Neural network based image compression

Rajashree Tapase

Assistant Professor in Electronics and Telecommunication Engineering department, Annasaheb Dange College of Engineering and Technology, Ashta, Maharashtra.

---***---

Abstract: *In the development of various multimedia computer services and telecommunication applications such as teleconferencing, digital broadcast codec and video technology; there is need of image compression. Image compression is a process of efficiently coding digital image, to reduce the number of bits required in representing image. Its purpose is to reduce the storage space and transmission cost while maintaining good quality. A number of neural network based image compression scheme have been proposed for this purpose, which are classified as lossless or lossy image compression technique. Image compression is done by back propagation neural network.*

***Key Words*: Image compression, Image compression techniques, back propagation neural network and performance parameters PSNR, NMSE.**

1. INTRODUCTION

Image compression is a process of efficiently coding digital image, to reduce the number of bits required in representing image. Its purpose is to reduce the storage space and transmission cost while maintaining good quality [1]. A number of neural network based image compression scheme have been proposed for this purpose, which are classified as lossless or lossy image compression technique. Ruther hart in his early work explored the potential of neural network to achieve data encoding /decoding, which later utilized by many researches for image compression employing the standard back propagation training algorithm. In most of the methods , an image is divided into number of non overlapping pixel blocks, and fed as patterns for network training .Image compression is achieved by encoding the pixel blocks into the trained weight set ,which is transmitted to the receiving side for reconstruction of the image.

As a model to simulate the learning function of human brains, neural networks have enjoyed widespread applications in telecommunication and computer science. Recent publications show a substantial increase in neural networks for image compression and coding. Together with the popular multimedia applications and related products, what role are the well-developed neural networks going to play in this special era where information processing and communication is in great demand. Although there is no sign of significant work on neural networks that can take over the existing technology, research on neural networks of image compression are still making steady advances. This could have a tremendous impact upon the development of new technologies and algorithms in this subject area.

1.1 Necessity:

Computer images are extremely data intensive and hence require large amounts of memory for storage. As a result, the transmission of an image from one machine 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.

Neural network can be used for the purpose of image compression. It is apparent that neural network derives its computing power through, first its massively parallel distribution structure and second, its ability to learn and therefore generalize. Generalization refers to the neural network producing reasonable outputs for inputs not encountered during training(learning).These two information processing capabilities make it possible for neural network to solve complex problem that are currently intractable.

2. The Concept of Image Compression

A *digital image I* is described by a function $f : Z \times Z \rightarrow \{0, 1, \ldots, 2^k - 1\}$ where *Z* is the set of natural numbers, and *k* is the maximum number of bits to be used to represent the gray level of each pixel. In other words, *f* is a mapping from discrete spatial coordinates (*x*,*y*) to gray level values. Thus, $M \times N \times k$ bits are required to store an $M \times N$ digital image. The aim of digital image compression is to develop a scheme to encode the original image *I* into *the fewest* number of bits such that the image *I*' reconstructed from this reduced representation through the decoding process is as similar to the original image as possible: *i.e.* the problem is to design a COMPRESS and a DECOMPRESS block so that $I \sim I'$ and $|I_c| << |I|$ where |.| denotes the size in bits as shown in Fig.1.

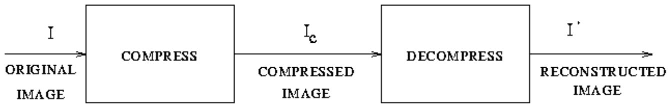

Fig-1: Image compression block diagram

3. Image compression techniques

For image compression there are two techniques lossless and lossy techniques.

3.1 Lossless image compression

Lossless compression is sometimes preferred for artificial images such as technical drawings, icons or comics. This is because lossy compression methods, especially when used at low bit rates, introduce compression artifacts. Lossless compression methods may also be preferred for high value content, such as medical imagery or image scans made for archival purposes. Lossy methods are especially suitable for natural images such as photos in applications where minor (Sometimes imperceptible) loss of fidelity is acceptable to achieve a substantial reduction in bit rate.

Compressing an image is significantly different than compressing raw binary data. Of course, general purpose compression programs can be used to compress images, but the result is less than optimal. This is because images have certain statistical properties which can be exploited by encoders specifically designed for them. Also, some of the finer details in the image can be sacrificed for the sake of saving a little more bandwidth or storage space. This also means that lossy compression techniques can be used in this area. The image compression technique most often used is transform coding.

3.2 Lossy image compression:

Lossy compression reduces a file by permanently eliminating certain information, especially redundant information. When the file is uncompressed, only a part of the original information is still there (although the user may not notice it). Lossy compression is generally used for video and sound, where a certain amount of information loss will not be detected by most users. Lossy methods are especially suitable for natural images such as photos in applications where minor (Sometimes imperceptible) loss of fidelity is acceptable to achieve a substantial reduction in bit rate. The JPEG image file, commonly used for photographs and other complex still images on the Web, is an image that has lossy compression. Using JPEG compression, the creator can decide how much loss to introduce and make a trade-off between file size and image quality.

4. Neural Network:

In this section we will see how the concept of biological neurons is used for forming artificial neurons.

4.1 Biological neurons

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems.

An Artificial Neural Network (ANN), like people, learns by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process.

Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of ANN as well.

In general a human brain neuron as shown in fig.2 is composed of a group or groups of chemically connected or functionally associated neurons. A single neuron may be connected to many other neurons and the total number of neurons and connections in a network may be extensive. Connections, called synapses, are usually formed from axons to dendrites, though dendrodendritic microcircuits and other connections are possible

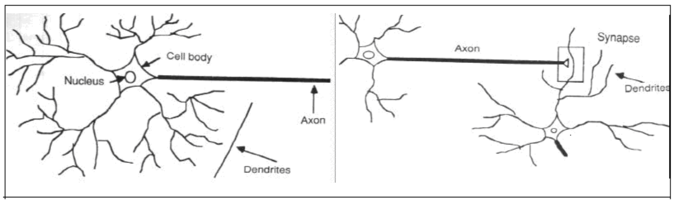

Fig-2: Human brain neuron

4.2 Artificial Neurons

ANN is based on the basic model of the human brain with capability of generalization and learning. The purpose of this simulation to the simple model of human neural cell is to acquire the intelligent features of these cells. The term "artificial" means that neural nets are implemented in computer programs that are able to handle the large number of necessary calculations during the learning process. ANN has gain a lot of interest over the last few years as a powerful technique to solve many real world problems. The basic computing element in the biological system is the neuron which receives electrochemical signals from different sources and generates electric impulses to be transmitted to other neurons.

Similar to the biological neural cell, the unit of structure of ANN is the neuron which consists basically of a summer and an activation function as shown in fig.3 where $I_1, I_2, I_3,\ldots, I_n$ are the inputs to the neuron with corresponding weights $w_1, w_2, w_3,\ldots, w_n$ which model the synaptic neural connections in biological nets and act in such a way as to increase or decrease the input signals to the neuron. Sometimes a bias term is added to the inputs.

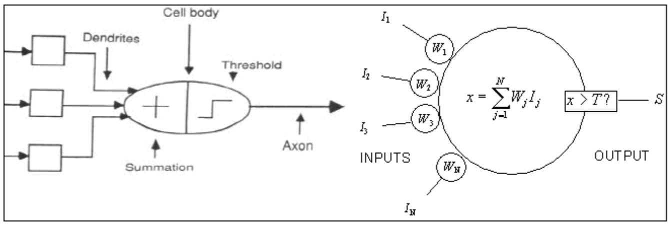

Fig-3: Human neurons to artificial neurons

Generally, inputs, weights, thresholds and neuron output could be real value or binary or bipolar. All inputs are multiplied by their corresponding weights and added together to form the net input to the neuron called net. The neuron behaves as activation or mapping function f (net) to produce an output y.Table.no.1.shows differentiation between biological neural network and artificial neural network.

5. BACK PROPAGATION NEURAL NETWORK

Back propagation is a systematic method for training multi-layer artificial neural network .It has a mathematical foundation that is strong if not highly practical. It is a multi-layer forward network using extend gradient descent based delta learning rule, commonly known as back propagation (of errors) rule. Back propagation provides computationally efficient method for changing the weights in a feed forward network, with differentiable activation function units, to learn a training set of input output examples.

Being a gradient descent method it minimizes the total squared error of the output computed by the net. The network is trained by supervised learning method. The aim of this network is to train the net to achieve a balance between the ability to respond correctly to the input patterns that are used for training and the ability to provide good responses to the input that are similar.

51. ARCHITECTURE

A feed forward back propagation neural network with one input layer with 64 neurons, one layer of z –hidden units with 16 neurons and output layer with 64 neurons is shown in the fig-4. The input layer is connected to hidden layer and hidden layer is connected to the output layer by means of interconnection weight, only the feed forward phase of operation is shown.

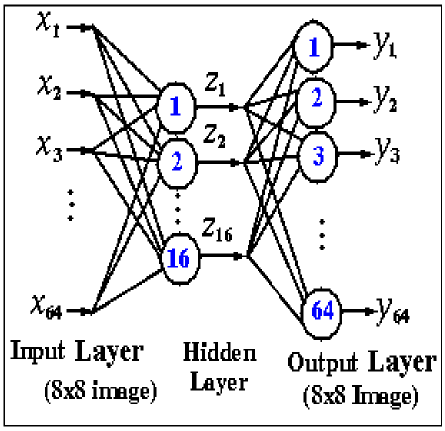

Fig-4: Architecture of back propagation network

5.2 PERFORMANCE PARAMETERS

The two acceptable measurements for the quality of reconstructed images which are PSNR (peak-signal-to noise ratio) and NMSE (normalized mean-square error). For a grey level image with *n* blocks of size *N*, or *n* vectors with dimension *N*, their definitions can be given, respectively, as follows:

$$\mathrm{PSNR} = 10\log\frac{255^2}{\frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N}(\overline{P_{ij}} - P_{ij})^2}\,(\mathrm{dB}),$$

NMSE:

$$\mathrm{NMSE} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{N}(\overline{P_{ij}} - P_{ij})^2}{\sum_{i=1}^{n}\sum_{j=1}^{N}P_{ij}^2},$$

Where P_{ij} is the intensity value of pixels in the reconstructed images; and P_{ij} the intensity value of pixels in the original images which are split up into *n* input vectors: $xi= \{P_{i1}, P_{i2},\ldots, P_{i6}\}$. Where P_{ij} is the intensity value of pixels in the reconstructed images; and P_{ij} the intensity value of pixels in the original images which are split up into *n* input vectors: $xi= \{P_{i1}, P_{i2},\ldots, P_{i6}\}$.

6. CONCLUSION

Image compression is done by back propagation neural network. In the architecture of back propagation neural network it consists of three layers, one input layer, one hidden layer, and one output layer. Both of input layer and output layer are fully connected to hidden layer. Activation function used is sigmoid function.

The architecture is used to train the net to achieve a balance between the ability to respond correctly to the input patterns that are used for training and the ability to provide good responses to the input that are similar.

REFERENCES

[1] J.Jiang (1999), "Image compression with neural networks – A survey", Signal Processing Image Communication 14 pp. vol.737 – 760.

[2] Rafael C.Gonzalez, Richard E.Woods (2002),"Digital image procession", Second Edition,by Prentice-Hall.

[3] Simon Haykin (1998), "Neural Networks-A comprehensive Foundation", Prentice-Hall of Indian private Limited, New Delhi.

[4] B.Yegnanarayana (1999), "Artificial Neural Networks", Prentice-Hall of Indian private Limited, New Delhi.

[5] James.A.Anderson (1995), "An Introduction to Neural Networks", Prentice-Hall of Indian private Limited, New Delhi.

Satish Kuma(2004), "Neural Network-A Classroom Approach", Tata McGraw-Hill Publishing company limited.