Semantic/instance segmentation, Detection
- A basic primitive for Auto. Nav.
- Other Metadata:
  - LIDAR scans
  - GPS/IMU data



Existing Datasets

- Camvid (2008)
  - ETHZ, Cambridge
- KITTI (2012-14)
  - Comprehensive set of data
  - MPI, TTI, KIT
- Cityscapes (2015-16)
  - Daimler, MPI, TU Darmstad
- Mapillary (2017-18)
  - User uploaded image from around the world
- BDD100K (2018)
  - Berkeley, Dashboard Cam

Other Segmentation Datasets

- MS COCO
- ADE20K
- SUN


Why

Some examples from Cityscapes

Diversity & Unstructured Conditions

Odd-shaped vehicles, challenging drivable areas

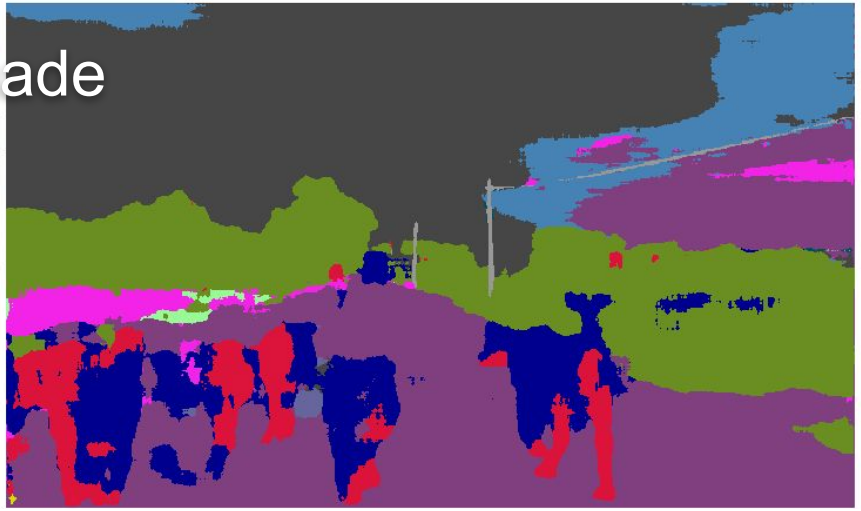


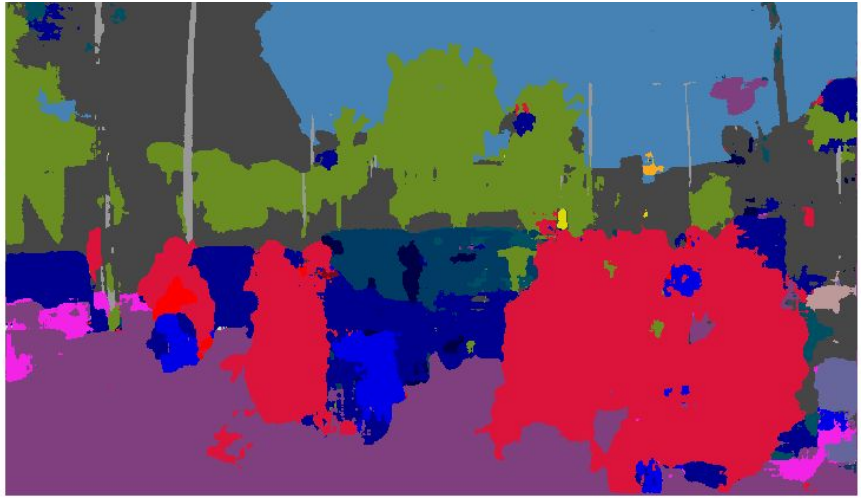
Pedestrians & jay-walkers



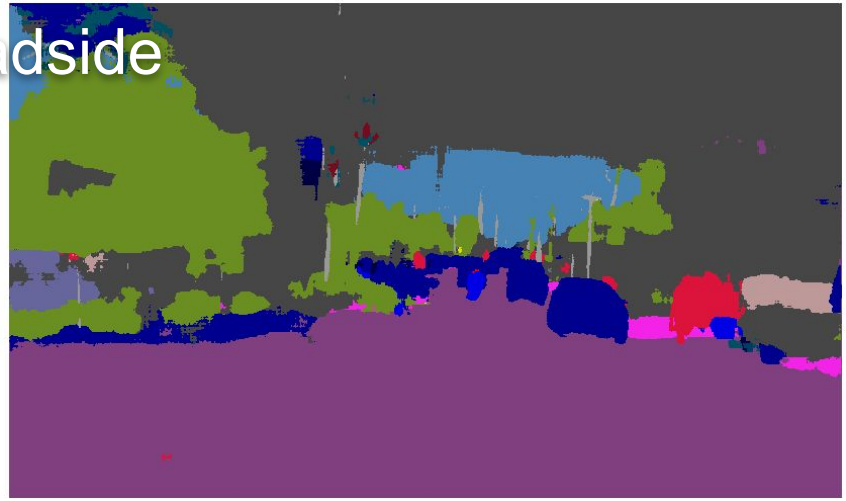


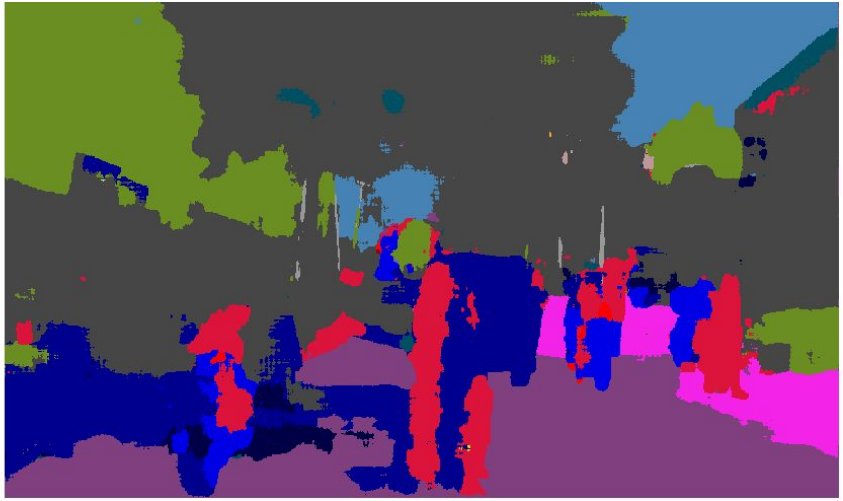

Shade



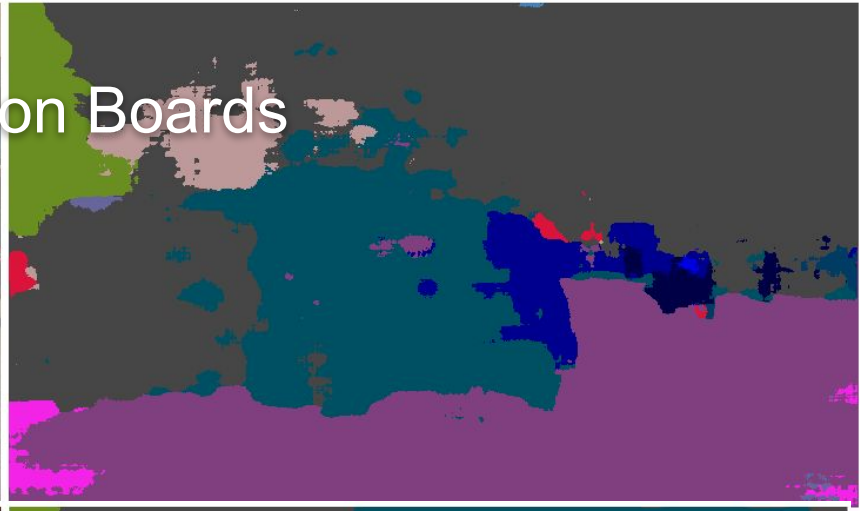


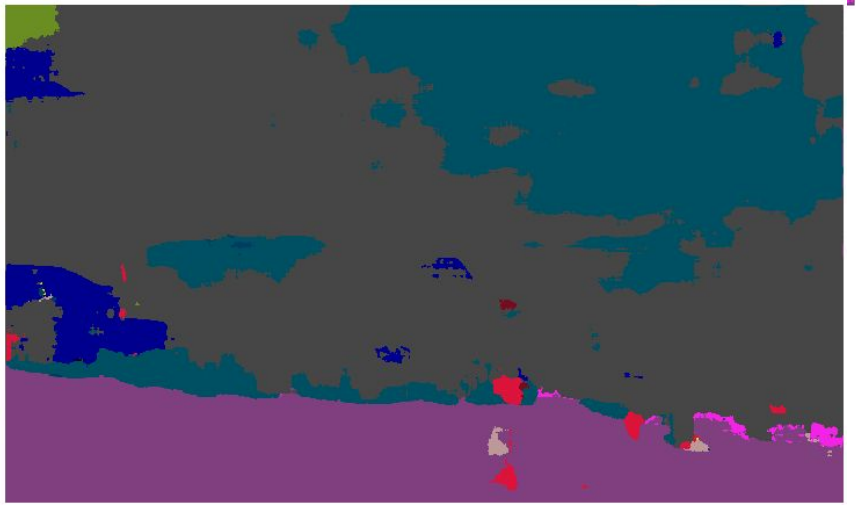
Roadside



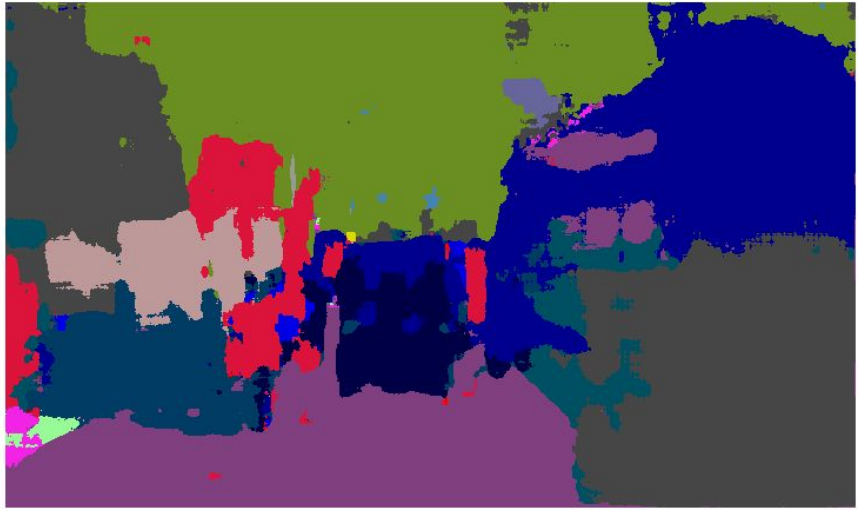


Information Boards





Unstructured Traffic



IDD

IDD: Vital Stats

- 10,004/50,000 images from 182 drive sequences annotated With pixel level/bounding box labels.
- From Bangalore and Hyderabad

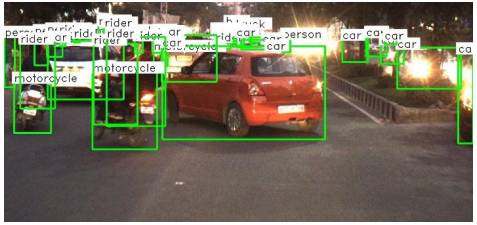
Dataset	Calibration	Nearby frames / Video	Distortion /Night	#Images/ #Sequences	#Labels Train/Total	Average Resolution
Cityscapes [5]	✓	✓		5K / 50	19/34	2048x1024
IDD	✓	✓		10K / 180	30/34	1678x968
BDD100K [26]		✓	✓	10K / 10K	19/30	1280x720
MVD [16]				25K / -	65/66	>1920x1080

Table 1. Comparison of semantic segmentation datasets for autonomous navigation.

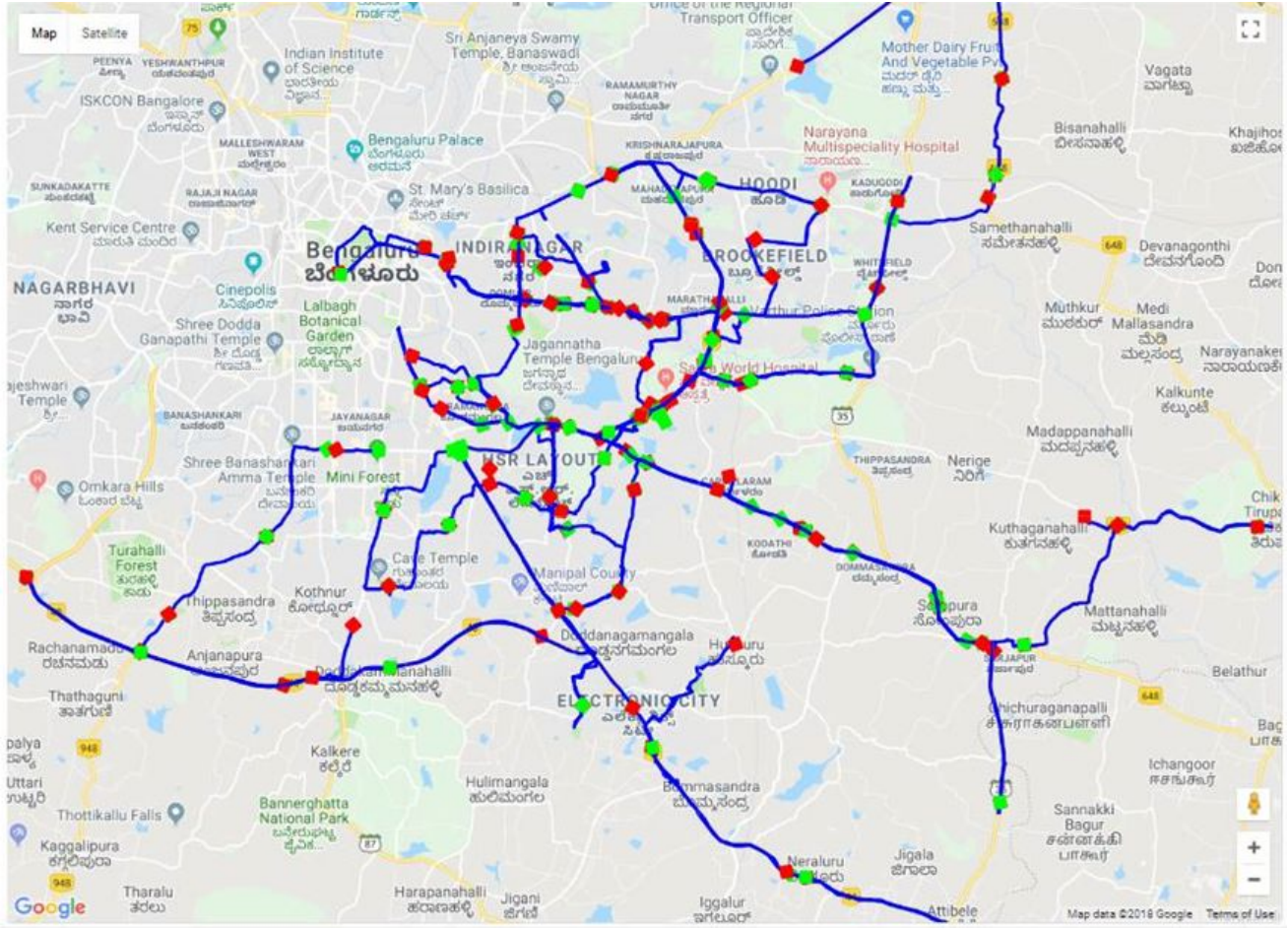
Unstructured Driving Conditions – Data Collection

Input

Coarse

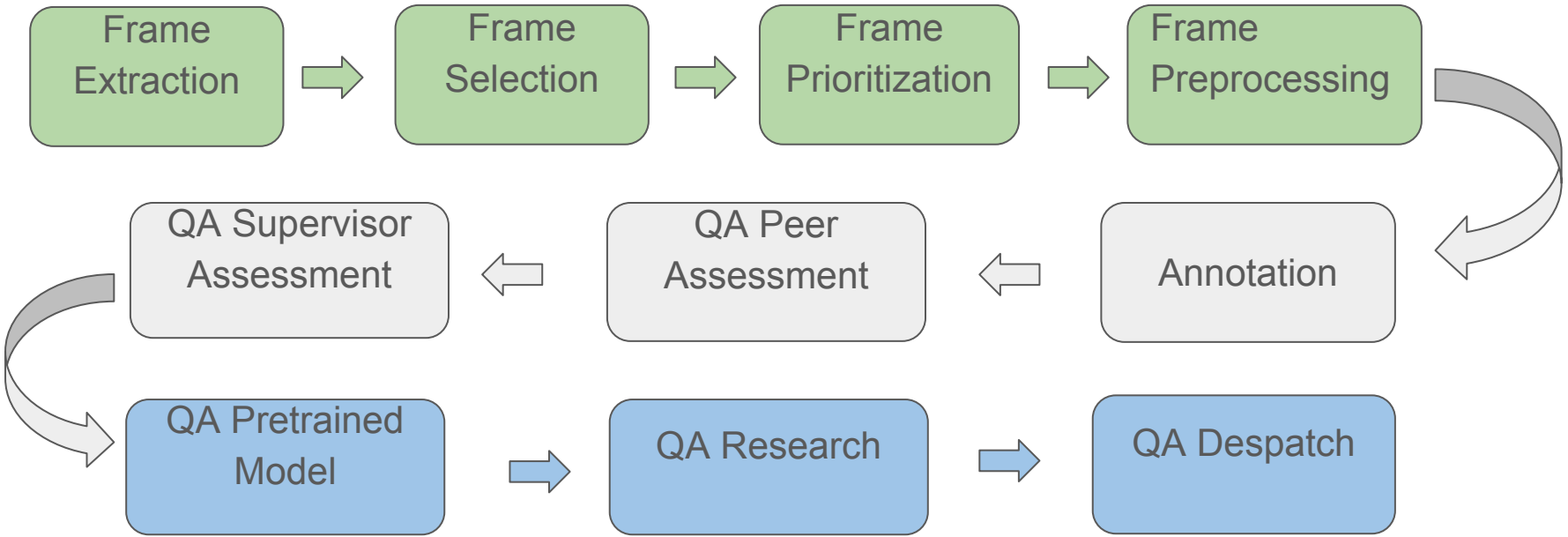


Fine

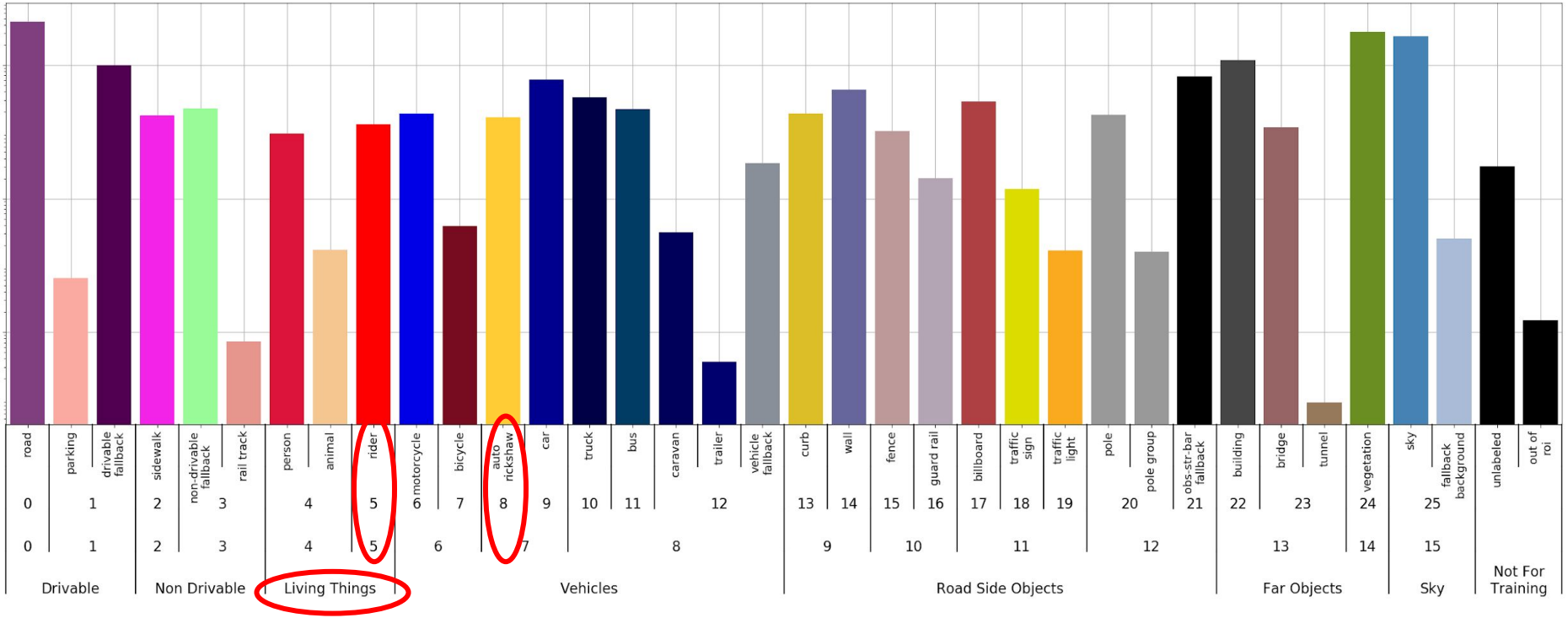


Annotation Pipeline

Annotators trained on Cityscapes.



AutoNUE Data Diversity – New Objects on Roads



http://cvit.iiit.ac.in/scene-understanding-challenge-2018/overview.php

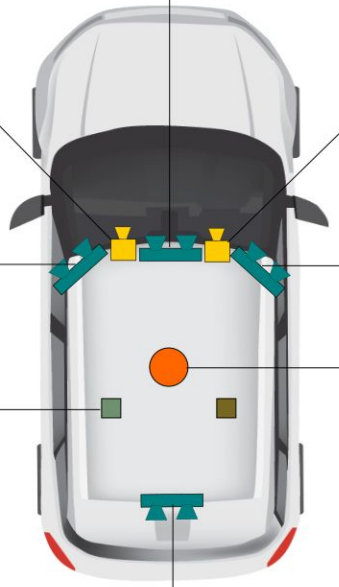
LEFT Front-Far

FRONT Near

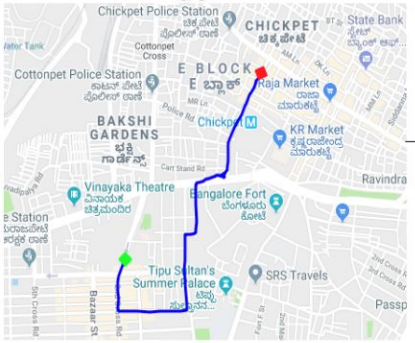
RIGHT Front-Far

LEFT Near

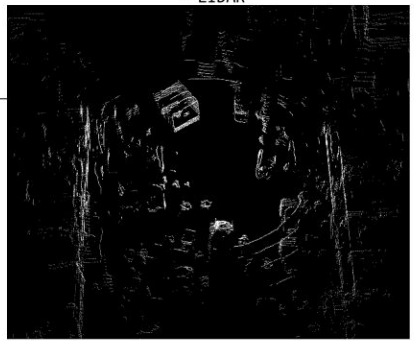
RIGHT Near



GPS



LIDAR



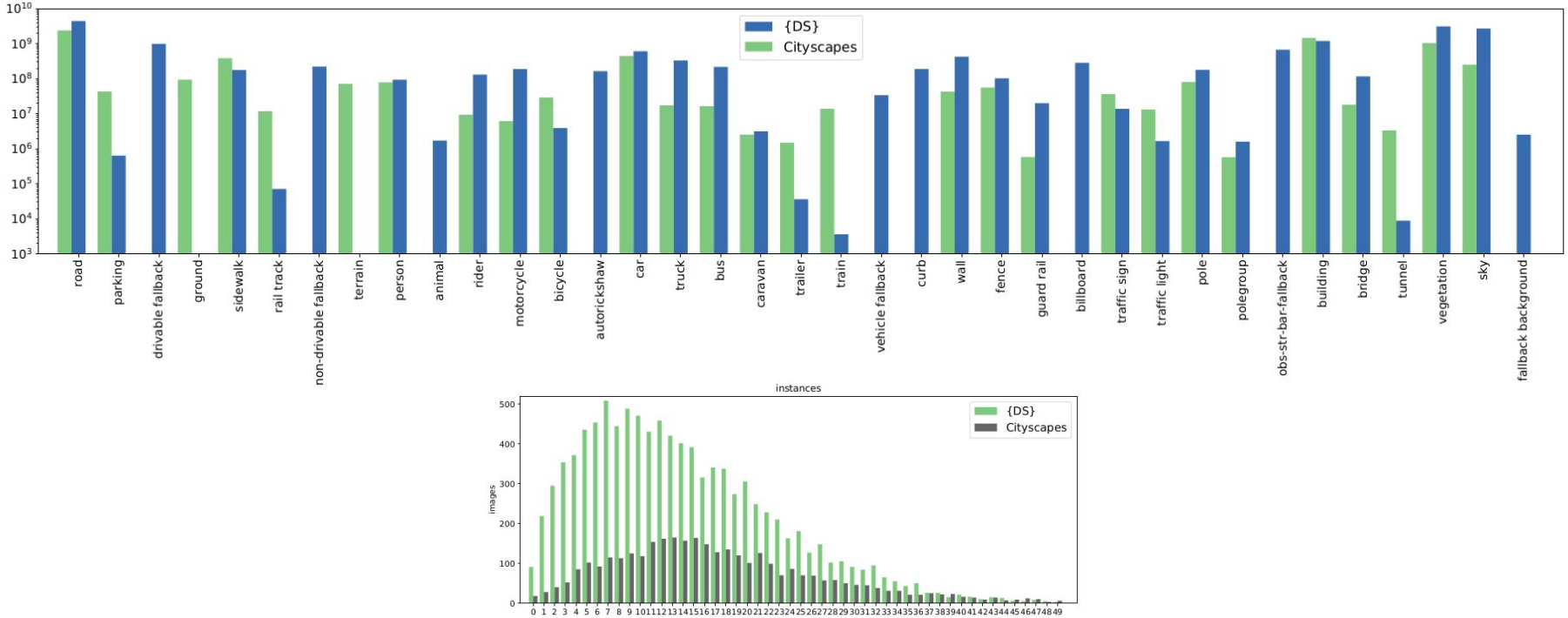
REAR Near

Timestamp = 2018-05-08 09:46:59

Label Hierarchy

- 4 Level label Hierarchy
  - L4: 30 labels
  - L3: 26 labels
  - L2: 16 labels
  - L1: 7 labels
- Splits
  - 7000 Train
  - 2000 Test
  - 1000 Validation
  - Procedure
    - Choose randomly subset of 182 sequences
    - Check if for each label pixels are split into 70%,20%,10%
    - If not repeat.

Comparison with Cityscapes



gure 8. Comparison of traffic participants in our dataset wi
tyscapes.

Trained Model

IoUs: Intersection over Union

Method	% mIoU at Levels		
	L1	L2	L3
ERFNet	-	-	55.4
DRN-D-38	85.9	72.6	66.6
*DeeplabV3+ [4]	89.8	78.0	74.0
*PSPNet [27]	89.9	78.0	74.1
*Wider Resnet-38, DeeplabV3 Decoder, Inplace ABN [20], Ensemble of 4	89.7	77.9	74.3

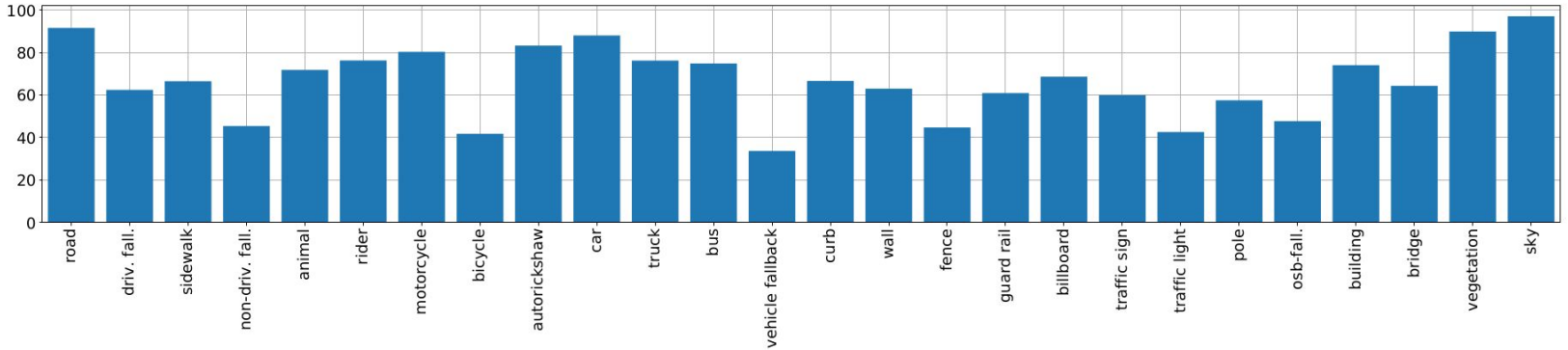


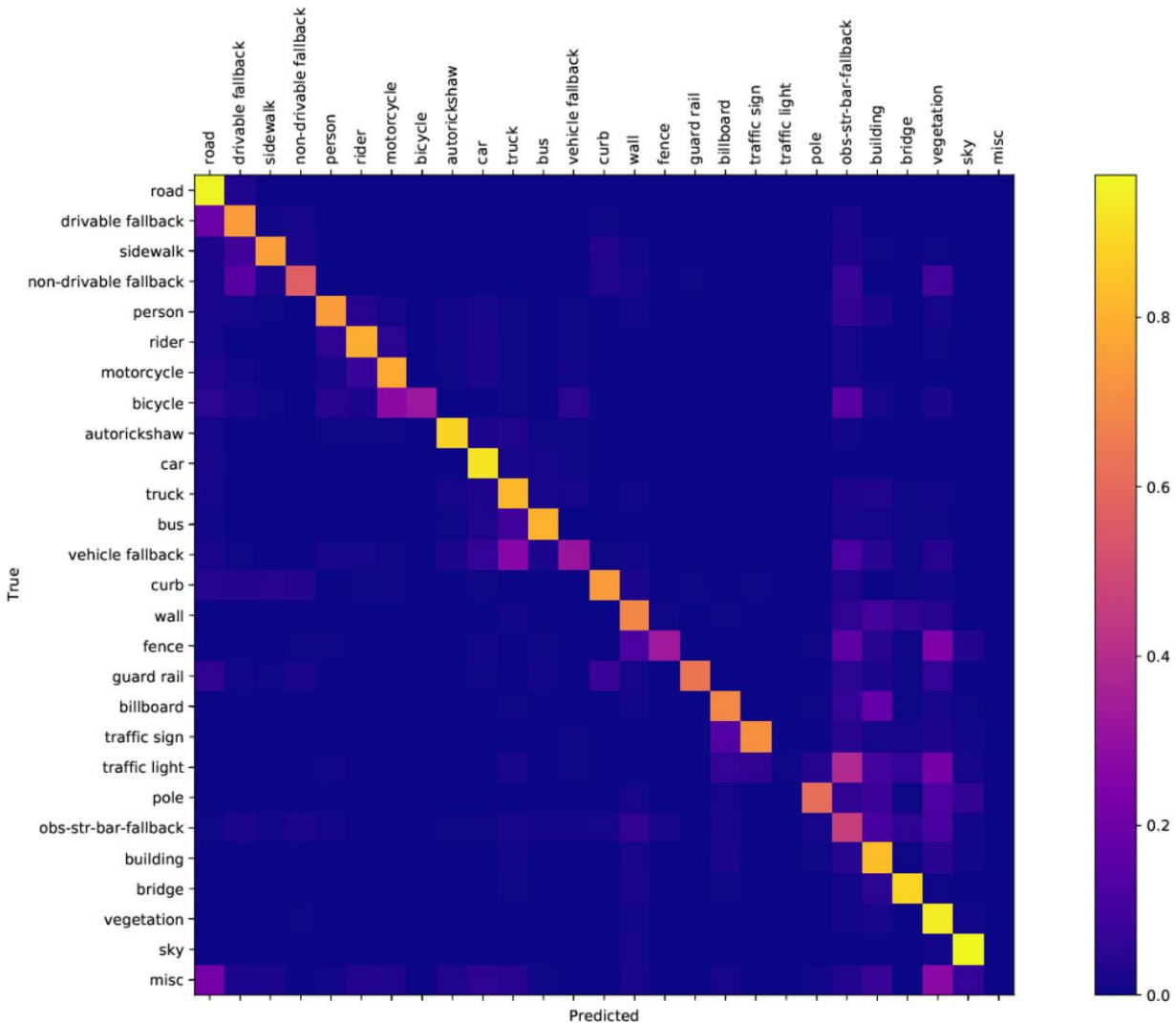
Figure 9. The IoUs for every class for the DRN D 38 model trained on IDDwith mIoU of 66.5%.

Domain Discrepancy

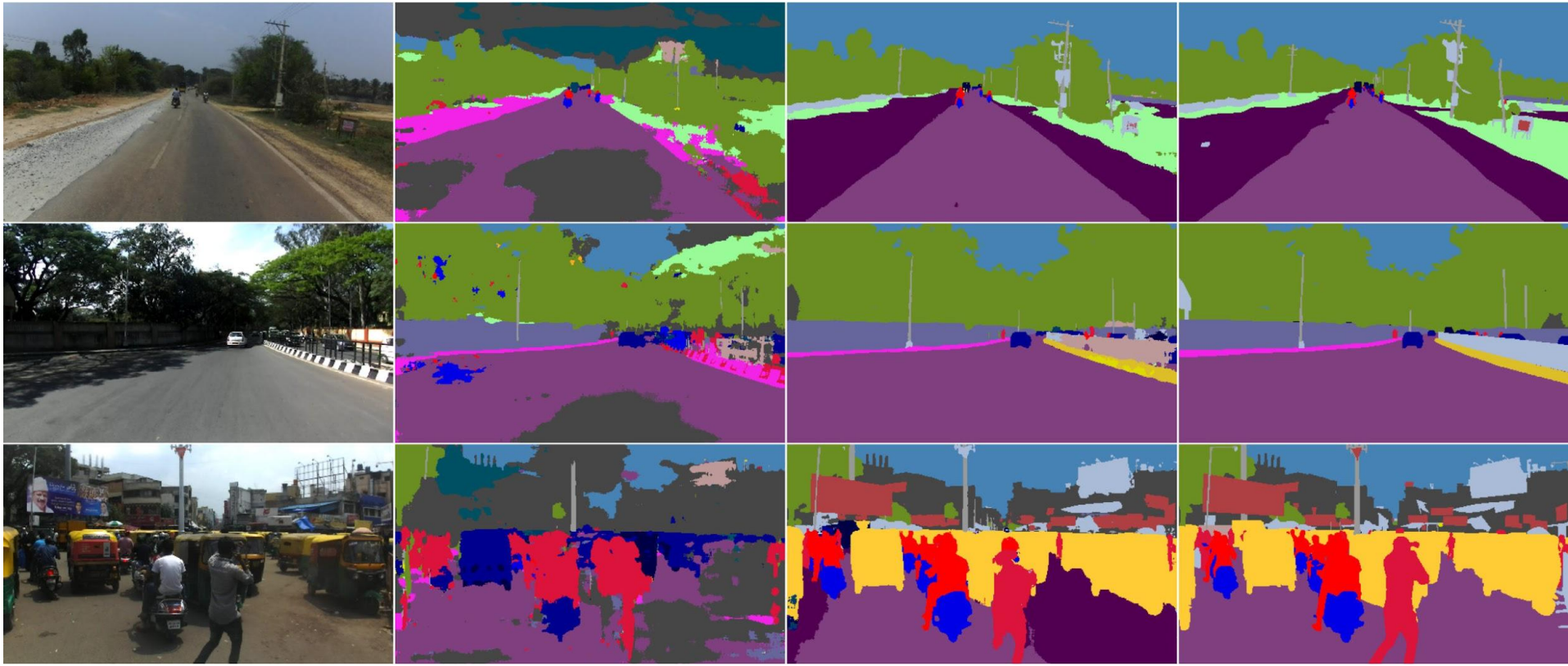
Train	Test	road	sidewalk	person	motorcycle	bicycle	car	truck	bus	wall	fence	traffic sign	traffic light	pole	building	vegetation	sky	mIoU of common labels
CS	DS	72	22	30	47	10	58	30	19	17	13	19	8	23	32	76	68	34
DS	CS	81	26	74	34	55	85	16	17	21	24	25	21	47	77	90	88	49
BD	ID	83	0	38	44	2	52	21	13	0	0	0	0	36	42	83	94	32
ID	BD	84	16	57	34	44	77	14	24	10	33	18	13	41	68	82	87	44
CS	CS	98	84	81	60	76	94	56	78	49	58	77	67	62	92	92	94	76
MV	MV	85	58	73	55	61	90	61	65	45	58	72	67	50	86	90	98	70
ID	ID	92	68	73	80	42	89	79	78	64	45	60	38	58	75	90	97	70
BD	BD	95	62	61	32	22	90	52	57	25	45	52	58	49	85	87	97	60

Table 3. The domain discrepancy between Cityscapes (CS) [5], Mapillary Vistas (MV) [16], Berkeley Deepdrive (BD) [26] Dataset and IDD (ID) using the DRN-D-38 Model [25]. Performance for only the common labels between the four datasets are used. First two rows compares the accuracy of a model trained on one of IDD or Cityscapes and tested on the other dataset. As can be seen, IDD trained model can predict CS and BD labels, better than predictions of trained models of the corresponding datasets on IDD. The bottom four rows gives the performance of models in each of the datasets. IDD dataset is harder than CS dataset and similar in hardness to MV on these 16 labels. BD is harder because i.) it has night scenes ii.) the images are take from a dash board cam, hence has reflections from inside the car as well as distortions like rain drops on the mirror.

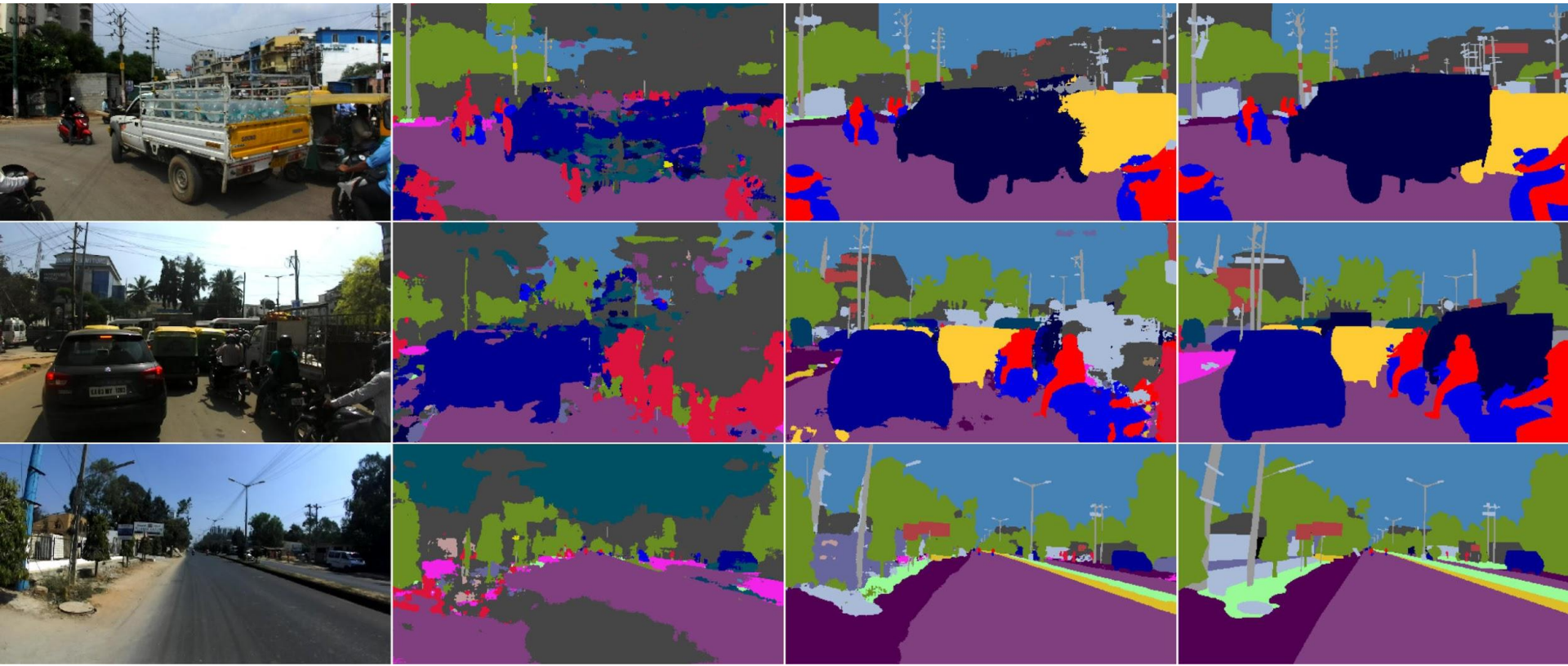
Confusion Matrix



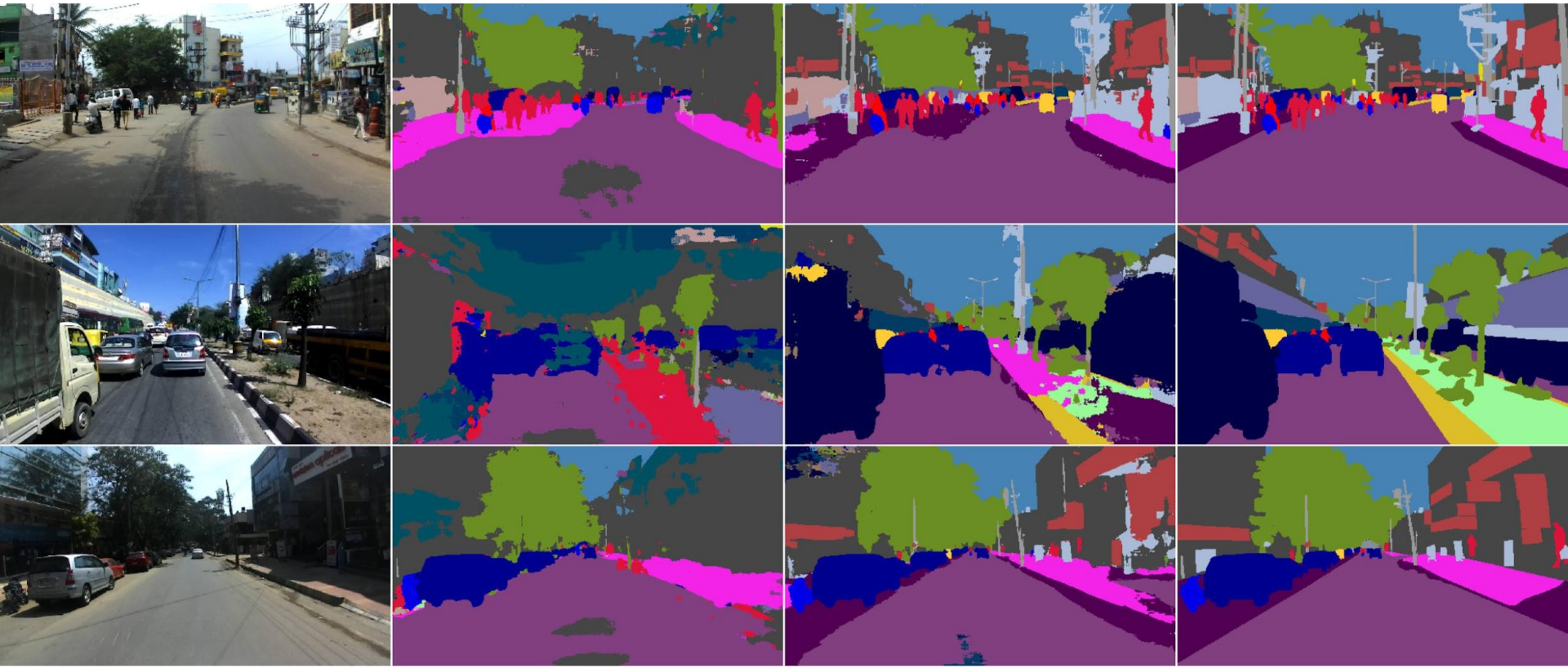
Qualitative Examples



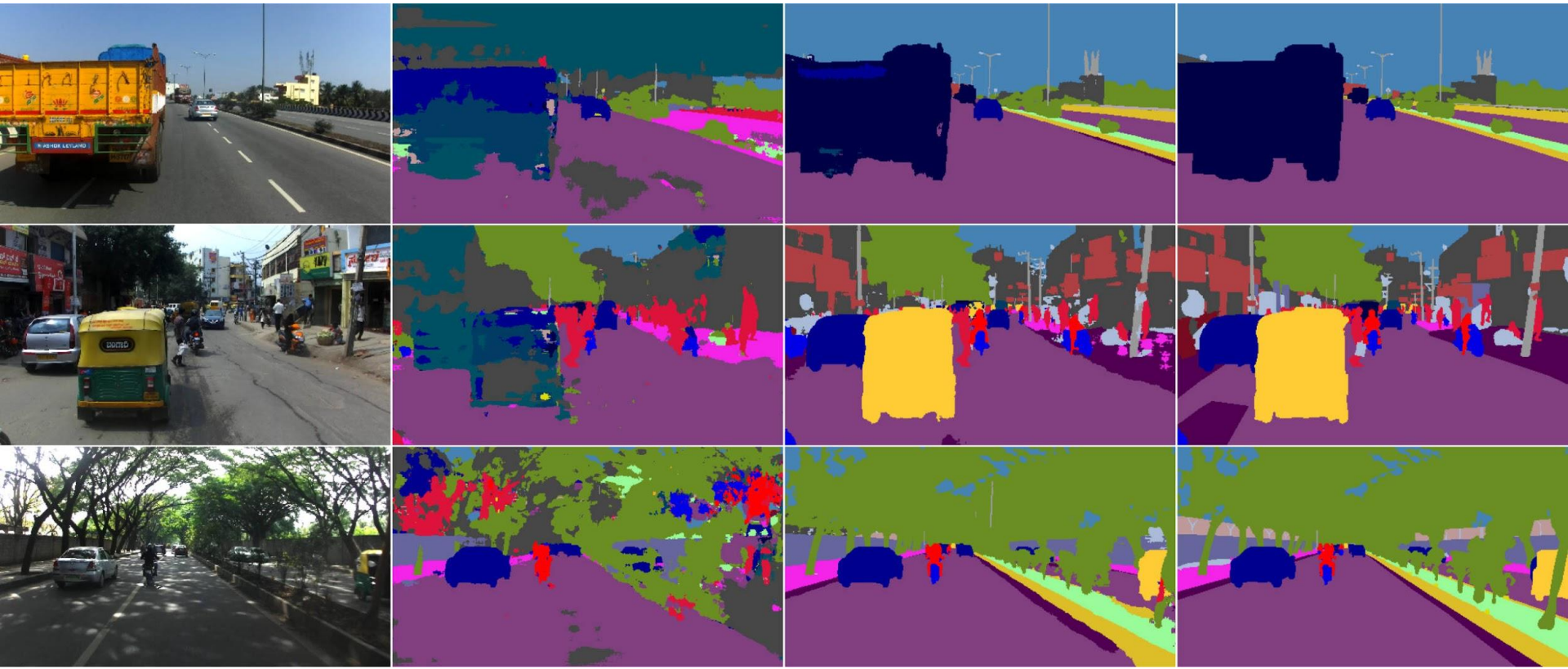
Qualitative Examples



Qualitative Examples



Qualitative Examples



Video

References

Research Paper:
IDD: A dataset for exploring problem of Autonomous Navigation in Unconstrained Environments. G. Varma, A. Subramanian, A. Namboodiri, M. Chandrakar, C. Jawahar. IEEE Winter Conference on Applications of Computer Vision (**WACV'19**)
http://idd.insaan.iiit.ac.in/publications/publications/idd-650.pdf

Workshop & Challenge:
AutoNUE, ECCV '18. Munich, Germany.
http://cvit.iiit.ac.in/scene-understanding-challenge-2018/
http://cvit.iiit.ac.in/autonue2018/

Other Research Problems

- Domain Adaptation
- Few shot segmentation
- Path planning
- Transport Density Estimation
- Road Safety Audit
- More Realistic Simulations