Training Neural Networks for Computer-Aided Diagnosis: Experience in the Intelligence Community

Paul Sajda and Clay Spence
National Information Display Laboratory, National Technology Alliance
Princeton NJ
E-mail: psajda@sarnoff.com
cspence@sarnoff.com

1. Abstract

Neural networks are often used in computer-aided diagnosis systems for detecting clinically significant objects. They have also been applied in the Intelligence Community to cue Image Analysts (IAs) for assisted target recognition and wide-area search. Given the similarity between the applications in the two communities, there are a number of common issues that must be considered when training these neural networks. Two such issues are: 1) exploiting information at multiple scales (e.g. context and detail structure) and 2) dealing with uncertainty (e.g. errors in truth data). Working with The University of Chicago, we address these two issues, transferring architectures and training algorithms, originally developed for assisting IAs in search applications, to improve computer-aided diagnosis for mammography. These include hierarchical pyramid/neural-network (HPNN) architectures that automatically learn and integrate multi-resolution features for improving microcalcification and mass detection in computer-aided diagnosis (CAD) systems. These networks are trained using an uncertain object position (UOP) error function for the supervised learning of image search/detection tasks when the position of the objects to be found is uncertain or ill-defined. Results show that the HPNN architecture trained using the UOP error function reduces the false positive rate of a mammographic CAD system by 30%-50% without significant loss in sensitivity. We conclude that the transfer of assisted target recognition technology from the Intelligence Community to the Medical Community can significantly impact the clinically utility of CAD systems.

2. Exploiting Multi-scale Information for Pattern Recognition

An important approach for detecting objects of interest in images is to exploit information at multiple scales. We have been developing pattern recognition architectures which integrate multi-scale information, via pyramid representations, together with neural networks to automatically learn scale space relationships for detecting objects of interest. These pattern recognizers, which we call hierarchical pyramid neural networks (HPNN), can be defined in terms of both a coarse-to-fine and a fine-to-coarse architecture (see Figure 1). The architectures can be used to detect both small and large objects by exploiting coarse-scale (low resolution) or fine-scale (high-resolution) information is associated with the target. For example in the coarse-to-fine HPNN networks operating at low resolution learn to detect *context* information. This is then passed to networks operating at high resolution to aid in information learned at the coarse resolution to detect the object of interest (i.e. the contextual inputs condition the probability of target present). For the fine-to-coarse HPNN architecture networks extract detail structure at fine resolutions of the image and then pass 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.

We have previously reported on how these HPNN architectures can improve detection 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 coarse-to-fine HPNN architecture has higher accuracy than both conventional neural network architectures and standard statistical classification techniques. In this paper we summarize our results of 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 developed by The University of Chicago (UofC). In both cases (microcalcification and mass detection) the HPNN acts as a post-processor of the UofC CAD system.

3. 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 100, micron resolution. A second set of data from the UofC clinical testing database included 47 true positives and 103 false positives, also 99x99 and sampled at 100/micron 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 the lowest 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, at low resolutions the microcalcifications are not directly detectable. 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*)[2]. To construct the IFP, we use 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 construct features which represent, at each pixel location, the maximum energy (energy at θ_{max}), the energy at the orientation perpendicular to θ_{max} (θ_{max} -90°), and the energy at the diagonal (energy at θ_{max} -45°). 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 specified the microcalcification positions often made errors in these positions of up to +/-2 pixels of the correct position. To take this uncertainty in position into account, we used the following error function :

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

which we have called the Uncertain Object Position (*UOP*) error function[4] (*y(x)* is the network's output when applied to position *x.)* It is essentially the cross-entropy error, but for positive examples the probability of generating a positive output *y(x),* has been replaced by the probability of generating at least one positive output in a region or set of pixels *p* in the image. In our case each *p* is a five-by-five pixel square centered on the location specified by the expert. To this we added the standard weight decay regularization term. The regularization constant was adjusted to minimize the ten-fold cross-validation error.

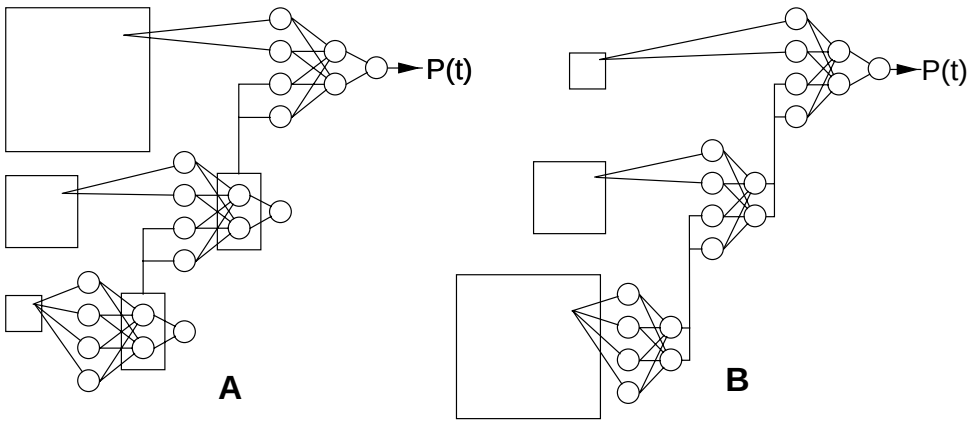

Figure 1: Hierarchical pyramid/neural network architectures for (A) detecting microcalcifications and (B) detecting masses. In (A) context is propagated from low to high resolution via the hidden units of low-resolution networks. In (B) small scale detail information is propagated from high to low resolution. In both cases the output of the last integration network is an estimate of the probability that a target is present.

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 a jackknife protocol[5], 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[6] 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 (σ_{Az}), 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 (σ_{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; pA_z=.0018, *pFPF*=.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.

Table 1 Comparison of HPNN and Chicago networks

Cc	HPNN				Chicago NN			
	A_z	σ_{Az}	*FPF* TPF=1.0	σ_{FPF}	A_z	σ_{Az}	*FPF* TPF=1.0	σ_{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

This is an interesting difference from our earlier work, in which the networks learned context associated with positive examples.

4. 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 applied a fine-to-coarse HPNN architecture to detect malignant masses in digitized mammograms. Radiologists often distinguish malignant from benign masses based on the detailed shape of the mass border and the presence of spicules along 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 straightforward. 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 (Equation 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 blurring operations. To take into account this loss of positional information as we move up the fine-to-coarse architecture, 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 $xe^{-\|x\|^2/2\sigma^2}$ where 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 easily 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.

Table 2 Sensitivity and specificity for fine-to-coarse HPNN for mass detection

Sensitivity	Specificity
100%	32%
95%	36%
90%	40%
80%	78%

Results for the fine-to-coarse HPNN system are shown in Table 2. The A_z value on the test set was 0.84. These results suggest a 30% reduction in false positive rate of the UofC mass detection system without loss in sensitivity. In summary the HPNN framework appears to be able to reduce the false positive rates of the UofC microcalcification and mass detection systems by roughly 30%-50%, potentially making CAD systems more acceptable to radiologists.

5. 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. In the case of microcalcifications, the coarse-to-fine HPNN architecture successfully discovered large-scale context information that improves the system's 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 mass detection, a fine-to-coarse HPNN architecture was used to exploit information from fine resolution detail which could be used to eliminate false positives. 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.

6. Acknowledgments

This work was funded by the National Information Display Laboratory and the Murray Foundation. We would like to thank Drs. Robert Nishikawa and Maryellen Giger of The University of Chicago for useful discussions and providing the data.

7. References

[1] Paul Sajda, Clay D. Spence, Steve Hsu and John C. Pearson. Integrating Neural Networks with Image Pyramids to Learn Target Context. *Neural Networks,* 8(7/8):1143-1152, 1995

[2] 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.

[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] C. Spence. Supervised learning of detection and classification tasks with uncertain training data. In *ARPA Image Understanding Workshop,* Palm Springs, CA, February 1996.

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

[6] 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.