Semantic Segmentation using MATLAB

Abhishek Thakur, P. Adithya Vardhan, Radhika Singh
SRM University, Neerukonda, A.P - India

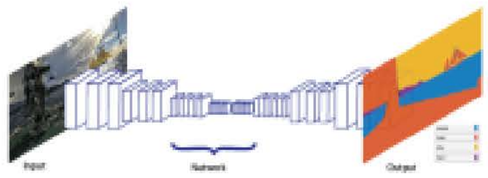
Abstract

Semantic segmentation refers to the process of linking each pixel in an image to a class label. For example, semantic segmentation is very crucial in self-driving cars and robotics because, it is important for the models to understand the context in the environment in which they're operating. Segmentation is an important stage of the image recognition system, as it extracts the objects of our interest, for further processing such as description or recognition. Segmentation techniques are used to isolate the desired object from the image in order to perform analysis of the object.

More specifically, 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. The semantic segmentation is a crucial part in automated vehicles and robotics. This helps the models to understand the context in their surrounding and where they are operating in order to have a safe operation.

Introduction

Semantic segmentation is very crucial in self-driving cars and robotics because it is important for the models to understand the context in the environment in which they're operating. Segmentation is an important stage of the image recognition system, as it extracts the objects of our interest, for further processing such as description or recognition. Segmentation techniques are used to isolate the desired object from the image in order to perform analysis of the object. The semantic segmentation plays vital role in automated vehicles and robotics. This helps the models to understand the context in their surrounding and where they are operating in order to have a safe operation. For example, let us say an automated car is operating in a crowded road. Therefore for it to operate in a safe mode it needs to understand the surroundings. The main goal for our model is to reach its destination or designated point keeping the surroundings and the environment in which it operates in safe. This project trains a type of Convolutional Neural Network Deeplab v3+ from an already pretrained network called resnet18 from deep learning. We have taken the image and label datasets from Cambridge university for training. These images are from the street level views which are captured while driving. This datasets provides 32 labelled classes which include car, pedestrian, building, sky,etc...



Working Process

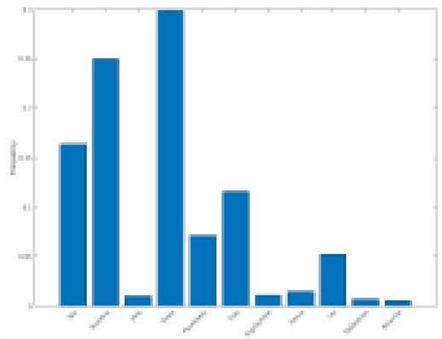
i) Import Network and Image Datasets

This project develops the Deeplab v3+ network with the weights which are initialized from a pre-trained Resnet-18 network. ResNet-18 is an efficient network that works well and efficient for applications with limited processing resources.

We need a labeled image database for the semantic segmentation with some classes. So, we have taken the Cambridge-driving Labeled Video Database (CamVid) is the first collection of videos with object class semantic labels, complete with metadata. The database provides ground truth labels that associate each pixel with one of 32 semantic classes

ii) Analyze Training Data for Segmentation

Preferably, all classes would have an equivalent number of observations. In any case, the classes in CamVid are imbalanced, which is a typical issue in automotive data-sets of street scenes. Such scenes have more sky, building, and street pixels than person on foot and bicyclist pixels since sky, structures and streets cover more territory in the picture. If not dealt with accuracy, this irregularity can be hindering to the learning cycle on the grounds that the learning is one-sided for the prevailing classes. Later on in this model, you will utilize class weighting to deal with this issue.



iii) Creating a Semantic Network

Now, we will be using the deeplabv3plusLayers function to create a DeepLab v3+ network based on ResNet-18.

As seen earlier, the classes in CamVid are not balanced as ideal ones. To improve the training, we can use class weighting to balance the classes. We use the pixel label counts computed earlier with countEachLabel function and calculate the median frequency class weights.

iv) Train a Semantic Network

The optimization algorithm which is used for training is called stochastic gradient descent with momentum (SGDM) in short form. The learning rate utilizes a piecewise plan. The learning rate is decreased by a factor of 0.3 each 10 epochs. This permits the organization to adapt rapidly with a higher beginning learning rate, while having the option to discover an answer near the neighborhood ideal once the learning rate drops. The organization is tried against the approval information each age by setting the 'ValidationData' boundary. The 'ValidationPatience' is set to 4 to quit preparing early when the approval precision unites. This keeps the organization from overfitting on the preparation dataset. A small group size of 8 is utilized to decrease memory utilization while preparing. Improving the network's accuracy by randomly transforming original data during training is done by the Data Augmentation. We will start training using the trainNetwork (Deep Learning Toolbox). If the doTraining flag is true, training will start. Otherwise, it will load a pretrained network which is resnet18 in our case.

v) Evaluate and Inspection of the results

After we start testing the network on an image we display the semantic segmented image with the pixel labeled image overlayed on the original image. We compare the results in C with the expected ground truth stored in pxdsTest. The green and magenta regions highlight areas where, the segmentation results differ from the expected ground truth.

Conclusion

We would like to conclude this project, by highlighting the points of the project. Starting of with Semantics Segmentation, this term refers to the process of linking each pixel in an image to a class label. This project uses semantics segmentation with deep lab v3+ network. Our projects explains how we use semantic segmentation on self driving cars and robotics as it is important for the models to understand the context of the environment in which they operate. Our goal here was to segment the image to label each pixel of an image, with a corresponding class of what is being represented. This project can help self diving car or automated vehicles understand the context of their surroundings and help in a safe operation. Wrapping all up we would like to say that this is how we can use semantic segmentation with deep lab v3+ and build models that perform the operation with safety.

Literature citations

[1] Chen, Liang-Chieh et al. "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation." ECCV (2018).

[2] Brostow, G. J., J. Fauqueur, and R. Cipolla. "Semantic object classes in video: A high-definition ground truth database." Pattern Recognition Letters. Vol. 30, Issue 2, 2009, pp 88-97.

[3] Segmentation and Recognition Using Structure from Motion Point Clouds, ECCV 2008,.

[4] Semantic Object Classes in Video: A High-Definition Ground Truth Database Pattern Recognition Letters.

[5] Mathworks Help center. *Semantic Segmentation using Deep Learning.* https://in.mathworks.com/help/vision/ug/semantic-segmentation-using-deep-learning.html