Comparisons of Neural Networks to Standard Techniques for Image Classification and Correlation

Justin D. Paola
Robert A. Schowengerdt

The Research Institute for Advanced Computer Science
is operated by Universities Space Research Association,
The American City Building, Suite 212, Columbia, MD 21044, (410) 730-2656

Work reported herein was partially supported by the National Aeronautics and Space Administration under Contract NAS 2-13721 to the Universities Space Research Association (USRA) and under Grant NAG 5-2198 to the University of Arizona Department of Electrical and Computer Engineering.

ABSTRACT

Neural network techniques for multispectral image classification and spatial pattern detection are compared to the standard techniques of maximum-likelihood classification and spatial correlation. The neural network produced a more accurate classification than maximum-likelihood of a Landsat scene of Tucson, Arizona. Some of the errors in the maximum-likelihood classification are illustrated using decision region and class probability density plots. As expected, the main drawback to the neural network method is the long time required for the training stage. The network was trained using several different hidden layer sizes to optimize both the classification accuracy and training speed, and it was found that one node per class was optimal. The performance improved when 3x3 local windows of image data were entered into the net. This modification introduces texture into the classification without explicit calculation of a texture measure. Larger windows were successfully used for the detection of spatial features in Landsat and Magellan synthetic aperture radar imagery.

INTRODUCTION

The neural network method has a potential advantage over maximum-likelihood in that it does not require any assumption about the underlying statistics of the data (Lippmann, 1987). Another feature of neural networks is the ease with which the algorithm can be adapted to handle the significantly different problem of spatial pattern detection. This paper presents a comparison of the neural network and maximum-likelihood classifiers for a sample classification problem, followed by two neural network spatial pattern detection examples, the second of which focuses on the use of a threshold to provide simple true/false results from the detection process.

MULTISPECTRAL CLASSIFICATION

Maximum-Likelihood Classifier

The maximum-likelihood multispectral image classifier assumes a particular multivariate probability density (usually Gaussian) in order to form a discriminant function for each class. These functions are then applied to each unknown pixel and the pixel is assigned to the class with the highest discriminant value. If the assumption of a normal distribution for each class is correct, then the classification has a minimum overall probability of error and the maximum-likelihood classifier is the optimal choice (Swain, 1978).

Neural Network Classifier

Operation of the neural network classifier is much like that of any standard classifier. Instead of estimating the class statistics from the training data, however, the interconnecting weights of the network nodes are adjusted in an iterative fashion, typically by the backpropagation method (e.g., Rumelhart *et al.*, 1986), until some targeted minimal error is achieved between the desired output (the training classes) and actual output of the network. For the classification phase, instead of calculating discriminant functions, as in maximum-likelihood, the network is used in a feed-forward mode like a hard-wired circuit. The entire image is fed into the net pixel-by-pixel, and a simple metric (such as the maximum) is used to process the network output to make a class selection for each pixel.

Classification Map Comparison

A standard backpropagation neural network was implemented with a slight modification to the training procedure to increase training speed. At fixed intervals during training the learning rate and momentum terms (see Rumelhart *et al.*, 1986) were adjusted based on the change in mean square error of the output pattern from that of the previous iteration. If the error increased, these values were reduced for subsequent training iterations, and if the error decreased, they were made larger. This allowed for accelerated convergence when the error was steadily decreasing, and led to faster and more stable training. It also decreased the importance of the initial values set by the user for these parameters.

The image data consisted of the six non-thermal bands of a Landsat Thematic Mapper scene of Tucson, AZ, acquired on April 1, 1987. Separate training and test regions were defined for each of 12 classes. A maximum-likelihood classifier was used to produce the map of Figure 1. This figure illustrates three of the classes in different shades of gray: white for 'urban residential', light gray for 'desert scrub', dark gray for 'foothills natural vegetation', and black for all other classes. There are some substantial errors in these classes. The 'urban residential' class is far too prevalent, particularly at the expense of the two vegetation classes. The classification accuracy of the test sites was 89.5%.

The map shown in Figure 2 was produced after 50,000 training iterations by a neural network with 6 input nodes, 18 hidden layer nodes, and 12 output nodes. The hidden layer size was set to 18 because that results in the same number of classification parameters as for maximum-likelihood. Thus, each classifier has the same number of unknowns (mean and covariance for maximum-likelihood, interconnecting weights for the neural network) with which to define the decision regions. The test site accuracy of the neural network

classification was 93.4%. The network did not have the same problem with the 'urban residential' class as maximum-likelihood. Both qualitatively and quantitatively, the neural network map better reflects the actual distribution of classes in the image.

Figure 1. Maximum-likelihood classification. See text for classes.

Decision Boundaries and Class Probability Plots

Visualization of the decision regions produced by the two classifiers leads to a better understanding of their capabilities. In order to view the decisions made over the entire feature space, the dimensionality of the original image was reduced to two bands, the red and near infrared. Both classifiers achieved a 77% test site accuracy with this reduced dataset, and the error with respect to the 'urban residential' class was still observed in the maximum-likelihood classification. The decision regions for the two methods are shown in Figure 3. In the maximum-likelihood case this plot shows the highest surfaces of the

intersecting probability density functions for each class. Figure 4 shows the probability densities of the 'urban residential' and 'desert scrub' classes on the same scale. Figure 5 contains plots of the neural network output values for the same two classes. For all three figures, the vertical axis represents the red band and the horizontal axis represents the NIR band.

Figure 2. Neural network classification. See text for classes.

As expected, spectral correlation results in most of the classes being clustered along the diagonal of the two spectral bands. The 'urban residential' class, which in reality is a mixture of many different surface types, has a high variance, and thus a wider, lower, probability density than the surrounding classes, as seen in Figure 4. Thus, in the maximum-likelihood classification, other classes are chosen only where their relatively narrow probability density functions protrude above the wide 'urban residential' density. This leads to the decision region (Figure 3) for 'urban residential' surrounding that of 'desert

scrub' and nearly surrounding that of 'foothills natural'. The same problem occurs in the 6-dimensional case and is responsible for the errors discussed previously.

The neural network produces a fundamentally different type of classification. The training data for a particular class in the maximum-likelihood case affects the statistics of the training class only. Training data in the neural network method, however, is used not only to boost the desired class output, but to suppress the other classes. Thus, the network produces mutually exclusive "probability" densities (Figure 5) that do not overlap nearly as much as those of maximum-likelihood. This gives the network the ability, in this case, to differentiate between classes that are difficult to separate using first and second order statistics alone.

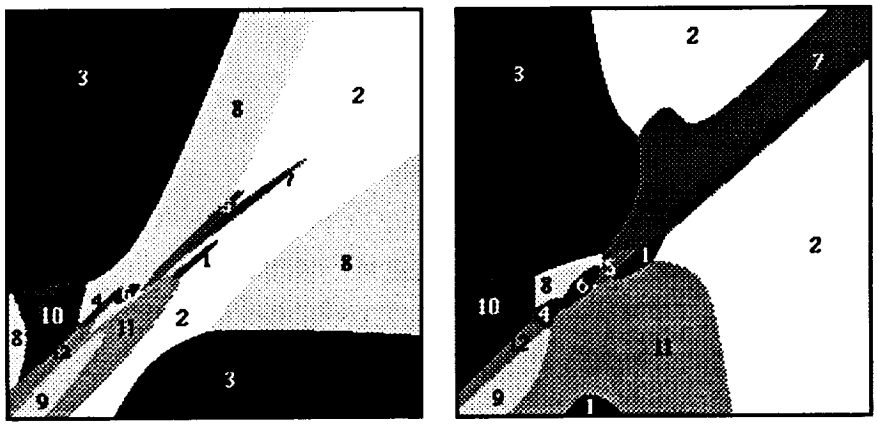

Figure 3. Decision regions for ML (left) and neural net (right).

1-Tarmac 2-Building 3-Grass 4-Foothills Natural 5- Sand
6-Desert Scrub 7-Bare Soil 8-Urban Residential 9-Asphalt
10-Riparian 11-Dense Urban 12-Shaded Foothills Natural

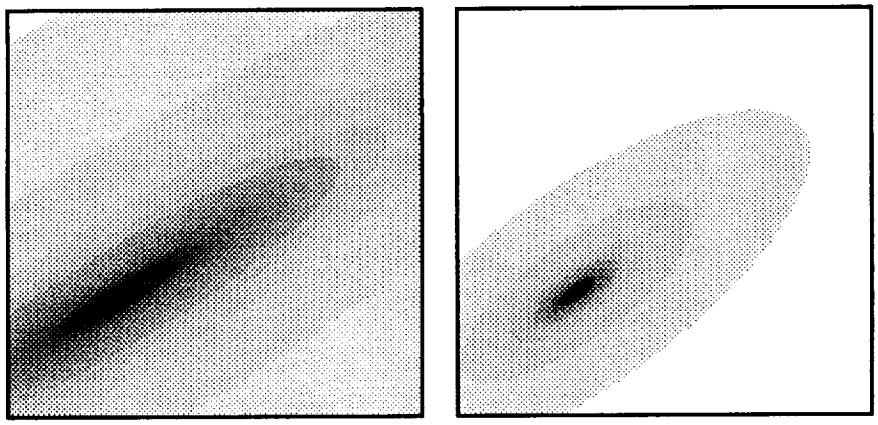

Figure 4. Probability density for 'urban residential', 'desert scrub'.

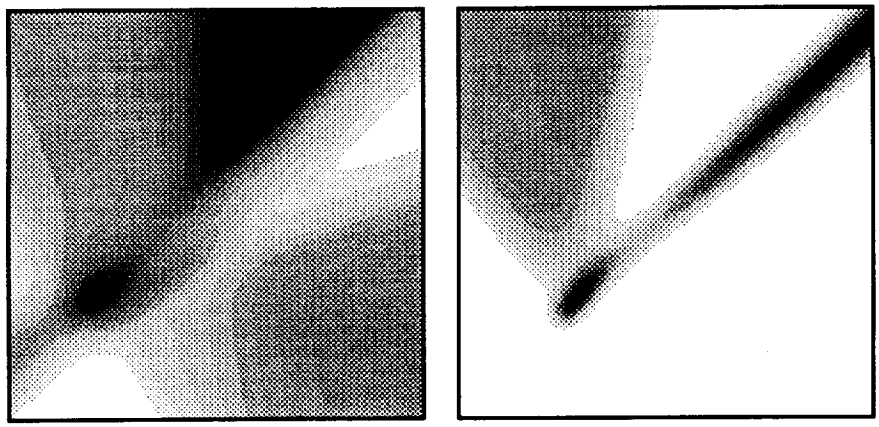

Figure 5. Neural net output for 'urban residential', 'desert scrub'.

Classification Speed

The two main disadvantages of the neural network classification method are the need for user supplied initialization variables and the length and inconsistency of the training phase. The main variable in using a neural network is the size of the hidden layer. It was shown previously that selecting one that gave the same number of classification parameters as maximum-likelihood was adequate. The maximum-likelihood classification took 590 seconds on a SUN Sparcstation 10. The 18 node hidden layer network was trained for 50,000 iterations, which required 53,605 seconds. Another 385 seconds were required for the classification stage. Figure 6 is a plot of the test site accuracy versus training iterations. It is shown in this plot that only 8000 training iterations were needed to achieve the accuracy of maximum-likelihood.

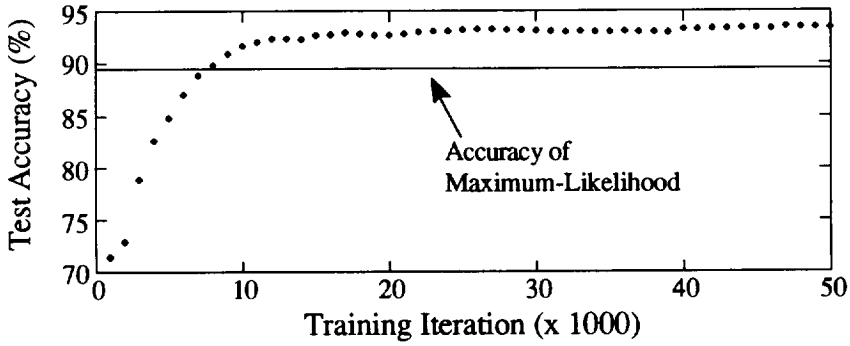

Figure 6. Test site accuracy for an 18 node hidden layer network.

Seven neural network training runs were carried out to 20,000 iterations for each of several hidden layer sizes ranging from 1 to 36 nodes with the hope that an optimal network structure would be found that would further reduce overall classification time. The multiple runs were necessary to investigate the repeatability of the procedure, since the network weights are initialized randomly, resulting in different paths for error convergence during each training run. Figure 7 shows the range and average test site accuracy of the seven runs for each hidden layer size. It is evident from this plot that a hidden layer size of 6 nodes and above can give accurate results. Also, it is clear that the consistency of the training phase increases with hidden layer size. Based on these results, the optimal choice for hidden layer size would be 12 nodes, which would give accurate, consistent results in less time than the 18 node case.

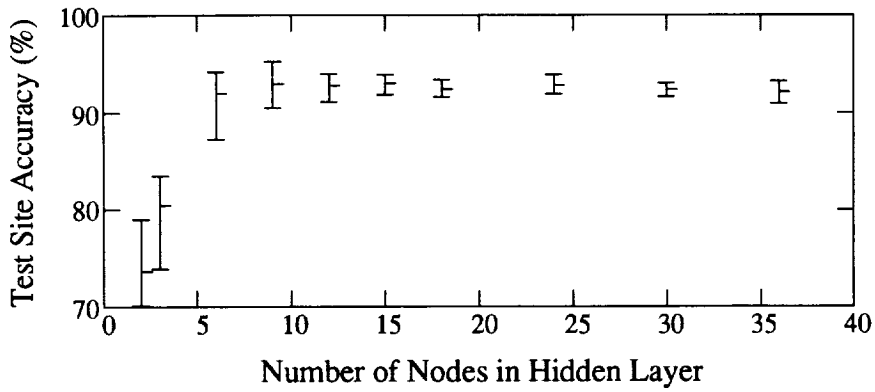

Figure 7. Test site accuracies obtained after 20,000 iterations.

Local Texture

The neural network structure is easily adaptable to include additional inputs, such as ancillary data or windows of pixels. The Tucson image classification was repeated using a 3x3 window of pixels in each band, with the hope that the texture information introduced by this window would result in a better classification. After 25,000 training iterations, an 18 node hidden layer network achieved a test site accuracy of 96.3%. The surprising result was that the network required only about 3,500 iterations and 6,000 seconds to achieve the accuracy of maximum-likelihood, despite a large increase in input nodes (from 6 to 54). This is about the same time as required for the 12 node hidden layer single pixel case.

SPATIAL PATTERN DETECTION

Another problem of interest in remote sensing is that of spatial pattern detection. Figure 8 shows the results of a detection of street intersections using a 49x49 window on band 4 (NIR) of the same Tucson image used for the multispectral classification. The network had 2401 input nodes, 2 hidden layer nodes, and 1 output node. The output was trained to a high value with four sample intersections, three of which are outlined on the original (upper) image. Sixteen non-matching patterns were used to train the output to a competing low value. The detection peaks created at the training sites are indicated by the arrows in the lower image, a map of network output.

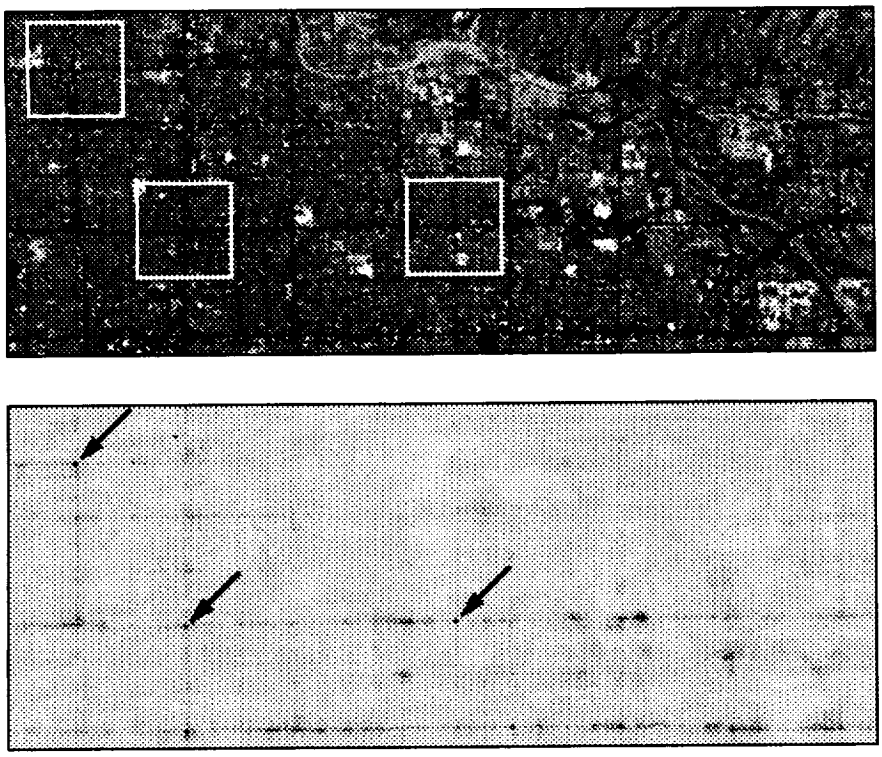

Figure 8. Detection of street intersections in TM imagery.

A second example involving the detection of craters in Magellan synthetic aperture radar (SAR) imagery of Venus illustrates the ability to apply a threshold to the network output and obtain true or false responses to the detection problem. Neural networks with different hidden layer sizes were trained for 10,000 iterations using 25x25 and 35x35 windows on 15 craters and 52 non-crater sites extracted from the browse images of Compressed-Once Mosaic Image Data Record (C1-MIDR) imagery. An additional 11 craters were then used to test the detection. To achieve similar results with a standard technique such as correlation, each training crater would have to be correlated separately, and the resulting correlation values combined in some way to give a final measure for each image pixel. This would involve a great deal of computer processing (Hall, 1979).

The top image in Figure 9 is a SAR image containing one of the training craters (top right) and one of the test craters (bottom left). The image at the lower left of the figure is the output of a 25x25 window, 2 node hidden layer neural network. The image at the lower right is the result of a true/false detection obtained by applying a threshold of 0.9 to the image on the left. This corresponds to the trained output value for the craters. If the threshold is lowered, more of the test craters are detected, at the expense of false detections

in the areas around the craters. The results of this exercise are summarized in Table 1. Depending on the desired level of detection versus false (bad) detections the threshold can be adjusted away from the "ideal" detection value (i.e., the trained output value).

Table 1. *True and false crater detections as a function of threshold.*

	25x25 win 2 h.l. nodes	25x25 win 5 h.l. nodes	25x25 win 8 h.l. nodes	25x25 win 15 h.l. nodes	35x35 win 5 h.l. nodes
Thr = 0.9	5/11, 0 bad	5/11, 0 bad	5/11, 0 bad	5/11, 0 bad	2/11, 0 bad
Thr = 0.86	6/11, 0 bad	6/11, 0 bad	6/11, 0 bad	5/11, 0 bad	9/11, 0 bad
Thr = 0.82	7/11, 2 bad	7/11, 1 bad	6/11, 1 bad	6/11, 0 bad	9/11, 1 bad
Thr = 0.78	9/11, 3 bad	10/11,2 bad	6/11, 3 bad	6/11, 1 bad	9/11, 2 bad
To detect all 11	Thr = 0.77 4 bad	Thr = 0.76 6 bad	Thr = 0.73 6 bad	Thr = 0.69 9 bad	Thr = 0.73 5 bad
Train Time	3623 sec	9170 sec	14,336 sec	26,531 sec	35,132 sec

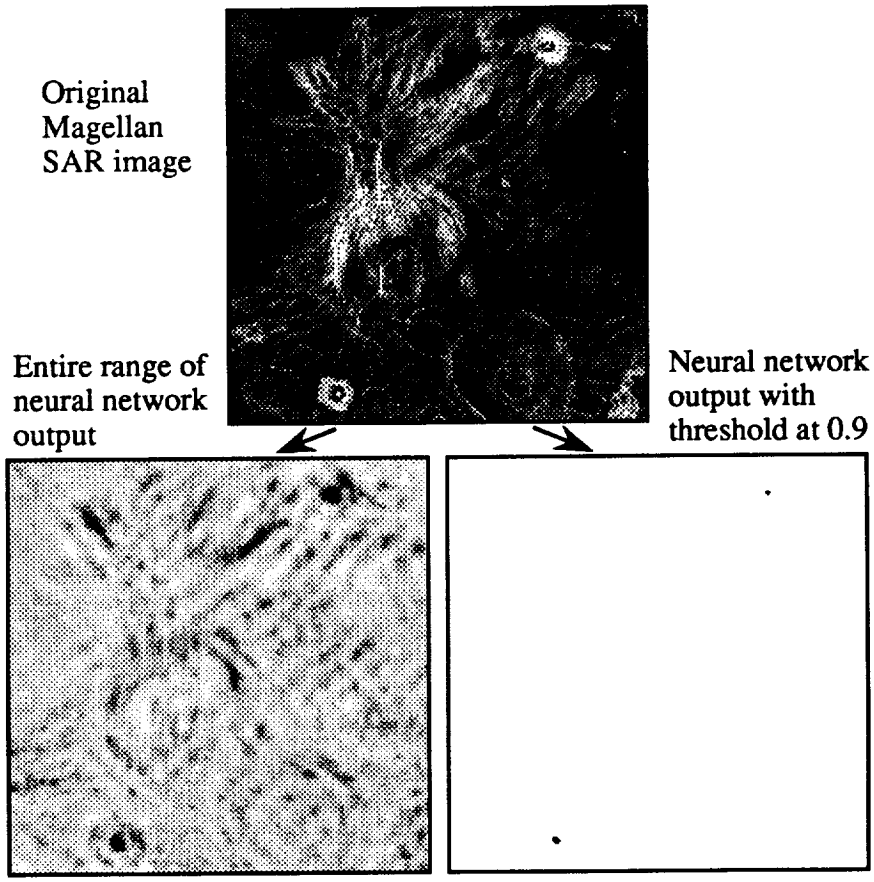

Figure 9. Crater detection in Magellan SAR imagery.

SUMMARY

A single neural network classifier can be applied to the very different problems of multispectral image classification and spatial pattern detection. The results of the multispectral classification were comparable to maximum-likelihood. It was found that for the sample Landsat TM image, the neural network was better able to differentiate classes with widely different variances, which can cause problems for the maximum-likelihood classifier. Incorporation of a 3x3 window of inputs in each band improved the accuracy without increasing the training time of the neural network. This method was then extended to the use of larger windows for the detection of spatial patterns in Landsat and Magellan imagery. A threshold with a value near the trained output value for the pattern was used to provide true/false detection information. This technique could prove useful for content-based image searching applications.

REFERENCES

Hall, E.L. *Computer Image Processing and Recognition.* New York: Academic Press, 1979, p. 483.

Lippmann, R.P. "An Introduction to Computing with Neural Networks." *IEEE ASSP Magazine*, April 1987: 4-22.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations* (D.E. Rumelhart, and J.L. McClelland, Eds.). Cambridge, Mass: The MIT Press, 1986, pp. 318-362.

Swain, P.H. *Remote Sensing: The Quantitative Approach* (P.H. Swain and S.M. Davis, Eds.). New York: McGraw-Hill, 1978, pp. 137-187.