Contents

1 Convolutions

The idea is that having an operator with "localized patterns to have an effect on the output regardless of their location in the image". This feature is called `translation invariant`. Convolutions have three features:

1. Local
2. Translation-invariant
3. Linear

Each convolution layer has `kernel` which consists of weights."The output image is created by translating the kernel on all input locations and performing the weighted sum. Same kernel, and thus each weight in the kernel, is reused across the whole image."
In Opposite to fully connected layers, the number of parameters is not dependent on the input size. Two factors determine the parameters:

1. Size of the convolution kernel
2. Number of convolution filters

1.1 Padding

In convolution layers, due to the idea of kernel, output image is smaller than the input. In order to solve this problem, we use padding which means "creating ghost pixels around the border that have value zero as far as the convolution is concerned". Two main reasons for padding are:

1. Not worrying about image size
2. Use-full in structure such as U-net (skip connections)

2 Down Sampling

One way to reduce the size of an image is down sampling. In this way, for reducing the image to $\frac{1}{n}$ size of real image, we need to consider each non-overlapping neighbourhood of $2n$ pixels and compute the output value with:

1. Average (average pooling)
2. Maximum (max pooling)
3. Stride Conv

2.1 Receptive Field

Combining max pooling layer and kernel idea increase the receptive field. This is the reason why we set the smaller kernel size (leads to consider larger portion of the input image).