The Performance of the Neocognitron with Various S-Cell and C-Cell Transfer Functions

David R. Lovell & Ah Chung Tsoi*
Intelligent Machines Laboratory, Department of Electrical Engineering,
University of Queensland, Queensland 4072, Australia

April 14, 1992

Abstract

When a neural network solution to a problem (*e.g.* handwritten character recognition) is proposed, it is important to know if the structure of the network and the function of the component neurons are well suited to the task. Recent research has examined the *structure* of Fukushima's neocognitron and the effect that it has on the classification of distorted input patterns. We present results which assess the classification performance of the neocognitron when the *function* of the component neurons is altered. The tests we describe demonstrate that using S-cells with a sigmoidal transfer function and modified activation function significantly enhances the classification performance of the neocognitron.

1 Introduction

Fukushima's neocognitron [1, 2, 3, 4] has received attention over the past decade as a partially shift invariant [5, 6], distortion tolerant handwritten character classifier. The performance of novel structures of the neocognitron has been investigated to a certain extent [7]. However, we are not aware of any studies that assess the effectiveness of different implementations of the component S and C-cells in the neocognitron.

In this paper, we compare the classification performance of Fukushima's neocognitron to identically structured networks that employ modified S and C-cells. Empirically, Fukushima's original implementation performs well but it is shown that the classification ability and noise tolerance of the network can be significantly improved by the use of a sigmoidal transfer function in the S-cell.

To employ a variety of S-cell transfer functions it was necessary to modify the activation function of those cells. The modification we propose allows supervised training of the S-cells to take place with a single presentation of all training vectors.

After our work on the performance of the modified versions of the neocognitron was completed, Hildebrandt [8] published a revealing analysis of the function of the S-cell as well as a method for tuning the *selectivity parameter* of those cells. We discuss the relation between our results and the work presented by Hildebrandt and show that the parameter tuning algorithm he presents does not improve the classification performance of the neocognitron.

Initially, the experiments we performed aimed to find out if the modified S and C-cell networks could yield better results than Fukushima's original neocognitron. We investigated 14 modified versions of the neocognitron and out of these, a network using sigmoidal S-cells and weighted average C-cells exhibited the best classification performance, significantly better than that of the original network. Once we had proved our conjecture of enhanced performance through modification of the component cells, the parameters of the *sigmoid + weighted average* network were tuned so as to maximize the performance of that network. The resulting network exhibited a correct classification rate of 72.5%, compared to 52.5% for the original neocognitron.

To place our experiments in the proper context, Section 2 gives a general overview of the operation of the neocognitron. Sections 3 and 4 examine, and propose modifications to, the functions of the S and C-cells, respectively. In Section 5, we outline the methods used to compare the performance of a number of versions of the neocognitron and present the results of our experiment in Section 6. The remainder of the paper discusses the results and their significance.

*The first author acknowledges the support of an Australian Postgraduate Research Award. The second author acknowledges support from the Australian Research Council. The comments and suggestions of T. Downs P. Bartlett and S. Young are gratefully acknowledged.

2 Basic Operation of the Neocognitron

The neocognitron performs classification of input through a succession of functionally equivalent stages [1, 2, 3, 4]. Each stage extracts appropriate features from the output of the preceding stage and then forms a compressed representation of those extracted features. The compressed representation preserves the spatial location of the extracted features and becomes the input to the following stage. Classification is achieved by steadily extracting and compressing feature representations until the input is reduced to a vector, each element of which corresponds to a similarity measure between the input and the different classes of input that the neocognitron has been trained to classify.

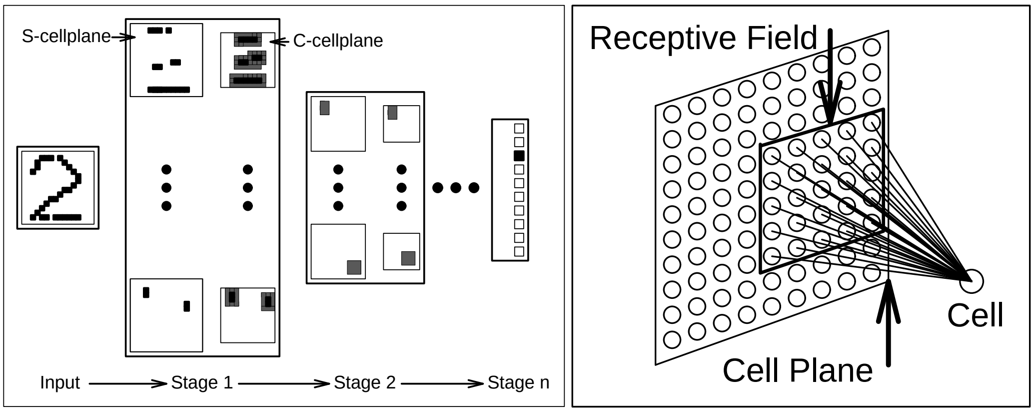

Figure 1: (a) The basic structure of the neocognitron (b) general structure of S and C cell planes.

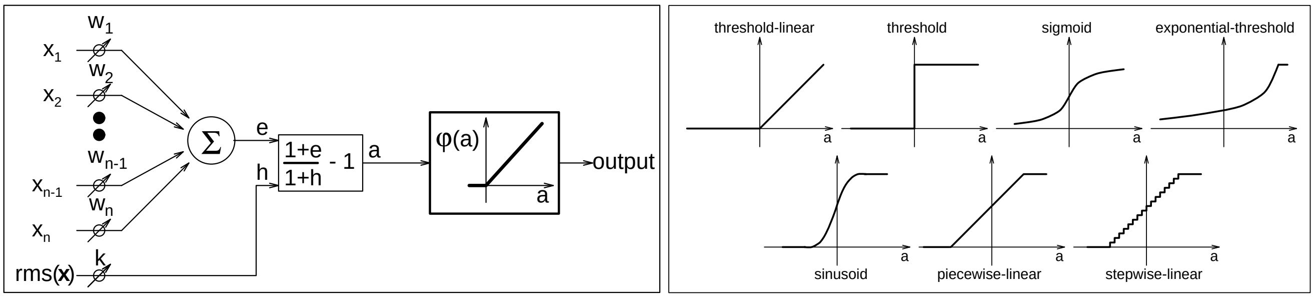

Figure 2: (a) The structure of an S-cell (b) The transfer functions used in this experiment.

Figure 1(a) shows the structure of the neocognitron. Feature extraction is performed by arrays of *S-cells* (called *S-planes*) that have been trained to respond to certain features that characterize the input patterns. Each S-cell is connected to a rectangular region of cells (known as a *receptive field*) in the preceding stage. The receptive fields (see Figure 1(b)) of the S-cells in the array uniformly cover the input cell plane. In any S-plane, the connection strengths between each cell and its receptive field are replicated. This ensures a translationally invariant response to features in the input cell plane.

The *C-cell* planes (*C-planes*) shown in Figure 1(a) compress the activity of the previous S-planes into a smaller representation. In doing this, the C-cells provide a *degree* of translational invariance [5, 6] to the responses of the preceding S-cells. Ultimately, this compression of activity reaches the stage where the input pattern is represented by a set of single C-cells, each corresponding to an input class that the neocognitron has been trained to recognize. At this stage, the C-cell with the highest activity represents the class to which the input belongs.

3 S-Cells: The Feature Extractors

The fundamental operation performed by an S-cell (see Figure 2) is a *weighted sum of inputs*. The connections between the S-cell and its receptive field inputs are the only adaptive elements in the neocognitron and the learning rule employs local Hebbian reinforcement.

The neocognitron can be trained in either a supervised or an unsupervised fashion. Since supervised training is completely reproducible and deterministic, it was used in the comparative experiments detailed in Sections 5 and 6. The reader is referred to [2] for a description of unsupervised learning in the neocognitron.

Supervised learning involves training each S-plane in a stage to respond to certain input features. Different S-planes are trained to extract different features. During training, an S-cell is exposed to the inputs that it is required to extract and once training is completed, the weight vector of that S-cell is equal to the sum of the inputs that have appeared within its receptive field.

Each S-cell also receives an inhibitory signal proportional to the root mean square (rms) activity present in its receptive field. The amount of inhibition that this signal causes is adjusted by the learning rule (which alters b_ℓ in Equation (2) below) to maximize the S-cell's output when a feature similar to the features the cell has been trained to recognize is presented.

The behaviour of an S-cell can be mathematically formulated as a function $\varphi(\cdot)$ of a cell's activation:

$$\begin{aligned} output(\boldsymbol{x}) &= r_\ell \cdot \varphi(a(\boldsymbol{x})), && (1) \\ a(\boldsymbol{x}) &= \frac{1+\boldsymbol{x}^T\boldsymbol{w}}{1+\frac{r_\ell}{1+r_\ell}\cdot b_\ell \cdot rms(\boldsymbol{x})} - 1, && (2) \end{aligned}$$

where $\boldsymbol{x}$ is the vector of activities present at the receptive field input, $\boldsymbol{w}$ is the vector of weights learned by the S-cell and r_ℓ is the *selectivity parameter*. The value of r_ℓ can be set by the experimenter or by a closed-form training algorithm (presented in [8]), and this value determines how closely the cell's input must correspond to the inputs it has been trained with in order to elicit a response. The root mean square activity of the input to the S-cell is defined by

$$rms(\boldsymbol{x}) = \sqrt{\sum_{i=1}^{N} c_i x_i^2}, \qquad (3)$$

where the vector $\boldsymbol{c} = [c_1, \ldots, c_N]^T$ describes a Gaussian kernel that serves to accentuate inputs towards the centre of the cell's receptive field, as well as implementing the arithmetic mean of the inputs. As noted previously, b_ℓ is a factor set by the learning rule to maximize the cell's response to any training feature.

3.1 Modifying the S-Cell Transfer Function

This paper is concerned with the effect that the function $\varphi(\cdot)$ has on the overall classification accuracy of the neocognitron. In the original neocognitron, Fukushima made $\varphi(\cdot)$ a threshold linear function of the form

$$\varphi_{ThreshLin}(a) = \begin{cases} 0 & \text{if } a < 0 \\ a & \text{if } 0 \leq a, \end{cases} \qquad (4)$$

thereby ensuring that the output of an S-cell is non-zero *only when the excitatory component of its input exceeds the inhibitory component.* This preserves the separation between *figure* and *ground* in the neocognitron. The ground, in this context, denotes the background in which the character (the figure) is presented.

To investigate the effectiveness of transfer functions other than $\varphi_{ThreshLin}(\cdot)$, we examined six functions (see Figure 2(b)): the threshold and the sigmoid, which are commonly used in connectionist networks, a sinusoidal approximation of the sigmoid, a piecewise-linear transfer function and its stepwise approximation, and an exponentially tailed threshold function.

$$\begin{aligned} \varphi_{Thresh}(a) &= \begin{cases} 0 & \text{if } a < 0 \\ 1 & \text{if } 0 \leq a, \end{cases} && (5) \\ \varphi_{Sig}(a) &= \frac{1}{1+e^{-\beta a}}, && (6) \\ \varphi_{Sin}(a) &= \begin{cases} 0 & \text{if } a < -1 \\ \frac{1}{2}(\sin a\frac{\pi}{2} + 1) & \text{if } -1 \leq a < 1 \\ 1 & \text{if } 1 \leq a, \end{cases} && (7) \\ \varphi_{PceLin}(a) &= \begin{cases} 0 & \text{if } a < -1 \\ \frac{1}{2}(a+1) & \text{if } -1 \leq a < 1 \\ 1 & \text{if } 1 \leq a, \end{cases} && (8) \end{aligned}$$

$$\varphi_{StepLin}(a) = \begin{cases} 0 & \text{if } a < -1 \\ \frac{1}{2}(\frac{\lfloor 10a \rfloor}{10} + 1) & \text{if } -1 \leq a < 1 \\ 1 & \text{if } 1 \leq a, \end{cases} \tag{9}$$

$$\varphi_{ExpThresh}(a) = \begin{cases} 2^{a-1} & \text{if } a < 1 \\ 1 & \text{if } 1 \leq a. \end{cases} \tag{10}$$

The threshold function (Equation (5)) may be directly applied to the activation in Equation (2), but the transfer functions in Equations (6)–(10) require a modified activation function[1] before they can be incorporated in the neocognitron.

For transfer functions like Equations (6)–(10) (odd functions with $\varphi(0) = 0.5$) we have devised the following activation function

$$a(\boldsymbol{x}) = \frac{\varepsilon_{sat}\left(\frac{|\boldsymbol{x}|}{|\boldsymbol{x}_{avg}|}\right)\cos\theta - \mathit{threshold}}{1 - \mathit{threshold}}. \tag{11}$$

where

$$\cos\theta = \frac{\boldsymbol{x}^T\boldsymbol{w}}{|\boldsymbol{x}||\boldsymbol{w}|}. \tag{12}$$

θ is the angle (in N-dimensional space where $\boldsymbol{x} \in \mathbf{R}^N$) between the input vector $\boldsymbol{x}$ and the weight vector $\boldsymbol{w}$. The weight vector is the sum of all T training patterns that the S-cell is to respond to

$$\boldsymbol{w} = \sum_{i=1}^{T} \boldsymbol{x}_i, \tag{13}$$

where $\boldsymbol{x}_i$ represents the i^{th} training vector for that S-cell. If $\boldsymbol{x}$ and $\boldsymbol{w}$ are similar in direction then $\cos\theta$ will approach 1. The *threshold* sets the value of $\cos\theta$ that is required to make the activation positive and satisfies $0 \leq \mathit{threshold} \leq 1$.

The function $\varepsilon_{sat}(\cdot)$ causes the S-cell to respond maximally to input features only when they are of comparable size to the average size of the training vector, $|\boldsymbol{x}_{avg}|$. We define $|\boldsymbol{x}_{avg}|$ as

$$|\boldsymbol{x}_{avg}| = \frac{1}{T}\sum_{i=1}^{T} |\boldsymbol{x}_i|. \tag{14}$$

$\varepsilon_{sat}(\cdot)$ is a simple linear-threshold function

$$\varepsilon_{sat}(z) = \begin{cases} z & z \leq 1 \\ 1 & 1 < z, \end{cases} \tag{15}$$

and prevents the S-cell from responding to small, spurious inputs that are similar in direction to the weight vector $\boldsymbol{w}$. For the case of a zero length input vector, $\boldsymbol{x} = \mathbf{0}$, (for which $\cos\theta$ is undefined), we define the product $\varepsilon_{sat}(|\,\mathbf{0}\,| / |\,\boldsymbol{x}_{avg}\,|)\cos\theta = 0$.

This formulation of the modified activation function is similar in some respects to the approach taken by Fukushima in [2]. It should be noted that an S-cell using Equation 11 can be trained by a single presentation of the training vectors. There is no learning rate parameter associated with the S-cell and no ambiguity as to how many presentations of the training vectors are required for adequate learning.

4 C-Cells: Compressing the Representation

The input to a C-cell from its receptive field is a subsampling of the activity present in the preceding S-plane. By subsampling this activity, a compressed representation of the S-plane output is obtained. C-cells also *blur* the activations of the preceding S-planes by performing a weighted sum of inputs, this time using fixed weights that

[1] Suppose there was no input to an S-cell with a sigmoid activation transfer function. The activation (in Equation (2)) would equal zero but with the sigmoid transfer function, this would result in an output of 0.5 and *figure-ground* separation would be lost.

describe a Gaussian kernel. If we denote the subsampled input as a vector $\boldsymbol{x}$ and the Gaussian kernel as $\boldsymbol{w}$ then the activation of a C-cell in Fukushima's original description of the neocognitron can be written as

$$a(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x}. \tag{16}$$

This *weighted mean* is then passed through a transfer function that limits the output of the C-cell to $[0, 1)$,

$$\psi_{Mean}(a) = \frac{a}{1+a}. \tag{17}$$

4.1 Modifying the C-Cell Transfer Function

Researchers have remarked that the blurring of S-plane activity by the C-cells is important in allowing the neocognitron to be tolerant of a considerable degree of input distortion [4, 5, 6].

Algorithms for blurring or *spreading* the activation include:

- *2-D convolution algorithms*, such as Equation (16) in Fukushima's original C-cell description
- *morphological algorithms ie* dilation of the incoming activations
- *order statistic filtering methods*, including median filtering

Since order statistic filters were straightforward to incorporate into the structure of the neocognitron, we developed a modified *ranked order filter* [11] version of the C-cell. This allows us to compare the effectiveness of Fukushima's approach with another standard blurring method. The output of this modified C-cell is given by

$$\psi_{Max}(\boldsymbol{x}) = \max_i \; x_i w_i. \tag{18}$$

$\boldsymbol{x} = [x_1, \ldots, x_N]^T$ is again the subsampled input vector and $\boldsymbol{w} = [w_1, \ldots, w_N]^T$ the Gaussian kernel. We refer to Equation (18) as a *weighted max* operation.

5 The Experimental Framework

The performance of a neocognitron that has undergone supervised learning is primarily dependent upon the feature set with which it was trained and the general structure of the network. The same feature set and network structure (as described in [2]) were used throughout this experiment so that a proper comparison of performance statistics could be obtained.

We performed three sets of tests. Test I estimates the classification performance of networks on a set of digits; Test II takes the four networks with the highest classification performance from Test I and examines their performance on data corrupted with salt and pepper noise; Test III looks at the performance of those four networks after they have been trained using Hildebrandt's closed-form algorithm [8].

5.1 Network Parameters

Versions of the neocognitron that used the threshold-linear and threshold transfer functions were implemented with the same parameters detailed in [4]. In all networks used in this experiment, the Gaussian kernels used to weight the inputs to the C-cells were calculated using the method and parameters described [4]. The value of β in Equation (6) was set to 4 to allow activation to range over the linear and nonlinear parts of the sigmoid.

In Tests I and II, for implementations of the neocognitron that used the S-cell transfer functions of Equations (6)–(10), the *threshold* parameters were given the values 0.68, 0.9, 0.7, 0.7 over stages 1 to 4 respectively.

The *threshold* parameters were *not* extensively tuned to maximize the network performance since, at the time that those tests were performed, there was no quantitative method for doing this. After Tests I and II were substantially complete, we learned of a feature dependent approach to setting the *threshold* parameter in neocognitron style networks [8, section III.D]. The performance of neocognitron style networks trained using this closed-form algorithm is investigated in Test III.

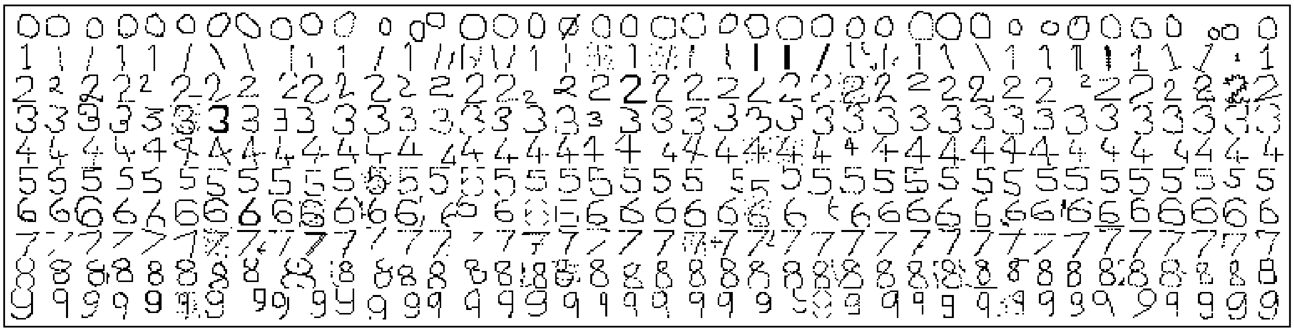

Figure 3: The test set of digits.

6 Experimental Results

To estimate the success rate of a classifier to within 5% accuracy, with 95% confidence, it is necessary to perform at least 385 trials [10]. We performed 400 trials on each network, using the digits[2] shown in Figure 3. The results shown in the tables below have been partitioned into three categories: the percentage of trials in which the input was correctly classified, the trials in which the input was misclassified and the trials in which the input was unclassified (*i.e.* there was no activity in any of the final stage units).

Test I			
S-Cell Transfer Function + C-Cell Transfer Function	% Classified	% Misclassified	% Unclassified
sigmoid + weighted average	62.50%	37.50%	0.00%
*sigmoid + weighted average	53.00%	23.50%	23.50%
threshold-linear + weighted average	52.50%	26.25%	21.25%
exponential-threshold + weighted max	47.75%	52.25%	0.00%
sigmoid + weighted max	45.75%	54.25%	0.00%
sinusoid + weighted average	41.00%	22.50%	36.50%
stepwise-linear + weighted average	40.00%	35.75%	24.25%
piecewise-linear + weighted average	39.50%	46.00%	14.50%
threshold-linear + weighted max	34.25%	33.00%	32.75%
threshold + weighted average	33.75%	25.50%	40.75%
piecewise-linear + weighted max	31.50%	12.50%	56.00%
threshold + weighted max	26.50%	33.50%	40.00%
sinusoid + weighted max	24.25%	12.50%	63.25%
stepwise-linear + weighted max	19.50%	7.25%	73.25%
exponential-threshold + weighted average	10.00%	90.00%	0.00%

Table 1: Classification performance statistics for various implementations of the neocognitron.

The high classification rate achieved by the *sigmoid + weighted average* network is counterbalanced by the increase in misclassification. Transfer functions that are always non-zero (such as Equations (6) and (10)) leave no input unclassified when class is chosen on the basis of highest output activity. If we impose the requirement that the highest activity should also exceed a threshold δ for positive identification of input, choosing $\delta = 0.15$ gives the performance statistics indicated by *.

7 Discussion

Before we discuss the results obtained, let us examine the significance of the modifications we have made to the S and C-cells. Fukushima's original method of training an S-cell involved repeated presentations of the training vectors and required a learning rate parameter q_ℓ to be set to an appropriate value. In his later work [4] Fukushima stated that if the value of q_ℓ was "large enough" supervised training of each S-cell could take place with only one

[2] This test set of digits is available via email request to the first author at `dlovell@s1.elec.uq.oz.au`

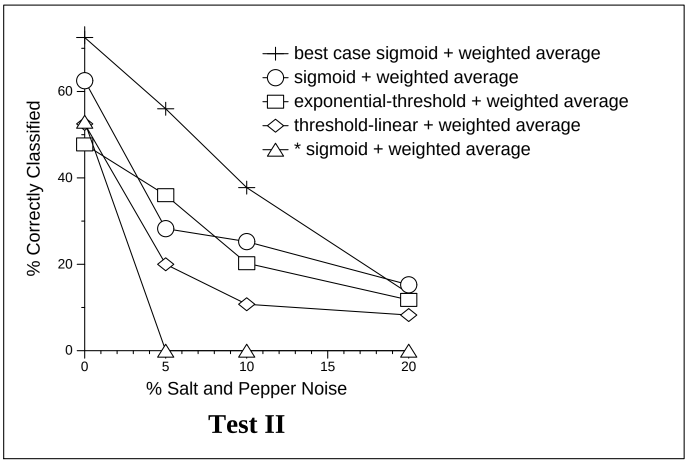

Figure 4: Classification performance statistics of the top four models from Table 1 for input corrupted with salt and pepper noise. The "best case" version of the *sigmoid + weighted average model* is discussed in Section 7.3.

Test III			
S-Cell Transfer Function + C-Cell Transfer Function	% Classified	% Misclassified	% Unclassified
sigmoid + weighted average	20.25%	79.75%	0.00%
*sigmoid + weighted average	0.00%	9.25%	90.75%
threshold-linear + weighted average	3.25%	10.00%	86.75%
exponential-threshold + weighted max	13.50%	86.50%	0.00%

Table 2: Classification performance statistics for implementations of the neocognitron trained using Hildebrandt's closed-form algorithm.

presentation of each training vector. It is not immediately apparent what constitutes a large enough value of q_ℓ (although Fukushima suggests 10^4), so we feel that it is an advantage that our formulation of the S-cell requires no learning rate parameter to be set, while guaranteeing that only single presentations of the training vectors are needed for training to proceed.

Our formulation of the S-cell's activation function (Equation (11)) makes explicit the relation between the S-cell's parameters and its response. We feel that this approach has more intuitive appeal than that of Fukushima (Equation 2).

As for our modified C-cell, the weighted max algorithm (Equation (18)) offers a more computationally efficient method of blurring and compressing feature representations in the neocognitron.

7.1 Indications of the Results

Test I (see Table 1) shows that it is possible to obtain a statistically significant improvement in classification performance by reformulating the behaviour of the S-cells in the neocognitron. As we noted in Section 6, the improvement achieved by the sigmoidal transfer function is accompanied by a corresponding increase in the misclassification of input (due to the fact that the input is *always* categorized).

It could, perhaps, be argued that this amounts to no improvement at all. However, the aim of this experiment was to establish whether or not the performance of the neocognitron could be enhanced by alterations to the component neurons alone. The results of Test I show that a coarsely tuned network employing sigmoidal S-cells can achieve a correct classification rate that is significantly higher than that of the original neocognitron[3]. We will show later that our results can be further improved (by adjusting the *threshold* parameters of the S-cells) to the point where the misclassification rate of the sigmoidal network is similar to that of Fukushima's neocognitron.

Our modification to the C-cells in the network was less effective. The generally poor performance of networks

[3] We presume that the parameters of the original neocognitron (as described in [4]) were set so that handwritten digits could be reasonably well classified. This implies that we are making a fair comparison between our modified networks and a neocognitron with close to maximum classification ability.

using the weighted max blurring algorithm implies that an effective blurring algorithm should respond to the average activity presented to a C-cell rather than the largest input alone. It would seem that the role of the C-cells in the neocognitron is more subtle than simply spreading the activation present in preceding layers.

Test II (see Figure 4) shows the performance of the four networks with the highest Test I classification rates when the input data is corrupted with salt and pepper noise. All but one of these networks showed graceful degradation of performance in the presence of noise. Adding a threshold δ to the output of the sigmoidal S-cell neocognitron gives similar performance to Fukushima's original network *only* in a low noise environment.

7.2 Hildebrandt's Closed-Form Training Algorithm

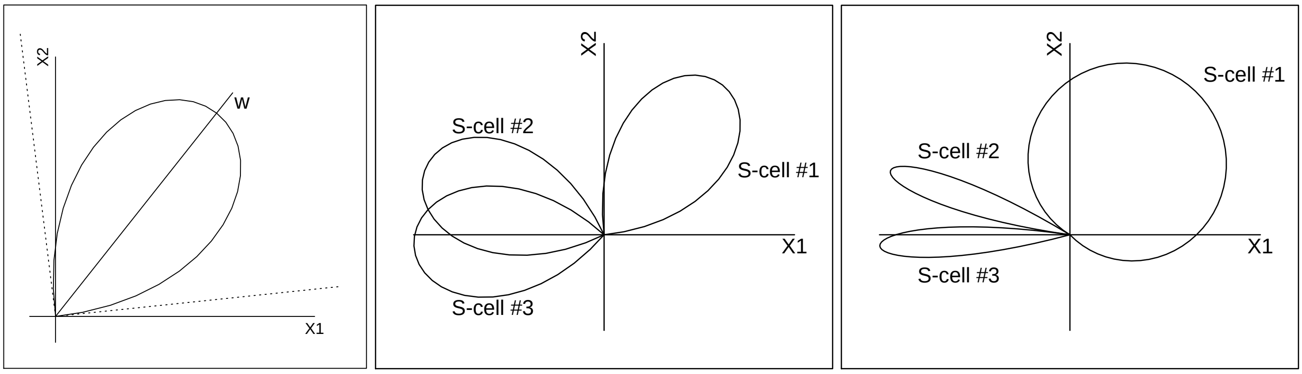

Figure 5: (a) The direction cosine of an S-cell response in 2 dimensions, DC($\boldsymbol{x}$), is given by the projection of the output onto the unit $\boldsymbol{x}$ vector: DC($\boldsymbol{x}$) = $output(\boldsymbol{x})(\boldsymbol{x}/|\boldsymbol{x}|)$ (b) Three S-cell response curves (c) Those response curves after optimal closed-form training.

Our third test was performed to evaluate the closed-form training algorithm for the neocognitron proposed by Hildebrandt [8, section III.D]. The theoretical basis of this algorithm allows the selectivity parameter r of an S-cell to be chosen in an optimal manner. As we point out in [12], however, the optimality of this method *does not imply high classification performance.* We shall briefly outline Hildebrandt's method using the notation of Section 3 and describe the difficulty with its practical application.

The correlation function implemented by an S-cell can be written

$$s(\boldsymbol{x}) = \frac{\boldsymbol{x}^T\boldsymbol{w}}{|\,\boldsymbol{x}\,||\,\boldsymbol{w}\,|}, \tag{19}$$

(see Equation (12), where $\boldsymbol{x}$ is the input and $\boldsymbol{w}$ is the weight vector of the S-cell. To define the behaviour of an S-cell with a zero length input vector, Fukushima employs a modified correlation function

$$s'(\boldsymbol{x}) = \frac{1+\boldsymbol{x}^T\boldsymbol{w}}{1+|\,\boldsymbol{x}\,||\,\boldsymbol{w}\,|}, \tag{20}$$

which is done under the assumption that $|\,\boldsymbol{x}\,||\,\boldsymbol{w}\,| \gg 1$ so that $s(\boldsymbol{x}) \approx s'(\boldsymbol{x})$. Hildebrandt shows that the output of an S-cell (Equation (2)) can be defined in terms of Equation (20) as

$$output(\boldsymbol{x}) = \varphi_{ThreshLin}((1+r_\ell)s'(\boldsymbol{x}) - r_\ell). \tag{21}$$

Note that this equation can be rearranged into a form similar to Equation (11) if we substitute $s'(\boldsymbol{x})$ for $\cos\theta$ and choose $threshold = r_\ell/(1+r_\ell)$.

If we plot the locus of the *direction cosine* of an S-cell's response to a two dimensional input we get a curve like that of Figure 5(a). This direction cosine shows how the response of the S-cell varies depending on the direction of the input vector $\boldsymbol{x}$. Equation (21) defines the *acceptance region* of input space that an input vector $\boldsymbol{x}$ must lie within in order to elicit a response from the S-cell. The acceptance region is conical with its apex at the origin and its axis along the weight vector $\boldsymbol{w}$. The boundary of the acceptance region is at an angle of $\theta = \arccos r_\ell/(r_\ell+1)$ from the region's axis (shown as the two dotted lines in Figure 5(a)).

Each stage of the neocognitron contains a number of S-planes, each of which can be described by the direction of its weight vector and the value of its selectivity parameter r_ℓ. A two dimensional example of the direction cosines of the responses of three S-planes can be seen in Figure 5(b). Put simply, Hildebrandt's optimal closed form training algorithm adjusts the thresholds of each of the S-cells in a given layer so that the acceptance regions of these cells are as large as possible without overlapping. This arrangement is optimal in the sense that it maximizes the generalization of the S-cells (*i.e.* coverage of input space) without compromising the discriminatory ability of those cells. Figure 5(c) shows the response of S-cells that have had their selectivities adjusted in this manner.

To test the practical merit of this algorithm, we applied Hildebrandt's methods to the training of the four networks with the highest Test I classification performances. The results obtained in Test III are shown in Table 2. In order to understand the poor performance of the networks in Test III, we examined their response to the patterns that they were trained to extract. It was found that the networks were frequently not responding to certain original training vectors.

The reason for this "training vector rejection" stems from the fact that weight vector of an S-cell, $\boldsymbol{w}$, is the vector sum of all the training vectors for that cell. This does not imply that $\boldsymbol{w}$ is in the same direction as any of the training vectors. Hence, it is possible for the selectivity of the S-cell to be adjusted by the closed-form algorithm so that one or more training vectors lie outside the acceptance region for that cell.

So, although the closed-form training algorithm produces a network that is, in a sense, optimal, it does not follow that this network will have high classification performance. No such suggestion was made by Hildebrandt. We felt that it was important to critically assess the practical value of this approach. The adjustment of the selectivity parameters in the neocognitron thus remains an open problem.

7.3 Best Performance Statistics Obtained

The aim of the experiments described here was to observe how modifications to the S and C-cells in the neocognitron affected the classification performance of the network. Our intention was to see if the performance of Fukushima's original implementation could be improved upon. Once it was demonstrated that the *sigmoid + weighted average* version of the neocognitron could correctly classify data significantly better than the original network, we decided to run one more test to see whether that result could be improved by different settings of the *threshold* parameters in the network.

The *threshold* parameter in each layer was adjusted so that the response of any S-cell in that layer to one of its training vectors was close to unity. The resulting *threshold* parameters were: 0.68, 0.8, 0.6 and 0.5 for layers 1-4, respectively. Using the same 400 digits as in the other tests, we obtained rates of 72.50% correct classification and 27.50% misclassification.

Since the misclassification rate is now reasonably close to that of Fukushima's neocognitron, we may conclude that the sigmoidal S-cell is a better implementation than that originally proposed. Figure 4 shows that this best case version of the neocognitron also gives superior performance in noisy conditions.

The superior performance of the sigmoidal S-cell network is in agreement with the performance of the mammalian visual system (which uses cells whose transfer functions are often empirically modelled by the sigmoid curve). In a paper describing the precursor to the neocognitron [13], Fukushima mentions the relationship between the transfer function of what he later terms the S-cell and the input-output function of a biological sensory neuron. Our explicit use of the sigmoidal transfer function therefore has strong biological motivation and is appropriate in a network whose structure and implementation has been heavily influenced by models of the mammalian visual pathway.

8 Conclusions

This paper details experiments that compare the classification performance of Fukushima's neocognitron to a number of modified versions of the neocognitron. In order to do this we have had to formulate a new activation function that can be used in neocognitron style networks (Equation (11)). This activation function makes explicit the relation between the output of the S-cell and the cell's parameters and requires a single presentation of the training patterns for supervised learning.

Our experiments have shown that a neocognitron that uses sigmoidal S-cells has a significantly better classification performance than Fukushima's original implementation of the neocognitron. The function of the C-cells in the neocognitron was also adapted to that of a ranked order filter but this did not enhance the test performance of the neocognitron.

Further tests showed that an optimal closed-form training algorithm for the neocognitron (proposed by Hildebrandt) does not improve the practical performance of the network.

Our main conclusion is that the high performance of the the neocognitron can be further improved by modifications to its component cells along the lines described.

References

[1] Fukushima, K. (1980). Neocognitron: A Self-Organizing Neural Network model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, **36**(4), 193-202.

[2] Fukushima, K. (1988). Neocognitron: A Hierarchical Neural Network Capable of Visual Pattern Recognition. *Neural Networks*, **1**, 119-130.

[3] Fukushima, K. (1989). Analysis of the Process of Visual Pattern Recognition by the Neocognitron. *Neural Networks*, **2**, 413-420.

[4] Fukushima, K., & Wake, N. (1991). Handwritten Alphanumeric Character Recognition by the Neocognitron. *IEEE Transactions on Neural Networks*, **2**(3), 355-365.

[5] Barnard, E., & Casasent, D. (1990). Shift Invariance and the Neocognitron. *Neural Networks*, **3**, 403-410.

[6] Menon, M. M., & Heinemann, K. G. (1988). Classification of Patterns Using a Self-Organizing Neural Network. *Neural Computation*, **1**, 201-215.

[7] Yamaguchi, T., & Kropatsch, W. G. (1990). Distortion-Tolerance Curve of the Neocognitron with Various Structures, including Pyramid. *Proceedings of the 10th International Conference on Pattern Recognition*, Atlantic City, New Jersey.

[8] Hildebrandt, T. H. (1991). Optimal Training of Thresholded Linear Correlation Classifiers. *IEEE Transactions on Neural Networks*, **2**(6), 577-588.

[9] Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture of the cat's visual cortex. *Journal de Physiologie*, **160**(1), 106-154.

[10] Hogg, R. V., & Ledolter, J. (1987). *Engineering Statistics*, New York: Macmillan, 163-166.

[11] Pitas, I., & Venetsanopoulos, A. N. (1990). *Nonlinear Digital Filters: Principles and Applications.* Massachusetts: Kluwer Academic Publishers, 128-131.

[12] Lovell, D. R., Tsoi, A. C., & Downs, T. (1992). A Note on a Closed-Form Training Algorithm for the Neocognitron. Submitted to *IEEE Transactions on Neural Networks.*

[13] Fukushima, K. (1975). Cognitron: A Self-organizing Neural Network. *Biological Cybernetics*, **20**, 121-136.