Development of a Neural Network for Anatomic and Reference Landmark Location on Radiographic Images

by

Brenda Baker

Development of a Neural Network for Anatomic and Reference Landmark Location on Radiographic Images

by

Brenda Baker

Abstract

Accurate and reliable data are crucial to research investigators and clinicians who evaluate growth and plan treatment on the basis of measurements from x-ray films. Current methods of data collection involve locating anatomical or reference features (landmarks) by hand on analog x-ray cephalograms. These operations are labor intensive, time-consuming, dependent on the characteristics of operator training, and prone to error. These deficiencies are the impetus for developing an automated system. One concern with most automated programs is that they rely on a written definition and a precise algorithm for landmark location. For this reason, they are confounded by the fact that landmarks vary in shape, structure and gray scale characteristics from patient to patient and at different stages of growth or treatment in the same patient.

Neural networks are a method for solving this problem. By using a set of training images, the network is trained to locate landmarks using the characteristics of the landmark and the surrounding image without an explicit definition or locating algorithm. Once trained, the network is able to process data outside of the training set with a high level of performance, particularly on images of later growth or treatment of other patients. The network in this thesis is distinguished by two dimensional layers, selective connectivity, supervised training and a standard backpropagation algorithm.

Two sets of experiments examined the network's ability to locate reference and anatomic landmarks. The first set of experiments with reference landmarks also investigated two different types of network connections and a study of the number of hidden layers necessary for optimal performance. The best results were achieved when the network used a 5x5 connection scheme and one hidden layer. The network trained to the characteristic distribution of the training judge. In fact the network picked the same pixel as the training judge in twenty seven of the thirty six images. This demonstrates that the network can be trained to the particular biases of an individual and it seems highly likely that a network trained to a better standard would have better results.

The second set of experiments explored the network's ability to locate a specific anatomic landmark: "upper incisor edge". The network located twenty nine of thirty six incisors despite variations in anatomy and patient age. The network was able to locate the incisor successfully on some but not all types of images not represented in the training set. This demonstrates that the training set need not be all inclusive in order to perform well.

The network was able to train to locate both the reference marker and the incisor despite differences in anatomy and patient age. While there is a significant difference in shape between anatomic landmarks, the quality of the network results is largely dependent on the design of the training set. Given a properly designed training set the network will perform well with other anatomic landmarks.

Chairman: Dr. Sheldon Baumrind
University of California, San Francisco

Acknowledgments

I would like to thank my mentor Dr. Sheldon Baumrind who whole heartily supported and encouraged my graduate research. His enthusiasm and energy were invaluable in creating an environment in which I could freely explore my ideas. He not only provided the opportunity to complete my research (enjoyably!) but also encouraged me to pursue other academic interests and to look beyond the immediate future to long range goals. I will always value our discussions on a variety of topics including research, history, politics, and human behavior. I am very grateful to my two other readers, Dr. Larry Stark and Dr. Jitendra Malik, who have been very helpful. My discussions with them always brought up new questions and ideas. I particularly appreciate my committee's help in making this thesis more coherent and unambiguous.

I would like to acknowledge the stimulating discussions and support of Dr. Robert Gould, Dr. Harry Genant, and Dr. Vojtech Licko. Randy Berndt has been an enormous help with so many details during my graduate studies; always coming up with a solution to every problem.

Finally I would like to thank my parents Juanita and Richard Baker, my husband Andy Liu, my friends Romella Husain and Kerstin Pfann, and my sister Desha Baker for their constant support and love. They have always encouraged me through the tough spots and celebrated with me over the triumphs!

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The treatment of craniofacial anomalies and dentofacial malformations relies heavily upon the interpretation of information from x-ray images of the skull. Measurements made from these images are used to evaluate and assess both normal and abnormal growth of bone and teeth structures. These measurements are dependent on the accurate and reliable location of anatomic landmarks. When examining x-ray cephalograms (images of the head) for diagnosis or research, anatomical features and reference markers (landmarks) need to be located on successive images. The current method of data collection involves locating landmarks, by hand, from analog x-ray cephalograms. Relevant distances and angles can then be measured for analysis purposes. These operations are labor intensive, time-consuming, dependent on the characteristics of operator training, and prone to error. These deficiencies are the impetus for an automated computerized system of landmark location. In addition, storage space for the large amount of medical data and an efficient means of retrieving images are leading investigators and clinicians towards the use of digital images. Another benefit of using digital images is the ability to enhance and process images to emphasize or clarify a particular aspect of the image. Digital image processing can then make diagnosis, treatment planning, and research, more efficient and more reliable for craniofacial clinicians and researchers.

Landmark Location

The work for this thesis was done in the Craniofacial Research Instrumentation Laboratory of University of California, San Francisco headed by Dr. Sheldon Baumrind. The standard method of landmark location in our laboratory involves making "acetate overlays" or tracings of the original x-rays upon which the location of landmarks is indicated (including reference markers). A Sumagraphics ID series digitizer is used to store the information in a data file. All acts of landmark location involve replicate independent assessments by different judges. Baumrind and Baker [1, 2] have explored a variety of forms of digital image formats and the reliability of human judges locating landmarks on a PC based system.

Often a successful way to teach a task that is difficult to put into words is through trial and error . Judges who locate landmarks are trained through a process of iteration. The definition of a landmark is discussed and the method for locating a landmark is learned. The judge then attempts to locate the landmark on a series of images. The locations selected by the judge are reviewed and discussed. If there are discrepancies the judge is retrained by discussing the definitions of landmarks as well as possible confounding issues of particular anatomy. Ambiguity in the definitions of landmarks can cause major problems with a judge's ability to reliably locate landmarks [1] One concern with automated algorithms is that since they rely on written and precise algorithms for location they will not perform well in cases when the landmark is difficult to define.

When human operators locate landmarks on x-ray images, systematic errors are associated with each landmark and each judge. These errors are the result of the specific features of the individual landmarks as well as the individual judges' conceptual and perceptual definition of the landmark. The systematic errors are characterized by envelopes of error for each landmark which differ both in size and in shape [3].

Furthermore, when several expert judges assess the same landmark on a series of images, the assessments of different judges typically show systematic differences with respect to the group mean. In point of fact, even the written definitions for conventional craniofacial radiographic landmarks can differ among different expert judges. Further, complications are introduced by the fact that landmarks vary in shape, structure and gray scale characteristics from patient to patient and at different stages of growth or treatment in the same patient.

Reliability Studies

Before adopting any new image medium, its ability to reliably replace the old medium must be examined. Baumrind and Baker [1] (1991 Reliability Study) compared the reliability of landmark location by five judges as a function of image type. Judges located twenty-seven landmarks on forty-four analog x-ray films and on corresponding (512 x 480 pixel) digital images. For many of the landmarks, the standard deviations of landmark location on the digital image approximated those on the x-ray films. However, digital images proved to be a poor approximation of the original x-ray films in regions where the gray scale values saturated, particularly in the interior regions of the skull. Thus an important finding from this study is that the digital medium was differentially acceptable for different landmarks.

Another conclusion of this reliability study was that regardless of image type, distributions of landmark locations vary from individual to individual. This variability can be attributed to the ambiguity and discrepancies in conventional verbal definitions of landmarks as well as the vagueness of the operational recipes for locating them. In fact, this result was also documented in a 1971 Reliability Study of landmark location [3].

Automated Landmark Location

The traditional approach to automated landmark location involves some form of template matching or feature based correspondence [4-7]. Template matching involves designing a filter or template which is passed over the image searching for a best match. This procedure is highly landmark dependent and sensitive to rotation or distortions in the image. For this reason, the variability of landmarks from patient to patient and the fact that landmarks change within patients through time confound the filtering problem.

In an attempt to avoid this problem, several systems using different degrees of rule-based or algorithmic definitions, have been proposed to automate landmark location. Levy-Mandel et al. [8] proposed a knowledge-based expert system for automatically locating landmarks. First, a global line-following or line-extracting technique is used to locate all the existing lines and edges of the image. Then, a knowledge base is used to select the relevant landmark location. Parthasarathy et al. [9] proposed an algorithm that used edge enhancement to locate lines but then used feature recognition techniques, in addition to a knowledge base, to locate landmarks. Both of these methods rely heavily on algorithmic definitions (the knowledge base). The process of converting a written or verbal definition into an algorithmic form invariably involves loss of information and may introduce errors. The algorithmic definition involves relating every landmark to an edge or contour occasionally losing the anatomic meaning in the process. In addition, such algorithms do not take age or patient variability into consideration.

Parthasarathy and Tong et al. [9, 10] have developed an algorithm for landmark detection based on feature recognition techniques. The essential aspect of their algorithm traces bony and soft tissue outlines. The program then defines and locates twenty seven landmarks with respect to these outlines. They tested their algorithms by comparing their results to those of an orthodontist on five images. Their expert determined that seventy five percent of the automatically located landmarks were acceptable. Criteria for acceptability were not given. In addition, the points located by the program were the closest approximation of a landmark on an edge, not necessarily carrying any meaning of the original landmark.

Neural Networks

The definitions of landmarks and the methods used to find them are frequently very difficult to put in algorithmic format. To avoid a dependency on the algorithmic definition of a landmark, I propose an alternative solution using neural networks to locate landmarks [11-13]. This method is based entirely upon the information contained in the images. The system I have designed identifies and locates anatomical or reference landmarks in a digital image without an explicit definition or locating algorithm. Instead the system uses the characteristics of the landmark and the surrounding image. Three properties of neural networks make them viable solutions to the problem of landmark location on cephalograms. First, by using a training set of images, the network can compute the weighting coefficients which allow identification and location of landmarks. This feature eliminates the need for the user to determine the necessary characteristics of a pattern or to describe it in an algorithmic format. Second, once trained, these algorithms are able to process data outside of the knowledge base or training set with a relatively high level of performance. This is important because once trained on one set of patients the network performs on either successive unseen time points or on other patient sets. Third, once the network has been trained, it is possible to disconnect or change many of the connections in a network with little affect on the accuracy of the system. This is important for images with noisy or variant data.

Neural networks have been widely applied to pattern recognition and classification problems. Investigators have described the use of neural networks in a variety of vision and image processing applications [14-19]. Applications include the XOR problem, speech synthesis and recognition, and visual pattern recognition [20-21]. In the field of medical imaging, neural networks have been used to either automate or aid in diagnosis. It has been demonstrated that neural networks perform pattern detection of simple test objects with better accuracy than human judges on radiological images [22]. Early work with multiple adaptive matched filters included automatic interpretation of electrocardiograms with the use of a self-organizing system. Okajima et. al. [23] and Yasui et. al. [24] designed a system to classify the QRS complex of the EKG curve. Neural networks have been used recently by Wang and Li [25] for edge detection on ophthalmoscopic images in a decision support system, by Silverman [26] to segment Ultrasonic images, and by Chen et. al. [27] to outline segmented areas on Magnetic Resonance images (MRI).

Only recently have neural networks been used to locate large structures on images. Hutchinson has used Multi Layer Perceptron (MLP) systems to locate eyes on facial images [28]. He was interested in transmitting only parts of images (eyes and mouth) in order to transmit video information in real time through low-bandwidth channels. The criterion for this system was to locate the center of the eye within a 4 x 4 pixel area around the true value. This large criterion was acceptable in this case because the task was to transmit a section of the image containing the eye not actually locate the center of the eye. Craniofacial research based on the location of landmarks requires a more specific criterion and a smaller error tolerance. Hutchinson also tried preprocessing images with Sobel edge detection, binary and median thresholds, and inverting the gray scale images to improve the network performance. The technique of inverting the gray scale worked best (fifty seven of sixty images) since the pixel values of the eye on the original images were low, and bright regions (such as glare from eye glasses) dominated eye information. The most interesting aspect was that averaging the output of networks independently trained on different preprocessing techniques improved results. Because there were no images on which all networks failed, averaging the results of the two most successful individual systems produced the best results: correctly locating the central region of the eye on fifty nine of sixty images.

Because there is considerable inter-individual variability in landmark characteristics, Curry [29] proposed a combined man-machine strategy for automating landmark location. Landmarks were located manually on the first film of a series. Boundary detection and transformation algorithms were then used to locate landmarks on succeeding images. Curry showed that incorporating information from an individual patient and not just a universal definition can improve results. This information might be data from a previous or a stereo image. For example, manually locating landmarks on one film of a patient set could be used to aid a trained network to incorporate the unique features of each patient.

Research Objectives

This thesis describes the considerations and development of a specific system designed to identify and locate landmarks of interest using neural networks. The automation of landmark identification would substantially reduce the total amount of time required to acquire data. This work was tailored to a specific problem of craniofacial analysis: the location of reference and anatomic landmarks on cephalograms. An additional requirement was that the system be able to track landmarks over a longitudinal series of images. Each series consists of cephalograms taken over a time period in which significant anatomical growth and treatment associated change may be occurring.

The research objectives were to:

(1) Develop a two dimensional neural network for landmark location;

(2) Identify the empirical problems of locating landmarks; and

(3) Develop and test methods for fiducial location; and

(4) Develop and test methods for locating an anatomic landmark.

The first phase of the project was the development of a two dimensional neural network for landmark location. The network used two dimensional spatial information as both input, hidden and output layers. Network design utilized backpropagation, selective connectivity, and gradient descent training. The network was tested with a test cross pattern on a sample image background.

The second step was the development of training sets for fiducial landmarks. I tried several different training sets and investigated different aspects of the network connections. This included a study of the number of hidden layers necessary for optimal performance and the differences between two inter-layer connection schemes. I also evaluated the effects of some preprocessing techniques.

The third step was the development of training sets for a specific anatomic landmark. The landmark "upper incisor edge" was used as a first test of the performance of the network on anatomical landmarks. I report how the application of thresholds, normalization, and edge detection filters used to enhance images prior to training affected neural network performance. A series of experiments evaluated the network's ability to locate upper incisor edge when trained with different size training images.

Chapter 2

Neural Network Design

Neural networks were first mentioned in the mathematical models of McCulloch and Pitts more than 40 years ago [30]. In the last decade, interest in neural networks has arisen from the advancement of computer technology which makes implementation of neural networks computationally feasible. Neural networks receive input data and produce an output, but are inherently different from rule-based approaches. *A priori* rules or descriptions of how an output is produced from a particular input are not required. Rather, neural networks are characterized by a training algorithm which through a series of iterations determines the appropriate correlation between a set of inputs and a set of outputs. One of the most commonly used training algorithms is the backpropagation algorithm developed by Rumelhart and McClellend [21]..

I will start this chapter by introducing some general concepts about neural networks relevant to this thesis. These include: network layers, nodes, connections, and the training algorithm. Following that, I will discuss the specific features of my network and the reasons behind my choices.

General Neural Network Description

Network Layers

Generally, neural networks contain layers of three types (input, hidden, and output) with a feed forward connectivity. The first layer is the input layer, followed by a series of hidden layers, and finally an output layer (Figure 1). Each layer contains a number of nodes which are the processing units of the network. In my model each pixel value in the image is represented by an individual node in the input layer, and by corresponding nodes in the hidden and the output layers. Thus each layer contains the same number of nodes as pixels in the image. The network analyzes the combined interaction of these nodes. Each node receives an input, processes it, and passes an output through a link to one or more nodes in the next layer of the network. Each link or connection in the network has associated with it a unique weighting coefficient which represents the amount of influence one node has on another. The input layer receives values from an input vector or image outside of the network, and passes these values on to the next layer in the network. The hidden layers receive and send outputs within the system. Each output node sends a single output value out of the system. In addition to incoming nodes, each hidden and output node is also connected to a bias node. The bias node provides the learnable bias for each hidden and output node. The bias node is always turned on (set equal to 1.0).

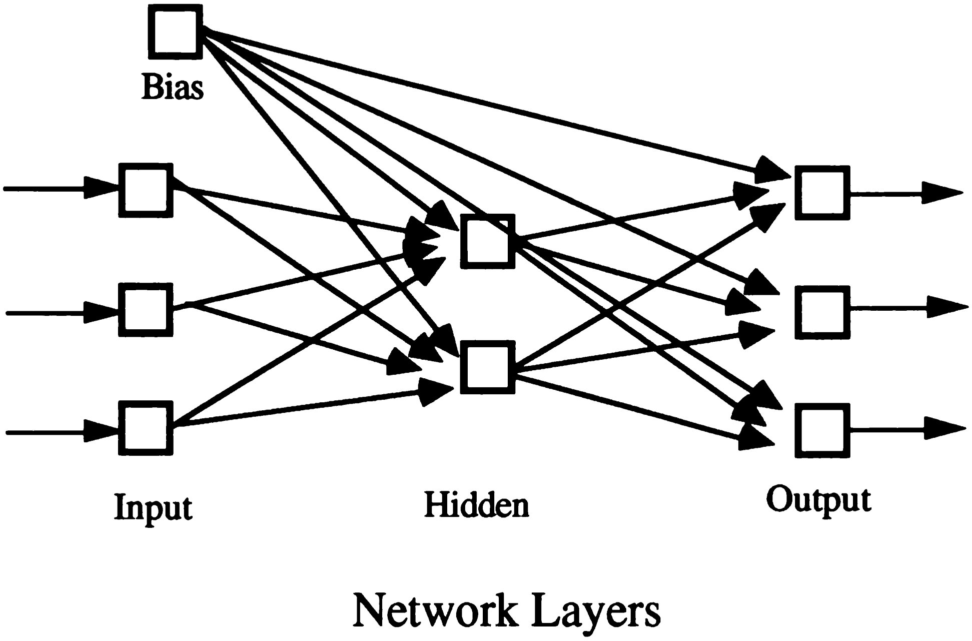


Figure 1. Diagram of a generic neural network with one hidden layer.

Activation level of nodes

The activation level of a node in the network can be calculated as a function of its initial state and the states of the external inputs. In my model activation values are continuous values in the interval [0,1]. The activity of any node in the hidden and output layers is usually a result of applying a squashing function to the weighted sum of the inputs into that node. Typically, the squashing function is a non-linear threshold, and in my model, a sigmoid response with an offset b:

$$f(x)=\frac{1}{1+e^{-(x-b)}}$$

Conceptually, processing in the network occurs in parallel which means that for each pass through the network the activation level of all nodes in a given layer is computed simultaneously. In reality, this is only possible on parallel machines and not on typical minicomputers or workstations.

The input stimulus is propagated through each successive layer of the network, resulting in a value at each node in the output layer which represents the likelihood value of the network. Eventually, the output layer produces the network's estimate of an output vector or image. This process is called forward propagation.

Network connections

The manner in which a node of one layer is connected to nodes in the adjacent layers has a strong effect on the responses of that node to an input. The two main parameters characterizing the connections are the weights of the links between nodes and the number of connections entering a node. The input to a given node is weighted by the coefficients of the links between that node and the nodes from the previous layer. Weight coefficients of the links in the network can be either positive or negative, corresponding to excitatory and inhibitory connections, or zero which indicates no connection between the two nodes. The absolute value of the weighting coefficient describes the strength of the link between the two nodes.

The number of links into any given node is an important consideration. A trade off exists between increased computation time due to an increased number of connections and the additional information available to the algorithm as a result of the increased number of connections. As the number of parameters, the nodes and connections in the network, increases the number of possible solutions also increases. It is therefore essential to attempt to keep the number of parameters as small as possible while still maintaining enough parameters to characterize the problem.

Learning Procedure

In order for the network to be useful, the appropriate weighting coefficients must be determined. The advantage of neural networks is that the initial weighting coefficients do not have to be predetermined. The process of changing these weights to the appropriate values is known as learning. The learning procedure of the network is a two step process using a series of training input/output pairs and both forward propagation and backward propagation. During the forward propagation step, the input stimulus (training input image) is propagated forward through the network, producing an output at each node in the network and finally resulting in the likelihood values of the location at the output layer.

The backward propagation algorithm is an iterative algorithm designed to minimize the mean square error between the actual output and the desired output. This involves the adjustment of weighting coefficients according to some training rule. The backpropagation algorithm used in this thesis is called the generalized delta rule [21]. First the residual absolute error between the desired output of the network (training output image) and the actual output of the network is calculated for each output node. Then this error value is passed to each unit in the network and the changes in the weighting coefficient for each connection into the output node are computed. The error in the next to last layer is then computed and propagated back to the preceding layer. This process is repeated for every layer in the network.. At each pass through the training set (an epoch), the order of training pairs is randomized so that no one pair has any more significance than any other. Through successive iterations the network learns to associate an output with a particular input configuration. This is similar to the way in which orthodontic students learn to locate a landmark on conventional analog films. The teacher shows the student a training set consisting of the images (training input image) and then points out the locations of the landmarks (training output image). In time, the student learns to associate a particular location in the image with a particular landmark. The ability to adapt or learn provides robustness to the network algorithm by compensating for minor variations in the characteristics of processing elements.

Network Model

Network Layers

In my model, the input layer consists of an n x n array of nodes that correspond to an n x n array of pixels in the image. The inputs to the system are the eight bit gray scale values of each pixel in the image. Each hidden layer is also represented by an n x n array of nodes. The output layer consists of a corresponding n x n image where the each value represents the likelihood that it is the location of the landmark (see Figure 2).

Alternatively, I could have designed the output layer to consist of a single node which indicated whether or not the landmark existed at the center of the input array. The value of the output node would decrease uniformly as the center of the input window moves away from the landmark position. This would have required much more care in designing training pairs -- one would not only need to locate the landmark but also calculate the distance from the center of the input window and an appropriately scaled output value. This is reasonable for locating a general area in an image or if any pixel at a given distance from the desired location is equally likely to be the correct location. However, the results from reliability studies [1-3] show that errors in human location of landmarks are not radially symmetrical around their centers but rather have unique distributions for different landmarks. Therefore, I chose an n x n output array which would designate the location of the landmark anywhere within the training input [12].

An advantage of using an n x n output array is that the incorrect responses the network might make may provide insight into which surrounding features contribute to the error. This information was in fact used to help design the training sets. I also thought the constraint of a single output node would be too limiting in a man-machine interface where the different likelihood values of a number of pixel locations would be valuable information. Using the network output as a weighted guide the human operator could be aided in his or her determination of the landmark.

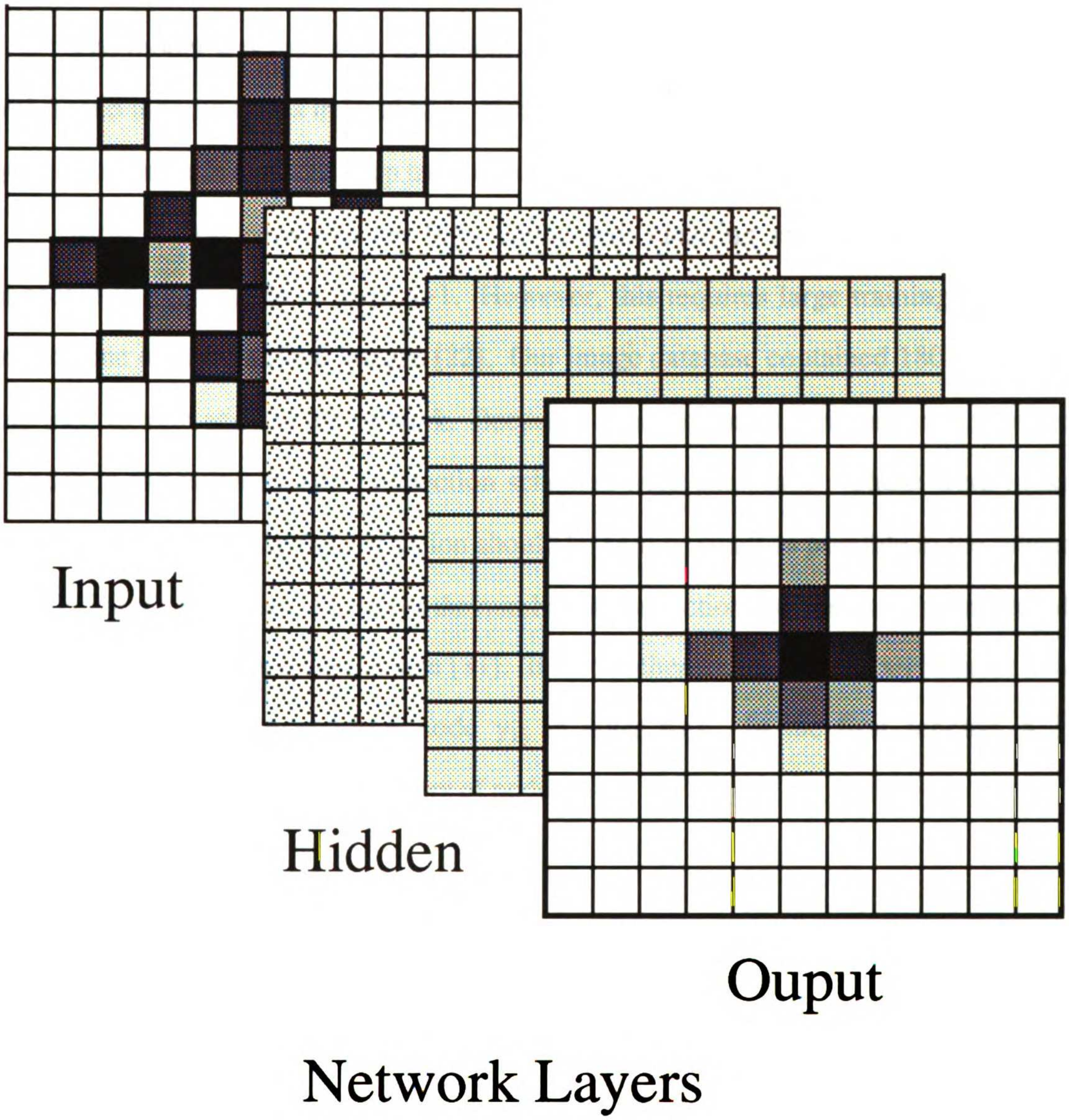


Figure 2. Diagram of the layers of the network model.

Inter-layer connectivity

The model was restricted to bottom-up processing. This means that only nodes in level n send input to nodes in level n+1. Nodes were not allowed to feedback into previous layers. This made it easier to keep track of which inputs were affecting a particular output.

Most neural networks are fully connected; meaning every node in layer n is connected to every node in layer n+1. However, this requires large training sets on the order of one or two hundred images [22]. Our image database contained 180 images that were used to train and test the network. Because I was concerned with limiting the number of nodes in the network, two connection schemes were examined. These were loosely based on the idea that the field of vision should be kept as small as possible in the interests of minimizing parameters. However the number of connections will have to be increased if the object of interest increases in size [31]. In both schemes, each node x_{ij} in layer n was connected to its corresponding node y_{ij} in layer n+1. In the first scheme, a 3x3 connection, the eight nearest neighbors of node x_{ij}

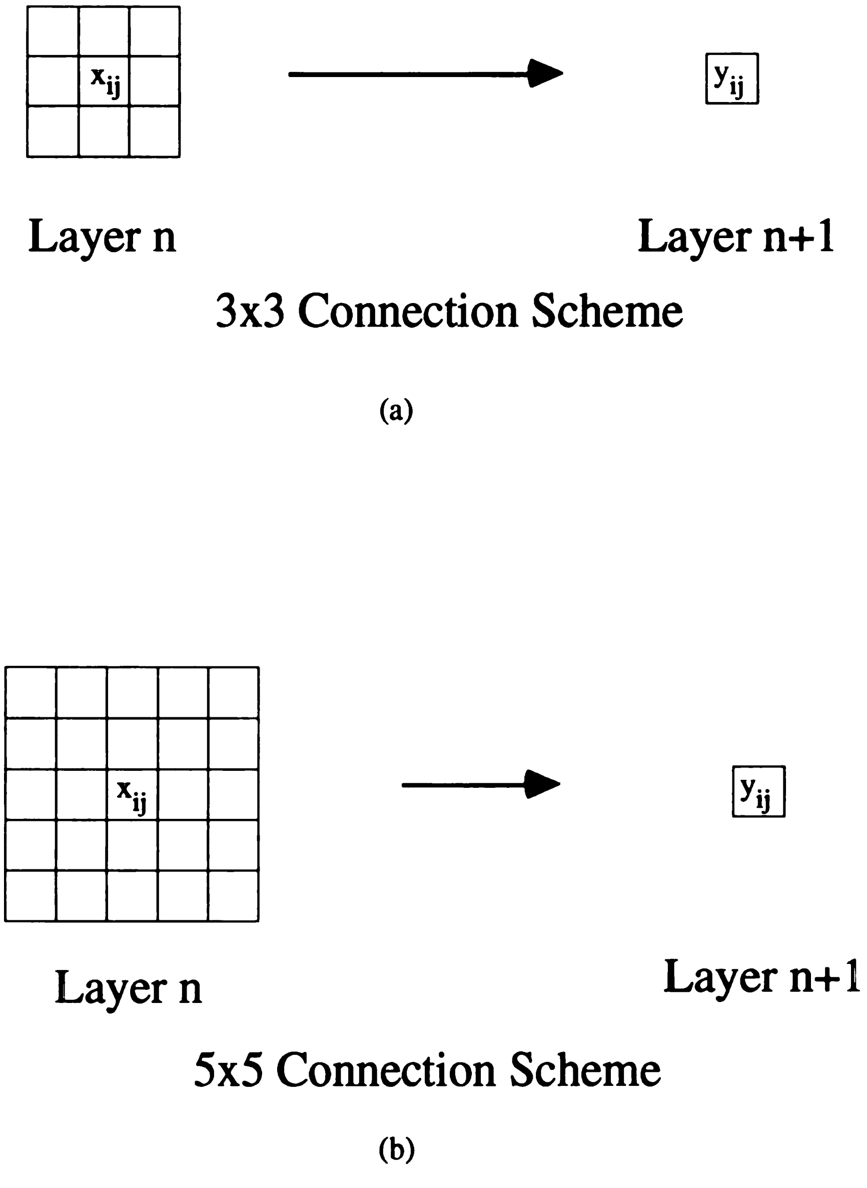
$$\begin{array}{ccc} x_{i-1j-1} & x_{ij-1} & x_{i+1j-1} \\ x_{i-1j} & & x_{i+1j} \\ x_{i-1j+1} & x_{ij+1} & x_{i+1j+1} \end{array}$$

were also connected to node y_{ij}.

In the second scheme, a 5x5 connection, the twenty four nearest neighbors of node x_{ij}

$$\begin{array}{ccccc}
x_{i-2j-2} & x_{i-1j-2} & x_{ij-2} & x_{i+1j-2} & x_{i+2j-2} \\
x_{i-2j-1} & x_{i-1j-1} & x_{ij-1} & x_{i+1j-1} & x_{i+2j-1} \\
x_{i-2j} & x_{i-1j} & & x_{i+1j} & x_{i+2j} \\
x_{i-2j+1} & x_{i-1j+1} & x_{ij+1} & x_{i+1j+1} & x_{i+2j+1} \\
x_{i-2j+2} & x_{i-1j+2} & x_{ij+2} & x_{i+1j+2} & x_{i+2j+2}
\end{array}$$

were also connected to node y_{ij}. These connection schemes are graphically depicted in Figure 3. While these connection schemes limit the amount of information each node can receive, they reduce the amount of time spent processing and the amount of memory necessary to encode the network.



(a)

(b)

Figure 3. The two connection schemes used in this thesis. (a) The 3x3 connection scheme. Pixel x_{ij} and its eight nearest neighbors are connected to pixel y_{ij} in the next layer. (b) The 5x5 connection scheme. Pixel x_{ij} and its 24 nearest neighbors are connected to pixel y_{ij} in the next layer.

The generalized delta rule

The generalized delta rule changes weighting coefficients in proportion to the difference between the desired output and the actual output. Assume that the actual output of the system is a vector $y_0, y_1, ...y_n$ and the desired output vector of our system is $d_0, d_1, ...d_n$. Initially by forward propagation, one output node (see Figure 4), y_j (j=0,...n) sums N weighted inputs x_j (j=0,...n), and passes the value on through a sigmoid non-linearity function

$$y_j = f(x) = \frac{1}{1 + \exp(-(b + \sum w_{ji} x_i))} \qquad \text{(EQ 1)}$$

where w_{ji} is the weight from node i to node j, and the node is characterized by an internal bias node b. Initially the weighting coefficients are set at random between +1.0 and -1.0.

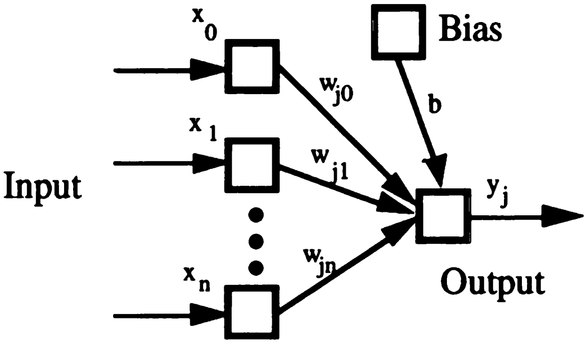


Figure 4. Diagram of the input into a single node. The node y_j could be either an output node, or a hidden node.

Learning occurs by adjusting the values of the weighting coefficients to minimize the error [32] :

$$E = \frac{1}{2}\sum_{m=0}^{n}(d_m - y_m)$$

This equation represents the error for one input/output training pair. The network is designed to minimize the sum of all such errors over all n input/output training pairs. The $(d_m - y_m)$ term is dependent on w_{ji} thus:

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial w_{ji}},$$

$$\frac{\partial E}{\partial w_{ji}} = -(d_j - y_j)\frac{\partial y_j}{\partial w_{ji}},$$

$$\frac{\partial E}{\partial w_{ji}} = -(d_j - y_j)f'(x_j)x_i.$$

Where x_i is the output from the preceding layer and f(x) was the exponential function given in EQ 1, and its derivative is

$$f'(x) = f(x)[1 - f(x)].$$

Therefore the error derivative becomes

$$\frac{\partial E}{\partial w_{ji}} = -(d_j - y_j)y_j(1 - y_j)x_i.$$

The delta rule then becomes

$$\delta_j = (d_j - y_j)y_j(1 - y_j).$$

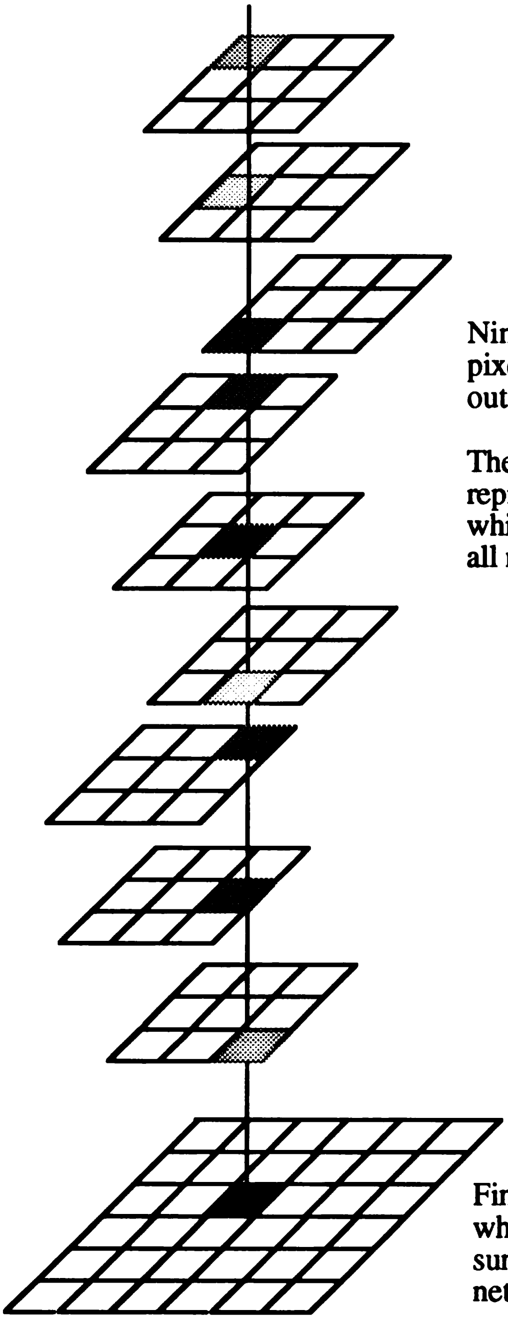
Learning occurs by starting at the output layer and working backward to the first hidden layer adjusting the weights by

$$w_{ji}(t+1) = w_{ji}(t) + \eta\delta_j x_i + \alpha\Delta w_{ji}$$

Where $w_{ji}(t)$ is the weight from node j to node i at time t, x_i is either the output of node i (and therefore the input to node j) or is an external input, η is the learning rate, δ is the error term for node j, and α is a weighting factor (set at 0.5) of the momentum term $\alpha\Delta w_{ij}$.

Application of the network

After training on n x n pixel images, the network can be applied to a larger M x M pixel image and producing an output M x M pixel image, where M > n. In addition, *a priori* knowledge of anatomic relationships can be used to limit the scope of the M x M pixel image, reducing the amount of time spent segmenting an image. Before applying the neural network, the M x M pixel image is segmented into sub-images of size n x n, the size of the training images. For every pixel p in the interior (M - n) x (M-n) of the image, an n x n sub-image surrounding that pixel is fed into the network. The pixel values from the sub-image are forward propagated through the network producing an output sub-image of size n x n. Each output sub-image is centered on and added to pixel p_{ij} and the surrounding n x n pixels of the output image . This way the value of each pixel in the final M x M pixel output image is the summation of its value from n^2 sub-images. For example, if M equals six and n equals three then sixteen subimages will be forward propagated through the network. The final output image is a summation of the values of all sixteen subimages. For a given pixel p (Figure 5) nine subimages contribute to the final output value. This method of adding the output from multiple subimages adds robustness to the network and makes the network somewhat shift invariant. For the large part though, this shift invariance was accounted for by the careful selection of training images.



Nine n x n (n=3) pixel network outputs.

The shaded pixel represents pixel p which is included in all nine subimages.

Final M x M (M=6) pixel Output Image where each pixel p (shown in black) is the sum of its corresponding pixels in nine network output subimages.

Figure 5. Application of the network to an M x M image. Every pixel p in the final output image is the result of the summation of results of pixel p's output in nine subimages that had been forward propagated through the network.

Chapter 3

Implementation

Materials

Images

An analog x-ray film is a thin transparent plastic sheet coated with silver salts embedded in a gelatin emulsion. Images are produced by a chemical transformation of the silver salts by the x-rays. The number of chemical transformations is related to the absorption characteristics of the body tissues. X-rays are emitted from a source, pass through the body, and strike the film producing a projection image. Since bone absorbs more x-rays than fat, fewer x-rays penetrate through the bone to the film. Fewer x-rays mean fewer chemical transformations and results in a lighter shade of gray on the bone region of the image compared to the region depicting fat or tissue. In other words, lighter areas of the film represent high-density body tissues, and darker areas represent low-density body tissues (Figure 6). Analog x-ray film can display in excess of 10^3 shades of gray. These shades -- which range from pure black to white -- are blended together continuously.

A digital image is a mathematical matrix containing both spatial and intensity information which can be displayed on a computer monitor. Acquiring a digital image is analogous to placing an invisible grid over a photograph or analog film, reading the information about the brightness of the image at each square of the grid, and assigning each corresponding matrix location (pixel) in the digital image with a single numerical value: real, non-negative, and limited in magnitude. Since computers can display only a discrete number of gray levels (in this study: 256), the brightness value stored in each pixel is scaled to one of these values. Each digital image is stored as a file in computer memory and can be displayed on a computer monitor. Duplicate images can be made at will by a simple computer command to copy a file.

The images used in this dissertation were lateral skull head films taken with the right side of the patient's head oriented toward the x-ray tube and the left side toward the film. The patient's head was placed in a head holder which oriented the patient in a fixed position relative to both x-ray tube and film. The central ray of the x-ray source passed through the central axis of the rod in the head holder and the porion-porion anatomic axis.

Our image database contains digitized x-ray images for thirty patients. For each patient, films were taken at different ages. There are approximately 180 images in the database spanning the growth period from eight to sixteen years at approximately annual intervals. Each image contains four optically placed reference markers in a specifically defined configuration. These images were used to train and test the neural network algorithm.

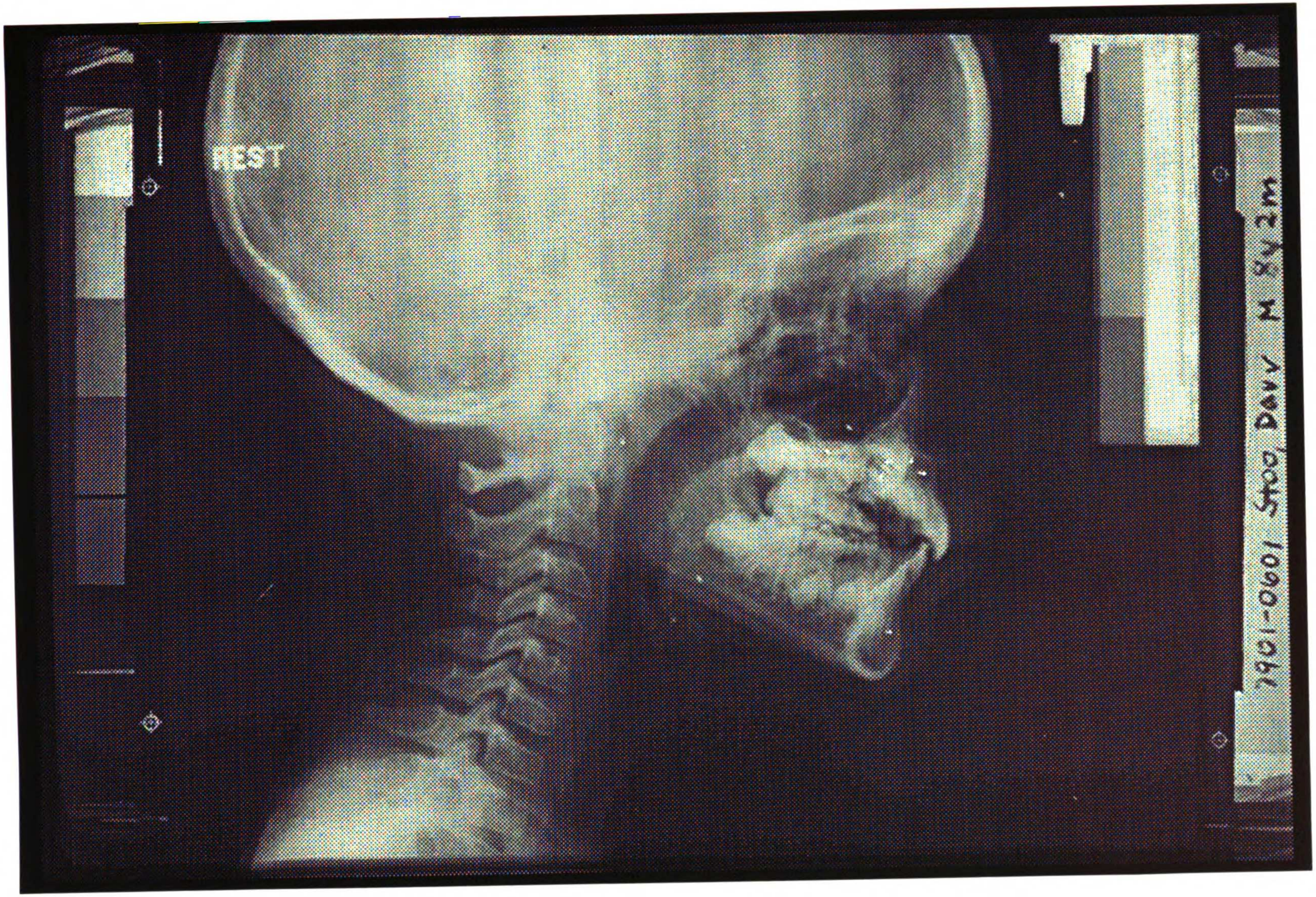


Figure 6. A digital x-ray cephalogram.

Hardware

The four major components of our system necessary for acquiring, manipulating, and displaying digital images from analog film are: (1) a personal computer (PC), (2) an image acquisition system, (3) a frame grabber, and (4) a display monitor. Our initial hardware system used a 386 Dell PC with an 80386 processor. The original x-ray cephalograms were photographed with a high resolution 70mm camera. The 70mm negative was then captured in an analog signal by a Pulnix TM-840 array camera. The output of the Pulnix camera was transformed to 512 x 480 pixel digital format with an eight bit PCplusVision frame grabber. The digital images have a pixel size of .28mm in the vertical direction and .33mm in the horizontal direction. The frame grabber has two main components: hardware for digitizing the analog signal and frame memory for temporary image storage. The digitizing hardware samples the analog signal and converts each sample to an eight bit integer value which represents the brightness or intensity level for one pixel in the image. The digital data is then temporarily stored in memory on the frame grabber board. This memory is accessible by software and can be copied to a file on the computer hard disk. The images are stored as binary random access files, with a header that identifies the patient, time point (which image in the age series) and the locations of four reference markers in the image. I developed a software package (BIBASE) on the PC to display the image via the PCplusVision frame grabber on an RGB SONY Trinitron monitor.

Software

The primary task of the BIBASE software is to acquire and display digital images. In addition to displaying images, the BIBASE software can also display other information from the database, including landmark location data and the associated angular and distance measurements [33]. The software also provides the ability to design the network training sets. Portions of the digital images can be easily selected and stored in files for later use as training or test images. Preprocessing of the images can be performed by BIBASE or by the neural network software.

The neural network software was developed in C on a Sun SPARCII which was networked to the PC. The software allows the user to vary the size of the training images, the size of the test images, the connection scheme, the number of hidden layers, and what kind of preprocessing (if any) should be done. Weighting coefficients can be stored in a file to be used again on successive runs of the network. The network segments the test images and produces network output images. The network output for each pixel in the image is the likelihood that it is the landmark location.

Method

This section describes the preprocessing techniques, calculation of standard deviations, definitions of the landmarks, and selection of training sets used in this thesis. I examined the following preprocessing techniques: Laplacian convolution, binary thresholding, normalization, and Sobel filtering. Standard deviation equations are used to assess multiple judges' estimating error in landmark location. The reference and anatomic landmarks used in this thesis are described and defined. The method of designing training sets and determining their size is described for each landmark.

Preprocessing and Enhancing Digital Images

In addition to providing clinical research investigators with the ability to duplicate and share records, an important advantage of using digital images is that they can be enhanced with a number of mathematical manipulations at no risk to the original. Images can be modified to improve their overall contrast or appearance or they can be filtered to bring out hidden or blurred features. The distribution of gray levels or pixel intensities in an image affects its appearance; therefore, any processing that changes this distribution will change its appearance. I will classify enhancing techniques into two categories: single-pixel processing which alters the intensity values of individual pixels and neighborhood processing which can sharpen certain features of the image, such as the edges of objects.

Single-Pixel Processing

Single-pixel processing techniques perform mathematical operations on each pixel in the image independent of other pixels in the image. I used two single-pixel processing techniques in this thesis: (1) binary thresholding and (2) normalization. If the landmark has intensities distinct from those of the background, binary thresholding can be used to create a binary image. Two threshold values are selected: a high and a low cut-off value. The pixel values above the high cut-off and below the low cut-off are set to 0, and the pixel values between the two values are set to 255. I used binary thresholding in an attempt to isolate the fiducials from the surrounding image.

The normalization of image intensities is used to expand the intensity range of an image to the full gray scale capabilities of the display. In this case, if the intensity range of an image is (min, max) a subset of (0,255) for an eight bit display, new pixel values (*newvalue*) can be calculated from old pixel values (*oldvalue*) with the following equation:

$$newvalue = 255 - \frac{(\max - oldvalue) * 255}{(\max - \min)}$$

The normalized image has an intensity range of (0,255), and an increased contrast between two pixels that vary in intensity.

Neighborhood processing

While single-pixel processing can be used to alter the intensity values of individual pixels, neighborhood processing techniques are most often used to change the overall sharpness of an image. Neighborhood processing differs from single-pixel processing because it calculates new values based on a group of pixels in the matrix. If a 3 x 3 filter is centered on pixel x(i,j) of the input image, the value of pixel y(i,j) in the new image is calculated as a function of x(i,j) and its eight nearest neighbors. Each of these nine pixels is multiplied by a corresponding coefficient in the filter, the values are summed, and the result placed in pixel y(i,j) of the output image (Figure 7). In effect, a weighted sum of the neighborhood is calculated for each pixel in the image. If the filter is moved over each pixel (i,j) of the image and applied in the manner described this process is called a convolution. By changing the weights of the filter coefficients, neighboring pixels will have a greater or lesser effect on the overall value.

a	b	c
d	e	f
g	h	i

Filter A

	1	2	3	4	5	6	7	8	9	10	11
1											
2											
3											
4											
5											
6											
7					x						
8											
9											
10											
11											

Image X

Convolution of position x(5,7) of Image X with Filter A:

$$\begin{gathered} y(5,7) = a * x(4,6) + b * x(5,6) + c * x(6,6) + \\ d * x(4,7) + e * x(5,7) + f * x(6,7) + \\ g * x(4,8) + h * x(5,8) + i * x(6,8) \end{gathered}$$

Figure 7. The convolution of a filter with an image.

Two common neighborhood processing techniques are high-pass and low-pass filtering. A high-pass filter emphasizes the edges or transitions in an image. In other words, it sharpens an image but also increases the amount of high-pass noise. If a high-pass filter is applied to areas with uniform intensity, it will have no effect on the image; on the other hand, in areas of the image where there is an intensity gradient, the filter will increase this gradient. A low-pass filter smoothes or blurs the sharp white to black (or black to white) transitions resulting in a blurred edge.

Averaging can be used to reduce the size of the network while still providing some useful information to the network. A border averaging scheme was used to reduce the number of nodes in the input images while still encoding the information from a larger section of the image. A 21 x 21 pixel region was reduced to a 13 x 13 pixel area by averaging the six outer layers while maintaining the inner 9 x 9 pixels at full resolution. This is depicted in Figure 8 where the two outer layers of the 13 x 13 pixel image are the result of averaging 3 x 3 pixel regions surrounding the full resolution center of the original image. Only the corner regions are shown but each pixel in the border of the 13 x 13 image is the average of a 3 x 3 pixel region in the original image.

This provides the network some of the border information while reducing the number of nodes by 272. Depending on the connection scheme of the network this can have a significant effect on the number of parameters in the system.

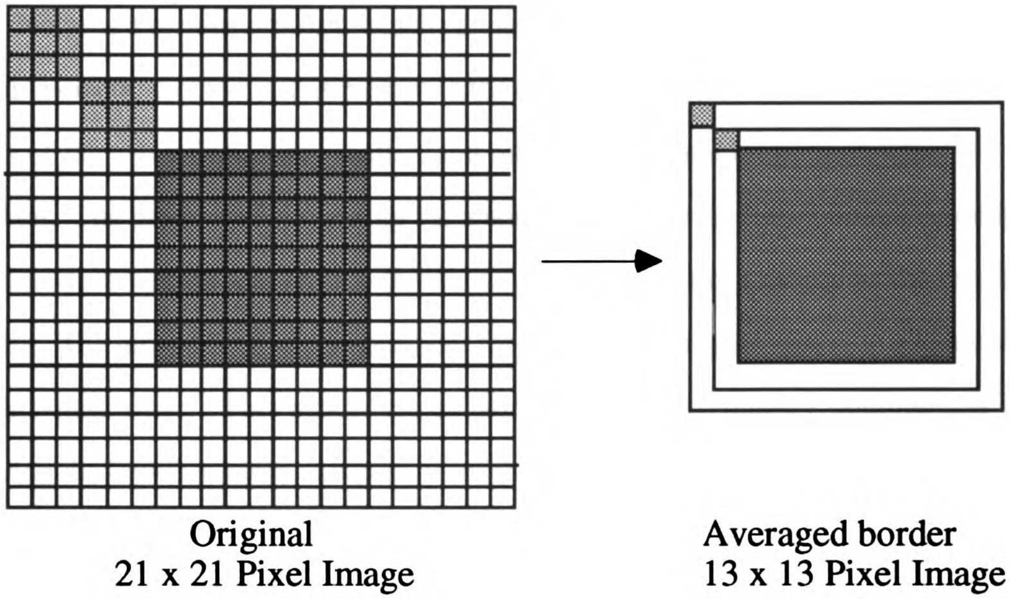


Figure 8. A border averaging scheme. The inner 9 x 9 pixels are maintained at full resolution (in dark gray) while the pixels in the border of the 13 x 13 image are the average of a 3 x 3 pixel region in the original image (in light gray). Only two pixels are shown here but all of the pixels in the 2 outer layers of the 13 x 13 image are the result of averaging a 3 x 3 pixel region in the original image.

Optimal Enhancements for Craniofacial X-ray Imaging

Noise

Noise is a serious problem which can make landmark location difficult and increase estimating errors. Some noise is present in the film itself due to the discontinuities in silver crystals of the x-ray film, x-ray scatter, and other artifacts and limitations when the x-ray was taken. The process of converting the analog x-ray films to digital images used in this study introduced additional error -- the digital images were not as sharp as the original x-ray films. The most effective enhancement for the digital x-ray images is to add a constant to a high-pass filter. This procedure enhances the high frequency components while preserving the low-frequency components of the image [34].

This technique was applied to images prior to training or testing by the network. At the very least, every image was convolved with the following filter:

$$\begin{matrix} -1 & -1 & -1 \\ -1 & 12 & -1 \\ -1 & -1 & -1 \end{matrix}$$

Thus the filter sharpened the edges in the image by convolving the original image with a Laplacian and a constant. By selectively amplifying the image's high spatial frequency components, the image contrast at the edges is enhanced in all directions. This convolution also tends to add some graininess to the image.

Edge detection

Edge detection routines were examined as a possible enhancement for the anatomic landmarks. The simplest edge detector uses gradient operators which compute the difference of intensity values along a line in the X or Y direction. In the X direction the simplest function would be

$$\Delta_x f(x,y) = f(x,y) - f(x-1,y).$$

The Sobel operator, a 3x3 edge detector, which was used to preprocess the anatomic images in one experiment can be described as a convolution with the following arrays:

$$\begin{array}{ccc} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{array} \qquad\qquad \begin{array}{ccc} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{array}$$

Calculations of Standard Deviations

Standard deviations and standard errors used in this thesis were calculated separately in their X and Y components. Standard deviations (SD) were calculated using

$$SD = \sqrt{\frac{\Sigma d^2}{N(K-1)}}$$

where N is the number of test images, K is the number of judges, and d is the deviation of an individual judge's estimate from the mean value for that particular image. The standard error of the mean (SE) is computed by

$$SE = \frac{SD}{\sqrt{N}}$$

The standard deviation of network results for N test images was calculated as

$$SD = \sqrt{\frac{\sum(x - best\_estimate)^2}{N-1}}$$

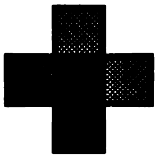
where x is the location found by the network and *best_estimate* is the location accepted as the "true" value of the landmark for a particular image. The determination of the *best estimate* of a landmark will be discussed in Chapter 4.

Landmark Definitions

Two categories of input were examined: reference landmarks such as the test cross and the fiducials, and anatomic landmarks which can be more conceptual in nature. Each of these classes of input has unique properties and problems associated with its location.

The test cross was designed to be a simple test of the network connectivity. This was not a landmark found on any of the images, but rather a pattern that was superimposed on a set of test images. The plus sign was not of uniform intensity, rather the values were random intensities between 150 and 255.

The shape was a simple five pixel form:



The network was trained to locate the single pixel at the center of the test cross.

Fiducials are also well defined landmarks which usually, though not always, exist overlaid upon a uniform background. Each fiducial consists of a cross within a circle. The task is to locate the position which corresponds to the center of the circle where the crosshairs intersect. The diameter of the fiducial is approximately eleven pixels wide. Since the edges of the crosshairs are wider than a single pixel, often the optimal location lies somewhere between pixels as seen in Figure 9(a).

Anatomic landmarks are more difficult to locate that reference markers because of the natural variability in the form and shape of the landmarks within and between patients. The majority of craniofacial landmarks are located on bony or soft tissue surfaces which appear in the image as an edge. The detectability of a given landmark to the human eye is highly dependent on image quality and the training of the judge.

The anatomic landmark studied in this thesis, Upper Incisor Edge (UIEdge), is defined as the tip of the incisal edge of the more anteriorly placed upper central incisor [3]. The human estimates for this landmark are good compared to other anatomic landmarks because there is generally a sharp contrast between the edge and the surrounding background. The edge also has a high degree of curvature. The standard deviation for UIEdge is smaller than for most other anatomic landmarks [3]. For this reason, UIEdge was chosen as a landmark for the network to locate.

Training sets

The training sets consist of input/output image pairs. The input training image is an n x n pixel section of the original image, where n was either 11 or 21. The output training image is an n x n array of nodes one of which specifies the location of the landmark in the input image (i.e. the node representing the landmark location is set to 1, all others are set to 0). Separate training sets were developed for each landmark. The training sets contained both positive and negative matches. Positive matches contain the landmark and negative matches are areas of the surrounding image that do not contain the landmark. The output training image for the negative matches contained nodes which were all set to zero.

The process of building a training set was iterative. An initial training set was designed. Positive and negative training pairs were selected from images of three patients. The training pairs were initially constructed under the assumption that definitive features used by a human to locate a landmark might also aid the algorithm. I attempted to incorporate the variability of image contrast as well as landmark form in the training pairs. In most cases, a positive match occurred within the center 3 x 3 region of the n x n matrix. Negative matches were selected from possibly confounding anatomy near the landmark. Some of the negative matches contained areas that might be considered part of the landmark, but the actual location of the landmark was not within the boundaries of these images. The network was then tested on four new images from the same three patients. If non-landmark sites were selected, these areas were added to the training set as negative matches and the network was retrained. This process was repeated until the network was able to locate the landmark on the training set of images. The size of the training sets varied but in general they contained approximately forty training pairs. The network was then tested on thirty six images from four entirely new patients. Throughout this study, the training and test images came from different patients.

When judges locate landmarks, systematic errors are introduced into their results [1-3]. These errors are dependent on landmark definition and judge training and experience. If a network is trained with a training set that has systematic errors these errors would be expected to affect network performance on data outside of the training set. Ideally, the training set should be a "standard" which could be achieved by averaging five expert judges' location of the landmark on the training set. I originally intended to train the network with the averaged results of two judges who had located the landmarks repeatedly on photographic enlargements of 70mm negatives of the original x-ray films. These were the same negatives which were used by our visual acquisition system to produce digital images. However scaling factors and human error locating the fiducials on the digital images made the transformation and fit of the data to the digital images less than optimal. Therefore in this thesis, the network was trained to the results of one judge's replicate locations of the landmark.

The more parameters (nodes and connections), there are in a network, the more solutions there are to which the network can converge unsuccessfully. With more parameters, larger and larger training sets are required to characterize the variability of the data set. In general it is important to keep the number of nodes as small as possible. 11 x 11 pixel images were used as the training pairs for the fiducial and UIEdge images. In the case of the fiducials, this was sufficient to contain the entire landmark. However in the case of UIEdge, only the lower portion of the tooth was included. For this reason, the network was also trained with 21 x 21 pixel training images, and 13 x 13 pixel training images that contained some discrete and some averaged information from 21 x 21 pixel images for the UIEdge landmark. Samples of these three types of training input images are shown in Figure 9.

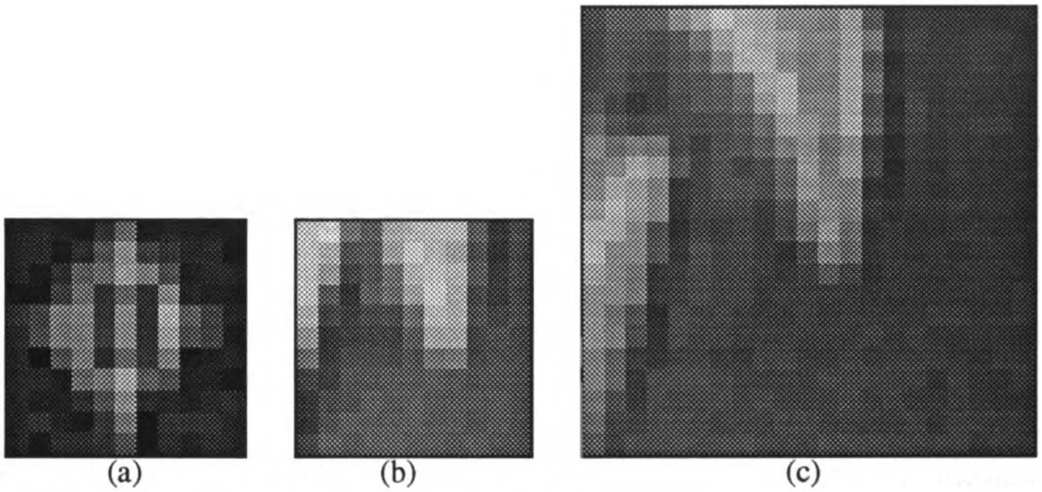


Figure 9. Examples of three training input images (a) fiducial, (b) 11 x 11 UIEdge, (c) 21 x 21 UIEdge.

Chapter 4

Criteria for Evaluating the Success of the Network

When the network was tested initially, an unambiguously defined test pattern was used. Unfortunately, real landmarks do not have an unambiguous location that all human judges can agree on. There are two reasons for this: (1) the desired value may lie somewhere between two pixels making it difficult to decide which pixel to select, or (2) judges may differ in their opinions of the location of the landmark in question. If judges disagree on a landmark's location then how does one determine whether the network has located the landmark? As a result of this uncertainty, there are two issues that need to be considered in order to evaluate network performance: (1) what is the *best estimate* of a landmark on any image, and (2) what is the acceptable error interval surrounding the *best estimate*.

Determination of the Best Estimate of Landmark Location on the Test Images

Since there is no completely unambiguous location for any landmark, it was necessary to define a *best estimate* of the landmarks location on each of the test images. In order to understand performance of the network, network results were compared to (1) the average of two independent estimates by the judge whose results were used to train the net (training judge mean) and (2) the averaged independent estimates of the training judge and two other judges (three judges mean). For the network to be a useful tool, the results must approximate the error distribution of human judges.

A note on the scale used in this thesis (e.g. Figure 10). The network treated each pixel in the image as a node and was designed to determine which node or pixel was the most likely location of the landmark. Since the performance of the network, and not the particular images used in this study, was being evaluated all results are reported with respect to a scale of the pixel size. This was done under the assumption that for higher resolution images, a higher degree of accuracy could be achieved. Later, if further accuracy or precision is desired, sub-pixel analysis techniques could be used [29, 35-37].

Assessment of test images by the Training Judge

The training judge repeated each landmark location procedure for all thirty six images and the mean value of the two estimates was considered the training judge's *best estimate* of the landmark for that image. In Figures 10 and 11 the mean value for each of the thirty six images has been translated to the origin of the scatter plot and each of the judge's two estimates has been plotted. The results for the fiducials are plotted in Figure 10; the estimations of UIEdge in Figure 11 (The graphs are plotted in units of pixels where one pixel corresponds to 0.33mm in X and 0.28mm in Y).

The standard deviations and standard errors of the training judge for fiducials and UIEdge are presented in Table 1. The 95 percent confidence intervals for fiducials are 0.92 pixels in X and 0.74 pixels in Y. The 95 percent confidence intervals for UIEdge are 1.40 pixels in X and 1.34 pixels in Y.

Standard Deviations (σ_T) for Training Judge's replicate estimates of Fiducial and UIEdge on 36 test images used in network experiments.

Landmark	SD	SDx	SDy	SE
Fiducial	0.58	0.46 (0.15mm)	0.37 (0.10mm)	0.17 ± 0.19
UIEdge	0.56	0.70 (0.23mm)	0.67 (0.19mm)	0.40 ± 0.30

Table 1. Training judge's standard deviations for replicate estimates of fiducial and UIEdge on thirty six test images. Results are reported in pixels with mm conversions in parentheses.

Distribution of Training Judge's Replicated Estimates of Fiducial
on 36 test images

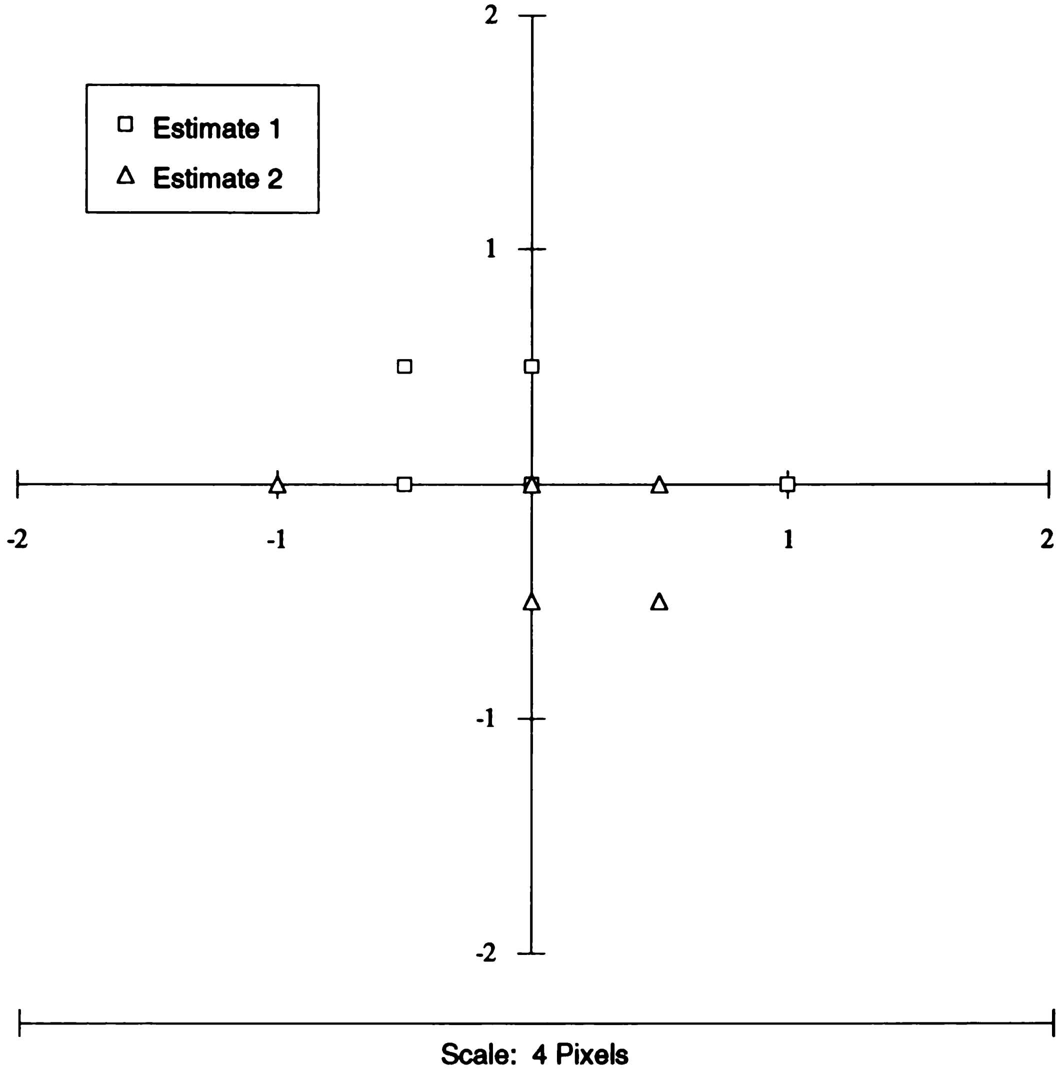


Figure 10. Training Judge's replicated estimates of fiducial on 36 test images. For each image, a mean of the two estimates was calculated and translated to the origin and the two estimates are plotted.

Distribution of Training Judge's Replicated Estimates of UIEdge on 36 test images

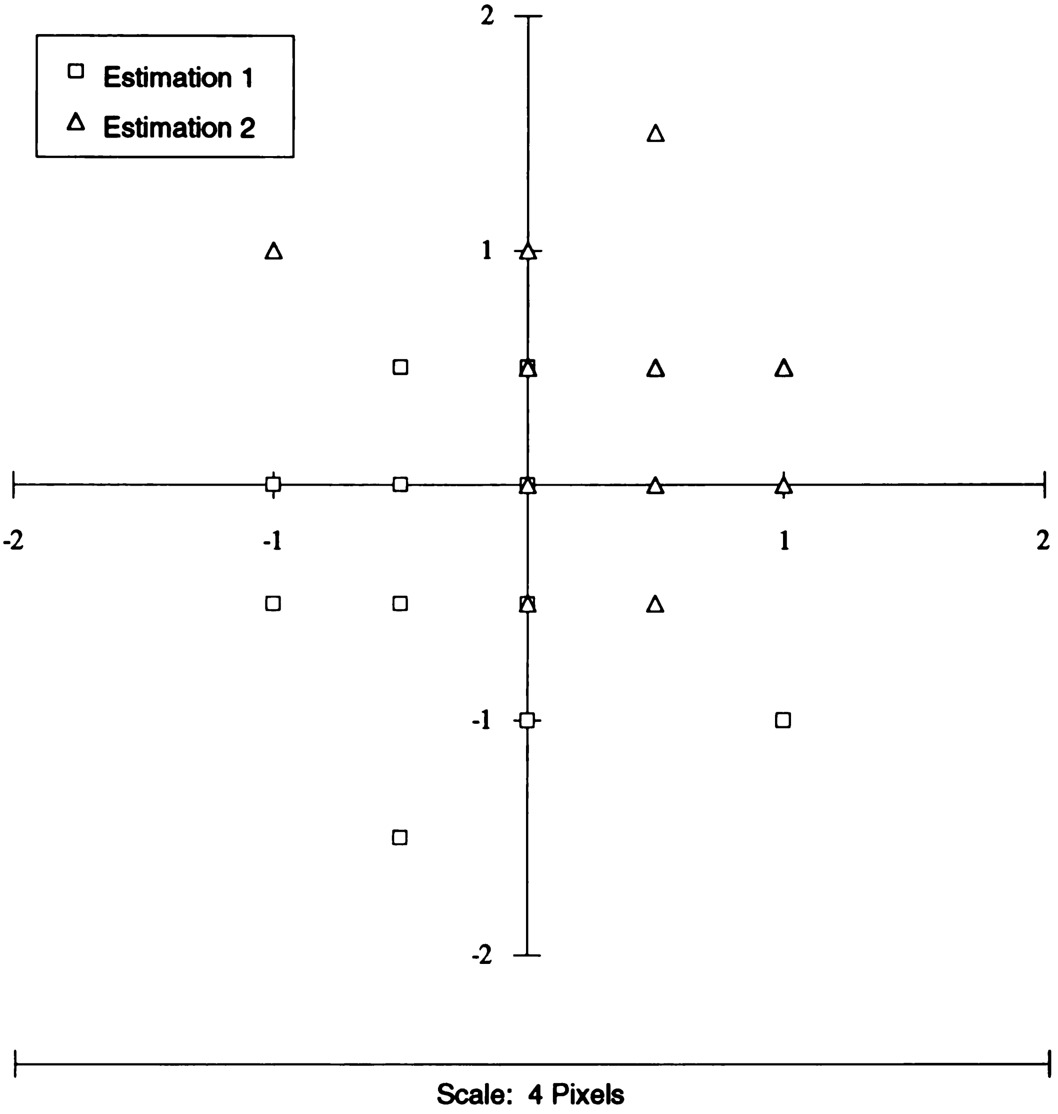


Figure 11. Training Judge's replicated estimates of UIEdge on 36 test images. For each image, a mean of the two estimates was calculated and translated to the origin and the two estimates are plotted.

Assessment of test images by Three Judges

For each of the thirty six test images used to test the network, three judges independently located the landmark of interest. For each image, the mean value of the three judges' estimates was considered the *best estimate* of the landmark. This mean, along with the means from the other test images, was then translated to the origin of the scatter plots shown in Figures 12 and 13. Judge number one was the training judge. It is evident in Figure 13 that each judge had a unique distribution of estimating errors, probably reflecting differences in the conceptual definition of UIEdge. These systematic biases and error distributions can also be partially explained by the unfamiliarity of the judges with the cursor; which was an arrow rather that the usual cross hair cursor. However, judges were required to digitize a control point in an attempt to minimize the possible cursor bias. Some of the error can also be attributed to the resolution of the images, as was also evident in the 1991 Reliability Study. Yet despite all of these factors, there remains the systematic bias of each judge.

The statistics of standard deviation and standard error for location of fiducials and UIEdge are presented in Table 2. The 95 percent confidence intervals (twice the standard deviation) are 1.18 pixels in X and 1.12 pixels in Y for fiducials. The 95 percent confidence intervals for UIEdge are 2.02 pixels in X and 1.62 pixels in Y. This provides a perspective within which the systematic bias of the estimating error of the training judge can be compared to the two other judges.

Standard Deviations for Three Judges' estimates of Fiducial and UIEdge on 36 test images used in network experiments.

Landmark	SD	SDx	SDy	SE
Fiducial	0.75	0.59 (0.19mm)	0.56 (.16mm)	0.37 ± 0.25
UIEdge	1.22	1.01 (0.33mm)	0.81 (0.23mm)	0.98 ± 0.41

Table 2. Three Judges' standard deviations of estimates for fiducial and UIEdge on thirty six test images. Results are reported in pixels with mm conversions in parentheses.

Distribution of Three Judges' Estimates of Fiducial
on 36 test images

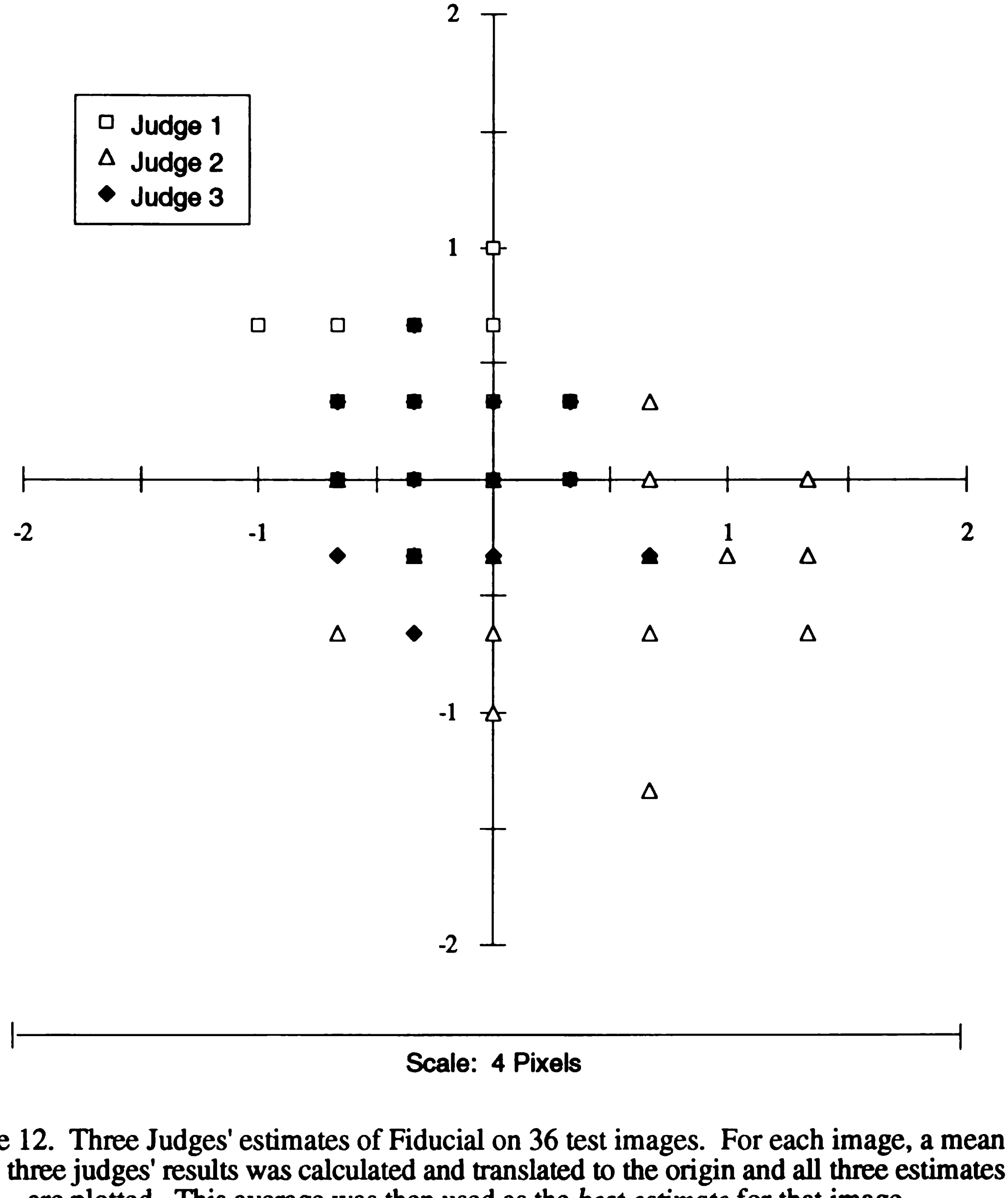


Figure 12. Three Judges' estimates of Fiducial on 36 test images. For each image, a mean of the three judges' results was calculated and translated to the origin and all three estimates are plotted. This average was then used as the *best estimate* for that image.

Distribution of Three Judges' Estimates of UIEdge.
on 36 test images

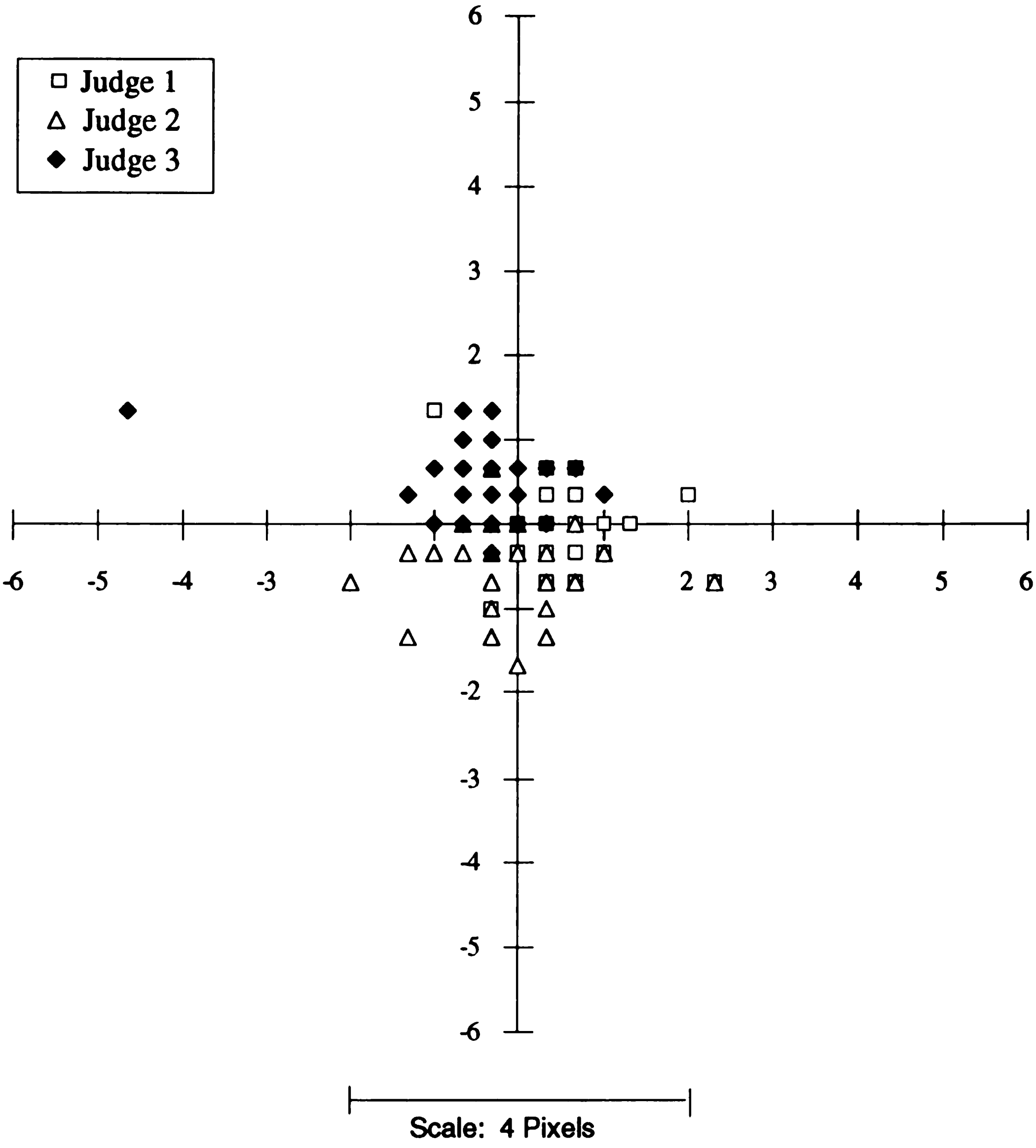


Figure 13. Three Judges' estimates of UIEdge on 36 test images. For each image, a mean of the three judges' estimates was calculated and translated to the origin and all three estimates are plotted. This average was then used as the *best estimate* for that image.

Determination of the Confidence Interval

Estimates of human error for each landmark were used to set confidence intervals for the network. If the network's result for a landmark falls within the confidence interval around the *best estimate*, the network is considered successful for that image. In addition to the standard deviation of the training judge, two other estimates of human errors from different reliability studies were used to evaluate results. The results from the 1971 Baumrind and Frantz Reliability study and the 1991 Baumrind and Baker Reliability study are reviewed in this section. Only the results pertaining to fiducials and UIEdge are presented.

1971 Baumrind and Frantz Reliability Study

The results from the 1971 Baumrind and Frantz Reliability Study for UIEdge are shown in Figure 14 [3]. The study did not report findings for fiducials. Separate X and Y components (SDx and SDy) of the total standard deviation (SD) were isolated and are presented along with the standard error (SE) in Table 3. One conclusion was that at approximately the 95 percent level of confidence the *best estimate* of the landmark will be within twice the standard error of the mean. In other words, the probability that a given estimate will differ from the true value by more than two times the standard deviations for that landmark is approximately 0.05. The study also indicated that given a task of locating sixteen landmarks, there is a fifty six percent probability that at least one value deviates by more than two standard deviations from the true value. However, by repeating the task of landmark location the task can, on average, be completed without detectable errors with 98 percent confidence.

Thus, one way to determine the success of the network is whether values fall within the 95 percent level of confidence range as determined by this study.

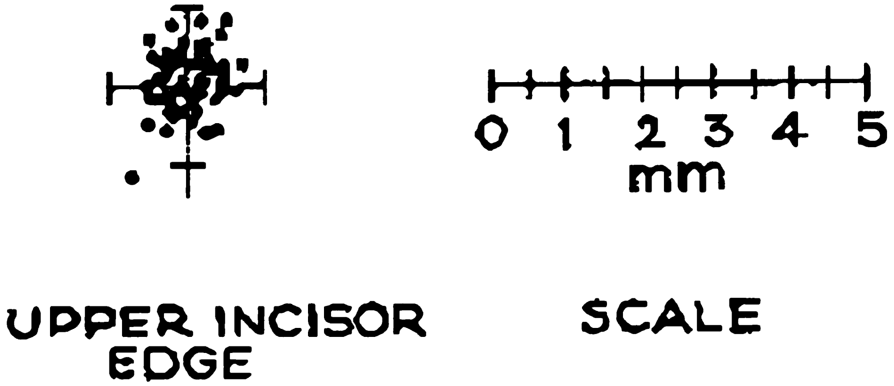


Figure 14. Distribution of estimating errors of the dental landmark UIEdge from the 1971 Baumrind and Frantz Reliability Study. (Source:[3]).

Standard Deviations (σ_{71}) for estimates of UIEdge as reported by the 1971 Baumrind and Frantz Reliability Study (reported in estimated pixels)

Landmark	SD	SDx	SDy	SE
UIEdge	1.16 (0.50 mm)	1.03 (0.34 mm)	1.29 (0.36 mm)	.86 ± .26 (0.37 ± 0.11 mm)

Table 3. Estimation of human errors for UIEdge on original x-ray films from the 1971 Baumrind and Frantz Reliability Study. (Source:[3]). Results were reported in millimeters (shown above in parentheses) and have been converted to estimated pixels. The study did not report results for fiducials.

1991 Baumrind and Baker Reliability Study

The 1991 Baumrind and Baker Reliability Study examined human performance on a number of different media including the digital images that were used in this dissertation [1, 2]. Some confounding factors in this study partially explain the increased estimations of human error compared to the 1971 study. However some of the error can be attributed to the fact that (1) the digital images used in the study had a lower resolution than the original x-ray films, and (2) judges were forced to pick the best pixel not as precise a position as they could have made with the digitizer on the original x-ray films.

Judges located twenty-seven landmarks on forty-four digital images and a magnified version (zoomed by a factor of two). Results for both versions are reported in Table 4, I will concentrate the discussion on the standard deviations of estimates for fiducials and UIEdge on the zoomed images because their values were smaller.

The study reported standard deviations of 0.1mm in X and Y for fiducials, 0.39mm in X and 0.44mm in Y for UIEdge. Converting these values to pixels, the 95 percent confidence intervals for fiducials are 0.60 pixels in X and 0.72 pixels in Y. The 95 percent confidence intervals for UIEdge are 2.36 pixels in X and 3.14 pixels in Y. This provides another criterion by which to judge the network results. The results from this study are useful because the images used were the same format of those used in this thesis.

Standard Deviations (σ_{91}) for estimates of Fiducial and UIEdge as reported by the 1991 Baumrind and Baker Reliability Study on Digital Computer Images

(reported in pixels)

Landmark	SD	SDx	SDy
Fiducial	.70 (0.30 mm)	0.39 (0.13 mm)	0.46 (0.13 mm)
UIEdge	1.33 (0.57 mm)	1.30 (0.43 mm)	2.71 (0.76 mm)

(a)

Standard Deviations (σ_{91}) for estimates of Fiducial and UIEdge as reported by the 1991 Baumrind and Baker Reliability Study on Zoomed Digital Computer Images

(reported in pixels)

Landmark	SD	SDx	SDy
Fiducial	.74 (0.32 mm)	0.30 (0.10 mm)	0.36 (0.10 mm)
UIEdge	.72 (0.31 mm)	1.18 (0.39 mm)	1.57 (0.44 mm)

(b)

Table 4. Estimation of human errors for fiducial and UIEdge on (a) digital images and (b) zoomed x 2 digital images from the 1991 Baumrind and Baker Reliability Study. Results were reported in pixels and millimeters (shown above in parentheses). (Source: [2]).

Summary

In order to evaluate the performance of the network it was necessary to determine the *best estimate* of each landmark for each image in the study. The network was ultimately compared to two different estimates. The first estimate was the mean of two independent estimations of the landmark by the training judge (training judge mean). The second estimate was the mean of three estimates: an additional estimation by the training judge and two other judges' estimation of the landmark (three judges mean). This mean provided a framework with which to evaluate both the training judge and the network's results. The results of the training judge and the three judges' estimations of the landmark on 36 test images are give in Figures 10 through 13.

In order to compare the network reliability with that of human judges I considered the data shown in Figures 10 through 13 along with the analogous findings from two earlier studies of the reliability of performance of human judges. The 1971 Reliability Study, performed on the original analog x-ray films was considered to be the most rigorous criterion because the judges used were well trained and experienced and the landmark location was performed on the original images. When network results were plotted against the three judges' mean, the standard deviation from the 1971 study was used to calculate confidence intervals for UIEdge. However, data from fiducials was not reported in the 1971 study. In this case, standard deviations from the zoomed images of the 1991 Reliability Study were used.

When network results were plotted with respect to the mean of the training judge, the training judge's standard deviations were used to set a confidence interval. A summary of these standard deviations is presented in Table 5. At approximately the 95 percent confidence level, values were accepted if they fell within two standard deviations of the *best estimate*. The standard deviations used to evaluate landmark results will be referred to as: σ_{71} = SD from the 1971 Reliability Study, σ_{91} = SD from the 1991 Reliability Study on zoomed images, and σ_{T} = SD from the training judge.

Summary of Estimations of Human Errors

Fiducial	**SDx**	**SDy**
1971 Reliability Study (x-rays)	not reported	not reported
1991 Reliability Study (digital)	0.39	0.46
1991 Reliability Study (zoom) (σ_{91})	**0.30**	**0.36**
Training Judge (σ_T)	**0.46**	**0.37**
Three Judges	0.59	0.56

UIEdge	**SDx**	**SDy**
1971 Reliability Study (x-rays) (σ_{71})	**1.21**	**1.29**
1991 Reliability Study (digital)	1.30	2.71
1991 Reliability Study (zoom)	1.18	1.57
Training Judge (σ_T)	**0.70**	**0.67**
Three Judges	1.01	0.81

Table 5. Summary of standard deviations reported in terms of pixels. Results in bold were used to set confidence intervals to evaluate network performance.

Chapter 5

Network Experiments -- Results and Discussion

This chapter is divided into three sections covering the experiments performed with respect to three landmarks: the test cross, the fiducial, and the UIEdge. Each section includes a description of the experiments, the results, and a discussion. A general discussion of all results follows in Chapter 6.

Location of the Test Cross

In order to test whether the network was set up and working properly, a test case was designed. The task was to locate multiple test patterns that had been added to an image. The test patterns, a 3 x 3 pixel plus sign, were superimposed on top of random sections of several images. A 3x3 connection scheme was used.

The network was trained on fourteen 3 x 3 pixel images. The training set consisted of seven positive training pairs each of which contained a plus sign, and seven negative pairs that were randomly selected from different sections of a training image. The network trained until the sum of differences in weighting coefficient values between successive iterations (stop error) was less than .0001. This value, chosen after a series of trials in which results did not improve with further training, meant that the network trained for approximately 1,000 epochs. After training, the network was tested on both the training set and a test set of four images, each of which contained four test patterns. Initially, the network successfully located all four of the patterns on the training images, but only ten of the sixteen patterns that were located on the test images. In addition there were three false positives. The false positives occurred on edges, and the algorithm failed to detect the test patterns in regions of intensity similar to that of the test pattern itself.

The training set was then expanded to twenty images by placing three additional test patterns in regions of high intensity, and selecting four additional regions containing edges as negative training patterns. The network was retrained and tested again on the four test images. The result was a positive location of all test patterns on all four images with no false positives. This iterative nature of developing the training set was used for both fiducials and reference markers as well.

The performance of the network was obviously dependent on the make up of the training set. This demonstrated the necessity of including as much variety in the training set as one might expect to see in the test data. Since the test pattern was used to test the functionality of the network and has no relevance to landmarks in actual images, no further testing was conducted.

Location of Reference Markers (Fiducials)

The first landmarks examined were the reference markers (fiducials) located on every image in the database. The network was trained on images from three patients until the stop error was less than 0.0001, generally 10,000 to 20,000 epochs. Forty 11 x 11 pixel training input/output pairs were generated. Twenty five were positive pairs and fifteen were negative training pairs. The test set of images consisted of 36 images from three different patients. The network was evaluated according to the criteria in Figure 15.

Fiducials

When the best estimate was:	The criterion interval in pixels was:
(1) Mean of Three Judges	$2\sigma_{91}$ = (.6 in X, .72 in Y)
(2) Mean of Training Judge	$2\sigma_{T}$ = (.92 in X, .74 in Y)

Figure 15. Criteria for evaluating network's estimates of fiducial: σ_{91} = standard deviations from 1991 Reliability Study and σ_T = Standard deviations of the Training Judge.

Results of Experiments with Fiducial

Examination of Connection Schemes and Hidden Layers

Two features of the network were examined: (1) how layers were connected and (2) the effect of varying the number of hidden layers. The connection schemes and the hidden layers were described in Chapter 2. The network was trained and tested for all combinations of connection scheme and number of hidden layers (zero, one, or two). Network results were plotted with respect to the best estimates of fiducial location by both the training judge and the three judges on the set of 36 test images. On each plot, the origin represents the ***best estimate*** for each image and the axes represent the distance in pixels in the horizontal (X-axis) and vertical (Y-axis) directions.

3x3 Connection Scheme

The results of the network with 3x3 connections and zero hidden layers are plotted in Figure 16. Seven of the thirty six values were within $2\sigma_{91}$ of the best estimate of the three judges. In addition, fifteen values were greater than $2\sigma_{91}$ from the best estimate but were located on some portion of the fiducial. The mean in X was -1.20 pixels and the mean in Y was 0.97 pixels. The standard deviation in X of network results was 3.64 pixels and the standard deviation in Y was 8.23 pixels.

When the same results are plotted with respect to the best estimate of the training judge, eight of the thirty six values were within $2\sigma_T$ of the best estimate (Figure 17). The mean in X was -0.97 pixels and the mean in Y was 0.61 pixels. The standard deviation in X of network results was 3.68 pixels and the standard deviation in Y was 8.24 pixels.

The results of the network with 3x3 connections and one hidden layer are plotted with respect to the three judges in Figure 18. Nine of the thirty six values were within $2\sigma_{91}$ of the best estimate. Thirty five of the thirty six values were positioned on the fiducial. The mean in X was -1.23 pixels and the mean in Y was 0.61 pixels. The standard deviation in X of network results was 4.00 pixels and the standard deviation in Y was 3.74 pixels.

When the results of the network are plotted with respect to the best estimate of the training judge (Figure 19), six of the thirty six values were within $2\sigma_T$ of the best estimate. The mean in X was -1.00 pixels and the mean in Y was 0.25 pixels. The standard deviation in X was 4.01 pixels and the standard deviation in Y was 3.74 pixels.

Network estimates of fiducial location plotted with respect to three judges' mean using 3x3 connection and zero hidden layers

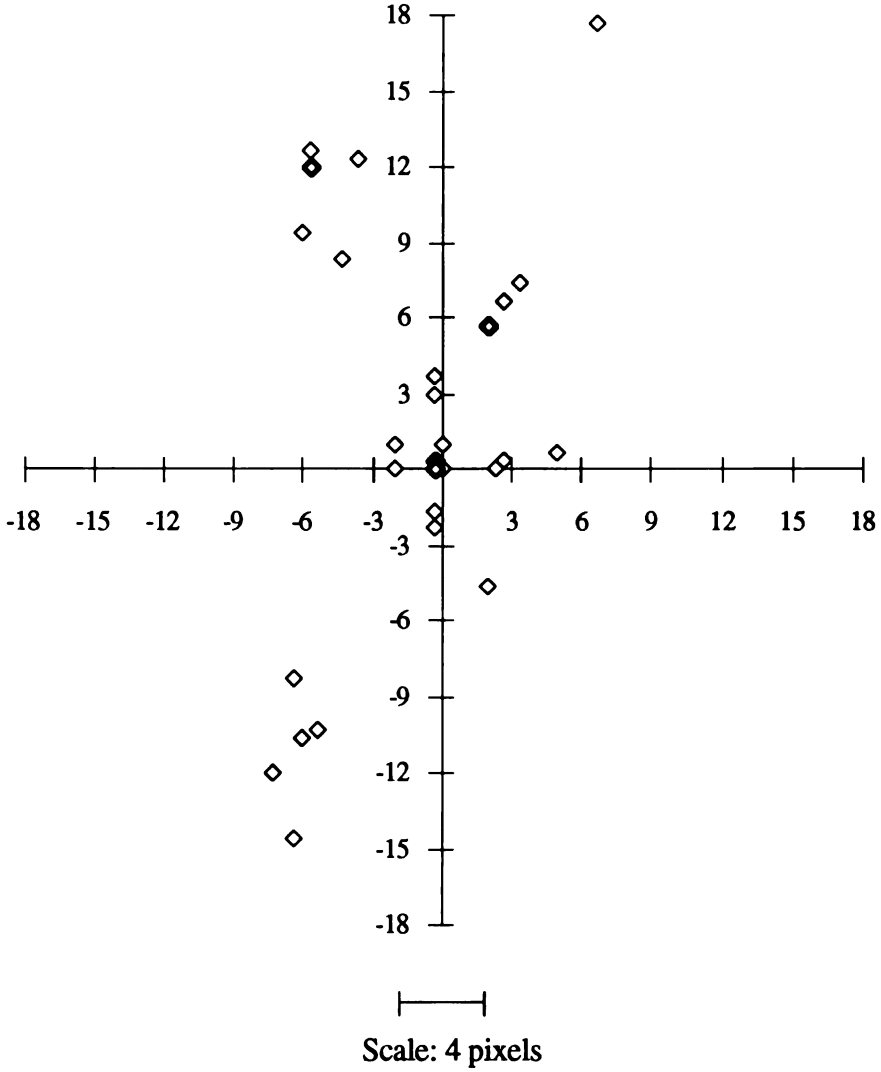


Figure 16. Network location of fiducial plotted with respect to three judges' mean for 36 test images. Network used 3x3 connection and zero hidden layers. Bold data points indicate the presence of two or more data points obscured by overlay. Axes are pixels in X and Y with the *best estimate* at the origin.

Network estimates of fiducial location plotted with respect to training judge's mean using 3x3 connection and zero hidden layers

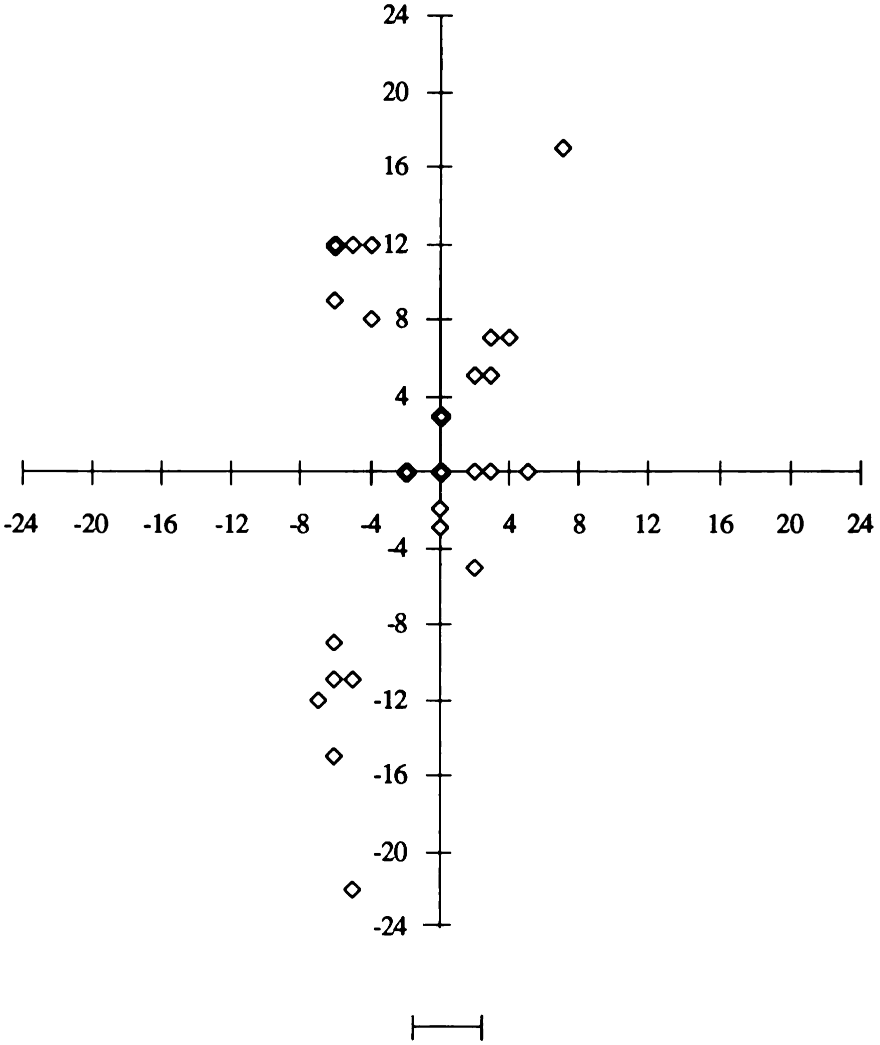


Figure 17. Network location of fiducial plotted with respect to training judge's mean for 36 test images. Network used 3x3 connection and zero hidden layers. Bold data points indicate the presence of two or more data points obscured by overlay. Axes are pixels in X and Y with the *best estimate* at the origin.

Network estimates of fiducial location plotted with respect to three judges' mean using 3x3 connection and one hidden layer

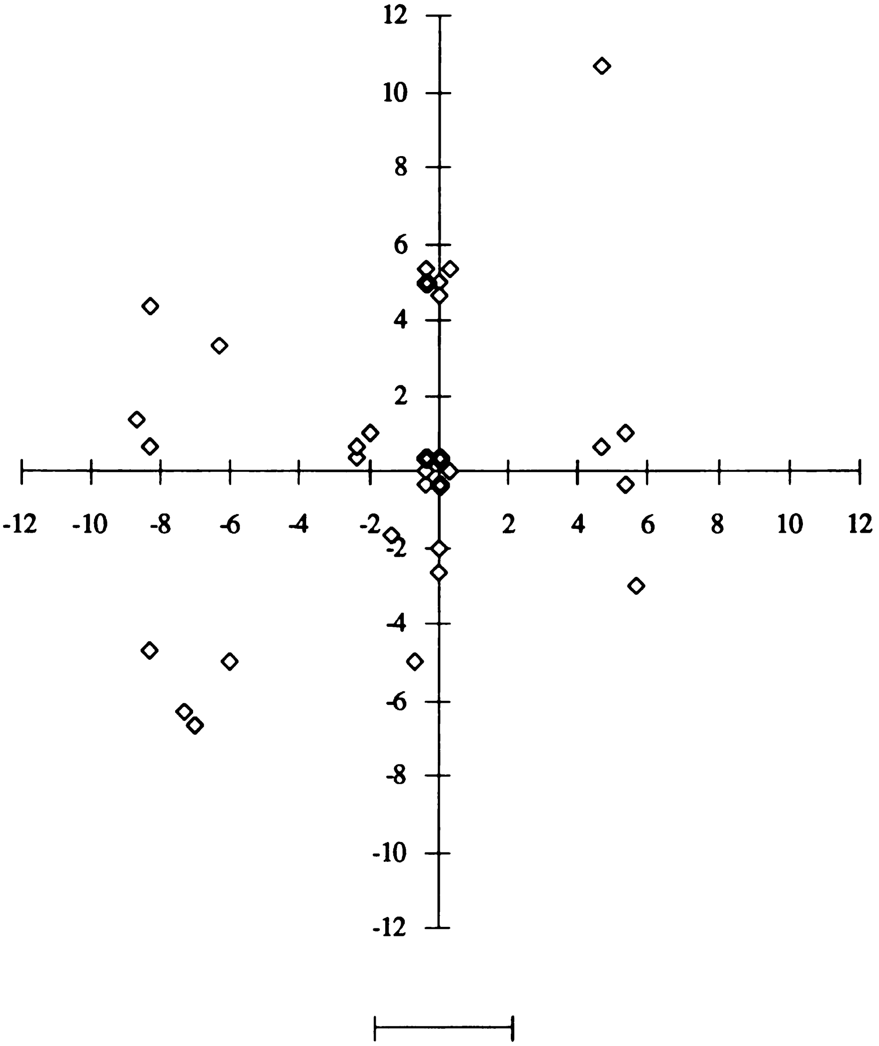


Figure 18. Network location of fiducial plotted with respect to three judges' mean for 36 test images. Network used 3x3 connection and one hidden layer. Bold data points indicate the presence of two or more data points obscured by overlay. Axes are pixels in X and Y with the *best estimate* at the origin.

Network estimates of fiducial location plotted with respect to training judge's mean using 3x3 connection and one hidden layer

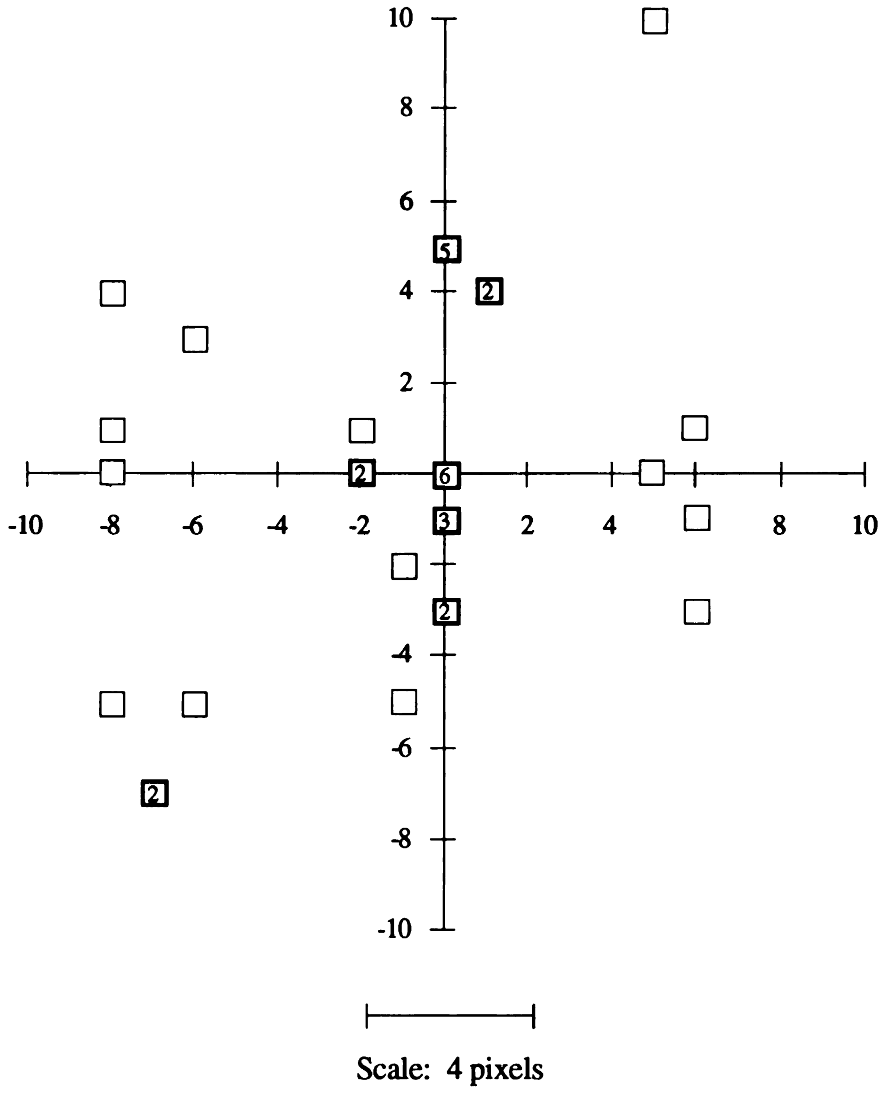


Figure 19. Network location of fiducial plotted with respect to training judge's mean for 36 test images. Network used 3x3 connection and one hidden layer. Numbers inside data points indicate the number of overlaid data points at that position. Axes are pixels in X and Y with the *best estimate* at the origin.

5 x 5 Connection Scheme

The results of the network with 5x5 connections with zero hidden layers have been plotted with respect to the best estimate of three judges in Figure 20. Twenty four of the thirty six values were within $2\sigma_{91}$ of the best estimate. The mean in X was -0.23 pixels and the mean in Y was 0.63 pixels. The standard deviation in X was 2.56 pixels and the standard deviation in Y was 3.20 pixels. When the results of the network were plotted with respect to the best estimate of the training judge (Figure 21), twenty six of the thirty six values were within $2\sigma_T$ of the best estimate. The mean in X was 0.00 pixels and the mean in Y was -1.00 pixels. The standard deviation in X was 2.60 pixels and the standard deviation in Y was 3.15 pixels.

The results of the network with 5x5 connections with one hidden layer have been plotted with respect to the best estimate of three judges in Figure 22. Twenty four of the thirty six values were within $2\sigma_{91}$ of the best estimate. Six additional values were within one pixel of the best estimate. The mean in X was -0.20 pixels and the mean in Y was 0.56 pixels. The standard deviation in X was 0.74 pixels and the standard deviation in Y was 1.10 pixels. When the results of the network were plotted with respect to the best estimate of the training judge (Figure 23) twenty seven of the values were within $2\sigma_T$ of the best estimate. Seven additional values were within one pixel of the mean. The mean in X was 0.03 pixels and the mean in Y was 0.19 pixels. The standard deviation in X was 0.60 pixels and the standard deviation in Y was 1.12 pixels.

Disregarding from both data sets the two values clearly in error, the mean values for results plotted with respect to the three judges were -0.35 in X and 0.31 in Y, the standard deviations were 0.40 pixels in X and 0.43 pixels in Y. The mean values when plotted with respect to the training judge were -0.08 in X and -0.06 in Y and the standard deviations were 0.29 pixels in X and 0.34 pixels in Y.

The results of the network with 5x5 connections with two hidden layers have been plotted with respect to the best estimate of three judges in Figure 24. Sixteen of the thirty six values were within $2\sigma_{91}$ of the best estimate. The mean in X was -2.65 pixels and the mean in Y was 0.44 pixels. The standard deviation in X was 3.36 pixels and the standard deviation in Y was 5.17 pixels. When the results of the network were plotted with respect to the best estimate of the training judge (Figure 25), eight of the thirty six values were within $2\sigma_T$ of the best estimate. Nine additional values were within one pixel of the mean. The mean in X was -2.42 pixels and the mean in Y was 0.08 pixels. The standard deviation in X was 3.46 pixels and the standard deviation in Y was 5.13 pixels.

Network results for 5x5 connections and zero, one, and two hidden layers have different characteristic network outputs. With zero hidden layers, the higher network output values were concentrated around the fiducial (Figure 26). Yet there were high values at what appeared to be random locations in the image as well. With one hidden layer, the two highest network output values were often adjacent pixels (Figure 27). The rest of the pixels in the output had values midway between the range of the maximum likelihood value and zero. When the network was trained with two hidden layers the output values tended to be concentrated in areas of high intensity either on the fiducial or on edges (Figure 28).

Binary Threshold

Binary thresholding by definition involves a loss of information. The network was trained and tested using images that had been filtered with a binary threshold filter with upper and lower threshold limits of 170 and 90. Under these conditions gray scale data was lost but the shape of the fiducial was maintained. Thresholding was done to assess the performance of the network when presented with only structural information. The network results have been plotted with respect to the three judges' mean in Figure 29. Twenty one of the thirty six values were within $2\sigma_{91}$ of the best estimate. The mean in X was -0.40 pixels and the mean in Y was 0.19 pixels. The standard deviation in X was 2.57 pixels and the standard deviation in Y was 2.57 pixels. Loss of gray scale information severely effected results.

Network estimates of fiducial location plotted with respect to three judges' mean using 5x5 connection and zero hidden layers

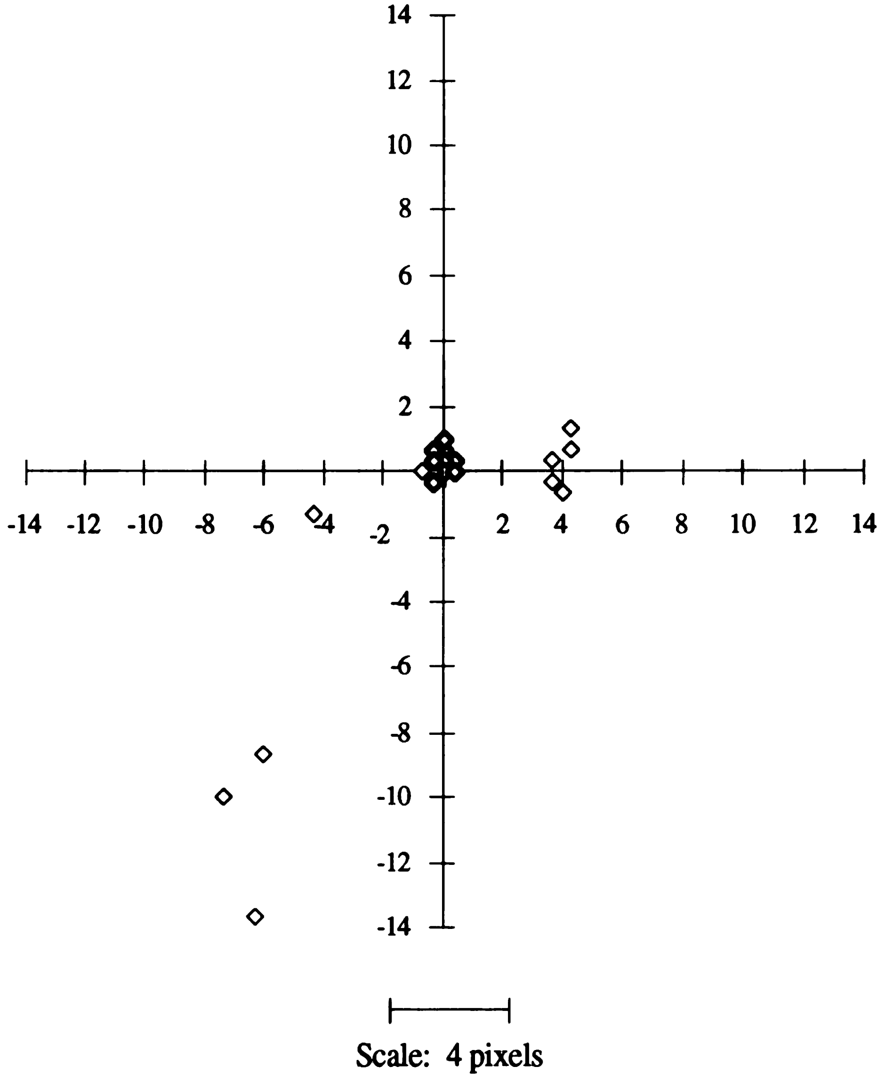


Figure 20. Network location of fiducial plotted with respect to three judges' mean for 36 test images. Network used 5x5 connection and zero hidden layers. Bold data points indicate the presence of two or more overlaid data points. Axes are pixels in X and Y with the *best estimate* at the origin.

Network estimates of fiducial location plotted with respect to training judge's mean using 5x5 connection and zero hidden layers

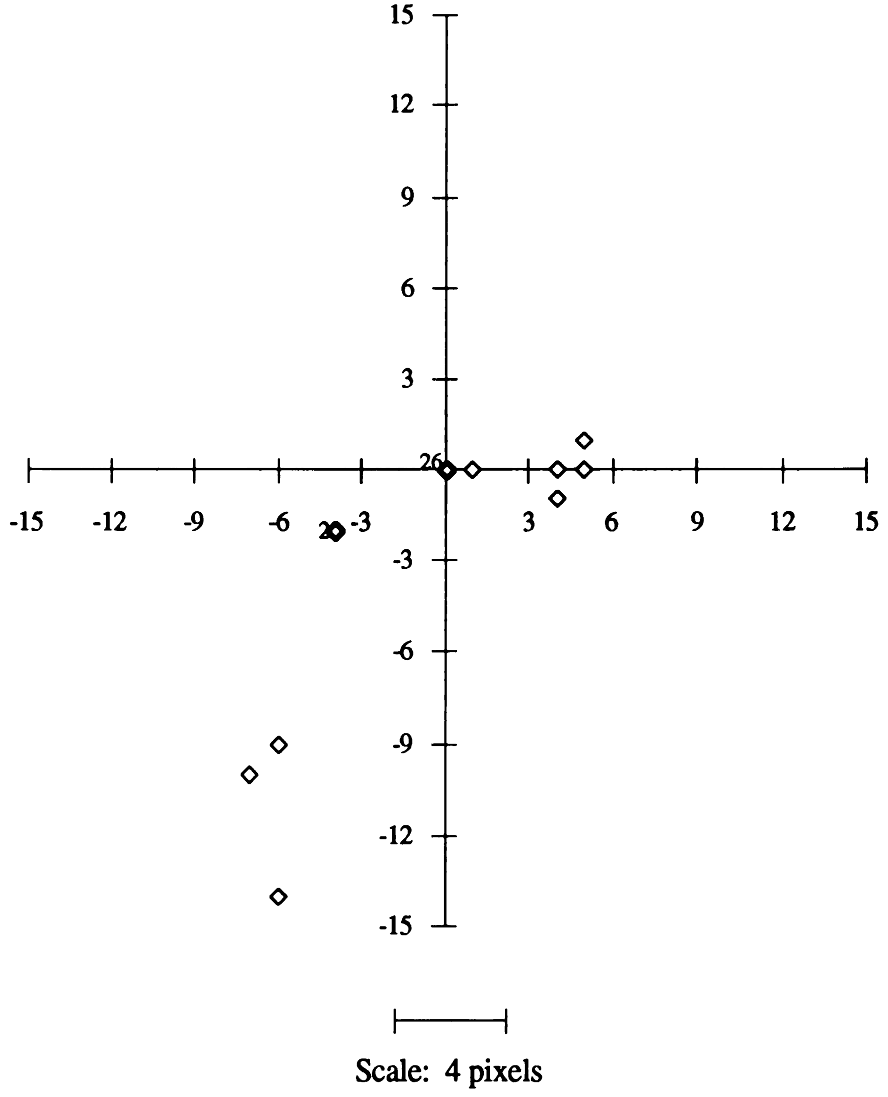


Figure 21. Network location of fiducial plotted with respect to training judge's mean for 36 test images. Network used 5x5 connection and zero hidden layers. Bold data points indicate the presence of two or more overlaid data points. Numbers next to data points indicate the number of data points overlaid at that location. Axes are pixels in X and Y with the *best estimate* at the origin.

Network estimates of fiducial location plotted with respect to three judges' mean using 5x5 connection and one hidden layer

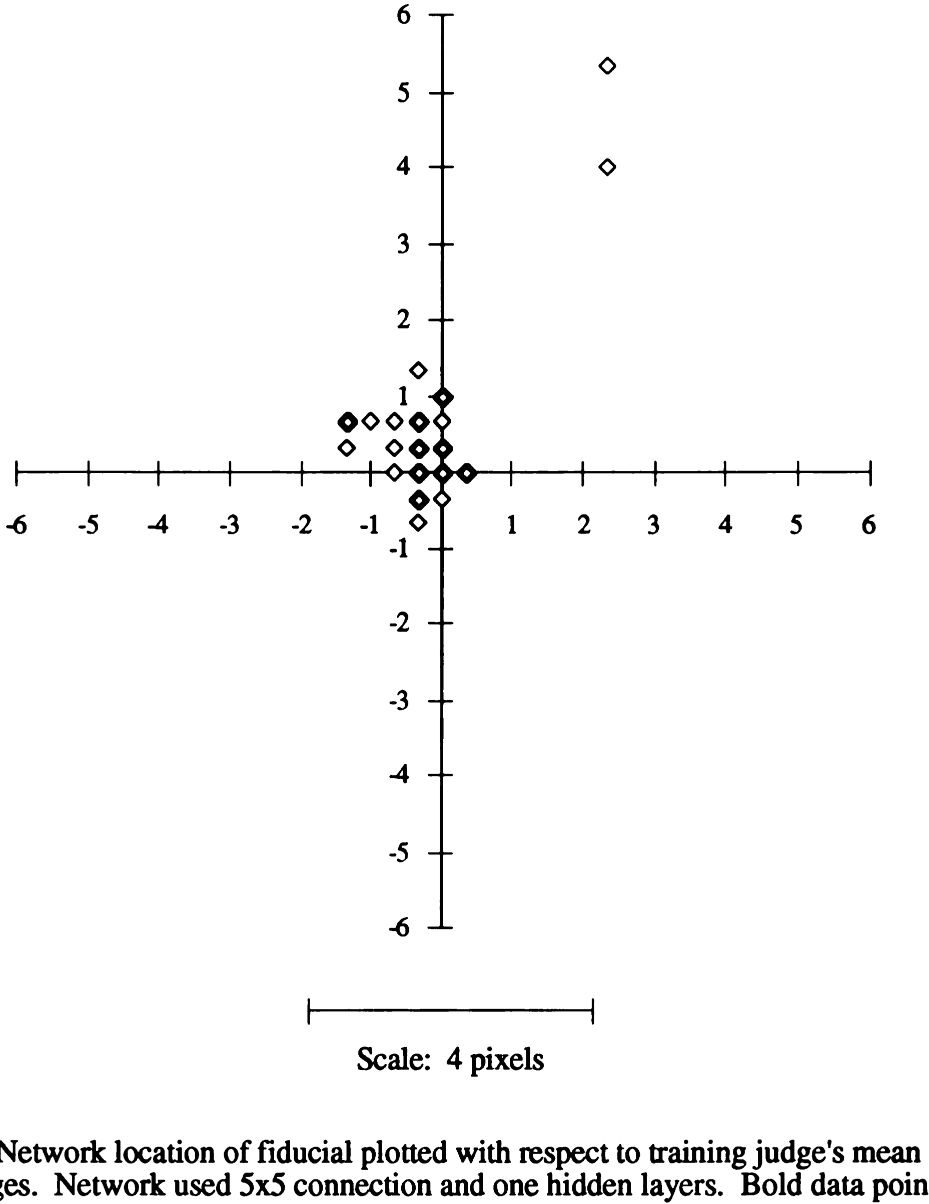
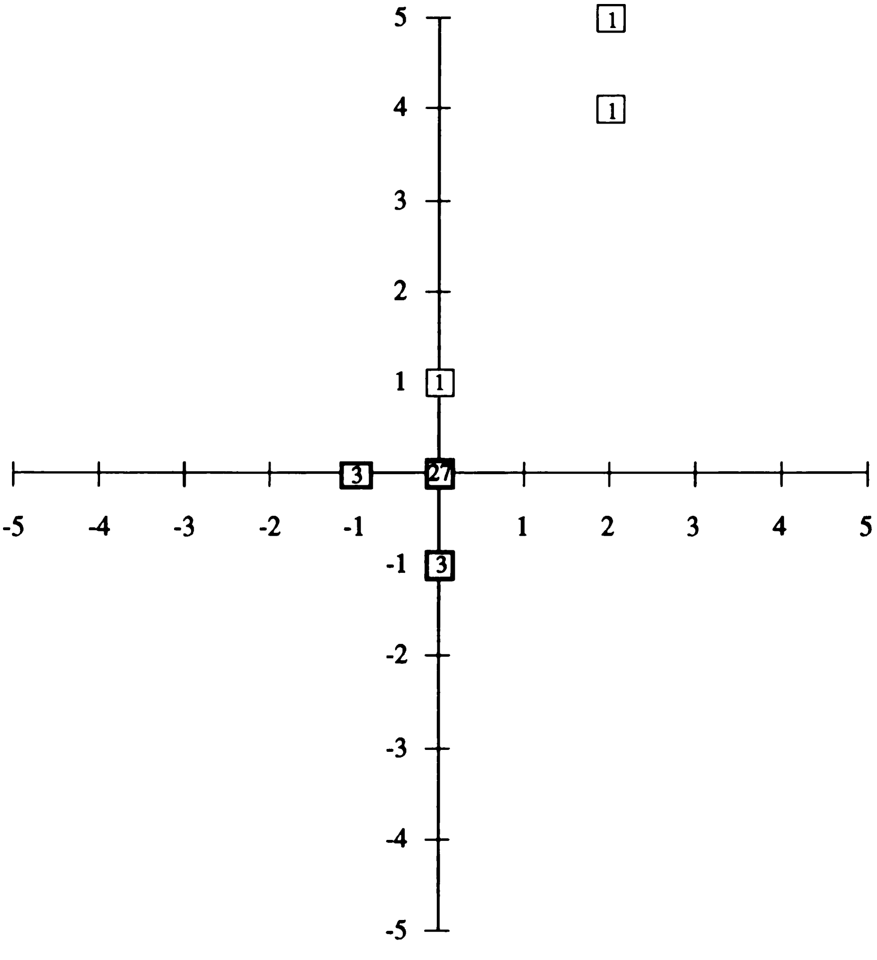


Figure 22. Network location of fiducial plotted with respect to training judge's mean for 36 test images. Network used 5x5 connection and one hidden layers. Bold data points indicate the presence of two or more overlaid data points. Axes are pixels in X and Y with the *best estimate* at the origin.

Network estimates of fiducial location plotted with respect to training judge's mean using 5x5 connection and one hidden layer



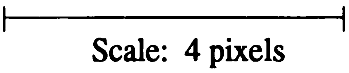


Figure 23. Network location of fiducial plotted with respect to training judge's mean for 36 test images. Network used 5x5 connection and one hidden layer. Numbers in data points indicate the number of data points overlaid at that location. Axes are pixels in X and Y with the *best estimate* at the origin.

Network estimates of fiducial location plotted with respect to three judges' mean using 5x5 connection and two hidden layers

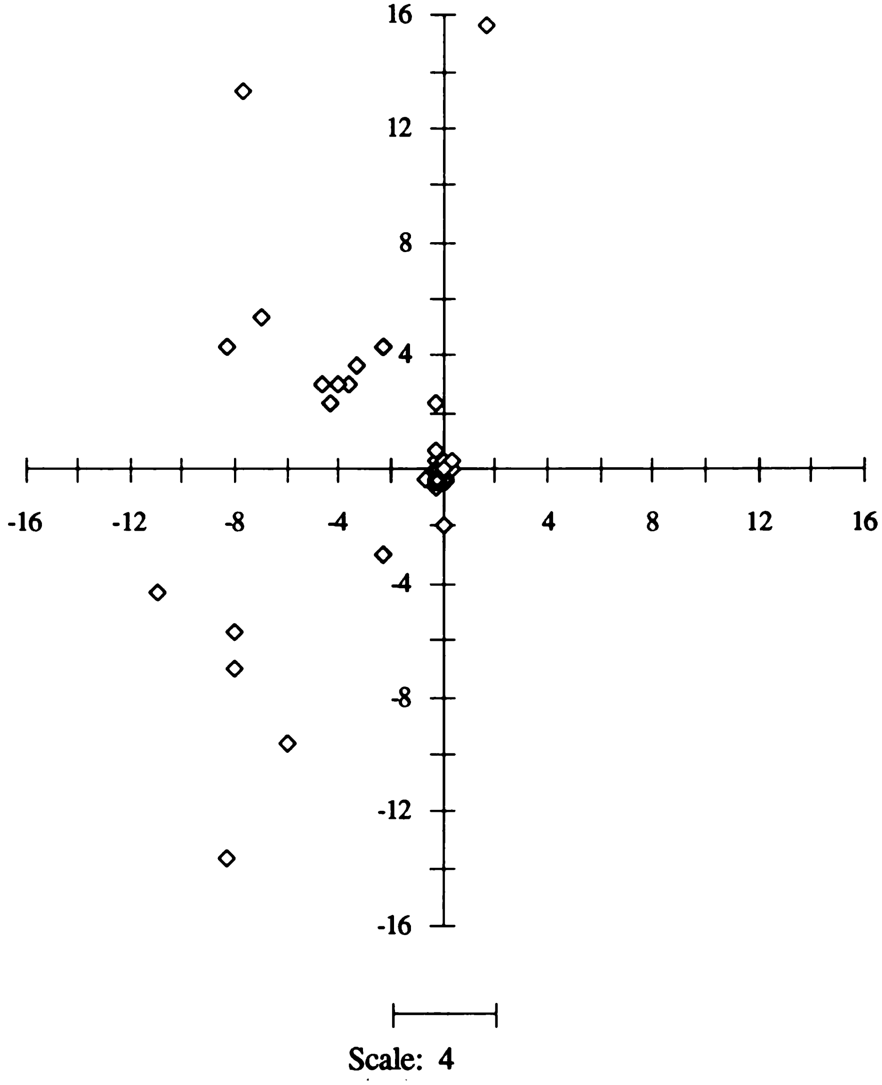


Figure 24. Network location of fiducial plotted with respect to three judges' mean for 36 test images. Network used 5x5 connection and two hidden layers. Bold data points indicate the presence of two or more overlaid data points. Axes are pixels in X and Y with the ***best estimate*** at the origin.

Network estimates of fiducial location plotted with respect to training judge's mean using 5x5 connection and two hidden layers

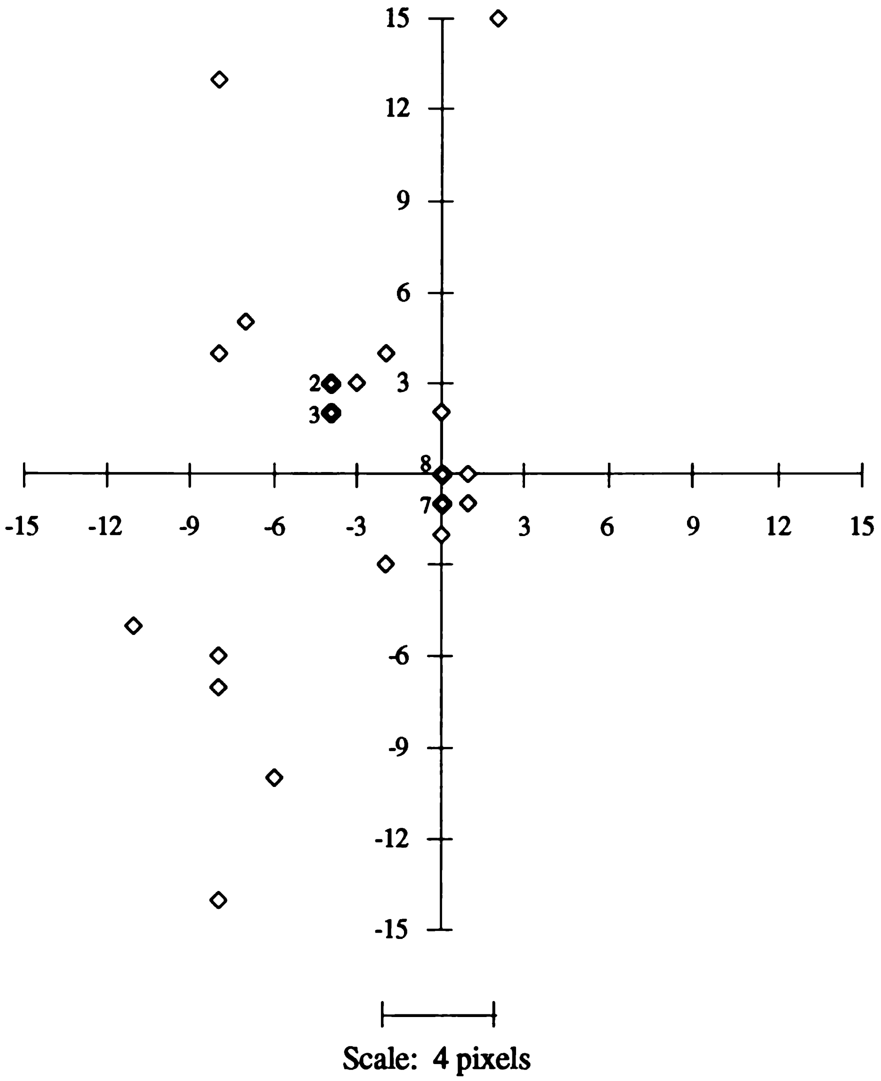


Figure 25. Network location of fiducial plotted with respect to training judge's mean for 36 test images. Network used 5x5 connection and one hidden layer. Bold data points indicate the presence of two or more overlaid data points. Numbers next to data points indicate the number of data points overlaid at that location. Axes are pixels in X and Y with the *best estimate* at the origin.

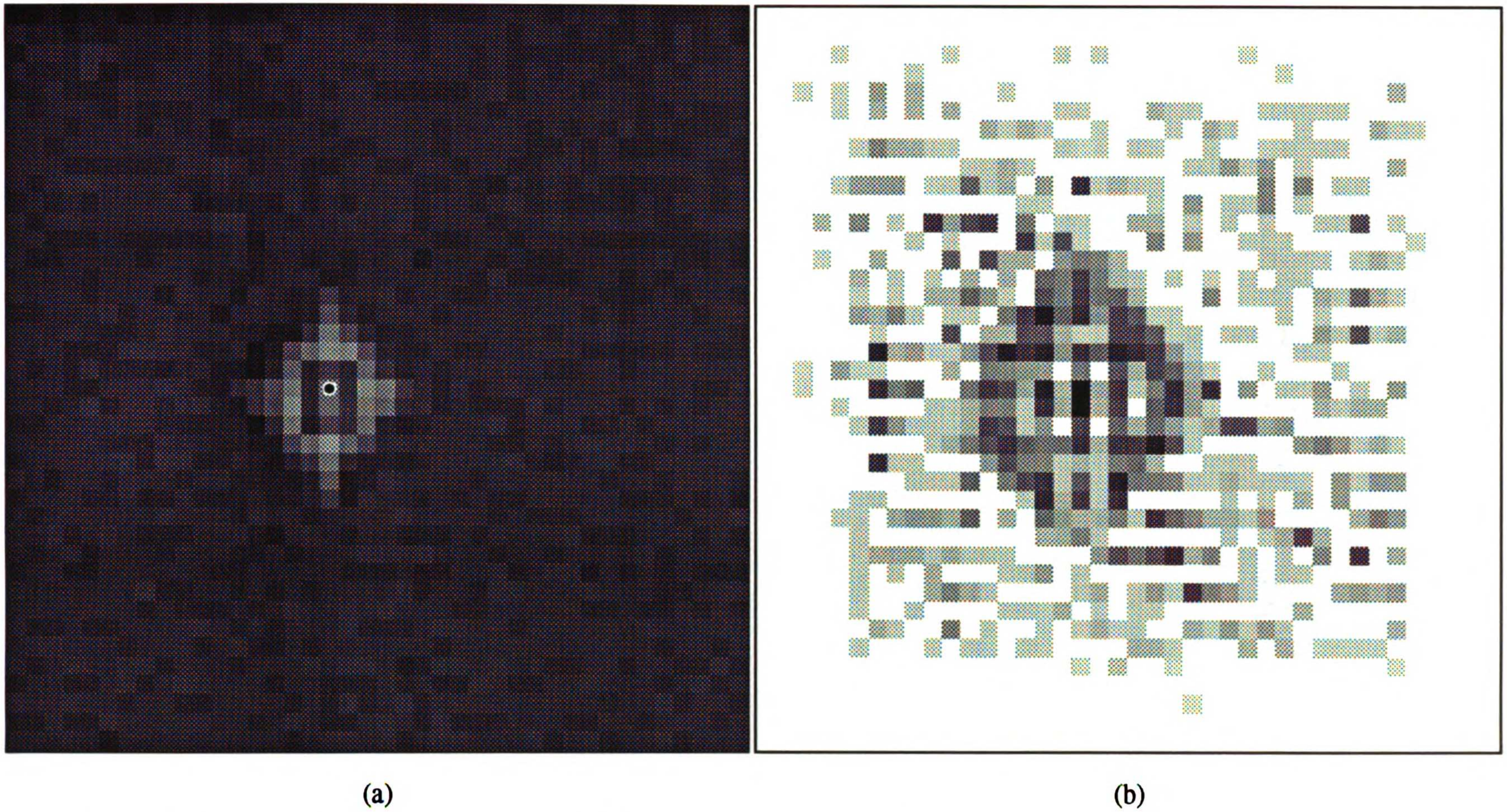
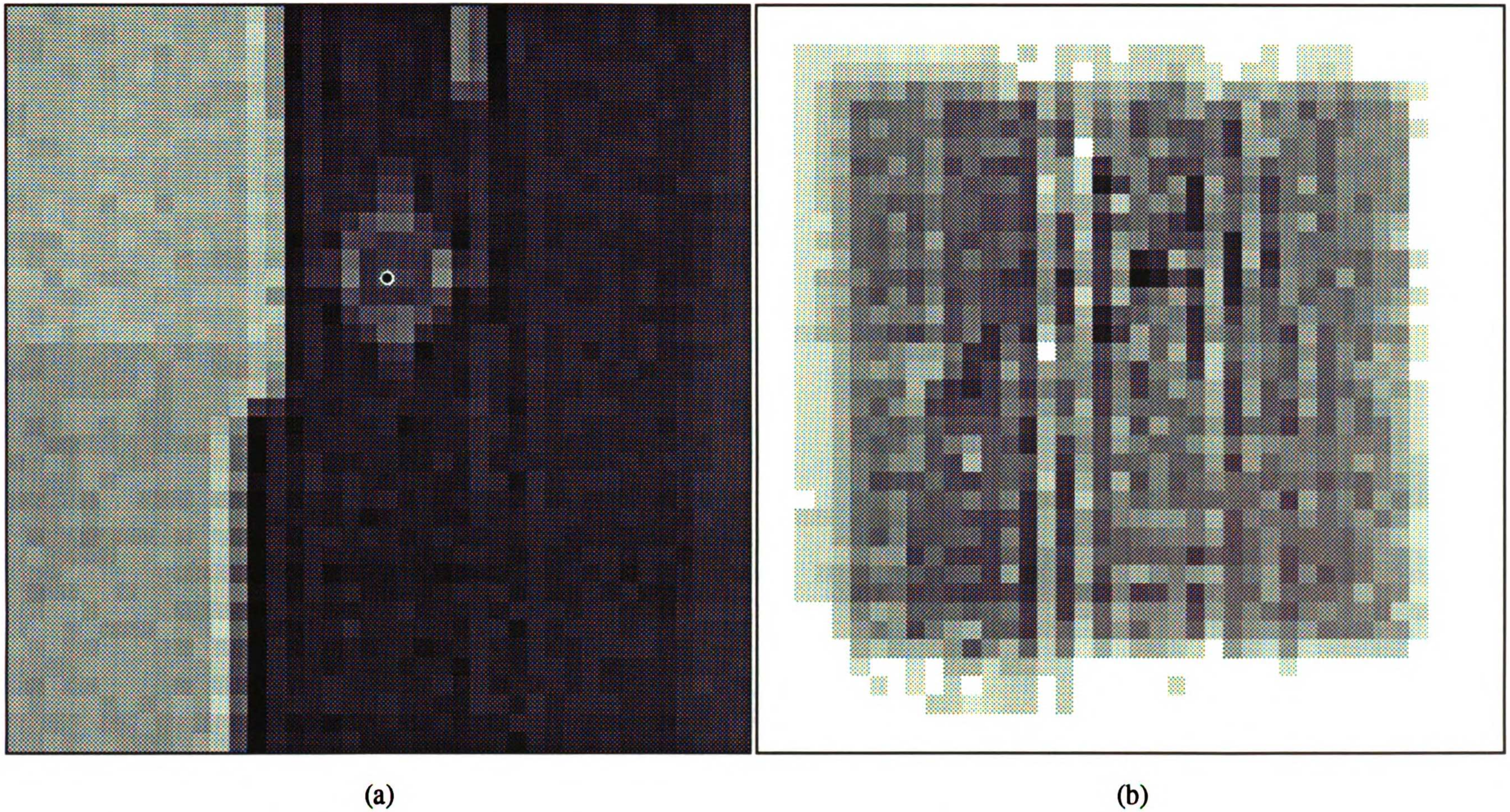
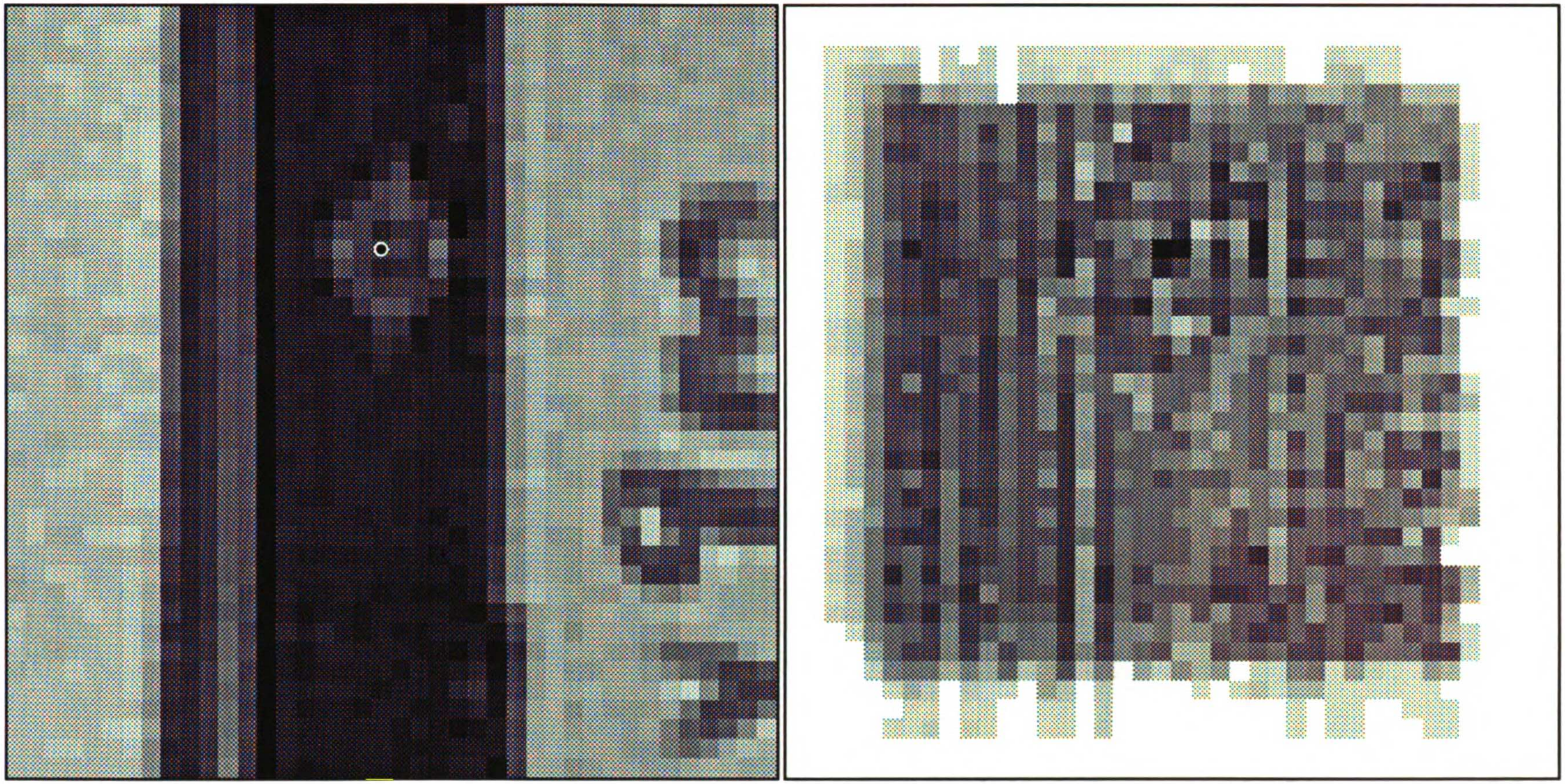


Figure 26. Results of network trained with 5x5 connections and no hidden layers.
(a) Test Image 00601a. Network location of fiducial is shown by the black dot. (b) Network output.



(a) (b)

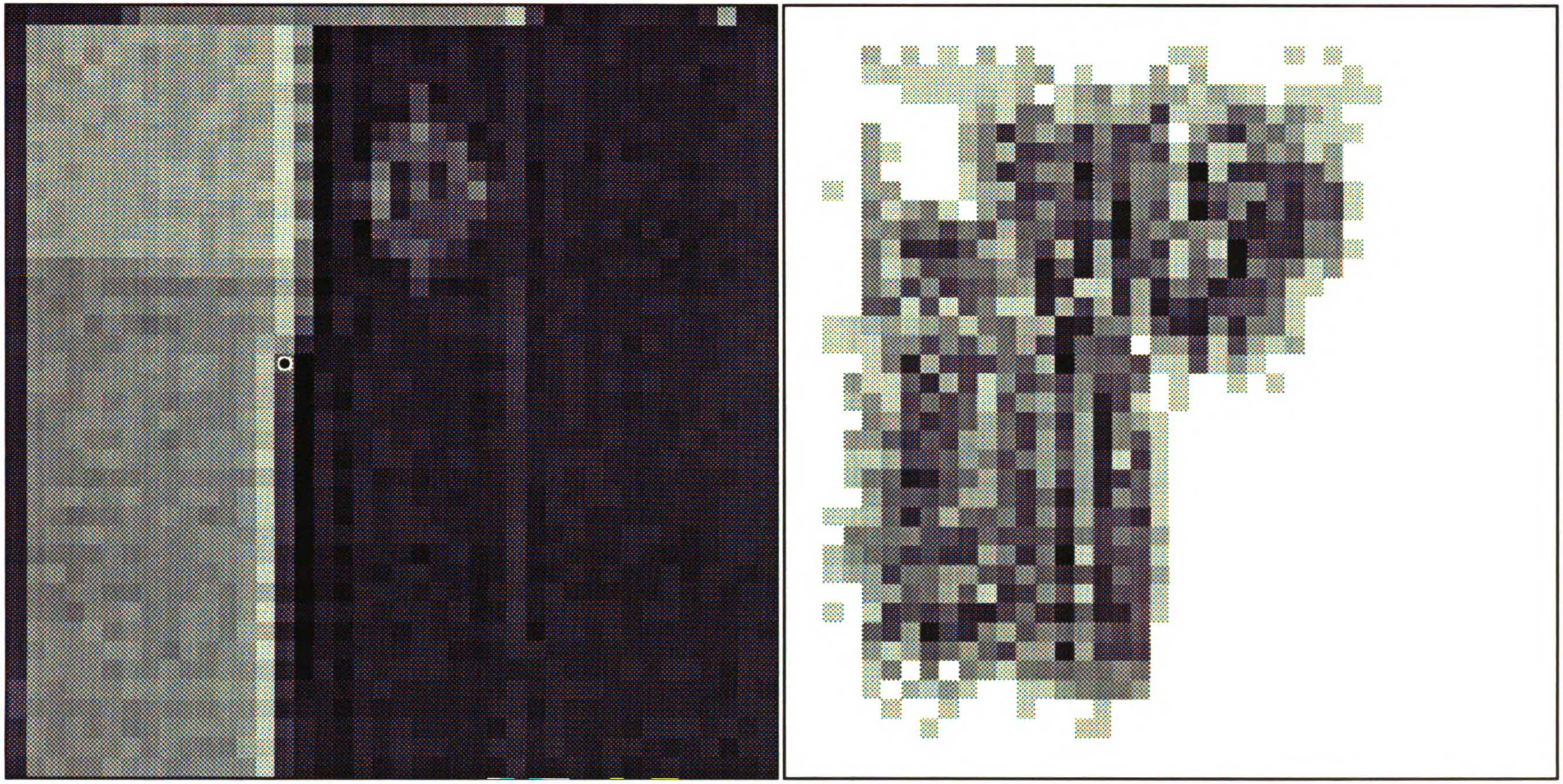
Figure 27. Results of network trained with 5x5 connections and one hidden layers.
(a) Test image 00803b. The network maximum likelihood of fiducial is shown by the black dot. (b) Output of the network.



(c)

(d)

(c) Test image 00803c. The network maximum likelihood of fiducial is shown by the black dot. (d) Output of the network.



(a)

(b)

Figure 28. Results of network trained with 5x5 connections and two hidden layers.
(a) Test Image 00501b. Network location of fiducial is shown by the black dot. (b) Output of the network.

Network estimates of fiducial location plotted with respect to three judges' mean using 5x5 connection, one hidden layer, and binary threshold images

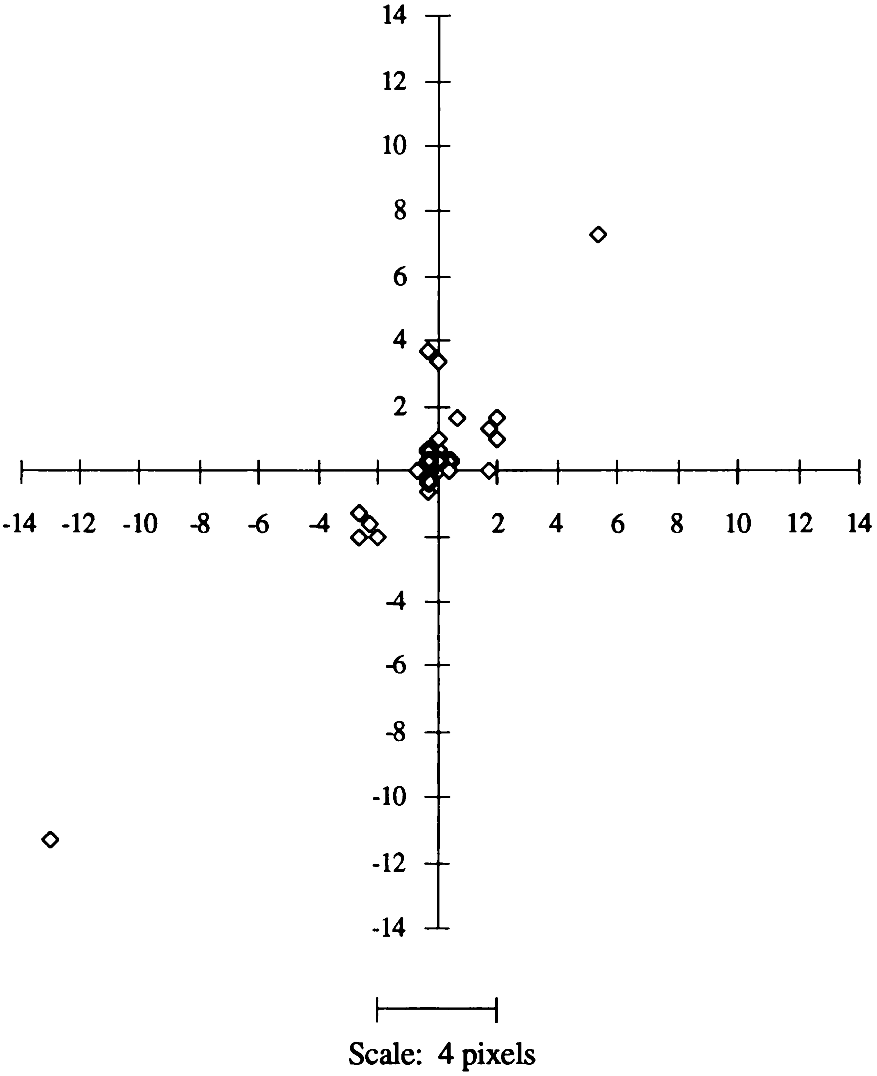


Figure 29. Network location of fiducial plotted with respect to training judge's mean for 36 test images. Binary thresholding was applied to both training and test images. Network used 5x5 connection and one hidden layers. Bold data points indicate the presence of two or more overlaid data points. Axes are pixels in X and Y with the *best estimate* at the origin.

Discussion of Experiments with Fiducial

The primary goal of the experiments on fiducial was to determine if a neural network could be trained to successfully locate reference markers on digital images. In the process of achieving this goal the effect of two different connection schemes and three conditions of hidden layers were examined. The best results were achieved when the network used a 5x5 connection scheme and one hidden layer; however there were several other interesting findings as well.

3x3 Connection Scheme

The 3x3 connection scheme with one hidden layer contains a total of 2420 connections (10 connections/node x 121 nodes x 2 layers) between nodes while the 5x5 connection scheme with one hidden layer has a total of 6292 connections (26 connections/node x 121 nodes x 2 layers). When training forty images for thousands of epochs this difference can lead to a consequential difference in training and segmentation time. The 3x3 connection scheme took fewer epochs both to train and to segment images.

While the 3x3 and the 5x5 connection schemes evaluated the same size image, the 3x3 connection did not provide enough information from the input image to each individual pixel in the network. Unlike the test cross, the fiducial spans a larger area which could not be fully characterized given the number of parameters and training data available in the 3x3 connection scheme.

The network results with zero hidden layers located many points positioned both above and below the fiducial. The standard deviations in Y were approximately twice those in X. The failures were always pixels along an edge, and the most distinct edges tended to be vertical (Figures 27-28). This may have been one of the local minima which with further training could be eliminated.

With a 3x3 connection scheme and one hidden layer, the network's location of the landmark was evenly spread around but usually physically on some part of the fiducial. The network was able to locate the general area of the fiducial in all but one case. However, only nine of thirty six values met the criterion $2\sigma_{91}$. These results suggest that the 3x3 connection scheme might be useful in segmenting larger images to locate the general area of the fiducial. Then a smaller image could be used as input to the 5x5 connection network. This would reduce the search area and the time required to segment images with the 5x5 connection network.

With two hidden layers and a 3x3 connection, the network was not able to converge. This is explained by examining the information that the network "sees" and how it is processed. With one hidden layer, each pixel in the output image is connected to nine pixels in the hidden layer and indirectly connected to twenty five pixels in the input image. With two hidden layers, each pixel in the output image is indirectly connected to forty nine pixels in the input image. The number of weighting coefficients increases from 2,420 to 3,630 (10 connections/node x 121 nodes x 3 layers). In addition, in the forward propagation phase with two hidden layers, a given node's value is modified through two squashing functions and three weighting coefficients before effecting an output node. Given the nature of the training set in the 3x3 connection scheme with two hidden layers, there are simply too many parameters for the network to converge to any solution.

The facts that the network was not able to converge with two hidden layers in the 3x3 connection scheme and that results worsened in the 5x5 connection scheme indicate that the number of parameters available to a system is not as important as the nature of the parameters. By adding an additional hidden layer, the chances of converging to a local minimum are also increased. The number of parameters must be large enough to characterize a problem, but not so large that the weighting coefficients can not be estimated from the training data.

When the network failed to locate the fiducial, it generally failed catastrophically. In such cases, the network tended to locate a point on another structure, i.e. on an adjacent name plate or the step wedge. When human judges fail catastrophically, they generally fail because they too have located an erroneous structure. This usually happens when judges are inexperienced or are locating landmarks on unfamiliar or unusual anatomy. Thus catastrophic failure by the network can be seen as inexperience and requiring further training. Catastrophic failure partially explains the large standard deviations we see for the 3x3 connection scheme.

5x5 Connection Scheme

The results of the network with 5x5 connections and zero hidden layers were better than any of the 3x3 connection trials. In fact, twenty six of thirty six values were located within the accuracy of the training judge. Increasing the number of nodes is similar to widening the field of view as opposed to increasing the depth by adding more layers. The fiducial is considerably larger than the test cross and requires more connections in order for the network to determine the location.

The best results were achieved when the network used a 5x5 connection scheme and one hidden layer. Under these conditions the network located twenty seven of the fiducials within two standard deviations of the training judge, and thirty four of the fiducials within one pixel of the best estimate. The second and occasionally the third highest network output was frequently an adjacent pixel. From this I can infer that sub-pixel processing techniques could be used to further improve results.

Results for 5x5 connection scheme with two hidden layers were similar to those with the 3x3 connection scheme. The network located erroneous structures requiring further training.

The Three Judges vs. the Training Judge

One of the more interesting results of this study emerged when comparing the results of the network (with 5x5 connections and one hidden layer) plotted with respect to the three judges' mean vs. the training judge's mean. Comparing the network's means and the distribution of points in Figures 22 and 23, I conclude that the network trained to the characteristic distribution of the training judge. In fact the network picked the same pixel as the training judge in 27 of the 36 images. This demonstrates that the network can be trained to the particular biases of an individual. From this I infer that a network trained to a better standard would have better results. In particular, training the network to the mean values of several experts might reduce the variability of the networks results due to variability found in the training judge's data.

Location of an Anatomic Landmark -- UIEdge

The UIEdge is one of the easier anatomic landmarks for humans to locate on lateral cephalometric x-ray images. This is evident in small standard deviations for the landmark (Chapter 3) and stems at least in part from the fact that this anatomical structure is more precisely defined than many others. Since 11 x 11 pixel images were used with good results to train the network to locate fiducials, I initially used this size image to train the network to locate the UIEdge as well.

Two preprocessing techniques were examined. First, the network was trained and tested on normalized images to compensate for the varying intensities on the incisor tip. Second, a Sobel filter was convolved with images to enhance the UIEdge's contours to assess the contribution of landmark shape.

One of the problems with the 11 x 11 pixel size training image was that the network occasionally located an unerupted upper lateral incisor rather than the UIEdge. By increasing the size of the training image, and thereby increasing the size of the subimage used during segmentation, larger areas of the image can be encoded into the networks knowledge. In an attempt to resolve this problem, the network was trained with 21 x 21 pixel images. Given a 5x5 connection scheme and one hidden layer, there are 22,932 weighting coefficients (26 connections/nodes x (21 x 21) nodes x 2 layers) when training on 21 x 21 pixel images compared to 6,292 weighting coefficients (26 connections/nodes x (11 x 11) nodes x 2 layers) with an 11 x 11 pixel image. I was skeptical about the performance of the network with this many parameters, but this size was necessary in order to incorporate the background information necessary to differentiate between different incisors.

Another way to provide the network with more information without quadrupling the number of parameters is to average the values of groups of pixels around the perimeter of the training image. This procedure is described in detail in Chapter 3. This way information from a 21 x 21 pixel image can be incorporated into a 13 x 13 training image with information loss only at the perimeter of the image. With a 5x5 connection scheme and one hidden layer this method reduces the number of weighting coefficients to 8,788 (26 connections/nodes x (13 x 13) nodes x 2 layers).

The test set of images consisted of thirty six images from seven patients. Recall that the training images and the test images came from different patients. The network results were evaluated according to the criteria in Figure 30.

UIEdge

When the best estimate was:	The criterion interval in pixels was:
(1) Mean of Three Judges	$2\sigma_{71}$ = (2.42 in X, 2.58 in Y)
(2) Mean of Training Judge	$2\sigma_T$ = (1.4 in X, 1.34 in Y)

Figure 30. Criteria for evaluating network's estimates of UIEdge: σ_{71} = standard deviations from 1971 Reliability Study and σ_T = Standard deviations of the Training Judge.

Results for Experiments with the UIEdge

The network was trained until the stop error was less than 0.0001 or the network had trained for 10,000 epochs. Early trials showed that training to 20,000 or even 30,000 epochs did not improve results. Network results were compared to the results of the training judge's mean and the three judges' mean for the landmark on set of thirty six test images. A mean difference and standard deviation in X and Y was calculated. The results for the 11 x 11 pixel images are described first, including those trained on images preprocessed with normalization and a Sobel filter. Next, the results from the 21 x 21 pixel training images and the 13 x 13 pixel training images are presented. In all cases, a 5x5 connection scheme with one hidden layer was used.

11 x 11 pixel training images

In the first set of experiments with The UIEdge, the network was trained on images from three patients. Forty one 11 x 11 pixel training input pairs were generated. Twenty two were positive pairs and nineteen were negative training pairs. The results of the network were plotted with respect to the three judges' mean in Figure 31. Twenty nine of the thirty six values were within $2\sigma_{71}$ of the best estimate. The mean in X was -0.77 pixels and the mean in Y was 1.25 pixels. The standard deviation in X was 3.14 pixels and the standard deviation in Y was 2.83 pixels. The results of the network were plotted with respect to the training judge's mean in Figure 32. Twenty seven of the thirty six values were within $2\sigma_T$ of the best estimate. The mean in X was -1.28 pixels and the mean in Y was 1.28 pixels. The standard deviation in X was 3.33 pixels and the standard deviation in Y was 2.87 pixels. Examples of some test images and corresponding network output are presented in Figures 33, 34, 35.

Normalized Images

The results of the network with normalized training and test images, were plotted with respect to the three judges' mean in Figure 36. Twenty four of the thirty six values were within $2\sigma_{71}$ of the best estimate. The mean in X was -1.60 pixels and the mean in Y was 1.86 pixels. The standard deviation in X was 4.39 pixels and the standard deviation in Y was 3.55 pixels. When the results of the network were plotted with respect to the training judge's mean (Figure 37), twenty of the thirty six values were within $2\sigma_T$ of the best estimate. The mean in X was -2.11 pixels and the mean in Y was 1.58 pixels. The standard deviation in X was 4.19 pixels and the standard deviation in Y was 3.70 pixels.

Network results greater than five pixels from the best estimate were the result of the network locating an alternative structure (as described in the 11 x 11 pixel section above). The mean estimating error of all values within five pixels of the best estimate was 0.68 in X and 0.87 in Y with respect to the three judges' mean, and 0.04 in X and 0.56 in Y with respect to the training judge's mean.

Sobel Images

The network was tested with test images that had been preprocessed with a Sobel filter. Though the 3 x 3 pixel Sobel filter has a limited applicable range, it did enhance the UIEdge contours. I expected to see results similar to those achieved with the fiducial images that had been convolved with a binary threshold. However, results were disappointing. The results are plotted with respect to the three judges' mean in Figure 38. Only one of the thirty six values was within $2\sigma_{71}$ of the best estimate. The mean in X was -3.27 pixels and the mean in Y was 1.86 pixels. The standard deviation in X was 7.86 pixels and the standard deviation in Y was 6.59 pixels.

21 x 21 pixel training images

Next, the network was trained on images from the same three patients as used in the experiment with the 11 x 11 pixel training images except that forty seven 21 x 21 pixel training input pairs were generated. Of these thirty were positive training pairs and seventeen were negative training pairs. This experiment with 21 x 21 pixel images was done in response to the results from the network trained with 11 x 11 pixel images. I wanted to determine if training on larger images would reduce the network's tendency to occasionally locate an unerupted upper lateral incisor rather than the UIEdge. The test set of images consisted of the same set as those tested with the network trained with the 11 x 11 pixel images.

The results of the network were plotted with respect to the three judges' mean in Figure 39. Eleven of the thirty six values were within $2\sigma_{71}$ of the best estimate. The mean in X was -0.94 pixels and the mean in Y was 3.19 pixels. The standard deviation in X was 3.54 pixels and the standard deviation in Y was 4.83 pixels. When the results of the network were plotted with respect to the training judge's mean (Figure 40), ten of the thirty six values were within $2\sigma_T$ of the best estimate. The mean in X was -1.44 pixels and the mean in Y was 3.22 pixels. The standard deviation in X was 3.41 pixels and the standard deviation in Y was 4.92 pixels.

13 x 13 pixel training images with averaged borders

Finally, the network was again trained on images from the same three patients except that forty eight 13 x 13 pixel training input pairs were generated. Of these twenty seven were positive pairs and nineteen were negative training pairs. The training image contained a 9 x 9 central image at full resolution and a two pixel border where each pixel was the average of a 3 x 3 pixel region of the original images (this procedure is detailed in Chapter 3). This was done in response to the results from the network trained with 21 x 21 pixel images. Data from a 21 x 21 pixel region were encoded into a 13 x 13 pixel training image reducing the number of parameters in the network. The test set of images consisted of the same set as tested with the network trained with 21 x 21 pixel images, except that the borders of the test images were averaged in the same way as the training images.

The network was trained for 10,000 epochs. While the network did not locate any upper lateral incisors, it was never able to perform consistently within criteria levels on the training data. The network often picked points on the lower incisor and points above the tip of the UIEdge.

The results of the network have been plotted with respect to the three judges' mean in Figure 41. Twelve of the thirty six values were within $2\sigma_{71}$ of the best estimate. The mean in X was -1.05 pixels and the mean in Y was 2.69 pixels. The standard deviation in X was 3.32 pixels and the standard deviation in Y was 3.05 pixels. When the results of the network were plotted with respect to the training judge's mean (Figure 42), seven of the thirty six values were within $2\sigma_T$ of the best estimate. The mean in X was -1.56 pixels and the mean in Y was 2.72 pixels. The standard deviation in X was 3.12 pixels and the standard deviation in Y was 3.07 pixels.

Network estimates of the UIEdge plotted with respect to three judges' mean using 11 x 11 pixel training images

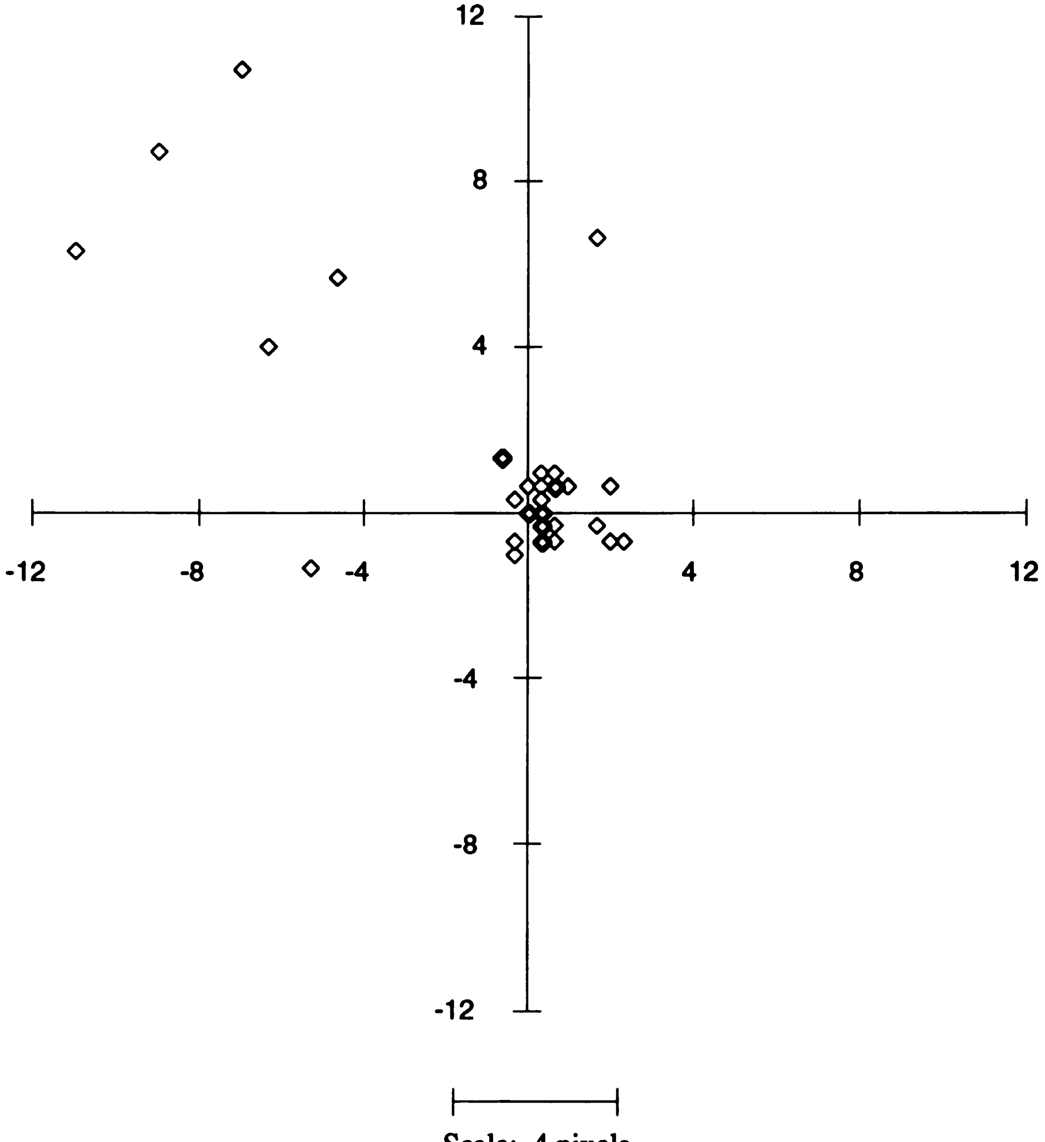


Figure 31. Network location of the UIEdge plotted with respect to three judges' mean for 36 test images. Network used 11 x 11 pixel training images. Bold data points indicate the presence of two or more overlaid data points. Axes are pixels in X and Y with the *best estimate* at the origin.

Network estimates of the UIEdge plotted with respect to training judge's mean using 11 x 11 pixel training images

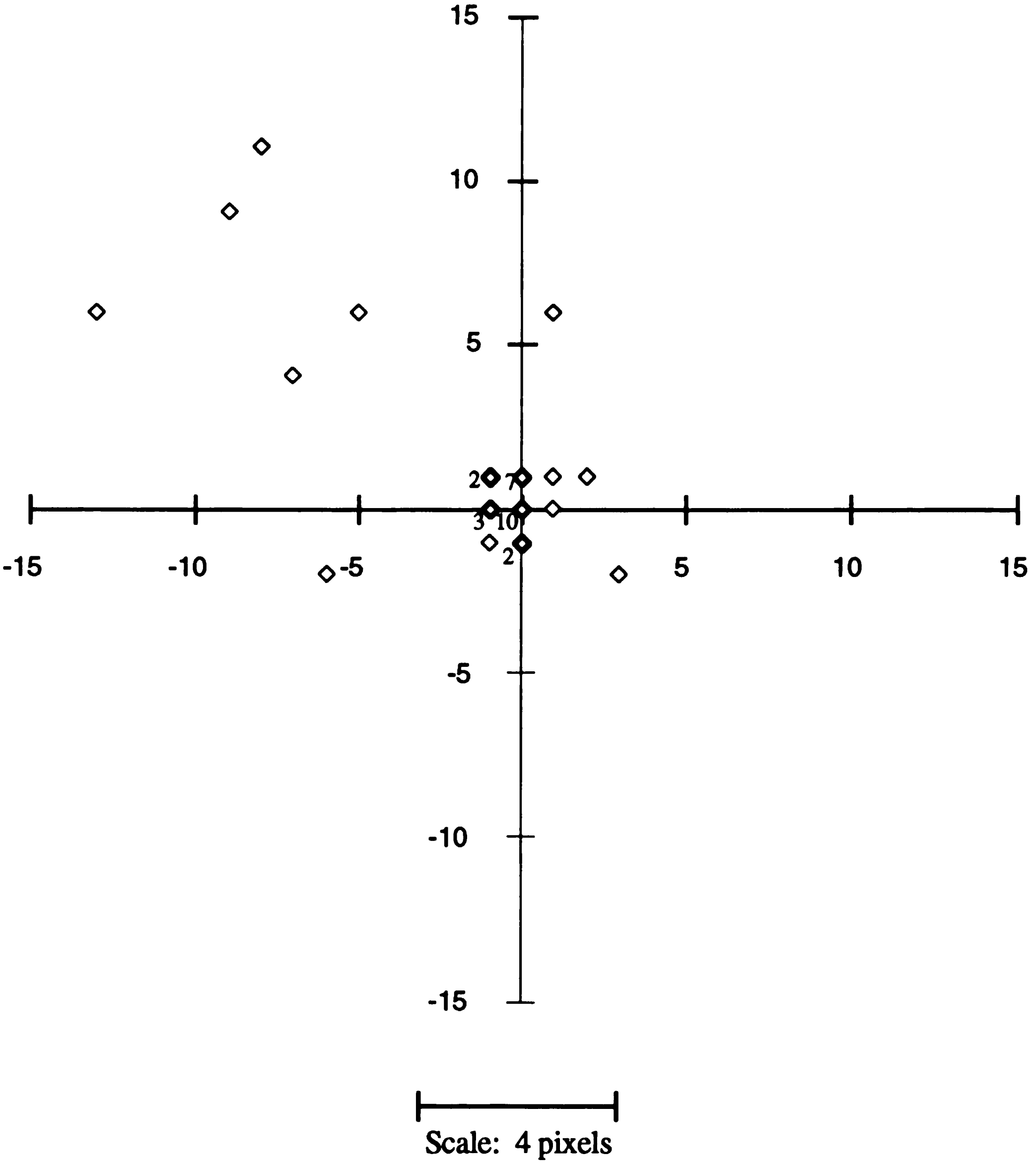


Figure 32. Network location of the UIEdge plotted with respect to training judge's mean for 36 test images. Network used 11 x 11 pixel training images. Bold data points indicate the presence of two or more overlaid data points. Axes are pixels in X and Y with the *best estimate* at the origin.

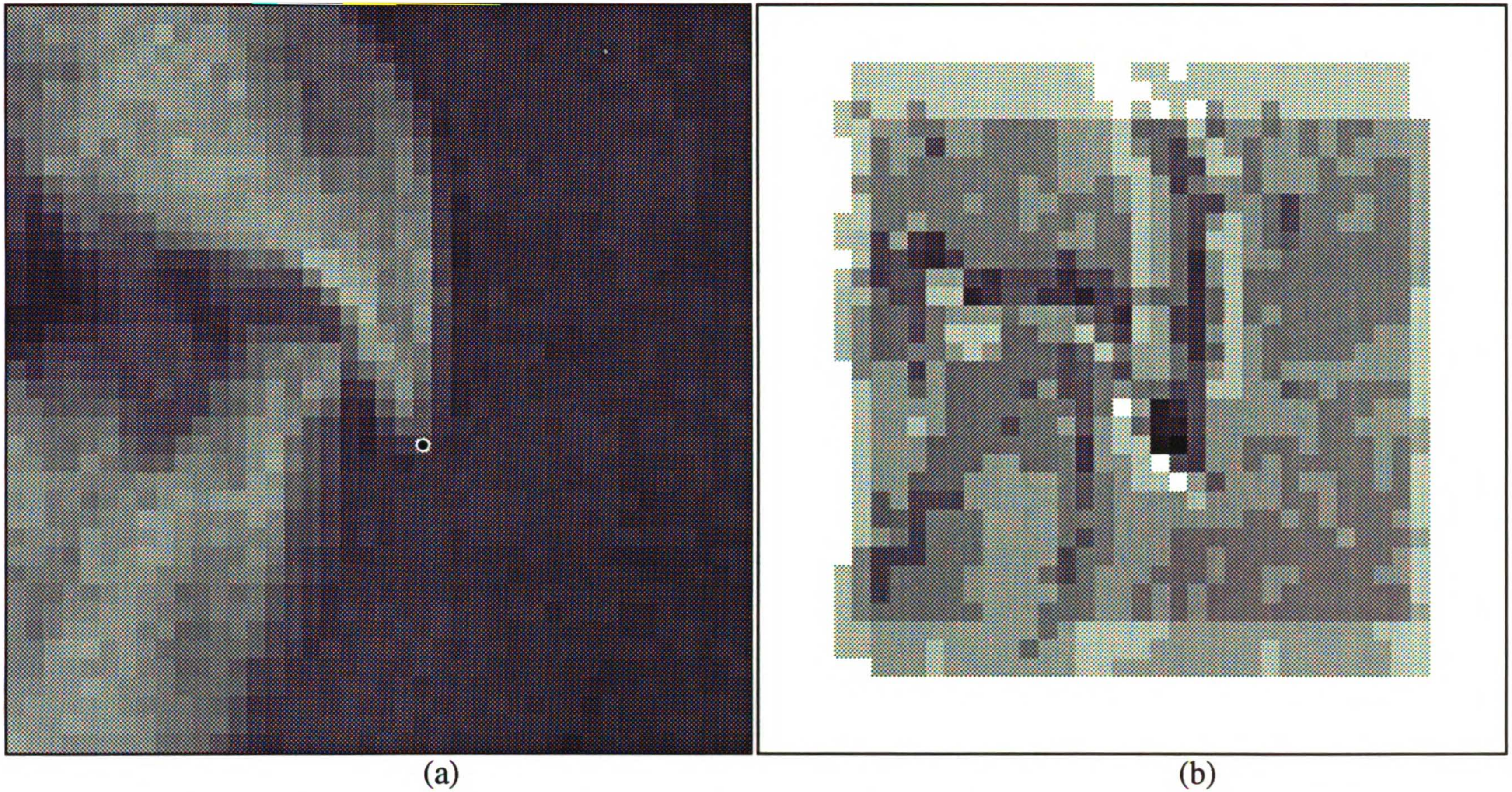
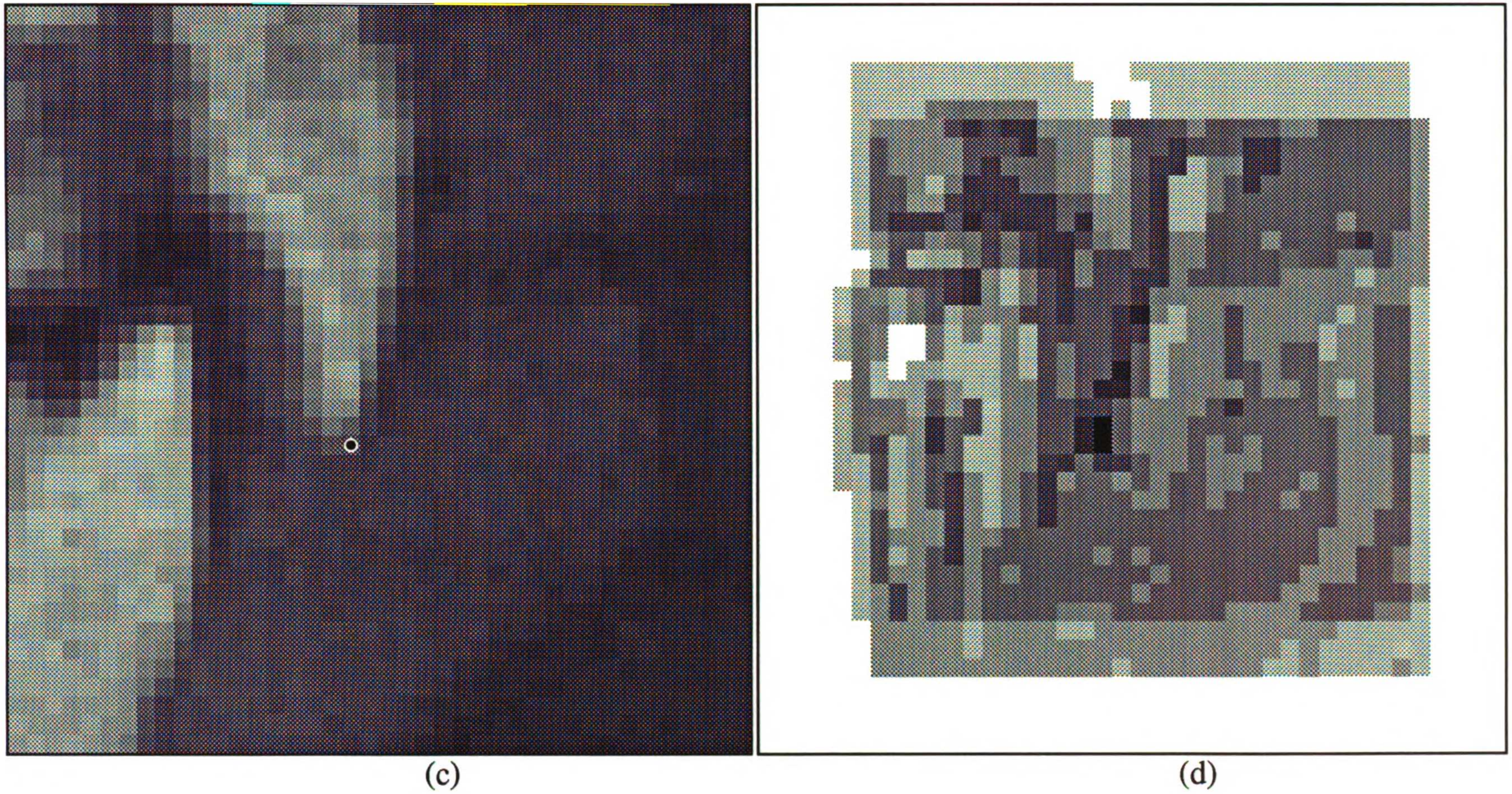
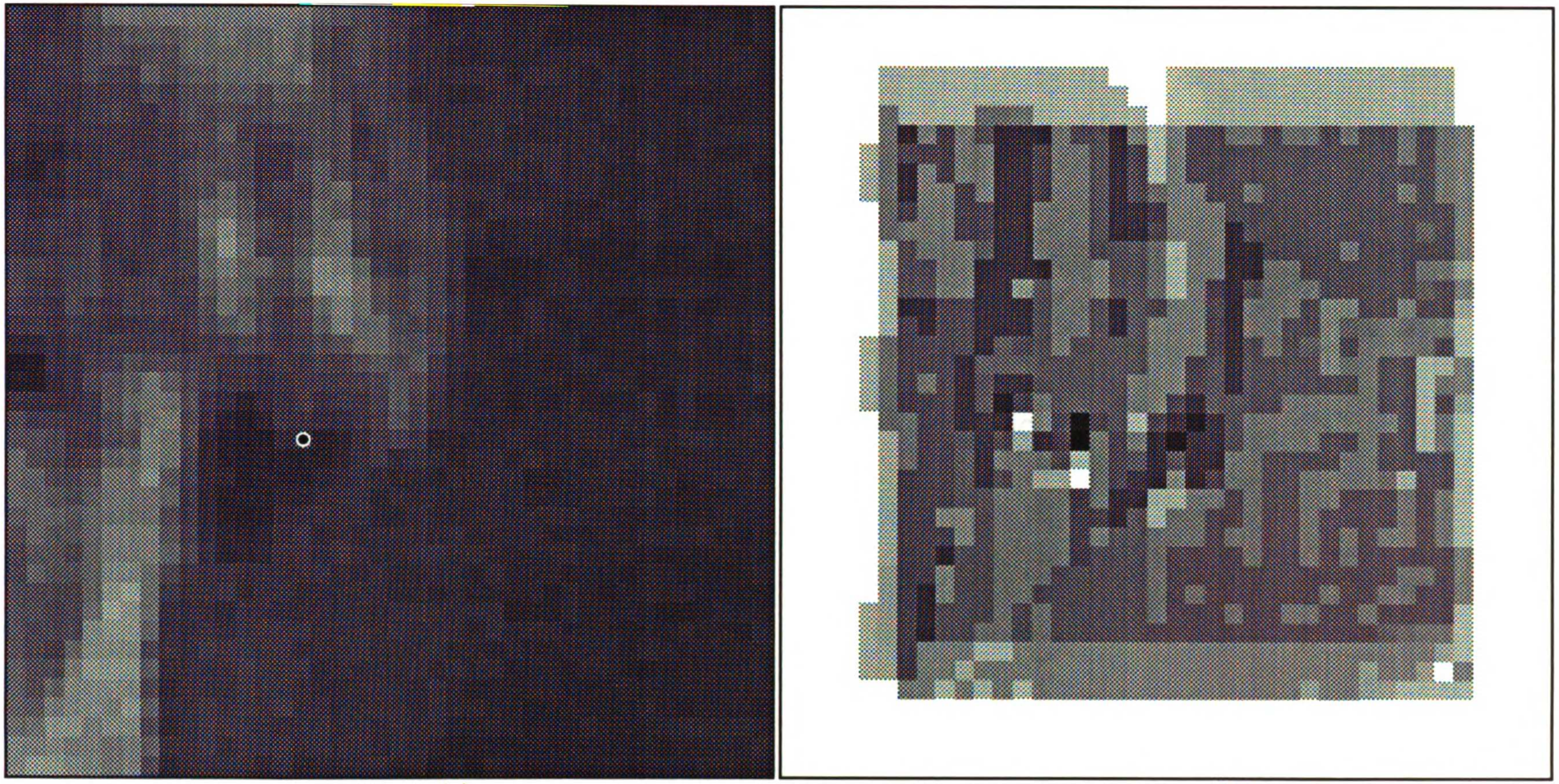


Figure 33 - Results of network with one hidden layer and 5x5 connection. (a) Test image 01205a with networks maximum liklihood output of UIEdge indicated by the black dot. (b) Network likelihood output.



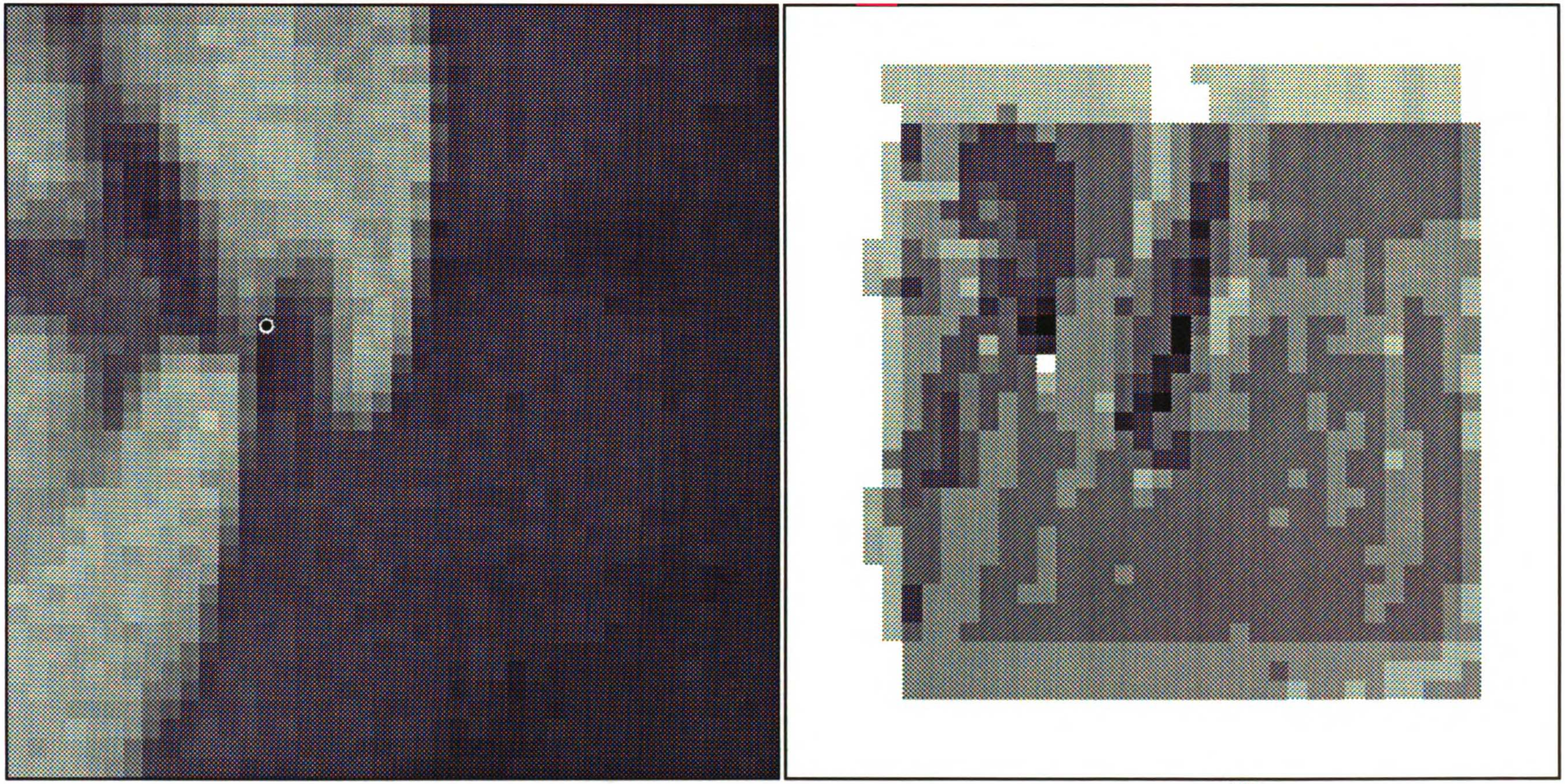
(c) Test image 01405a with networks maximum liklihood output of UIEdge indicated by the black dot. (d) Network likelihood output.



(a)

(b)

Figure 34 - Results of network with one hidden layer and 5x5 connection. Network located the edge of the posterior incisor. (a) Test image 01401a with networks maximum liklihood output of UIEdge indicated by the black dot. (b) Network likelihood output.



(c)

(d)

(c) Test image 01503a with networks maximum liklihood output of UIEdge indicated by the black dot. (d) Network likelihood output.

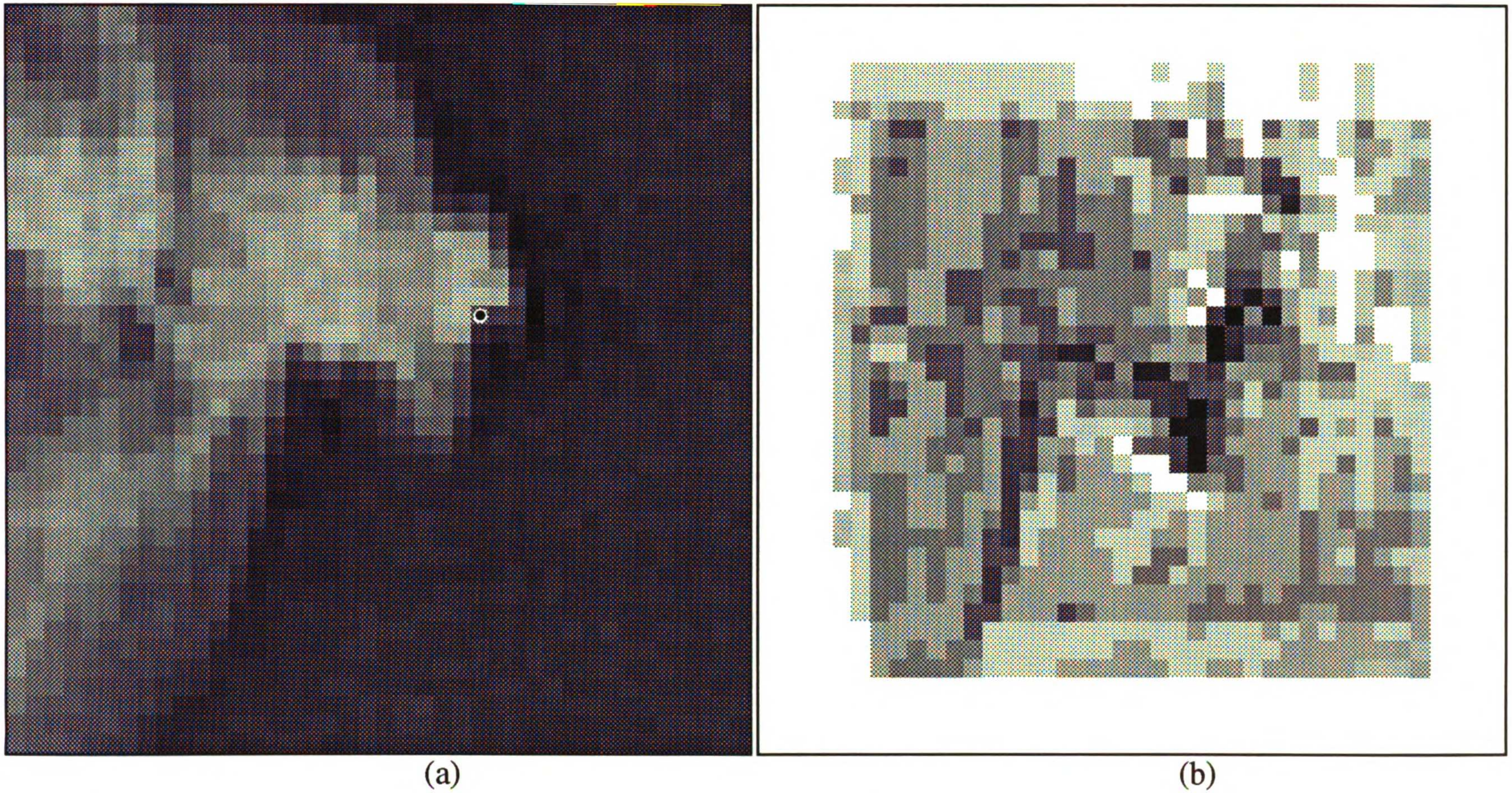


Figure 35 - Results of network with one hidden layer and 5x5 connection. This patient has braces. (a) Test image 01702a with networks maximum liklihood output of UIEdge indicated by the black dot. (b) Network likelihood output.

Network estimates of the UIEdge plotted with respect to three judges' mean using normalized 11 x 11 pixel training images

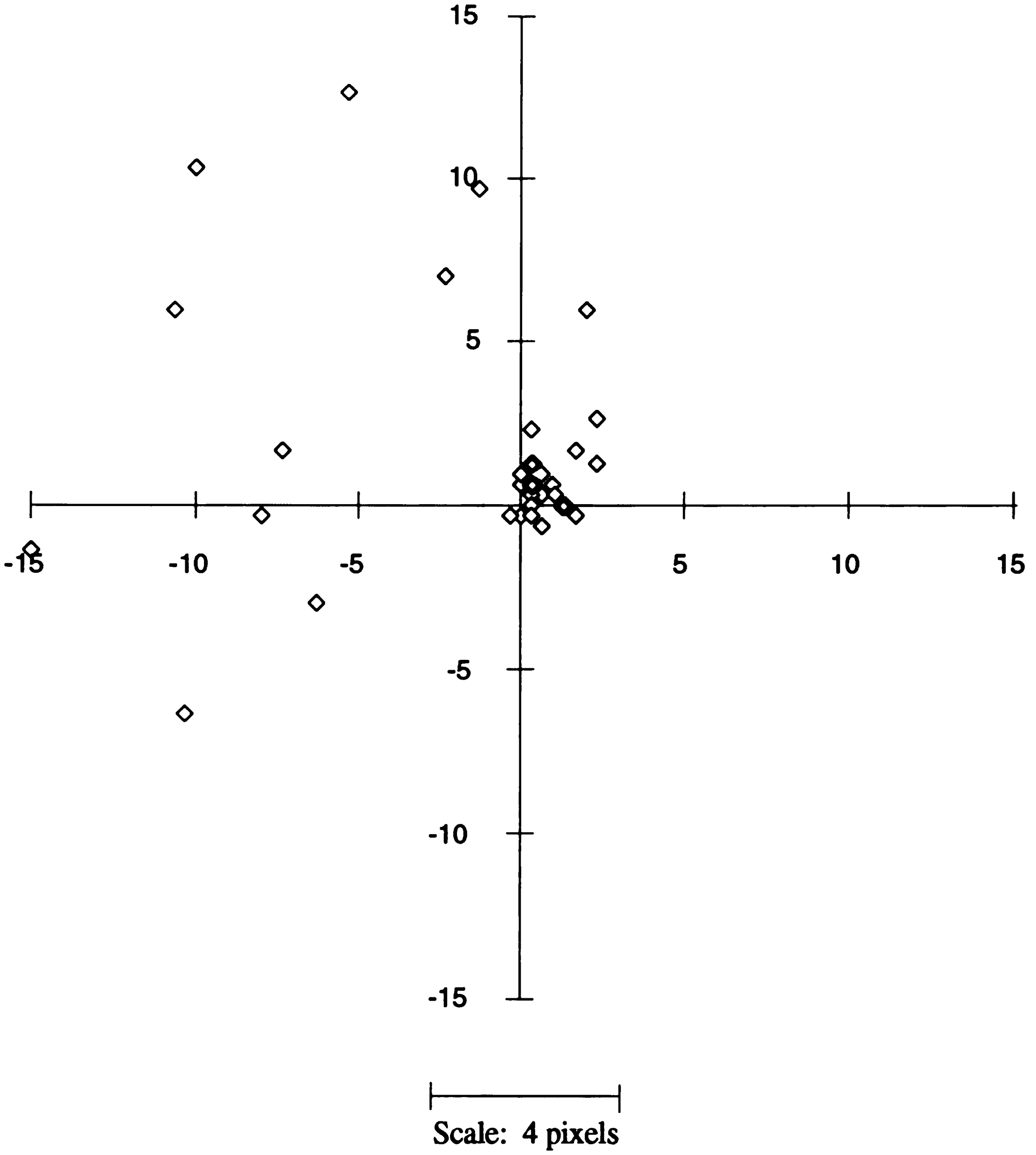


Figure 36. Network location of the UIEdge plotted with respect to three judges' mean for 36 test images. Training and test images were normalized to full gray scale capabilities. Network used 11 x 11 pixel training images. Bold data points indicate the presence of two or more overlaid data points. Axes are pixels in X and Y with the *best estimate* at the origin.

Network estimates of the UIEdge plotted with respect to training judge's mean using normalized 11 x 11 pixel training images

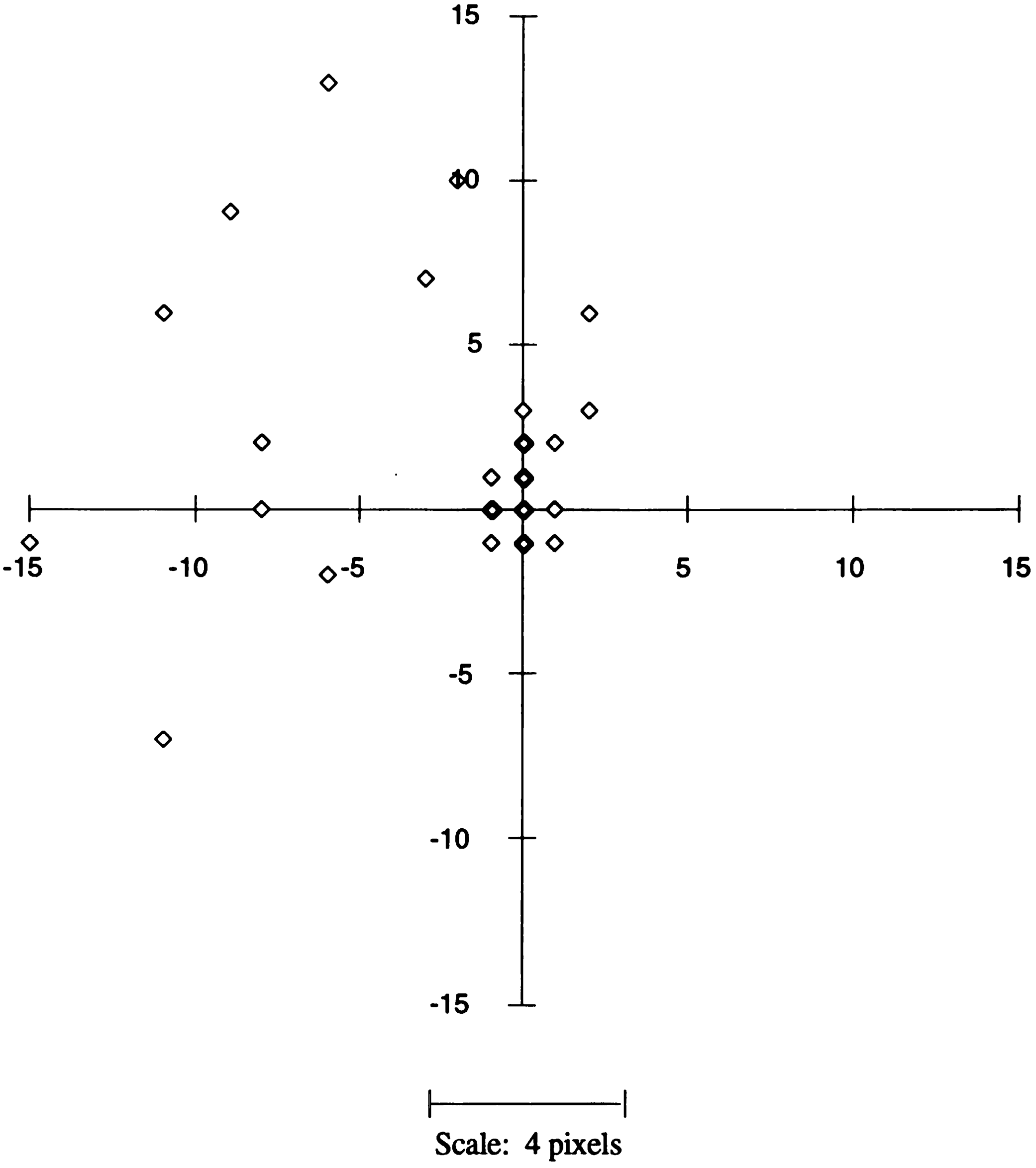
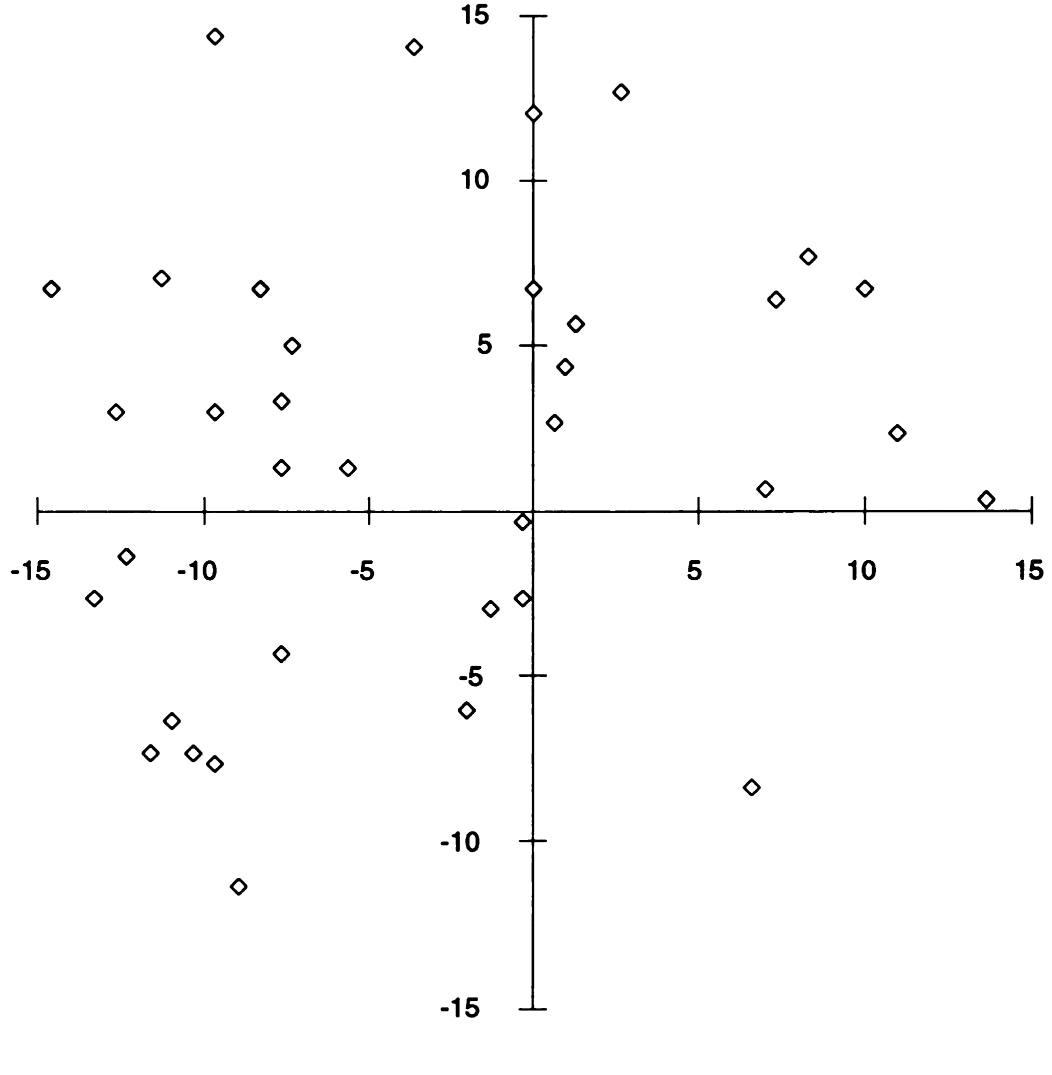


Figure 37. Network location of the UIEdge plotted with respect to training judge's mean for 36 test images. Training and test images were normalized to full gray scale capabilities. Network used 11 x 11 pixel training images. Bold data points indicate the presence of two or more overlaid data points. Axes are pixels in X and Y with the *best estimate* at the origin.

Network estimates of the UIEdge plotted with respect to three judges' mean using Sobel filtered 11 x 11 pixel training images



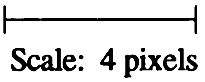


Figure 38. Network location of the UIEdge plotted with respect to three judges' mean for 36 test images. Training and test images were convolved with a Sobel filter. Network used 11 x 11 pixel training images. Axes are pixels in X and Y with the ***best estimate*** at the origin.

Network estimates of the UIEdge plotted with respect to three judges' mean using 21 x 21 pixel training images

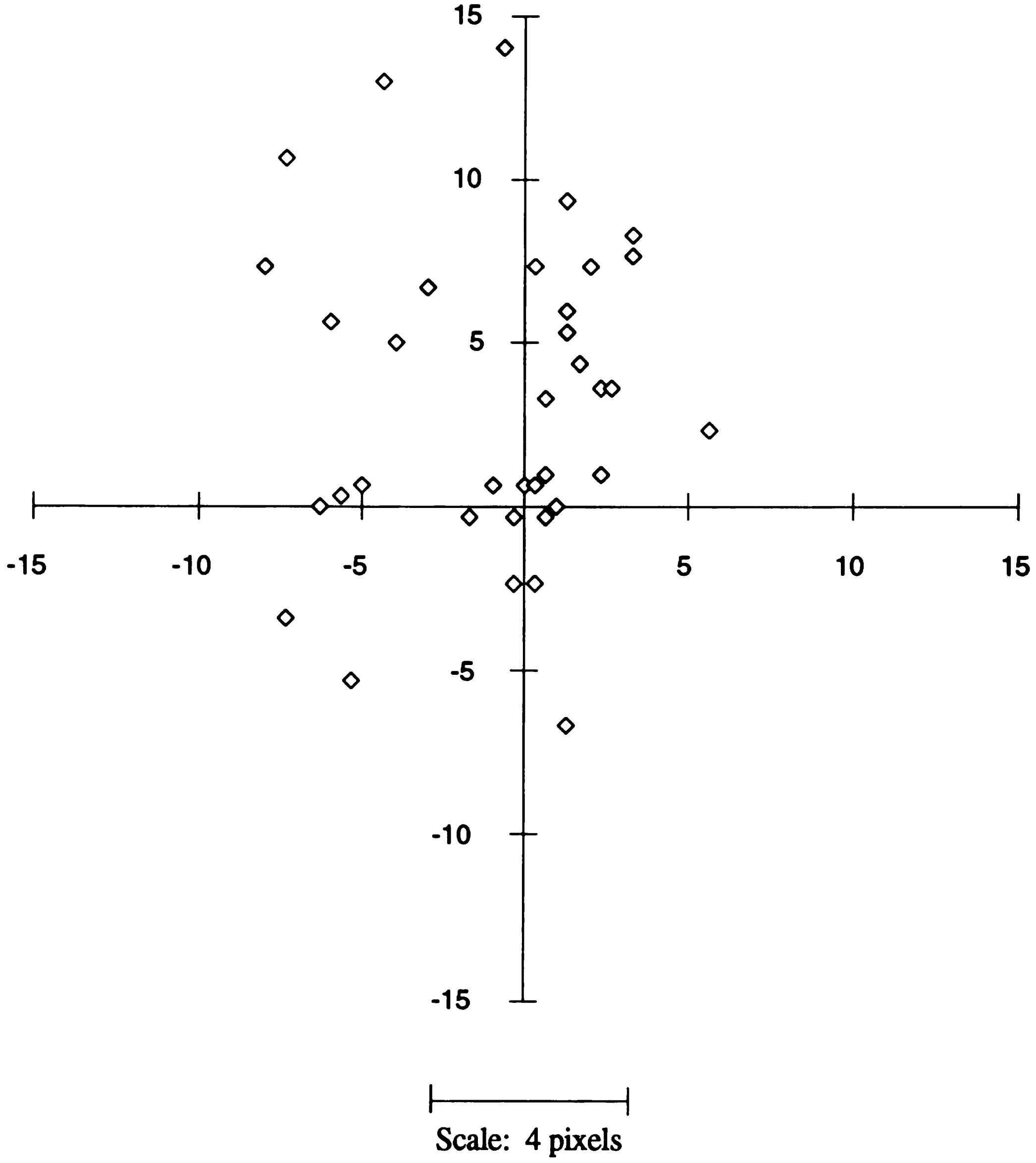


Figure 39. Network location of the UIEdge plotted with respect to three judges' mean for 36 test images. Network used 21 x 21 pixel training images. Axes are pixels in X and Y with the *best estimate* at the origin.

Network estimates of the UIEdge plotted with respect to training judge's mean using 21 x 21 pixel training images

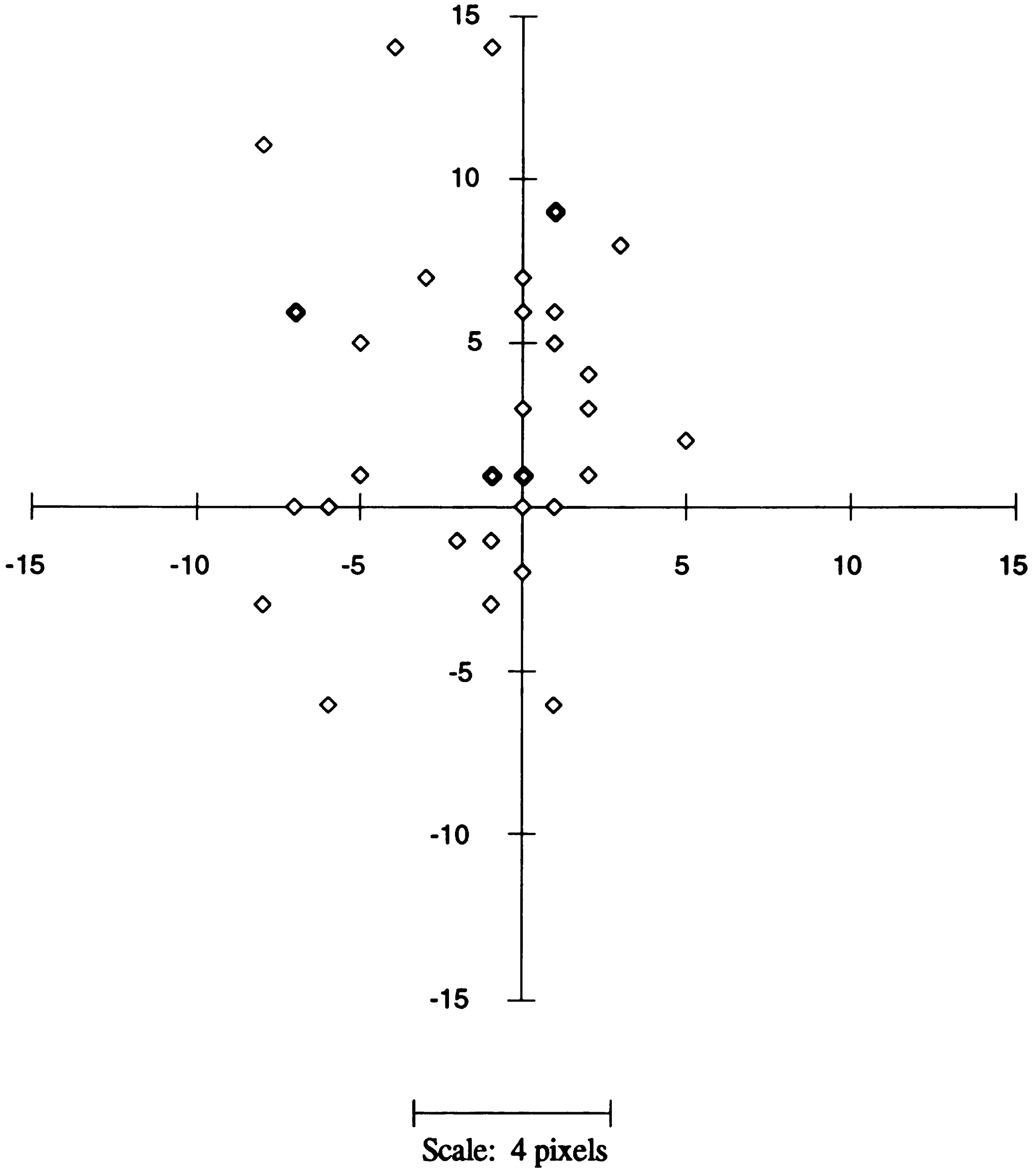


Figure 40. Network location of the UIEdge plotted with respect to training judge's mean for 36 test images. Network used 21 x 21 pixel training images. Bold data points indicate the presence of two or more overlaid data points. Axes are pixels in X and Y with the *best estimate* at the origin.

Network estimates of the UIEdge plotted with respect to three judges' mean using 13 x 13 pixel averaged border training images

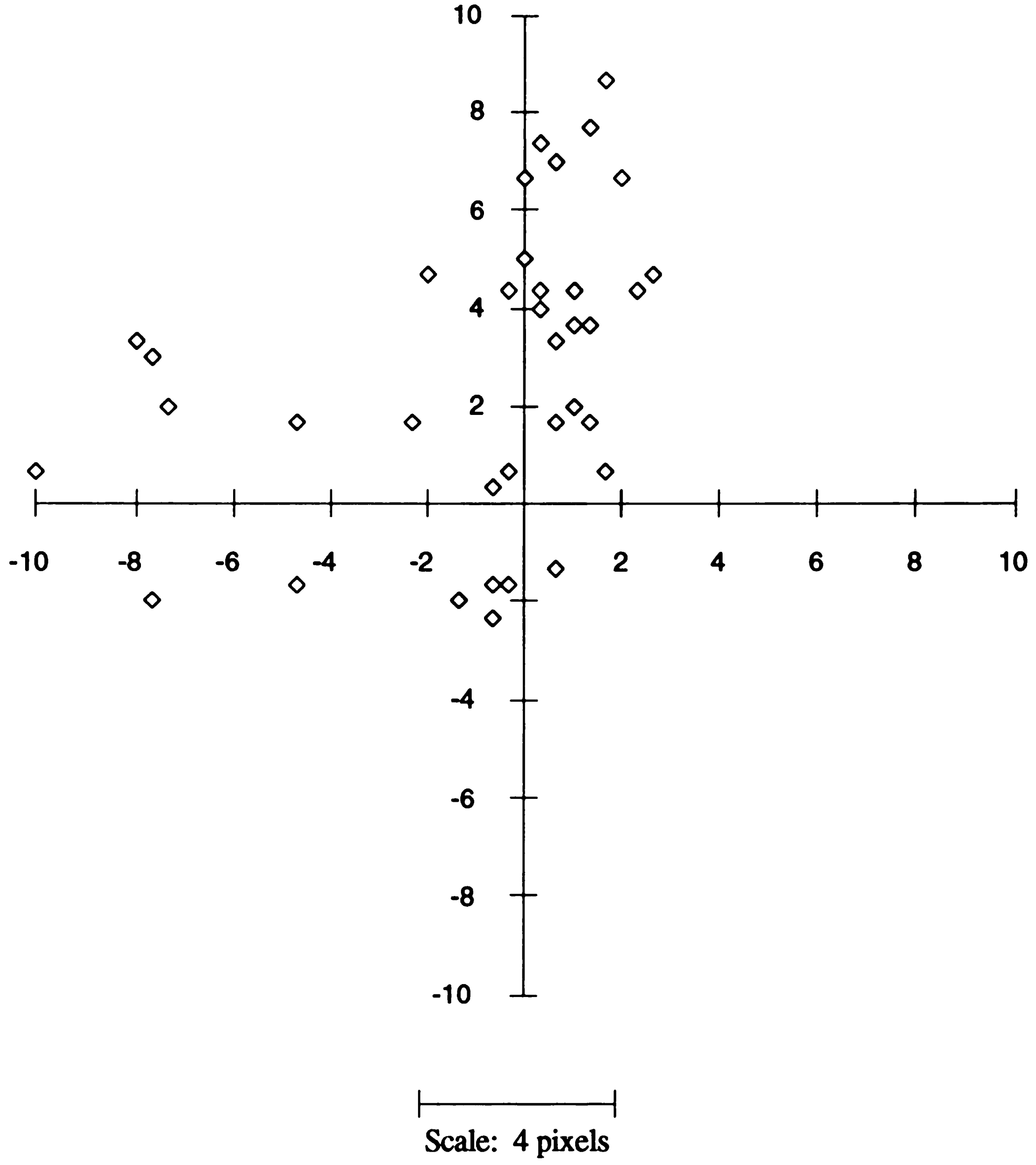


Figure 41. Network location of the UIEdge plotted with respect to three judges' mean for 36 test images. Network used 13 x 13 pixel averaged border training images. Axes are pixels in X and Y with the *best estimate* at the origin.

Network estimates of the UIEdge plotted with respect to training judges' mean using 13 x 13 pixel averaged border training images

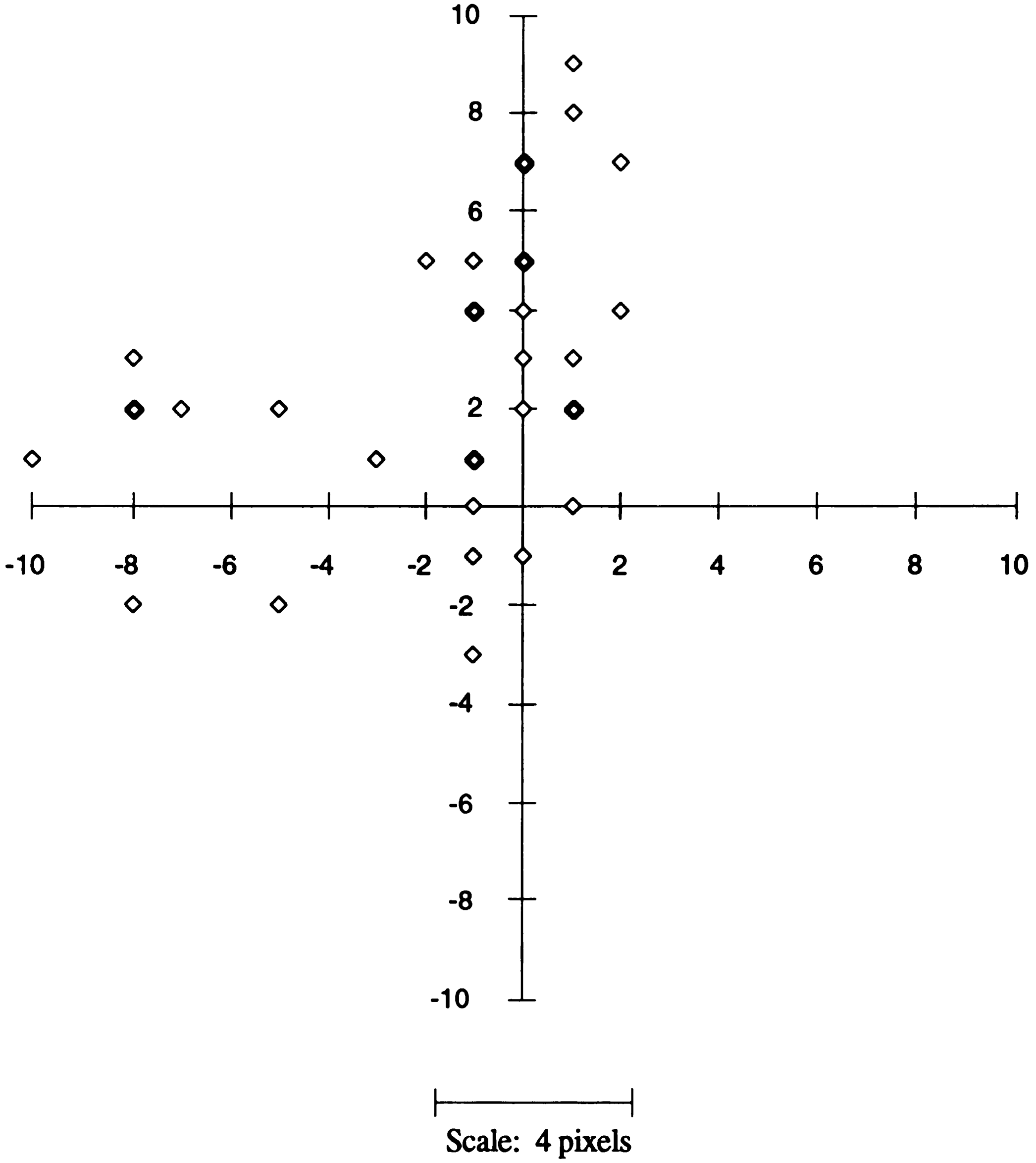


Figure 42. Network location of the UIEdge plotted with respect to training judge's mean for 36 test images. Network used 13 x 13 pixel averaged border training images. Bold data points indicate the presence of two or more overlaid data points. Axes are pixels in X and Y with the *best estimate* at the origin.

Discussion of Experiments with the UIEdge

The test set of the UIEdge images was specifically chosen to represent the variability in the image database. It contained a number of types of images which were not represented in the training set. Specifically, two of the images were of patients with orthodontic fittings (Figure 35(a)) and two of the images were of poor quality (Figure 34(a)). In twelve of the test images the unerupted upper lateral incisor edge was discernible (Figure 34(c)).

The network trained on 11 x 11 pixel images located twenty nine of the incisors within two standard deviations of the three training judges' mean. But looking at the negative results more carefully, the networks responses are consistent with the characteristics of the training set. Of the seven images in which the network was unable to locate the landmark successfully, two were the images of poor quality, three were images in which the network located the lateral incisor edge, and one was an image in which the network located the bottom edge of the orthodontic fitting on the upper incisor. Interestingly enough, the network's second likelihood value on these six images was the UIEdge. Clearly the network had problems with some of the images outside of its knowledge base. On the other hand, the network did locate the UIEdge on one image of a patient with orthodontic fittings and on nine images with discernible lateral incisors. Expanding the training set to include a greater representation of both positive and negative training pairs should improve results. Considering that the network segmented the test image into 11 x 11 images, it is not surprising that a lateral incisor was occasionally picked over the UIEdge. There were no negative training pairs of a lateral incisor and at the level of 11 x 11 pixel images I suspect that even a human judge would be hard pressed to determine whether an edge was the lateral incisor or the UIEdge.

Normalizing the UIEdge images actually reduced network performance to twenty four values within $2\sigma_{71}$ of the best estimate. This can be explained because the training set was designed for the original gray scale images and preprocessing of the images may require a different training set in order to optimize performance. It is interesting to note that the network's results conform more closely to the training judge's mean compared to the three judges' mean. As in the case of the fiducial, the network was trained to the systematic biases of the training judge. I expected the network to only be able to perform within the limitations of its training set and that a training set with more variability would yield better results.

Convolving the images with a Sobel filter to enhance the differential pixel intensities of the UIEdge did not improve results but in fact worsened them. Only one of the thirty six values was within $2\sigma_{71}$ of the best estimate. Apparently, the Sobel convolved training set did not include enough information in order for the network to characterize the UIEdge's location.

The network was trained with 21 x 21 pixel images to determine if capturing a larger section of the image could reduce the problems of erroneous selection of a lateral incisor's edge. However, increasing the size of the images increased the number of parameters (nodes and weight coefficients) dramatically and results did not improve. Clearly the training set did not contain enough variability for the network to be able to characterize the problem.

When the network was trained with 13 x 13 pixel averaged border images, performance was better than with the 21 x 21 pixel images. However, the network located only twelve of the UIEdges within $2\sigma_{71}$. When the network had been trained under other conditions, if the correct structure was located the accuracy was fairly good. Given that the network could not perform within the criteria on the training set it is not surprising that it did not perform within the criteria on the test images. Presumably, expanding the training set should give the network a better fix on the UIEdge.

Chapter 6

General Discussion and Recommendations

General Discussion

In order for neural networks to be used as landmark locators with sufficient accuracy, results must be within criteria set for human judges. Using the criterion of the 1971 Reliability Study, the network was able to locate twenty six of thirty six (74 percent) of fiducials and twenty nine of thirty six (81 percent) of the UIEdge landmarks. This compares favorably with the results achieved by Parthasarathy and Tong et al.[9, 10]. I believe these results could be further improved by expanding the training sets, by adding sub-pixel processing, and by the use of higher resolution images.

The network clearly trained to the characteristic distribution of the training judge for both the reference and anatomic landmarks. This is very encouraging because it suggests that the network can be trained to locate any landmark given both variety and consistency in the training set.

The network tended to fail on anatomy outside of its training set. By increasing the comprehensiveness of the training set, the network can learn to locate these landmarks.

This does point out the benefit of a learning algorithm as opposed to the fixed definition algorithms. Though the training set is ideally designed to be as comprehensive as possible, there may always be unusual cases that are not included in the training set. With a neural network algorithm, these cases can be incorporated into the training set, at any time, followed by retraining, thus expanding the knowledge of the network.

The results of this study imply that the optimal number of connections is determined by the size and shape of the landmark. Larger landmarks require that more information be available to each node. The 3x3 connection scheme was sufficient to characterize and locate the test pattern but a 5x5 connection scheme was required to locate the fiducial with any consistency. While different connection schemes were not examined for the UIEdge, it seems likely that a larger connection scheme might improve results.

One hidden layer was found to be the optimal number of layers for the 3x3 and 5x5 connection schemes. Another way to vary the connection scheme and the number of parameters in the network is to vary the number of hidden nodes in a layer. One alternative is to have fewer hidden nodes but a fully connected network. This is the approach most often used by networks designed for pattern recognition. However, if the number of hidden nodes in a layer does not match the number in the input image, a more complex connection scheme would have to be designed. This makes understanding what is happening in the network much more difficult and reduces our ability to reverse engineer the solution to the problem. I chose to approach the problem of hidden layers by hidden nodes corresponding to pixels in the input image because this mimicked the parallel nature of the input and output layers, and allowed a connection scheme similar to fields of vision. In addition, three-layer perceptron models have been shown to be able to form arbitrarily complex decision regions [38]. This evidence suggests that networks with two hidden layers and larger and more comprehensive training sets would give better results.

Using an output layer with nodes that corresponded to nodes in the input layer did provide the ability to evaluate which anatomic features were conflicting with the "correct" location of the landmark. This was helpful in designing the training sets, as well as evaluating network results. It also provides a means for further processing -- namely sub-pixel localization to give a closer approximation to the *best estimate* of the landmark.

The network was able to train to locate both the fiducial and the UIEdge despite differences in shape and size. While there is a significant difference in shape between anatomic landmarks, the quality of the network results is largely dependent on the design of the training set. Thus, I believe that from the results from the UIEdge studies I infer that the network will apply well to other anatomic landmarks.

The networks were robust enough to generalize across patients and across age differences. This variability was incorporated into the training set and thus was evident in the results. However, the training set did not include the full range of anatomic variability. For example, there were no patients with severe overbite or Class III over jet. I suspect that the network would be able to locate the landmark on some images of unusual anatomy but not others, just as it located the UIEdge correctly on one image with orthodontic fittings but not on another. Alternatively, with severe differences in anatomy, it may be necessary to train the network selectively. For example, more than one training set would be developed to account for severe cases of anatomy.

In the cases of both the fiducial and UIEdge, the actual location of the landmark was a point on some edge of a structure. A Sobel filter was used in this study but different gradient operators such as the Kirsch or Presitt operators could be used in different orientations to enhance edges [34, 39, 40] . There are a few landmarks however, such as sella, the midpoint of the pituitary fossa, which though defined by edges are not physically on an edge. These landmarks will require some careful consideration of the connection scheme of the network. If the pituitary fossa is fifteen pixels across, a 5x5 connection scheme will not provide each node in the network with enough information. Instead, I would either try a larger connection scheme, or in the interests of minimizing parameters, pick a semicircular region of connection that corresponds to the pituitary fossa.

Summary

Though this study was limited to the evaluation of network results for two landmarks, there are two conclusions from which we can reasonably infer that the network can be applied successfully to other landmarks as well. First, despite variations in anatomy and patient age, the network was successfully able to locate the UIEdge on patients not in the training set. In particular, the network was able to locate the UIEdge with accuracy on some but not all of the types of images not represented in the training set. This implies that the training set need not be completely comprehensive in order to perform well. Second, the network trained to the systematic bias of the training judge. It appears intuitively reasonable that if the network can be trained to a single judge it certainly can be trained to a "gold standard".

Recommendations for Future Work

For the immediate future, one research direction should be an investigation of the extent to which one can increase the connection scheme while still characterizing the landmark location. I have preliminary results which indicate that, given the training data available, fully connected networks contain too many parameters to converge to unique solutions. I recommend that further studies be done to investigate the results of 7x7 and 9x9 connection schemes. For landmarks such as sella, where key features appear to be on the periphery of the connection scheme, altering the shape of the connection scheme to match the landmark configuration may tailor the network more appropriately to the landmark.

Expanding the available training data could provide more conclusive evidence of the usefulness of a second hidden layer. Expanding the training set would also improve the network's ability to segment images currently outside of the training set. The training set used in this study generally consisted of "normal" patients. It would be worthwhile to examine the degree of "abnormality" that can be incorporated into the training set while still maintaining the accuracy desired. I also recommend that further processing techniques, such as sub-pixel localization, be explored as a means of improving network results.

While it is important to design the training set to incorporate the variety of anatomy, it is equally important that the training set have a consistency and accuracy that the locating task demands. The network should be trained with a "gold standard" for other landmarks. This "gold standard" would consist of the mean of three to five experienced judges' estimates of the landmark on the training images. This would reduce individual judge bias in the training set.

Finally, the network should be tested with a wide variety of landmarks in order to assess the generality of my results. This would also include landmarks from photographic images of patients, study casts of patient teeth, and stereo images.

In the next five years with advancements in research and development, including computer image technology, network design considerations, and the development of fully comprehensive training sets, neural networks can replace many of the labor intensive landmark locating tasks. In the short term, I propose a combined man-machine strategy for automating landmark location. The network would segment the image, then display its first choice for each landmark. A human judge would then determine whether a given landmark's location is acceptable. If not, there are several alternatives, the network's second choice can be presented, the judge can redirect the network to segment a particular section of the image, or the judge could select the landmark location. If the network is consistently missing the landmark, the training set should be modified to reflect the discrepant locations. As in the case of human judges, replicated assessments are necessary to provide an acceptable confidence level. This can be achieved by independently training two or more networks and comparing results. This may alleviate the need for human judge

Bibliography

1. Baumrind, S. and B. Baker, Reliability of Human Performance in Locating Landmarks on Analog and Digital Duplicate X-ray Images. In *Proc. of the IEEE/EMBS*, pp. 342-343, Orlando, FL, 1991.

2. Baker, B., *et al.*, Reliability of Cephalometric Landmark Location Directly on a Computer Monitor. In *Proc. of the IEEE/EMBS*, p. 333, Orlando, FL, 1991.

3. Baumrind, S. and R. Frantz, The Reliability of Head Film Measurements. 1. Landmark Identification. *Am. J. Orthod*, vol. Vol. 60, no. , pp. 111-127, 1971.

4. Medioni, G. and R. Nevatia, Matching Images Using Linear Features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 675-685, 1984.

5. Li, X. and R.C. Dubes, The First Stage in Two-Stage Template Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, pp. 700-707, , 1985.

6. Forstner, W., *Chapter 16: Image Matching,* in *Computer & Robot Vision,* R. Harelick and L. Shapiro, Editor. 1993, Addison-Wesley: p. 289-378.

7. Forstner, W., A Feature Based Correspondence Algorithm for Image Matching. *International Archives of Photogrammetry and Remote Sensing*, vol. 26, no. 3, pp. 150-166, 1986.

8. Levy-Mandel and et. al., Knowledge-based landmarking of cephalograms. *Computers in Biomedical Research*, vol. 19, pp. 282, 1986.

9. Parthasarathy, S., *et al.*, Automatic landmarking of cephalograms. *Computers and Biomedical Research*, vol. 22, pp. 248-269, 1989.

10. Tong, W., *et al.*, An Algorithm for Locating Landmarks on Dental X-rays. In *Proc. of the IEEE Engineering in Medicine and Biology Society*, pp. 552-4, Seattle, WA, November, 1989.

11. Baker, B., S. Baumrind, and S. Curry, Neural Network Approach to Computerized Analysis of Digitized Cephalograms. *Journal of Dental Research (Special Issue)*, vol. 69, no. 339, pp. Abstract #1846, 1991.

12. Baker, B., S. Curry, and S. Baumrind, A Neural Network Method for Solving Pattern Recognition Problems in Craniofacial X-ray Image Analysis. In *Proc. of the IEEE/EMBS*, Seattle, WA, 1989.

13. Baker, B., S. Baumrind, and S. Curry, Neural Network Approach to Computerized Analysis of Digitized Cephalograms. In *Proc. of the International Association of Dental Research*, pp. Poster #1846, Cincinnati, Ohio, 1989.

14. Cortes, C. and J.A. Hertz, A Network System for Image Segmentation. In *Proc. of the International Joint Conference on Neural Networks*, pp. 121-125, 1989.

15. Fukishima, K., A Neural Network Model for Selective Attention in Visual Pattern Recognition. *Biological Cybernetics*, vol. 55, pp. 5-15, 1986.

16. Koch, C., J. Marrochin, and A. Yuille, Analog 'Neuronal' Networks in Early Vision. *Proc. National Academy of Science*, vol. 83, pp. 4263-4267, 1986.

17. Kohonen, T., *Self Organization and Associative Memory*. 1984, Berlin: Springer-Verlog.

18. Grossberg, S. and E. Mingolla, Neural Dynamics of Perceptual Grouping: Textures, Boundaries, and Emergent Segmentations. *Perception and Psychophysics*, vol. 38, no. 2, pp. 141-171, 1985.

19. Grossberg, S., Nonlinear Neural Networks: Principles, Mechanisms, and Architectures. *Neural Networks*, vol. 1, pp. 17-61,1988.

20. Lippmann, R.P., An Introduction to Neural Nets. *IEEE ASSP Magazine*, vol. , April, pp. 4-22, 1987.

21. Rumelhart, D.E. and J.L. McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol. 1. 1986, MIT Press.

22. Boone, J.M., V.G. Sigillito, and G.S. Shaber, Neural networks in radiology: An introduction and evaluation in a signal detection task. *Medical Physics*, vol. 17, no. 2, pp. 234-241, 1990.

23. Okajima, M., *et al.*, Computer Pattern Recognition Techniques: Some results with Real Electrocardiographic Data. *IEEE Transactions on Bio-Medical Electronics*, vol. BME-10, no. 3, pp. 106-114, July, 1963.

24. Yasui, S., G. Whipple, and L. Stark, Comparison of Human and Computer Electrocardiographic Wave-form Classification and Identification. *American Heart Journal*, vol. 68, no. 2, pp. 236-242, August, 1964.

25. Wang, S.L. and P.Y. Li, Neural Networks for Medicine: two cases. In *Proc. of the Electro International Conference*, pp. 586-90, New York, NY, April, 1991.

26. Silverman, R.H., Segmentation of Ultrasonic Images with Neural Networks. *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 5, no. 4, pp. 619-628, 1991.

27. Chen, C.T., E.C. Tsao, and W.C. Lin, Medical Image Segmentation by a Constraint Satisfaction Neural Network. *IEEE Transaction on Nuclear Science*, vol. 38, no. 2, pp. 678-686, April, 1991.

28. Hutchinson, R.A., Development of an MLP Feature Location Technique using Preprocessed Images. In *Proc. of the International Neural Network Conference*, pp. 67-70, Paris, 1990.

29. Curry, S., *A Point Location and Tracking System Employing Digital Images For Use In Biostereometrics*. PhD in Engineering thesis, University of California, Berkeley, 1985.

30. McCulloch, W.S. and W. Pitts, A Logical Calculus of the Ideas Imminent in Nervous Activity. *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115-133, 1943.

31. Guyton, A.C., *Medical Physiology*. 1981, Philadelphia: W.B. Saunders Company.

32. Rummelhart, D.E., G.E. Hinton, and R.J. Williams, Learning Representations by Back-Propagating Errors. *Nature*, vol. 323, pp. 533-536 1986.

33. Baker, B., A. Liu, and S. Baumrind, Graphic Display of Craniofacial x-ray Data Overlaid on Digital Images. In *Proc. of the IEEE-EMBS*, pp. 334, Philadelphia, 1990.

34. Gonzalez, R.C. and P. Wintz, *Digital Image Processing*. ed. R. Kalaba. 1977, Reading, Massachusetts: Addison-Wesley Publishing Company. 431.

35. Akey, M.L. and R. Mitchell, Detection and Subpixel Location of Objects in Digitized Aerial Images. In *Proc. of the Seventh Conference on Pattern Recognition*, pp. 411-414, Montreal, 1984.

36. Berenstein, C.A., *et al.*, A Geometric Approach to Subpixel Registration Accuracy. *Computer Vision, Graphics, and Image Processing*, vol. 40, pp. 334-360, 1987.

37. Tian, Q. and M.N. Huhns, Algorithms for Subpixel Registration. *Computer Vision, Graphics, and Image Processing*, vol. 35, pp. 220-233, 1986.

38. Duda, R.O. and P.E. Hart, *Pattern Classification and Scene Analysis*. 1973, New York: John Wiley & Sons.

39. Rosenfeld, A. and A.C. Kak, *Digital Image Processing*. 1982, New York: Academic Press.

40. Levine, M.D., *Vision in Man and Machine*. 1985, New York: McGraw-Hill.

Appendix A. Software User's Guide

There are two key pieces of software necessary to implement the neural network: BIBASE and Network. Currently, BIBASE is used to develop the training sets and the test images for the neural network. Once training sets are developed, two files are required: a file specifying the names of the training images and a file specifying the names of the test images. The network is then set up and trained and images are tested using software called Network.

BIBASE

was developed on an 80386 PC to display images via the PCplusVision frame grabber on an RGB SONY Trinitron monitor. The software is written in C and makes use of library functions that accompany the frame grabber.

The primary task of the software is to acquire and display digital images. In addition to displaying images, the software can also display other information from the database, including landmark location data and the associated angular and distance measurements. The software also provides the ability to design the network training sets. Portions of the digital images can be easily selected and stored in files for later use as training and test images. Preprocessing of the images can be performed by or by the neural network software.

The size of the training and test images, the starting search position, and landmark name is hard coded in variables in the file nn.c. Changing the name of the landmark allows access to database values of its position.

The procedure for designing a training pair within is as follows:

(1) Retrieve the desired image

(2) Select training set option from the neural nets menu (under the annotate menu)

(3) Move box around screen until desired location of input image has been selected then hit return to save image.

(4) Select appropriate time point and output image will be saved.

The procedure for designing a test image within is as follows:

(1) Retrieve the desired image

(2) Select save image option from the neural nets menu (under the annotate menu)

(3) Move box around screen until desired location of test image has been selected then hit return to save image.

Listing Files

Once the training set and the test images have been formed, two listing files must be generated. The format of the listing files is the following:

Training Images List

# of training images

<input training image> <output training image>

For example the file might look like this:

```
3
train1.in train1.out
train2.in train2.out
train3.in train3.out
```

Test Images List

# of test images

<test image>

For example the file might look like this:

```
4
test1
test2
test3
test4
```

Network

The neural network software was developed in C on a Sun SPARCII which was networked to the PC. Modifications can be made to the network through a global variable file (externs.h), or by compiling different versions of the network code. The software gives the user the ability to vary the size of the training images, the size of the test images, the connection scheme, the number of hidden layers, and what kind of preprocessing if any should be done. Weighting coefficients can be stored in a file to be used again on successive runs of the network. The network segments the test images and produces network output images. The network output for each pixel in the image is the likelihood that it is the landmark location.

Network prompts for the following information:

(1) Random number seed

(2) Number of hidden layers

(3) Name of the training file list

(4) Want to retrieve old weights? If yes then weight file name.

(5) Number of epochs and the stop error

Network now proceeds to train... When network has finished training:

(6) Name of the test images file list

Network proceeds to segment images....

(7) Do you want to save weights? If yes then weight file name.

Network places the output of each image in a file. If the test image is called test1 then the output is placed in a file called test.out. Input and output files can be viewed with a program called display. Once results are evaluated, the training set may need to be redesigned. This can be done by repeating the steps outlined above.