Name : T. Guru Pradeep Reddy

Reg.No : 14CO246

Mob.No : 8123289195

Email-id :gurupradeept@gmail.com

Fully Convolutional Networks for Semantic Segmentation

Evan Shelhamer, Jonathan Long, and Trevor Darrell

Essence:

- Instead of segmenting images in form of bounding boxes, with the help of fully convolution networks we can segment image at pixel level, i.e we give label to each and every pixel which is called semantic segmentation.
- Doesn't use any fully connected layers making it more robust and also making it possible for the model to take the input of arbitrary size and produce correspondingly sized outputs with efficient interference.
- This method is simple and efficient both asymptotically and absolutely. It uses a simple model and eliminates all complicated design which was used earlier. No pre and post processing techniques are needed for this method.
- It uses a network model which was trained on ImageNet and converts it to fully convolutional network, and then the model will be trained on segmentation for fine tuning to get the final network thereby using the power of transfer learning.
- In addition to that it introduced a notion of skip architecture which adds skip connections between the layers to make the segmentation smooth. They help to preserve the spatial context in a better way. It uses deconvolutional layers and unpooling layers for up sampling.

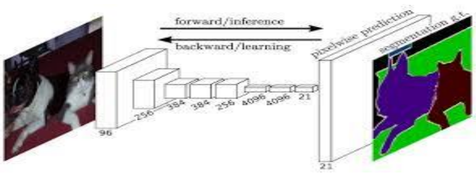
Work Described:

- Other approaches for semantic segmentation such as patch wise classification and their drawbacks. Motivation for introduction of the fully convolutional network architecture and extension to FCNS.
- Design of fully convolutional networks and it involves description the usage and importance deconvolution layers and unpooling layers. Importance of skip layers and their effect on the performance of the model.
- Application of paradigm of transfer learning to get better results and the performance of model on various datasets such as PASCALVOC2011-2,NYUDv2, SIFT Flow, and PASCAL-Context and analysed various design choices and stated the areas where Fully convolutional layers can be used.

Importance:

- It popularized the use of end to end convolution networks for semantic segmentation. This work has been extended to new tasks such as contour detection, optical flow and weakly supervised semantic segmentation. It inspired the U-Net model which is used for pixel labelling of microscopy images. This achieved a very high performance both in terms of accuracy and time. Various models which have used CRFs along with this model like DeepLab and got improved results. All in all it considered both global and local information into consideration which are important for predicting the location and classification of image.

Methodology:



The entire work can be divided into three parts. First part is application of transfer learning paradigm, second step is to tune the model to perform segmentation task and the third part is to introduce the skip connections to improve the segmentation performance. For first part we take a ImageNet pre trained model and we remove all the fully connected layers of that model and we replace them with convolutional layers to perform the same task which ensures that model is fully convolutional now. After that the feature maps still need to be upsampled because of the pooling operations of in CNNs. Instead of using simple bilinear interpolation they used deconvolution layers to learn the interpolation. However, upsampling(even with deconvolutional layers) produces a coarse segmentation maps because of the loss of information during pooling. To overcome this problem they introduced shortcut/skip connection from high resolution feature maps. They help us to preserve semantic information.

Pros:

- It is a simple and scalable implementation which is efficient both asymptotically and absolutely, unlike the previous methods it doesn't use any fully connected layers which are expensive and it can be trained on any arbitrary sized input. It doesn't use any pre-processing and post processing techniques which tend to make the model complicated and add additional overhead for computation.
- Usage of paradigm of transfer learning makes the model learn the features in a better way. Usage of deconvolutional layers helps the model learn interpolation properly, and the introduction of skip connections improved the coarseness of upsampling. This also helps in transferring the information captured in the initial layers and was required for re construction during the upsampling. So the information that we had in the primary layers can be fed explicitly to the later layers using the skip architecture. It achieved an accuracy of 30 percent better than the models which existed at that time in segmentation task.

Cons:

- As it is difficult to train FCN from scratch they use a model trained on ImageNet, the model is pre trained with low resolution images whereas the input segmentation images will be of high resolution, this techniques doesn't handle this domain gap, making this model less optimised. The feature maps that are used for classification in FCN have limited contextual fields, so predictions are inconsistent for local ambiguous regions. Down sampling will reduce the performance on small sized objects.
- Pooling helps in classification networks because receptive field increases. But this is not the best thing to do for segmentation because pooling decreases the resolution. Even though skip connections are there, they can't handle the significant variance of object scales across different images. So still semantic information is lost despite of skip connections. And also max pooling indices are not transferred from during the feature extraction stage.

Future Work:

- As of now fully convolutional layered network is tested only on natural images, we can check what are the other tasks which can be accomplished with this model like to find contours and segmentation of medical images e.t.c.
- To in-corporate post processing models such as random fields to improve the performance and also weakly supervised training to reduce the cost of labelling every pixel.

Work that will be demonstrated:

- Will take any ImageNet trained model and it will be tuned to perform the task of semantic segmentation on PASCAL VOC dataset using Fully convolutional network architecture proposed in the paper. The no of categories which will be considered in the PASCALVOC dataset will depend on the computation feasibility.

Other Observations:

- Even with the help of skip connection some amount of semantic information is lost due to convolutions and max pooling. Techniques like transferring max pooling indices to corresponding un pooling layers will help in preserving the semantic information to certain extent. An alternative to this techniques like Dilated convolutions can be used to solve this problem
- No pre-processing or post processing techniques are used here. Conditional random Markov fields(CRFs) can be combined along with fully convolutional networks to achieve better performance in semantic segmentation.