A compact CMOS face detection architecture based on shunting inhibitory convolutional neural networks

XiaoXiao Zhang  
*Hong Kong University of Science & Technology*

Amine Bermak  
*Hong Kong University of Science & Technology*

Farid Boussaid  
*University of Western Australia*

Abdesselam Bouzerdoum  
*University of Wollongong, bouzer@uow.edu.au*

Follow this and additional works at: [https://ro.uow.edu.au/infopapers](https://ro.uow.edu.au/infopapers)

Part of the Physical Sciences and Mathematics Commons

**Recommended Citation**

Zhang, XiaoXiao; Bermak, Amine; Boussaid, Farid; and Bouzerdoum, Abdesselam: A compact CMOS face detection architecture based on shunting inhibitory convolutional neural networks 2008.  

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au
A compact CMOS face detection architecture based on shunting inhibitory convolutional neural networks

Abstract
In this paper, we present a compact, low cost, real-time CMOS hardware architecture for face detection. The proposed architecture is based on a VLSI-friendly implementation of Shunting Inhibitory Convolutional Neural Networks (SICoNN). Reported experimental results show that the proposed architecture can detect faces with 93% detection accuracy at 5% false alarm rate. A VLSI Systolic architecture was considered to further optimize the design in terms of execution speed, power dissipation and area. Potential applications of the proposed face detection hardware include consumer electronics, security, monitoring and head-counting.

Disciplines
Physical Sciences and Mathematics

Publication Details

This conference paper is available at Research Online: https://ro.uow.edu.au/infopapers/3207
A compact CMOS Face Detection architecture based on Shunting Inhibitory Convolutional Neural Networks

Xiaoxiao Zhang\(^1,2\), Amine BERMAK\(^1\), Farid Boussaid\(^2\), A. Bouzerdoum\(^3\)

\(^1\)Dept. of ECE, Hong Kong University of Science & Technology, Hong Kong
\(^2\)School of EECE, University of Western Australia, Australia
\(^3\)SECTE, The University of Wollongong, Australia

Email: zhangxx@ust.hk

Abstract—In this paper, we present a compact, low cost, real-time CMOS hardware architecture for face detection. The proposed architecture is based on a VLSI-friendly implementation of Shunting Inhibitory Convolutional Neural Networks (SICoNN). Reported experimental results show that the proposed architecture can detect faces with 93% detection accuracy at 5% false alarm rate. A VLSI Systolic architecture was considered to further optimize the design in terms of execution speed, power dissipation and area. Potential applications of the proposed face detection hardware include consumer electronics, security, monitoring and head-counting.

I. INTRODUCTION

Face detection is defined as the process of identifying the existence of human faces within an image regardless of its position, orientation, and of the environmental conditions in the scene. It is a necessary task for a wide range of applications such as surveillance, security, and consumer electronics. Current state-of-the-art face detection software implementations are very effective in detecting faces and can process up to 15 images per second [1]. The actual hardware implementation of face detection algorithms have so far been mainly considered on FPGAs, microcontrollers or multiprocessor platforms [2][3][4][5]. Recent advances in the microelectronic industry standard CMOS process, have enabled the concept of a camera-on-chip, in which face detection processing could be implemented on a single chip together with the imaging device. The fully integrated camera-on-chip, promises to offer significant advantages in terms of manufacturing cost, system volume and weight, power dissipation and increased built-in functionalities. In this paper, we explore this avenue with an integrated CMOS face detection architecture, based on Shunting Inhibitory Convolutional Neural Networks, which allow for robust localization and positioning of human face, even in the presence of partial occlusion, poor illumination conditions and facial expression changes [6].

In the next Section, we will provide an overview of Shunting Inhibitory Convolutional Neural Networks (SICoNN) and describe a VLSI-friendly algorithm to enable on-chip SICoNN silicon integration. Section III presents the proposed hardware architecture together with synthesis results for a 0.18µm CMOS process. Finally, a conclusion is given in Section IV.

II. SICoNN

The convolutional neural networks (CoNN) approach to face detection uses a class of hierarchical neural networks to perform feature extraction as well as classification [6]. Shunting inhibitory neurons have been used in a conventional feedforward architecture for classification and nonlinear regression and were shown to be more powerful than multilayer perceptrons (MLPs) [7][8]. i.e., they can approximate complex decision surfaces much readily than MLPs.

In a SICoNN, the input layer is a 2D square array of arbitrary size. Each hidden layer consists of several planes (feature maps) of shunting inhibitory neurons. Each feature map has a unique set of weights connecting them to different locations in the input image (receptive field), and each feature map has a unique receptive field. This arrangement allows neurons in a plane to extract elementary visual features from the previous layer. The same receptive field size is used to connect from one layer to another throughout the network architecture. Three possible connection strategies have been reported [6]: full-connection, toelplitz-connection, and binary-connection.

In this architecture, the activation of the hidden neurons is governed by the steady-state response of the feedforward shunting inhibitory neuron model, which can be generalized to the following:

\[ Z_j = \frac{g \left( \sum w_{ij}I_i + b_j \right)}{a_j + f \left( \sum c_{ij}I_i + d_j \right)} \]  

where \( r_j \) is the activity of the \( j \)th neuron, \( I_i \)'s are the external inputs, \( a_j \) is the passive decay rate, \( w_{ij} \) and \( c_{ij} \) are the connection weights from the \( i \)th neuron to the \( j \)th neuron, \( b_j \) and \( d_j \) are constant biases, and \( f \) and \( g \) are activation functions.

The output layer is a set of linear or sigmoid neurons (perceptrons). The response of an output neuron (Fig.1) is a weighted sum of the input signals added to a bias term,
and the result is passed through an appropriate activation function (linear or sigmoidal). Mathematically, the response of an output neuron is given by

\[ y = h \left( \sum w_z z + b \right) \]  

(2)

where \( h \) is the output activation function, \( w_z \)'s are the connection weights, \( z \)'s are the inputs to the neuron, and \( b \) is the bias term.

Based on four large test sets with 1269 persons, this system correctly detects and localizes 90.8% of the faces. When tested on segmented images, the convolutional network achieves an overall accuracy of 98.36%, and has a 99% correct classification rate at 5% false alarm rate. Some of the detection results are shown in Fig.2.

The aforementioned algorithm uses 3 hidden layers with 14 feature maps (2 in the first layer, 4 in the second layer, and 8 in the third layer). The implementation of this advanced processing on-a-chip, requires a large number of arithmetic blocks (e.g. multipliers, adders, etc), making the size and cost of the chip prohibitively high.

To enable a compact and low cost implementation, we limit the network to a single layer and evaluate the loss in performance, for the same test set of 1269 persons. Both architectures perform equally well in absence of partial occlusions, and in constrained environments. Training on a larger set should improve the performance of single layer SICoNN under occlusions.

III. HARDWARE IMPLEMENTATION

Fig.3 and Fig.4 show the general block diagram of the proposed on-chip face detection system and its computation unit respectively. We chose to use 8 feature maps, each having two weight matrices associated with it, to perform (Eq.2) face detection (Fig.4a). The output of the feature map is multiplied by a weight and then added to the result of feature maps, before being compared to a threshold value (Fig.4b). The threshold can be made equal to zero, so that a positive result indicates a face and a negative result a non-face.

The proposed face detection hardware is based on a VLSI systolic architecture, which enables the implementation of cost-effective silicon subroutines for computations such as pattern matching, error correcting, data base processing, or signal and image processing [9]. It results in cost-effective, high-performance special purpose systems applicable to a wide range of problems.

In the proposed algorithm, 2-D convolution is the most frequently used operation and will be executed over 5000 times for eight 7×7 window size feature maps convolutions on a 32×32 image frame. With the selected systolic architecture, we can decompose the 2-D convolution into several small and fast 1-D convolution cells [10]. The use of several systolic cells in a processor allows concurrent multiplications and additions. Furthermore, depending on the speed and area specifications, it is possible to trade-off speed for silicon area (i.e. cost).

The system was designed in Verilog HDL and synthesized using Synopsys Design Compiler using a TSMC CMOS 180nm technology. Table 1 shows the synthesis results for a face detection system, integrating a 32×32 pixel array. The clock frequency is 125MHz, and a total of 13570 cycles are needed to complete a full detection cycle (Fig.5). Face detection processing can thus be carried out at 9211 frames per second, enabling real-time applications. Requirements on
area and power are minimum with $0.77 \, \text{mm}^2$ and 44mW at 1.8V, respectively.

IV. CONCLUSION

In this paper, we propose a CMOS face detection architecture based on a single layer Shunting Inhibitory Convolutional Neural Network. The proposed hardware offers a good trade-off between detection performance and implementation complexity. To enable a low cost real-time on-chip integration, a VLSI systolic architecture was adopted. Potential applications of the proposed face detection hardware include security, user authentication, assistance and monitoring of the elderly, to name a few.

V. ACKNOWLEDGEMENT

This work was supported in part by an Emerging High Impact Area Grant ref. HIA05/06.EG03 and an RGC research Grant Ref. 610405. Support was also provided under the Australian Research Council Discovery Projects funding scheme (DP0664909)

REFERENCES

