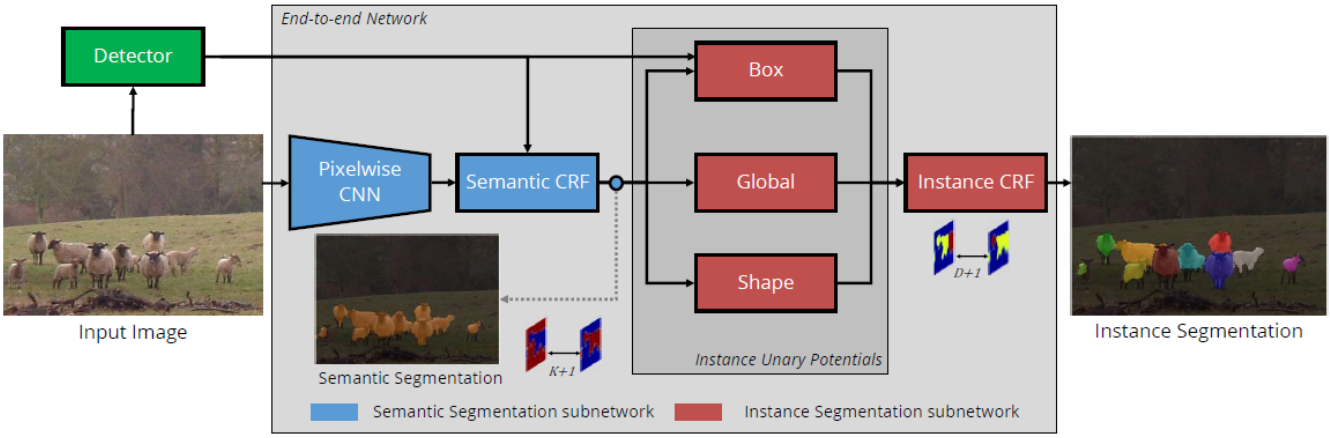
Pixelwise Instance Segmentation with a Dynamically Instantiated Network [1]

Prathyusha Akundi(2018701014) Sowmya Aitha(2018702007) Tanu Sharma(2018702012)

Abstract:

We are aware that Semantic segmentation task has no notion of number of instances in the image and object detection task operates at a coarse bounding box level. An Instance Segmentation system that produces a segmentation map where each pixel is assigned an object class and instance identity label is proposed. This method is based on an initial semantic segmentation module, which feeds into an instance subnetwork. This subnetwork uses the initial category-level segmentation, along with cues from the output of an object detector, within an end-to-end CRF to predict instances. This part of our model is dynamically instantiated to produce a variable number of instances per image.

Network Architecture:



This network has two subnetworks: Semantic segmentation and Instance Segmentation. The intermediate category-level segmentation, along with the outputs of an object detector, are used to reason about instances. This is done by instance unary terms which use information from the detector's bounding boxes, the initial semantic segmentation and also the object's shape. A final CRF is used to combine all this information together to obtain an instance segmentation

Completed Tasks:

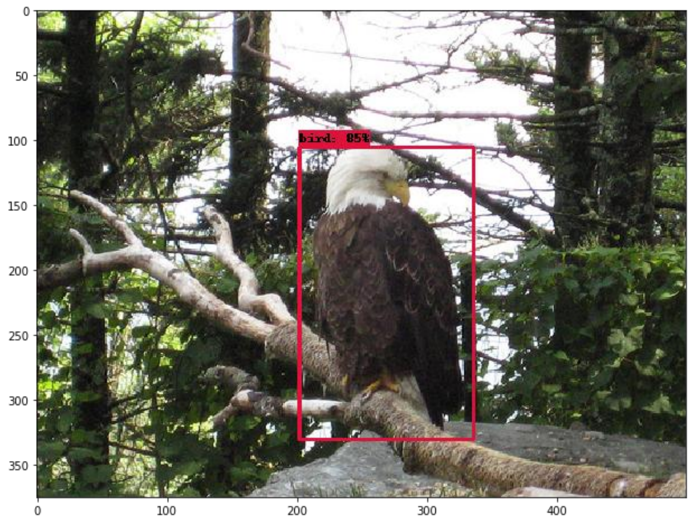
- ### Object Detector:

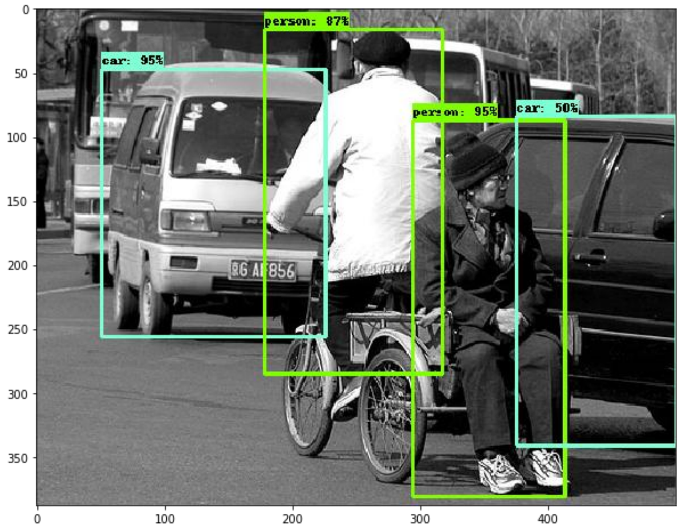
  We have used existing Tensorflow Object Detection API to detect the objects in the given image. We have considered MS COCO 2017 dataset which has a total of 90 different classes, for training. We made use of pre-trained MobileNet model with the following configurations: Dropout: 0.8, Regularizer: L2, IOU threshold: 0.6, Learning rate: 0.004, Momentum: 0.9.

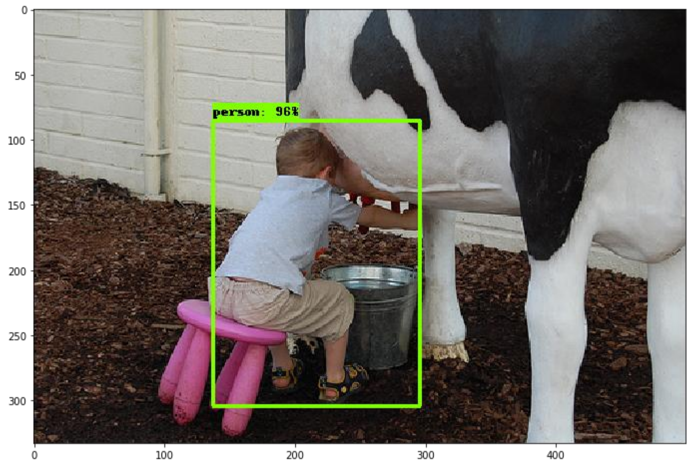
  We have tested the model for images in PASCAL VOC 2011 dataset. The following are some of the results:

Results:

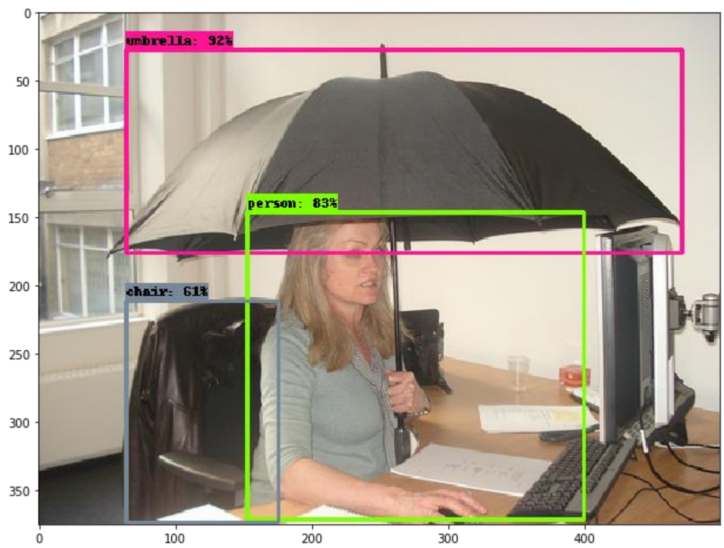

airplane: 65%
tv: 71%

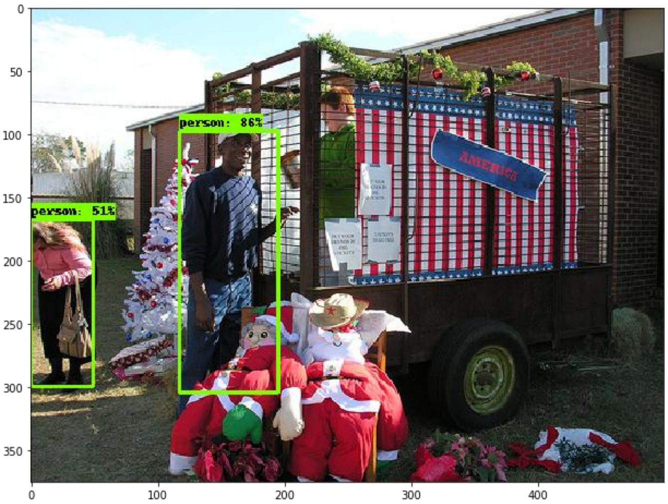

bird: 85%

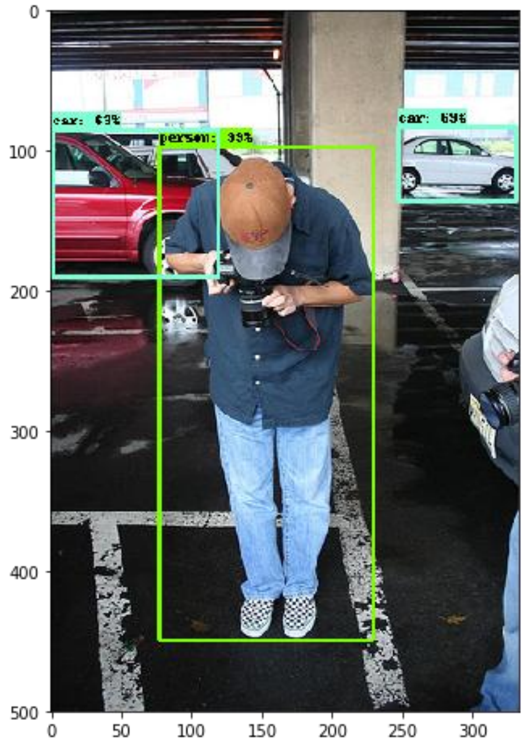

person: 87%
car: 95%
person: 95%
car: 50%

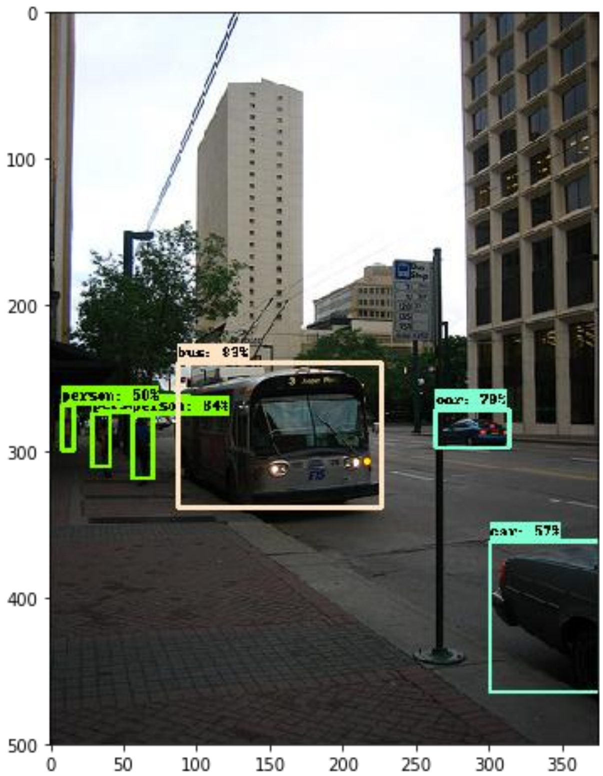

person: 96%

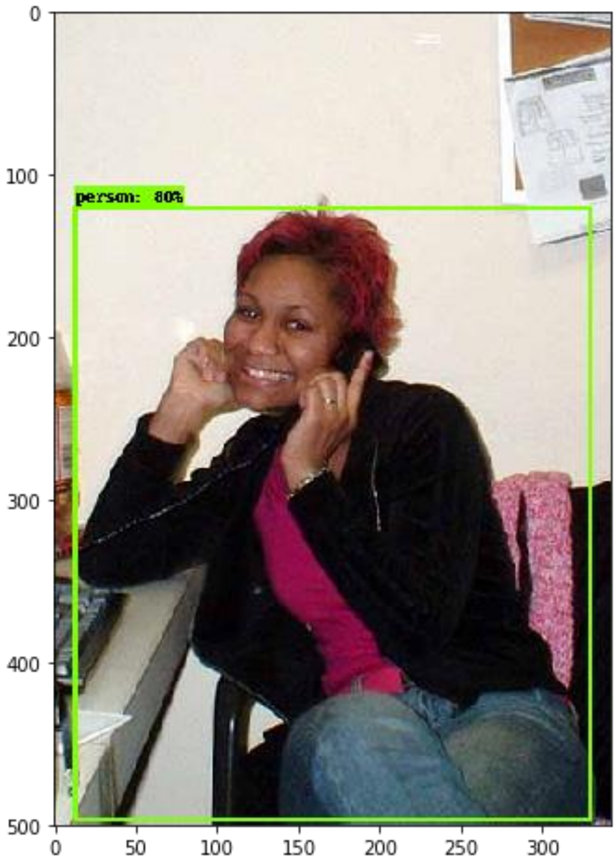

train: 71%


umbrella: 92%
person: 83%
chair: 61%


person: 86%
person: 51%
AMERICA


person: 99%
car: 69%

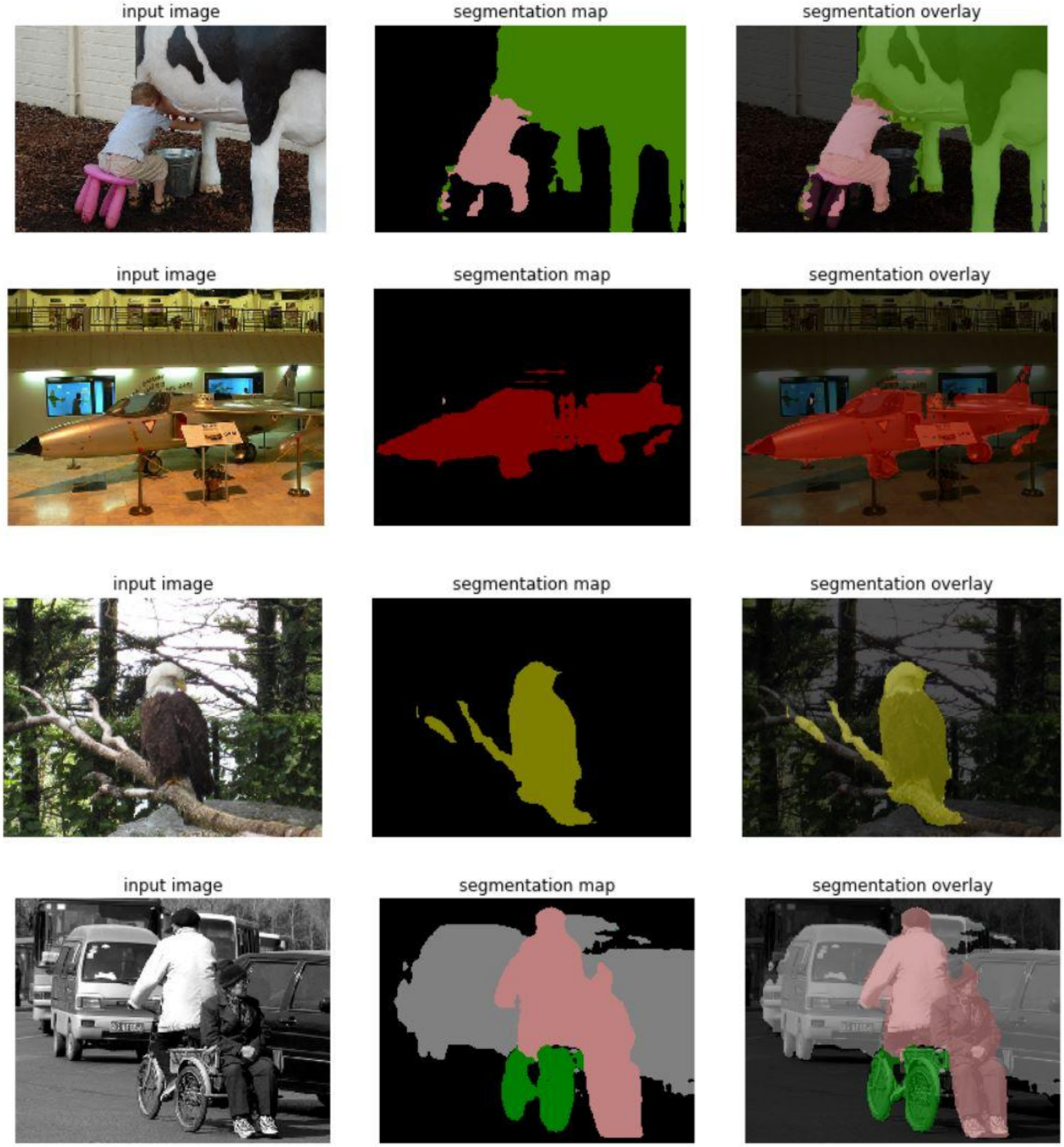

bus: 83%
person: 50%
person: 84%
car: 79%
car: 57%

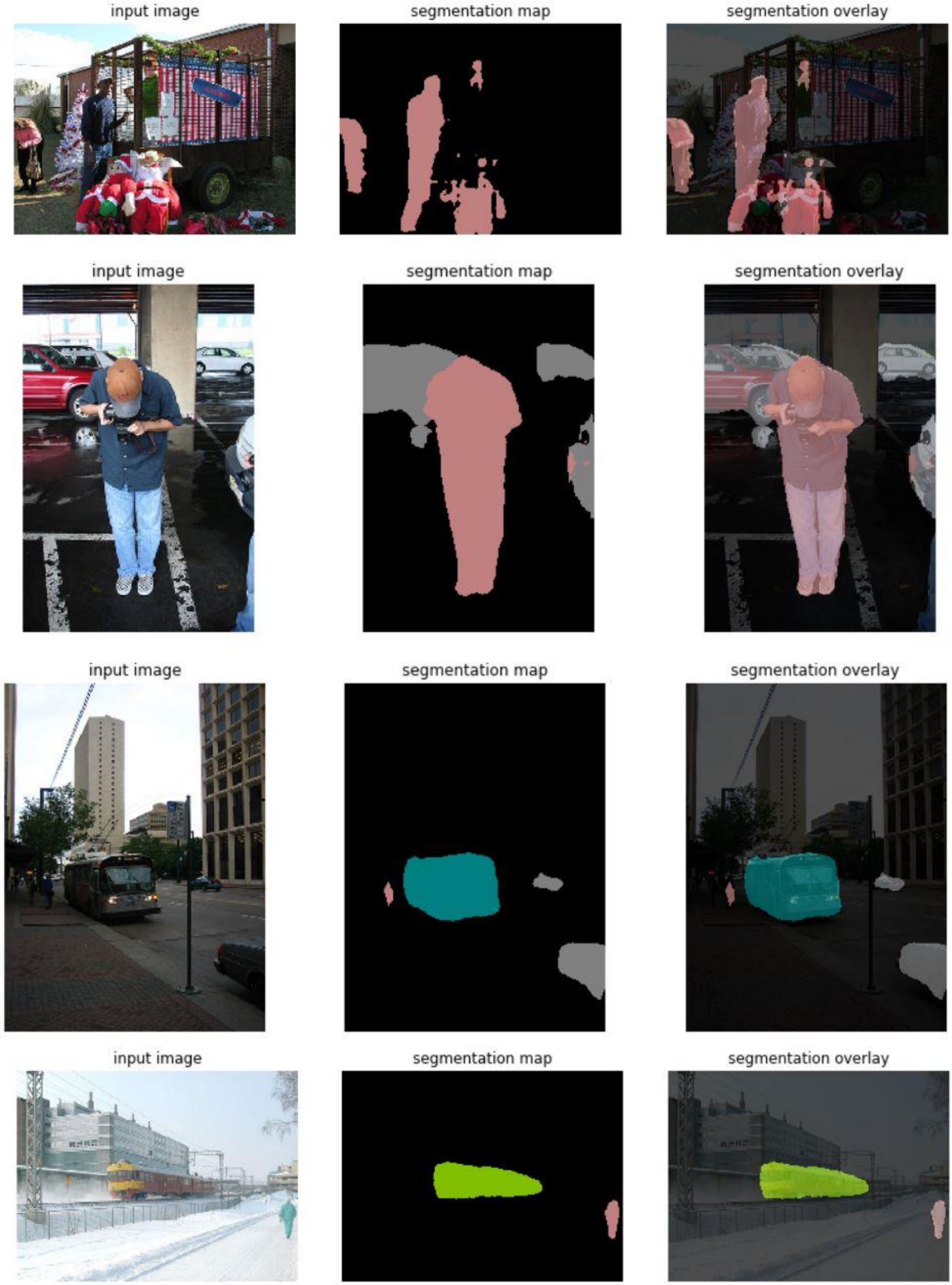

person: 80%

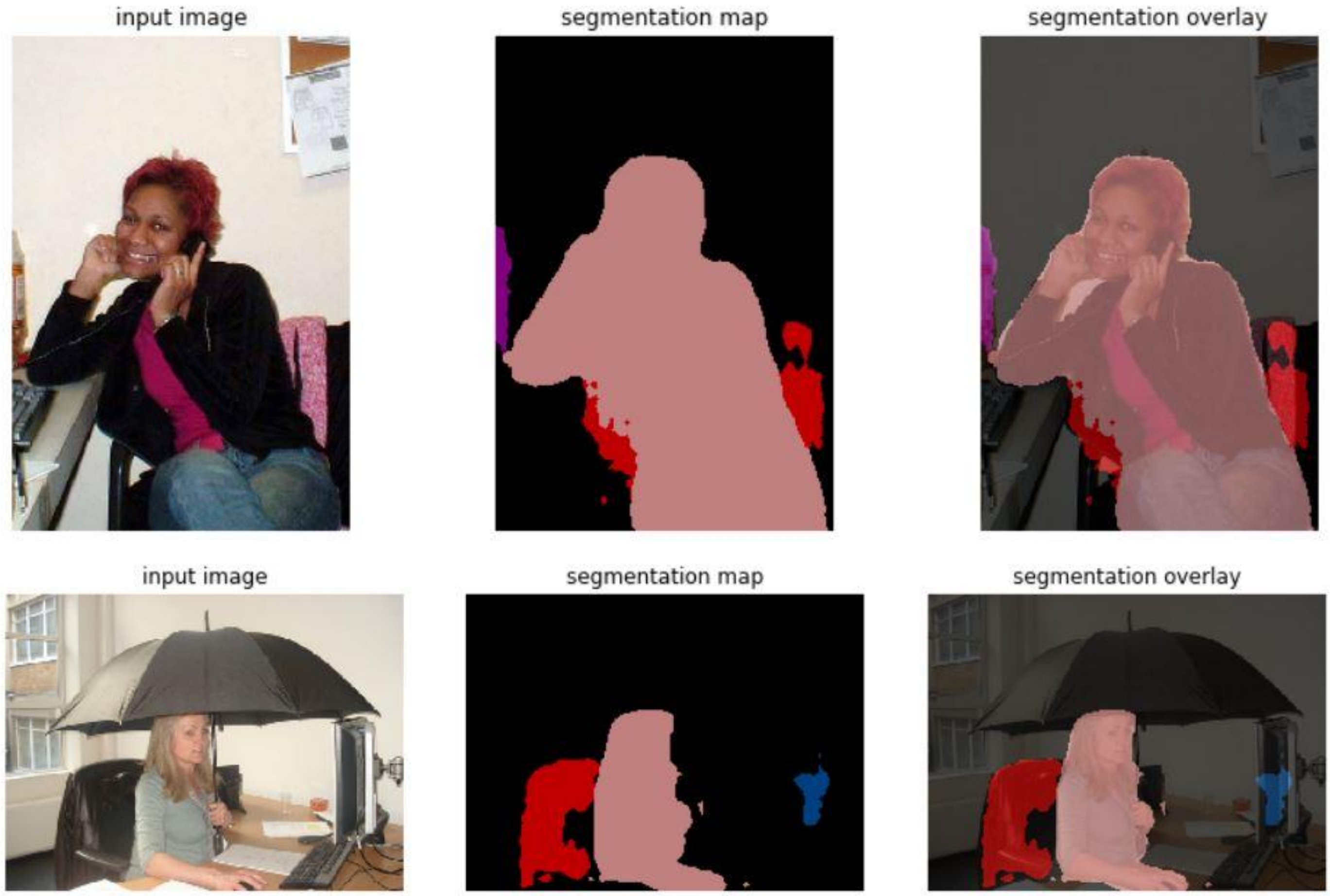
- ## Semantic Segmentation:

  We have used FCN-8 architecture to perform semantic segmentation on PASCAL VOC 2011 dataset. We utilized pre-trained weights for the network. It has a total of 46 layers. The following are the results for semantic segmentation.

  ### Results:




input image
segmentation map
segmentation overlay
input image
segmentation map
segmentation overlay
input image
segmentation map
segmentation overlay
input image
segmentation map
segmentation overlay



Task in Progress:

Shape Term: We consider shape priors to help us reason about occlusions involving multiple objects of the same class. Because, a prior on the expected shape of an object category can help us to identify the foreground instance within a bounding box. The following is the algorithm to calculate shape priors:

Algorithm [3]:

1. Learning soft mask dictionary
    a. Cluster all bounding boxes of training set into five aspect ratio clusters using K-means.
    b. Extract mask for each bounding box and resize mask into low resolution thumbnail.
    c. Hierarchical K-means clustering in mask pixel space for every unique (aspect ratio, class) pair that exists in the data set.
    d. Pool all the mask clusters for a given aspect ratio into a single set.
2. Learning a soft mask classifier - Learn a linear SVM classifier using HoG features to differentiate between exemplars in each soft mask cluster.
3. Localization with superpixels - Integrate soft mask over underlying superpixel and normalize by the area of each superpixel.

Tasks to be done:

1. Post processing semantic segmentation network with CRFs using *CRF as RNN* [2] paper.
2. Complete shape term calculation
3. A final CRF (Instance segmentation subnetwork) to combine: Box, Global and Shape information to obtain an instance segmentation.

Github link: https://github.com/Prathyusha-Akundi/Pixelwise_Instance_Segmentation.git

References:

[1] A. Arnab and P. H. S. Torr. Pixelwise instance segmentation with a dynamically instantiated network. In CVPR, 2017

[2] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. Torr. Conditional random fields as recurrent neural networks.

[3] D. Weiss and B. Taskar. Scalpel: Segmentation cascades with localized priors and efficient learning