Semantic Segmentation - Milestone 1

Thursday, March 12, 2020   5:23 PM

Topic: Semantic Segmentation and Pedestrian movement prediction
Team: Nikita Goswami 50320909 | Vimal Kumarasamy 50321558

Dataset: Camvid
- Details:
  - The images are from videos filmed with CCTV-style cameras, from the driver's perspective
  - 421 training images of dimensions 960 X 720 paired with 421 labels dataset
  - Every pixel in the image pertains to one of the 32 semantic classes

Animal	LaneMkgsNonDriv	SUVPickupTruck
Archway	Misc_Text	TrafficCone
Bicyclist	MotorcycleScooter	TrafficLight
Bridge	OtherMoving	Train
Building	ParkingBlock	Tree
Car	Pedestrian	Truck_Bus
CartLuggagePram	Road	Tunnel
Child	RoadShoulder	VegetationMisc
Column_Pole	Sidewalk	Void
Fence	SignSymbol	Wall
LaneMkgsDriv	Sky	

  - Test dataset contains 161 such images and labels and validation dataset contains 112 images and labels

Project Progress:
- Image procurement and loading on to disc - **Done**
  - It's an open data source, downloaded and placed in personal Google Colab server - GPU availability
- Programming language - Python - **Done**
- Data Treatment - **Done**
  - The data is procured predominantly from Europe and some minimal representation from USA, Asia etc.
  - Images might pertain to a particular driving conditions, so some annotation is required
  - Python functions are written to rotate, flip, crop, modify intensities on the images based on the user provided parameters
  - However these annotations are not sufficient to generalize the images for harsh weather conditions, challenging lighting conditions etc.
  - The initial results of the prediction can be studied before making further modifications on the images to improve generalizability
- Modelling architectures - **In progress**
  - The plan is to leverage any state of art modelling architectures to learn the features - transfer learning
  - Exploring techniques to leverage the pretrained weights from the below architectures and passing the training data to further refine/include more elements
    - ResNet - Neural networks with skip connections to tackle the vanishing gradients

problem in deep neural networks
- Atrous Convolution based DeepLab V3 - Google's research on semantic segmentation to identify the entities in an image, with Atrous convolutions that is capable of downsizing and upsizing pyramids without multiple iterations. This will save training time
- The existing architectures and weights in the initial feature layers would be leveraged, and the some modification would be done on the networks to solve the specific problem at hand
- Tracking pedestrians - **Yet to start**
  - Once we have a method to identify the segments in an image, the successive images can be scored on the model and movement among the pedestrians can be studied
  - Presence or absence of traffic signal coupled with pedestrian movement can help us make decisions
  - This piece would be started once we have a sound architecture based on which the segments can be identified from an image accurately

References:
http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/
https://arxiv.org/abs/1802.02611
https://arxiv.org/abs/1512.03385