PSP-Net

Why we choose PSP-Net?

PSP-Net is used to aggregate different levels of features. There are already 2 different levels. So why we need this structure?

Actually the real receptive field is far more less than the theoretical one. So our network may not be able to learn enough global information from the image.

The PSP-Net can solve this problem.

Here is how the Pyramid pooling module works.

There are 4 size of pooling. The red one is global average pooling. It simply get the average of all the values in one dimension. The next 3 pooling levels are to separate the feature map into different parts and calculate the average for each parts. For example, the second level is to separate the feature map into 4 parts and got 4 averages.

So the result of the poolings contain the global information. And we use a 1-by-1 convolutional kernel to reduce the dimension to 1. Then we do up-sampling so that the final result are the same size as the feature map. Finally the fusion layer combine them together.

Our net:

Our net may look like these. All the things is the same as before except the last layer. Before the last layer, we add a Pyramid pooling module. So the Decoder here its output has 32 dimension. Then we calculate these layers and add them to the 32-dimension result. Then we do the last convolution to get our result.