

SEGMENTATION OF ULTRASONIC IMAGES WITH NEURAL NETWORKS

RONALD H. SILVERMAN

*The Margaret M. Dyson Vision Research Institute
Cornell University Medical College
1300 York Avenue
New York, NY 10021, USA*

Neural networks differ from traditional approaches to image processing in terms of their ability to adapt to regularities in image structure and to self-organize so as to implement directed transformations. Biomedical ultrasonic images are often degraded in quality by noise and other factors, making enhancement techniques particularly important. This paper describes use of back propagation and competitive learning for enhancement and segmentation of ultrasonic images of the eye. Of particular interest is the extension of these techniques to segmentation of three-dimensional data sets, where simple thresholding and gradient operations are not entirely successful.

Keywords: Neural network; Back propagation; Competitive learning; Image processing; Ultrasound; Segmentation.

1. INTRODUCTION

Neural networks with their massively parallel architectures are a natural method for image processing. A number of investigators have described applications of neural networks for a variety of vision and image processing operations.¹⁻⁴ This paper describes the use of back propagation and competitive learning for the processing of biomedical ultrasound images.

Ultrasonic imaging is based on generation of a compressional pulse by a voltage transient placed across a piezoelectric crystal, propagation of the pulse through tissue, and reconversion of echoes into voltages by the same crystal. Display brightness is a function of echo amplitude, and display position is determined by the range to each echo and transducer orientation. Range is determined by the time interval between pulse emission and echo return. The transducer may be either scanned in an arc or linearly translated to sweep out a two-dimensional image.

Ultrasonic reflection occurs at tissue interfaces between regions having differing acoustic impedances, where impedance is the product of tissue density and the speed of sound. Reflection amplitude is also affected by Snell's Law, resulting in poor visualization of surfaces aligned obliquely to the beam propagation axis. Other factors

This work was supported in part by NIH grant EY03183, the Dyson Foundation, the St. Giles Foundation, the Rudin Foundation and Research to Prevent Blindness. Computations for this project were performed in part at the Cornell National Supercomputer Facility, which is supported by the National Science Foundation, New York State, the IBM Corporation, and members of the Corporate Research Institute.

which may complicate the interpretation of ultrasonic images include backscatter by unresolvable tissue components, low signal-to-noise ratios, attenuation by intervening tissue, and the range dependence of lateral resolution for focused transducers. For the above reasons, generally in ultrasonic images there is poor contrast between different tissue components, surfaces often appear discontinuous, and solid tissue is subject to speckle.

Improvements in acoustic image quality have proceeded from several directions, including better "front-ends" (i.e. transducers, amplifiers, displays), application of signal processing techniques (i.e. bandpass filtering, deconvolution), and post-processing of digital images. This will be concerned primarily with the last approach.

An extensive library of image processing operators exists. Operators may act in either the time (spatial) or frequency (Fourier) domain. Such operators carry out specific image transformations, e.g. smoothing, sharpening. Two or more such operations may be applied to an image sequentially, but such transformations are not commutative. The choice of which operators to use and in what order is, therefore, as much art as science.

The recent introduction of devices for the acquisition of serial ultrasound scans from parallel planes has lent further urgency to the need for better acoustic image segmentation techniques. Serial scans permit three-dimensional reconstruction of tissue structures. Such reconstructions are most conveniently produced on computer workstations using object-oriented software. The low intrinsic contrast between the soft tissues imaged with ultrasound makes the two most commonly used 3-D segmentation operators, thresholding and gradient detection, less than satisfactory.

This paper describes the use of neural networks for boundary enhancement and image segmentation, including extension to three-dimensional data sets. The two techniques that will be described are back propagation and competitive learning.

Back propagation⁵ is a form of directed learning involving multistage feedforward networks comprising units having non-linear (usually sigmoidal) activation functions. During training, the global error (i.e. the squared difference between the teaching and output vectors) is minimized by determination of the dependence of the global error on individual link weights throughout the network. Link weights are modified so as to reduce this error. Because of the combination of non-linearity and the use of hidden layers, functions of arbitrary complexity can be implemented by such networks, making them of interest as (among other things) image processing operators.

Competitive learning procedures have been described by a number of investigators.⁶⁻⁹ The conceptual basis of this technique is preferential learning by units which are most well adapted to patterns present in a teaching set. For each input pattern, the most sensitive unit "wins" and gets to modify its afferent connections, while all others are inhibited. The winning unit thus becomes both more well adapted and more specialized. Through this mechanism, competing units gradually segment pattern space. As a form of undirected learning with a single modifiable layer of connections between input and output, competitive learning will be limited to the detection of less complex patterns than back propagation, but will converge much

more quickly. This is not a trivial concern given that biomedical images generated by modern instrumentation are typically 512×512 or more in dimension.

2. METHODS

Ultrasonic images of the human eye were acquired with weakly focused 10 MHz transducers. The digital acquisition unit had a 50 MHz sample rate. On each scan were stored 128 scan lines consisting of 2048 8-bit samples. Serial scans were acquired as rectilinear slices 25.5 mm wide separated by 0.5 mm intervals. Rectification and averaging were used to reduce scan data to 128×128 images. (This sacrifices axial resolution, but reduces image size to a manageable level for experimentation.) Pixel values were converted to real values ranging from 0 to 1.

The back-propagation procedure, as described by Rumelhart, Hinton and Williams,⁵ was used as a directed learning procedure. All back-propagation networks described in this paper accepted square input regions from the image plane as input, had two hidden layers, and a single output unit. During processing, the network input region was scanned over the entire image, and the value of the output unit at each position was interpreted as the transformed pixel intensity of the center of the square input region. During training, the dependencies of the global error on link weight values were accumulated and then adjusted using gradient descent with a momentum term.⁵

In competitive learning, input was accepted from square regions of images as described above. Ten competing units were completely connected to the input region. On each stimulus position, the "excitation" of each competing unit was determined by

$$x_j = \sum_i (y_i - w_{ij})^2 \quad (1)$$

where y_i refers to the value of the i th element of the input vector, x_j indicates the excitation of competing unit j , and w_{ij} is the link weight between these elements. The excitation is thus the sum of squares of the differences between inputs and corresponding link weights. The unit with the *minimum* excitation (i.e. the minimum difference in the least squares error between the input vector and the weight vector) was the "winner". The winning unit on each trial adapted by adjusting link weights as shown in the following equation

$$\Delta W_i = k_w (y_i - w_{ij}) \quad (2)$$

where k is a small positive constant. In fact, k_w was allowed to decrease with a shallow exponential function during training so as to gradually decrease the learning rate.

Losing units were also allowed to adapt using the same rule as given in Eq. (2), but with a constant k_l , such that $k_l \ll k_w$. Upon the initiation of training, $k_l = 0.01k_w$. k_l , however, was allowed to increase in units which had very low firing rates, and to decrease in units with very high firing rates. This assured that all units fell towards valid stimulus patterns and eventually became sensitive to some subset of pattern space.

Competitive learning was applied to both single scans and serial scan series. Typically, about 24 slices were acquired during a serial scan series. For serial scan data, the learning procedure was similar to that used for processing of single slices, but the input region consisted of a rectangular solid of $5 \times 9 \times 9$ voxels, where the dimensions indicate that 9×9 regions across five scans form an input vector. This is equivalent to a volume measuring $2.5 \times 1.8 \times 2.2$ millimeters.

3. RESULTS

Figure 1 illustrates the capacity of back propagation to simulate the mode of action of a classical image processing operator. The image in the top left of the figure is the original gray-scale image of the eye. The orientation is such that the anterior of the eye is on the left and the posterior is to the right. The bright feature on the left is the anterior sclera. In the plane of this scan, the cornea, iris, and lens do not appear. The central portion of the image is the vitreous compartment, which is black, because it does not (normally) produce acoustic reflections. On the right are the posterior coats of the eye (retina, choroid, sclera) and orbital fat. The elevated region extending into the vitreous is an ocular tumor.

The image on the top right results from application of a span-5 median filter to the original image. The image on the bottom left results from use of a span-2 gradient filter on the median image.

A neural net consisting of a 2×2 input zone, two hidden layers of four units each (plus bias), and a single output was trained to simulate the function of the gradient operator by using the median images as input and the gradient image as teacher. The result after 1000 iterations, shown on the bottom right of Fig. 1, closely matches the action of the gradient operator.

While this result illustrates the capability of back propagation to transform images, of greater interest is the inverse problem of creating a filter of unspecified structure to carry out some arbitrary operation. This is illustrated in Fig. 2. In this case, the teaching image (top right) consists of manually delineated boundaries of interest, i.e. the sclera, retina, and tumor. A total of 500 iterations were performed. In the first 250 iterations, a learning step occurred after each scan line. Subsequently, one learning step occurred after each pass over the entire image. The input regions, in this case, were 9×9 in dimension. The network contained two hidden layers of nine units each (plus bias). The result after training is shown on the bottom left. Even with a single image as a training set, we found that the application of the trained network to other images gave satisfactory results.

Figure 3 is an example of tumor localization through use of back propagation. In this case a training set of ten examples of ultrasonic images of eyes with ocular tumors was used. Paired with each input image was a training image in which only positions corresponding to the tumor were set as active. The network, again, accepted 9×9 input regions and had two hidden layers of nine units each. Input regions, however, were taken at three image scales: 128×128 , 64×64 , and 32×32 . This procedure increased tolerance for variations in scale. The output of the trained network is shown

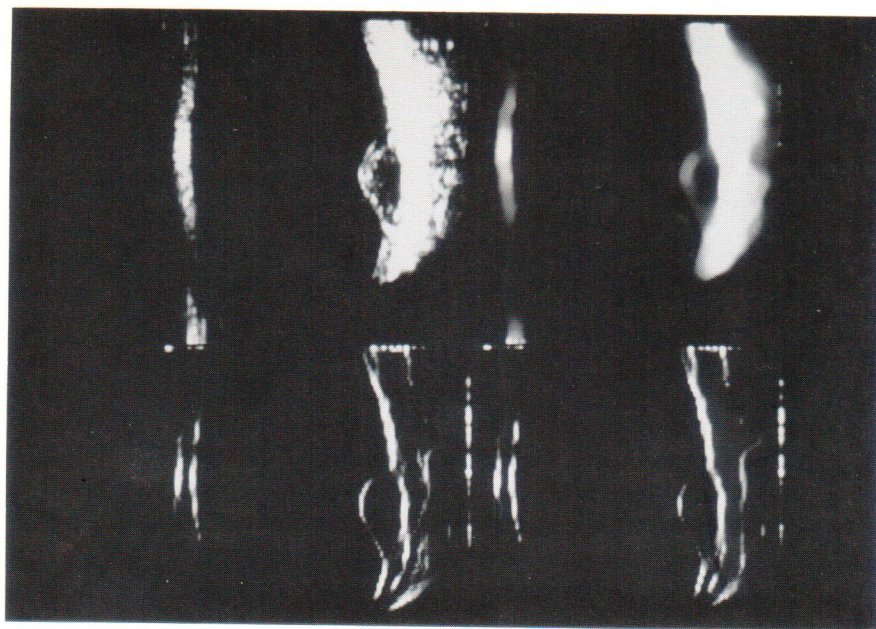


Fig. 1. Top left: Ultrasonic image of eye. Orientation is anterior to posterior from left to right. Elevation extending into anechoic (black) vitreous compartment from posterior wall is choroidal malignant melanoma. Top right: Ultrasonic image after median filter transformation. Bottom left: Gradient transform of median image. Bottom right: Output after training of back-propagation network with median image as input and gradient image as teacher.

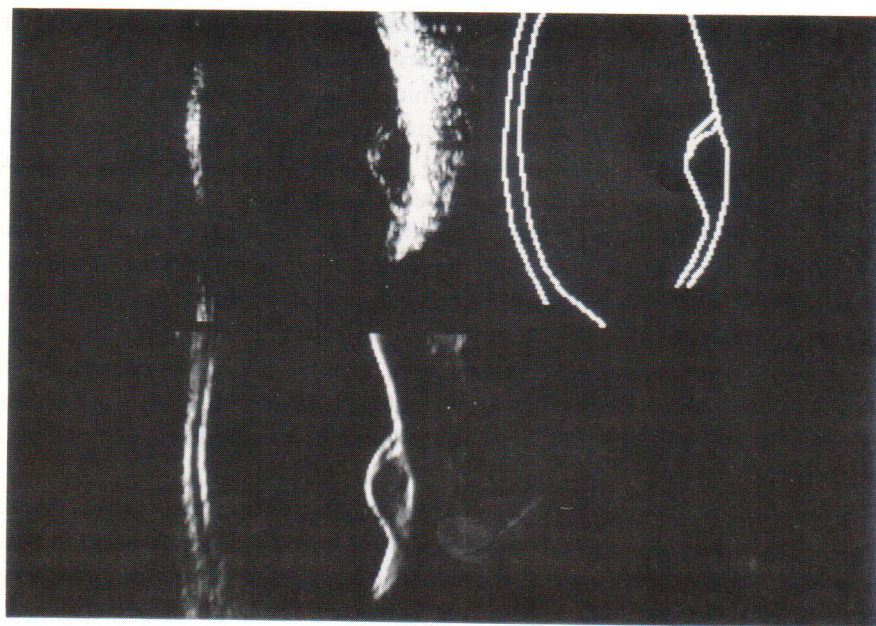


Fig. 2. Top left: Ultrasonic image input to back-propagation network. Top right: Teaching image, defined by manually delineated boundaries. Bottom left: Network output after training.

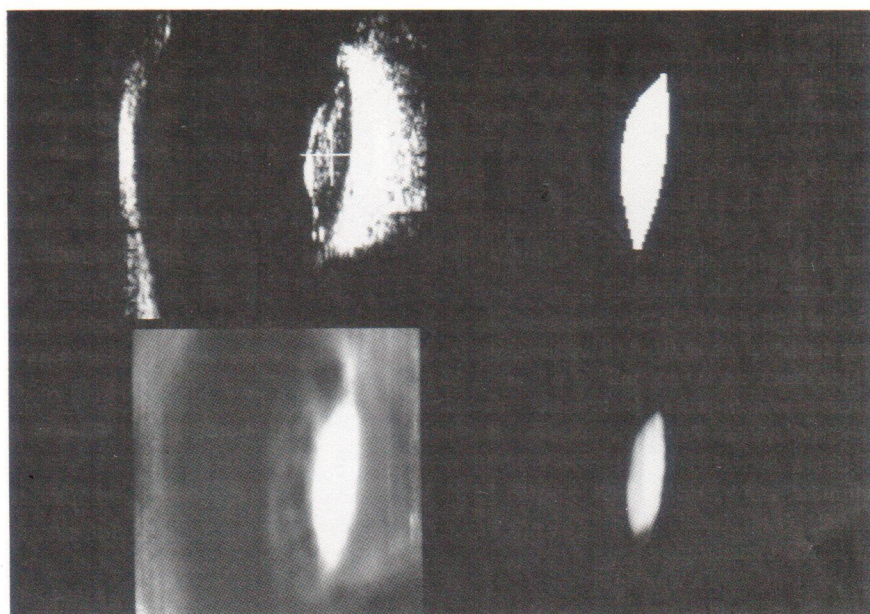


Fig. 3. Top left: Ultrasonic image input to back-propagation network. Top right: Teaching image, defined by manually delineated region corresponding to intraocular tumor. Bottom left: Network output after training. Bottom right: Output after thresholding.

on the bottom left, and the output after thresholding, on the bottom right. This technique has been previously described in greater detail.¹⁰

An example of the use of competitive learning in image segmentation is shown in Fig. 4. The original gray-scale image is shown on the top left. The network was trained with five iterations of the image (14 400 patterns per iteration). The image immediately to the right of the gray-scale input image is the color-coded, superimposed output of all the competing units. The other images correspond to the individual outputs of each of the ten competing units. Although this learning modality is undirected, the network was found to have segmented the input image into anatomically interesting features.

Unit #9 (dark green, bottom, second from right) has largely demarcated the boundaries of the tumor. Unit #3 (orange, middle left) corresponds to the choroid. Orbital fat is largely segmented by units #2 (salmon, upper right) and #7 (dark blue, bottom left).

The extension of competitive learning to a three-dimensional scan series is illustrated in Fig. 5. In this case, 25 sequential scans were acquired of an eye in which an intraocular malignant melanoma was present. In Fig. 5 (top left) the gray-scale data has been displayed using volume-rendering software. The serial slices are stacked along the Z-axis. The X-axis refers to axial distance, so the anterior of the eye is near $X = 0$. The Y-axis corresponds to lateral distance. The orientation of the eye in Fig. 5

is such that the viewer's perspective is from above and behind the globe. In the gray-scale rendering, the globe has been sectioned across the Z-axis so as to expose a cross section of the tumor. The black circular feature seen opposite and anterior to the tumor is the pupil.

Figure 5 (top right) is the output of the network after one iteration across the entire scan series. In Fig. 5 (bottom left) the globe has been sectioned in a manner similar to the gray-scale image, so as to expose a lateral tumor cross section. In Fig. 5 (bottom right), a second cut has been made in a plane perpendicular to the Y-axis, to expose an additional cross section through the tumor.

In the gray-scale rendering, the tumor appears to be relatively homogeneous, despite gray-scale remapping for increased contrast. In the images generated by the neural network, the tumor appears to be quite heterogeneous, especially in its anterior aspect. This characteristic is consistent with the tortuous vasculature that is often seen in fast-growing, pedunculated lesions of this type.

4. CONCLUSION

In the paper it is shown that neural networks using either directed or undirected learning may be used to segment ultrasonic images of the eye. These techniques are applicable to other radiologic images and to processing of gray-scale images in general.

Back propagation has the advantage of being a directed learning procedure and of having the capability of detecting very complex patterns. Its disadvantage is its slow convergence rate which necessitates (especially for image processing) high speed or special purpose computers. The tumor segmentation problem as described in this paper, for instance, was run on an IBM 3090/600.

Competitive learning, though undirected, converges rapidly. Simulations for this study were run in a few hours on a MicroVax II and a Personal Iris/25 workstation.

The color-coded segmentation of images created by competitive learning networks can be expected to be superior to simple color-recoding of the gray-scale for contrast enhancement. This is because the network segments on the basis of both gray-scale intensity and spatial structure. Although competitive learning does not appear to be capable of fully isolating the tumor, it does appear to improve display of subtle structural features as compared to the gray-scale image.

Although identification of anatomic or pathologic features is not necessarily difficult for trained personnel, as in other image understanding problems, it is a challenge for the computer. In this paper our intention was not demonstration of the superiority of the neural network over radiologists in the detection of lesions or other anatomic components, but rather illustration of the network's capacity to automate this procedure for processing of multiple scan planes. Applications include both presentation of segmented images as an aid to anatomic comprehension, and quantification of volumes of specific anatomical entities. Future directions for this line of research will include extension of directed learning to three dimensions.

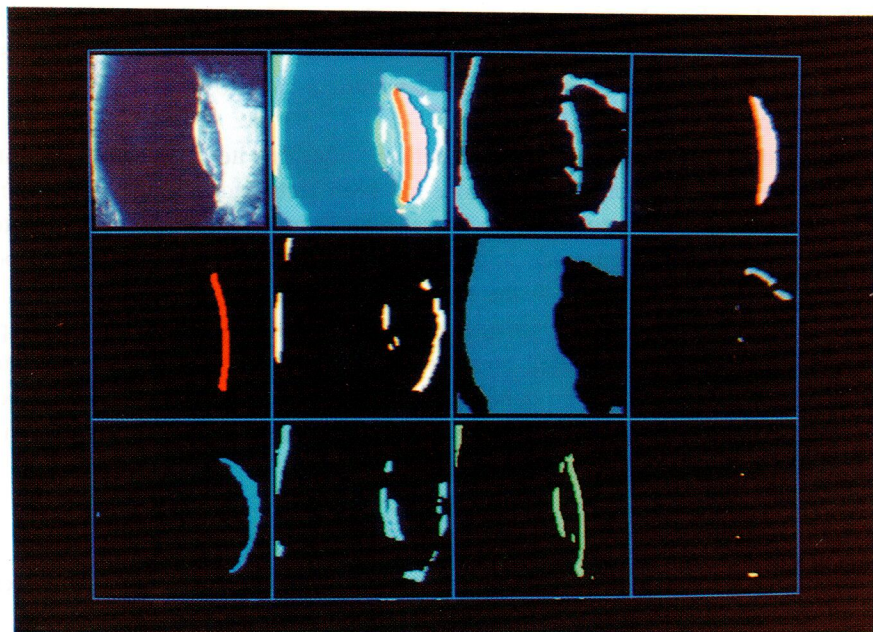


Fig. 4. Segmentation of ultrasonic image (top left) by competitive learning. Image to right of gray-scale input is composed of superimposed, color-coded outputs of ten competing units. Remaining images correspond to outputs of individual units.

REFERENCES

1. T. Kohonen, *Self Organization and Associative Memory*, Springer-Verlag, Berlin, 1984.
2. K. Fukushima, "A neural network model for selective attention in visual pattern recognition", *Biol. Cybern.* **55** (1986) 5-15.
3. C. Koch, J. Marrochin and A. Yuille, "Analog 'neuronal' networks in early vision", *Proc. National Academy of Science U.S.A.* **83** (1986) 4263-4267.
4. T. J. Sejnowski, C. Koch and P. S. Churchland, "Computational neuroscience", *Science* **241** (1988) 1299-1306.
5. D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning representations by back-propagating errors", *Nature* **323** (1986) 533-536.
6. C. von der Malsberg, "Self-organizing of orientation sensitive cells in the striate cortex", *Kybernetik* **14** (1973) 85-100.
7. D. E. Rumelhart and D. Zipser, "Feature discovery by competitive learning", *Cogn. Sci.* **9** (1985) 75-112.
8. S. Grossberg, "Adaptive pattern classification and universal recoding: Part I. Parallel development and coding of neural feature detectors", *Biol. Cybern.* **23** (1976) 121-134.
9. T. Kohonen, "Clustering, taxonomy, and topological maps of patterns", *Proc. Sixth Int. Conf. on Pattern Recognition*, ed. M. Lang, Silver Spring, 1982, IEEE Computer Society Press, pp. 114-125.
10. R. H. Silverman and A. S. Noetzel, "Image processing and pattern recognition in ultrasonograms by back propagation", *Neural Networks* **3** (1990) 593-603.

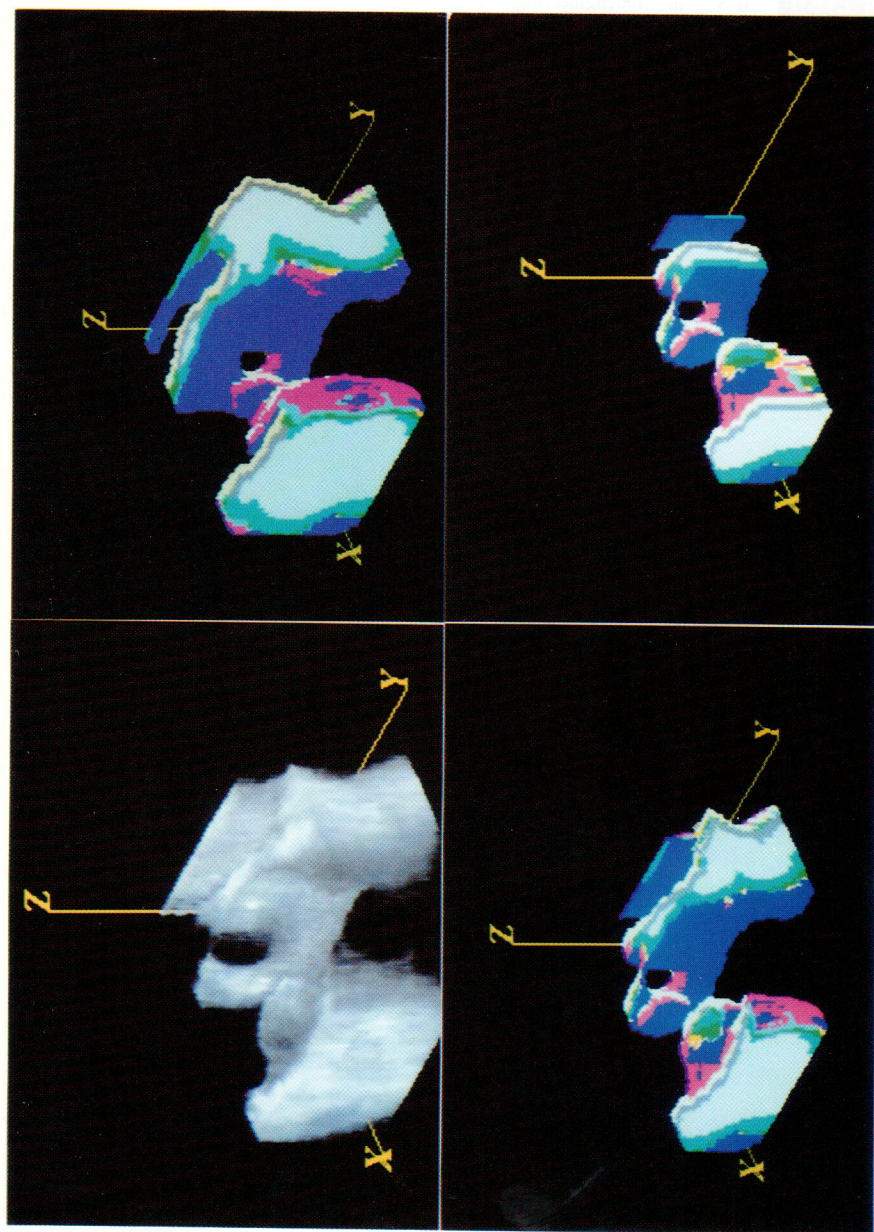
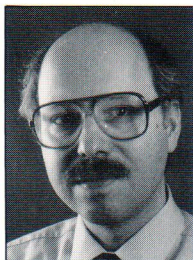


Fig. 5. Top left: Volume rendering of three-dimensional scan series. Top right: Output of competitive learning network after training. Bottom left: Network output sectioned transversely so as to expose tumor. Bottom right: Segmented image after additional sagittal cut.

Received 1 March 1991.



Ronald H. Silverman received a B.S. in chemistry from Brooklyn College in 1971. He received an M.S. in bioengineering (1979) and a Ph.D. in computer science (1990) from the Polytechnic University. He is currently Assistant Professor of

Computer Science in Ophthalmology at Cornell University Medical College, where he has worked since 1981. His research interests include biomedical signal and image processing, diagnostic and therapeutic ultrasound and neural network theory and applications.