Week 3 Reading Reflection

Hao Deng

October 2023

1 Introduction

In the article this week, A Comprehensive Introduction to Different Types of Convolutions in Deep Learning, the author explained different types of convolution in a concise and lucid way. The intelligible illustrations of convolution make the concepts easy to understand. I don't have any problem understanding this article, and therefore I would take this reflection as a reading note.

2 Notes

2.1 Convolution v.s. Cross-correlation

The convolution of two function f and g is defined as, where f is the input and g is the filter:

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$$

In the convolution, g is reversed as we are integrating $g(t-\tau)$. However, in deep learning, g is not reversed. Rigorously speaking, it's cross-correlation instead of Convolution.

2.2 Convolution in Deep Learning

Convolution is used to extract useful features from the input in Deep Learning.

2.3 Kernel and Filter

Kernel refers to a 2D array of weights while filter is 3D structures of multiple kernels stacked together.

2.4 Transposed Convolution

One can use transposed convolution to perform up-sampling, like generating high-resolution images, mapping low dimensional feature map to high dimensional sapce, etc.

Checkerboard artifacts result from "uneven overlap" of transposed convolution.

Uneven overlap exists when the filter size is not divisible by the stride.

Unevenly overlapped region tends to be more extreme in two dimensions, where two patterns are multiplied together and the unevenness get squared consequently.

2.5 Dilated Convolution

Dilated convolutions "inflate" the kernel by inserting spaces between the kernel elements.

Dilated Convolution enables "observing" a large receptive field without adding additional costs.

2.6 Separable Convolution

There are two types of separable convolution, spatially separable convolution and depthwise separable convolution.

Spatially separable convolution divides the original kernel to two smaller kernel. Compared with convolution, it requires less parameters and operations. However, it is rarely used because not all kernels can be divided into two smaller kernels.

Depthwise separable convolution consists of two steps: depthwise convolution and 1×1 convolutions. The advantage of this method is efficiency. The disadvantage is that the model capacity may deteriorate, making the model sub-optimal.

2.7 Flattened Convolution

Intuitively, the idea is to apply filter separation. Instead of applying one standard convolution filter to map the input layer to an output layer, we separate this standard filter into 3 1D filters.

2.8 Grouped Convolution

Separate filters into different groups and each group is responsible for a certain depth of the output.

It has following three advantages:

1. Allow parallelization.
2. The number of model parameter decreases as the number of filter group increases.
3. May generate better model.

2.9 Shuffled Grouped Convolution

One drawback of grouped convolution is that it only handles information passed down from the fixed portion in the previous layers.

Shuffled Grouped Convolution divides channels in different filter groups and shuffle them before applying next convolution.

2.10 Pointwise Grouped Convolution

Pointwise Grouped Convolution is using grouped convolution for 1×1 convolution. It can reduce computation cost while maintaining accuracy.