**Semantic Segmentation - Indian Driving Dataset**

**Project Proposal – Group2**

**Author(s): Hemangi Kinger, Varun Shah, Ishan Kuchroo**

**11/8/2022**

**Abstract**

We can define Deep Learning as a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks. It is a technique that teaches computers to do what comes naturally to humans: learn by example.

The final project details the implementation of tools, libraries and techniques acquired during the course.

**Introduction**

Semantic image segmentation is a computer vision task in which we label specific regions of an image according to what's being shown. Mostly used in Autonomous vehicles and Medical Image diagnostics, the goal of semantic image segmentation is to label each pixel of an image with a corresponding class of what is being represented. Because we're predicting for every pixel in the image, this task is commonly referred to as dense prediction. One important thing to note is that we're not separating instances of the same class: we only care about the category of each pixel. In other words, if you have two objects of the same category in your input image, the segmentation map does not inherently distinguish these as separate objects.

IDD is a dataset for road scene understanding in unstructured environments used for semantic segmentation and object detection for autonomous driving.

**Data Overview**

**Overview**:

IDD consists of 10,004 images, finely annotated with 34 classes collected from 182 drive sequences on Indian roads. The label set is expanded in comparison to popular benchmarks such as Cityscapes, to account for new classes. It also reflects label distributions of road scenes significantly different

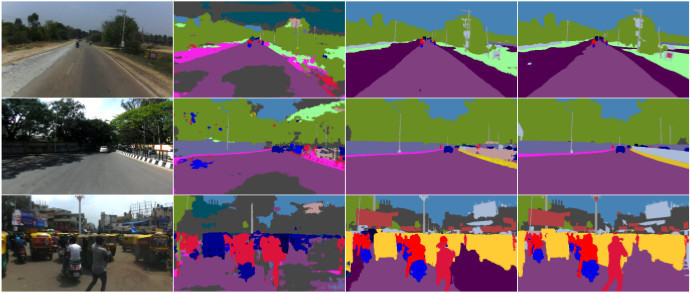
from existing datasets, with most classes displaying greater within-class diversity. Consistent with real driving behaviors, it also identifies new classes such as drivable areas besides the road. The dataset is inspired from the one used in the 2018 paper *[IDD: A Dataset for Exploring Problems of](https://arxiv.org/pdf/1811.10200v1.pdf)*

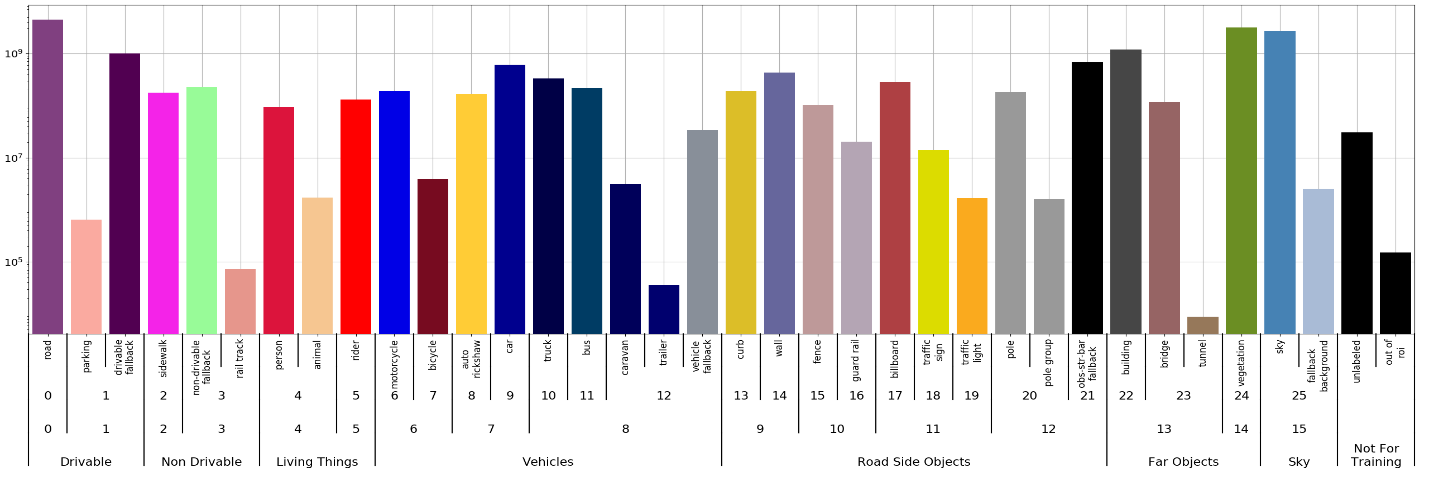
*[Autonomous Navigation in Unconstrained Environments](https://arxiv.org/pdf/1811.10200v1.pdf)*

**Source**:

<http://idd.insaan.iiit.ac.in/dataset/details/>

**Snapshot**:





**Computer Vision – Methods and Procedures**

For this project, we model unstructured environments using Convolution Neural Networks, and show that they achieve state-of-the-art performance.

**Convolution Neural Networks**:

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. A CNN can capture the Spatial and Temporal intricacies in an image through the application of relevant filters. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and reusability of weights.

Diagram, engineering drawing

Description automatically generated

**Reference**: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>

**Procedures and Packages**:

1. **TensorFlow**:

TensorFlow is an open-source collection of tools and libraries that allows researchers to build and deploy cutting-edge ML models. We’ll be using various methods from TensorFlow like “image.resize”, “cast” etc.

**Reference**: <https://www.tensorflow.org/>

1. **Keras**:

Keras is an open-source high-level neural network library that runs on top of TensorFlow. It provides a Python interface for artificial neural networks. We’ll use Keras to import and run a pre-trained Neural Network called “ResNet50”.

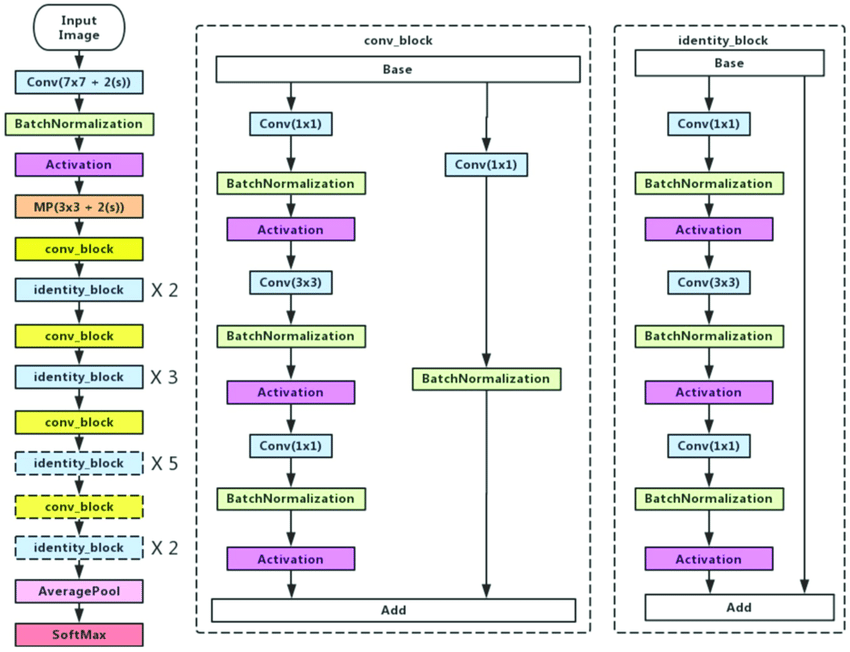
**Reference**: <https://keras.io/>

1. **ResNet50**:

ResNet or Residual Network is an innovative neural network that was first introduced by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun in their 2015 computer vision research paper titled ‘Deep Residual Learning for Image Recognition’.

**Resnet50** is one of the many ResNet variants that can work with 50 neural network layers.

**Reference**: <https://viso.ai/deep-learning/resnet-residual-neural-network/>



**Metrics**: Paperwithcode website for this dataset has various benchmarks. We’ll try to build/improve existing models and compare the accuracy of the model with existing benchmarks.

**Appendix**

GitHub Link - <https://github.com/IshanKuchroo/IDD-Indian-Driving-Dataset>

**References**

<https://paperswithcode.com/dataset/idd>

<https://www.jeremyjordan.me/semantic-segmentation/>

<https://arxiv.org/pdf/1811.10200v1.pdf>