

SegNet

SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

IPAMI 2015

4255+ citations

- Previous methods like FCN, CRF etc., tried to directly adopt deep architectures designed for category prediction(classification) to pixel-wise labelling, these approaches work but produce coarse results because the deep architectures have max-pooling layer due to which we reduce resolution and also spatial information.
- The key component in SegNet is the deconvolution network which use **Unpooling** using **indices produced during max-pooling**
- Advantages of using max-pool indices:
 - improves boundary delineation.
 - reduces the number of parameters enabling end-to-end training.
 - upsampling using pool-indices can be incorporated into any encoder-decoder architectures with little modification.

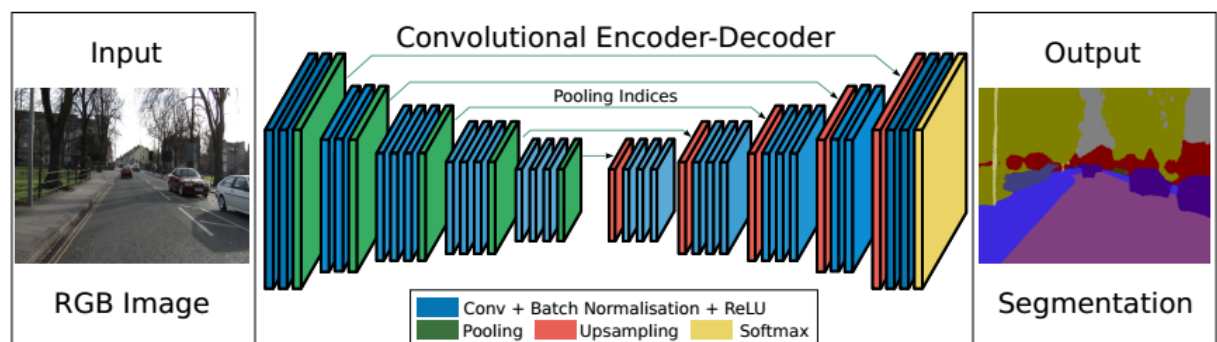


Fig. 2. An illustration of the SegNet architecture. There are no fully connected layers and hence it is only convolutional. A decoder upsamples its input using the transferred pool indices from its encoder to produce a sparse feature map(s). It then performs convolution with a trainable filter bank to densify the feature map. The final decoder output feature maps are fed to a soft-max classifier for pixel-wise classification.