CS783 VISUAL RECOGNITION
Assignment 4

Praphul Kumar Singh - 150513
Shashank Singh - 150663

1 Introduction

In the given assignment, we have been given a vehicle dataset. The dataset has been retrieved from the IIT Kanpur surveillance cameras and consists of a variety of vehicle classes like, cars, buses, bicycles, motorbikes and pedestrians. This data can be used to implement various computer vision tasks and we have explored the problem of segmenting the dataset. We have used two different methods as described below :-

2 Approaches

2.1 Unsupervised Image Segmentation by BackProapagation

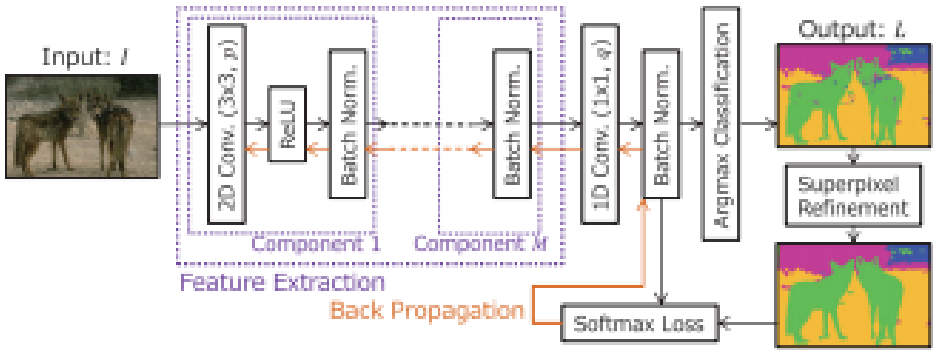
Given an RGB image where each pixel is a 3-dimensional vector, this method computes a feature vector for each pixel by passing it through a convolutional network and then the pixels are assigned labels using the method of k-mean clustering.

Let x_n be the feature vector for the n_{th} pixel in the image and $f(x_n)$ be a function which predicts the cluster label c_n for the particular pixel. Now we have three things, x_n, $f(x_n)$ and c_n which need to be trained. We do this by alternately fixing parameters for two things and training the third function. For example, if c_n is being predicted we keep x_n and $f(x_n)$ constant. The weights are updated by using backpropagation method using stochastic gradient optimizer.

For good segmentation, certain characteristics are required for the cluster label c_n. Instance of any object contains patches of similar texture patterns. This is taken into account while performing the segmentation. Hence, spatially continuous pixels that have similar color and texture patterns should be

grouped together. On the other hand, different object instances should be categorized separately. To facilitate this cluster separation, the number of cluster labels is desired to be large. CNN architecture is used to extract the pixel features. This CNN assigns the cluster labels to image pixels and updates the convolutional filters for better separation of clusters. Backpropagation of softmax loss is used to update the network.

The model architecture is given below:

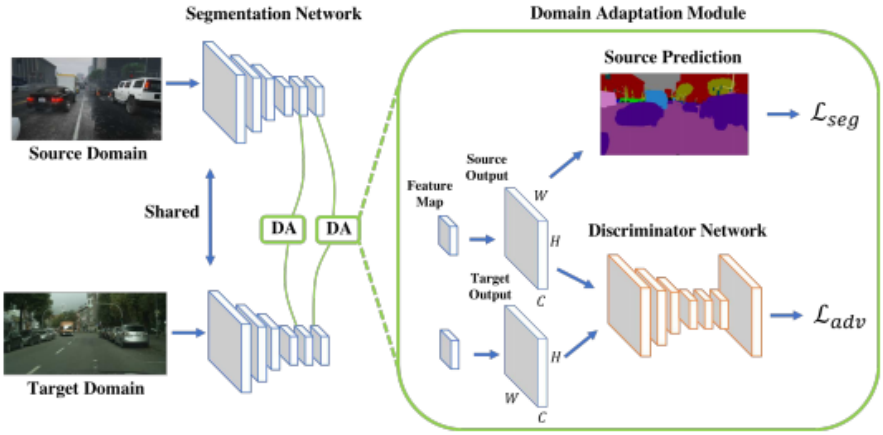


2.2 Domain Adaptation for Segmentation

In this method, knowledge transfer or domain adaptation is done to close the gap of distributions of source and target domains. Based on the General Adversial Network, this model consits of two parts:
1) Segmentation Network to predict the segmentation of the images, and
2) Discriminator Network to tell whether the input image is from source domain or target domain.

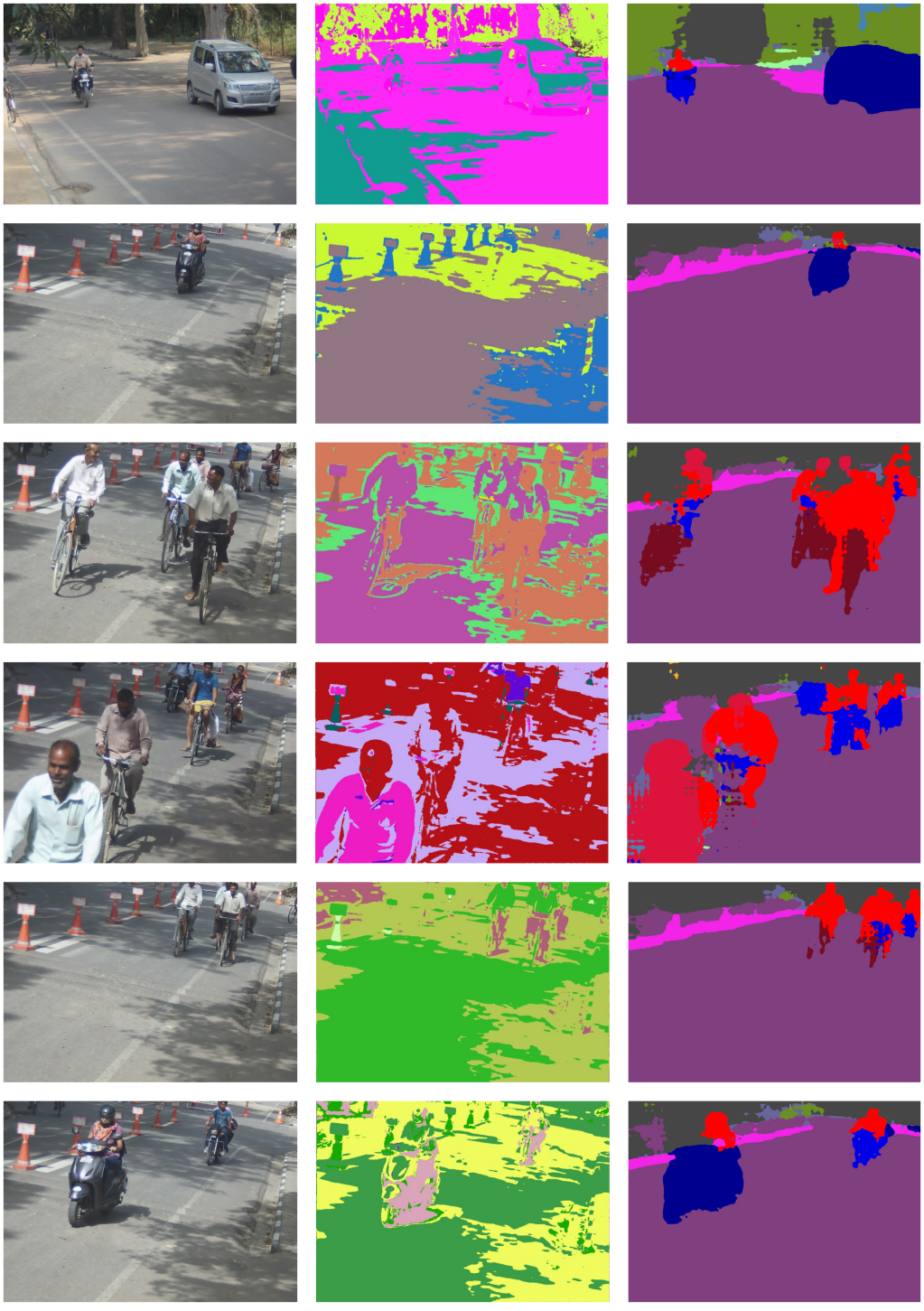
We first used a pretrained model adapted on Cityscapes dataset through GTA dataset directly to visualise the results which were not significant. Then we used the GTA dataset as the source domain and the IITK dataset as the target domain to train our separate model.

The model architecture is as follows:



On the one hand the loss for the segmentation is minimized while simultaneously the loss for the discriminator network is maximised so that it becomes hard for the discriminator network to distinguish between the source domain and target domain and thus the segmentation network can easily use the transferred knowledge from GTA dataset(which is labelled) to predict the segmentation for the IITK Dataset.

3 Results Obtained By Both Methods Respectively:



References

[1] Richter 2016 ECCV, author = Stephan R. Richter and Vibhav Vineet and Stefan Roth and Vladlen Koltun, title = Playing for Data: Ground Truth from Computer Games, booktitle = European Conference on Computer Vision (ECCV), year = 2016, editor = Bastian Leibe and Jiri Matas and Nicu Sebe and Max Welling, series = LNCS, volume = 9906, publisher = Springer International Publishing, pages = 102–118

[2] @articleKanezaki2018UnsupervisedIS, title=Unsupervised Image Segmentation by Backpropagation, author=Asako Kanezaki, journal=2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), year=2018, pages=1543-1547

[3] https://github.com/kanezaki/pytorch-unsupervised-segmentation

[4] https://github.com/wasidennis/AdaptSegNet

[5] @articleTsai2018LearningTA, title=Learning to Adapt Structured Output Space for Semantic Segmentation, author=Yi-Hsuan Tsai and Wei-Chih Hung and Samuel Schulter and Kihyuk Sohn and Ming-Hsuan Yang and Manmohan Krishna Chandraker, journal=2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, year=2018, pages=7472-7481