Spatial Pyramid Pooling

Badrinath Singhal

April 2021

Pooling layers are generally used to down sample feature maps by summarizing features in patches. Common pooling methods are average pooling and max pooling. Pooling down samples feature maps considering all the neighbouring features so that the down sample features contains the relation of neighbouring features [1]. I recently read about Spatial Pyramid Pooling and I will try to put a short summary here.

Spatial Pyramid Pooling

This pooling strategy is used to eliminate the requirement of fixed input image size. The idea proposed here [2]. In general CNN can take input of arbitrary size and and produce outputs of variable sizes. But for classifiers or fully connected layer, we need fixed length vectors. Here spatial pyramid pooling comes into picture, it maintains spatial information by pooling in local bins. To maintain fix vector size at the output the spatial bins sizes are proportional to the image size which makes number of bins fixed regardless of image size.

In spatial pyramid pooling layer, response of each filter are pooled. The output are kM dimensional vector where M is denoted as no. of bins and k is the number of filters. Below is figure which demonstrates spatial pyramid pooling.

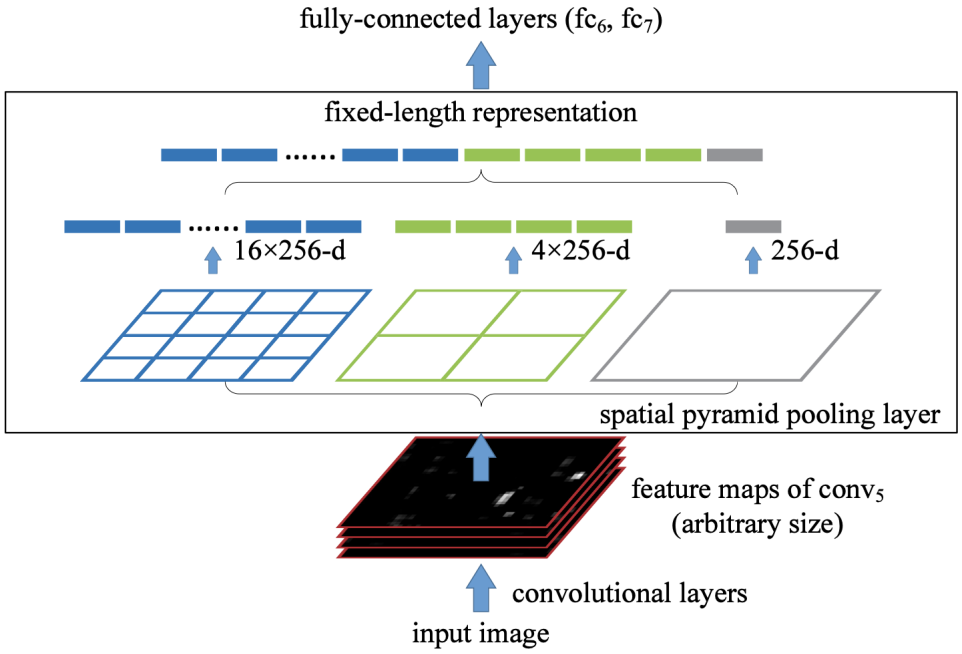

Figure 1: Spatial Pyramid Pooling Layer

End

This document is a part of making a short notes whenever I come across something interesting so I can retain it for longer period of time.

References

[1] A Gentle Introduction to Pooling Layers for Convolutional Neural Networks

[2] Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition