Artificial Neural Network Approach for Image Compression: A Survey

Trupti V. Tokarkar[1] Dr. Kirit R. Bhatt[2]
[1]PG Scholar [2]Professor
[1,2]Department of Electronics & Communication
[1,2]S.V.I.T, Vasad, Anand

Abstract— Digital images require large amounts of memory for storage. Thus, the transmission of an image from one computer to another can be very time consuming.By using data compression techniques, it is possible to remove some of the redundant information contained in images, requiring less storage space and less time to transmit. Artificial Neural networks can be used for the purpose of image compression.Apart from the existing technology on image compression represented by series of JPEG, MPEG and H.26x standards, new technology such as neural networks and genetic algorithms are being developed to explore the future of image coding.This paper discusses various neural network architectures for image compression. Among the architectures presented are the Back-Propagation networks (BPN),Kohonen self-organized maps (KSOM),Hierarchical Self-organized maps (HSOM), Modular Neural networks (MNN), Wavelet neural Networks, Fractal Neural Networks, and Cellular Neural Networks (CNN).

Key words: KSOM, Compression, CNN

I. INTRODUCTION

Image compression is a key technology in the development of various multimedia computer services and telecommunication applications such as teleconferencing, digital broadcast codec and videotechnology, etc. At present, the main core of imagecompression technology consists of three important processing stages: pixel transforms quantization and entropy coding. In addition to these keytechniques, new ideas are constantly appearingacross different disciplines and new research fronts[1].Over the last decade, numerous attempts have been made to apply artificial neural networks (ANNs) for image compression [2,3, and4].Alarge number of techniques have been developed [5] to make the storage and transmission of images economical. Research on neural networks for image compression is still making steady advances. In this paper, various neural network architectures for compression of still images are discussed. This paper consists of fivesections.. Section2 describes back propagation and Adaptive back propagation neural networks for image compression. Section 3 discusses the SOMs and HSOMs for image compression using VQ. In Section 4modular neural networks are discussed. Waveletand Fractal neural networks are discussed insection 5. Cellular neural networks are describedin section 6. Finally, Section 7 gives the conclusion.

II. BACK-PROPAGATION NETWORKS

A. *Basic Back Propagation Neural Network:*

Back-propagation neural networks are directlyapplied to image compression [2, 6]. The neuralnetwork structure is shown in Figure 1. Threelayers, one input layer, one output layer and onehidden layer are designed. The input layer and output layer are fully connected to the hidden layer. Compression is achieved by designing the value of K, the number of neurons at the hiddenlayer less than that of neurons at both input and theoutput layers. The input image is split up into blocks or vectors of 8x8, 4x4 or 16x16 pixels.When the input vector is referred to as N-dimensional which is equal to the number of pixels included in each block, all the coupling weights connected to each neuron at the hidden layer can be represented by $\{w_{ij}, i = 1,2, \dots . N, j = 1,2, \dots . K\}$ which can also be described by a matrix of order KxN. From the hidden layer to the output layer, the connections can be represented by $\{w'_{ij} : 1 \leq i \leq N, 1 \leq j \leq K\}$ which is another weight matrix of order NxK. Image compression is achieved by training the network in such a way that the coupling weights, $\{w_{ij}\}$, scale the input vector of N-dimension into a narrow channel of K-dimension(K<N) at the hidden layer and produce optimum output value which makes the quadratic error between input and output minimum

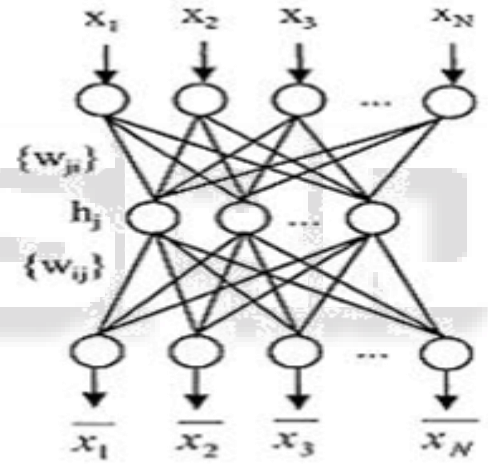

Fig. 1: Back Propagation Neural Network

B. *Hierarchical Back propagation Neural Network:*

The basic backpropagation network is furtherextended to construct a hierarchical neural network by adding two more hidden layers into the existing network as proposed in [6]. The Hierarchical neural network structure is illustrated in Fig. 2 inwhich the three hidden layers are termed as thecombiner layer, the compressor layer and the de-combinerlayer. The idea is to exploit correlationbetween pixels by inner hidden layer and to exploit correlation between blocks of pixels by outer hidden layers

From the input layer to the combiner layer and from the de-combiner layer to the output layer, local connections are designed which have thesame effect as *M* fully connected neural subnetworks.As seen in Fig. 4, all three hidden layersare fully connected. The basic idea is to divide aninput image into *M* disjoint sub-scenes and each sub scene is further partitioned into T pixel blocks of size p x p. For a standard image of 512x512, asproposed [6], it can be divided into 8 sub-scenes and each sub-scene has 512 pixel blocks of size 8x8. Accordingly, the proposed neural

network structure is designed to have the following parameters:

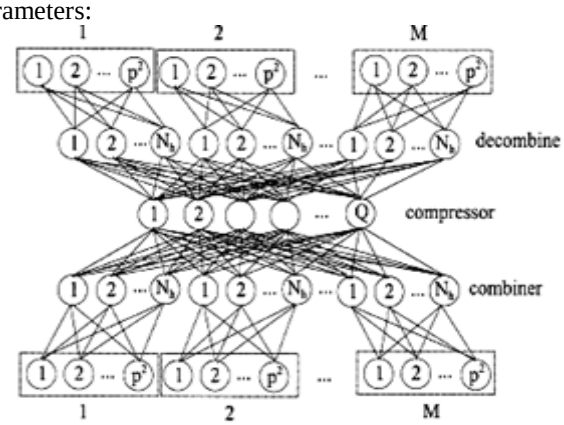

Fig. 2: Hierarchical Neural Network Structure

The total number of neurons at the input layer isMxp^2 = 8x64 =512. Total number of neurons at thecombiner layer is *MxNh*=8x8=64. Total number ofneurons at the compressor layer is *Q* = 8. Thetotal number of neurons for the de-combiner layerand the output layer is the same as that of thecombiner layer and the input layer, respectively. A socalled nested training algorithm (NTA) isproposed to reduce the overall neural networktraining time, which comprises the followingsteps:

1) Step 1: Outer loop neural network (OLNN)training.
2) Step 2: Inner loop neural network (ILNN) training.
3) Step 3: Reconstruction of the overall neural networks.

After training is completed, the neural network isready for image compression in which half of thenetwork acts as an encoder and the other half as adecoder. The neuron weights are maintained thesame throughout the compression process.

C. *Adaptive Back propagation Neural Network:*

Further to the basic narrow channel back-propagation image compression neural network, a number of adaptive schemes are proposed [2, 7] based on the principle that different neural networks are used to compress image blocks with different complexity. The general structure for the adaptive schemes are shown in Figure-3 in which agroup of neural networks with increasing number of hidden neurons, *(*h_{min}, h_{max}), is designed. The basic idea is to classify the input image blocks into a few sub-sets with different features according totheir complexity measurement. A fine tuned neural network then compresses each sub-set. Four schemes are proposed [2] to train the neural networks which are classified as parallel training, serial training, activity-based training and activity and direction based training schemes.

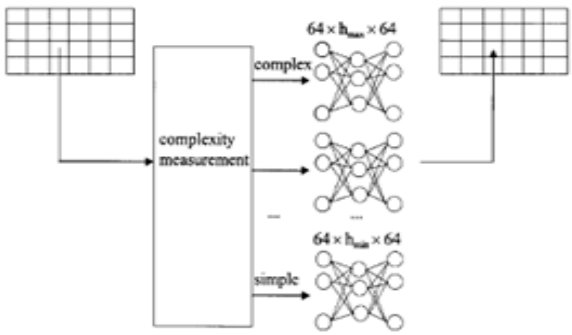

Fig. 3: Adaptive Neural Network Structure

III. SELF-ORGANIZING MAPS (SOMS)

A. *Kohonen's Self Organizing Maps (KSOM):*

One of the common methods to compress images is to code them through vector quantization (VQ) techniques [8]. VQ is a lossy compression technique. First, the image is split in to squareblocks of size τ x τ (4x4 or 8x8) pixels; each block is considered as a vector in a 16 or 64 dimensional spaces. Second, a limited number *(*l) of vectors (code words) in this space is selected in order to approximate as much as possible, the distribution of initial vectors extracted from the image. Third,each vector from the original image is replaced by its nearest codeword. Finally, during transmission, the index of the codeword is transmitted.Compression is achieved if the number of bits usedto transmit the index ($\log_2$ l) is less than the number of initial bits of the block (τ xτ x m);where *m* is the number of bits per pixel. Kohonen's self-organized feature map (KSOM) isa reliable and efficient method to achieve VQ for image compression. The basic structure of KSOM is shown in Figure 4. Compression ratios of 10:1to 100:1 have been reported using the SOM [9].The search complexity is of order O (N).

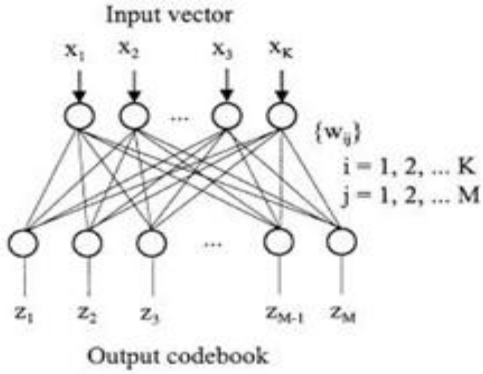

Fig. 4: SOM Neural Network for Vector Quantization

B. *Hierarchical Self Organizing Maps (HSOM)*

HSOM is an extension of the conventional SOM.A tree-structure is defined, where each node is a SOM, trained with one determined data set. The map in level-1 is trained with the full set of data and in accordance with the quantization of each neuron, the map children are trained with subgroups of this. Fig.5 illustrates the configuration of the HSOM. The trained HSOM is executed sequentially, i.e., from the highest to the lowest level of the tree. The complexity of search is reduced to O(log N) in the HSOM.

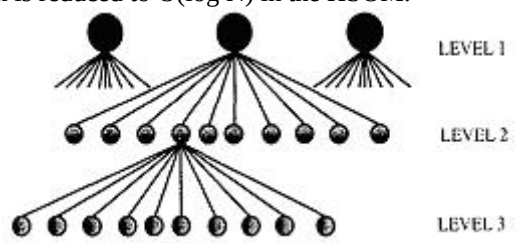

Fig. 5: Configuration of Hierarchical SOM

IV. MODULAR NEURAL NETWORKS

Though a single neural network can compress the average characteristic for the image data, it is difficult to compress both the edge and flat regions with the same precision. For this, Rahim et al. have proposed the compression method using two neural networks [10]. One of the neural network

is used for the compression of the original image and the other is used for the compression of the residual image. However, it is effective to compress for each region, which is divided in to the edge and flat regions. A modular structured neural network consisting of multiple neural networks with different block sizes (the number ofinput units) for region segmentation has been proposed by Watanbe and Mori [11]. By the region segmentation, each neural network is assigned to each region such as the edge or the flat region. From simulation results, it is shown that the proposed method yields a better compression compared with the conventional compression technique using a single neural network.

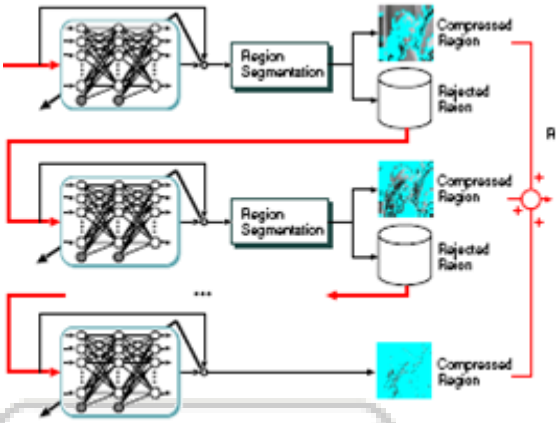

Fig. 6: Modular Neural Network

V. NEURAL NETWORK DEVELOPMENT OF EXISTING TECHNOLOGY

In this section, we show that the existing conventional image compression technology can be developed right into various learning algorithms to build up neural networks for image compression. This will be a significant development in the sense that various existing image compression algorithms can actually be implemented by one neural network architecture empowered with different learning algorithms.Two conventional techniques are covered in this section, which include wavelet transforms and fractal coding.

A. Wavelet Neural Networks:

Based on wavelet transforms, a number of neural networks are designed for image processing and representation [12],[13]. When a signal s(t) is approximated by daughters of a mother wavelet h(t), for instance, a neural network structure can be established as shown in Fig. 7 [12],[13]. Here,

1) Step1 computes a search direction **[**s] at iteration i.
2) step2 computes the new weight vector using a variable step size α ,

By simply choosing the step size α as the learning rate, the above two steps can be constructed as a learning algorithm for the wavelet neural network in Fig.7. Experiments reported [14] on a number of image samples support the wavelet neural network by finding out that Daubechie's wavelet produces a satisfactory compression with the smallest errors.

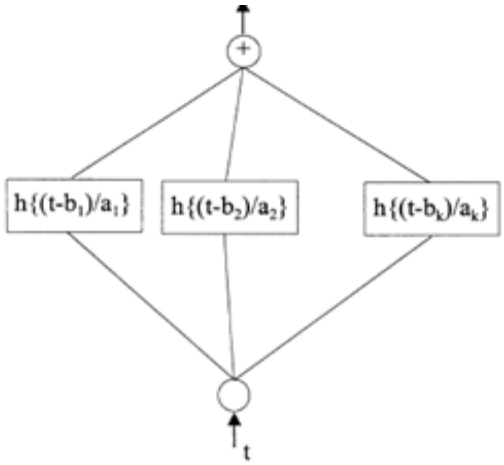

Fig. 7: Structure of Wavelet Neural Network.

B. Fractal Neural Networks:

Fractal configured neural networks [15,16], based on iterated function system (IFS) codes[17],represent another example along the direction of developing existing image compression technology into neural networks. Its conventional counterpart involves representing images by fractals and each fractal is then represented by so-called IFS, which consists of a group of affined transformations. To generate images from IFS, random iteration algorithm is the most typical technique associated with fractal based image decompression [17]. Hence, fractal based image compression features higher speed in decompression and lower speed in compression.By establishing one neuron per pixel, two traditional algorithms of generating images using IFSs are formulated into neural networks in which all the neurons are organized as a topology with two dimensions [16]. The network structure is illustrated in Fig. 8 in which $w_{ij,i'j'}$is the couplingweight between (ij)th neuron to $(i'j')$th one, and $S_{ij}$is the state output of the neuron at position (i,j).The training algorithm is directly obtained from the random iteration algorithm in which the coupling weights are used to interpret the self-similarity between pixels [64]. In common with most neural networks, the majority of the work operated in the neural network is to compute and optimize the coupling weights,$w_{ij,i'j'}$. Once these have been calculated, the required image can typically be generated in a small number of iterations. Hence, the neural network implementation of the IFS based image coding system could lead to massively parallel implementation on a dedicated hardware for generating IFS fractals.

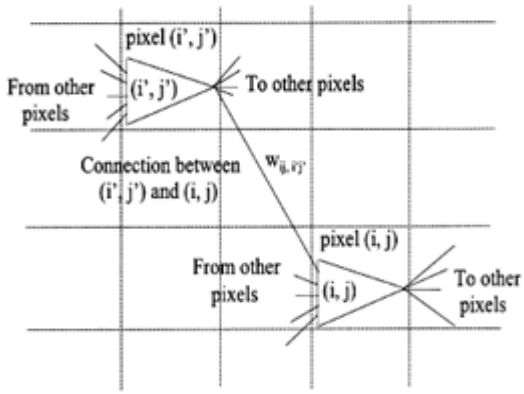

Fig. 8: Fractal Neural Network

VI. CELLULAR NEURAL NETWORKS

Recently, a novel class of information-processing system called cellular neural networks has been proposed [18]. Like neural network[19], it is a large-scale nonlinear analog circuit which processes signals in real time. It is made of a massive aggregate of regularly spaced circuit clones, called cells, which communicate with each other directly only through its nearest neighbors. Each cell is made of a linear capacitor, a nonlinear voltage-controlled current source, and a few resistive linearcircuit elements. Cellular neural networks share the best features of both worlds; its continuous time feature allows real-time signal processing found wanting in the digital domain and its local interconnection feature makes it tailor made for VLSI implementation.

The CNN is inherently local in nature, so it cannot be expected to efficiently perform global operations of a coding scheme, e.g., entropycoding. However, due to the highly parallel nature of the structure, its speed outperforms traditional digital solutions. The price for this high execution speed is the lower precision of the analog device.

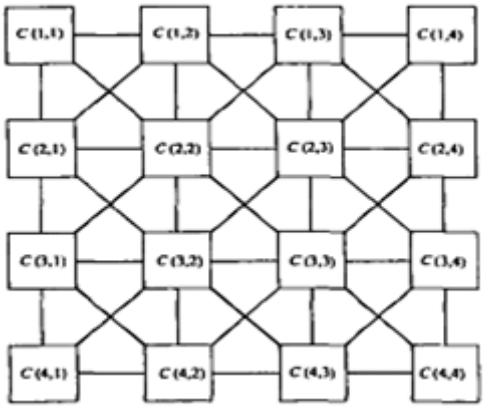

Fig. 9: Two Dimensional Cellular Neural Networks

Spatial subband coding algorithm is also well suited for the CNN architecture, which is superior to the JPEG for lossless compression both in terms of speed and compression efficiency.

VII. CONCLUSION

This paper discusses various neural network architectures for image compression, which are classified into different categories based on the nature of their application and design. These include direct development of neural learning algorithms for image compression, and neural network implementation of traditional image compression algorithms. With dedicated hardware implementation, the massive parallel computing nature of neural networks is quite obvious due to the parallel structure and arrangement of the neurons within each layer. In addition, neural networks can also be implemented on general purpose parallel processing architectures or arrays with programmable capability to change their structures and hence their functionality. Future work in image compression neural networks can be considered by designing more hidden layers to allow the neural networks go through more interactive training and sophisticated learning procedures. Accordingly, high performance compression algorithms may be developed and implemented in those neural networks.

REFERENCES

[1] J. Jiang, "Image compression with neural networks : A survey", Signal Processing: Image Communication 14 (1999) 737-760
[2] R.D. Dony, S. Haykin, "Neural network approaches to image compression", Proc.IEEE, Vol. 83, No. 2, Feb. 1995, pp 288-303.
[3] P. V. Rao,SuhasMadhusudana, Nachiketh S S, and KusumaKeerthi, "ImageCompression using Artificial NeuralNetworks",Second International Conference on Machine Learning and Computing,ICMLC, Feb 2010, Bangalore, India.
[4] PratimDuttaet.al,"Digital Image Compression Using Neural Networks," International Conference on Advances in Computing, Control, and Telecommunication Technologies,ACT2009, Tiruvananthapuram, India.
[5] R. Hecht Nielsen, "Neurocomputing", Addison Wesley, Reading, MA, 1990.
[6] A. Namphol, S. Chin, M. Arozullah, "Image compression with a hierarchical neural network", IEEE Trans. Aerospace Electronic Systems, Vol 32, No.1, January 1996, pp 326-337.
[7] G. Candotti, S. Carrato et al., "Pyramidal multiresolution source coding for progressive sequences", IEEE Trans. Consumer Electronics,Vol. 40, No.4, Nov.1994, pp 789-795.
[8] A.K. Krishnamurthy et al., "Neural Networks for VQ of speech and Images", IEEE Jl, on selected areas in comm., Vol.8, No.8, Oct 1990.
[9] Mark Nelson, Jean Loup Gailly, "The data compression Book", 2 ed., M&T Pub.Inc.,USA, 1996.
[10] N.M.Rahim, T.Yahagi, "Image Compression by new sub-image bloc classification techniques using Neural Networks", IEICE Trans. On Fundammentals, Vol. E83-A, No.10, pp 2040-2043, 2000.
[11] EijiWatanbe and Katsumi Mori, "Lossy Image compression using a modular structured neural network", Proc. of IEEE Neural Network and Signal Processing Society, XI, Sept 2001
[12] T.K. Denk, V. Parhi, Cherkasky, "Combining neural networks and the wavelet transform for image compression", IEEE Proc. ASSP,Vol. 1, 1993, pp. 637-640
[13] S. G. Mallat, "A theory for multi resolution signal decomposition: The wavelet representation", IEEE Transactions on Pattern Analysis & Machine Intelligence,Vol. 11, No.7, July 1989, pp 674-693.
[14] S.B. Lo, H. Li et al., "On optimization of Orthonormal wavelet decomposition:implication of data accuracy, feature preservation and compression effects", SPIE Proc., Vol. 27, 07, 1996, pp. 201-214.
[15] A W H Lee, and L M Cheng, "Using Artificial Neural Network for the calculation of IFS Codes, Proc. of IEEE, 1994, pp 4038-4043 ICMLC2010, Feb 2010, Bangalore, India.
[16] S J Lee et. al, "Fractal Image compression using Neural Networks", Proc.. IEEE, 1998,pp 613-618.

[17] Yuval Fisher, "Fractal image Compression Theory and application", Springer Verlag,New York, 1995.
[18] L. O. Chua and L. Yang, "Cellular neural networks: Theory and applications," IEEE Trans. Circuits Systems., vol. 35, pp. 1257–1290, Oct. 1988.
[19] L. O. Chua and T. Roska, "The CNN paradigm," IEEE Trans. Circuits Syst. I, Vol.40, pp. 47–156, Mar. 1993.
[20] W.C. Fang, et. al, "A VLSI neural processor for image data compression using self-organization networks", IEEE Trans. Neural Networks, Vol. 3 , No. 3, 1992, pp 506-519
[21] Christophe Amerijckx et.al,"Image Compression by Self-organized Kohonen maps", IEEE Trans. on Neural Networks , Vol.9, No.3, May 1998
[22] M. Mougeot, R. Azencott, B. Angeniol, "Image compression with back propagation: improvement of the visual restoration using different cost functions", Neural Networks, Vol. 4, No. 4, 1991, pp 467-476

