Implementation of a CNN Layer

A Convolutional Neural Network (CNN) is a neural network typically used for object recognition in images. A convolutional layer is the part of this neural network in which a convolution operation between a matrix and a filter is performed. In particular, a convolutional layer is composed of:

- 1 input channel C_{in}
- N output channels $C_{out}(i)$. Each output channel is generated using an associated filter $f_i \in \mathcal{R}^{w \times h}$.

Each element of $C_{out}(i)$ is calculated with the convolution between $w \times h$ elements of the input matrix and the $w \times h$ elements of the associated filter f_i:

$$C_{out}(i)(x,y) = \sum_{j=0,k=0}^{j=w,k=h} f_i(j,k) \cdot C_{in}(x_i + j, y_i + k)$$

Where x_i and y_i are offset in the input matrix which depend on the element of the output matrix to calculate.

In particular, as described in the example of Figure 1, the dimension of the output matrix is lower than the dimension of the input one. The size reduction depends on the size of the filter. The filter f_i, whose elements are indicated in red, is overlapped with a portion (in yellow) of the input matrix (in green) and multiplied with the corresponding elements. The sum of these multiplication is the output. To calculate each element of the output matrix, the operation is repeated by moving the filter (like a moving window in 1D digital filter) along x and y dimensions.

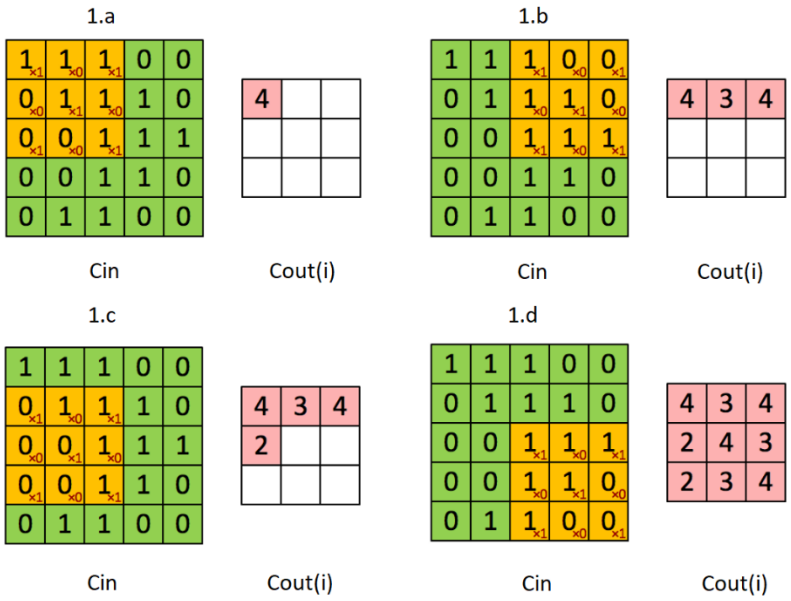

Figure 1

It is required to design a digital circuit for implementing a convolutional layer which calculate a single output channel, given an input NxN matrix and the MxM filter matrix. Use VHDL generic whenever is possible. The interface of the circuit to be designed is as follows:

Input image and filter must be provided with *x* input signal and are both 8-bit wide. *i/f* bit is used to indicate to the circuit if the data on *x* is a pixel of the image, or an element of the filter. Inputs are considered valid and evaluated only when the input *x_valid* signal is 1, otherwise the circuit state is retained. The correct procedure to use the circuit is to load the filter first (*i/f* = 1), and then process the input image provided pixel by pixel. Output values must be considered valid only for one cycle (*y_valid* = 1). Input pixels can be provided in any order you consider convenient.

You are requested to deal with the various possible error situations, documenting the choices made. In particular, it is necessary to take into consideration:

- Unexpected changes of *i/f* signal
- Partial loading of filter elements

The final project report must contain:

- Introduction (circuit description, possible applications, possible architectures, etc.)
- Description of the architecture designed (block diagram, inputs/outputs, etc.)
- VHDL code (with detailed comments) to be attached to the report.
- Test strategy (Test-plan) and related Testbench for verification; a detailed, though not exhaustive, verification is required, including error situations and borderline cases of functioning
- Interpretation of the results obtained in the automatic synthesis/implementation on a Xilinx FPGA platform in terms of maximum clock frequency (critical path), elements used (slice, LUT, etc.) and estimated power consumption. Comment on any warning messages.
- Conclusions