CAPTURING CONTEXTUAL DEPENDENCIES IN MEDICAL IMAGERY USING HIERARCHICAL MULTI-SCALE MODELS

Paul Sajda

Columbia University
Department of Biomedical Engineering
New York, NY 10027

Clay Spence and Lucas Parra

Sarnoff Corporation
Vision Technologies Lab
Princeton, NJ 08540

ABSTRACT

In this paper we summarize our results for two classes of hierarchical multi-scale models that exploit contextual information for detection of structure in mammographic imagery. The first model, the hierarchical pyramid neural network (HPNN), is a discriminative model which is capable of integrating information either coarse-to-fine or fine-to-coarse for microcalcification and mass detection. The second model, the hierarchical image probability (HIP) model, captures short-range and contextual dependencies through a combination of coarse-to-fine factoring and a set of hidden variables. The HIP model, being a generative model, has broad utility, and we present results for classification, synthesis and compression of mammographic mass images. The two models demonstrate the utility of the hierarchical multi-scale framework for computer assisted detection and diagnosis.

1. INTRODUCTION

Robust analysis of imagery, including object detection, recognition and segmentation, often depends on the exploitation of *contextual cues*. For image analysis, "context" can be defined as that which surrounds an object and determines its meaning. This definition implies that the spatial (and/or temporal and/or spectral, etc.) neighborhood of an object plays a major role in its interpretation. Radiologists performing medical image analysis, particularly diagnostic screening, utilize contextual cues derived from the imagery in their decision process. The location of suspicious structure relative to the surrounding anatomy and the appearance of surrounding tissue are just two examples of these contextual cues that are exploited. Computer-aided diagnosis (CAD) systems have been developed to assist radiologists in medical image analysis by providing a low-cost "second-reader"[1]. An important challenge in CAD is developing systems which automatically learn to exploit contextual cues to maximize overall sensitivity and specificity.

We have developed two classes of pattern recognition models which utilize multi-scale decompositions for learning contextual dependencies in images. In previous work, we have presented the details of these two models [2][3]. The first is a discriminative model, called the hierarchical pyramid neural network (HPNN), that utilizes a multi-resolution pyramid structure to integrate scale information either coarse-to-fine or fine-to-coarse, depending on the natural scale of the object. The second model, termed the hierarchical image probability model (HIP), is a probabilistic generative model of the image distribution (or features extracted from the image). Similar to the HPNN, HIP utilizes a multi-resolution pyramid decomposition to capture both short-range and long-range dependencies (i.e. context) in the image. In this short paper we summarize our results for these two models, particularly as they relate to analysis of mammographic data.

2. A MULTI-SCALE DISCRIMINATIVE MODEL: HPNN

The first model is the hierarchical pyramid neural network (HPNN). In this paper we summarize our results of applying the HPNN framework to two problems in mammographic CAD; that of detecting microcalcifications in mammograms and that of detecting malignant masses in mammograms. A more detailed description of the model has been published elsewhere [2].

Two architectures for the HPNN are illustrated in figure 1. The coarse-to-fine HPNN architecture is well-suited for the microcalcification problem, while the fine-to-coarse HPNN is best suited for mass detection. We evaluate the performance and utility of the HPNN framework by considering its effect on reducing false positive rates in a well-characterized CAD system developed by The University of Chicago (UofC). In both cases (microcalcification and mass

This work was funded by the U.S. Army Medical Research and Material Command (DAMD17-98-1-8061). P.S. was also supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research under Grant N00014-01-1-0625.

detection) the HPNN acts as a post-processor of the UofC CAD system.

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 regions of interest (ROIs). These ROIs are 99x99 pixels and digitized at 100micron 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 100micron resolution.

We trained the coarse-to-fine HPNN architecture in figure 1 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)[4]. To construct the IFP, we use steerable filters[5] to compute local orientation energy. The steering properties of these filters enable 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} ($\theta_{max} - 90^o$), and the energy at the diagonal (energy at $\theta_{max} - 45^o$). The resulting features are input into the coarse-to-fine network hierarchy.

The coarse-to-fine HPNN was applied to each input ROI, and a probability map was constructed from the output network in the hierarchy. This map represents the network's estimate of the probability that a microcalcification is at a given pixel location. 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.

ROC results were computed for the HPNN and compared to another neural network that had been used in the UofC CAD system [6]. The area of the ROC curve (A_z) for the HPNN was $A_z^{HPNN} = 0.94$ vs. $A_z^{UofC} = 0.91$. In terms of reduction in false positives, at 100% sensitivity the HPNN had a false positive rate of 21% vs 43% for the UofC network. Thus for this dataset, the HPNN had a higher A_z than the UofC network while also halving the false positive rate. This difference, between the two networks' A_z and FPF values, is statistically significant (z-test; $p_{Az} = .0018$, $p_{FPF} = .00001$).

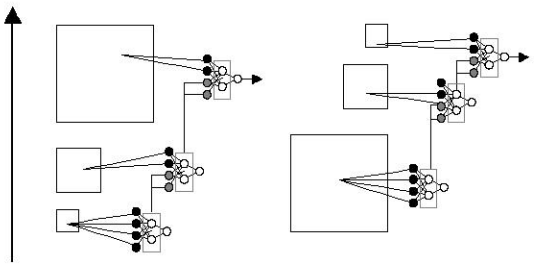

Fig. 1. Coarse-to-fine (left) and fine-to-coarse (right) HPNN architectures. Large arrow on left shows the direction of training and processing.

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 UofC 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 that would otherwise be false positives for the UofC network. One possible reason for this is that the coarse-to-fine HPNN learns context for the false positives. The UofC 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.

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 1.

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 200micron 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 straightforward. Back-propagating error through the network units is the same as in conventional networks.

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. 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 in the pyramid 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. Evaluation of the fine-to-coarse HPNN system resulted in an A_z value on the test set of 0.85 and a 52% reduction in false positive rate of the UofC mass detection system without loss in sensitivity. Thus the HPNN framework is able to reduce false positive rates of both the UofC microcalcification and mass detection CAD systems by approximately 50%

3. A MULTI-SCALE GENERATIVE MODEL: HIP

The second model we describe is a hierarchical image probability (HIP) model, which is a generative model that explicitly models the probability distribution of an image (or image features) for a given class of images. It is beyond the scope of this paper to describe the details of the HIP model and we instead refer the reader to [3]. In this paper we briefly describe one of the key aspects of the model, namely coarse-to-fine factoring of image distributions using a pyramid representation. We then present results of evaluating HIP within the context of its generative utility, specifically with regard to 1) mammographic mass classification, 2) synthesis and 3) compression.

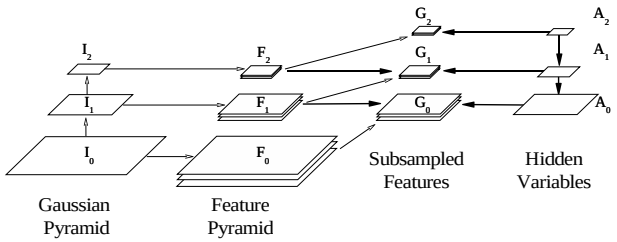

Fig. 2. Structure of the HIP model. Conditioning is shown with bold arrows while construction of features is shown with light arrows.

Fig. 3. Mammographic ROI images synthesized from positive and negative HIP models. Synthesized positive ROIs (left) tend to have more focal structure, with more defined borders and higher spatial frequency content. Negative ROIs (right) tend to be more amorphous with lower spatial frequency content.

3.1. Coarse-to-fine factoring of image distributions

In developing the HIP model our goal is to write the image distribution $Pr(I)$ as a product of conditional distributions in a form similar to $\Pr(I) \sim \Pr(\mathbf{F}_0 \mid \mathbf{F}_1) \Pr(\mathbf{F}_1 \mid \mathbf{F}_2) \ldots$, where $\mathbf{F}_l$ is the set of feature images at pyramid level l. To do this consider building a Gaussian pyramid of image I. From each Gaussian level I_l we extract some set of feature images $\mathbf{F}_l$ (see figure 2). Sub-sample these feature images to get the feature images $\mathbf{G}_l$ so that the images in $\mathbf{G}_l$ have the same dimension as I_{l+1}. Denote the set of images $\{I_{l+1}$,$\mathbf{G}_l\}$ by $\tilde{\mathbf{G}}_l$, and the mapping from I_l to $\tilde{\mathbf{G}}_l$ by $\tilde{\mathcal{G}}_l$. If $\tilde{\mathcal{G}}_l$ is invertible for all l we can show that

$$\Pr(I) = \left[\prod_{l=0}^{L-1} |\tilde{\mathcal{G}}_l| \Pr(\mathbf{G}_l \mid I_{l+1})\right] \Pr(I_L) \qquad (1)$$

In order to factor $\Pr(\mathbf{G}_l | I_{l+1})$ over positions we introduce hidden variables, of which several different choices are discussed in[7]. The model, including hidden variables, can be fit to data using the Expectation-Maximization (EM) algorithm.

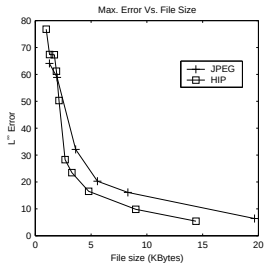

Fig. 4. Maximum error (L^∞ norm) vs. size of compressed file, JPEG and HIP.

3.2. Classification

We applied HIP to the problem of detecting masses in ROIs taken from the UofC CAD system for mass detection. We trained a HIP model on a set of 36 positive examples (ROIs containing masses) and a second HIP model on 48 negative examples (ROIs without masses). The likelihood ratio was used as a test criterion-i.e. a threshold on this ratio is used to decide which ROIs will be classified as masses. The true and false positive rates were measured on an independent test set of 36 positives and 48 negatives. Results show a reduction in false positive rate of 25% with no loss in sensitivity.

3.3. Synthesis

Since the HIP model is a generative model, we can sample the model and synthesize new images. In practice, this property might be best utilized for image compression or noise reduction. Within the context of ROI classification, synthesized images can give insight into what features the model is extracting and representing for both positive and negative ROIs. Using the same ROI database used for classification, we constructed HIP models for positives (masses) and negatives (no masses). The trained HIP models were sampled to synthesize new ROI images. Figure 3 shows examples of these images. Inspection of the synthesized positive ROIs shows more focal structure, with more well-defined borders and higher spatial frequency content than the negative ROIs.

3.4. Compression

Given an image and a HIP model, we compute the most likely value of each hidden label and code each feature vector for a given image. Figure 4 shows the maximum errors versus the size of the resulting compressed file, respectively. Note that this result is for one randomly-chosen mass ROI image, which was not part of the training set of the HIP model. The HIP model gives maximum errors that are lower than JPEG.

4. CONCLUSION

We have presented two HPNN architectures and results which demonstrated their ability to reduce false positive rates for microcalcification and mass detection in mammographic CAD. A coarse-to-fine HPNN has been directly integrated with the UofC CAD system for microcalcification detection and the complete system has undergone clinical evaluation[2]. Recently we have combined the coarse-to-fine and fine-to-coarse structures into a single generalized framework, allowing for the simultaneous integration of information both up and down the pyramid structure [2].

The second model presented was the hierarchical image probability (HIP) model, trained to estimate $\Pr(Image|Class)$. There are several attractive features of the HIP framework which could have a major impact on the design, and development of mammographic CAD systems. Since HIP computes $\Pr(Image|Class)$, we can detect unusual images and reject them rather than trust the classifier; something that is not possible with models of $\Pr(Class|Image)$. Building confidence measures into CAD systems is an open area of research and the HIP model provides a mechanism by which to generate these measures.

5. REFERENCES

[1] K. Doi et al., "Digital radiography: A useful clinical tool for computer-aided diagnosis by quantitative analysis of radiographic images," *Acta Rad.*, vol. 34, pp. 426–439, 1993.

[2] P. Sajda and C. Spence, "Learning contextual relationship in mammograms using a hierarchical pyramid neural network," *IEEE TMI*, 2002.

[3] C.D. Spence, L. Parra, and P. Sajda, "Detection, synthesis and compression in mammographic image analysis using a hierarchical image probability model," in *IEEE MMBIA 2001*, M. Staib, Ed., 2001, pp. 3–10.

[4] P. Burt, "Smart sensing within a pyramid vision machine," *Proc. IEEE*, vol. 76, no. 8, pp. 1006–1015, 1988.

[5] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE TPAMI*, vol. 13, no. 9, pp. 891–906, 1991.

[6] W. Zhang et al., "Computerized detection of clustered microcalcifications in digital mammograms using a shift-invariant artificial neural network," *Med. Phys.*, vol. 21, no. 4, pp. 517–524, April 1994.

[7] C.D. Spence, L. Parra, and P. Sajda, "Development and application of a hierarchical image probability model," in *IEEE Trans. Image Proc.*, submitted.