# Applications of multi-resolution neural networks to mammography

**Clay D. Spence and Paul Sajda**
Sarnoff Corporation
CN5300
Princeton, NJ 08543-5300
{cspence|psajda}@sarnoff.com

## Abstract

We have previously presented a coarse-to-fine hierarchical pyramid/neural network (HPNN) architecture which combines multiscale image processing techniques with neural networks. In this paper we present applications of this general architecture to two problems in mammographic Computer-Aided Diagnosis (CAD). The first application is the detection of microcalcifications. The coarse-to-fine HPNN was designed to learn large-scale context information for detecting small objects like microcalcifications. Receiver operating characteristic (ROC) analysis suggests that the hierarchical architecture improves detection performance of a well established CAD system by roughly 50%. The second application is to detect mammographic masses directly. Since masses are large, extended objects, the coarse-to-fine HPNN architecture is not suitable for this problem. Instead we construct a fine-to-coarse HPNN architecture which is designed to learn small-scale detail structure associated with the extended objects. Our initial results applying the fine-to-coarse HPNN to mass detection are encouraging, with detection performance improvements of 20% to 40%. We conclude that the ability of the HPNN architecture to integrate information across scales, both coarse-to-fine and fine-to-coarse, makes it well suited for detecting objects which make have contextual clues or detail structure occuring at scales other than the natural scale of the object.

## 1   Introduction

In a previous paper [6] we presented a coarse-to-fine hierarchical pyramid/neural network (*HPNN*) architecture that combines multi-scale image processing tech-

niques with neural networks to detect small objects in images (see figure 1A). Networks operating at low resolution or coarse scale learn to detect *context* information. This is then passed to networks operating at high resolution to aid in the detection task. Each of the networks receives information directly from only a small part of several feature images, and so each of the networks can be relatively simple. The network at the highest resolution integrates the contextual information learned at the coarse resolution to detect the object of interest (i.e. the contextual inputs condition the probability of target present).

The HPNN architecture can be extended by considering the implications of inverting the information flow in the coarse-to-fine archiecture. This fine-to-coarse HPNN would have networks extracting detail structure at fine resolutions of the image and then passing this detail information to networks operating at coarser scales (see figure 1B). For many types of objects, information about the fine detail structure is important for discrimination between different classes. The fine-to-coarse HPNN is therefore a natural architecture for exploiting fine detail information for detecting extended objects.

In this paper, we present our experiences in applying the HPNN framework to two problems in mammographic Computer-Aided Diagnosis ($CAD$); that of detecting microcalcifications in mammograms and that of detecting malignant masses in mammograms. The coarse-to-fine HPNN architecture is well-suited for the microcalcification problem, while the fine-to-coarse HPNN is suited for mass detection. We evaluate the performance and utility of the HPNN framework by considering its effects on reducing false positive rates in a well characterized CAD system.

The University of Chicago (UofC) has been actively developing mammographic CAD systems for microcalcification and mass detection (REFS) and has been evaluating their performance clinically (REFS). A general block diagram showing the basic processing elements of these CAD systems is shown in figure 2. First, a pre-processing step is used to segment the breast area and increase the overall signal-to-noise levels in the image. Regions of interest ($ROIs$) are defined at this stage, representing local areas of the breast which potentially contain a cluster of calcifications or a mass. The next stage typically involves feature extraction and rule-based/heuristic analysis, in order to prune false positives. The remaining ROIs are classified as positive or negative by a statistical classifier or neural network. One of the key requirements of CAD is that false positive rates be low enough that radiologists will not ignore the CAD system output. Critical therefore is to reduce false positive rates of CAD systems without significant reductions in sensitivity. In this paper we evaluate the HPNN framework within the context of reducing the false positive rates of the UofC CAD systems for microcalcification and mass detection. In both cases the HPNN acts as a post-processor of the UofC CAD system.

## 2  Microcalcification detection

Microcalcifications are calcium deposits in breast tissue that appear as very small bright dots in mammograms. Clusters of microcalcifications frequently occur around tumors. Unfortunately microcalcification clusters are sometimes missed, since they can be quite subtle and the radiologists can only spend about a minute evaluating a patient's mammograms. CAD systems are being developed to serve as second readers, aiding radiologists with this problem.

Data used for the microcalcification experiments was provided by The University of Chicago. The first set of data consists of 50 true positive and 86 false positive ROIs. These ROIs are 99x99 pixels and digitized at 100micron resolution. A second
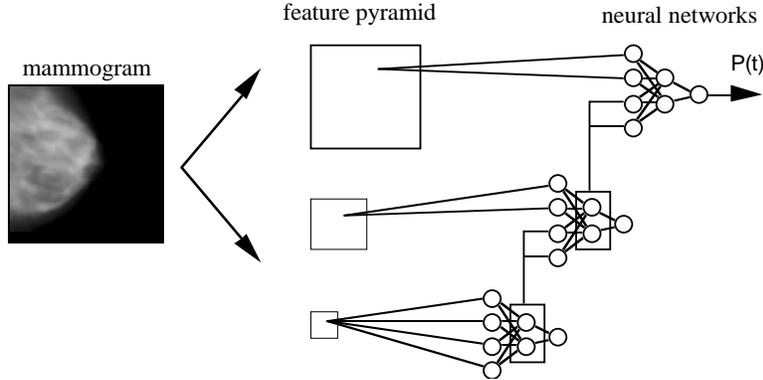
Figure 1: Hierarchical pyramid/neural network architectures for (A) detecting microcalcifications and (B) detecting masses. In (A) context is propagated from low to high resolution via the hidden units of low resolution networks. In (B) small scale detail information is propagated from high to low resolution. In both cases the output of the last integration network is an estimate of the probability that a target is present.
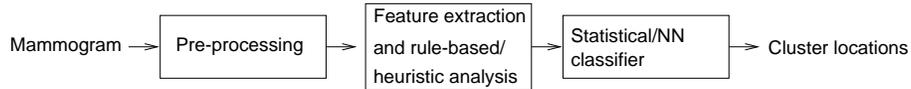


Figure 2: Block diagram for a typical CAD detection system.

set of data from the UofC clinical testing database included 47 true positives and 103 false positives, also 99x99 and sampled at 100micron resolution.

We trained the coarse-to-fine HPNN architecture in figure 1A as a detector for individual calcifications. For each level in the pyramid a network is trained, beginning with the network at low resolution. The network at a particular pyramid level is applied to one pixel at a time in the image at that resolution, and so produces an output at each pixel. All of the networks are trained to detect microcalcifications, however, due to the reduced scale of the mammogram at low resolutions the microcalcifications are not directly detectable at those pyramid levels. To achieve better than chance performance, the networks at those levels must learn something about the context in which microcalcifications appear. To integrate context information with the other features the outputs of hidden units from low resolution networks are propagated hierarchically as inputs to networks operating at higher resolutions.

Input to the neural networks come from an integrated feature pyramid (*IFP*) [1]. To construct the IFP, we used steerable filters [3] to compute local orientation energy. The steering properties of these filters enables the direct computation of the orientation having maximum energy. We constructed features which represent, at each pixel location, the maximum energy (energy at $\theta_{max}$), the energy at the orientation perpendicular to $\theta_{max}$ ($\theta_{max} - 90^o$), and the energy at the diagonal (energy at $\theta_{max} - 45^o$).[1] The resulting features are input into the coarse-to-fine network hierarchy.

In examining the truth data for the ROI data set, we found that the experts who

---

[1]We found that the energies in the two diagonal directions were nearly identical.

| cc | HPNN | | | | Chicago NN | | | |
|---|---|---|---|---|---|---|---|---|
| | $A_z$ | $\sigma_{A_z}$ | $FPF$ TPF=1.0 | $\sigma_{FPF}$ | $A_z$ | $\sigma_{A_z}$ | $FPF$ TPF=1.0 | $\sigma_{FPF}$ |
| 1 | .93 | .03 | .24 | .11 | .88 | .04 | .50 | .11 |
| 2 | .94 | .02 | .21 | .11 | .91 | .02 | .43 | .10 |
| 3 | .94 | .03 | .39 | .19 | .91 | .03 | .48 | .19 |
| 4 | .93 | .03 | .48 | .15 | .90 | .05 | .56 | .21 |
| 5 | .93 | .03 | .51 | .06 | .88 | .05 | .68 | .21 |

Table 1: Comparison of HPNN and Chicago networks.

specified the microcalcification positions often made errors in these positions of up to $\pm 2$ pixels of the correct position. To take this uncertainty in position into account, we used the following error function

$$E_{UOP} \quad = \quad -\sum_{p \in Pos} \log \left( 1 - \prod_{x \in p} (1 - y(x)) \right) - \sum_{x \in Neg} \log(1 - y(x)) \qquad (1)$$

which we have called the Uncertain Object Position ($UOP$) error function [5]. It is essentially the cross-entropy error, but the probability of generating a positive output on a particular example (the network's output on the example) has been replaced by the probability of generating at least one positive output in a region or set of pixels $p$ in the image. To this we added the standard weight decay regularization term. The regularization constant was adjusted to minimize the ten-fold cross-validation error.

The coarse-to-fine HPNN was applied to each input ROI, and a probability map was constructed from the output of the Level 0 network at each pixel. This map represents the network's estimate of the probability that a microcalcification is at a given pixel location. Training and testing were done using as jackknife protocol [4], whereby one half of the data (25 TPs and 43 FPs) was used for training and the other half for testing. We used five different random splits of the data into training and test sets.

For a given ROI, the probability map produced by the network was thresholded at a given value to produce a binary detection map. Region growing was used to count the number of distinct detected regions. The ROI was classified as a positive if the number of regions was greater than or equal to a certain cluster criterion.

Table 1 compares ROC results for the HPNN and another network that had been used in the University of Chicago CAD system [7] using five different cluster criterion (cc). Reported are the area under the ROC curve ($A_z$), the standard deviation of $A_z$ across the subsets of the jackknife ($\sigma_{A_z}$), the false positive fraction at a true positive fraction of 1.0 ($FPF@TPF = 1.0$) and the standard deviation of the FPF across the subsets of the jackknife ($\sigma_{FPF}$). $A_z$ and $FPF@TPF = 1.0$ represent the averages of the subsets of the jackknife. Note that both networks operate best when the cluster criterion is set to two. For this case the HPNN has a higher $A_z$ than the Chicago network while also halving the false positive rate. This difference, between the two networks' $A_z$ and $FPF$ values, is statistically significant (z-test; $p_{A_z} = .0018$, $p_{FPF} = .00001$).

A second set of data was also tested. 150 ROIs taken from a clinical prospective study and classified as positive by the full Chicago CAD system (including the Chicago neural network) were used to test the HPNN. Though the Chicago CAD

system classified all 150 ROIs as positive, only 47 were in fact positive while 103 were negatives. We applied the HPNN trained on the entire previous data set to this new set of ROIs. The HPNN was able to reclassify 47/103 negatives as negative, without loss in sensitivity (no false negatives were introduced).

On examining the negative examples rejected by the coarse-to-fine HPNN, we found that many of these ROIs contained linear, high-contrast structure which would otherwise be false positives for the Chicago network. One possible reason for this is that the coarse-to-fine HPNN also learns context for the false positives. The Chicago neural network presumably interprets the "peaks" on the linear structure as calcifications. However because the coarse-to-fine HPNN also integrates information from low resolution it can associate these "peaks" with linear structure at low resolution and thus determine that these peaks are not microcalcifications. This is an interesting difference from our earlier work, in which the networks learned context associated with positive examples.

# 3   Mass detection

Although microcalcifications are an important cue for malignant masses in mammograms, they are not visible or even present in all cases. Thus mammographic CAD systems include algorithms to directly detect the presence of masses. We have started to apply a fine-to-coarse HPNN architecture to detect malignant masses in digitized mammograms. Radiologists often distinquish malignant from benign masses based on the detailed shape of the mass border and the presence of spicules alone the border. Thus to integrate this high resolution information to detect malignant masses, which are extended objects, we apply the fine-to-coarse HPNN of figure 1B.

The experimental paradigm is similar to the microcalcification experiments in that we apply the HPNN as a post-processor to the UofC CAD system for mass detection. The data in our study consists of 72 positive and 100 negative ROIs. The negative ROIs are false-positives of the earlier stages of the CAD system. These are 256-by-256 pixels and are sampled at 200 micron resolution.

At each level of the fine-to-coarse HPNN several hidden units process the feature images. The outputs of each unit at all of the positions in an image make up a new feature image. This is reduced in resolution by the usual pyramid blur-and-subsample operation to make an input feature image for the network units at the next lower resolution. We trained the entire fine-to-coarse HPNN as one network instead of training a network for each level, one level at a time. This training is quite straighforward. Back-propagating error through the network units is the same as in conventional networks. We must also back-propagate through the pyramid reduction operation, but this is linear and therefore quite simple, up to some implementation details, such as propagating the error properly at the borders of the image. In addition we use the same UOP error function 1 used to train the coarse-to-fine architecture. The rationale for this application of the UOP error function is that the truth data specifies the location of the center of the mass at the highest resolution. However, as we integrate up the fine-to-coarse architecture the coordinate information becomes uncertain due to the subsample and bluring operations. To take into account this loss of positional information as we move up the fine-to-coarse archiecture, we use the UOP error function.

The features input to the fine-to-coarse HPNN are radial and tangential gradient components at each resolution, relative to the mass center. The center coordinates are generated by the earlier stages of the CAD system. The gradients are generated

by first-derivative of Gaussian filters, proportional to $\vec{x}\exp(-\|\vec{x}\|^2/2\sigma^2)$ where $\vec{x}$ is the relative displacement from the filter center. In addition to the filter outputs, we add the squares of the filter outputs so the local radial and tangential image energies are available to the network. Unlike the microcalcification coarse-to-fine HPNN, we did not reduce these images in size to all lower resolution pyramid levels. For example, the gradient features extracted from level 2 are provided as input only to the hidden units at level 2. Information from this level passes to level 3 only through the hidden unit outputs.

The fine-to-coarse HPNN is quite similar to the convolution network proposed by Le Cun, et al [2], however with a few notable differences. The fine-to-coarse HPNN receives as input preset features extracted from the image (in this case radial and tangential gradients) at each resolution, compared to the network by Le Cun, which receives preset features only at the highest resolution. Secondly, in the fine-to-coarse HPNN, the inputs to a hidden unit at a particular position are the pixel values at that position in each of the feature images, one pixel value per feature image. Thus the hidden units do not learn linear filters, as they do in the Le Cun network, except as linear combinations of the filters used to form the features. Finally the fine-to-coarse HPNN is trained using the UOP error function, which is not used in the Le Cun network.

Currently our best performing fine-to-coarse HPNN system for mass detection has two hidden units per pyramid level. This gives an ROC area of $A_z = 0.82$ and eliminates 32 % of the false-positives at a cost of missing 10 % of the actual positives. To improve performance further, we are investigating different regularizers, richer feature sets, and more complex architectures, i.e., more hidden units. In practice the HPNN's threshold and those in the previous stages of the CAD system would have to be adjusted together to optimize performance, so even this network may give a useful improvement in the CAD system's performance.

## 4 Conclusion

We have presented the application of multi-resolution neural network architectures to two problems in computer-aided diagnosis, the detection of microcalcifications in mammograms and the direct detection of malignant masses in mammograms. A summary of the performance of these archiectures is given in TABLE 2. In the case of microcalcifications, the coarse-to-fine HPNN architecture successfully discovered large-scale context information that improves the systems performance in detecting small objects. A coarse-to-fine HPNN has been directly integrated with the UofC CAD system for microcalcification detection and the complete system is undergoing clinical evaluation.

In the case of malignant masses, a fine-to-coarse HPNN architecture was used to exploit information from fine resolution detail which could be used to differentiate malignant from benign masses. The results of this network are encouraging, but additional improvement is needed. In general, we have found that the multi-resolution HPNNs are a useful class of network architecture for exploiting and integrating information at multiple scales.

## 5 Acknowledgments

of Chicago for useful disccusions and the providing of the data.

# References

[1] Peter Burt. Smart sensing within a pyramid vision machine. *Proceedings IEEE*, 76(8):1006–1015, 1988. Also in *Neuro-Vision Systems*, Gupta and Knopf, eds., 1994.

[2] Y. Le Cun, B. Boser, J. S. Denker, and D. Henderson. Handwritten digit recognition with a back-propagation network. In David S. Touretzky, editor, *Advances in Neural Information Processing Systems 2*, pages 396–404, 2929 Campus Drive, San Mateo, CA 94403, 19901. Morgan-Kaufmann Publishers.

[3] William T. Freeman and Edward H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13(9):891–906, 1991.

[4] Charles Metz. Current problems in ROC analysis. In *Proceedings of the Chest Imaging Conference*, pages 315–33, Madison, WI, November 1988.

[5] Clay D. Spence. Supervised learning of detection and classification tasks with uncertain training data. In *Image Understanding Workshop*. ARPA, 1996. This Volume.

[6] Clay D. Spence, John C. Pearson, and Jim Bergen. Coarse-to-fine image search using neural networks. In Gerald Tesauro, David S. Touretzky, and Todd K. Leen, editors, *Advances in Neural Information Processing Systems 7*, pages 981–988, Massachusetts Institute of Technology, Cambridge, MA 02142, 1994. MIT Press.

[7] W. Zhang, K. Doi, M. L. Giger, Y. Wu, R. M. Nishikawa, and R. Schmidt. Computerized detection of clustered microcalcifications in digital mammograms using a shift-invariant artificial neural network. *Medical Physics*, 21(4):517–524, April 1994.