Name : T. Guru Pradeep Reddy

Reg.No : 14CO246

Mob.No : 8123289195

Email-id :gurupradeept@gmail.com

Fully Convolutional Networks for Semantic Segmentation

Evan Shelmar, Jonathan Long, Trevor Darrell.

Finalized Work:

- Any standard model like VGG, Alex Net, ResNet that is trained on ImageNet dataset for task of classification will be taken.
- Fully connected layers of the model will be replaced by appropriate convolutions, which help in reducing the parameters to be trained and also helps us in upsampling.
- The model is then upsampled with the help of deconvolution layers to get back to the original size and skip connections will be introduced to make the eliminate coarse segmentation.
- The model will be then trained for task of semantic segmentation on PASCAL VOC 2012 dataset and performance will be evaluated.
- No of classes that will be considered for the task of semantic segmentation will be based on computation feasibility.

Work Completed:

- PASCAL VOC 2012 segmentation class data has been used for training.
- The Images are preprocessed and labelled by using one hot encoding i.e classifying each pixel to one of the 21 categories.
- VGG net has been used to implement fully convolutional network.
- The model is already trained on the Imagenet for task of classification
- The classification layer of the model is discarded and all fully connected layers are converted to convolutions. Then 1*1 convolutions with channel dimension 21 were used to predict pascal classes, followed by a deconvolution layer to bilinearly upsample the coarse outputs to pixel-dense outputs.
- FCN-16s (Deconvolution layers with stride 16) has been implemented. It has one skip connection which is used to transfer the features of low level layers to up sampling layers.
- The model is then evaluated on sample images of the same dataset.

Libraries, Dataset and Implementation Approach :

- Keras library is used for implementation of the Deep learning model. Keras is a model-level library, providing high-level building blocks for developing deep learning models.
- Used Tensorflow as backend for keras, which implements all the low level functionality.
- In keras following layers have been used Input, Dropout, Permute, Add, add,Convolution2D, ZeroPadding2D, MaxPooling2D, Deconvolution2D, Cropping2D, Merge.
- Input layer is used to specify the input size, Deconvolutional layers have been used for upsampling. Max pooling layers are used to reduce image size and Merge layer along with Add layers have been used to implement the skip connections.
- Cropping and Zero padding layers are used to transform the image into required size for proper functioning of the model.
- Scio.io library is used to load the weight of VGG model from the matlab file.
- Matplotlib is used for displaying images and loading images.

- PASCAL VOC 2012 dataset has been used for training purpose. It has 2913 training images.
- The entire training set is divided into training, validation and testing sets. The ratio of 0.7:0.15:0.15 has been used for this purpose.
- Then the model is trained and validated accordingly, certain hyper parameters are adjusted based on the performance on the validation test.

.

Portion of Work remaining:

- As of now only FCN-16 has been implemented. FCN-32 and FCN-8 are yet to be implemented.
- No data augmentation techniques have been used now. They can be used to improve the size of the training data which gives the model more scope for better prediction.
- Hyper parameters for all the layers and the gradient have not been tuned properly due to lack of time. Those things have to be taken care of.
- Full dataset was not considered for training due to lack of resources and the amount of time taken for training.
  .

Novelty and Extra Information:

- The model can be trained with datasets like Microsoft COCO which has lot more labelled images, which can be used to improve the performance of the model.
- If time permits the like VGG the model will also be implemented with other ImageNet trained models like AlexNet and performance will be compared.
- The model doesn't use any pre or post processing techniques. Techniques like CRF (Conditional Random Fields) can be used to improve the accuracy. CRFs fall into the sequence modeling family. Whereas a discrete classifier predicts a label for a single sample without considering neighboring samples, a CRF can take context into account. By considering the neighbouring pixels into account CRF can be used as post processing technique to improve the performance of the model.