INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF NEURAL ENGINEERING
J. Neural Eng. **1** (2004) 228–237 PII: S1741-2560(04)85485-5

Determining the receptive field of a neural filter

Kenji Suzuki

Kurt Rossmann Laboratories for Radiologic Image Research, Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637, USA

E-mail: suzuki@uchicago.edu

Received 26 August 2004
Accepted for publication 22 October 2004
Published 2 December 2004
Online at stacks.iop.org/JNE/1/228
doi:10.1088/1741-2560/1/4/006

Abstract

In this paper, a method for determining the receptive field and the structure of hidden layers of a neural filter (NF) was developed and evaluated. With the proposed method, redundant units are removed from input and hidden layers in an NF based on the influence of removal of units on the error between output and teaching images. By performing the removal of units and retraining for recovery of the loss of the removal repeatedly, the receptive field and a reduced structure of hidden layers are determined. Experiments with NFs were performed for acquiring the function of a known filter, for the reduction of noise in natural images and for the reduction of noise in medical image sequences. By use of the proposed method, redundant units were able to be removed from NFs, while the performance of the NFs was maintained. Experimental results suggested that, with the proposed method, a reasonable receptive field for a given image-processing task could be determined, i.e., the receptive field of the NF trained to obtain the function of a filter corresponded to the kernel of the filter, and the receptive fields of the NFs for noise reduction gathered around the object pixels in the input regions of the NFs.

1. Introduction

In the field of image processing, supervised nonlinear filters based on artificial neural networks (ANNs) called neural filters (NFs) have been studied for obtaining a desired image-processing function from examples (data) (Arakawa and Harashima 1990, Yin *et al* 1993, Suzuki *et al* 1998a, 1998b, 2002a, 2002b). NFs can learn about the desired function that is determined by the relationship between input images and the corresponding teaching images through training. For example, when noisy input images and noiseless teaching images are presented to an NF, the NF can learn about the reduction of noise in images without the spatial blur of signals. NFs are useful for the reduction of a specific type of noise which it may be difficult to represent by a simple model, e.g., the performance of an NF in the reduction of quantum mottle (specific noise caused by medical devices) in medical x-ray images, the modeling of which is complex, was superior to that of conventional nonlinear filters (Suzuki *et al* 2002a, 2002b). Recently, NFs were extended to enhancement of edges in noisy images (Suzuki *et al* 2003), and also to enhancement of edges traced by a medical doctor in medical images (Suzuki *et al* 2004).

Because we cannot determine the structure of an NF which is optimal for a given image-processing task prior to training, a relatively large structure is usually used to learn the task sufficiently. A trained NF, therefore, could have many redundant units (model neurons) in the structure. This could cause a low generalization ability (performance for non-training samples) due to overfitting of training samples. Because the function of an NF is determined by data, analysis of the trained NF is very important for understanding the function obtained. It would be useful for understanding the trained NF to determine the receptive field (Hubel and Wiesel 1962) (i.e., the effective input units in which input signals lead to the output response) of the NF which is adequate for a given task. However, many redundant units may make the analysis of the trained NF difficult, and they also may prevent a proper determination of the receptive field. In addition, redundant units cause the computational cost to increase. Therefore, it is important to remove redundant units as much as possible from the input and hidden layers in an NF in order to determine the receptive field and the structure of hidden layers which are

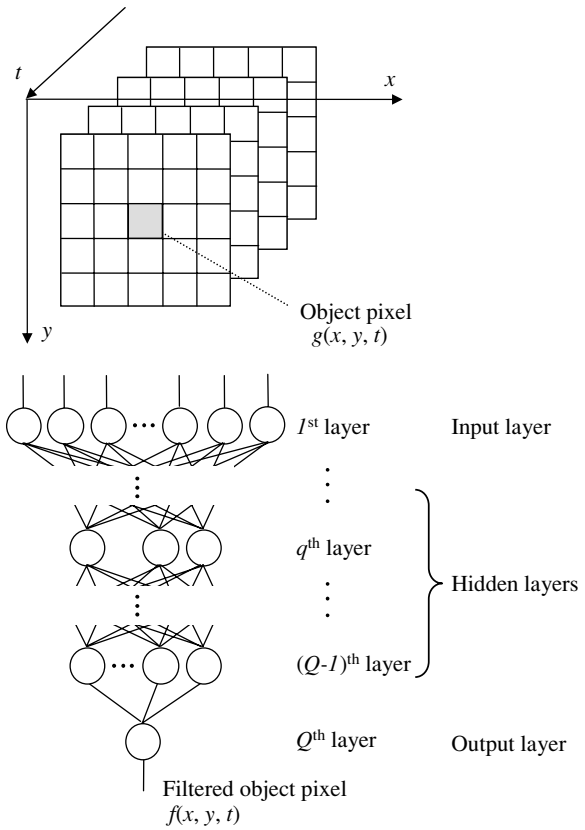


Figure 1. Architecture of an NF.

adequate for a given task, while the performance of the NF is maintained.

The purpose of this study was to develop and evaluate a method for determining the receptive field and the structure of hidden layers of an NF which are adequate for a given image-processing task.

2. Method for determining the receptive field of a neural filter

2.1. Architecture of a neural filter

An NF consists of a linear-output multilayer ANN model (Suzuki *et al* 1995, 2003) which is capable of operating on image data directly. The linear-output multilayer ANN model employs a linear function instead of the commonly used sigmoid function as the activation function of the output unit for outputting continuous values. The activation functions of units in the input, hidden and output layers of the NF are an identity function, a sigmoid function and a linear function, respectively. The architecture of the NF is shown in figure 1. The inputs of the NF are the pixel values in a spatiotemporal local region R (which corresponds to an input region of the NF) in input images, i.e., an object pixel value $g(x, y, t)$ and spatiotemporally neighboring pixel values. The pixel values of input images are normalized from zero to one. The output is a continuous value, which corresponds to the center pixel in the local region, represented by

$$f(x, y, t) = NN(\mathbf{I}_{x,y,t}), \tag{1}$$

where

$$\mathbf{I}_{x,y,t} = \{g(x - i, y - j, t - k) | i, j, k \in R\} \tag{2}$$

is the input vector to the NF, x and y are the spatial coordinates, t is the temporal coordinate, and $NN(\mathbf{I})$ is the output of the linear-output multilayer ANN. Note that only one unit is employed in the output layer. The input vector can be rewritten as

$$\mathbf{I}_{x,y,t} = \{I_1, I_2, \ldots, I_m, \ldots I_{N^{(1)}}\}, \tag{3}$$

where m is an input unit number, and $N^{(1)}$ is the number of input units. Because the activation functions of the units in the input layer are an identity function $f_I(\cdot)$, the output of the nth unit in the input layer can be represented by

$$O_n^{(1)} = f_I(I_n) = I_n. \tag{4}$$

The output of the nth unit in the qth layer is represented by

$$O_n^{(q)} = f_S \left\{ \sum_{m=1}^{N^{(q-1)}} \left(W_{mn}^{(q)} \cdot O_m^{(q-1)} \right) - W_{0n}^{(q)} \right\}, \tag{5}$$

where $W_{mn}^{(q)}$ is a weight between the mth unit in the $(q-1)$th layer and the nth unit in the qth layer, $W_{0n}^{(q)}$ is an offset of the nth unit in the qth layer, $N^{(q-1)}$ is the number of units in the $(q-1)$th layer, and $f_S(u)$ is a sigmoid function:

$$f_S(u) = \frac{1}{1 + \exp(-u)}. \tag{6}$$

The output of the unit in the output (Qth) layer is represented by

$$NN(\mathbf{I}_{x,y,t}) = f_L \left\{ \sum_{m=1}^{N^{(Q-1)}} \left(W_m^{(Q)} \cdot O_m^{(Q-1)} \right) - W_0^{(Q)} \right\}, \tag{7}$$

where $W_m^{(Q)}$ is a weight between the mth unit in the $(Q-1)$th layer and the unit in the output layer, $W_0^{(Q)}$ is an offset of the unit in the output layer, and $f_L(u)$ is a linear function,

$$f_L(u) = u + \tfrac{1}{2}. \tag{8}$$

The error to be minimized by training is defined by

$$E = \frac{1}{N} \sum_{x,y,t} \{T_C(x, y, t) - f(x, y, t)\}^2, \tag{9}$$

where $T_C(x, y, t)$ is a teaching image, and N is the number of training samples. The NF is trained by a linear-output back-propagation (BP) algorithm (Suzuki *et al* 1995, 2003), which was derived for a linear-output ANN model by use of the steepest descent method, in the same way as in the derivation of an original BP algorithm (Rumelhart *et al* 1986a, 1986b),

until the convergence criterion is fulfilled, e.g., the number of training epochs (times) exceeds a predetermined number or the error E becomes smaller than or equal to a predetermined error E_P.

2.2. Proposed method

The basic concept of the proposed method for determining the receptive field and the structure of hidden layers of an NF is as follows: first, an NF with a structure that is large enough to learn a given image-processing task is prepared. Next, the weights of the NF are initialized with small random numbers. The NF is trained until the convergence criterion is fulfilled. There are many redundant units in the trained NF. Redundant units are removed from the input layer and hidden layers based on the influence of removal of each unit on the error between the output images and teaching images. The unit with the smallest influence is removed first. Then, the NF is retrained to recover the potential loss due to this removal. This process is performed repeatedly until every unit is examined (convergence criterion of this method), resulting in a reduced structure from which redundant units are removed. Thus, the convergence is determined experimentally in this method, because it would be difficult to determine the convergence criterion for various applications theoretically.

A flow chart of the proposed method is shown in figure 2. Let $E^{(n,q)}$, $F^{(q)}$ and E^{MAX} define an error after the removal of the nth unit in the qth layer, an indication of a target layer for removal, and a high enough value against the error E, respectively. The proposed method consists of the following steps:

Step 1. Train a large NF until $E \leqslant E_P$.
Step 2. Remove the nth unit in the qth layer experimentally and calculate $E^{(n,q)}$.
Step 3. If every unit in a target layer is examined by removing it experimentally and calculating $E^{(n,q)}$, then go to step 4; otherwise, go back to step 2.
Step 4. Remove the ath unit in the bth layer, where $E^{(a,b)}$ is the minimum among $E^{(n,q)}$.
Step 5. Retrain the NF, where the ath unit has been removed from the bth layer, by use of a linear-output BP algorithm until the convergence criterion is fulfilled.
Step 6. If $E \leqslant E_P$, then memorize the weights and the structure, and go back to step 2; otherwise, go to step 7.
Step 7. Replace the current network by the previous one, i.e., the structure before removal of the ath unit in the bth layer, and exclude the bth layer from candidates for removal.
Step 8. If every layer is examined (convergence criterion of this method), then finish the steps; otherwise, go back to step 2.

It is easy to remove a unit experimentally by use of the following property of an ANN: the output of an NF where a certain unit is removed equals the output of an NF where the weights connected to the output of the unit are set to zero (or the output of the unit is set to zero).

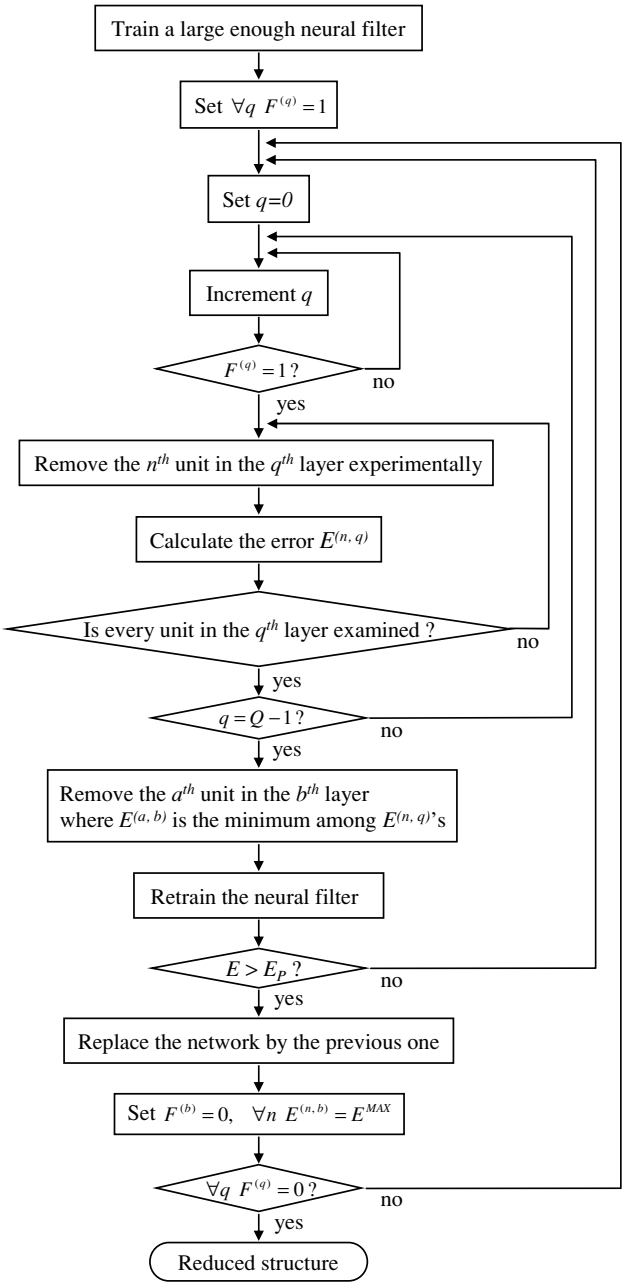


Figure 2. Flow chart of the proposed method for determining the receptive field and the structure of hidden layers of an NF.

3. Validation of the proposed method

3.1. Experiment with an NF to learn the function of a known filter

To validate the proposed method, an experiment was performed with a spatial NF where the input region is two dimensional to learn the function of a known filter. A Laplacian filter was used as a known filter in this experiment. An input image (512 × 512 pixels) containing white uniform random noise was used, because random noise contains various intensities and frequencies, and is used for the analysis of a

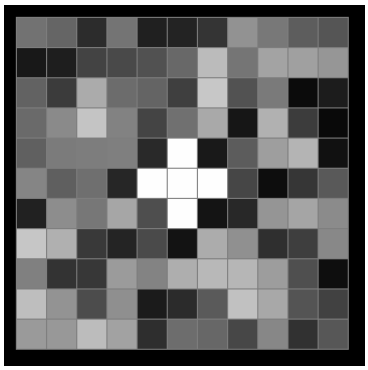


Figure 3. Receptive field obtained by use of the proposed method for the NF trained to obtain the function of a Laplacian filter. Each small square indicates an input unit in the NF. White squares indicate a remaining unit, and gray or black squares indicate a removed unit. The order of the removal is represented by the gray tone. The darker squares indicate units which are removed earlier.

linear system. A teaching image was obtained by filtering the input image with a Laplacian filter, represented by

$$L(x, y) = 4 \cdot g(x, y) - g(x, y-1) - g(x-1, y) - g(x+1, y) - g(x, y+1). \quad (10)$$

The input region of the spatial NF consisted of 11×11 pixels which were sufficient to learn the function of a Laplacian filter. The numbers of units in the input, hidden and output layers were 121, 50 and 1, respectively (here referred to as 121–50–1). The spatial NF was trained for 100 000 epochs with training pixels in rectangular regions (50×100 pixels) in the input and teaching images. A mean absolute error (MAE) has been used in the studies on NFs (Suzuki *et al* 1998a, 1998b, 2002a, 2002b), because an MAE is more sensitive to the degradation of details of objects in images compared to the use of a mean squared error. An MAE is defined by

$$E_A = \frac{1}{N_E} \sum_{x,y \in R_E} |T_C(x, y) - f(x, y)|, \quad (11)$$

where R_E is a region for evaluation, and N_E is the number of pixels in R_E. Note that $T_C(x, y)$ and $f(x, y)$ are normalized such that the maximum value of the gray scale is one and the minimum value is zero. The training converged with an MAE of 0.017. The CPU execution time of the training was 39.2 h on a workstation (UltraSPARC II, 300 MHz, Sun Microsystems, CA).

3.2. Determining the receptive field

The proposed method was applied to the trained NF (original NF). By use of the proposed method, the receptive field was determined to have the same form as the input kernel of a Laplacian filter used for training, as shown in figure 3. This result suggested that the proposed method could remove redundant units from the input layer effectively. In order to compare the generalization ability of the original NF with that of the NF obtained by use of the proposed method, error maps were obtained by subtracting the output images of the NFs from the teaching image. In the error map for the original NF (figure 4(*a*)), errors (MAE of 0.017) are small only in the training region due to overtraining, whereas errors (MAE of 0.017) are small in both training and non-training regions in the error map for the NF obtained by use of the proposed method (figure 4(*b*)). The average error (MAE of 0.017) in the non-training region of the NF obtained by use of the proposed method was about three times smaller than that (MAE of 0.053) of the original NF. Thus, the generalization ability of the NF was improved by use of the proposed method.

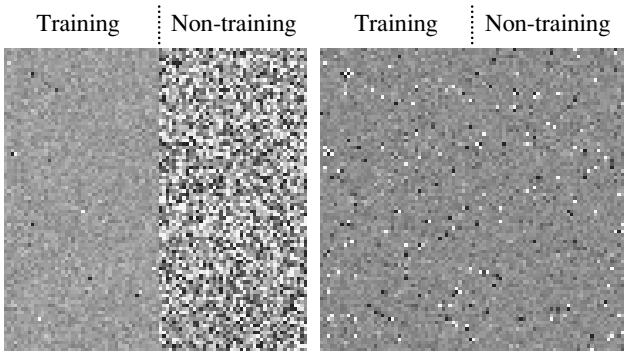


Figure 4. Comparison of the generalization ability of the original NF with that of the NF obtained by use of the proposed method. These images are error maps obtained by subtracting the output images of the NFs from the teaching image. (*a*) Error map for the original NF. MAEs in the training region (left half region) and the non-training region (right half region) were 0.017 and 0.053, respectively. (*b*) Error map for the NF obtained by use of the proposed method. MAEs in the training region and the non-training region were 0.017 and 0.017, respectively.

3.3. Comparison with a conventional method

Various methods for determining the structure of an ANN have been proposed (Reed 1993), which can be classified into three groups:

(1) Units in an ANN are removed from hidden layers based on a performance index, i.e., an ANN is trained, the performance index is calculated for units in hidden layers, and the unit with the smallest performance index is removed from a hidden layer. This process is repeated until the smallest performance index exceeds a predetermined value (Hagiwara 1989, 1990, Sietsma and Dow 1991, Kameyama and Kosugi 1991, Castellano *et al* 1997).

(2) ANNs with various numbers of units in hidden layers are trained, trained ANNs are evaluated by use of the information criteria, and the ANN with the maximum score is determined to be an optimal one (Kurita 1990, Fogel 1991, Murata *et al* 1994).

(3) Weights are removed from an ANN based on a performance index, or values of weights are gradually diminished during training by use of a training algorithm with a penalty term (Chauvin 1989, LeCun *et al* 1990, Ji *et al* 1990, Weigend *et al* 1991, Ishikawa and Uchida 1992, Nowlan and Hinton 1992, Hassibi and Stork 1993, Ishikawa 1994, 1996).

(*a*) (*b*)

Figure 5. Natural image ('Lena') used for the training of a spatial NF for the reduction of noise. (*a*) Noisy input image. (*b*) Noiseless teaching image.

Table 1. Comparison of the proposed method with a conventional method (OBD).

Neural filter	Execution time (h)	Structure
Original NF	39.2[a]	121–50–1
NF obtained by OBD	26.0	5–43–1
NF obtained by the proposed method	20.2	5–5–1

[a] Note that the execution time of the original NF corresponds to the time for training.

The proposed method was compared with a conventional method called optimal brain damage (OBD) (LeCun *et al* 1990), which is a well-known method cited in many papers (Ishikawa and Uchida 1992, Reed 1993, Gorodkin *et al* 1993, Ishikawa 1994, 1996, Castellano *et al* 1997, Dumitras and Kossentini 2000, Chandrasekaran *et al* 2000, Castellano and Fanelli 2000). With OBD, weights in an NF were removed based on the damage estimated from the second derivative of the error with respect to weights. When all output weights from a certain unit were removed, the unit itself was removed. The results are summarized in table 1. The proposed method eliminated more units from the hidden layer than did the OBD.

4. Experiment with a spatial NF for noise reduction

4.1. Training a spatial NF

An image (size, 512 × 512 pixels; number of gray levels, 256) called 'Lena' from the University of Southern California (USC) image database was used for training a spatial NF in this experiment. For the reduction of quantum noise in images, a noisy image was synthesized by addition of quantum noise (which is signal-dependent noise) to a noiseless original image $g_S(x, y)$, represented by

$$g(x, y) = g_S(x, y) + g_N(\sigma_{g_S(x,y)}), \tag{12}$$

where $g_N(\sigma_{g_S(x,y)})$ is noise and where the standard deviation $\sigma_{g_S(x,y)} = k_N\sqrt{g_S(x, y)}$, and k_N is a parameter determining the amount of noise. A noisy image (figure 5(*a*)) and a noiseless original image (figure 5(*b*)) were used as input image and as teaching image, respectively, where a k_N of 5% of the maximum gray level was used for synthesizing the input image. For a sufficient reduction of noise, the spatial input region of the NF consisted of 11 × 11 pixels. A three-layer structure was used for the NF, because a three-layered ANN can realize any continuous mapping approximately (Funahashi 1989, Barron 1993). The structure of the NF was 121–50–1. For training features in the entire image efficiently, 5000 training pixels were extracted randomly from the input and teaching images. The training of the NF was performed for 100 000 epochs, and it converged with an MAE of 0.018.

4.2. Comparison of the performance of the proposed method with that of conventional methods

For comparison of the proposed method with conventional methods, Hagiwara's method (Hagiwara 1990) and OBD (LeCun *et al* 1990) were selected from groups (1) and (3) (see section 3.3), respectively. In the Hagiwara method, units are removed based on a performance index defined by use of the back-propagation error. The proposed method and the conventional methods were evaluated by use of the execution time, the number of removed units and the performance of the NF obtained. The results are shown in table 2. The execution time was defined as the total CPU execution time for the removal of units and retraining. The execution time of Hagiwara's method was three times greater than that for the training of the original NF, whereas the execution time of OBD and that of the proposed method were comparable.

The method for removing units from an NF can be evaluated by the use of two measures: (1) the number of units removed from an NF; (2) the image quality of the output images of the obtained NF. Measure (1) is the figure of merit of a method for removing units from an NF, which is related to the reduction of the computational cost of the NF. Measure (2) is the figure of merit (or performance) of the obtained NF. I used the improvement in signal-to-noise ratio

(*a*) (*b*)

Figure 6. Comparison of the output image of the original NF with that of the NF obtained by use of the proposed method. (*a*) Output image of the original NF. (*b*) Output image of the NF obtained by use of the proposed method.

Table 2. Comparison of the proposed method with conventional methods for spatial NFs for the reduction of noise in natural images.

Neural filter	Execution time (h)	Structure	ISNR (dB)
Original NF	48.0[a]	121–50–1	7.6
NF obtained by Hagiwara's method	161.6	121–27–1	7.5
NF obtained by OBD	47.6	61–8–1	7.5
NF obtained by the proposed method	40.4	22–8–1	7.6

[a] Note that the execution time of the original NF corresponds to the time for training.

(ISNR) (Banham and Katsaggelos 1997, Brailean *et al* 1995) as measure (2), which is defined by

$$\mathrm{ISNR} = 10\log_{10}\frac{\sum_{x,y}\{T_C(x,y)-g(x,y)\}^2}{\sum_{x,y}\{T_C(x,y)-f(x,y)\}^2}. \qquad (13)$$

The ISNR represents the improvement of the signal-to-noise ratio of the output images from the signal-to-noise ratio of the input images.

As for measure (1), the proposed method removed a larger number of units from the input layer than did the conventional methods, as shown in table 2. The output images of the original NF and the NF obtained by use of the proposed method are very similar, as shown in figure 6. The ISNR of the NF obtained by use of the proposed method was the same as that of the original NF, as shown in table 2. This result showed that the proposed method did not degrade the image quality by the removal of units. The NFs were applied to non-training images obtained from the USC standard image database. The output images of the original NF and of the NF obtained by use of the proposed method are very similar, as shown in figure 7. Some of the edges in the output images of the NF obtained by use of the proposed method seem to be slightly sharper. The ISNRs of the NF obtained by use of the proposed method are greater than those of other NFs, as shown in table 3. Thus, the NF obtained by use of the proposed method worked similar to the training image for non-training images.

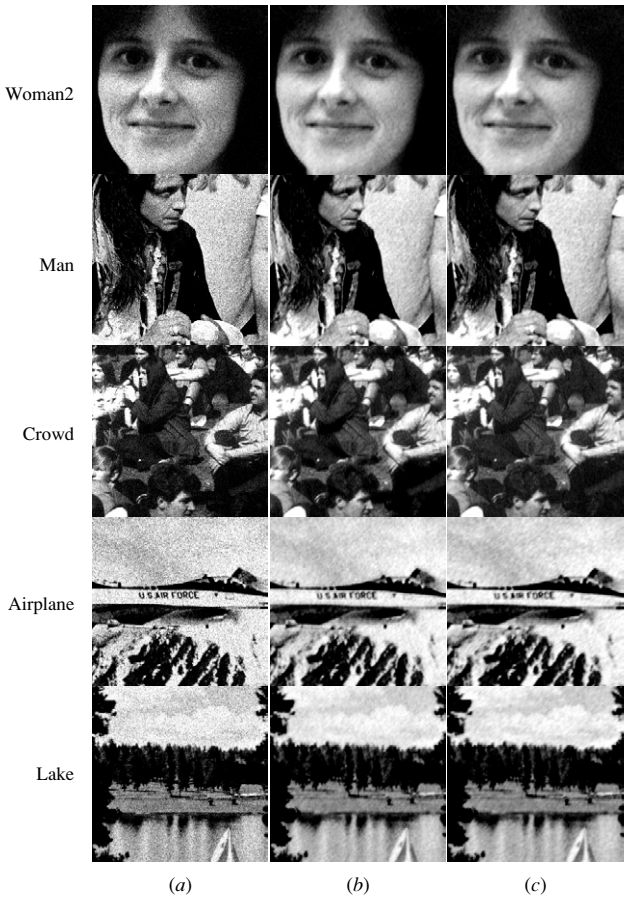
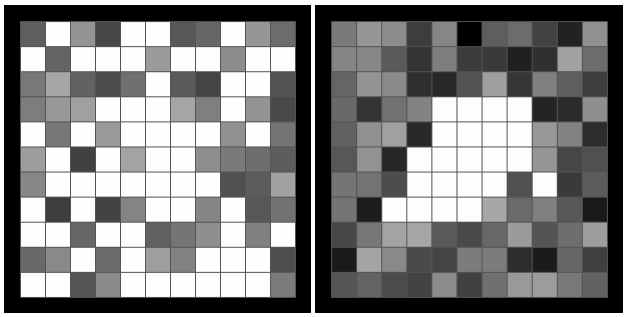


Figure 7. Comparison of the output images of the original NF and those obtained by use of the proposed method for non-training images. (*a*) Input images. (*b*) Output images of the original NF. (*c*) Output images of the NF obtained by use of the proposed method.

The receptive fields determined by use of OBD and the proposed method are shown in figure 8. In the receptive field obtained by use of OBD, the remaining units are dispersed, whereas those in the receptive field obtained by use of the



(*a*) (*b*)

Figure 8. Comparison of the receptive field obtained by use of OBD with that obtained by use of the proposed method for the NF for the reduction of noise in the image shown in figure 5. (*a*) Receptive field obtained by use of OBD. (*b*) Receptive field obtained by use of the proposed method. (See figure 3 for explanation.)

Table 3. Comparison of the performance of the NF obtained by use of the proposed method with that of the NFs obtained by use of conventional methods.

	ISNR (dB)			
Image	Original NF	NF obtained by Hagiwara's method	NF obtained by OBD	NF obtained by the proposed method
Woman2	5.1	6.7	7.6	8.3
Man	3.1	3.7	4.1	4.1
Crowd	1.4	2.2	2.7	3.2
Airplane	6.1	5.8	5.8	6.3
Lake	2.3	3.0	3.7	3.9

proposed method gather around the object pixel (center pixel) in the input region. Because in the reduction of noise in images, the correlation of each pixel with the object pixel would be in inverse proportion to the distance from the object pixel, the receptive field obtained by use of the proposed method seems reasonable. The receptive field obtained by use of the proposed method extends from the lower left to the upper right, probably because the training image (figure 5) contains mainly directional patterns from lower left to upper right.

5. Experiment with a spatiotemporal NF for processing of image sequences

For processing of image sequences (time-varying images), a spatiotemporal NF with a three-dimensional input region that consists of 5 × 5 pixels in each of four consecutive frames was employed. The structure of the spatiotemporal NF was 120–50–1. A medical x-ray image sequence (containing 20 frames, the size of which was 512 × 512 pixels) of the stomach was used in this experiment; such images are used for diagnostic examination for stomach cancer. An example of frames in the sequence is shown in figure 9(*b*). Because the image sequence was acquired at a high x-ray exposure level, the images contained little noise. A noisy image sequence was synthesized from the noiseless image sequence by addition of quantum noise. The object frame (frame 0) in the noisy image sequence used for training is shown in figure 9(*a*). The corresponding noiseless original image was used as a teaching image (figure 9(*b*)). Four thousand five hundred training pixels were extracted from the noisy input and noiseless teaching images. The spatiotemporal NF was trained for 80 000 epochs with these images. The training converged with an MAE of 0.0237.

Table 4. Comparison of the proposed method with conventional methods for spatiotemporal NFs for the reduction of noise in image sequences.

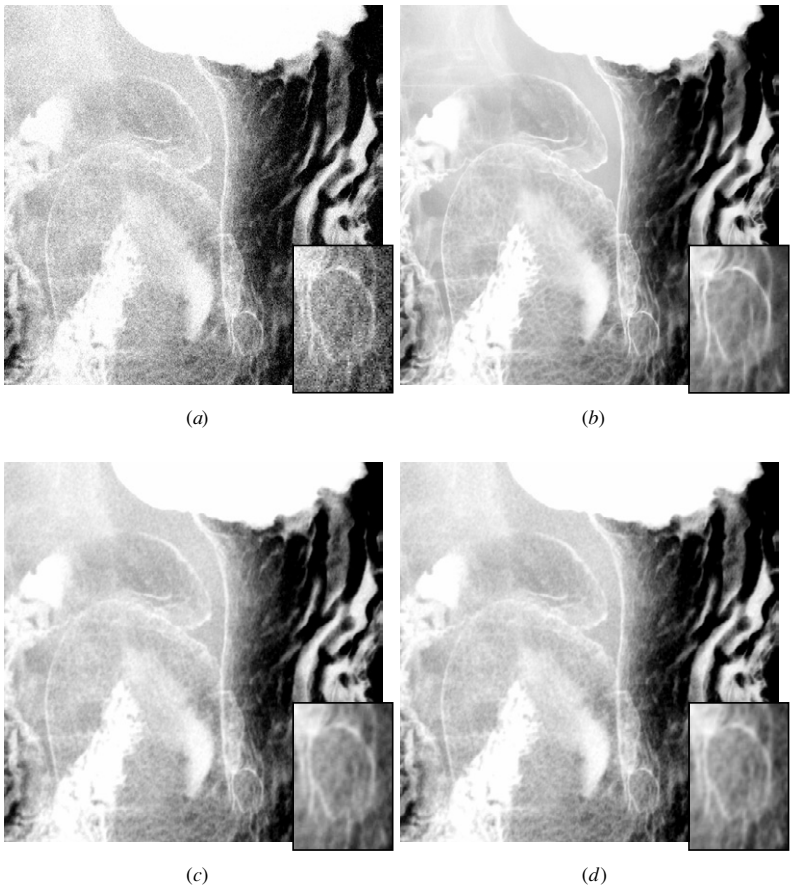
Neural filter	Execution time (h)	Structure	Average ISNR (dB)
Original NF	39.0[a]	125–50–1	1.6
NF obtained by Hagiwara's method	91.6	124–50–1	1.4
NF obtained by OBD	90.2	58–16–1	1.5
NF obtained by the proposed method	73.2	40–9–1	1.6

[a] Note that the execution time of the original NF corresponds to the time for training.

The results of determination of the structures of the NFs are summarized in table 4. The results showed that the performance of the proposed method was superior to that of conventional methods. The output images of the original NF and of the NF obtained by use of the proposed method are very similar, as shown in figures 9(*c*) and (*d*), and as can be seen from the average ISNRs over 20 frames in table 4. The receptive field obtained by use of the proposed method is shown in figure 10. The remaining units gather around the object pixel (center pixel in the region in the object frame). The receptive field obtained by use of the proposed method seems reasonable, because in the reduction of noise in image sequences, the correlation of each pixel with the object pixel in the spatiotemporal input region of the NF would be in inverse proportion to the distance from the object pixel. The remaining units in the receptive field, however, seem to be somewhat dispersed, e.g., the remaining unit in frame 3 is not located at the center, probably because there was some bias in the motion of the stomach wall in the training images. The dispersion in the receptive field would be smaller if a large number of frames was used for training.

6. Discussion

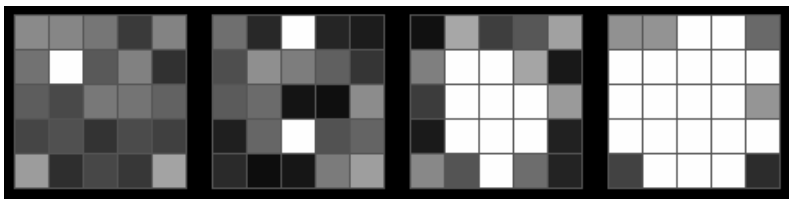
Most studies on determining the structure of an ANN have focused on determining the number of units in hidden layers, because the structure of the input and output layers is determined automatically by a problem to be solved. For example, when an ANN is applied to single-digit number recognition, i.e., from 0 to 9, the input units should be of the matrix size of an input image. The number of output units should be ten. Indeed, conventional methods in group (1) (section 3.3), except for Hagiwara's method (Hagiwara 1989, 1990), used the output value of a unit in the definition of the performance index. Because the output values of units in the input layer in an NF are pixel values of training data, they would not be suited to the performance index of units



(*a*) (*b*)

(*c*) (*d*)

Figure 9. Medical x-ray images (stomach) used for the training of a spatiotemporal NF for the reduction of noise in image sequences, and the output images of NFs. (*a*) Object frame (frame 0) in a noisy input image sequence. (*b*) Noiseless teaching image. (*c*) Output image of the original NF. (*d*) Output image of the NF obtained by use of the proposed method. A region of interest was enlarged and appears on the lower right of each image.



frame 3 frame 2 frame 1 frame 0

Figure 10. Receptive field obtained by use of the proposed method for the spatiotemporal NF for the reduction of noise in the medical image sequence shown in figure 9. (See figure 3 for explanation.)

in the NF. Therefore, Hagiwara's method was selected from group (1) for comparison. In the conventional methods of group (2), training of ANNs with various structures is required. The number of combinations of units in a three-layered NF is directly proportional to the following order:

$$N_{CON} = O(2^{N^{(1)}} \cdot N^{(2)}). \tag{14}$$

A tremendous number of NFs needs to be trained, e.g., $O(1.33 \times 10^{38})$ for the NF in section 4. Therefore, the execution time would be a serious problem. Because the conventional methods in group (3) basically do not remove units directly, but remove weights, the direct removal of units would be suitable for the determination of the receptive field of an NF.

In the experiment in section 3, the number of hidden units determined for the NF trained to obtain the function of a Laplacian filter was five, which is not optimal theoretically (a Laplacian filter can be realized by an NF with one hidden unit). An additional experiment was performed to investigate an adequate number of hidden units. NFs were trained with various numbers of hidden units, where the structure of units in the input layer was the same as the kernel of a Laplacian

filter. The training of NFs with four or fewer hidden units did not converge. This result indicated experimentally that the minimum number of hidden units was five.

I used a three-layered NF in the experiments in this study, because it has been proved theoretically that a three-layered ANN can realize any continuous mapping approximately (Funahashi 1989, Barron 1993). The proposed method can be applied to an NF with more than three layers. Addition of one hidden layer could reduce the overall number of units for the same performance in some applications, but could increase it in other applications.

The receptive fields obtained by use of the proposed method seemed to depend on the training images used. When a different training image is used, the structure of the receptive field would be changed, and it would depend on the patterns in the training image. Because an NF is a supervised nonlinear filter trained with training patterns, it would be reasonable that the characteristics of the trained NF depend on the training patterns. It is interesting to note that this influence of training images is reminiscent of the receptive fields of various simple units in the cat and monkey cerebral cortex, as discovered by Hubel and Wiesel (1962). In the cat and monkey, these biological neural filters are acquired during the critical period just after birth (Blakemore and Cooper 1970). When a cat is limited to seeing only vertical patterns during the critical period, the cat does not respond to stimuli of horizontal light. The cat grows up to be permanently unable to recognize horizontal patterns.

7. Conclusion

A method for determining the receptive field and the structure of hidden layers of an NF was developed and evaluated. By use of the proposed method, redundant units were able to be removed from NFs, while the performance of the NFs was maintained. Experimental results suggested that the proposed method could determine a reasonable receptive field for a given image-processing task.

Acknowledgments

The author is grateful to K Ishikawa and to S Ikeda of the Hitachi Medical Corporation for preparation of the medical images, and to Ms E F Lanzl for improving the manuscript.

References

Arakawa K and Harashima H 1990 A nonlinear digital filter using multi-layered neural networks *Proc. IEEE Int. Conf. Communications* vol 2 pp 424–8

Banham M R and Katsaggelos A K 1997 Digital image restoration *IEEE Signal Process. Mag.* **14** 24–41

Barron A R 1993 Universal approximation bounds for superpositions of a sigmoidal function *IEEE Trans. Inf. Theory* **39** 930–45

Blakemore C and Cooper G F 1970 Development of the brain depends on the visual environment *Nature* **228** 477–8

Brailean J C, Kleihorst R P, Efstratiadis S, Katsaggelos A K and Lagendijk R L 1995 Noise reduction filters for dynamic image sequences: a review *Proc. IEEE* **83** 1270–91

Castellano G and Fanelli A M 2000 Variable selection using neural-network models *Neurocomputing* **31** 1–13

Castellano G, Fanelli A M and Pelillo M 1997 An iterative pruning algorithm for feedforward neural networks *IEEE Trans. Neural Netw.* **8** 519–31

Chandrasekaran H, Chen H and Manry M T 2000 Pruning of basis functions in nonlinear approximators *Neurocomputing* **34** 29–53

Chauvin Y 1989 A back-propagation algorithm with optimal use of hidden units *Adv. Neural Inf. Process.* **1** 519–26

Dumitras A and Kossentini F 2000 Feedforward neural network design with tridiagonal symmetry constraints *IEEE Trans. Signal Process.* **48** 1446–55

Fogel D B 1991 An information criterion for optimal neural network selection *IEEE Trans. Neural Netw.* **2** 490–7

Funahashi K 1989 On the approximate realization of continuous mappings by neural networks *Neural Netw.* **2** 183–92

Gorodkin J, Hansen L, Krogh A, Svarer C and Winther O 1993 A quantitative study of pruning by optimal brain damage *Int. J. Neural Syst.* **4** 159–69

Hagiwara M 1989 Supervised learning with artificial selection *Proc. Int. Joint Conf. Neural Networks* vol 2 pp 611–6

Hagiwara M 1990 Novel back propagation algorithm for reduction of hidden units and acceleration of convergence using artificial selection *Proc. Int. Joint Conf. Neural Networks* vol 2 pp 625–30

Hassibi B and Stork D G 1993 Second order derivatives for network pruning: optimal brain surgeon *Adv. Neural Inf. Process.* **5** 164–71

Hubel D H and Wiesel T N 1962 Receptive fields, binocular interaction and functional architecture in the cat's visual cortex *J. Physiol.* **160** 106–54

Ishikawa M 1994 Structural learning and its applications to rule extraction *Proc. Int. Conf. Neural Networks* 354–9

Ishikawa M 1996 Structural learning with forgetting *Neural Netw.* **9** 509–21

Ishikawa M and Uchida H 1992 A structural learning of neural networks based on an entropy criterion *Proc. Int. Joint Conf. Neural Networks* vol 2 pp 375–80

Ji C, Snapp R R and Psaltis D 1990 Generalizing smoothness constraints from discrete samples *Neural Comput.* **2** 188–97

Kameyama K and Kosugi Y 1991 Neural network pruning by fusing hidden layer units *Trans. IEICE* **74** 4198–204

Kurita T 1990 A method to determine the number of hidden units of three layered neural networks by information criteria *Trans. IEICE* **73** 1872–8

LeCun Y, Denker J S and Solla S A 1990 Optimal brain damage *Adv. Neural Inf. Process.* **2** 598–605

Murata N, Yoshizawa S and Amari S 1994 Network information criterion—determining the number of hidden units for an artificial neural network model *IEEE Trans. Neural Netw.* **5** 865–72

Nowlan S J and Hinton G E 1992 Simplifying neural networks by soft weight-sharing *Neural Comput.* **4** 473–93

Reed R 1993 Pruning algorithms—a survey *IEEE Trans. Neural Netw.* **4** 740–7

Rumelhart D E, Hinton G E and Williams R J 1986a Learning internal representations by error propagation *Parallel Distributed Processing* vol 1 (Cambridge, MA: MIT Press) pp 318–62

Rumelhart D E, Hinton G E and Williams R J 1986b Learning representations of back-propagation errors *Nature* **323** 533–6

Sietsma J and Dow R J F 1991 Creating artificial neural networks that generalize *Neural Netw.* **4** 67–9

Suzuki K, Horiba I, Ikegaya K and Nanki M 1995 Recognition of coronary arterial stenosis using neural network on DSA system *Syst. Comput. Japan* **26** 66–74

Suzuki K, Horiba I and Sugie N 2002a Efficient approximation of a neural filter for quantum noise removal in x-ray images *IEEE Trans. Signal Process.* **50** 1787–99

Suzuki K, Horiba I and Sugie N 2003 Neural edge enhancer for supervised edge enhancement from noisy images *IEEE Trans. Pattern Anal. Mach. Intell.* **25** 1582–96

Suzuki K, Horiba I, Sugie N and Nanki M 1998a Noise reduction of medical x-ray image sequences using a neural filter with spatiotemporal inputs *Proc. Int. Symp. Noise Reduction for Imaging and Communication Syst.* pp 85–90

Suzuki K, Horiba I, Sugie N and Nanki M 1998b A recurrent neural filter for reducing noise in medical x-ray image sequences *Proc. Int. Conf. Neural Inf. Process.* **1** 157–60

Suzuki K, Horiba I, Sugie N and Nanki M 2002b Neural filter with selection of input features and its application to image quality improvement of medical image sequences *IEICE Trans. Inf. Syst.* **85** 1710–8

Suzuki K, Horiba I, Sugie N and Nanki M 2004 Extraction of left ventricular contours from left ventriculograms by means of a neural edge detector *IEEE Trans. Med. Imaging* **23** 330–9

Weigend A S, Rumelhart D E and Huberman B A 1991 Generalization by weight-elimination applied to currency exchange rate prediction *Proc. Int. Joint Conf. Neural Networks* **1** pp 837–41

Yin L, Astola J and Neuvo Y 1993 A new class of nonlinear filters—neural filters *IEEE Trans. Signal Process.* **41** 1201–22