Image Stream Processing on a Packet-Switched Discrete-Time CNN

Suleyman Malki[a], Lambert Spaanenburg[a] and Nick Ray[b]

[a] LTH/IT, P.O.Box 118, 22100 Lund (Sweden)
[b] Interay B.V., Mr. W.M. Oppedijk van Veenweg 8, 9251 GA Burgum

The full version of this paper appeared as: S. Malki, L. Spaanenburg and N. Ray. *Neural vision sensors for surface defect detection*. Proceedings IJCNN'04, Vol. IV, Budapest, pp. 3155 – 3160, 2004.

1. Introduction

Cellular Neural Networks (CNNs) are widely used for real-time image processing applications. Though the Cellular Neural Network as a concept is characterized by a strict locality of operation, the large-scale digital implementation has been far from trivial. The intense interaction of a CNN node with all others within a specified neighbourhood poses severe interconnection requirements. Already 8 input and output values need to be communicated for the minimal 1-neighbourhood. The attached problem is the need for massively parallel and accumulated multiplications to implement the basic cell functionality.

The earliest CNN realization was targeted on exploiting the benefits of analogue circuitry to handle high communication bandwidth requirements as well as dense footprint multiplication. Added to the analogue core is a digital framework to universally integrate the CNN machine in a programmable environment. Today this technology style has created an impressive 128 by 128 capacity [1].

In a first digital realization, Szolgay discusses the use of the 2nd generation Field-Programmable Gate-Array (FPGA). From an analysis on the mapping of the major arithmetic blocks on the FPGA function blocks, he concludes that a further increase in packing density can be achieved in future generations [2]. This is confirmed by ILVA, where the memory and multiplier macro facilities on the Virtex-II are exploited [3]. Such leaves the impression that even more can be gained by moving the local broadcast of data from circuit switching to packet switching.

2. The Effect of a Network-on-Chip

The operation of a DT-CNN on images covers many dimensions. The local operation is performed in a two-dimensional plane (width & length) and iterates in time. Due to the limited capacity of the CNN implementation this has to be repeated over image slices and iterates over the surface to handle potential wave propagation. Finally the operation is performed on sequences of images. All this has to be facilitated on the two-dimensions in a Field-Programmable Gate-Array. Consequently, the dominating architectural question is: how to reduce the dimensions from the functional requirements to the platform facilities?

The key issue seems to be whether access to image information stored off-chip can be kept outside the inner loops of the computation. This is clearly exemplified in the original ILVA architecture [3], where the computation is unrolled on the nodal iteration dimension at the expense of the on-chip image salvage. The consequence is that image stream manipulations will involve a bandwidth problem with respect to the external image RAM.

The principle of broadcasting processing elements, loosely coupled through a packet switching network retains the potential of image stream handling. Of course, in the present generation of FPGAs, the amount of distributed memory seems large enough to store a number of images. The new designs provide a similar high speed of 500 Mpixels per second as the original ILVA design after detailed optimization, but the capacity has been raised from a pipeline of 20 nodes in a line to a full parallel network of 128 neurons. The designs are developed using ISE, ModelSim, and Synplify; they are targeted for the Virtex-II Pro P30 on a Memec FF1152 development board.

References

[1] G. Linan et al. “ACE16K: A 128 x 128 focal plane analog processor with digital I/O”. In R. Tetzlaff, editor, *Proceedings 7th IEEE Workshop on CNNs and their Applications*, pp. 132-139, 2002.

[2] Z. Nagy and P. Szolgay. Configurable Multi-Layer CNN-UM Emulator on FPGA. In R. Tetzlaff, editor, *Proceedings 7th IEEE Workshop on CNNs and their Applications*, pp. 164 – 171, 2002.

[3] S. Malki, and L. Spaanenburg. *CNN Image Processing on a Xilinx Virtex-II 6000*. Proceedings ECCTD’03 (Krakow) pp. 261-264, 2003.