http://www.moderntechno.de/index.php/meit/article/view/meit25-02-019
DOI: 10.30890/2567-5273.2023-25-02-019

UDC 004.2

USING NEURAL NETWORKS IN THE PROCESS OF IMAGE COMPRESSION

Breus B.V.
Postgraduate student
Mohylnyi O.A.
Postgraduate student
Cherkasy state technological university, Cherkasy, bul. Shevchenka, 460,18000

Abstract. *The main attribute of neural networks is determining relationships among different parameters of information, which allows providing it more densely through the detection and rejection of unnecessary information (lossless compression) or through rejection of a part of the information that is less important and which loss we will barely notice (lossy compression). The neural network of data compression consists of developing and training a neural network to receive a layer with less number of neurons than in an input one. The received layer should allow restoring the input (lossless compression) or something close to it (lossy compression). As a result of the compression, the image size is reduced. It reduces the time of image transmission via the network and saves data storage space. In this article, we will consider data compression algorithms that are used in a widespread TIFF file format.*

Key words: *visual information, neural networks, image compression, neural network training, algorithm*

Introduction

During all time of the existence of digital technologies the task of compact data transmission through information channels was, is and most likely will be relevant.

A distinct example of such information is visual information which static images can be attributed to. With the development of digital data, processing requirements for image size, compression speed and quality of its transmission through channels of communication have increased but all existing algorithms to date can not satisfy such requirements. That is why it is necessary to search for new methods of compression which perform intellectual data analysis and which are more effective in certain compression characteristics. Under such circumstances, the neural network approach to data compression looks promising.

Today significant scientific and practical interest in computational structures of a new type — artificial neural networks — is indisputable.

It is caused by a number of successful applications of this new technology which allowed to develop effective approaches to solving problems that were thought to be complicated for implementation at traditional computers.

A significant contribution to the study of the properties of neural networks was made by such scientists as V.A. Duke [1], I.D. Mandel [4], A.A. Starikov[5], N.G. Erushkina [7].

There is a certain field of action for the use of neural networks in the problem of image compression regarding the specifics and effectiveness of the such method.

The main advantage of neural networks in the problem of image compression is a high processing speed made possible by a parallel implementation. That is why software development for image compression based on a neural network is quite relevant.

The main property of neural networks is to determine the relationships between various parameters of information, which allows us to present it more compactly, by identifying and rejecting redundant information (lossless compression), or discarding that part of information that is not important and we hardly notice its loss (compression with losses).

Therefore there is a need for researching artificial neural networks as a tool for image compression. In this study information technology for image, compression will be proposed and implemented.

Main text

When compressing graphic information, different techniques are used to reduce the number of bytes needed to represent an image. Of course, a lot depends on the compression method and the content of the graphic file (some files are compressed much better than others), but the case is quite ordinary when a large graphic file is compressed five times or more. There are methods that compress even more, but with a loss of properties - when restoring the image, information is lost. As a result, the unpacked picture may become slightly blurred and discolored. Lossy methods give higher compression ratios, but do not allow to reproduce the original image with pixel accuracy. The human eye does not perceive all subtle color shades in an ordinary raster image. Thus, some details can be omitted without visible violation of the informational content of the picture.

Neural network data compression consists in designing and training a neural network in such a way as to obtain a layer with fewer neurons than in the input, such that it is possible to restore the input signal (lossless compression) or close to it (lossy compression).

As a result of compression, the size of the image is reduced, which reduces the time of image transmission over the network and saves storage space.

The whole purpose of training a neural network is to minimize costs. Neural networks do this using a process called back propagation. It seems like a complicated word, but it's quite simple. As I mentioned, forward propagation is when you run information through a neural network to get a result. Backward distribution is literally the same, but backward. You simply start from the original layer and run the neural network backwards to optimize the weights and biases.

A type of artificial neural networks that can be used to compress data and images is Kohonen neural networks (Fig. 1).

It is shown in [6] that such networks can give better results than networks with inverse error propagation during compression. Kohonen networks are one of the main types of self-organizing neural networks. The ability to self-organize provides new opportunities - adaptation to previously unknown input data. It seems that this is the most natural way of learning, which is used in our brain, where there are no specific patterns. These regularities are formed in the process of training combined with normal work. Kohonen networks are synonymous with a whole group of networks that use the method of self-organization, the competitive method of training. We set signals at the inputs of the network, and then select the winning neuron that best matches the input vector. The exact pattern of competition and later modifications of synaptic wages can take different forms. There are many subtypes based on

competition, which differ among themselves in the exact algorithm of self-organization.

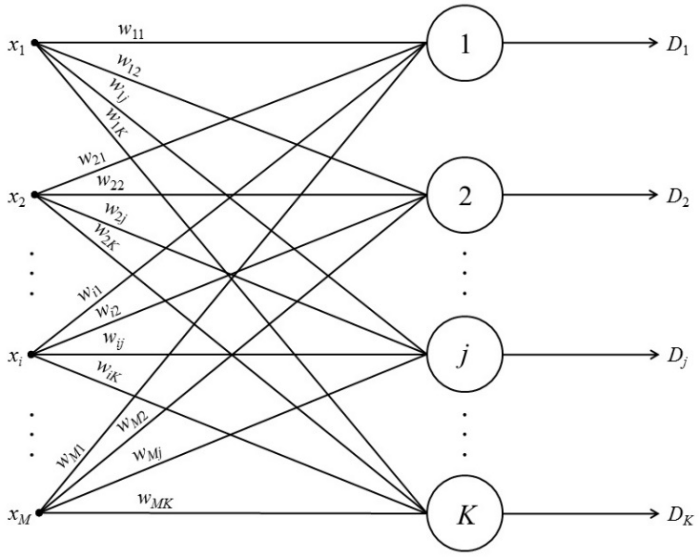

Fig. 1 – Kohonen network model

As opposed to traditional methods of compression – mathematical computation and removing redundancies – in solving the problem of compression neural network relies on a lack of resources. Network topology and its learning algorithm demand that high-dimensional data should be transferred from the input of the neural network to its outputs through a relatively small channel. To make such compression multi-layer perceptron can be used. Its architecture will be the following: the number of neurons in the input and output layer is equally equal to the dimensionality of compressed data. Between these layers, there are one or more smaller intermediate layers. A number of intermediate layers define the level of complexity of data conversion. For example network with three intermediate layers can compress training data better but can result worse in a real situation. It happens because some dependency that bears no relation to reality may occasionally be formed in initial data.

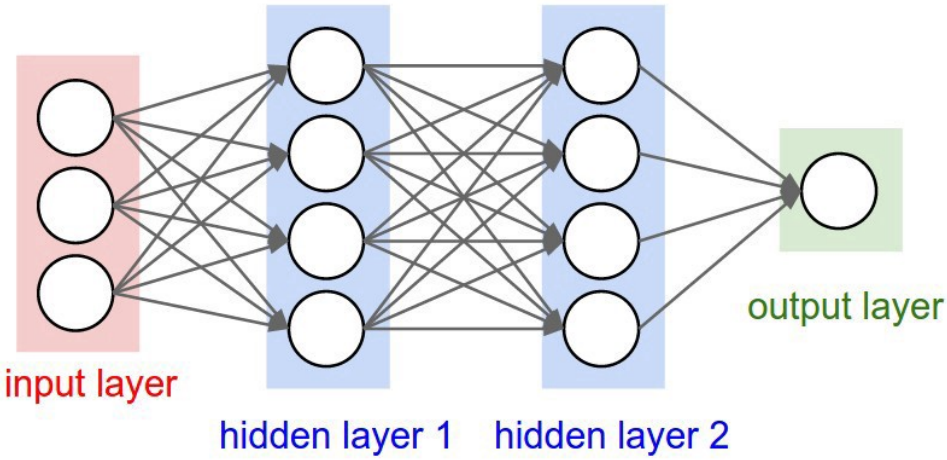

Fig. 2 – Layers of the neural network

Output data for the network are made in such a way that outputs always have the same set of signals as inputs. During its work backpropagation algorithm minimizes

the error. It means that weights of associations from the input layer of neurons to approximately its middle layer will act to compress the signal and the rest of the layers will act to decompress it. In the real world acquired network is split into two. The output of the first network is transmitted through a communication channel and fed to the input of the second network responsible for decompression.

Neural network training is usually made step by step. Such steps are usually named cycles or epochs. At each step input of the neural network is fed with all elements of the training set, then input values are calculated which in return are compared with target (training) values and function error (training criterion) is calculated.

Image with .tif format is selected as input data. In each specific case images with size (1:64, 1:64) pixels are used. The network will be developed with 4 neurons in the first layer (compression) and 16 neurons in the second layer (decompression) but selected values may vary depending on desired results. A sigmoid function will be used for the first layer. A linear function will be used for the second layer. Then training using an optimization algorithm is conducted for solving many complicated tasks including image compression. Backpropagation is one of the methods of training of multi-layer feed-forward neural network. Multi-layer perceptrons are successfully used to solve many complicated tasks including image compression.

Compression with this method will occur faster because of the use of an optimization algorithm during the training of the neural network. Compression of 250-270 KB images will occur in 2-3 seconds. Such information technology is better than its alternatives if an image is no more than 300 KB in size.

Summary and conclusions.

Have been considered the algorithm of image compression that based on a feedforward neural network using an optimization training algorithm is proposed. Software based on this algorithm will be developed. It will have advantages over existing services in such parameters as speed, convenience, accessibility, image processing quality and, most importantly, it will have a higher compression ratio, and therefore it will have a higher compression efficiency.

References

1. Duke V.A. Data Mining - data mining // Information technologies: website. – URL: http://www.inftech.webservis.ru/it/database/datamining/ar2.html

2. Xianjun Ni Research of Data Mining Based on Neural Networks // World Academy of Science, Engineering and Technology. - 2008. - No. 39. - P. 381-384.

3. Xu R. and Wunsch D. II. Survey of Clustering Algorithms. IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 16, NO. 3, MAY 2005, pp. 645ñ678.

4. Mandel I.D. cluster analysis. Moscow: Finance and Statistics, 1988. No176 p.

5. Starikov A. Practical application of neural networks for classification (clustering) problems, http://www.basegroup.ru/neural/practice.htm, January 2000.

6. Olefir A.O. Neural network model of data compression– International conference "Intellectual analysis of information". – Kyiv: Prosvita, 2015 – pp. 152 – 157.