Neural Networks and Prior Knowledge Help the Segmentation of Medical Images

Guido Valli [1], Riccardo Poli[2], Stefano Cagnoni [1*] and Giuseppe Coppini [3]

[1] Department of Electronic Engineering, University of Florence, Italy
[2] School of Computer Science, The University of Birmingham, UK
[3] CNR Institute of Clinical Physiology, Pisa, Italy

This paper describes some achievements in the segmentation of medical images using artificial neural networks. We have identified three main sources of *a priori* information available to help perform the task of medical image segmentation: anatomical knowledge about the imaged region, the physical principles of image generation and the "regularities" of biological structures. The exploitation of each of these forms of knowledge can be effectively achieved with suitable neural architectures, three of which are described in the paper. An important lesson learnt from using these architectures is that different kinds of knowledge unavoidably induce different limitations in the resulting segmentation systems, either in terms of generality or of performance. Our experience indicates that in several applications some of such limitations can be overcome through a careful exploitation and integration of available knowledge sources via proper neural modules.

Keywords: Medical images, Computer vision, Image segmentation, Artificial neural networks, Knowledge representation

Introduction

In computer vision, segmentation is a crucial step in building systems for understanding the imaged "world". In the case of medical images, the general objective of segmentation is to find regions which represent single anatomical structures (both normal and pathological). For example, the availability of regions which represent single structures makes tasks such as interactive visualization and automatic measurement of clinical parameters directly feasible. In addition, segmented images can be further processed with computer vision techniques [Ballard & Brown, 1982] to perform higher-level tasks such as shape analysis and comparison, recognition and clinical decision-making.

Unfortunately, segmenting medical images is a very challenging task for the following two reasons. First, standard computer vision techniques cannot always be applied satisfactorily to the segmentation of medical images because the physics of "natural-scene" image generation on which such techniques rely is quite different from the physics of medical image generation. Second, medical images have a number of unusual features [Macovski, 1983, Webb, 1991], such as high noise intensity, the presence of semi-transparent structures, biological shape variability, tissue inhomogeneity, imaging-chain anisotropy and variability, which severely hamper their segmentation.

In order to overcome these problems most researchers have adopted the strategy of exploiting different kinds of *a priori* information about the imaged structures. However, segmentation systems based on conventional algorithmic techniques or on symbolic knowledge-representation and processing have often shown a limited robustness and, in most cases, have required considerable efforts for eliciting knowledge.

Artificial Neural Networks (ANNs) can partially overcome these drawbacks thanks to the following properties: capability to learn from examples and to generalize what has been learnt

* Stefano Cagnoni is now with the Department of Computer Engineering, University of Parma, Italy

(as in the case of feed-forward nets), noise rejection, fault tolerance, optimum-seeking behavior (which is typical of most recurrent nets) [Rumelhart & McClelland, 1986, Kohonen, 1989, Hopfield & Tank, 1985].

In our research in the area of medical imaging, we have explored the features of ANNs to help improve the performance and reduce the development time of image segmentation systems. Several neural architectures for medical-image segmentation have been developed, which exploit different kinds of *a priori* information. Available knowledge sources are: the anatomy of the imaged region and the structure(s) of interest inside it, the physics on which the adopted imaging modality is based, and the typical "regularities" of biological structures [Marr, 1982, Reuman & Hoffman, 1986]. In the following section we discuss how these forms of knowledge can be used to drive the segmentation of medical images.

In the third section we describe three architectures, which exploit such sources of knowledge, and give some examples of the results they produce. The first architecture exploits anatomical knowledge as a source of *a priori* information to train a set of feed-forward neural modules operating at different resolutions. Knowledge about the physics underlying the imaging technique is exploited by the second feed-forward neural system which uses a tissue classification strategy based both on statistical properties of Magnetic Resonance (MR) multi-spectral images and on anatomical knowledge. A strategy based on the exploitation of visual "regularities" derived from properties of natural structure has been used in the third neural architecture based on Hopfield's neural networks.

In the fourth section of this paper, we summarize the properties of these architectures. As will be shown, different kinds of knowledge unavoidably induce different limitations in the resulting segmentation systems either in terms of generality or of performance. Our experience indicates that in several applications some of such limitations might be overcome through a careful exploitation and integration of different knowledge sources via different neural modules.

Knowledge Sources for Medical Image Segmentation

Anatomical Knowledge

Anatomical knowledge about the typical shape and appearance of the structures being imaged is the most obvious, and widely used, source of *a priori* information on medical images. However, the use of anatomical knowledge is not easy and requires the solution of two difficult problems.

First, in order to exploit anatomical knowledge for medical-image segmentation it is necessary to define adequate models that represent the structures being considered. Such models should combine the appropriate level of detail with a representation as invariant to changes of scale, translation, rotation and deformation as possible.

Second, the input data must be correctly matched against an internal model which is used to produce *a priori* expectations about the image content.

Symbolic systems based on anatomical knowledge that were proposed in the past by several authors [Stansfield, 1986, Raya, 1990] have partially solved these problems. However, in general, those systems have not achieved fully satisfactory results. As recognized by most authors (for example, see [Sonka *et al.*, 1996]) this is mainly due to the intrinsic complexity and variability of biological objects which heavily hamper the elicitation and the use of knowledge. For this reason, alternative approaches to the exploitation of anatomical knowledge have been largely investigated (e.g. see [Bajcsy & Kovacic, 1989]).

Physics of Image Generation

Some imaging techniques, such as MR, allow for reproducible measurements of parameters which characterize different tissues. Such measurements can be exploited in developing classification algorithms for image segmentation.

For example, MR spin-echo image sequences are generated by a physical process that can

be summarized by the following approximate equation [Webb, 1991]:

$$I = k(\rho, T_1, TR)\, e^{-\frac{TE}{T_2}}$$

where I is the measured signal intensity in one pixel, ρ is the proton density, T_1 and T_2 are the relaxation times, TE is the echo time, TR is the excitation-sequence repetition time and k is a coefficient which depends on tissue and acquisition parameters. The parameters ρ, T_1 and T_2 are tissue-dependent, while TE and TR depend on the acquisition sequence and can be set by the operator. The above equation shows that the characteristics of spin echo images strongly depend on T_2 and that contrast between different tissues can be adjusted with a suitable choice of TE. Given a set of MR images of the same slice acquired with different TEs (multi-echo sequence), T_2 can be estimated for each pixel [MacFall *et al.*, 1986], thus allowing for T_2-based tissue classification.

Several authors have reported encouraging results obtained on the segmentation of MR images of the brain using statistical, fuzzy, and neural-network approaches based on this idea [Özkan *et al.*, 1990, Gerig *et al.*, 1992, Hall *et al.*, 1992]. However, a critical problem to be tackled is the presence of different tissues of similar appearance, which hampers classification based only on MR physics. In the case of brain tissues, misclassifications mainly affect sub-cutaneous fat and white matter, as reported in [Piraino *et al.*, 1991] and as we verified in our early experiments. In theory, this problem can be solved by integrating knowledge about the physical principles of the imaging device with some other form of *a priori* knowledge about the imaged structures [Lundervold & Storvik, 1995, Wells III *et al.*, 1996].

Perceptual Regularities

Biological vision is ruled by principles such as perceptual grouping, selection, discrimination, etc. which mostly depend on regularities of nature such as cohesiveness of matter or existence of bounding surfaces. These properties are certainly valid also for the anatomical structures contained in medical images, and can be exploited to build general-purpose segmentation systems for that kind of images.

From the perceptual standpoint, the optimum segmentation algorithm for medical images should be sensitive to small-size and low-contrast structures (high discriminating power), and robust with respect to noise, texture and slow intensity-changes (high grouping power).

These requirements counteract each other and a trade-off solution is necessary. The trade-off can be set at design-time and embedded in a segmentation algorithm, as it is often done in standard computer vision algorithms [Ballard & Brown, 1982], or can be optimized for each image so as to obtain maximum performance in terms of grouping and discriminating power. In this case the problem of medical image segmentation can be formulated as a problem of combinatorial optimization. This will be clarified in the following section.

Neural Architectures for Medical Image Segmentation

Segmentation Based on Anatomical Knowledge

Feed-forward ANNs can naturally integrate anatomical knowledge with the information contained in the images without requiring the formulation of explicit descriptions of objects. In fact, the output of a trained neural network relies both on the input data and on the *a priori* expectations that have been stored in the network connections during the learning phase. Thanks to this property ANNs can effectively face the problems encountered in knowledge-based segmentation of medical images.

On the basis of these ideas we developed a system (based on feed forward–networks trained with the back–propagation algorithm [Rumelhart & McClelland, 1986]) for the segmentation of target structures in tomographic images, and of lung nodules in standard projection radiography.

The system consists of a set of basic modules (one for each kind of structure to be segmented) such as the one shown in Fig. 1. Each module includes three major blocks: a retina, an Attention Focuser (AF) and a Region Finder (RF). The retina is the input section of the system. It

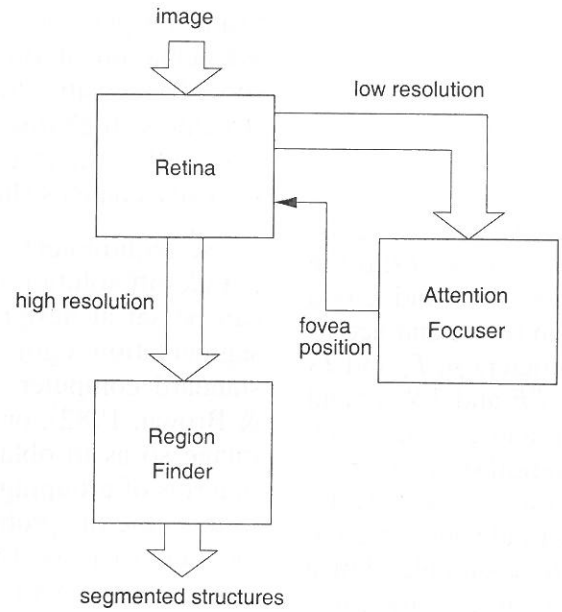

Fig. 1. Basic-module architecture of a segmentation system based on anatomical knowledge.

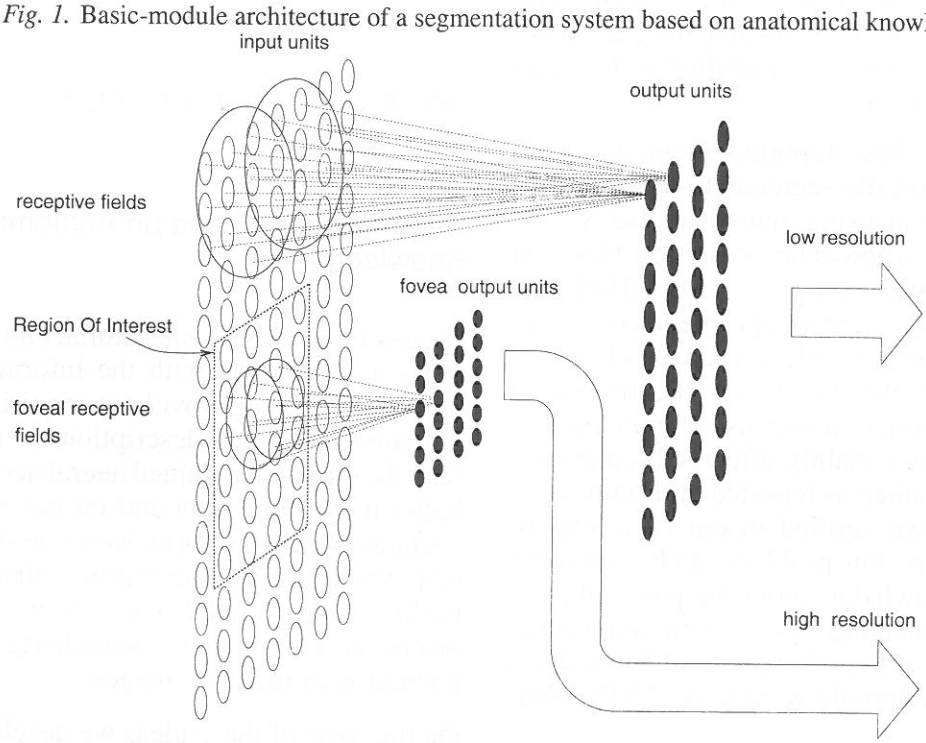

Fig. 2. Structure of the retina.

preprocesses the input image to produce a low-resolution output picture, utilized by AF to locate the desired structure, and a high-resolution output picture, used for segmentation by RF.

The retina is composed of an input layer including as many neurons as image pixels, and an output layer with a reduced number of neurons (see Fig. 2). As in biological retinas, the connections between the input and the output neurons are local and are arranged in overlapping receptive fields centered on each output neuron. Let $w(x, y)$ be the connection weight between a given output unit and the input unit at position (x, y). For MR and Computed Tomography (CT) images we use receptive fields with Gaussian weights:

$$w(x, y) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{x^2 + y^2}{2\sigma^2}\right]$$

For X-ray images we utilize Laplacian of Gaus-

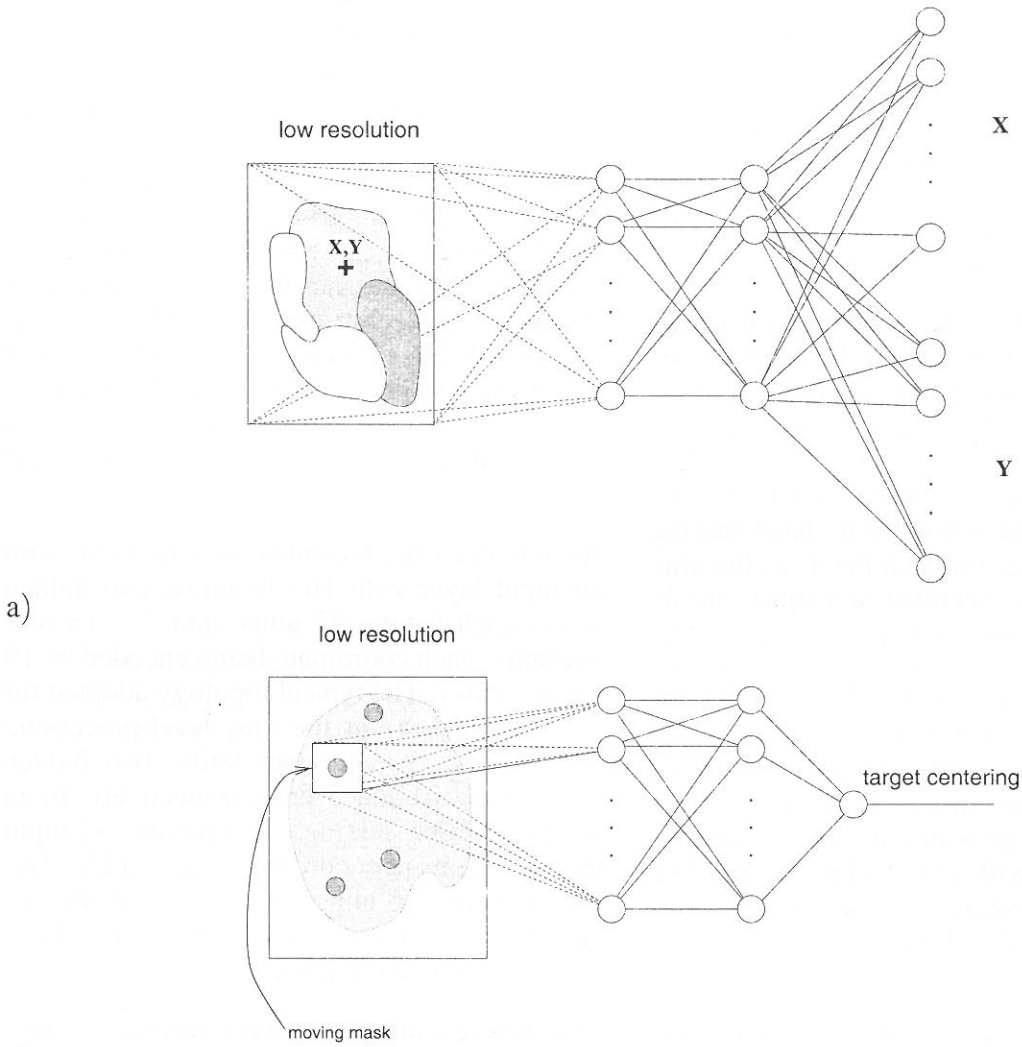

Fig. 3. Two alternative implementations of the Attention Focuser module.

sian (LoG) weights:

$$w(x,y)=\frac{1}{4\pi\sigma^4}\left[2-\frac{x^2+y^2}{\sigma^2}\right]\exp\left[-\frac{x^2+y^2}{2\sigma^2}\right]$$

The shape of both types of receptive fields is tuned through the σ parameter. Gaussian-shaped receptive fields allow for the smoothing of input images with minimal distortion in both space and frequency domain [Marr, 1982]. On the other hand, the LoG-weighted receptive fields are useful to filter out the low-frequency components of X-ray images which are primarily responsible for background variability. The retina also includes a moving region, called the fovea, which performs the same operation at a higher spatial resolution on a Region of Interest (ROI) selected by AF.

As to AF, we have designed two different structures illustrated in Fig. 3. When a single entity has to be segmented, AF is a fully connected net with: a) as many input neurons as the number of pixels of the low-resolution image produced by the retina, b) two hidden layers and c) an output layer which encodes the coordinates of the centroid of the structure under consideration (Fig. 3a). When multiple instances of the structure of interest may be present, AF net has: a) an input layer arranged as a square mask which scans the image, b) two hidden layers, and c) one output neuron whose output is high when a target structure is centered in the input mask (Fig. 3b). The network scans the low-resolution picture like a convolution operator. The points where AF output is high (with respect to a predefined threshold) are marked as *attention foci* and indicate possible lesions.

It is worth noting that the use of an attention-focusing mechanism has two main advantages: a) the overall computation load is reduced (only the ROI is processed at high spatial resolution), b) the produced segmentation is insensitive to translation. Moreover, the use of overlapping receptive fields provides a certain degree of tolerance to shape distortion [Hubel, 1988].

Once the centroid of the structure considered has been computed by AF, the fovea processes the ROI at the higher spatial resolution and RF extracts the pixels belonging to the structure of interest. The topology of RF is depicted in Fig. 4. RF is a block-connected multi-layer net which operates like a nonlinear space-varying filter by processing, for each pixel: a) the gray level of the pixels contained in a square mask, b) the position of the mask (given by an appropriate encoding of the coordinates of its central pixel). As shown in the figure, these different kinds of information are processed independently by two different fully connected sub-networks joined in a common layer. This network topology provides an optimal integration of input data with *a priori* knowledge. The activation of the output unit indicates whether the processed pixel belongs to the considered structure.

The architecture described has been used to segment MR and CT images [Coppini *et al.*, 1992] and to detect lung nodules, which appear as small low-contrast blobs in standard chest radiographs [Coppini *et al.*, 1993b].

Tomographic Images

In the experiments with tomographic images (represented by 256×256 matrices), we have considered the segmentation of a) the brain from MR head slices, and b) the spinal column from CT thorax scans. In both cases, the retina produces 16×16 Gaussian filtered low-resolution images (typically we used $\sigma = 4.5$). A fovea region with 256×256 input units and 128×128 output units and $\sigma = 1.5$ was used for brain images. For spinal-column images, the fovea has 80×80 input units, 40×40 output units and $\sigma = 1.5$.

AF has been implemented as a network with an input layer with 16×16 units, two hidden layers each having 32 units, and 2×16 output units (each coordinate being encoded by 16 output units). The typical topology adopted for RF includes: a) in the gray-level-processing sub-network, 9×9 input units, two hidden layers with 10 and 5 units respectively, b) in the position-processing sub-network, 32 input units (16 units per coordinate), two hidden layers each with 16 units. The output of the two sub-networks converge into a three-units layer which, in turn, feeds a single output unit.

Two different image-sets were utilized: the first one composed of 80 MR brain images from 6 patients, the second one including 120 CT thorax images from 8 patients. An expert radiologist labelled all the images of both sets by providing

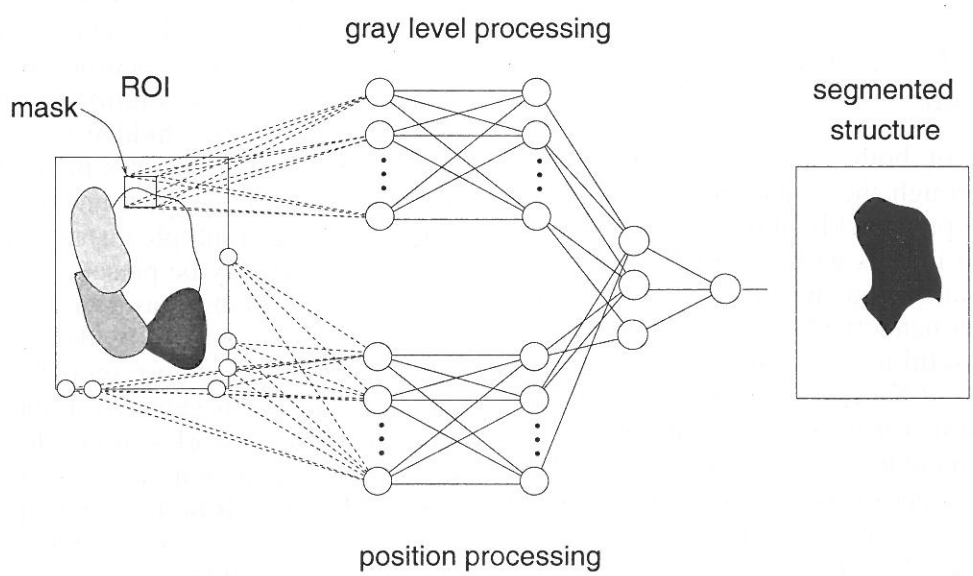

Fig. 4. Region Finder topology.

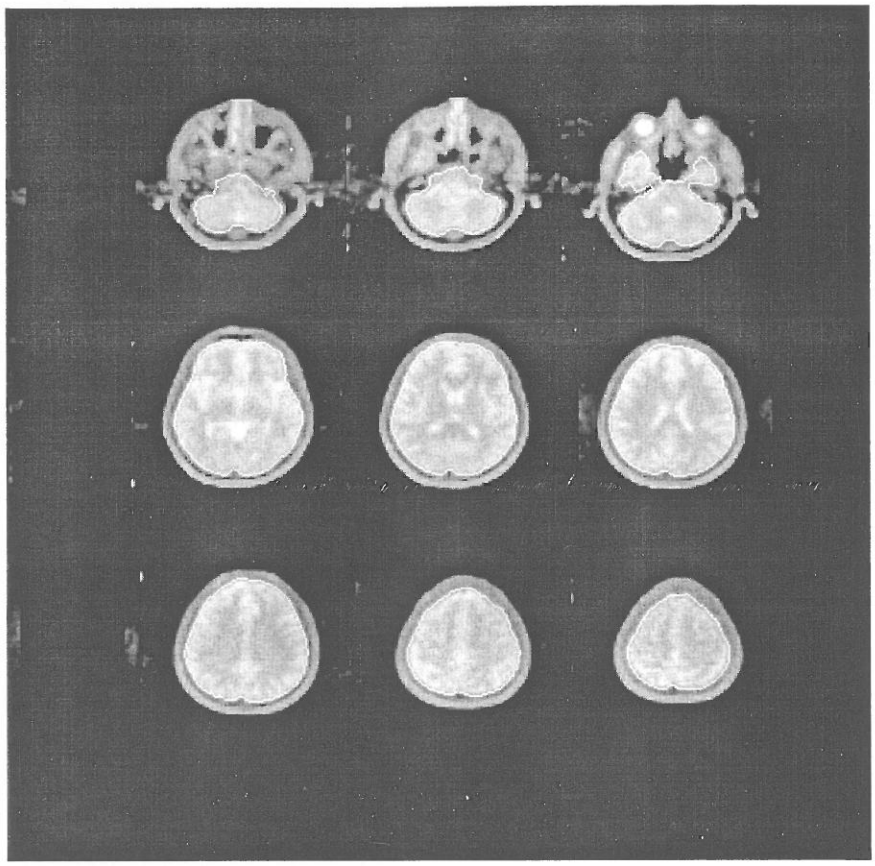

Fig. 5. Segmentation of the brain from an MR image sequence using the architecture presented in Fig. 1.

the position of the considered organ and tracing the related boundary using a graphic interface. Afterwards, we utilized half of the images in each set to train a corresponding segmentation system. The remaining halves were used as test sets. In all cases, the rate of correct classifications with respect to radiologist's segmentation was above 95%, with remarkable sensitivity and specificity. In the case of brain slices, the rate of correct pixel-classification was 97%, sensitivity being about 95% and specificity 98%. In the case of spine images, we observed a rate of correct classifications of about 96% (sensitivity 95%, and specificity 97%).

In Fig. 5 we show a sequence of MR tomograms of the head: the contours of the brain as segmented by the system are superimposed on the original images.

Chest Radiographs

In the experiments with chest radiographs (each represented by a 768×768 matrix), the retina produces LoG filtered low-resolution 256×256 images, with $\sigma = 8$. The fovea operates at full resolution and has 60×60 output units, with $\sigma = 2.5$. AF has been implemented as a network with 19×19 input units, two hidden layers each having 32 units, and a single output unit.

RF analyzes the ROIs around the *attention foci* produced by AF and, for each of them, produces a binary output which represents the segmentation of nodule-like patterns. The grey-level-processing sub-network includes 19×19 input units, three hidden layers with 10, 6 and 3 units, while the position-processing sub-network has 12 input units (6 per coordinate), and three hidden layers with 6, 6, and 4 units respectively. As in the case of tomographic images, the two sub-networks share the final layers including 3 units and a single output neuron. The segmentation can then be analyzed by a neural recognition system (a three-layer fully-connected network with 21×21 input units, two hidden layers with 4 and 2 units, respectively, and 1 output unit), which labels each region as nodule-like or normal.

The data set used to train and test the system includes 62 standard antero-posterior chest radio-

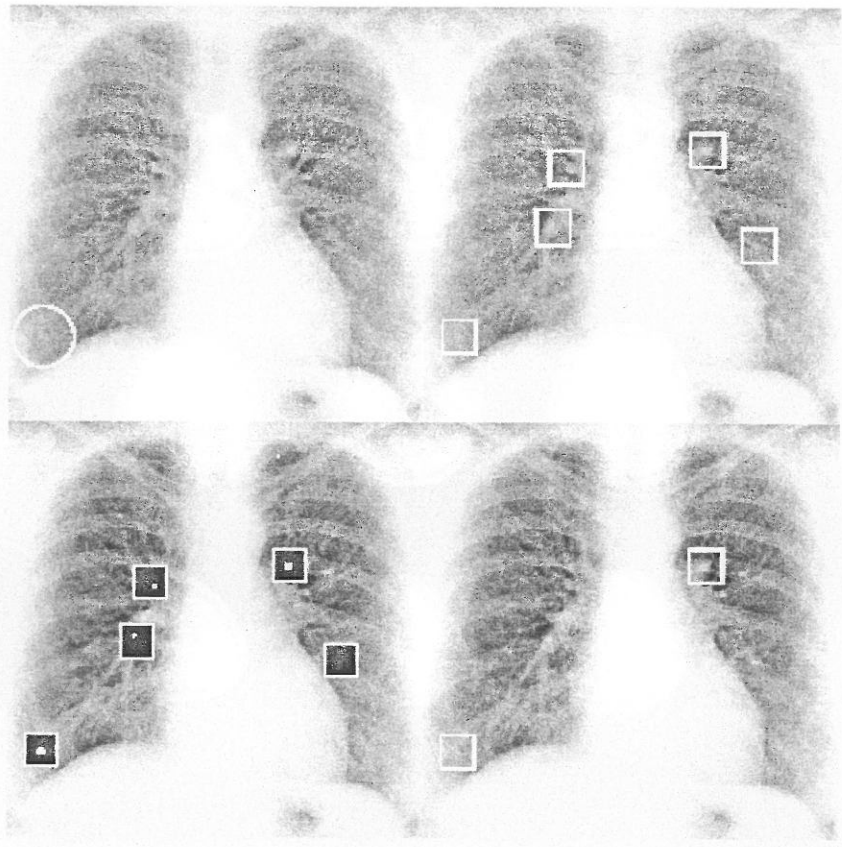

Fig. 6. Steps of the segmentation of lung nodules exploiting anatomical knowledge: original image analysed by a radiologist (*top left*); output of the Attention Focuser (*top right*); regions segmented by the Region Finder (*bottom right*); structures classified as nodular (*bottom right*).

grams. All images were examined by an expert radiologist who identified and manually segmented nodular lesions (having an approximate size from 3 to 30 mm, corresponding to a range from 5 to 55 pixels) . A sub-set of 32 images (with 92 nodules) was adopted as training set. The remaining radiograms were used as a test set, 6 of them coming from normal subjects and 16 with lesions (for a total of 18 nodules). Experimental results indicate good sensitivity and reasonable specificity in detecting parenchyma lesions. In particular, as to the segmentation phase, the rate of correct pixel-classifications has been above 93%, with a sensitivity of about 94% and a specificity of 96%. As concerns the global system performance, all the nodules were correctly identified with no false negative and a total number of 7 false positives. In the images from normal patients we have observed 4 false alarms.

In Fig. 6 we illustrate the typical operation of this system. The four panels show (top to bottom, left to right): a radiogram with a malignant nodule encircled by the radiologist, the *attention foci* produced by AF, the regions segmented by RF and the structures classified as pathological.

Combined Use of Physics of Image Generation and Anatomical Knowledge

In the previous section we pointed out that the physics of image generation is an important source of prior knowledge. Even if its sole use can lead to poor segmentation results, its use combined with anatomical knowledge can provide more accurate results [Sonka *et al.*, 1996, Lundervold & Storvik, 1995]. We explored this idea in a neural system for the segmentation of MR spin echo images of the brain [Cagnoni *et al.*, 1993] where the ambiguity between fat and white matter was removed thanks to the use of anatomical knowledge.

The system includes three main modules (see Fig. 7) which accomplish the following computational processes: a) enhancement and analysis of the information provided by signal decay over time; b) detection of the brain parenchyma; c)

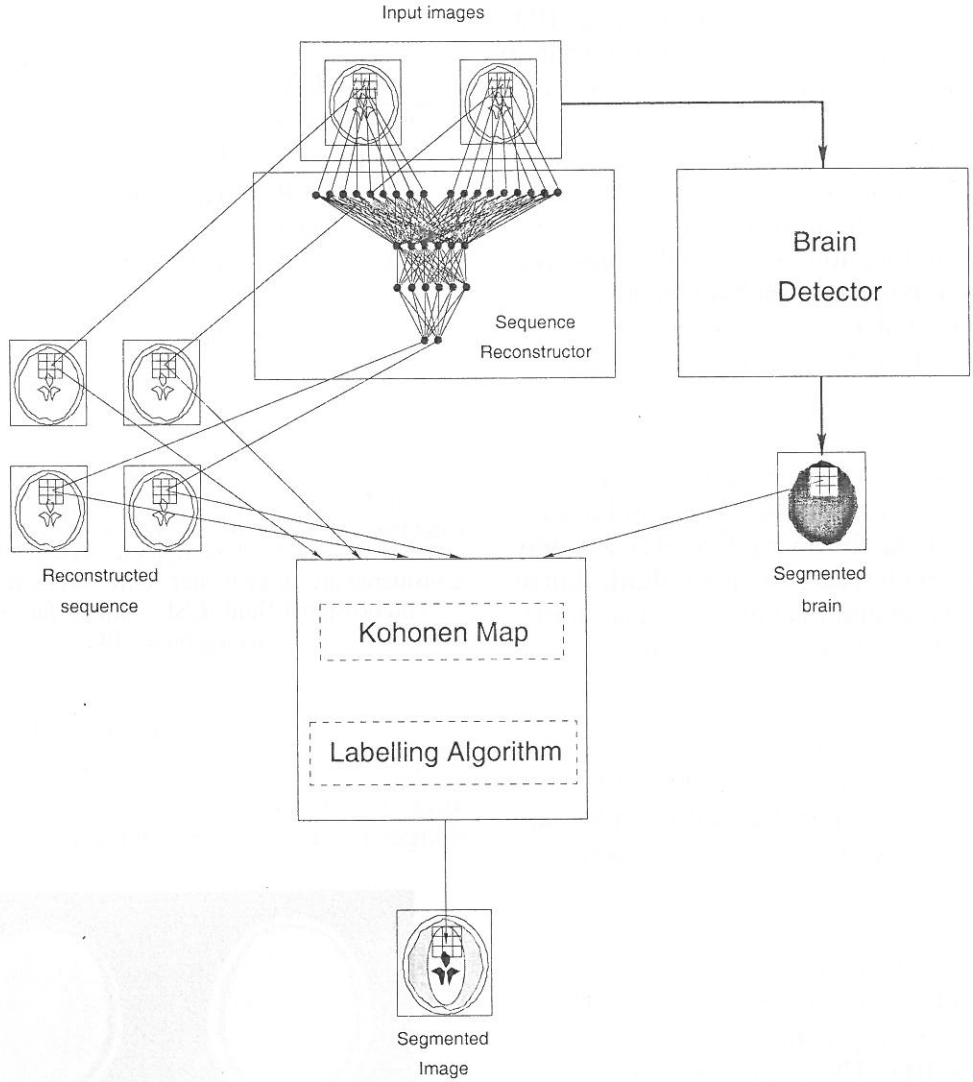

Fig. 7. Overall structure of a segmentation system that uses anatomical knowledge and the physics of image generation.

pixel classification into five predefined tissue groups. The first two modules are based on feed-forward networks trained with the back-propagation algorithm. The neural classifier which performs the actual segmentation is implemented with a Kohonen topology-preserving map [Kohonen, 1989], whose units have been labeled after training according to the class of patterns to which they respond maximally.

The first module, that we have called Sequence Reconstructor (SR), enhances information related to signal decay and improves the signal-to-noise ratio of the image, thus emphasizing the differences among different tissues [Cagnoni *et al.*, 1992]. SR generates two low-noise long-TE (150, 200 ms) images, which could not be reliably obtained otherwise (e.g. via conventional extrapolation techniques or direct acquisition). The neural network that performs such a task consists of four layers with 18, 6, 6, and 2 units. The inputs are the gray levels of two 3×3 windows taken in the same position from two short-TE (50, 100 ms) images of a given slice. The outputs represent the estimated intensity values of the central pixel of such windows in two images with long TEs. Therefore, by successively processing pairs of 3×3 windows centered on each pixel of the input images, the network synthesizes whole long-TE images. The four-image sequence produced by this network is considerably less noisy than the one obtainable via direct acquisition. This can be justified by considering that the number of degrees of freedom of the network is much smaller than the number of pixels used to train it, so that the network cannot (over)fit the noise.

The second module, termed Brain Detector (BD), supplies the neural classifier with *a priori* anatomical knowledge. The architecture of BD is described in the previous subsection. BD has been trained to produce an image in which pixels belonging to brain parenchyma are enhanced. As pointed out in the introduction, knowledge about the brain location is essential to discriminate between the tissues (subcutaneous fat and white matter) that have similar ρ, T_1 and T_2, and therefore have very similar gray levels in all images of the sequence.

The outputs of SR and BD are processed by a neural classifier. This module performs the segmentation of the sequence by assigning each pixel to one of the following five classes: gray matter, white matter, cerebrospinal fluid, skin or sub-cutaneous fat and background. The implemented classifier is based on a one-dimensional 256-unit Kohonen self-organizing map used jointly with a unit labeling algorithm [Wolpert, 1992]. This paradigm has the advantage of faster training with respect to standard back-propagation and of reliable encoding of the statistical properties of the training set [Kohonen, 1995].

A total of 2500 examples were taken from 25 images coming from three spin-echo multislice sequences (TE = 50, 100 ms, TR = 2000 ms) and from the corresponding segmented image, generated by BD. These examples were classified by an expert radiologist. Half of them (50 examples per image) were used to build the training set. The other 1250 were used to test the system.

Table 1 shows the confusion matrix calculated on the test set. In the matrix, the elements C_{ij} with $i = j$ indicate the percentage of patterns belonging to class i that have been correctly classified; elements C_{ij} with $i \neq j$ indicate the percentage of patterns belonging to class i that have been misclassified as belonging to class j. The test resulted in a global accuracy of 94%. The classifier's best performance (99%) was obtained (thanks to the anatomical information) on skin and sub-cutaneous fat, while the worst (90%) was obtained on cerebrospinal fluid. Not surprisingly, if the output of BD is not used, performance is remarkably worse, as the accuracy drops to about 70% in the case of patterns representing skin and sub-cutaneous fat.

	GM	WM	CSF	S/F	BG
GM	91.8	6.1	1.6	0.5	0
WM	6.9	91.4	1.0	0	0.7
CSF	3.9	1.3	89.5	5.3	1.3
S/F	0.5	0	0	99.0	0.5
BG	0.4	1.0	0.4	0.4	97.8
Global accuracy %					94.24

Table 1. Segmentation of MR spin-echo images: confusion matrix for the test set. The element C_{ij} is the percentage of image pixels belonging to class i that have been classified as belonging to class j. The classes considered are: grey matter (GM), white matter (WM), cerebrospinal-fluid (CSF), skin or fat (S/F), and background (BG).

The results we have obtained by the system are illustrated in Fig. 8, in which the two input images are shown in the top row, the two long-*TE* images produced by SR in the mid row, and the

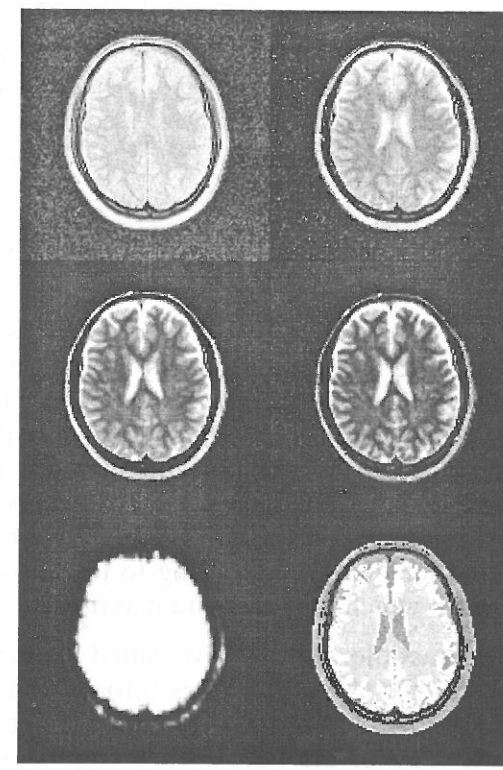

Fig. 8. Segmentation combining anatomical knowledge and physics of image generation: input images (*top row*); output of the Sequence Reconstructor (*mid row*); output of the Brain Detector (*bottom left*); final segmentation (*bottom right*).

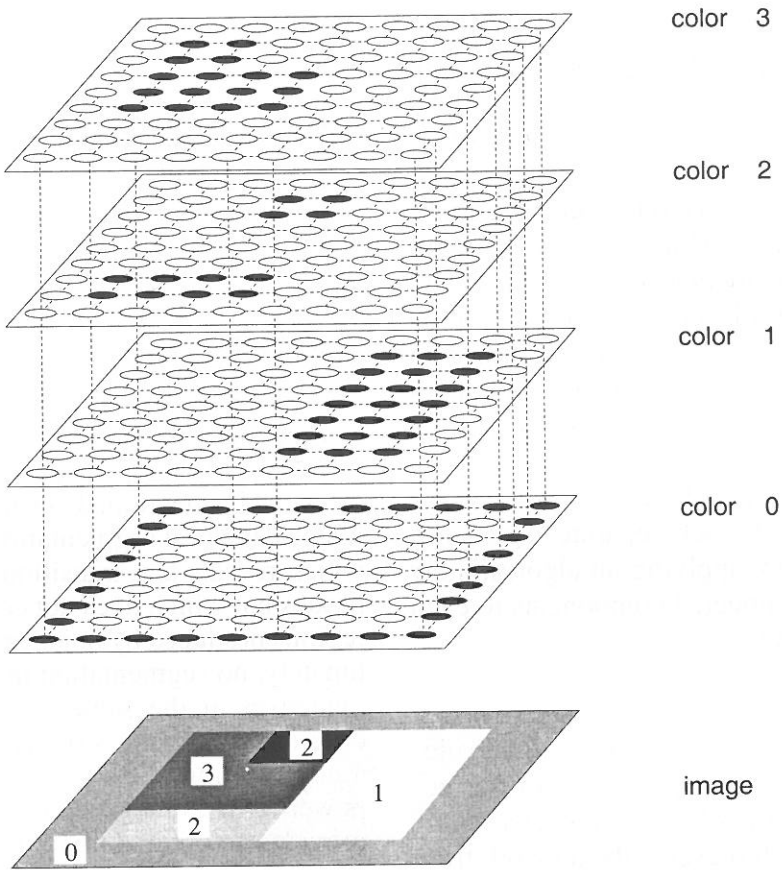

Fig. 9. Structure of the Hopfield's segmentation network.

output of BD and the final segmentation in the bottom row.

Segmentation Based on Perceptual Principles.

The requirements of maximum discriminating power and maximum grouping power for an ideal segmentation algorithm counteract each other and a trade-off solution is always necessary. If this needs to be achieved for each image, rather than being built into the algorithm, a quantitative criterion of goodness of segmentation needs to be explicitly defined. Once this is available, it can be optimized by the segmentation procedure for any specific image. Unfortunately, for any given image the space of possible segmentations is huge and cannot be explored effectively with standard optimization procedures.

Continuous Hopfield's networks [Hopfield, 1984] are dynamic systems evolving towards stable states which are the minima of an energy function E_{net} of the form:

$$E_{net} \approx -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} T_{ij} v_i v_j - \sum_{i=1}^{N} i_i v_i.$$

where v_i is the output of neuron i, i_i is its external input and T_{ij} is the weight of the connection from neuron j to neuron i. Thanks to this minimum-seeking dynamics, Hopfield's networks can be used to solve optimization problems [Hopfield & Tank, 1985, Hopfield & Tank, 1986]. Following similar approaches in the field of natural scene segmentation [Bilbro *et al.*, 1987, Darrell *et al.*, 1990, Darrell & Pentland, 1991, Reed, 1992, Wang *et al.*, 1992], we decided to use Hopfield's neural networks to solve the medical-image-segmentation optimization problem [Poli & Valli, 1997]. In the following we describe the steps required to do this.

The first step is to find a binary representation for segmentations, so that they can be mapped into the states of the neurons of a Hopfield's network. We have adopted a representation which has been suggested by the analogy between the

process of segmentation and the one of coloring geographic maps. A well-known theorem in graph theory states that in order to represent the states in a geographic map (regions of an image), only at most 4 colors are needed, as far as different colors are given to bordering states (connected regions). Thus, we can represent the results of the segmentation of an image with 4 bit maps (layers of neurons) each of which represents a different color. As illustrated in Fig. 9, the segmentation of an image can therefore be represented by the activations of a three-dimensional array of neurons. The activations are converted into a standard region-based segmentation (in which each separate region has a different label) by applying an algorithm for the detection of connected components to each layer (blob coloring).

The next step is to define a cost function E_{net} whose minimization provides an optimal solution to the segmentation problem. In our approach E_{net} is the sum of two terms: i) a *syntax energy* E_{syntax} which prevents the network from settling into non-binary states or states which cannot be mapped back to segmentations, ii) a *goodness energy* $E_{goodness}$ which drives the network towards points in state space which represent good segmentations.

The syntactic correctness of the solutions requires that one and only one neuron be active among the neurons representing a given pixel. This syntax constraint can be restated mathematically as: $\forall x, y \, \exists! \, l : v_{xyl} = 1$, where v_{xyl} denotes the activation of the neuron which represents the presence of label l for the pixel of position xy. In order to enforce this constraint, the syntax energy should be devised so that states which do not respect syntax rules have a very high energy. This can be obtained by including, for each pixel, the energy term $\sum_{l_1} \sum_{l_2 \neq l_1} v_{xyl_1} v_{xyl_2}$ and a corrective term $\left(\sum_l v_{xyl} - 1\right)^2$, which prevents the network from settling into the non-valid null solution $v_{xyl} = 0, \forall l$.

The syntax energy is obtained by summing up those terms for all the pixels in the image:

$$E_{syntax} = \frac{1}{2} \sum_x \sum_y \left[K_1 \sum_l \sum_{\hat{l} \neq l} v_{xyl} v_{xy\hat{l}} + K_2 \left(\sum_l v_{xyl} - 1 \right)^2 \right]$$

where K_1 and K_2 are constant values.

The goodness energy has to drive the network towards segmentations which are as good as possible from the perceptual point of view. As we already mentioned, in the case of medical images, the best segmentation would be the one which reveals any transition between different tissues but which does not contain any spurious regions produced by noise or by texture. Unfortunately, no segmentation method reaches both objectives at the same time. Therefore, any criterion of goodness of segmentation must be a combination of two terms: a discriminating power term and a grouping power term, which provides an optimal trade-off.

The discriminating power term of $E_{goodness}$ should make the network reveal any transition between different tissues, i.e. any change in the image gray-levels. In order to obtain this effect, we must include terms which increase when neighboring pixels lying across a boundary have the same label. To this end, we can use the following energetic term:

$$\sum_l v_{xyl} v_{\hat{x}\hat{y}l} \frac{dI(x,y)}{d\vec{n}(x,y,\hat{x},\hat{y})}$$

where (x, y) and $(\hat{x}, \hat{y})$ are two neighboring pixels, and $\frac{dI(x,y)}{d\vec{n}(x,y,\hat{x},\hat{y})}$ is the directional derivative of the image $I(x, y)$ in the direction $\vec{n}(x, y, \hat{x}, \hat{y}) = \frac{(\hat{x},\hat{y})-(x,y)}{\|(\hat{x},\hat{y})-(x,y)\|}$. This expression must be present for all pixels lying in a neighborhood $\mathcal{B}^{xy}$ of (x, y). $\mathcal{B}^{xy}$ should not contain pixels which are too close to or too far from (x, y). We have adopted the simplest neighborhood which meets these requirements:

$$\mathcal{B}^{xy} = \left\{ (\hat{x}, \hat{y}) \;\middle|\; 2 \leq \sqrt{(\hat{x}-x)^2 + (\hat{y}-y)^2} \leq 2\sqrt{2} \right\}.$$

The aim of the grouping power term is to force the net to construct large regions which have a

Noise σ	0	5	10	20	40	80
Wrong assignments (%)	1.05	1.17	1.32	1.46	16.38	52.88

Table 2. Segmentation of synthetic tomograms: wrong pixel assignments (%) vs. noise standard deviation.

high probability of representing single anatomical structures. We can obtain this effect with the following constraint: pixels which are close to each other should have the same label. In order to implement it we require the energetic term $-\sum_l v_{xyl} v_{\hat{x}\hat{y}l}$ to be minimum, for all the pixels $(\hat{x}, \hat{y})$ in the 4-connected neighborhood $\mathcal{N}^{xy}$ of each pixel (x, y).

By summing up the above terms for all pixels, we get the complete expression of the goodness energy:

$$E_{goodness} = \frac{1}{2} \sum_x \sum_y \left(K_3 \sum_{(\hat{x},\hat{y}) \in \mathcal{B}^{xy}} \sum_l v_{xyl} v_{\hat{x}\hat{y}l} \times \right.$$

$$\left. \times \frac{dI(x,y)}{d\vec{n}(x,y,\hat{x},\hat{y})} - K_4 \sum_{(\hat{x},\hat{y}) \in \mathcal{N}^{xy}} \sum_l v_{xyl} v_{\hat{x}\hat{y}l} \right)$$

where K_3 and K_4 are constant values.

Once the energy function E_{net} has been defined, the weights T_{ij} and inputs i_i of the network can be computed by direct term-matching as in [Hopfield & Tank, 1985]. The calculations show that two kinds of connections are present: a) excitatory and inhibitory intra-layer connections which implement perceptual grouping and discrimination principles, and b) inhibitory inter-layer connections which enforce syntactic correctness. Syntactic correctness is also favored by the excitatory external inputs which prevent the network from settling into the meaningless null state.

The segmentation network described is implemented by numerically integrating the motion equation of Hopfield's nets until a stable state is reached. As mentioned before, this state is then mapped back into a segmentation using blob coloring techniques.

The network described has been tested quantitatively on synthetic tomographic images and qualitatively on real tomographic ones.

Synthetic images were generated by simulating the operation of a real tomographic device on an ellipsoidal organ (grey level 230) surrounded by a homogeneous tissue (grey level 30). In order to test the robustness of the method, in addition to the blurring caused by the finite thickness of the slices (partial-volume effect), Gaussian white noise with zero mean and increasing standard deviation σ was included in the images. The resulting images were segmented using the network described above and then compared with the exact segmentation obtained manually with images in which noise and partial-volume effect were absent. Table 2 shows the average errors obtained in these experiments for several different values of σ. The table reveals that the method is quite insensitive to noise until this reaches relatively high levels (σ =40–80). (For $\sigma < 40$ the misclassification errors can be entirely attributed to the partial volume effect.)

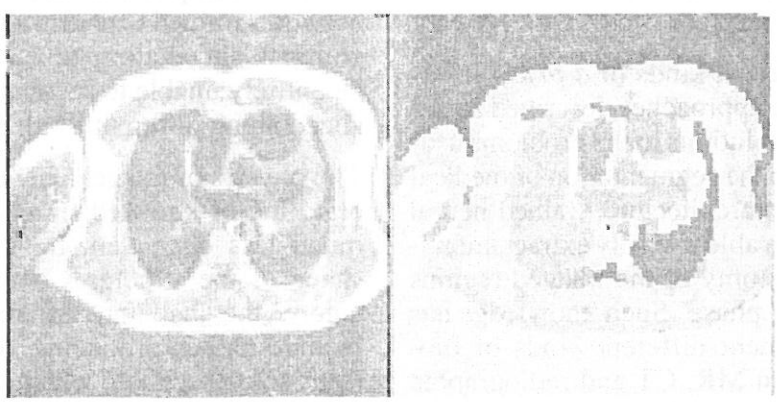

Fig. 10. Segmentation of an MR image of the thorax using a Hopfield's network that exploits perceptual regularities.

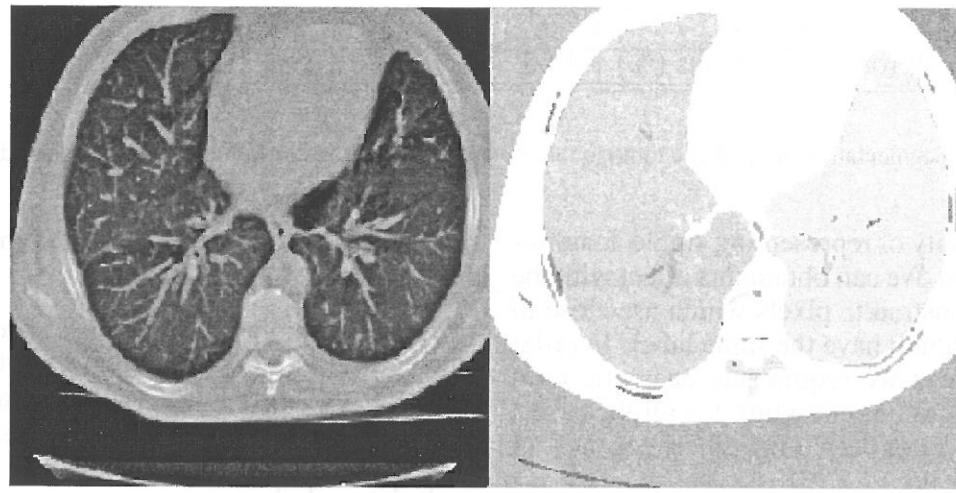

Fig. 11. Segmentation of a CT image of the thorax using a Hopfield's network as in the case of Fig. 10.

The good accuracy shown by the method in the experiments with synthetic images has been confirmed qualitatively by numerous experiments with real tomograms. Fig. 10 illustrates the results obtained on an MR image of the thorax. The segmentation of the original image (left) is shown after blob coloring (right). The algorithm has correctly segmented most of the anatomical structures of clinical interest such as lungs, sub-cutaneous fat, muscular tissue, right atrium, right ventricle, backbone, pulmonary artery, etc. In Fig. 11 we display an original CT image of the thorax and the related segmentation. The segmented image contains four main regions: the two lungs, the soft tissue and the background. There are also a number of small regions (rib borders, main tracts of bronchi, part of backbone, etc.) which, due to their high contrast, have not been grouped with surrounding tissue.

Discussion and Conclusions

By exploiting different kinds of *a priori* information, each of the approaches described above provides different solutions to the problems usually encountered in the segmentation of medical images. In the first architecture, trained neural networks have been able to easily extract knowledge about the anatomy of the imaged regions during the learning phase. Such knowledge has been used to segment different kinds of biological structures in MR, CT and radiographic images. In the second architecture, anatomical knowledge has been integrated with knowledge about the physics of image generation. In the third example, perceptual principles of grouping and discrimination have been implemented by a Hopfield's neural network which segments images through a relaxation process.

An important lesson learnt from using these architectures is that different kinds of knowledge unavoidably induce different strengths and limitations in the resulting segmentation systems, either in terms of generality or of performance.

Anatomical knowledge alone seems to be sufficient to guarantee a reliable segmentation of medical images. However, most representations of such knowledge (including the subsymbolic one used in this paper) are point-of-view dependent. This means that systems based on such representations can work properly only if the orientation, shape and appearance of the structures to be segmented are sufficiently similar to those used to build the knowledge representation adopted. In the first system we presented in the previous sections this means that networks trained to perform the segmentation of an organ sliced along transversal planes would be entirely unable to segment the same organ if sliced along differently oriented planes.

The physics of image generation is a more general kind of knowledge, and is expected to be much less dependent on the orientation and shape of the structures considered. Unfortunately, the fact that different structures may include tissues producing very similar signals leads to ambiguities which cannot be resolved without the use of other sources of knowledge. In our system we did this using anatomical

knowledge. Nevertheless, it must be pointed out that changes of image acquisition parameters or acquisition devices may render this type of knowledge unusable.

On the other hand, trained networks are characterized by an intrinsic simplicity of knowledge acquisition (the training sets can directly be generated by experts through simple graphic interfaces). To a certain extent, this could be exploited to override some of the previously mentioned limitations by using a different module, based on the same architecture, for different structures of interest, for different points-of-view, and for different image sequences or imaging devices. However, in general, the construction of the corresponding training sets might become long and tedious.

The regularities of nature are an even more general form of knowledge. They do not depend either on the point-of-view or on the shape of the structures to be segmented, or on the imaging parameters. When noise is not too strong and texture not too marked, this form of knowledge leads to general-purpose segmentation systems. In many kinds of tomographic images this is true and, indeed, our third system performed good segmentations on such images. However, the experiments with synthetic images have shown that this generality may be paid with a reduced robustness with respect to systems in which more specialised forms of knowledge are used.

Our experience with the second system indicates that a careful exploitation and integration of available knowledge sources via proper neural modules can lead to the systems which overcome some of the limitations implicit in each source. A lot more can be done in this direction. We will devote our future research to this.

In any case, the properties shown by ANNs in our work lead us to believe that they are superior to the symbolic segmentation methods we developed in the past [Calamai *et al.*, 1990, Coppini *et al.*, 1993a]. Therefore, we think neural nets should be favored when choosing the architectures to exploit the sources of knowledge available in medical images. In addition we wish to point out that, besides image segmentation, ANNs can help solve other difficult problems of medical computer vision, such as the recovery of 3-D shape from incomplete data [Coppini *et al.*, 1995] and the classification of biological structures [Rucci *et al.*, 1995].

On the basis of our experience, we cannot imagine ways of obtaining maximum performance and generality at the same time. This is after all the well known strong-methods vs. weak-methods dilemma which has been afflicting AI search techniques for decades [Russel & Norvig, 1995]. However, it is possible to imagine that, for different applications, a different combination of the three sources of knowledge we have identified and exploited above could give optimum performance/generality trade-offs.

Acknowledgments

This work has been partially supported by the Italian Ministry of University and Scientific and Technological Research (MURST), the Italian National Research Council (CNR) and the British Council/MURST CRUI agreement.

References

[Bajcsy & Kovacic, 1989] R. BAJCSY, S. KOVACIC, Multiresolution elastic matching, *Computer Vision, Graphics, and Image Processing*, **46** (1989), 1–21.

[Ballard & Brown, 1982] D. H. BALLARD, C. M. BROWN, *Computer Vision*, Prentice–Hall, Englewood Cliff, NJ, 1982.

[Bilbro *et al.*, 1987] G. L. BILBRO, M. WHITE, W. SNYDER, Image segmentation with neurocomputers, In *Neural Computers* (R. ECKMILLER, C. V.D. MALSBURG, EDS), (1987), Springer–Verlag, Berlin.

[Cagnoni *et al.*, 1992] S. CAGNONI, D. CARAMELLA, R. DE DOMINICIS, G. VALLI, Neural network modelling of spin echo multiecho sequences, *Journal of Digital Imaging*, **5** (1992), 89–94.

[Cagnoni *et al.*, 1993] S. CAGNONI, G. COPPINI, M. RUCCI, D. CARAMELLA, G. VALLI, Neural network segmentation of magnetic resonance spin echo images of the brain, *Journal of Biomedical Engineering*, **15** (1993), 355–362.

[Calamai *et al.*, 1990] R. CALAMAI, G. COPPINI, M. DEMI, R. POLI, G. VALLI, A computational approach to medical imaging, *Journal of Nuclear Medicine and Allied Sciences*, **34** (1990), 42–50.

[Coppini *et al.*, 1992] G. COPPINI, R. POLI, M. RUCCI, G. VALLI, A neural network architecture for understanding 3D scenes in medical imaging, *Computer and Biomedical Research*, **25** (1992), 569–585.

[Coppini *et al.*, 1993a] G. COPPINI, M. DEMI, R. POLI, G. VALLI, An artificial vision system for X-ray images of human coronary trees, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15** (1993), 156–162.

[Coppini *et al.*, 1993b] G. COPPINI, R. POLI, R. LEGITTIMO, R. DE DOMINICIS, G. VALLI, A neural network system for detecting lung nodules in chest radiograms, In *Computer Assisted Radiology, CAR'93* (H. U. LEMKE, K. INAMURA, C. C. JAFFE, R. FELIX, EDS), (1993) pp. 594–599, Springer–Verlag, Berlin.

[Coppini *et al.*, 1995] G. COPPINI, R. POLI, G. VALLI, Recovery of the 3–D shape of the left ventricle from echocardiographic images, *IEEE Transactions on Medical Imaging*, **14** (1995), 301–317.

[Darrell & Pentland, 1991] T. Darrell, A. Pentland, Discontinuity models and multi-layer description networks, Technical Report 162, Vision and Modeling Group, The Media Lab, M.I.T., 1991.

[Darrell *et al.*, 1990] T. DARRELL, S. SCLAROFF, A. PENTLAND, Segmentation by minimal description, In *IEEE International Conference on Computer Vision III* (1990), Osaka, Japan.

[Gerig *et al.*, 1992] G. GERIG, J. MARTIN, R. KIKINIS, O. KUBLER, M. SHENTON, F. A. JOLESZ, Unsupervised tissue type segmentation of 3D dual-echo MR head data, *Image and Vision Computing*, **10** (1992), 349–360.

[Hall *et al.*, 1992] L. O. HALL, A. M. BENSAID, L. P. CLARKE, R. P. VELTHUIZEN, M. S. SILBUGER, J. C. BEZDEK, A comparison on neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain, *IEEE Transactions on Neural Networks*, **3** (1992), 672–682.

[Hopfield & Tank, 1985] J. J. HOPFIELD, D. W. TANK, "Neural" computation of decisions in optimization problems, *Biological Cybernetics*, **52** (1985), 141–152.

[Hopfield & Tank, 1986] J. J. HOPFIELD, D. W. TANK, Computing with neural circuits: a model, *Science*, **233** (1986), 625–633.

[Hopfield, 1984] J. J. HOPFIELD, Neurons with graded response have collective computational properties like those of two-state neurons, *Proceedings of the National Academy of Sciences*, **81** (1984), 3088–3092.

[Hubel, 1988] D. H. HUBEL, *Eye, Brain, and Vision*, Scientific American Books, New York, 1988.

[Kohonen, 1989] T. KOHONEN, *Self-Organization and Associative Memory*, Springer–Verlag, Berlin, 1989.

[Kohonen, 1995] T. KOHONEN, *Self-Organizing Maps*, Springer–Verlag, Berlin, 1995.

[Lundervold & Storvik, 1995] A. LUNDERVOLD, G. STORVIK, Segmentation of brain parenchyma and cerebrospinal fluid in multispectral magnetic resonance images, *IEEE Transactions on Medical Imaging*, **14** (1995), 339–349.

[MacFall *et al.*, 1986] J. R. MACFALL, S. J. RIEDERER, H. Z. WANG, An analysis of noise propagation in computed T2, pseudodensity, and synthetic spin-echo images, *Medical Physics*, **13** (1986), 285–292.

[Macovski, 1983] A. MACOVSKI, *Medical Imaging Systems*, Prentice Hall, Englewood Cliffs, 1983.

[Marr, 1982] D. MARR, *Vision*, W. H. Freeman & Co., New York, 1982.

[Özkan *et al.*, 1990] M. ÖZKAN, H. G. SPRENKELS, B. M. DAWANT, Multi-spectral magnetic resonance image segmentation using neural networks, In *International Joint Conference on Neural Networks, Vol. I* (1990) pp. 429–434.

[Piraino *et al.*, 1991] D. PIRAINO, S. SUNDAR, B. RICHMOND, J. SCHILS, J. THOME, Segmentation of magnetic resonance images using a backpropagation neural network, In *Annual International Conference of the IEEE EMBS* (1991) pp. 1466–1467.

[Poli & Valli, 1997] R. Poli, G. Valli, Hopfield neural nets for the optimum segmentation of medical images, In *Handbook of Neural Computation* (E. FISHER, R. BEALE, EDS), chapter G5.5. IOP and Oxford University Press 1997.

[Raya, 1990] S. P. RAYA, Low-level segmentation of 3–D magnetic resonance brain images — a rule-based system, *IEEE Transactions on Medical Imaging*, **9** (1990), 327–337.

[Reed, 1992] T. R. Reed, Region growing using neural networks, In *Neural Networks for Perception, Vol. 1* (H. WECHSLER, ED), pp. 386–397, Academic Press San Diego, CA 1992.

[Reuman & Hoffman, 1986] S. R. Reuman, D. D. Hoffman, Regularities of nature: the interpretation of visual motion, In *From Pixels to Predicates* (A. P. PENTLAND, ED), pp. 201–226, Ablex Norwood, New Jersey 1986.

[Rucci *et al.*, 1995] M. RUCCI, G. COPPINI, I. NICOLETTI, D. CHELI, G. VALLI, Automatic analysis of hand radiographs for the assessment of skeletal age: a subsymbolic approach, *Computers and Biomedical Research*, **28** (1995), 239–256.

[Rumelhart & McClelland, 1986] D. E. RUMELHART, J. L. MCCLELLAND, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1–2*, MIT Press, Cambridge, MA, 1986.

[Russel & Norvig, 1995] S. RUSSEL, P. NORVIG, *Artificial Intelligence: a Modern Approach*, Prentice Hall, Upper Saddle River, 1995.

[Sonka *et al.*, 1996] M. SONKA, S. K. TADIKONDA, S. M. COLLINS, Knowledge-based segmentation of MR brain images, *IEEE Transactions on Medical Imaging*, **15** (1996), 443–452.

[Stansfield, 1986] S. A. STANSFIELD, Angy: A rule-based expert system for automatic segmentation of coronary vessels from digital subtracted angiograms, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **8** (1986), 188–199.

[Wang *et al.*, 1992] T. WANG, X. ZHUANG, X. XING, Robust segmentation of noisy images using a neural network model, *Image and Vision Computing*, **10** (1992), 233–240.

[Webb, 1991] S. WEBB, *The Physics of Medical Imaging*, Institute of Physics Publishing, Bristol and Philadelphia, 1991.

[Wells III *et al.*, 1996] W. M. WELLS III, W.E.L. GRIMSON, R. KIKINS, F. A. JOLESZ, Adaptive segmentation of MRI data, *IEEE Transactions on Medical Imaging*, **15** (1996), 429–442.

[Wolpert, 1992] D. H. WOLPERT, Stacked generalization, *Neural Networks*, **5** (1992), 241–259.

Received: June, 1997
Revised: March, 1998
Accepted: April, 1998

Contact address:
G. Valli
Department of Electronic Engineering
University of Florence
Via S. Marta, 3
I-50139 – Italy
phone : +39–055–4796–378
fax: +39–055–494569
e-mail: valli@diefi.die.unifi.it

GUIDO VALLI received his degree in Physics from the University of Padua in 1963. He is professor of Biomedical Technology at the Department of Electronic Engineering of the University of Florence, where he has taught since 1980. Previously, he was with the Italian Research Council (CNR). His main research interests are medical imaging and computer vision systems for the understanding of medical images. He has authored or co-authored more than 120 papers. He is a member of the GNB (Gruppo Nazionale di Bioingegneria), the AEI (Associazione Elettrotecnica ed Elettronica Italiana), and the AIIMB (Associazione Italiana di Ingegneria Medica e Biologica).

GIUSEPPE COPPINI received his degree in Electronic Engineering from the University of Florence in 1980. He is a researcher at the Institute of Clinical Physiology of the Italian Research Council (CNR) in Pisa. Since 1980, Dr. Coppini has collaborated with the Department of Electronic Engineering of the University of Florence. His research interests include medical imaging, cardiovascular imaging, computer vision, medical-image understanding, neural-networks architectures for vision systems. He is author or co-author of more than 80 papers. Giuseppe Coppini is a member of the IEEE, the IEEE Engineering in Medicine and Biology Society, and the AEI (Associazione Elettrotecnica ed Elettronica Italiana).

RICCARDO POLI received his degree in Electronic Engineering with Summa Cum Laude (in 1989) and the PhD in Biomedical Engineering (in 1993) from the University of Florence, Italy, earning the prize for the best Italian PhD thesis in the field. Since 1994 he has been with the School of Computer Science of The University of Birmingham teaching artificial intelligence. Dr Poli is the founder and coordinator of the Evolutionary and Emergent Behaviour Intelligence and Computation (EEBIC) group at Birmingham, which is an associate node of the European Network of Excellence in Evolutionary Computation. Dr. Poli is author of more than 60 publications in computer vision, medical image analysis, neural networks, artificial intelligence and evolutionary computation.

STEFANO CAGNONI received his degree in Electronic Engineering and the Ph.D. in Biomedical Engineering from the University of Florence. In 1994 he was a visiting scientist at the Whitaker College Biomedical Imaging and Computation Laboratory of the Massachusetts Institute of Technology. He is presently an assistant professor at the Department of Computer Engineering of the University of Parma, Italy.