## Research paper-1: IDD: A Dataset for Exploring Problems of Autonomous Navigation in Unconstrained Environments

Source: <https://arxiv.org/pdf/1811.10200.pdf>

## Summary

Introduction:

* Autonomous navigation is rapidly maturing towards becoming a mainstream technology, with even consumer deployment by major automobile manufacturers.
* A significant contributor to this progress has been the availability of large scale datasets for sensing and scene understanding
* The author of the paper proposes a dataset that helps to solve problems involving Autonomous navigation, road scene understanding, and Few Shot learning.
* The data set consists of more than 10,004 images, finely annotated with 34 classes collected from 182 drive sequences on Indian roads that are captured in unstructured environments.
* Data consists of four-level label hierarchy with the number of labels as 30 (level 4), 26 (level 3), 16 (level 2), and 7 (level 1) labels, respectively, giving different complexity levels for training models
* Similar Data sets that are available are Cityscapes and KITTI but they are captured in well-delineated infrastructure such as lanes, a small number of well-defined categories for traffic participants, strict adherence to traffic rules, etc.

Challenges in Unstructured Environments:

* **Ambiguous Road Boundaries:** Roadsides can have muddy terrain, while also being drivable to some extent. Roads themselves can be covered by dirt or mud, making the boundaries very ambiguous
* **Diversity of Vehicles and Pedestrians:** The road consists of a variety of vehicles that include Buses, Trucks, Auto rickshaws and motorcycle with multiple riders where they are less likely to stick to traffic rules
* **Extensive Use of Information Boards:** Information boards display information like directions and landmark but they can also be present on the vehicle as advertisements.
* **Diversity of Ambient Conditions:** Data can include images with lighting variation where the image is captured at various times of the day, like mid-day, night and image can also include fog, dusk, etc.

Data Acquisition:

* The data is collected from Bangalore and Hyderabad cities in India and their outskirts and images have a mix of urban and rural, highway, single lane, and double lane roads with a variety of traffic and with large diversity.
* Data is highly instructed due to

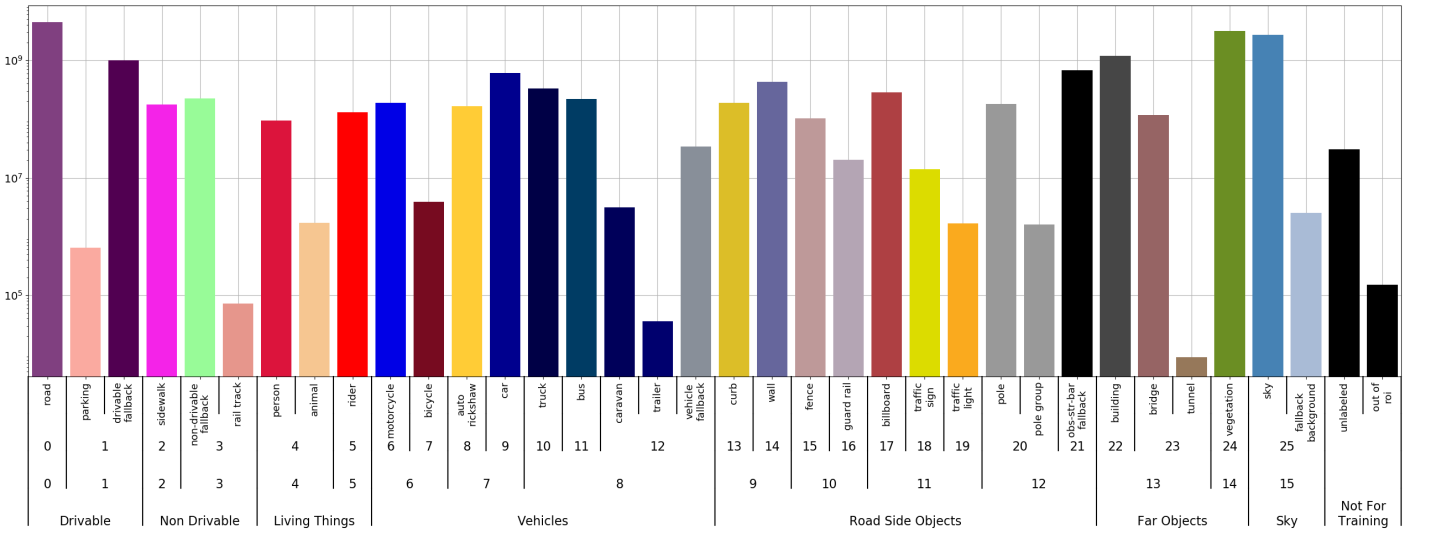
1. Cities are rapidly growing and have a lot of construction
2. Road boundaries are not well defined
3. The variety of vehicle and high density of motorcycle
4. Traffic participants don’t follow traffic rules most of the time.

* A total of 182 drive sequences were used for the preparation of the dataset.

**Frame Selection:**

* The image is chosen from forward-facing cameras on a car.
* Images are sampled at varying rates from videos from all drive sequences and they are annotated very finely, by layered polygon masks similar to Cityscapes.

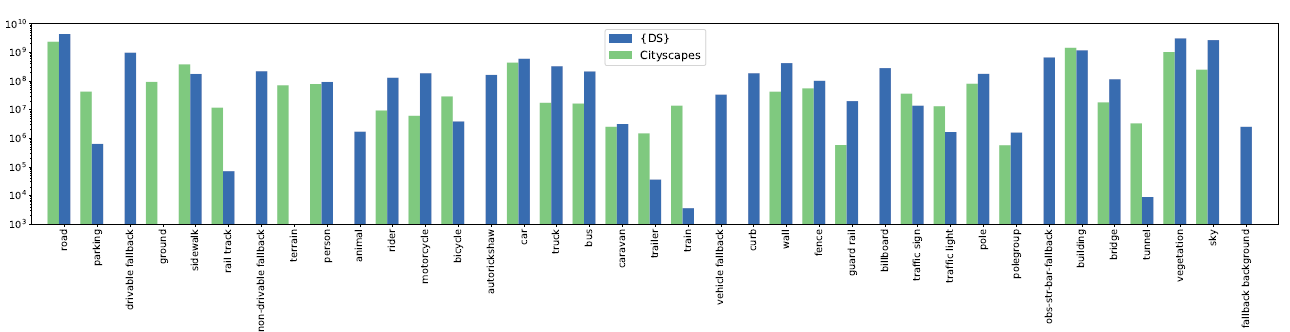
**Label Hierarchy & Annotation:**

****

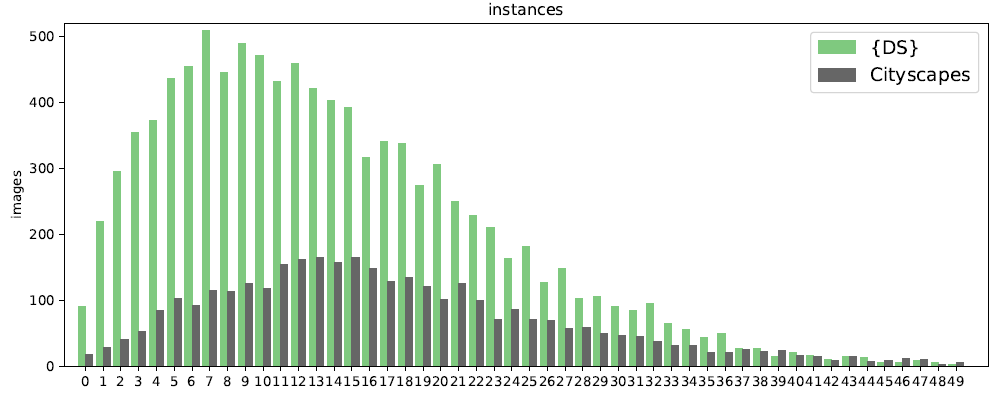
* The label hierarchy consists of 4 levels as above having 7 (level 1), 16 (level 2), 26 (level 3), and 30 (level 4) labels.
* Fine Annotations consists of 34 labels but there was some ambiguity while labeling the image For example labels like parking, caravan, or trailer cannot be precisely defined due to the diversity of the scenes and vehicles in the data collected.
* The higher level like level 4 has some ambiguity among the labels the lower levels have lesser ambiguity
* Data consists of new labels like auto-rickshaw, billboards, animal, curb, and road labels consists of drivable and non-drivable fall back indicating safety which is not present in similar datasets like cityscapes.

**Statistical Analysis and Dataset Splits:**

**Comparison of the Pixel count in the dataset with Cityscapes**



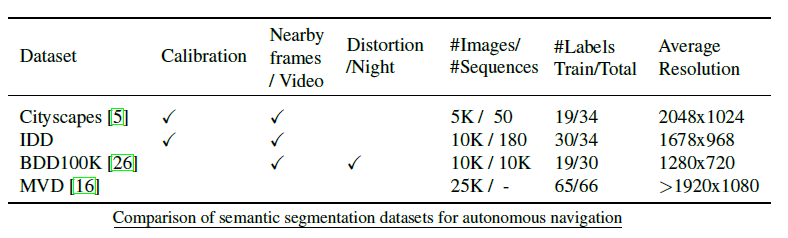
* The y-axis represents the pixel count is in log scale and x-axis represents labels in the data.
* The pixel statistics of the data areas above the labels in level 4 have high class imbalance, Labels like parking, animal, caravan, or traffic light have much fewer pixels. Since they have very few pixels that mostly fell within a few drive sequences.
* The lower levels labels are is well balanced and has a sufficient amount of pixels
* The Train, Validation, Test split is performed to 70:10:20 percent of total data where data is taken with this ratio from all drive sequences.
* Imbalance data problem is handled while splitting by ensuring labels with fewer pixel goes to all splits.



Comparison of traffic participants in the dataset with Cityscapes.

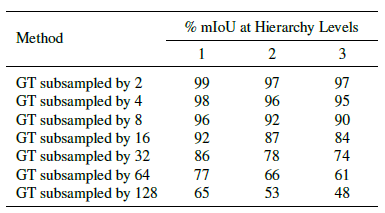
* The data contain more Traffic participants than compared to the cityscape dataset and have right-skewed distribution.
* The most common number of Traffic participants found is between 5 to 12 for the proposed dataset and 8 to 20 for cityscape dataset.
* The above distribution is taken from a small number of random samples from the dataset.

**Comparison with Other Datasets:**



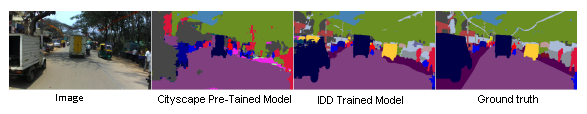
* There are many other similar datasets available on the internet like are Cityscapes, MVD, BDD100K, KITTI, Camvid, etc. but they all are captured in structured environments and focus on well-defined infrastructure.
* Some of the datasets suffer from distortion as they are captured from dashcam from inside.
* Pascal VOC, COCO, SUN are some of the datasets for the general image segmentation task.

**Control Experiments:**

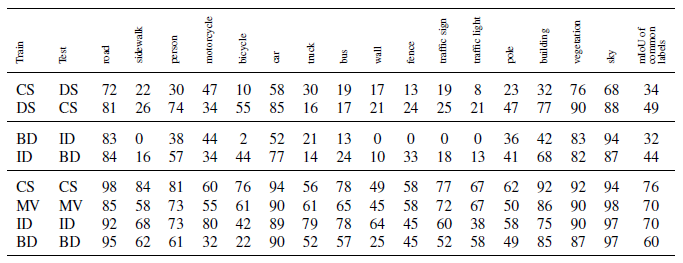


* Control experiments are performed for IOU scores from predictions at a given factor of the input resolution. Where the label is downsampled the by some factor and then upsampled to its original image size for evaluation.
* Results show that low-resolution processing contributes significantly to the overall degradation of segmentation results.

**Domain Discrepancy:**



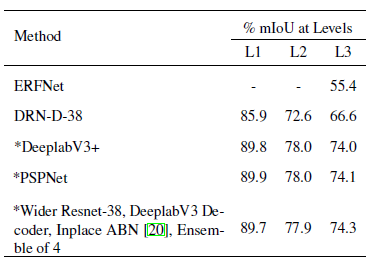
* A model trained on one domain dataset can perform predictions for other similar datasets that have different data distribution but have the same labels; Domain discrepancy studies the quantitative shift in data distributions between datasets.



* The 4 datasets that are considered as above where there are 16 common labels among them.
* A model trained on IDD can predict better than other models for the different domains as above.
* From a result, it is clear that the IDD dataset problem is hard when compared to other datasets.

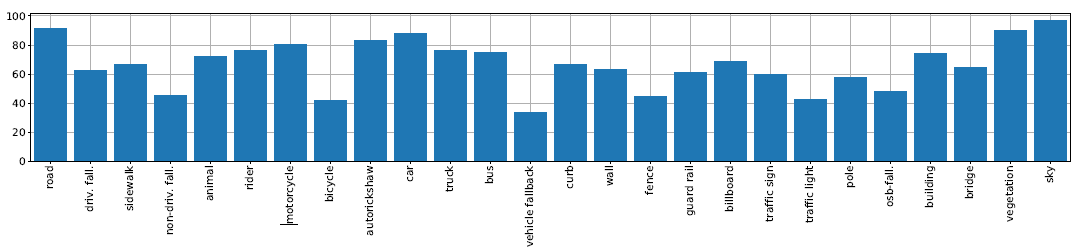
**Semantic Segmentation Benchmark:**

* The semantic segmentation benchmark on our dataset quantifies the mean Intersection over Union (mIoU) scores
* Level-4 labels like a traffic light, parking, or animal for which the number of labeled pixels are very few and to get rid of ambiguity the prediction is restricted to level-3.

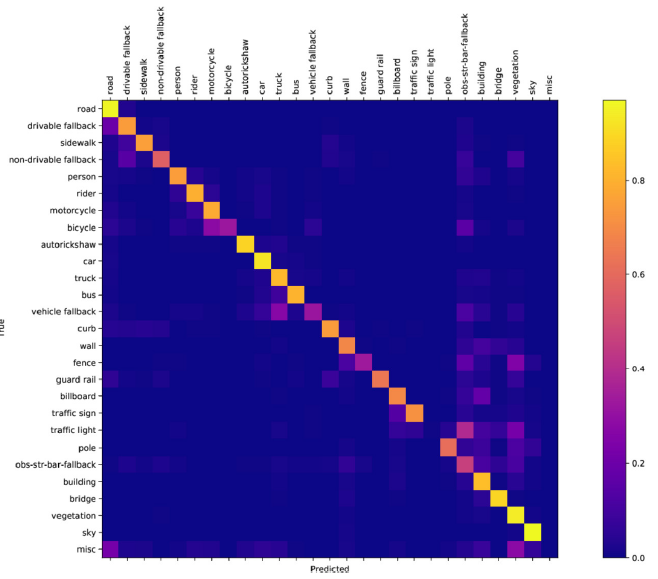


* The author has used DRN-D-38 and the ERFNet for benchmark and there is also challenges conducted and submissions are evaluated which use some of the state-of-the-art models.
* Pspnet outperforms all the models at level-1 and level-2 for the Data with best Miou Score.

**Class IoUs and Confusion Matrix:**



* The IoUs for every class for the DRN D 38 model trained on IDD with mIoU of 66.5%.
* IoUs is lower than 25% for bicycle, traffic light, vehicle fall-back, and fence labels because of the low pixel counts.



* There is significant confusion between:
  1. Motorcycle and bicycle.
  2. Billboard and traffic sign.
  3. Vegetation and traffic light.
  4. Building and billboard.
  5. Vegetation and wall, pole, fence.
  6. Drivable, non-drivable

**Conclusion:**

* This Dataset brings an ideal opportunity to solve new problems such as domain adaptation, few-shot learning, and behavior prediction in road scenes.
* This is a novel dataset for studying problems of autonomous navigations in unstructured driving conditions and which helps to build autonomous systems.
* This dataset covers several drawbacks of existing datasets, such as distinguishing safe or unsafe drivable areas and also has new labels with a large diversity.
* Finally Thanks to the authors and the Intel India team for the efforts that they have taken to capture the data and for making it available to everyone for experimentation.