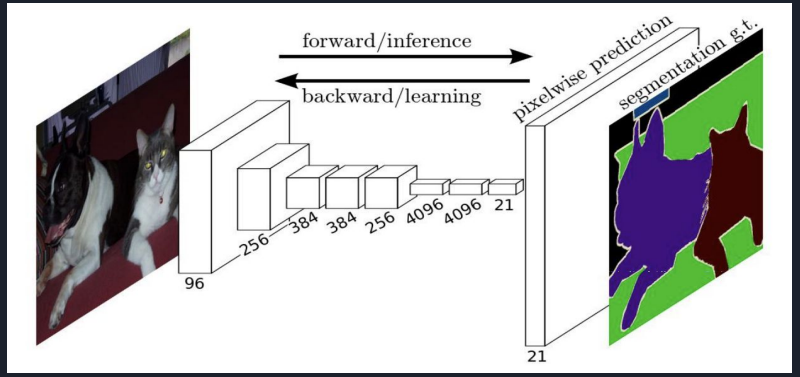
Semantic Segmentation

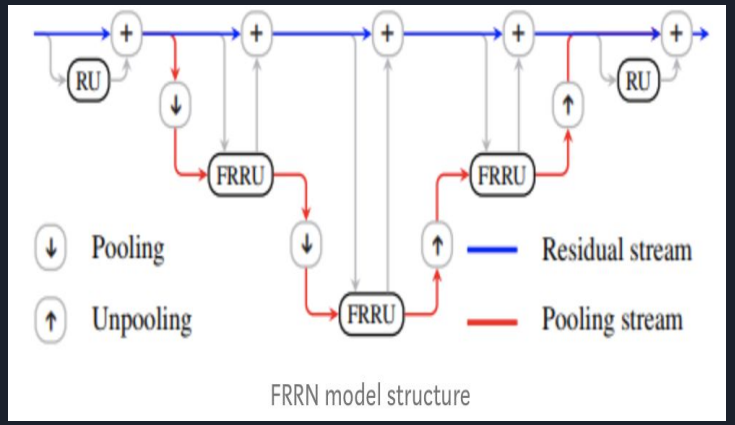
Satish Kumar Anbalagan, Divyang Teotia,
Varun Sahasrabudhe, Daniel Uvaydov

Semantic Segmentation: Overview

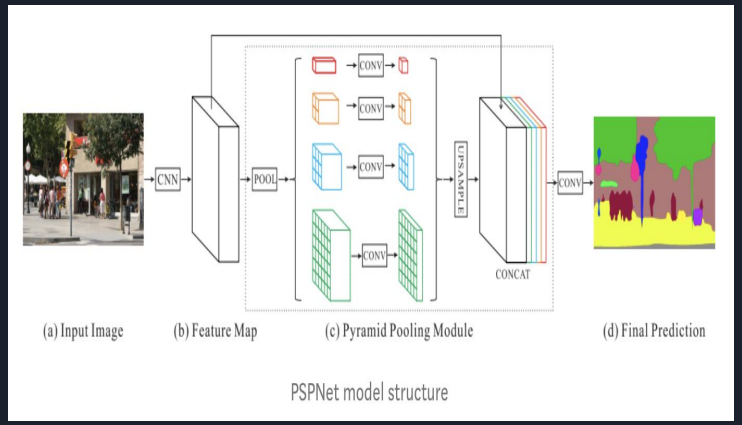
- What ? Classify and clustering each and every pixel in the image which belong to same object class
- How ? Includes segmenting by a feature extraction network trained for image classification like VGGNet, ResNets, DenseNets, MobileNets, NASNets etc
- Cityscapes data sets are used
- Why ? autonomous driving, medical, HCI, photo editing tools, robotics vision and understanding



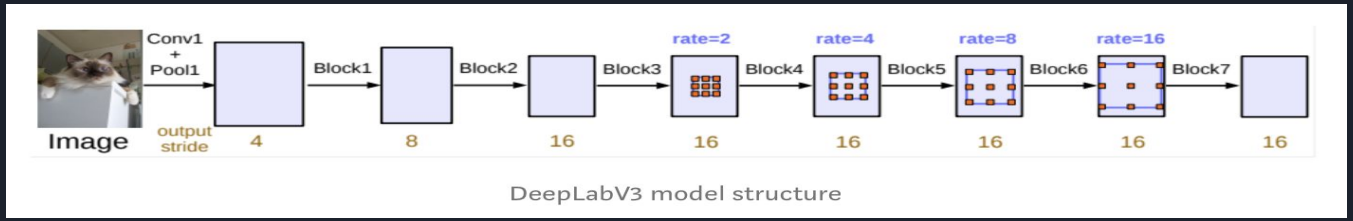
Related Work



FRRN model structure



PSPNet model structure



DeepLabV3 model structure

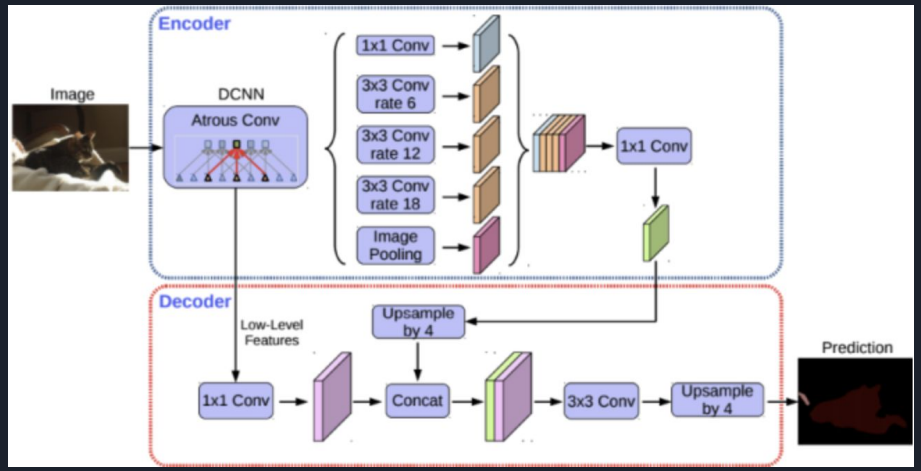
Challenges

- Tradeoffs between accuracy vs speed per memory, while maintaining the efficiency of the network during classification
- Network model encounters objects of many different sizes that require features processing at different scales
- Improving localization of object boundaries
- Segmenting and existence of objects at multiple scales
- Reduced feature resolution caused by a repeated combination of max-pooling and downsampling
- Better refining of feature map using Channel attention

DeepLabs v3

Encoder-Decoder Architecture:

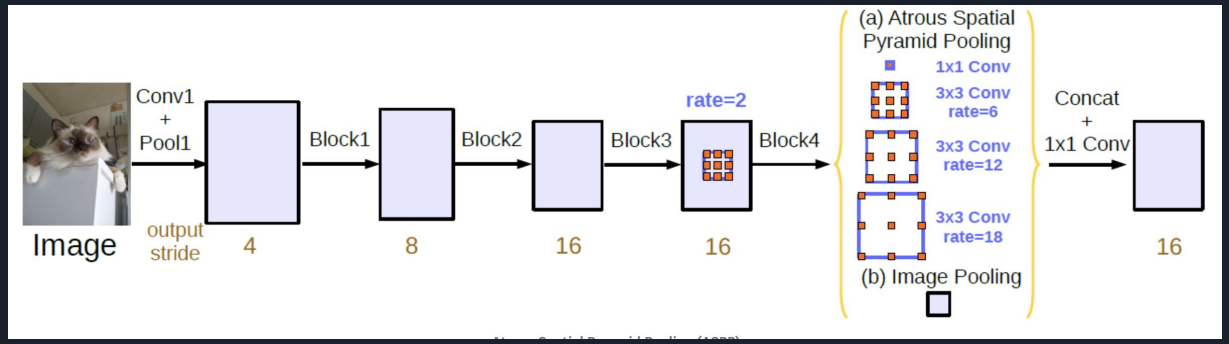
- Deeplabs prevents signal decimation and learns multi scale contextual features



- Uses an ImageNet pretrained Resnet as its main feature extractor with atrous conv in the last block

- Uses Atrous Spatial Pyramid Pooling (ASPP) on top of Resnet to classify regions of an arbitrary scale and decoder upsamples the output in stages

ASPP



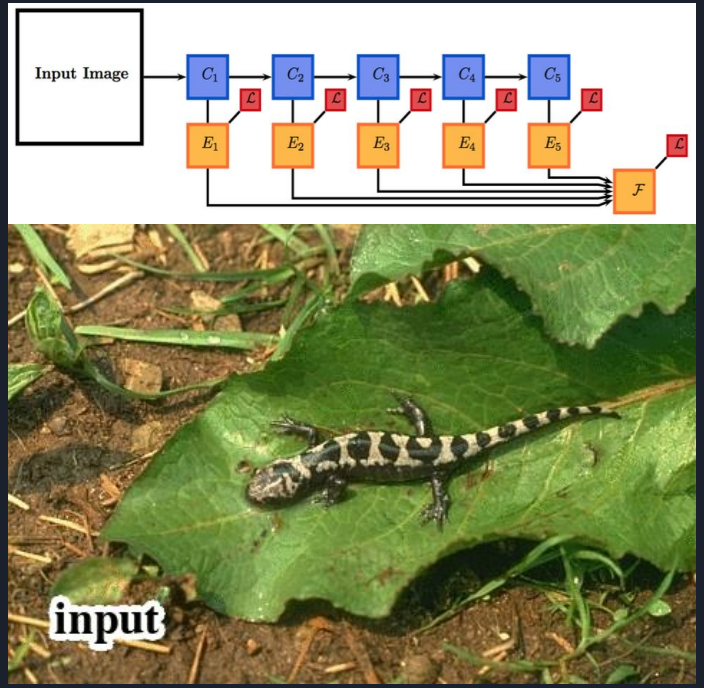
- Provides the model with multi scale information using a series of atrous convolutions with different dilation rates to capture long range context.
- To add global context information, ASPP incorporates image level features via Global Average Pooling
- Finally, all the multiple scales are concatenated along with global features and followed by a 1x1 convolution to feed to the decoder.

Holistic Edge Detection

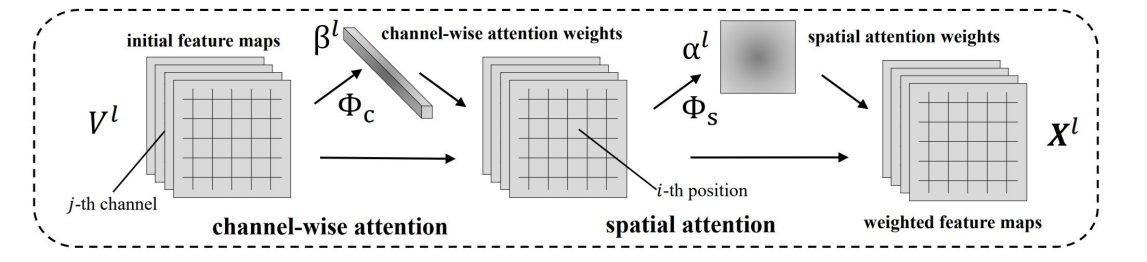
- Uses Deep Supervised Network training to fine tune VGG for the task of boundary detection

- We will supplement our input with an extra channel using the output of a pre-trained HED

- Add a long skip connection from input to decoder and reiterate boundary information right before decoder output.



Channel Attention



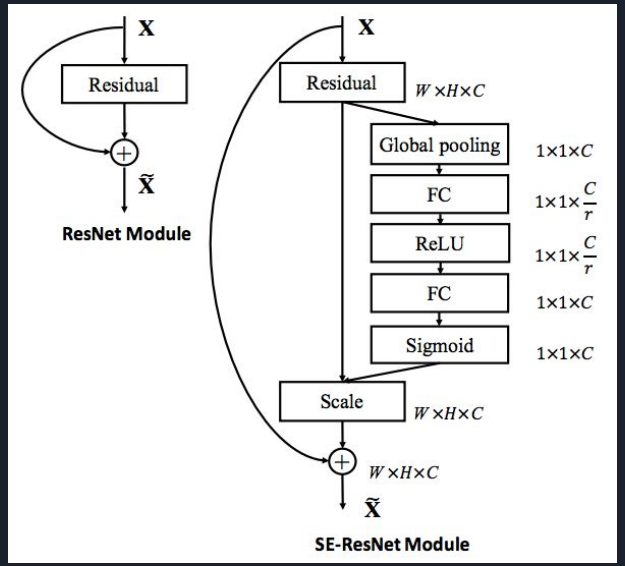
- Channels in the feature maps of a layer are all traditionally **treated or weighted equally**

- Channel attention attempts to give a hierarchy to the channels within a feature map by multiplying weights for each channel to scale them adaptively

- This is ideal in image processing applications as different channels may have more prominent features for a classification than others

Squeeze and Excitation Blocks

- Squeeze and excitation blocks attempt to map the channel interdependencies
- Squeezes all feature maps to single values (per channel), extracts channel features through FC layers, and weights each channel value
- Original feature map is then scaled with weighted channel values that are continuous



Our Contribution

- Add SE to DeepLabs Resnet
  - We add a content aware mechanism to weight each channel adaptively. Thus, providing us with the same accuracy as ResNet 101.
  - Test softmax in replacement of sigmoid for last layer of SE block to establish stricter channel hierarchy

- Creates channel attention which hasn't been popular in semantic segmentation
  - The attention maps will amplify the relevant region, thus demonstrating super generalisation over several dataset (here, cityscapes dataset).
  - Attention allows the network to focus on the most relevant features without additional supervision, avoiding the use of multiple similar feature maps and highlighting salient features that are useful for a given task

- Enhancing segmentation task using end-to end Holistic Edge Detection(HED) technique

Future Steps

- Implementation of Channel Attention Module, Position Attention Module, Deep Attention Module to further improve the accuracy of the model.

- Implementation of Boundary Detection with the help of Encoder-Decoder architecture

References

- https://heartbeat.fritz.ai/a-2019-guide-to-semantic-segmentation-ca8242f5a7fc
- https://towardsdatascience.com/semantic-segmentation-with-deep-learning-a-guide-and-code-e52fc8958823
- https://towardsdatascience.com/squeeze-and-excitation-networks-9ef5e71eacd7
- Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
- Chen, Long, et al. "Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- https://www.analyticsvidhya.com/blog/2019/02/tutorial-semantic-segmentation-google-deeplab/
- Xie, Saining, and Zhuowen Tu. "Holistically-nested edge detection." Proceedings of the IEEE international conference on computer vision. 2015.