

CNP: An FPGA-based Processor for Convolutional Networks

Clément Farabet¹, Cyril Poulet¹, Jefferson Y. Han², and Yann LeCun¹

¹ Courant Institute of Mathematical Sciences, New York University

² Perceptive Pixel Inc.

<http://www.cs.nyu.edu/~yann>

Over the next decades, embedded vision systems will become part of an increasing number of applications that include autonomous or semi-autonomous vehicles, consumer electronics (cameras and mobile phones), and toys. Because of their limited payload capacity and energy capacity, toys, micro-UAVs, and small domestic robots cannot use active imaging sensors and must rely on passive cameras and vision for navigation, obstacle avoidance, and object/target recognition. We describe an implementation of a complete vision/recognition system on a single low-end Field-Programmable Gate Array (FPGA), which can be the basis of low power, lightweight, and low cost vision systems for the above applications.

The design requires no external hardware, other than a few memory chips, and can be integrated onto a small 5×5 cm printed circuit board. The system is programmable, and can implement any vision system in which the bulk of the computation is spent on convolutions with small-size kernels. The design is specifically geared towards Convolutional Networks [5, 6], but can be used for many similar architectures based on local filter banks and classifiers, such as HMAX [11, 8], and HoG methods [2].

Convolutional Networks (ConvNets) are feed-forward architectures composed of multiple layers of convolutional filters, interspersed with point-wise non-linear functions [5, 6]. ConvNets are used in several commercial and experimental applications, including video surveillance, OCR [6], face/person detection [3, 9], object recognition [10], and robot navigation [7, 4]. Because they can easily be trained for a wide variety of tasks, ConvNets have many potential applications in micro-robots and other embedded vision systems that require low cost and high-speed implementations.

Pre-trained ConvNets are algorithmically simple, with low requirements for memory bandwidth and arithmetic precision. Several hardware implementations of ConvNets have been proposed in the past, including on FPGAs [1], but fitting it into the limited-capacity FPGAs of the time required the use of extremely low-accuracy arithmetic. Modern FPGAs have considerably higher capacity as well as

large numbers of hard-wired multiply-accumulate units.

the system described here is a *programmable ConvNet processor*, which can be thought of as a RISC (Reduced Instruction Set Computer) processor, with an instruction set that matches the elementary operations of a ConvNet. These elementary operations are highly optimized, and make extensively use of the parallelism inherent in the hardware. Implementing a particular ConvNet simply consists in *compiling* a set of instructions for this RISC processor. This provides a high degree of flexibility. The entire system uses a single low-end Xilinx Spartan-3A DSP 3400 FPGA with an external DRAM module, and no extra parts. A ConvNet face detection system was implemented and tested. The face detector runs at 4 frames per second at 512×384 resolution in the Xilins development board. The speed is entirely limited by the narrow memory bandwidth available on the development board. Performance with the custom board being designed is expected to reach 20 frames per second, corresponding to an average performance of about 8×10^9 connections per second.

References

- [1] J. Cloutier, E. Cosatto, S. Pigeon, F. Boyer, and P. Y. Simard. Vip: An fpga-based processor for image processing and neural networks. In *Fifth International Conference on Microelectronics for Neural Networks and Fuzzy Systems (MicroNeuro'96)*, pages 330–336, Lausanne, Switzerland, 1996. 1
- [2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. of Computer Vision and Pattern Recognition*, 2005. 1
- [3] C. Garcia and M. Delakis. Convolutional face finder: A neural architecture for fast and robust face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11):1408–1423, 2004. 1
- [4] R. Hadsell, A. Erkan, P. Sermanet, J. Ben, K. Kavukcuoglu, U. Muller, and Y. LeCun. A multi-range vision strategy for autonomous offroad navigation. In *Proc. Robotics and Applications (RA'07)*, 2007. 1
- [5] Y. LeCun. Generalization and network design strategies. In R. Pfeifer, Z. Schreter, F. Fogelman, and L. Steels, editors,

Connectionism in Perspective, Zurich, Switzerland, 1989. Elsevier. an extended version was published as a technical report of the University of Toronto. [1](#)

- [6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, November 1998. [1](#)
- [7] Y. LeCun, U. Muller, J. Ben, E. Cosatto, and B. Flepp. Off-road obstacle avoidance through end-to-end learning. In *Advances in Neural Information Processing Systems (NIPS 2005)*. MIT Press, 2005. [1](#)
- [8] J. Mutch and D. Lowe. Multiclass object recognition with sparse, localized features. In *CVPR*, 2006. [1](#)
- [9] M. Osadchy, Y. LeCun, and M. Miller. Synergistic face detection and pose estimation with energy-based models. *Journal of Machine Learning Research*, 8:1197–1215, May 2007. [1](#)
- [10] M. Ranzato, F. Huang, Y. Boureau, and Y. LeCun. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In *Proc. Computer Vision and Pattern Recognition Conference (CVPR'07)*. IEEE Press, 2007. [1](#)
- [11] T. Serre, L. Wolf, and T. Poggio. Object recognition with features inspired by visual cortex. In *CVPR*, 2005. [1](#)