A HIERARCHICAL NEURAL NETWORK ARCHITECTURE THAT LEARNS TARGET CONTEXT: APPLICATIONS TO DIGITAL MAMMOGRAPHY

Paul Sajda, Clay Spence, and John Pearson

David Sarnoff Research Center
Princeton, NJ 08543-5300
sajda@sarnoff.com

1. INTRODUCTION

An important problem in image analysis is finding small objects in large images. The problem is challenging because 1) searching a large image is computationally expensive and 2) small targets (on the order of a few pixels in size) have relatively few distinctive features which enable them to be distinguished from non-targets. To overcome these challenges we have developed a hierarchical neural network architecture which combines multi-resolution pyramid processing with neural networks. The advantages of the architecture are 1) both neural network training and testing can be done efficiently, and at a reduced computational cost, through a coarse-to-fine paradigm and 2) such as system is capable of learning low-resolution *contextual information* to facilitate the detection of the small target objects.

We have previously reported on the improved accuracy of this architecture for problems in automatic target recognition (ATR)[1]. For example, we have shown that for the problem of detecting small buildings in aerial imagery, the hierarchical neural network architecture has higher accuracy than both conventional neural network architectures and standard statistical classification techniques (e.g. Fisher's linear discriminant). In this paper we discuss the application of our hierarchical neural network architecture to the problem of detecting microcalcifications in digital mammograms. Microcalcifications are cues for breast tumors. 30% to 50% of breast carcinomas have microcalcifications visible in mammograms while 60% to 80% of all breast tumors eventually show microcalcifications via histology [2]. Similar to the building/ATR problem, microcalcifications are generally very small point-like objects (< 10 pixels in mammograms) which are hard to detect. Radiologists must often exploit other information in the imagery (e.g. location of blood vessels, ducts, etc.) in order to detect these microcalcifications. In this paper, we examine how well our hierarchical neural network architecture learns and exploits contextual information in mammograms.

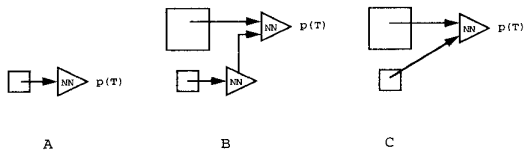

Figure 1: Three architectures that are compared in this paper.

2. GENERAL NETWORK ARCHITECTURES

Three neural network architectures are trained and tested, as indicated in Figure 1. The input to the networks are features at two different levels of an image pyramid (levels 2 and 3, with level 0 being full-resolution) with the outputs, $p(T)$, representing the probability that a target is present at a given location in the image. Network A consists of a single net processing data from level 3 features (one-eighth of the linear extent of the original image). Network B is a hierarchical architecture, constructed by adding to network A a second net processing level 2 data. In this architecture, information is propagated hierarchically, with the outputs of the hidden units of the level 3 net serving as inputs to the level 2 net. Network C is the "control" architecture. It is a single net that receives input directly from both level 2 and level 3.

This work was was funded under ARPA contract No. N00014-93-C-0202. and a grant from the Murray Foundation. Digital mammogram data was provided by Dr. Robert M. Nishikawa of The University of Chicago.

0-8186-7310-9/95 $4.00 © 1995 IEEE

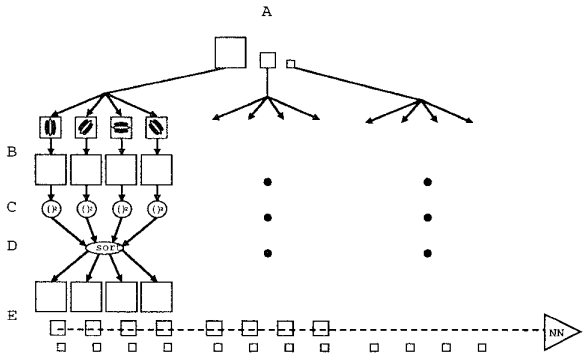

Figure 2: Construction of image features.

3. FEATURES

Input to the neural networks are image features constructed at several scales. Choosing features at multiple scales allows the network to take advantage of coarse-to-fine search and operate on only a small region of the entire image. We use sorted, oriented "energies" at several image scales as our features (see Figure 2) . These were derived by constructing the Gaussian pyramid of the mammogram (Figure 2A), and applying four oriented high-pass filters to each image of the pyramid (Figure 2B). The pixel values in these images are then squared to get the energies (Figure 2C). This ensures that when the resolution is reduced by low-pass filtering, the resulting image features are present [3]. Orientation-invariant features are constructed by sorting the energy images by their magnitude at each pixel location (Figure 2D). The resulting features are useful because the relative size of the minimum energy compared with the maximum energy indicates the degree to which the local image detail is oriented. Gaussian pyramids of these feature images are then computed (Figure 2E), with a neural network integrating the features across a given level.

4. SPECIFICS OF THE ARCHITECTURES AND TRAINING PROTOCOL

The neural nets used in the three architectures are multi-layer perceptrons, having one hidden layer with four hidden units. All units in a network perform a weighted sum of their inputs, subtracting an offset or threshold from that sum to get the activation;

$$a = \sum_i w_i x_i - \theta \tag{1}$$

This activation is transformed into a unit's output, y, by passing it through the sigmoid function;

$$y = \sigma(a) = \frac{1}{1+e^{-a}} \tag{2}$$

The networks are trained using a sequential quadratic programming optimization routine from the commercial Numerical Algorithms Group's library. The error function used was the cross-entropy error;

$$\epsilon(y,d) = -d\log y - (1-d)\log(1-y) \tag{3}$$

where $d \in \{0,1\}$ is the desired output. To obtain the objective function for the optimization routine, we computed the total error on the training examples, adding to it a regularization term;

$$r = \frac{\lambda}{2}\sum_i w_i^2 \tag{4}$$

This type of regularization is commonly referred to as "weight decay", and is used to prevent the network from becoming "over-trained". λ was adjusted to minimize the cross-validation error. Cross-validation error was computed by dividing the training data into five separate disjoint subsets, whose union is the entire set. The network was first trained on all of the training data, and then, starting from this set of weights, the network was retrained on the data with one of the five subsets left out. The resulting network was tested on the "holdout" subset. This retraining and testing with a holdout set was repeated for each of the five subsets, and the average of the five errors on the five subsets is the cross-validation error, an unbiased estimate of the average error on new data.

Each network receives as input a single pixel from the same location in each of the feature images at the resolution being searched (see figure 2). Network B also receives hierarchical contextual input (i.e. output of the hidden units of the level 3 net are inputs to the level 2 net). Network C receives inputs from level 2 features and also level 3 features (level 3 features are expanded to level 2 resolution). The output of each of the three networks is an estimate of the probability that a microcalcification is present at a given position, conditioned on its input. We trained the network on five mammograms, each of which had one or two clusters of about twenty microcalcifications, for a total of 97. The results given below were measured on five other mammograms with one cluster each, for a total of 95 microcalcifications.

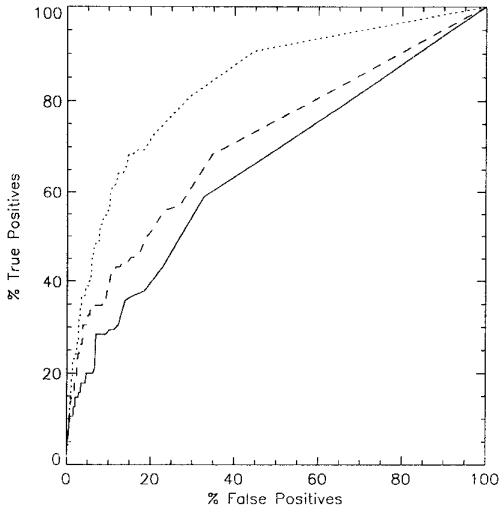

Figure 3: ROC curves for the three networks. Solid curve is for Network A, the dotted curve is for Network B (the hierarchical network with context inputs) and the dashed curve is for Network C.

5. RESULTS

Results for the three networks are shown as ROC curves in figure 3. These are the parametric curves of true-positive rate vs. false-positive rate, where the parameter being varied is the threshold on the network's output. Notice the improvement at the higher resolution level, and especially the very large improvement when using context (hierarchical architecture of Network B). We examined whether the system was in fact taking advantage of context information by looking at the representations developed by various hidden units in the network. Figure 4 shows examples of two classes of hidden units which were found. The first class (figure 4 left) appears to be representing point-like structure, similar to the structure of an individual microcalcification. The second class of hidden unit (figure 4 right) seems to be representing something very different. In this case, the unit is selective to long, extended, and oriented structure, which may be breast tissue, mammalian ducts or vasculature. It is widely known in mammography that clustered microcalcifications which are associated with carcinomas will tend to be located near ducts or blood vessels. Therefore the network appears to be extracting this context automatically, without any predefined notion of what the context might be.

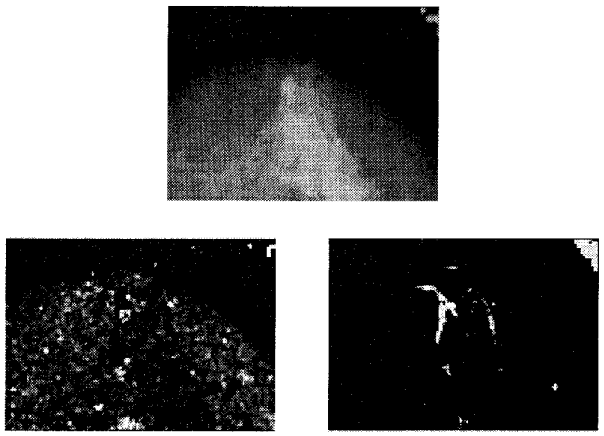

Figure 4: *(top)* Digital mammogram *(left bottom)* One class of hidden unit representing point-like structure. *(right bottom)* Second class of hidden unit representing elongated structure.

6. SUMMARY

We have described a hierarchical neural network architecture for detecting small targets in large images. We compared its detection accuracy with two other architectures for the problem of finding microcalcifications in digital mammograms. Using ROC analysis, we find that the hierarchical network has much better accuracy than its non-hierarchical counterparts, presumably due to the passing of contextual information from coarse-to-fine levels of processing. This learning of contextual information is supported by our findings that certain hidden units appear to represent information about the location of ducts and/or blood vessels, implying that the network utilizes context to increase microcalcification detection accuracy.

7. REFERENCES

[1] C. Spence, P. Sajda, S. Hsu and J. Pearson (1994) Neural network/pyramid architectures that learn target context, *Image Understanding Workshop, 1994* Monterey CA, pp. 853-862.

[2] R. Mills, R. Davis, and A. Stacey (1976) The detection and significance of calcification in the breast: A radiological and pathological study, *British Journal of Radiology*, 49, pp. 12-26.

[3] P. Burt (1988) Smart sensing with a pyramid vision machine, *Proceedings of the IEEE*, 76:8 pp. 1006-1015.