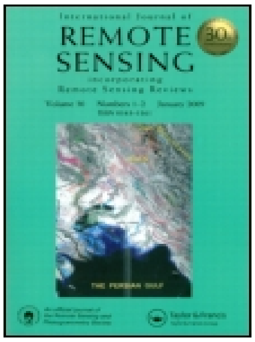


International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: http://www.tandfonline.com/loi/tres20

An improved backpropagation neural network for detection of road-like features in satellite imagery

U. Bhattacharya & S. K. Parui

INT. J. REMOTE SENSING, 1997, VOL. 18, NO. 16, 3379–3394

An improved backpropagation neural network for detection of road-like features in satellite imagery

U. BHATTACHARYA† and S. K. PARUI

Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, 203, B. T. Road, Calcutta-700 035, India

(*Received 17 October 1996; in final form 24 April 1997*)

Abstract. This paper presents an application of backpropagation neural network for the detection of linear structures in remote-sensing images. The purpose of the approach is two-fold. First, to exploit the advantages of a neural network classifier over the tranditional ones. Second, to avoid the strategic phases of enhancement and thresholding. Once the network is learnt, the classification scheme is real-time. Two critical issues in the present approach are the selection of the network architecture and the rate of convergence of learning. Solutions to these two problems are proposed. Experimental results on IRS and SPOT images are presented. Satisfactory classification results have been obtained using the network.

1. Introduction

Remote sensing image data of the Earth's surface acquired either from aircraft or from spacecraft platforms are now widely available in digital format. Researchers from various disciplines are involved in processing such remotely-sensed image data for solving different real-life problems. Detection of linear and curvilinear structures (like roads, runways etc.) contributes significantly towards the understanding of such images. In the present paper, we will address the problem of delineating roadlike structures in such remotely-sensed images using a widely used neural network model. A roadlike structure means a segment of an image which contrasts sufficiently with its background (either lighter or darker), has a reasonably uniform width, and sufficient length (Vasudevan *et al.* 1988).

There are a variety of techniques currently available in the literature (see § 2.2) for automatic detection of roadlike structures in remotely-sensed images and these are mostly based on either heuristics or knowledge-based methods or statistical techniques or low level enhancement operators. In these techniques, images go through certain intermediate steps and finally the segmented image is obtained. We have attempted to develop a method based on the multi-layer perceptron (MLP) network, which does not require the framing of a set of heuristics, or a knowledge base, or determining certain enhancement operators, but instead can obtain the desired features directly from the signal domain. Our approach is similar to the apparent technique of human visual perception of an image which possibly uses only the spatial distribution of pixel values for the recognition purpose.

† Partially supported by research fellowship grant from UGC.

We have tested the proposed MLP-based approach to roadlike structure determination using IRS (Indian Remote-sensing Satellite) and SPOT images considering only a single band. The results obtained are encouraging. Moreover, this method has several advantages over the existing methods which are described later.

2. Background

2.1. *Neural network-based approaches to remotely-sensed image classification*

Since 1986 when Rumelhart *et al.* (1986) published the backpropagation (BP) algorithm, the MLP network has become extremely popular due to the fact that such a network can be successfully applied to a wide variety of problems. In fact, in recent years numerous attempts have been made to employ an MLP network equipped with the backpropagation training algorithm to general pattern recognition problems.

Neural network based classification tools have some advantages over the traditional classifiers. They can amicably solve the problem of classifying complex data sets. They are non-parametric in nature unlike the traditional classifiers. In fact, such a tool does not require any knowledge of the underlying distribution. It performs better when the classes are strongly non-Gaussian (Lippmann 1987) and can form arbitrary decision boundaries. Moreover, a neural network, once trained, can perform classification tasks in almost real time. Paola and Schowengerdt (1995) made a comprehensive review of the research work on classification of remotely-sensed images using MLP networks. They observed that the neural network approach is feasible for classifying remotely-sensed images. Among the neural network based remotely-sensed image classification schemes, most are based on an MLP network along with the original backpropagation algorithm or its modifications (see Hepner 1990, Heerman and Khazenie 1992, Bischof *et al.* 1992, Wilkinson *et al.* 1992, Civco 1993, Dreyer 1993). There also exist several approaches (Benediktsson 1990, Liu and Xiao 1991, Schaale and Furrer 1995) for classification of remotely-sensed image data using various other neural network models.

Benediktsson *et al.* (1990) studied both statistical (parametric) and neural network (distribution free) based approaches to classification of multi-source remote sensing data and observed the advantages and disadvantages of the two classification schemes. Augusteijn *et al.* (1995) studied neural networks for the classification of satellite images and arrived at the conclusion that a neural network can be a powerful instrument in the classification of ground covers in satellite images. Also, Lee *et al.* (1990) observed that a neural network classifier has the potential of delivering an improved classification accuracy for geophysical data using a very small training data set.

Though numerous examples of neural network based approaches to classification of remotely-sensed images exist, no such attempt for extracting the line-like features from such imagery has been reported in the literature. Extracting the road-like structures from an image can be thought of as a classification problem where the image-pixels are to be classified into road and non-road pixels. In this paper, we will discuss the potential of an MLP network in performing this task and also some related issues.

2.2. *Existing approaches to line detection in remote sensing images*

Bajcsy and Tavakoli (1976) described an algorithm for computer recognition of real roads and roadlike objects from satellite pictures. The authors considered a certain low-level operator and a world model. Vanderbrug and Rosenfeld (1978) described a method for a global representation of the curves in an image by linking together the responses of local line detection operations. The method has been applied to map linear features in satellite imagery. Fischler *et al.* (1981) considered the problem of tracking roads in clear imagery of rural scenes at low resolution. They proposed a general paradigm for the integration of information from multiple image operators and knowledge sources. Also, they described its application for the detection of road-like structures. Groch (1982) presented a method for the extraction of line shaped objects from grey level pictures semiautomatically. In this approach, the author used an operator for analysing the profiles of grey level diagrams, detected starting points, considered methods for line following and finally combined all these in a system for extraction of line shaped objects from grey level images.

Vasudevan *et al.* (1988) developed an expert system for aerial image interpretation. Initially a low level module extracted roadlike feature segments through the use of templates. Then the isolated blobs of clutter were removed. In the next stage, the task of partitioning and connecting the broken roadlike segments using heuristics based upon knowledge of the domain, was performed. Zhu and Yeh (1986) also considered the problem of road network detection from aerial photographs. They first applied edge operation to detect linear segments. Then small curved elements were removed by applying a thresholding technique. A few antiparallel segments were used for generating longer road pieces by using certain heuristics. For road tracking in aerial imagery McKeown and Denlinger (1988) proposed a multi-level architecture to image analysis. This considered the cooperation among low-level processes and the aggregation of information was considered by high-level analysis. The high level module generates a symbolic description of the road in terms of a number of attributes of the road.

Parui *et al.* (1991) described a parallel algorithm for detecting road-like structures in satellite images. In their approach the curvilinear or line-like structures were enhanced using masks of different orientations. The enhancement operator highlighted only thin structures and suppressed others, including edges. Mukherjee *et al.* (1994) proposed an adaptive thresholding technique based on statistical models for segmenting the linear features in the enhanced image. A robust parameter estimation technique has been developed for this purpose. Mukherjee *et al.* (1996) presented an algorithm for finding road-like structures, which consisted of line enhancement, segmentation and linking. For the segmentation purpose the authors have proposed successive eliminations of non-road pixels from the original image.

3. Pattern recognition using multilayer perceptrons

3.1. *Multilayer perceptrons*

The architecture of an MLP is a layered one. In such a network the processors, called nodes, are arranged in three or more layers. The nodes in the second or higher layers are connected with only the nodes in the layer just below it. The connection strengths are the connection weight values. The nodes in the lower most layer are called input nodes and they only feed the input pattern vectors to the network. The nodes in the upper most layer are called output nodes and this layer provides the network response to the input pattern vector. The nodes in the intermediate layers are called hidden nodes. These are used to generate the coding of the input pattern vector on a space of much lower dimension than the input space. The architecture of an MLP network is shown in figure 1. The activity of a hidden node or an output

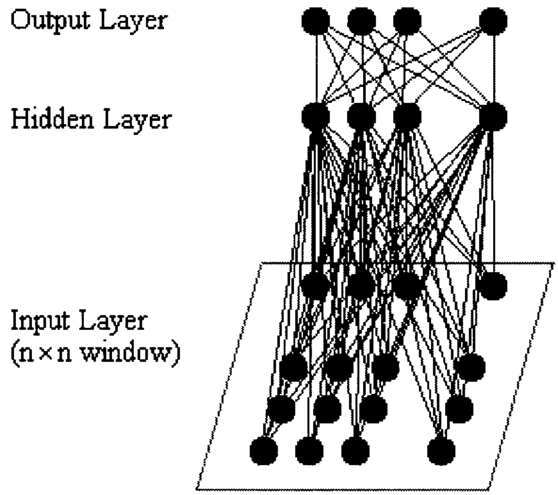


Figure 1. Architecture of a typical MLP network.

node is computed as some nonlinear bounded monotonically increasing function of a weighted sum of the node activities in the layer just below it.

3.2. *Approximation capability of MLP networks*

An MLP network can be described as a universal approximator of continuous functions. In fact, a single hidden layer feedfoward network with adequate number of hidden nodes and with sigmoidal type activation at hidden nodes, can approximate any continuous function $f: R^m \rightarrow R^n$ to any desired degree of accuracy, the output nodes being considered as linear. This result has been proved using the classical theorem on function approximation due to Kolmogorov (Hecht-Nielsen 1990). Thus a single hidden layer MLP network with a single linear output unit can be successfully employed in a pattern recognition problem but this requires encoding of the class labels via quantized levels of the single linear output unit. However, this is not practical as this imposes unnecessary constraints on the weights connecting the hidden nodes and the output node. In practical implementations, local encoding of classes is commonly used, where a number of sigmoidal-type nodes in the output layer is considered, each being responsible for representing a unique class. This scheme considerably relaxes the constraints on the connection weights feeding to the output nodes.

4. Proposed method

In the proposed approach, a small square window around each pixel in the original image is presented to the input layer of an MLP network. The network itself calculates certain distinguishing feature values in its hidden nodes and finally, classification is performed at the output node. The result obtained after presentation of the whole image to the MLP network is a binary image in which a road pixel is black and a non-road pixel is white.

Here it is to be noted that an MLP can detect the road-like structures in a given satellite image only after training it properly. The training has been done by using a certain variant of the original backpropagation algorithm. The training of an MLP involves a number of considerations which are described below. Moreover, for a successful training of an MLP network, a right choice of the network architecture is also important.

4.1. *Training of MLP network*

The connection weight values in an MLP network are obtained by means of a supervised training procedure. For this, a set of training patterns which can represent the two underlying pattern classes (road and non-road classes), is formed in which there are a sufficient number of training samples from each class. The training set is presented to the network repeatedly and after presentation of each training sample, the connection weights are modified. The process is continued until a stopping criterion is satisfied. This training procedure is popularly known as backpropagation algorithm.

4.1.1. *Selection of training set of patterns*

The patterns in the training set corresponding to a particular class must be representative of the class. The separability among the training patterns from the two classes in the feature space is captured by the MLP network during its training and this knowledge is used during recognition phase.

The present problem is a two-class problem. A pixel in the image is to be classified as a road or non-road pixel and the classification is performed using local information only. A window of size $n \times n$ around a pixel constitutes a training pattern. For the selection of such training patterns, a few issues are important.

1. Each class has several subclasses. For example, a non-road pixel may be a pixel from a crop area or a concrete area or a lake. Similarly, a road-like structure may be a city road or an airport runway or a highway. The training set should consist of patterns from each of these subclasses. If any of these subclasses is not properly represented in the training set, the pixels from that subclass may not be classified properly during the recognition phase.
2. The selection of the size of the window around each pixel is very crucial for the successful training. A larger window implies more memory and computational burden while a small window may not capture the distinguishing features. However, in higher resolution images, roads appear thicker and a larger window size is needed for the purpose. A rule of window size selection is that (*i*) it should not be less than the thickness of the road, (*ii*) it should be just large enough so that a window containing road pixels should also contain some backround pixels around the road.
3. The road-like structures in an image have different orientations and windows around pixels on road-like structures of different orientations should be included in the training set.

Keeping the above factors in mind, we selected a sub-image (128 pixels by 128 pixels) of the satellite image where a wide variety of both roadlike and non-road structures are present. From this sub-image, 25 pixels of each class are manually selected so that different subclasses get proper representation in the training set. Around each of these 50 pixels, a window of size $s \times s$ ($s=3$ and 5 for IRS and SPOT images respectively), is considered and the pixel grey values (arranged in the raster scan order) in such a window form a training pattern vector. To reduce the direction dependency in the training set, a window (corresponding to a chosen pixel) is presented to the training set with four orientations, e.g., 0° (the actual window), 90°, 180° and 270°. Their mirror images also are considered. Each component of such a training pattern vector is normalized with respect to the maximum possible grey value, i.e., 255.

4.1.2. *Backpropagation algorithm*

The backpropagation algorithm (Rumelhart *et al.* 1986) is normally used to train an MLP network. This algorithm performs steepest descent (corresponding to each pattern p in the training set) in the connection weight space on an error surface defined by

$$E_p = \frac{1}{2} \sum_k (t_{pk} - o_{pk})^2 \tag{1}$$

where $\{t_{pk}\}$, $\{o_{pk}\}$ are respectively, the target and output vectors corresponding to the pth input pattern. The system error is defined as

$$E = \frac{1}{P} \sum_p E_p \tag{2}$$

where P is the total number of patterns in the training set.

In backpropagation algorithm weight modification rules are given by

$$w_{jk}(t+1) = w_{jk}(t) - \eta \frac{\partial E_p(t)}{\partial w_{jk}(t)} \tag{3}$$

and

$$w_{ij}(t+1) = w_{ij}(t) - \eta \frac{\partial E_p(t)}{\partial w_{ij}(t)} \tag{4}$$

where $w_{jk}(t)$ is the weight connecting a hidden node j to an output node k while $w_{ij}(t)$ is the weight connecting an input node i to a hidden node j, each at time t, η is a positive constant, called the learning rate. The initial weights (at time $t=0$) are usually taken as random values.

As the backpropagation algorithm performs steepest descent on a hyper surface, called error surface, in the weight space, there is no guarantee that the global minimum of that surface will be reached (even approximately) after a moderate number of sweeps, and the algorithm may get stuck at a local minimum. The difficulty in reaching the global minimum on the error surface from a random initial position is dependent on the nature of the surface. To tackle the problem Rumelhart *et al.* (1986) suggested a modification of the above weight modification rule by including a momentum term and the rules (3) and (4) become

$$w_{jk}(t+1) = w_{jk}(t) - \eta \frac{\partial E_p(t)}{\partial w_{jk}(t)} + \alpha \Delta w_{jk}(t-1) \tag{5}$$

$$w_{ij}(t+1) = w_{ij}(t) - \eta \frac{\partial E_p(t)}{\partial w_{ij}(t)} + \alpha \Delta w_{ij}(t-1) \tag{6}$$

where $\Delta w(t-1)$ is the change in the corresponding weight at time $t-1$, and $0 < \alpha < 1$ is a constant, called the momentum factor.

In many situations, the inclusion of the momentum term increases to an extent the convergence rate of the algorithm without leading to oscillation. In the present problem, we have observed that the momentum term in the weight modification rule does contribute to faster convergence. Now, the selection of a proper value of the learning rate (η) is an important issue in the successful convergence of the back-propagation algorithm. This problem of selecting a suitable value of η (or step size) is common to all steepest descent methods. Bhattacharya and Parui (1995 a) observed that the single layer perceptron learning algorithm can show improved performance if the learning rate is varied over time under some constraints. This observation is also valid for multi-layer perceptrons (Jacobs 1988).

4.1.3. *Self-adaptation of learning rates*

During our extensive study on the convergence of the backpropagation algorithm, we have made the following observations:

1. Every weight of the network should have its own individual learning rate.
2. Every learning rate should be allowed to vary over time.
3. When the error derivative with respect to a weight possesses the same sign for several consecutive steps, the learning rate for that weight should be increased.
4. When the sign of the derivative with respect to weight alternates for several consecutive steps, the learning rate for that weight should be decreased.
5. The modifications to the learning rates at any time should be entirely based on the shape of the error surface at the present position.
6. The overall convergence performance of the algorithm should not depend much on the choice of the parameter values involved in the modification rule for the learning rates.
7. The values of the learning rates should not be allowed to increase indefinitely in order that the weight values do not explode.

The first four of the above observations were also made by Jacobs (1988). A newly suggested modification (Bhattacharya and Parui 1995 b) of the backpropagation

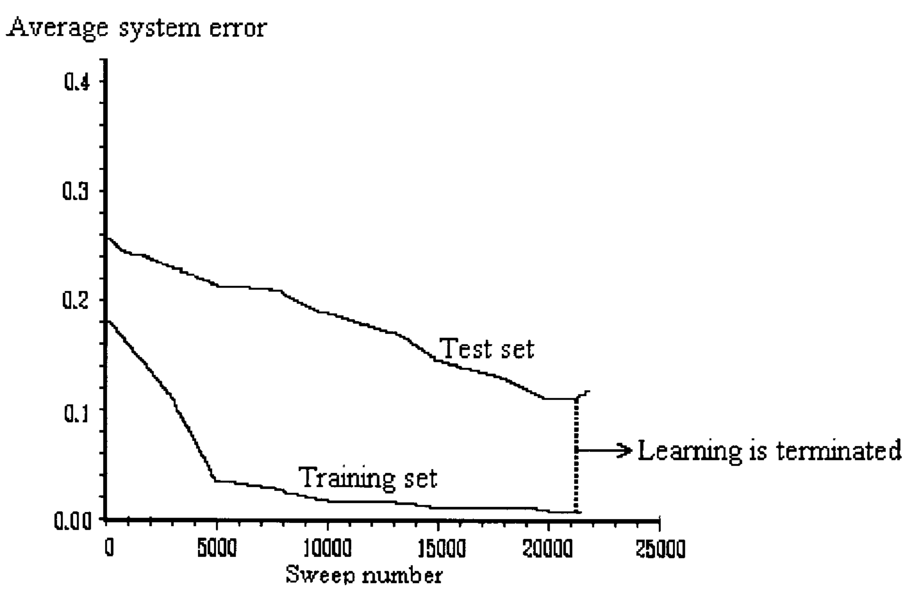


Figure 2. The behaviour of learning in an MLP. The learning is continued till the average system error (corresponding to both the training and test sets) decreases with the increase in the number of sweeps. It is terminated as the error corresponding to the test set of patterns starts increasing.

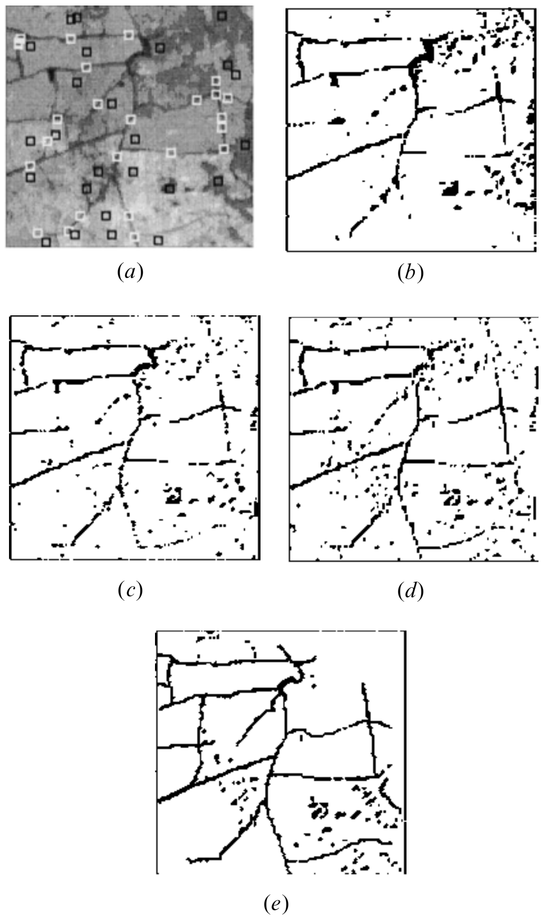


Figure 3. (*a*) A sub-image (near-infrared-band-IRS) of size 128 by 128 used for selecting training and test samples. The road and non-road pixels selected in the training set for learning of the concerned MLP network have been shown by enclosing them by square boxes of black and white colours respectively. (*b*) The result of segmentation of the image in figure 3 (*a*) using the MLP when average system error on the training set is 0·075. (*c*) The segmentation result when the error is 0·05. (*d*) The segmentation result when the error is 0·025. (*e*) The final result of segmentation on the image in figure 3 (*a*) after the training of the MLP is terminated.

algorithm involves self-adaptation of learning rate values satisfying the above set of observations. In this method, each of the learning rates is adapted dynamically by performing steepest descent on the surface defined by equation (1). Using the self-adaptive learning rates, the weight modification rules given by equations (5) and (6), become

$$w_{jk}(t+1) = w_{jk}(t) - \beta_{jk} \frac{\partial E_p(t)}{\partial w_{jk}(t)} + \alpha \Delta w_{jk}(t-1) \tag{7}$$

$$w_{ij}(t+1) = w_{ij}(t) - \beta_{ij} \frac{\partial E_p(t)}{\partial w_{ij}(t)} + \alpha \Delta w_{ij}(t-1) \tag{8}$$

where $\beta_{jk} = h(\eta_{jk})$ and $\beta_{ij} = h(\eta_{ij})$. $h(x) = d/1 + e^{-x+d/2}$ is called the effective value

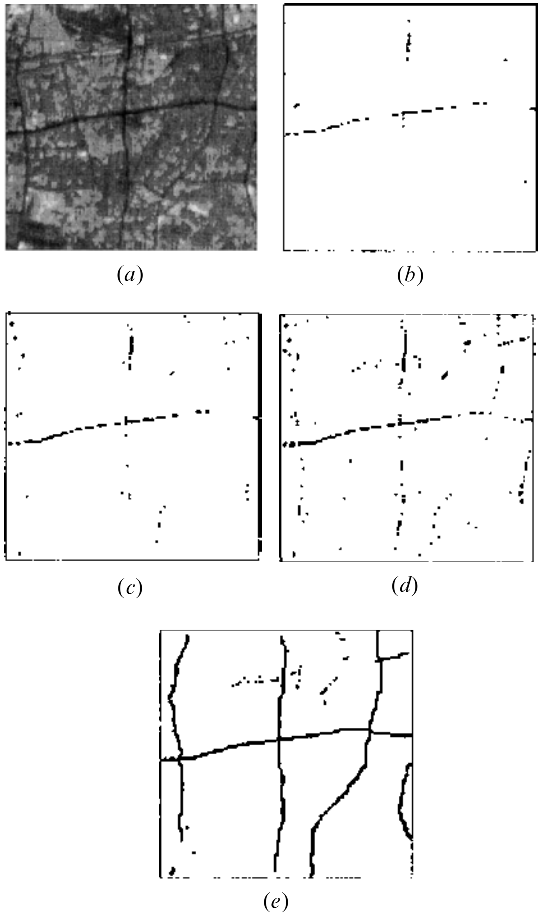


Figure 4. (*a*) A sub-image (near-infrared-band-IRS) of size 128 by 128 used for showing the effectiveness of the MLP classifier when a part of the image is considered from where no training or test samples have been selected. (*b*) The result of segmentation on the image in figure 4 (*a*) using the MLP when average system error on the training set is 0·075. (*c*) The segmentation result when the error is 0·05. (*d*) The segmentation result when the error is 0·025. (*e*) The final result of segmentation on the image in figure 4 (*a*) after the training of the MLP is terminated.

function and $d(>0)$ is a constant. The effective value function has been considered to avoid the danger of self-adaptive learning rates growing very large (Bhattacharya and Parui 1995 b). The modification rules for learning rates are:

$$\eta_{jk}(t+1) = \eta_{jk}(t) + \Delta_p \eta_{jk}(t) \tag{9}$$

and

$$\eta_{ij}(t+1) = \eta_{ij}(t) + \Delta_p \eta_{ij}(t) \tag{10}$$

where

$$\Delta_p \eta_{jk}(t) = \frac{\gamma}{d} \frac{\partial E_p(t)}{\partial w_{jk}(t)} \frac{\partial E_p(t-1)}{\partial w_{jk}(t-1)} \beta_{jk}(d - \beta_{jk})$$

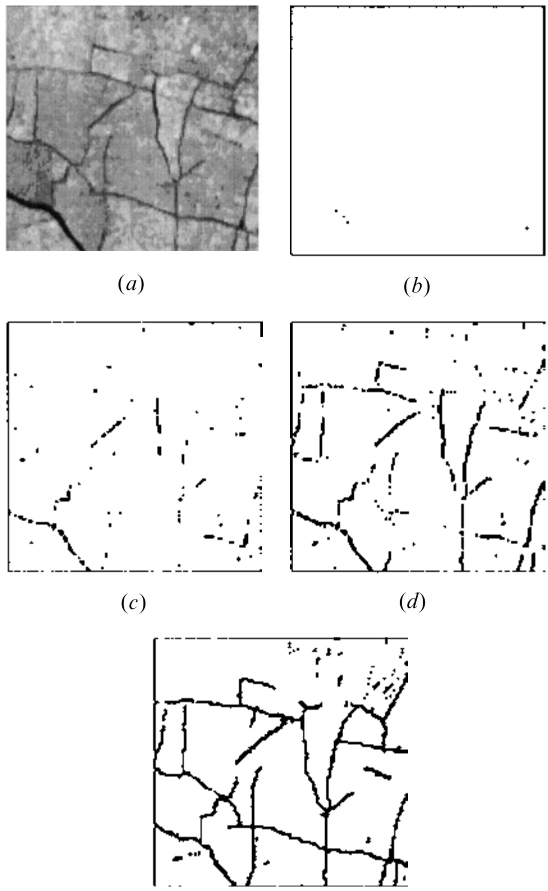


Figure 5. (*a*) Another sub-image (near-infrared-band-IRS) of size 128 by 128 used for showing the effectiveness of the MLP classifier when a part of the image is considered from where no training or test samples have been selected. (*b*) The result of segmentation on the image in figure 5 (*a*) using the MLP when average system error on the training set is 0·075. (*c*) The segmentation result when the error is 0·05. (*d*) The segmentation result when the error is 0·025. (*e*) The final result of segmentation on the image in figure 5 (*a*) after the training of the MLP is terminated.

and

$$\Delta_p \eta_{ij}(t) = \frac{\gamma}{d} \frac{\partial E_p(t)}{\partial w_{ij}(t)} \frac{\partial E_p(t-1)}{\partial w_{ij}(t-1)} \beta_{ij}(d - \beta_{ij})$$

where $\gamma > 0$ is a constant of proportionality. The initial learning rate $\eta_{\text{start}} > 0$ is taken to be the same for all the connection weights. They are allowed to differ during learning.

4.1.4. *Termination of the training session*

Overtraining is a common problem of backpropagation algorithm and setting up of an appropriate condition for the termination of the training session is very important. To avoid overtraining, we have used another set of patterns called the test set. The test set has been constructed in a manner similar to the training set.

The test set of patterns has been used as an indicator to determine the termination point of the training session. During the training session, the system errors on the training and test set is computed after regular intervals of time, for five consecutive learning sweeps. Initially, it is found that this error on both the training and test sets is decreasing. But after a period of learning, the error on the test set starts increasing though the error on the training set still decreases (figure 2). The point of time when the error on the test set increases for at least three consecutive sweeps for the first instance is noted and the weight values before the error started increasing, are stored.

4.2. *Network architecture selection*

Only one hidden layer in the MLP network is used for the present problem of detection of road-like structures. In fact, it is seen that the use of multiple hidden layers does not improve and sometimes even deteriorates the training performance of the network in terms of its rate of convergence. We have studied in detail the selection of the number of nodes in the input and hidden layers of the network. The input pattern consists of pixel grey values within a square window and hence the number of input nodes is specified by the size of such a window. As the present problem is a two-class problem, we can easily use only a single node at the output layer. For quantizing the output levels of the two classes, the following step is taken— if the computed value at the sigmoidal output node is less than 0·5, it is class 0 and otherwise, it is class 1.

4.2.1. *Selection of input window size*

The choice of input window size is an important aspect here. We have studied the effect of various window sizes on the performance of the proposed method using SPOT and IRS images and observed that even 3 by 3 windows provide acceptable results for IRS images while 5 by 5 windows are necessary for SPOT images.

4.2.2. *Selection of the number of hidden nodes*

The capacity of an MLP to approximate a given mapping has been investigated by Hornik (1989). But in real-life applications of such a network, a good choice of the number of hidden units cannot in general be obtained in a straightforward manner and the selection of this number involves conflicting interests (Kruschke 1988). A small number of hidden units is good with respect to the generalization, computational efficiency and interpretation of network. But an MLP with a restricted number of hidden units often does not converge. So, a standard approach to this problem is to start with a large hidden layer and then to reduce its size. Several researchers have proposed a number of ways for the reduction of the size of the hidden layer. In the present work we considered the approach of Hagiwara (1994) for the removal of superfluous hidden units. In this approach, 'weights power' for the jth unit in the hidden layer is defined as follows:

$$w_j = \sum_i (w_{ij})^2 + \sum_k (w_{jk})^2 \tag{11}$$

The hidden unit having the least value of 'weights power' is considered as the least important unit and is removed. The network is then retrained. If the network can achieve a predetermined level of performance, removal of another hidden unit takes place and the process is continued until the network cannot achieve the desired degree of performance in which case the network and the connection weights before the last removal are final.

5. Simulation results

We have simulated the proposed MLP-based technique for detection of road-like structures on images obtained from IRS and SPOT satellites. Images from IRS-1A satellite consists of four spectral bands of which three are in the visible range (0·45–0·52 μm, 0·52–0·59 μm and 0·62–0·68 μm) and one is in the near-infrared range (0·77–0·86 μm). In our study, we have considered only the near-infrared band of IRS images, since the information on linear structures is not significant in the first three bands. In the case of SPOT, PLA images are considered which are only single band. The spatial resolutions for IRS and SPOT images are 36·25 m and 10 m respectively.

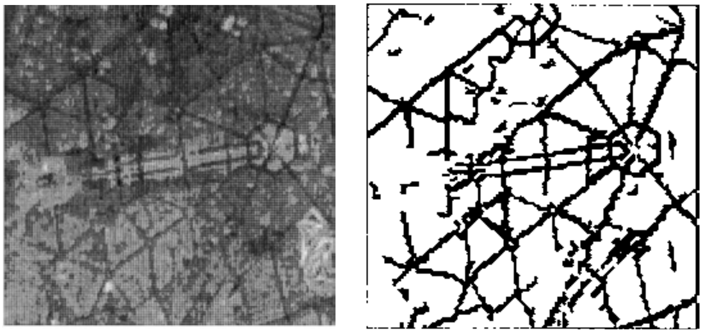
For supervised learning of the MLP network, we have used the modified back-propagation algorithm and also compared its training performance with that of the original algorithm. The network should be separately trained for images from different satellites and this is due to the difference in sensor parameters.

In figures 3 (*a*), 4 (*a*) and 5 (*a*) three sub-images (near-infra-red-band-IRS), each of size 128 pixels by 128 pixels, are shown. Both the training and test sets of patterns have been selected from the sub-image in figure 3 (*a*) where training road and non-road pixels are indicated by small white and black squares respectively. (No training or test patterns have been selected from images in figures 4 (*a*), 5 (*a*).) After successful termination of the training process, the network is used as a classifier and the final results of classification on the three images in figures 3 (*a*), 4 (*a*) and 5 (*a*) are shown in figures 3 (*e*), 4 (*e*) and 5 (*e*) respectively. At the time of termination of training, the system error is 0·0094. We also saved the network when the system error was 0·075, 0·05 and 0·025. The classification results of these three networks are shown in (*b*), (*c*) and (*d*) respectively for each of the input images 3 (*a*), 4 (*a*) and 5 (*a*).

For IRS images input windows of size 3 by 3 provide acceptable classification performance. Using 5 by 5 windows cannot improve the performance considerably while it increases the computational burden significantly. So, in such cases, the choice of 3 by 3 windows is suggested. The results in figures 3, 4 and 5 correspond to the choice of 3 by 3 input windows. Further, as for the selection of hidden layer size, we have always started with seven hidden nodes. And after applying Hagiwara's hidden node reduction technique, the final size of the hidden layer is reduced to two only.

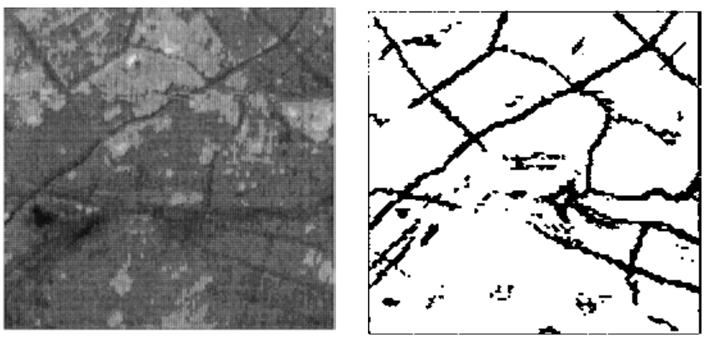
For the SPOT images as in figures 6 (*a*) and 7 (*a*), we have formed both the training and test sets of patterns from the image in figure 6 (*a*) and no training or test patterns have been taken from the image in figure 7 (*a*). In this case, 3 by 3 input windows failed to provide good training performance. So, we selected a larger input window of size 5 by 5. To start with, we considered 20 hidden nodes and after applying the above hidden node reduction technique, it is also reduced to four only. The final classification results are shown in figures 6 (*b*) and 7 (*b*) respectively. The system error at termination of the training process here is 0·012.

The classification results shown in figures (3), (4), (5), (6) and (7) are obtained by training the network using our modified backpropagation algorithm. We have also trained the network with the original backpropagation algorithm and observed that though the classification performances in the two cases are comparable, the modified backpropagation algorithm needs less training time. The best choices for the learning parameter values of both the learning algorithms are given in table 1.



(*a*) (*b*)

Figure 6. (*a*) A sub-image of a PLA image (obtained from the French satellite SPOT) of size 128 by 128 used for selecting training and test set of patterns. (*b*) The result of segmentation on the image in figure 6 (*a*) using the MLP after the training is terminated.



(*a*) (*b*)

Figure 7. (*a*) Another sub-image of the PLA image (obtained from the French satellite SPOT) of size 128 by 128 used for showing the effectiveness of the MLP classifier when a part of the image is considered from where no training or test samples have been selected. (*b*) The result of segmentation on the image in figure 7 (*a*) using the MLP after the training is terminated.

Table 1. Best choices of learning parameter values.

	Original BP		Modified BP			
Satellite	η	α	η_{start}	α	d	γ
IRS-1A	0·5	0·25	0·9	0·7	1·0	0·1
SPOT	0·25	0·1	0·5	0·25	1·0	0·1

These best choices have been determined on the basis of the average performance over 20 different random initial starts in the weight space.

Here it is to be noted that the convergence of the original backpropagation algorithm depends heavily on the proper choice of the learning parameter values (values of η and α) but the modified one is robust with respect to the choice of these parameter values (η_{start} and α). The detailed convergence performance of the two algorithms is shown in table 2. (In all the trials, the values of d and γ are taken as 1 and 0·1 respectively for the modified backpropagation algorithm.) The second

Table 2. Learning statistics.

Satellite	η/η_{start}	α	Average no. of sweeps: Original BP	Average no. of sweeps: Modified BP
IRS-1A	0·1	0·1	19 704	4333
		0·25	18 211	4121
		0·7	8000	3765
	0·25	0·1	11 417	4012
		0·25	11 166	3887
		0·7	6345	3612
	0·5	0·1	8574	3665
		0·25	5945	3587
		0·7	9112	3112
	0·9	0·1	12 333	2845
		0·25	12 988	2222
		0·7	15 843	1668
SPOT	0·1	0·1	21 532	6552
		0·25	20 777	6322
		0·7	18 045	5987
	0·25	0·1	14 980	6225
		0·25	16 622	5224
		0·7	17 055	4155
	0·5	0·1	15 221	3325
		0·25	17 550	2765
		0·7	18 114	3126
	0·9	0·1	16 221	3525
		0·25	16 872	3845
		0·7	22 943	5111

column of this table shows the values of η and η_{start} for the original and modifed backpropagation algorithms respectively. From this table it can be seen that the average number of sweeps needed to reach the termination of learning, is much smaller in the modified backpropagation than in the original backpropagation. The average is taken over 20 different random initial weight vectors. These results are based on the training sets and test sets obtained from images in figures 3 (*a*) and 6 (*a*).

6. Conclusions

In the present approach, a human operator can select and label a few training samples from road-like and non-road structures in a satellite image obtained from a particular sensor. The network is then trained on these samples. The trained network can extract road-like features from other image data obtained from the same sensor.

The backpropagation training algorithm is sometimes rather slow and in this paper, we have proposed an improvement of the backpropagation algorithm through self-adaptation of learning rates. Aside from the simplicity involved in training the network in this application, it can provide an acceptable labelling of images which exhibit complex distributions in the original signal space. We are currently working on post-processing of the output image for removing noise elements and linking road segments.

References

AUGUSTEIJN, M. F., CLEMENS, L. E., and SHAW, K. A., 1995, Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural network classifier. *IEEE Transactions on Geoscience and Remote Sensing*, **33,** 616–626.

BAJCY, R., and TAVAKOLI, M., 1976, Computer recognition of roads from satellite pictures. *IEEE Transactions on Systems, Man and Cybernetics*, **6,** 623–637.

BENEDIKTSSON, J. A., SWAIN, P. H., and ERSOY, O. K., 1990, Neural network approaches versus statistical methods in classification of multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, **28,** 540–541.

BHATTACHARYA, U., and PARUI, S. K., 1995 a, On the rate of convergence of perceptron learning. *Pattern Recognition Letters*, **16,** 491–497.

BHATTACHARYA, U., and PARUI, S. K., 1995 b, Self-adaptive learning rates in backpropagation algorithm improve its function approximation performance. *Proceedings of the IEEE International Conference on Neural Networks, Australia, December 1995* (San Diego: IEEE), pp. 2784–2788.

BISCHOF, H., SCHNEIDER, W., and PINZ, A. J., 1992, Multispectral classification of landsat images using neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, **30,** 482–490.

CIVCO, D. L., 1993, Artificial neural networks for land-cover classification and mapping. *International Journal of Geographic Information Systems*, **7,** 173–186.

DREYER, P., 1993, Classification of land cover using optimized neural nets on SPOT data. *Photogrammetric Engineering and Remote Sensing*, **59,** 617–621.

FISCHLER, M. A., TANENBAUM, J. M., and WOLF, H. C., 1981, Detection of roads and linear structures in low-resolution aerial imagery using a multisource knowledge integration technique. *Computer Graphics and Image Processing*, **15,** 201–223.

GROCH, W. D., 1982, Extraction of line shaped objects from aerial images using a special operaor to analyze the profiles of functions. *Computer Graphics and Image Processing*, **18,** 347–358.

HAGIWARA, M., 1994, A simple and effective method for removal of hidden units and weights. *Neurocomputing*, **6,** 207–218.

HECHT-NIELSEN, R., 1990, *Neurocomputing* (Reading, MA: Addison Wesley).

HEERMAN, P. D., and KHAZENIE, N., 1992, Classification of multispectral remote sensing data using a backpropagation neural network. *IEEE Transactions on Geoscience and Remote Sensing*, **30,** 81–88.

HEPNER, G. F., 1990, Artificial neural network classification using a minimal training set: Comparison to conventional supervised classification. *Photogrammetric Engineering and Remote Sensing*, **56,** 469–473.

HORNIK, K., STINCHCOMBE, M., and WHITE, H., 1989, Multilayer feedforward networks are universal approximators. *Neural Networks*, **2,** 359–366.

JACOBS, R. A., 1988, Increased rates of convergence through learning rate adaptation. *Neural Networks*, **1,** 295–307.

KRUSCHKE, J. K., 1988, Creating local and distributed bottlenecks in hidden layers of back-propagation networks. *Proceedings of the 1988 Connectionist Models Summer School*, edited by D. Touretzky, G. Hinton, and T. Sejnowski, (San Mateo, CA: Morgan Kaufmann), pp. 120–126.

LEE, J., WEGER, R. C., SENGUPTA, S. K., and WELCH, R. M., 1990, A neural network approach to cloud classification. *IEEE Transactions on Geoscience and Remote Sensing*, **28,** 846–855.

LIPPMANN, R. P., 1987, An introduction to computing with neural networks. *IEEE ASSP Magazine*, April, 4–22.

LIU, Z. K., and XIAO, J. Y., 1991, Classification of remotely-sensed image data using artificial networks. *International Journal of Remote Sensing*, **12,** 2433–2438.

MCKEOWN, D. M., JR., and DENLINGER, J. L., 1988, Cooperative methods for road tracking in aerial imagery. *DARPA, Image Understanding Workshop, 1988* (San Mateo: Morgan Kaufmann), **1,** 327–341.

MUKHERJEE, A., PARUI, S. K., CHAUDHURI, B. B., and KRISHNAN, R., 1994, Detection of linear features in satellite imagery using robust estimation. *Proceedings of 12th International*

Conference on Pattern Recognition, Jerusalem, 1994 (California: IEEE Computer Society Press), **1,** 514–516.

Mukherjee, A., Parui, S. K., Chaudhuri, D., Chaudhuri, B. B., and Krishnan, R., 1996, An efficient algorithm for detection of road-like structures in satellite images. *Proceedings of 13th International Conference on Pattern Recognition, Austria, 1996* (California: IEEE Computer Society Press), **III,** 875–879.

Paola, J. D., and Schowengerdt, R. A., 1995, A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery, *International Journal of Remote Sensing*, **16,** 3033–3058.

Parui, S. K., Umashankar, B., Mukherjee, A., and Dutta Majumder, D., 1991, A parallel algorithm for detection of linear structures in satellite images. *Pattern Recognition Letters*, **12,** 765–770.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J., 1986, Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*, edited by D. E. Rumelhart and J. L. McClelland (Cambridge, MA: The MIT Press), pp. 318–362.

Schaale, M., and Furrer, R., 1995, Land surface classification by neural networks. *International Journal of Remote Sensing*, **16,** 3003–3031.

Vanderbrug, G., and Rosenfeld, A., 1978, Linear feature mapping. *IEEE Transactions on Systems, Man and Cybernetics*, **SMC-8,** 768–774.

Vasudevan, S., Cannon, R. L., Bezdek, J. C., and Cameron, W. L., 1988, Heuristics for intermediate level road finding algorithms. *Computer Vision, Graphics, and Image Processing*, **44,** 175–190.

Wilkinson, G. G., Kanellopoulos, I., Kontoes, C., and Megier, J., 1992, A comparison of neural network and expert system methods for analysis of remotely-sensed imagery. *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS'92), Houston, TX, May 1992* (Piscataway, NJ: IEEE), pp. 62–64.

Zhu, M. L., and Yeh, P. S., 1986, Automatic road network detection on aerial photographs. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1986* (California: IEEE Computer Society Press), pp. 1986.