Formatting instructions for NIPS 2016

Anirudhan J. Rajagopalan
Department of Computer Science
New York University
New York, NY.
`ajr619@nyu.edu`

1 Variants of Pooling

Pooling is an operation in Convolutional Neural Networks which is performed after a convolutional layer. We decide on a size of the region to do pool our convolutional filters over. Then we divide the convolved features into disjoint regions and take the max, mean or norm of the feature activations depending on the type of pooling operation. Generally the stride length used in pooling is equal to the pooling window size.

1.1 Three types of pooling

There are three types of pooling operations that are used

Max pooling Max pooling takes the maximum of the feature activations. Implemented by SpatialMaxPooling.

Average pooling Average pooling takes the average of the feature activations. Implemented by SpatialAveragePooling.

Lp pooling LpPooling takes the l2 norm of the feature activations. Implemented by SpatialLPPooling.

1.2 Mathematical formulas for pooling

Suppose we have a NxN convolutional layer (filters) and a pooling layer of size mxm where $m < N$ The pooling operation is generally implemented using a stride of m. Therefore the size of filters after pooling will be $\frac{N}{m} * \frac{N}{m}$ and each of the mxm pooling operations gives a single value. The single value for a pooling operation can be represented by

$$y = \sigma(x_{ij})$$

where $1 \leq i \leq m$ and $1 \leq j \leq m$.

1.2.1 MaxPooling

Incase of max pooling

$$\sigma = \max(x_{ij}) \qquad 1 \leq i \leq m and 1 \leq j \leq m$$

1.2.2 Average Pooling

Incase of average pooling

$$\sigma = \frac{\sum_{i=1}^{m} \sum_{j=1}^{m} x_{ij}}{m * m}$$

1.2.3   LPPooling

Incase of LP pooling where $p \geq 2$.

$$\sigma = (\sum_{i=1}^{m} \sum_{j=1}^{m} x_{ij}^{p})^{\frac{1}{p}}$$

1.3   Reason for using in deep learning

Pooling is used for a number of reasons.

1. Computational tractability.   The number of features after convolution increases by $numFIltersx(N-m+1)x(N-m+1)$. Generally it will be computationally intractable to perform learning on this huge data.

2. Overfitting. Since the number of features are high, this will lead to overfitting. Pooling helps in reducing the number of features.

3. Stationary property. Generally images have stationary property. This means that features that are useful in one region will be useful in other regions also. Pooling operation helps in aggregating the features at various locations.

We generally use Max-pooling for images as the performance of Max pooling has been emperically shown to be better than Average or Lp pooling for image workflows. This can be explained by the stationary property of the image as the most important activation in a small patch can be found by using max pooling.

2   Convolution

2.1   Number of values after forward propogation

When we apply a convolution on a $N \times N$ image with a kerenel of size $m \times m$ we get $(N - m + 1) \times (N - m + 1)$ output values after the convolution.

Here, N = 5, m = 3. Therefore we will get $(5 - 3 + 1) \times (5 - 3 + 1) = 9$ values.

2.2   Values after forward propogation

The resulting filter obtained by applying the convolutional kernel is:

109	92	72
108	85	74
110	74	79