A new class of convolutional neural networks based on shunting inhibition with applications to visual pattern recognition

Fok Hing Chi Tivive

University of Wollongong, tivive@uow.edu.au

A New Class of Convolutional Neural Networks based on Shunting Inhibition with Applications to Visual Pattern Recognition

A thesis submitted in fulfilment of the requirements for the award of the degree

Doctor of Philosophy

from

University of Wollongong

by

Fok Hing Chi Tivive

B. Eng. (Hons.)

Supervisor : Prof. Abdessalam Bouzerdoum

School of Electrical, Computer and Telecommunications Engineering

March 2006

Certification

I, Fok Hing Chi Tivive, declare that this thesis, submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submitted for qualifications at any other academic institution.

Fok Hing Chi Tivive
March 2006

Abstract

In the contemporary era of increased information overload, there is a growing interest in a new class of computational intelligence systems. These systems have been proven as powerful and versatile computational tools for solving certain types of problems that are too complex to be analyzed with traditional analytical means. Inspired by the computational mechanism of the human brain, many researchers have looked into neurobiology for new inspiration to solve more complex problems than those based on traditional computational techniques. Artificial neural networks, evolving from neuro-biological insights, give computer systems an amazing capability to actually learn from input data to generate solutions for problems that are too abstract to be understood or too resource-intensive to tackle. Although neural networks have been applied with success in many industries, there is a continuing demand for new types of hierarchical artificial neural networks that can overcome some of the drawbacks of the earlier models.

This thesis presents a new class of convolutional neural networks based on the physiologically plausible mechanism of shunting inhibition with its various systematic connection schemes. The network has a generic architecture in which shunting inhibitory neurons are used as feature extraction elements. A series of training algorithms, ranging from first-order gradient methods to Quasi-Newton and hybrid methods, have been implemented to adapt the synaptic weights of the developed networks; all of them have been successfully used to train the convolutional neural networks for a classification task.

To demonstrate their capability in real life applications, the convolutional networks are employed in a face detection system and a handwritten digit recognition system. The face detector has 383 trainable network parameters and achieves a detection rate of 98% for detecting human faces on a large set of unconstrained and complex images. The handwritten digit recognition system, on the other hand, has 2722 trainable parameters, and its classification rate is 97.3% for recognizing human

handwritten numerals. Besides these two applications, the developed network is analyzed for its built-in invariance, and it is implemented as a rotation invariant face classifier. The network achieves a classification rate of 97.3% in the rotation range $\pm 90^0$, and for 360^0 in-plane rotation, it has a correct detection rate of 93.6% at 5% false detection rate. These classification results demonstrate that the new class of convolutional neural networks has excellent generalization capability and achieves rotation invariance by adapting its connection weight matrices (receptive fields) as invariant feature detectors.

Acknowledgements

This thesis could not be written without the help and support of the following people. First, I would like to express my deepest appreciation and gratitude to my supervisor, Prof. Abdessalam Bouzerdoum, for his advice and permanent guidance throughout my research. I was fortunate to have him as my supervisor for both undergraduate and postgraduate studies and have his support during all these years. I would like to thank my best friend, Dr. Son Lam Phung, for his expert advice in Matlab, his help in image processing and his time for proof-reading this thesis. A special thank goes to Dr. Farid Boussaid who has shared his past experience of how to manage the time for writing a Ph.D thesis and his motivating phone calls. I would like to thank my girlfriend Bonnie for her love and daily phone call and emails which have motivated me to complete my Ph.D. And finally, I would like to dedicate this thesis to my parents, brothers and my closest aunty Lee for their support and encouragements.

Publications

- F. H. C. Tivive and A. Bouzerdoum, "Handwritten digit recognition based on shunting inhibitory convolutional neural networks," in *Proc. of the Workshop on Learning Algorithms for Pattern Recognition in conjunction with the 18th Australian Joint Conference on Artificial Intelligence*, 2005, pp. 72-77.

- F. H. C. Tivive and A. Bouzerdoum, "A fast neural-based eye detection system," in *Proc. of the International Symposium on Intelligent Signal Processing and Communications Systems* , 2005, pp. 641-644.

- F. H. C. Tivive and A. Bouzerdoum, "An eye feature detector based on convolutional neural network," in *Proc. of the Eighth International Symposium on Signal Processing and its Applications*, 2005, pp. 90-93.

- F. H. C. Tivive and A. Bouzerdoum, "Efficient training algorithms for a class of shunting inhibitory convolutional neural networks," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 541-556, 2005.

- F. H. C. Tivive and A. Bouzerdoum, "A face detection system using shunting inhibitory convolutional neural network," in *Proc. of the International Joint Conference on Neural Networks*, vol. 4, 2004, pp. 2571-2575.

- F. H. C. Tivive and A. Bouzerdoum, "A new class of convolutional neural networks (siconnets) and their application to face detection," in *Proc. of the International Joint Conference on Neural Networks*, vol. 3, 2003, pp. 2157-2162.

Contents

List of Figures

List of Tables

Acronyms

2-D	2-Dimensional
ANN	Artificial Neural Network
BFGS	Broyen Fletcher Goldfarb and Shanno
CAD	Computer-Aided Diagnosis
CDF	Cumulative Density Function
C–S	Convolutional and Sub-sampling
CoNN	Convolutional Neural Network
DY	Dai and Yuan
DVS	Double Verification Strategy
FCR	Face Classification Rate
FT	Fourier Transform
GLDS	Gray Level Difference Statistic
HS	Hestenes-Stiefel
LM	Levenberg Marquardt
LMS	Least Mean Squares
LS	Least Squares
MACE	Minimum Average Correlation Energy
MSE	Mean Square Error

MLP	Multilayer Perceptrons
MNIST	Modified National Institute of Standard and Technology
NFCR	Nonface Classification Rate
NIST	National Institute of Standard and Technology
NNE	Normal Network Evaluation
OCR	Optical Character Recognition
PDF	Probability Density Function
PR	Polak-Ribière
RBF	Radial Basis Function
ROC	Receiver Characteristic Curve
ROI	Region Of Interest
Rprop	Resilient Backpropagation
SCG	Scale Conjugate Gradient
SGLD	Spatial Gray Level Dependence
SIANN	Shunting Inhibitory Artificial Neural Network
SICNN	Shunting Inhibitory Cellular Neural Network
SICoNNet	Shunting Inhibitory Convolutional Neural Network
SOM	Self Organization Map
SSE	Sum Square Error
SVM	Support Vector Machine
TAP	Target Aim Point
TDNN	Time-Delay Neural Network

Nomenclature

Throughout this thesis, the following mathematical nomenclature has been used to denote the components of the new convolutional neural network architecture and the derivation of its training algorithms.

$\alpha(k)$	step length or the learning rate at the kth iteration
$\Delta\vec{W}(k)$	weight update at the kth iteration
$\delta_{N-1,i}$	sensitivity of the ith neuron in the $(N-1)$th layer
ℓ_L	size of the receptive field of the shunting neuron at the Lth layer
$\mathbf{A}$	Hessian matrix
$\mathbf{G}(k)$	an approximation to the Hessian at the kth iteration
$\mathbf{I}$	identity matrix
$\mathbf{W}$	a matrix
$\mid c \mid$	absolute value of the scalar of c
$\parallel \vec{c} \parallel$	Euclidean norm or least-square norm of the vector $\vec{c}$
Ψ_L, Φ_L	activation functions at the Lth layer
$\vec{d}(k)$	search direction at the kth iteration
$\vec{g}(k)$	gradient vector of the error function at the kth iteration. The gradient vector is an n-dimensional column vector given by: $$\vec{g}(k) = [g_1(k), g_2(k), \cdots, g_n(k)]^T,$$ where $g_i(k) = \partial E(k)/\partial w_i \quad (i = 1, \cdots, n)$ is the local gradient

$\vec{V}^T\vec{W}$	the inner product of two vectors
$\vec{W}$	a vector
$\vec{W}(k)$	a n-dimensional column vector containing all n free parameters (i.e., adaptable weights) of the network at the kth iteration: $$\vec{W}(k) = [w_1(k), w_2(k), \cdots, w_n(k)]^T$$
A^T	the transpose of matrix A
A^{-1}	the inverse of matrix A
$a_{L,r}$	passive decay rate constant of the neuron in the rth feature map of the Lth layer
$b_{L,r}$, $d_{L,r}$	bias parameters of the neuron in the rth feature map of the Lth layer
$C_{L,k}(x,y)$	excitatory weight at location (x,y) in the receptive field of the shunting neuron in the kth feature map of the Lth layer
$D_{L,k}(x,y)$	shunting inhibitory weight at location (x,y) in the receptive field of the shunting neuron in the kth feature map of the Lth layer
E, f	cost function or error function
$f'(x)$	the first derivative of function $f(x)$
F_L	size of the feature map at the Lth layer
h	activation function at the output layer
n	number of weights in the network
$net_{L,i}$	net input or weighted sum of inputs for the ith neuron in the Lth layer
P	number of training patterns
S_I	number of pixels in an image
S_N	number of output neurons
S_T	number of training iterations or epochs

s_{L+1} number of neurons in the $(L+1)$th layer

$sgn(x)$ the sign of a scalar x

$t^j_{L,i}$ target value of the ith output neuron in the Lth layer due to the jth input pattern

$w_{L,ij}$ connection weight from the jth neuron in the $(L-1)$th layer to the ith neuron in the Lth layer

$z^j_{L,i}$ output response of the ith output neuron in the Lth layer due to the jth input pattern

"$*$" 2D convolution operator