Capstone Proposal

Domain background :
Scene Understanding using computer vision

Problem Statement:
The image semantic segmentation challenge consists in classifying each pixel of an image (or just several ones) into an instance, each instance (or category) corresponding to an object or a part of the image (road, sky, …). This task is part of the concept of scene understanding: how a deep learning model can better learn the global context of a visual content ?

Datasets and Input:
The dataset is Cityscapes (https://www.cityscapes-dataset.com/) which is an urban scene understanding dataset and could be used for various tasks like pixel-level, instance-level and panoptic level semantic labelling.

There are 20 classes including background and images are divided in 3 categories (train/val/test). These are high quality images and for each image there is a segmented map which should be used as ground truth.

There are 2975 high resolution Images in the training set.

Solution Statement and Project design:
For this task, I plan to use Unet and PSPnet for semantic segmentation, Unet has done well in medical domain segmentation so it will be interesting to see if it will do well in this case with so many classes present in each image. PSPnet was a winner of Imagenet scene parsing challenge (2016) and seems to provide good accuracy.

The roadmap of the project looks as follows:

- Data Understanding - this includes visualization of images and class distribution of the dataset
- Data preparation for training - this includes implementation of data pipeline with data augmentation and required data format for training
- Model Architecture- this includes implementation of Unet and PSPnet in pytorch
- Training and evaluation - training the dataset and evaluating it on validation set using mIoU metric.

Note - for training I don't plan to use sagemaker.

Benchmark model and Evaluation metrics:
Semantic segmentation is particularly tough compared to object classification and object detection as it is dense pixel level prediction and involves many challenges like fine boundary predictions, multi scales of objects and context understanding.

Cityscapes dataset contains official benchmarks, however it does not contain Unet benchmark. Since Unet performs well for medical domain where we have very few classes compared to this scenario, I am expecting to obtain 0.50 mIoU and using PSPnet should result in better segmentation.

Evaluation metric for this task will be mIoU (mean IoU over all classes).

References:

Unet (https://arxiv.org/abs/1505.04597)
Pyramid Scene Parsing Network (https://arxiv.org/abs/1612.01105)