Back-Propagation Network and its Configuration for Blood Vessel Detection in Angiograms

Reza Nekovei, *Student Member, IEEE,* and Ying Sun

***Abstract*— A neural-network classifier for detecting vascular structures in angiograms was developed. The classifier consisted of a multilayer feed-forward network window in which the center pixel was classified using gray-scale information within the window. The network was trained by using the back-propagation algorithm with the momentum term. Based on this image segmentation problem, the effect of changing network configuration on the classification performance was also characterized. Factors including topology, rate parameters, training sample set, and initial weights were systematically analyzed. The training set consisted of 75 selected points from a 256 × 256 digitized cineangiogram. While different network topologies showed no significant effect on performance, both the learning process and the classification performance were sensitive to the rate parameters. The best result was obtained with a small learning rate (0.05) and a medium momentum rate (0.5). The three-layer (121-17-2) network was adequate for the problem and showed good generalization to the entire cineangiogram and other images including direct video angiograms and digital subtraction angiograms. In a comparative study, the network demonstrated its superiority in classification performance; its classification accuracy was 92%, as compared to 68% from a maximum likelihood estimation method and 83% from a method based on iterative ternary classification. It was also shown that the trained neural-network classifier was equivalent to a generalized matched filter with a nonlinear decision tree.**

I. Introduction

THE computational structure using neural networks has similarities to that of the human vision system [1]. However, past research has shown very few applications of using neural networks to classify objects in images directly based on the image gray-scale data. Preprocessing is usually done on the image raw data or a set of features is extracted before applying a neural-network classifier [2]–[4]. The preprocessing stage reduces the amount of data from the image pixels and, therefore, makes the computation for neural-network classifiers more tractable. In this paper, it is our interest to evaluate the performance of neural-network classifiers that take the image gray-scale data as direct input. The specific problem under investigation is the identification of vascular structures from angiograms. Using this example, we also study how different network configurations affect the classification performance.

Angiography is used during various catheterization procedures for diagnosis and treatment of patients with blood vessel abnormalities, especially in the area of coronary artery disease [5], [6]. The time sequence of angiograms contains much information about vessel lumen geometry, dimensions, and blood flow. However, due to the noise induced by the complex imaging chain and the dynamic nature of the process (heart motions and infusion of the x-ray contrast agent), extraction of such information is not a trivial task. Digital angiography and computer algorithms have contributed to the reduction of inter-observer and intra-observer variability in performing measurement such as the percent diameter stenosis [7]. The need for automated vessel detection remains critical in computer based systems which perform complex and computation-intensive tasks such as assessing coronary blood flow [8] and reconstructing 3-D vascular structures [9]. Past research on automated detection of vascular structures has been based on statistical and heuristic methods [10]–[13].

Although tracking algorithms have significantly improved vessel identification [14], [15], they often require the operator's interventions for selecting parameters such as starting and ending search points. Some algorithms [16], [17] lack the capability of detecting the entire vascular network. None of the aforementioned algorithms has taken advantage of distributed parallel processing.

This study has two purposes: (1) to develop a practical approach based on neural network computing for the segmentation of coronary arteriograms, and (2) to study the effect of network configuration on classification performance as applied to the previously mentioned image segmentation problem. The paper is organized as follows. Following a statement of the problem, the design of the neural network classifier is described. Next, a systematic analysis of different network configurations is presented. Then, the results of angiogram segmentation are shown, including a performance comparison with two existing algorithms. Finally, the paper concludes with a discussion of the results and findings.

II. Statement of the Problem

An angiogram is a sequence of x-ray images obtained during the injection of an x-ray contrast agent into the circulatory system to be imaged. This is usually done during a catheterization procedure and the contrast agent is introduced via a catheter. Conventionally, angiograms are recorded on film. The film-based angiogram is termed cineangiogram. To input a cineangiogram to the computer, the film is projected and digitized off-line. In this paper, this type of digital images is referred to as the digitized cineangiogram (DCA). A modern angiographic system is often equipped with a video camera which focuses directly on the x-ray image intensifier,

Manuscript received November 21, 1992; revised December 6, 1993.

R. Nekovei is with the Remote Sensing Laboratory, University of Rhode Island/Bay Campus, Narragansett, RI 02882 USA.

Y. Sun is with the University of Rhode Island, Department of Electrical Engineering, Kingston, RI 02881 USA.

IEEE Log Number 9400114.

acquires the images, and sends them to a computer for on-line digitization. This is referred to as the direct video angiogram (DVA). A common practice to remove background for a better visualization of blood vessels is subtracting digitally two images, which are obtained before and after the injection of the contrast agent [18]. The resulting image is called the digital subtraction angiogram (DSA).

DCA's usually contain the highest level of noise due to the involvement of the complex imaging chain. DSA's usually have the highest signal-to-noise ratio but may contain subtraction artifacts caused by missed registration of the two images during subtraction. In this study, a DCA is used for network training because of its low image quality and the fact that segmentation of DCA's is generally more difficult. However, to cover a broad spectrum of image quality in the generalization test, the neural-network classifier designed for DCA's is also applied to DVA's and DSA's.

In this study, we demonstrate the suitability of using a back-propagation model for coronary artery detection in angiograms. We also investigate various multilayer networks to determine the optimal configuration for such application. The network is designed to identify blood vessels in cineangiograms, while avoiding false-positive detection of noise structures arising from background intensity variations, irrelevant anatomical details, quantum mottles, and other types of interference.

III. Architecture of the Classifier

Because of digital angiogram's typical high resolution, it is physically impossible to develop (or simulate) a multilayer neural network that classifies the entire angiogram at once. Therefore, we designed a network window classifier which classifies the center pixel of a relatively small area in the image. To extract the vascular structures, the window must slide, pixel by pixel, over the entire image. A multilayer feed-forward network is used as the window classifier. Its input units receive the gray-scale values from a square region of the angiogram. The input to the neural network is the raw digital data from cineangiogram without preprocessing. We avoided the use of feature extraction because (1) preprocessing merely alters the feature space and does not enhance the information content and (2) determining a set of meaningful features itself is a cumbersome task.

The generalized delta rule with a momentum [19] is selected as the learning paradigm for network training, because it shows several desirable properties as a classifier including its landmark success at generalization, complex decision boundary, and the fact that it makes no statistical assumption about the input. The network weights are trained over a set of points from a target list using the back-propagation algorithm. When the training converges and the system error decreases below an acceptable threshold, the window classifier is considered trained and then applied over the entire angiogram. The state of output units identifies the vessel from the background and creates the segmented image. Fig. 1 illustrates the architecture of the window classifier and how it is applied to the image.

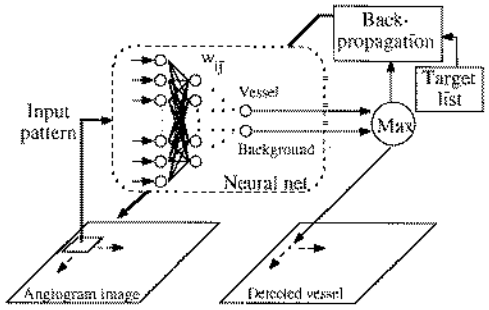


Fig. 1. Architecture of the neural-network window classifier.

IV. Analysis of Network Configuration

In contrast to the elegant mathematical basis for network training based on back-propagation, there is no general guideline for choosing the appropriate network configuration for a given problem. Therefore, we designed an experiment to address this issue. The configuration is systematically adjusted when the neural-network classifier is applied to the segmentation problem described above. The suitability of a particular configuration is judged on the basis of learning rate and classification performance. These results not only lead to the selection of the optimal network configuration but also give us an insight to the sensitivity of adjusting network configuration. The study of the network configuration is broken down into four categories: network topology, rate parameters, training sample set, and initial weights.

A. Network Topology

To investigate the role of network topology in this classification problem, we evaluate the performance of the window classifier using different network topologies. Let N be the total number of weights for a given network topology. The degree of freedom for the classifier increases as N increases and so does the computational complexity. To eliminate this variable in the comparison, we attempt to keep N constant for different topologies. All neural networks that we use belong to the feed-forward type. By definition, any neuron in a hidden layer must receive inputs from all neurons in the previous layer and feed all neurons in the next layer. Let n be the number of layers in the network and L_i be the number of neurons in the ith layer of the network. The total number of weights is given by

$$N = \sum_{i=1}^{n-1} L_i L_{i+1}. \tag{1}$$

The following facts should be noted. First, there are $(n+1)$ parameters available to configure the network topology, i.e., n and $L_i, i = 1, 2, \cdots, n$. Second, because N is determined by (1), it is impossible to keep the total number of weights exactly the same for different topologies. We can only configure a topology with N as close to a desirable number as possible. Third, the learning capacity (C) is related to N but a constant N does not guarantee a fixed learning capacity. The learning capacity is an indicator of how much a system can learn and its ability to generalize outside the training set [20], [21]. For single-layer networks, the learning capacity is twice the

TABLE I
NETWORK TOPOLOGIES AND PERFORMANCE

Network Topology	N	Performance on Training Sets	Performance on Testing Sets	Number of Training Iterations
121-17-2	2091	86.7%	92.2%	764
121-16-10-2	2116	86.7%	92.2%	963
121-14-24-2	2078	86.7%	92.2%	881
121-16-4-16-2	2096	84.0%	92.2%	1236
121-14-14-14-2	2114	86.7%	92.1%	1276
121-13-13-13-13-2	2106	85.3%	91.9%	3047

system's degree of freedom [22], i.e., $C = 2N$. For three-layer networks with h hidden nodes and n input nodes, the learning capacity is bounded by [23]

$$\frac{nh}{\log(nh^2)} < C < nh\log(h). \tag{2}$$

While an exact expression of C for networks with more than three layers has not yet been discovered, a crude estimate of the upper bound is on the order of $2N$.

It is computationally prohibitive to explore all possible topologies. To reduce the computational effort, some of the network parameters must be determined *a priori* and held constant. The degree of freedom (N) is set around 2100. The output layer contains two nodes (i.e., $L_n = 2$): one for vessel and one for background. The two output values are compared to determine the final classification. A preliminary experiment was performed with three-layer networks whose input window size varied from 5×5 to 15×15. As expected, the performance showed a direct relation to the average vessel size under investigation. For the angiograms under investigation, the average vessel width is about 9 pixels. Any window larger than 9×9 showed a reasonable performance. An 11×11 input window (i.e. $L_1 = 121$) was selected for the remaining experiments.

It has been shown [24] that the deep, shallow, and bottleneck network topologies present different characteristics in their classification performance. Thus, we alter the number of hidden layers and the numbers of nodes in the hidden layers in such a way that the three aforementioned topologies are represented. As shown in the left most column of Table I, six topologies with n ranging from 3 to 6 are investigated. Each of these networks are trained and their outputs are compared with the "correct" vascular structures delineated by a human operator.

B. Rate Parameters

The feed-forward network with back propagation learning has been commonly used in neural network computing. The network is trained by the delta rule with a momentum term which was introduced by Rumelhart [19]. At $(k+1)$th iteration during training, the update of network weight Δ_{ij} is based on two components. The first component is proportional to the effect of error signal δ_i on output O_j for the neuron sending the activation signal. The second component is proportional to the amount of weight change in the previous iteration (the momentum term).

$$\Delta_{ij}(k+1) = \beta(\delta_i O_j) + \alpha\Delta_{ij}(k). \tag{3}$$

The learning rate (β) controls the rate of convergence for the training process. The momentum term (α) helps prevent the oscillation problem near the solution point. These two parameters are the major factors that affect learning; they also control the tradeoff between the system's stability and classification quality. Although some heuristic-based algorithms exist for adapting these rate parameters [25], [26], an optimal adaptation scheme for the general case has not yet been found. In this study, the value for θ_o which controls the nonlinearity of the neuron's sigmoidal transformation is set to unity. Each of the node thresholds (θ_i) is trained as if it was the weight for a neuron which always outputs a constant value of one. When, the neuron's sigmoid function is given by

$$f(x_i) = \frac{1}{1 + e^{\frac{-(x_i+\theta_i)}{\theta_o}}}. \tag{4}$$

C. Training Sample Set

There is no theoretical guideline regarding the number of samples required for adequate learning which cover all possible patterns of vessel, background, and noise. In this study, we address the following questions based on empirical findings: (1) How many samples are sufficient for a satisfactory classification? (2) How should the samples be selected? In particular, the effects of random selection versus manual selection on the training and classification performance are studied.

D. Network Initial Weights

Although, learning by back-propagation has been successful for many applications, the possibility of being trapped in a local minimum of the error function during training does exist [27]. Thus, theoretically, an inappropriate choice of the initial weights could result in a failure in network training and/or poor classification performance. To examine the effect of initial weights on network training, each experiment is repeated five times with initial weights obtained from five different sets of random numbers.

V. SIMULATION RESULTS

The network window classifier shown in Fig. 1 was implemented on a VAX/VMS computer (Digital Equipment Cor-

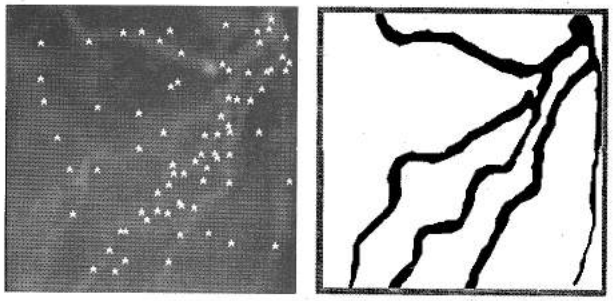


Fig. 2. Test image with overlaid training samples (left) and target image defined by human operator.

poration, Maynard, MA). In the preliminary study on the selection of training set, failures of vessel detection were observed for certain training sets. When a small set (25 samples) of randomly selected samples was used, the network was trained to ignore instead of detect the vascular structures. This was because the area occupied by the vessels was small in comparison to the background area in the angiogram. As a result, a randomly selected sample set usually contained very few samples (about 1 or 2) representing vessel and most of the training occurred over the background. Therefore, as the training proceeded, the network became biased towards background and tended to ignore vessels. To resolve this problem, the number of randomly selected training samples was progressively increased. As expected, performance increased with the number of samples. It soon became obvious that in order to achieve an adequate performance, a substantial number of training samples was required. This was especially true when the probability for selecting a vessel pixel over a background pixel was low (i.e., vessels were relatively thin). Thus, a manually selected training sample set with a roughly equal distribution of vessel and background pixels was used. The classification performance was significantly improved. Even with a training set as small as 25 samples, a satisfactory performance was observed. The training set that we finally decided to use for the remaining experiments contained 75 manually selected samples. In Fig. 2 (left) the angiogram (a DCA of human left anterior descending coronary artery) under investigation is shown with the 75 samples marked by asterisks. Fig. 2 (right) shows the target image in which the entire vascular structure, including the 75 sample pixels, was delineated by a human operator.

Based on the previously described training set, each network under investigation was either allowed up to 3500 representations of the training set or considered converged when normalized system error was less than 0.15. The system error (E) is defined by

$$E = \frac{1}{2 \times 75} \sum_{s=1}^{75} \sum_{p=1}^{2} (T_{ps} - O_{ps})^2 \tag{5}$$

where T_{ps} and O_{ps} are, respectively, the target and the network output patterns for sample s. Once the network has been trained, its classification performance (φ) over the entire image is measured by a pixel-to-pixel comparison between the network classification and the target image

$$\varphi = \frac{1}{2} + \frac{1}{2\mathrm{MN}} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} X_{ij} Y_{ij} \tag{6}$$

where X_{ij} and Y_{ij} are the target and network classification with values of 1 for the vessel and -1 for the background at (ij) location of the $M \times N$ image. Notice that this performance index is between 0 and 1, with 1 representing the best achievable performance.

A. Performance and Complexity

The learning rate (β) and momentum rate (α) were systematically adjusted for every topology in Table I. For each setting of topology and rate parameters, the network was trained and its performance was recorded. The results are summarized in Fig. 3 with each row representing a different topology. In the first column of Fig. 3, the complexity contours indicate how long it took for the network to converge for different (α, β) values. In the second column, the performance contours indicate the detection quality for different (α, β) values. In the third column, the rate parameters were fixed at ($\beta = 0.05, \alpha = 0.5$) and the system error was plotted versus the number of iterations. The results indicate that as the number of hidden layers increases, the convergence becomes slower and training difficulty appears. In the last column, the classification of the entire image by the trained network is shown. Notice that the six-layer network failed completely to classify the image with the given rate parameters (0.05). In this case, the learning momentum was too slow to converge in 3500 iterations. In fact, this network only converged at one test point, i.e., $\beta = 0.05$ and $\alpha = 0.9$. A larger learning rate becomes necessary as the number of hidden layers increases, however, a learning rate which is too large often results in oscillations. The best performance given by each topology was very close to one another. The only major difference was in the number of iterations (training time) required to achieve such performance. The 121-17-2 network, with the highest ratio of performance over learning time, is considered the best topology for this problem. The results also suggest that, to obtain fast learning and high performance, one should use a small learning rate and a medium momentum rate (e.g., $\beta = 0.05$ and $\alpha = 0.5$).

The initial weights had a negligible effect on the network training process. Fig. 4 illustrates an example of learning rate plots for five different sets of initial values applied to the 12-17-2 network. The classification performance (φ) was relatively insensitive to the setting of initial weights. In the worse case, the decrease in performance was observed on the order of 0.005.

B. Generalization

The generalization test determines how the trained network performs on data outside the training set. This test was conducted at two levels. First, the test image itself was used. Excluding the training samples, the test image contained new data of 60 441 pixels, i.e., $246 \times 246 - 75$. Only the central

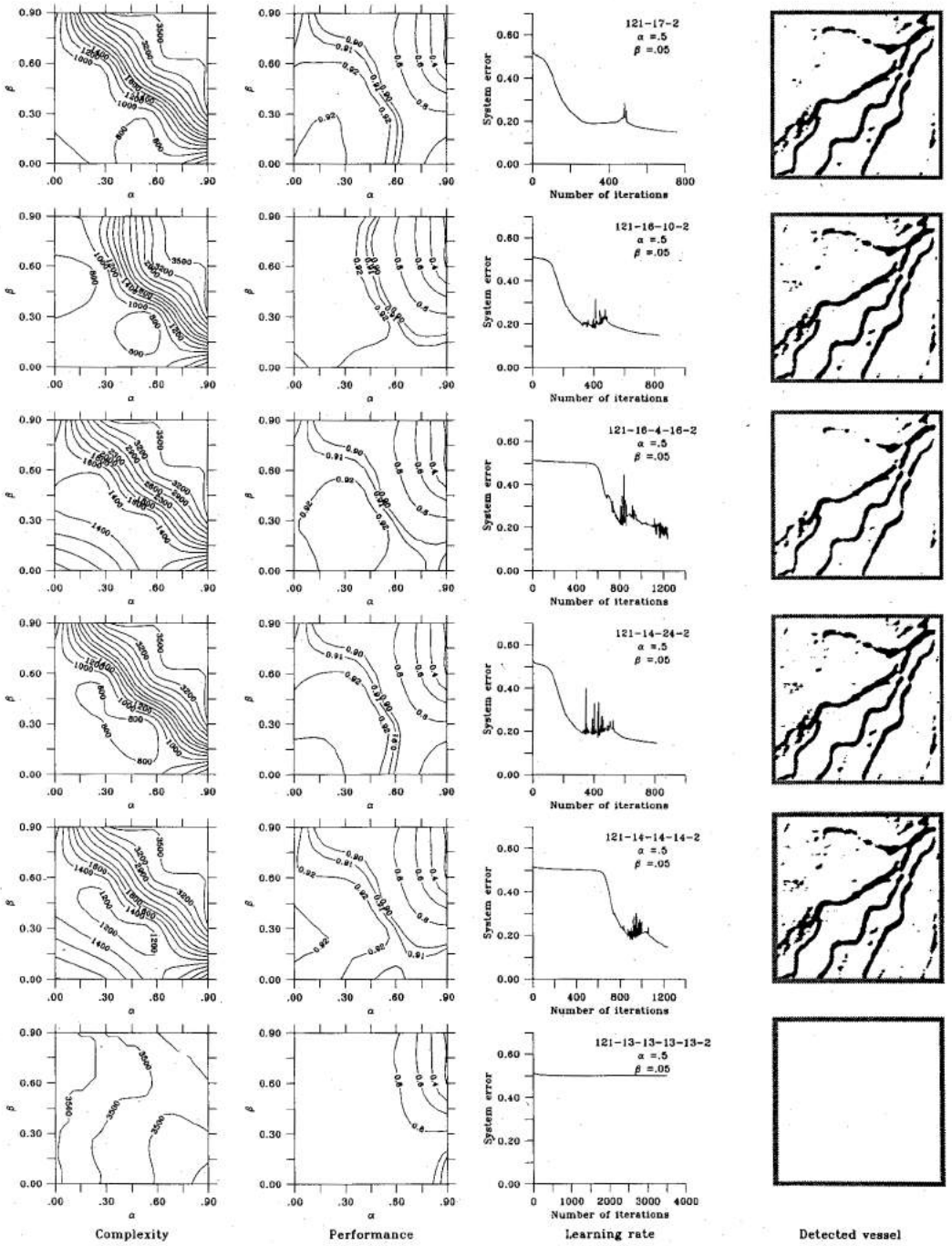


Fig. 3. Experimental evaluation of training and classification performance for different topologies and rate parameters.

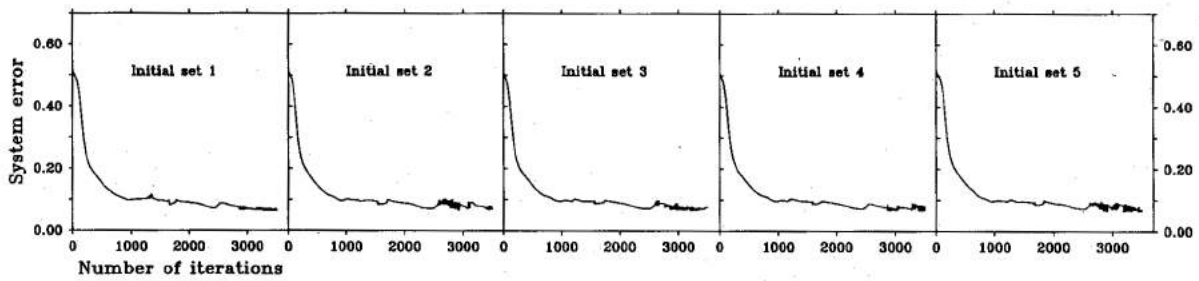


Fig. 4. Effects of different initial weights on training for the 121-17-2 network with $\alpha = \beta = 0.05$.

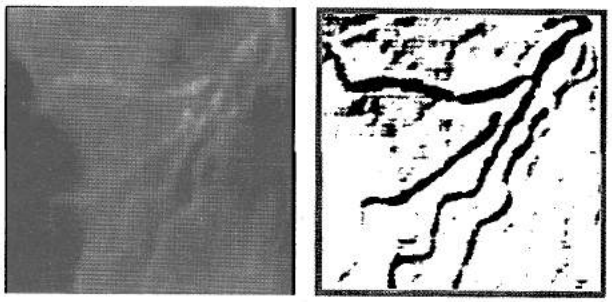


Fig. 5. A different frame from the same sequence of the DCA in Fig. 2 and the result of neural-network classification.

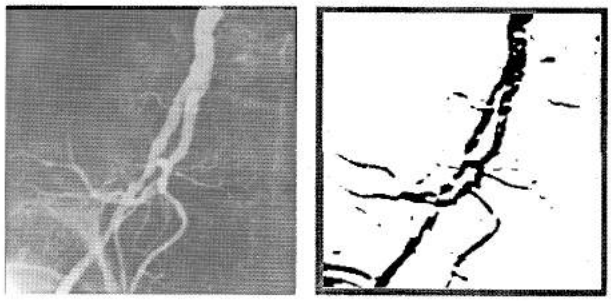


Fig. 6. A DVA of human right iliac artery and the result of neural-network classification.

246×246 pixels in the 256×256 image were examined because the border pixels were also excluded. The test results for each topology operating at its best performance are summarized in Table I. The performance on the original training set was also included for comparison. It is interesting to observe that performance on the test set was superior than that on the training set. This was because the training set was intentionally selected to include more cases which were difficult to classify. These cases corresponded to pixels near the edge of a vessel. On the other hand, the test set contained many background pixels which were easy to classify. It was also observed that sometimes a better performance was achieved on the test set by stopping the training process earlier to prevent the network becoming overspecific and "memorizing" the training samples. The networks trained to system errors below 0.05 adjusted their behavior more closely to what was depicted by the training set. Although they performed better on the training samples, they could perform rather poorly on the test set.

For the second level of generalization test, different types of angiograms were used. The 121-17-2 network trained with the 75 target points of the DCA image was applied to another DCA image. Next, this network was trained and applied to a DVA image and a DSA image. These results are shown, respectively, in Figs. 5–7. In these figures, the original angiogram is shown on the left with the neural-network classification result shown on the right.

C. Computational Time

The combined computational time for the entire study was on the order of 5000 VAX 11/780 CPU hours using several systems. The training time for each network using the 75 target points took between 2 and 10 CPU hours, depending on the number of iterations required to reach the specified system error. Once the window classifier was trained, the average time to slide it over a 256 × 256 image was 9 minutes.

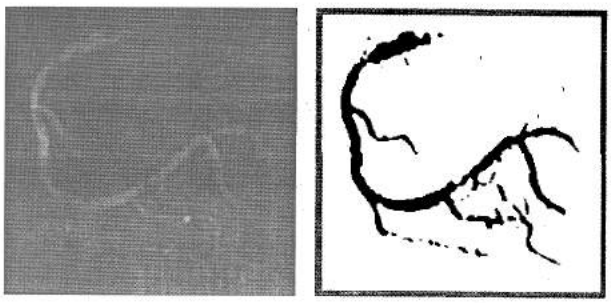


Fig. 7. A DSA of human right coronary artery and the result of neural-network classification.

VI. Comparison with Other Methods

It is important to assess the performance of the neural-network classifier relative to some other existing methods on this vessel-detection problem. In this section, we present a comparative study of classification performance. Two other algorithms are evaluated: One is based on the traditional Bayesian approach in which a maximum likelihood estimator is employed. The other is based on the iterative ternary classification developed by Kottke and Sun [10].

The maximum likelihood classifier (MLE) is derived by assuming that patterns in each of the M classes have a multivariate Gaussian distribution. The probability density function for pattern $\mathbf{x}$ in class ω_i is given by

$$P(\mathbf{x} \mid \omega_i) = \frac{1}{2\pi^{\frac{n}{2}}|\mathbf{C}_i|^{\frac{1}{2}}} e^{[-\frac{1}{2}(\mathbf{x}-\mathbf{m}_i)^T \mathbf{C}_i^{-1}(\mathbf{x}-\mathbf{m}_i)]}, \quad i = 1, 2, \cdots, M \tag{7}$$

where n is the dimension of the pattern vector. The covariance matrix ($\mathbf{C}_i$) and the mean vector ($\mathbf{m}_i$) of class ω_i are estimated from selected samples of the target image. A distance measure (D_i) is derived on the basis of the Bayesian maximum likelihood decision rule [28]:

$$D_i = P(\mathbf{x} \mid \omega_i) P(\omega_i). \tag{8}$$

The pixel is assigned to class ω_i for which distance D_i is the smallest. After simplification, we have

$$D_i = \ln P(\omega_i) - \frac{1}{2} \ln |\mathbf{C}_i| - \frac{1}{2}[(\mathbf{x} - \mathbf{m}_i)^T \mathbf{C}_i^{-1} (\mathbf{x} - \mathbf{m}_i)], \quad i = 1, 2, \cdots, M. \tag{9}$$

The above equation, with $M = 2$, was used to classify the vessel class and the background class in the test angiogram described earlier. In real-life situations, the *a priori* probabilities of vessel and background, $P(\omega_i)$, are unknown. Without this *a priori* knowledge, the classification result was quite poor. As shown in Fig. 8 (left), when we overestimated the *a priori* probability for vessel, a large background area was segmented as vessel. On the other hand, an underestimation of the *a priori* probability for vessel resulted in the missed detection of some vessel segments, as shown in Fig. 8 (right).

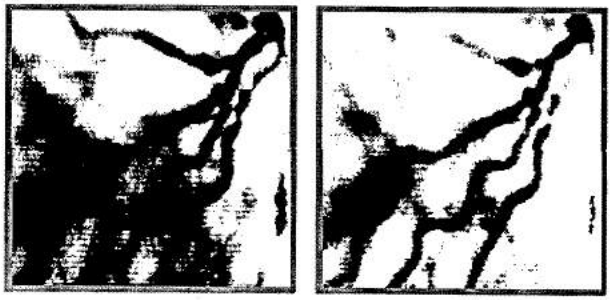


Fig. 8. Over and under classification by MLE.

TABLE II
COMPARISON OF CLASSIFICATION ACCURACY FOR THREE METHODS

Algorithm	Performance
Bayesian Maximum Likelihood	68%
Iterative Ternary Classification	83%
Back Propagation Network	92%

For the final comparison, we actually measured the *a priori* probabilities by counting the pixels in the two classes in the target image. This, of course, was unfair but insured the best performance from the MLE.

The iterative ternary classification (ITC) is an adaptive segmentation algorithm designed for classification of coronary arteries in angiograms. The algorithm iteratively adapts two gray-scale thresholds, T_a and T_b, to classify each pixel in the angiogram to one of three classes (artery, background, and undecided). It employs a learning algorithm in conjunction with line structure and a consistency measure from undecided pixels to update the two thresholds. This process iterates until the number of undecided pixels falls below a predetermined threshold.

In Table II, we present the classification accuracy (φ) for neural network, MLE, and ITC. The neural network showed the best performance with a 92% accuracy. In Fig. 9, the original test image and the classification results from neural network, MLE, and ITC are shown together for comparison.

VII. DISCUSSION

In this paper, we investigate the use of neural networks for segmentation of blood vessels in angiograms. It has been demonstrated that a window classifier based on a multilayer neural network outperforms the classic method based on maximum likelihood estimation and a modern method based on iterative ternary classification. The neural network classifier takes image data directly as input and, thus, avoids the cumbersome steps for feature selection and feature extraction. However, the configuration of the network is crucial for obtaining a satisfactory performance within an acceptable training time.

Specific findings from this study include the following: First, the network topologies under investigation, ranging from three-layer to six-layer, are all suitable for segmenting the angiograms and show about the same classification performance.

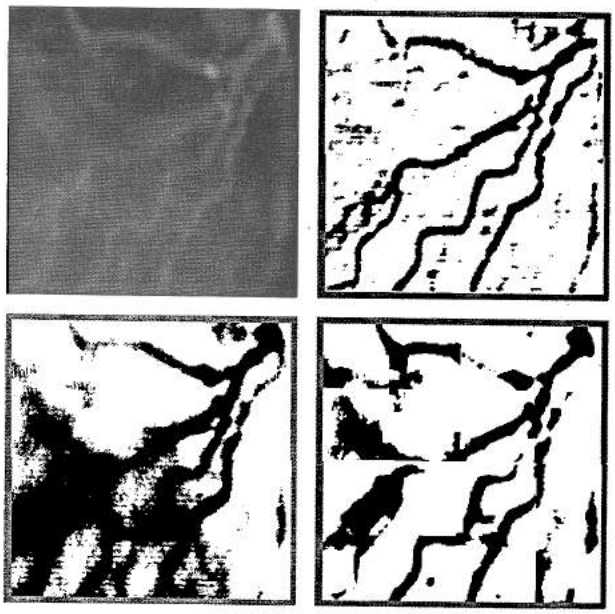


Fig. 9. Test angiogram (upper-left), classification by neural network (upper-right), classification by MLE (lower-left) and classification by ITC (lower-right).

Second, the three-layer network (121-17-2) generalizes well for all three types of angiograms (DCA, DVA, and DSA). Increasing the number of hidden layers only results in training difficulty and fails to gain in classification performance. Thus, we conclude that the three-layer network is sufficient for the segmentation problem. Third, classification performance is sensitive to the choice of training samples. Manual selection of those samples that best characterize the desirable performance can significantly reduce the necessary number of samples as compared to random selection. Fourth, the rate parameters for the back-propagation learning affect not only the learning rate but also the classification performance. The appropriate choice for our problem is a small learning rate ($\beta = 0.05$) combined with a medium momentum rate ($\alpha = 0.5$). Fifth, network training is not sensitive to the selection of the random initial weights. Sixth, the neural network shows superiority with a 92% classification accuracy, as compared to 83% given by the ITC algorithm and 68% given by the MLE algorithm.

The fact that the neural-network classifier achieves the highest performance over the other methods may contribute to its ability to form highly nonlinear decision boundaries for non-Gaussian distributed data. The excellent ability of generalization has also motivated us to examine the anatomy of the weight patterns further. In search of some insight to the true nature of the network and the basis behind its high performance, the weight patterns in the trained 121-17-2 network and their relations to the network architecture are shown in Fig. 10. The connection between the input and the first layer is illustrated as images of spatial filters, in which gray indicates a zero weight, dark gray indicates a positive weight, and light gray indicates a negative weight. A logarithmic scale has been used on these images for contrast enhancement. After a careful examination of the classifier's architecture, one can find its resemblance to a nonlinear

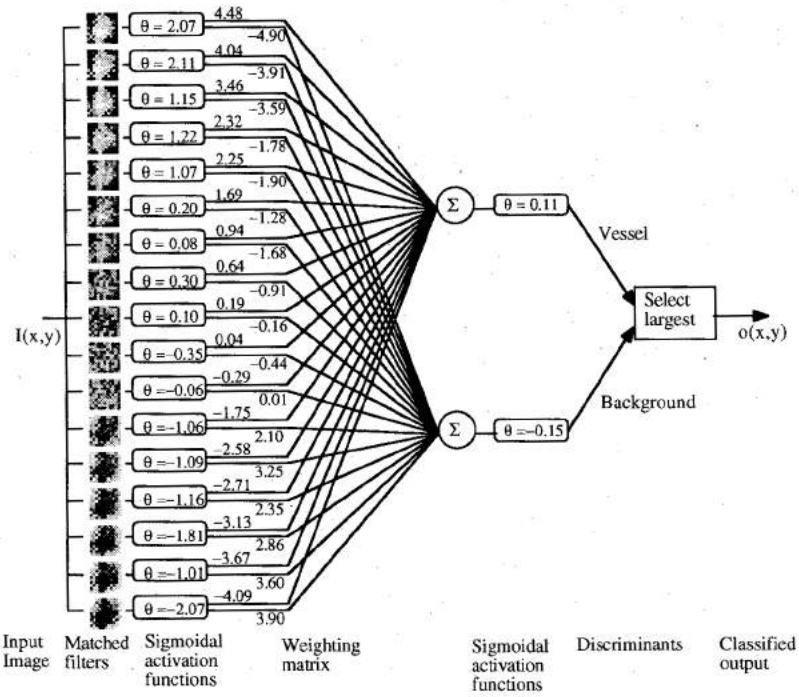


Fig. 10. Weight patterns and architecture for the 121-17-2 network.

matched filter. From this point of view, the neural-network classifier is in essence an automated and efficient way to implement a generalized matched filter. The matched filter contains two parts, a convolution operator followed by a nonlinear decision tree. The templates, which are the first-layer weight matrices determined by training, are convolved with the input image. The convolution results are processed by the decision tree to reach the final classification. The parameters in the decision tree are the second-layer weights and the threshold values in the sigmoidal functions, which are also determined by training.

A further examination of the weight images in Fig. 10, however, fails to reveal a simple geometric interpretation of how the vascular structure is extracted by the neural-network classifier. Each weight image is roughly symmetric with respect to the center, nevertheless, they do show a slight pattern along the direction from upper-right to lower-left. This may be explained by the fact that vessels in the training image (Fig. 2) lie predominantly in that direction. The top six weight images are roughly the complement of the bottom six. Thus, there seems to be a spatial differential operation, but the exact mechanism is difficult to trace due to the complex nonlinear decision tree.

This has been a dual-purpose study. The first purpose, which was to develop a neural-network classifier for the segmentation of angiograms, has been successfully achieved. The second purpose was to study the effect of network configuration on training and classification performance based on this image-segmentation problem. The results from this study has given us a valuable insight to the roles of network topology, rate parameters, training sample set, and initial weights. However, the pursuit for a general guideline to configure neural networks remains a challenge for on-going research.

REFERENCES

[1] D. H. Ballard, G. E. Hinton, and T. J. Sejnowski, "Parallel visual computation," *Nature*, vol. 306, pp. 21–26, 1983.

[2] I. Solheim, T. L. Payne, and R. Castain, "The potential in using back-propagation neural networks for facial verification systems," *Simulation*, vol. 58, pp. 306–310, 1992.

[3] P. Sajda and L. H. Finkel, "Simulating biological vision with hybrid neural networks," *Simulation*, vol. 59, pp. 47–55, 1992.

[4] J.-O. Eklundh, A. Lansner, and R. Wessblad, "Classification of multi-spectral images using associative nets," in *Proc. Eighth Int. Conf. Pattern Recognition*, Paris, France, 1986, pp. 1240&ndash ;1243.

[5] B. G. Brown, E. Bolson, M. Frimer, and H. T. Dodge, "Quantitative coronary arteriography: estimation of dimensions, hemodynamic resistance, and atheroma mass of coronary artery lesions using the arteriogram and digital computation," *Circulation*, vol. 55, pp. 329–337, 1977.

[6] P. Spiller, F. K. Schmiel, B. Politz, M. Block, U. Fremor, W. Hackbarth, I. Jehle, R. Korfer, and H. Pannek, "Measurement of systolic and diastolic flow rates in the coronary artery system by x-ray desitometry," *C irculation*, vol. 68, pp. 337–347, 1983.

[7] T. Sandor, A. D'Adamo, W. B. Hanlon, and J. R. Spears, "High precision quantitative angiography," *IEEE Trans. Med. Imaging*, vol. MI-6, pp. 258–265, 1987.

[8] W. A. Bateman and R. A. Kruger, "Blood flow measurement using digital angiography and parametric imaging," *Med. Phys.*, vol. 11, pp. 153–157, 1984.

[9] I. Liu and Y. Sun, "Fully Automated reconstruction of 3-D vascular tree structures from two orthogonal views using computational algorithms and production rules," *Optical Eng.*, vol. 31, pp. 21 97–2207, 1992.

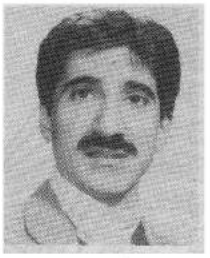
[10] D. Kottke and Y. Sun, "Segmentation of coronary arteriograms by iterative ternary classification," *IEEE Trans. Biomed. Eng.*, vol. 37, pp. 778–785, 1990.

[11] R. Nekovei and Y. Sun, "An adaptive algorithm for coronary artery identification in cineangiograms," In *Proc. IEEE Eng. Medicine and Biology Soc. 12th Ann. Int. Conf.*, Philadelphia, PA, 1990, pp. 1402–1404.

[12] T. V. Nguyen and J. Sklansky, "Computing the skeleton of coronary arteries in cineangiograms," *Computers and Biomed. Res.*, vol. 19, pp. 428–444, 1986.

[13] D. L. Parker, D. L. Pope, R. van Bree, and H. W. Marshall, "Three-dimensional reconstruction of moving arterial beds from digital subtraction angiography," *Computers and Biomed. Res.*, vol. 20, pp. 166–185, 1987.

[14] Y. Sun, "Automated identification of vessel contours in coronary arteriograms by an adaptive tracking algorithm," *IEEE Trans. Medical Imaging*, vol. 8, pp. 78–88, 1989.
[15] K. Hoffmann, K. Doi, S. Chen, and H. Chan, "Automated tracking and computer reproduction of vessels in DSA images," *Investigative Radiology*, vol. 25, pp. 1069–1075, 1990.
[16] T. Pappas and J. Lim, "A new method for estimation of coronary artery dimensions in angiograms," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, 1501–1513, 1988.
[17] P. H. Eichel, E. J. Delp, K. Koral, and A. J. Buda, "A method for a fully automatic definition of coronary arterial edges from cineangiograms," *IEEE Trans. Med. Imaging*, vol. 7, pp. 315– 320, 1988.
[18] R. Kruger, C. Mistretta, and A. Crummy, "Digital k-edge subtraction radiography," *Radiology*, vol. 125, pp. 243–245, 1977.
[19] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Exploration in the Microstructure of Cognition.* Cambridge, MA: MIT Press, 1986, ch. 8.
[20] V. N. Vapnik and A. Y. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory Prob. Appl.*, pp. 264–280, 1971.
[21] E. B. Baum and D. Haussler, "What size net gives valid generalization?," *Neural Computation*, pp. 151–160, 1989.
[22] T. M. Cover, "Geometrical and statistical properties of systems of linear inequalities with application in pattern recognition," *IEEE Trans. Electron. Comput.*, pp. 326–334, 1965.
[23] S. Akaho and S. Amari, "On the capacity of three-layer networks," in *IEEE-INNS Int. Joint Conf. Neural Net.*, 1990, vol. 3, pp. 1–6.
[24] J. K. Kruschke, "Improving generalization in back-propagation networks with distributed bottlenecks," in *IEEE-INNS Int. Joint Conf. Neural Net.*, 1989, vol. 1, pp. 443–447.
[25] R. A. Jacobs, "Increasing rates of convergence through learning rate adaptation," *Neural Net.*, vol. 1, no. 4, pp. 295–308, 1988.
[26] J. P. Cater, "Successfully using peak learning rates of 10 (and greater) in back-propagation networks with the heuristic learning algorithm," in *IEEE-INNS Int. Joint Conf. Neural Net.*, 1987, v ol. 2, pp. 645–651.
[27] M. Gori and A. Tesi, "On the problem of local maxima in backpropagation," *IEEE Trans. Pattern Recogn. Mach. Intell.*, vol. 14, pp. 76–86, 1992.
[28] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*. Reading, MA: Addison-Wesley, 1974, ch. 4.



Reza Nekovei (S'84) received the B. S. and M. S. degrees, both in electrical engineering, from University of Maine in 1986 and 1988, respectively. He is currently a Ph.D. candidate in the Electrical Engineering Department of the University of Rhode Island.

Since 1988 he has been with the Remote Sensing Laboratory at University of Rhode Island, where he has been involved in satellite oceanographic research. His research interests include adaptive signal processing, pattern recognition, and medical imaging.

Mr. Nekovei is a member of Eta Kappa Nu.

Ying Sun received the B. S. degree from the National Taiwan University in 1978, the M. S. degree from the University of Rhode Island in 1982, and the Ph. D. degree from the Worcester Polytechnic Institute in 1985, all in electrical engineering.

He joined the Department of Electrical Engineering at the University of Rhode Island as an Assistant Professor in 1985, and has been an Associate Professor since 1990. During 1992–93 he was on sabbatical leave, visiting at Linköping University in Sweden and University of Vienna in Austria. His research in medical imaging, microprocessor-based instrumentation, and modeling of the cardiovascular system has been supported by industry and by the National Science Foundation. He has published over 80 articles to date in professional conferences and in a variety of journals such as IEEE TRANSACTIONS (BIOMEDICAL ENGINEERING, MEDICAL IMAGING, PATTERN ANALYSIS AND MACHINE INTELLIGENCE), *Optical Engineering*, and the *American Journal of Physiology*. He has also served as a consultant to hospitals and the medical device industry for over ten years.