INTERNATIONAL COMPUTER SCIENCE INSTITUTE

1947 Center St. • Suite 600 • Berkeley, California 94704-1198 • (510) 643-9153 • FAX (510) 643-7684

Recognition of Handwritten Digits and Human Faces by Convolutional Neural Networks

Claus Neubauer

TR-96-058

December 1996

Abstract

Convolutional neural networks provide an efficient method to constrain the complexity of feedforward neural networks by weightsharing. This network topology has been applied in particular to image classification when raw images are to be classified without preprocessing. In this paper two variations of convolutional networks - Neocognitron and Neoperceptron - are compared with classifiers based on fully connected feedforward layers (i.e. Multilayerperceptron, Nearest Neighbor Classifier, Autoencoding network) with respect to their visual recognition performance. Beside the original Neocognitron a modification called Neoperceptron is proposed which combines neurons from Perceptron with the localized network structure of Neocognitron. Instead of training convolutional networks by time-consuming error backpropagation in this work a modular procedure is applied, whereby layers are trained sequentially from the input to the output layer in order to recognize features of increasing complexity. For a quantitative experimental comparison with standard classifiers two very different recognition tasks have been chosen: handwritten digit recognition and face recognition. In the first example on handwritten digit recognition the generalization of convolutional networks is compared to fully connected networks. In several experiments the influence of variations of position, size and orientation of digits is determined and the relation between training sample size and validation error is observed. In the second example recognition of human faces is investigated under constrained and variable conditions with respect to face orientation and illumination and the limitations of convolutional networks are discussed.

1 Introduction

Convolutional neural networks with local weightsharing topology gained considerable interest both in the field of speech and image analysis [17]. Their topology is more similar to biological networks based on receptive fields and improves tolerance to local distortions. Additionally the model complexity and the number of weights is efficiently reduced by weightsharing. This is an advantage when images with highdimensional inputvectors are to be presented directly to the network instead of explicit feature extraction and data reduction which is usually applied before classification [18], [20], [25]. Weightsharing can also be considered as an alternative to weight elimination [26] in order to reduce the number of weights [3]. Moreover networks with local topology can more effectively be migrated to a locally connected parallel computer than fully connected feedforward networks [14].

The Neocognitron [8], [9], [10], [11], which can be considered as the first realization of convolutional networks has been introduced by Fukuskima. In the Neocognitron model for the first time receptive fields are used extensively, which have been discovered in the cat's visual cortex by Hubel and Wiesel [12], [13]. Fukushima applied the Neocognitron primarily to handwritten digit recognition. Later variants of convolutional networks have been applied for example to large scale zip code recognition and face recognition [16], [4].

Within a convolutional architecture there are several possibilities to combine different kinds of neurons and learning rules. One method is to use Mc-Culloch Pitts neurons, which calculate a weighted sum plus sigmoid nonlinearity, and to train the whole network by error backpropagation [27]. This approach has been applied for zip code recognition [16]. In contrast within the Neocognitron neurons calculate a weighted sum normalized by the incoming signal which results in a normalized convolution [10]. Furthermore weights are trained layer by layer independently by reinforcement and a winner takes all rule. This approach has been shown to be feasible for binarized images of digits but has not been verified on large data sets. In this work a third method is proposed which combines advantages of both approaches by using McCulloch Pitts neurons instead of more complicated neurons based on Neocognitron [19]. The network is trained layer by layer similarly to Neocognitron and thus time-consuming error backpropagation is avoided. This approach is called Neoperceptron in this context. Based on the previous work in this field two questions arise:

Can Neoperceptron and Neocognitron as examples of convolutional neural networks be used for general recognition tasks like face recognition and digit recognition on large scale databases?

How do they compare quantitatively with feedforward networks based on complete connectivity between layers without topological constraints?

In previous work the performance of Neocognitron has not yet been determined for more sophisticated problems like face recognition but only for character or digit recognition on a limited datasample. Thus one goal of this work is, to compare the Neocognitron with several fully connected networks and with the Neoperceptron. A comprehensive evaluation considering the influence of varying training sample size and other key parameters is subject of this paper (see also [22]).

2 Convolutional Neural Networks: Neocognitron and Neoperceptron

2.1 Network Structure

A detailed description of the Neocognitron architecture can be found in [8]. In Fig. 1 the topology of convolutional networks used in this work is shown. The raw image is feed into the input layer (1C) and determines the size of the inputvector. Neurons perform local feature extraction and therefore each neuron is connected by a receptive field to a small area of the previous layer. In this work a four layered network is applied. The hidden layers consist of S-sublayer and C-sublayer (see [8] for details) and each sublayer itself consists of several planes. The input layer is first mapped onto multiple planes of the 2S-sublayer. Each plane of a layer contains neurons which are extracting a particular local feature like a oriented bar or edge. The weights of the neurons in the S-sublayers are modified by training. Neurons of the same plane share the same weights in order to achieve some degree of tolerance to shift and deformation. The receptive field size has been chosen to be 5 x 5 pixels throughout this paper. A mapping from one plane to the next can be considered as a convolution since all neurons of one plane have got the same receptive fields. From one layer to the next the spatial resolution is reduced by two and a blurring filter (3x3 receptive field with fixed weights of uniform shape) avoids subsampling. On the other hand the number of planes is increased from layer to layer in order to detect more specific features of higher complexity (curved shapes). The same topology is used for the Neoperceptron but with a different kind of S-neuron.

2.2 Model of Neurons

In the Neocognitron the S-neurons (together with the V-neurons) are described by (1). Is has been shown in [10] that (1) including the normalization term (V-neurons u_{vl}) approximates a convolution normalized by the length of the weightvector and the inputvector. The selectivity of the S-neurons can by adjusted by hand with the parameter r. r is set to 0.5 in this work, so that significant activation of a few neurons remains, while small neuron activations are suppressed.

$$\begin{aligned} u_{sl}(n,k) &= r_l(k)\phi\left(\frac{1+\sum_\kappa\sum_\nu a_l(\nu,\kappa,k)u_{cl-1}(n+\nu,\kappa)}{1+\frac{r_l(k)}{r_l(k)+1}b_l(k)u_{vl}(n)}-1\right) \\ u_{vl}(n) &= \left(\sum_\kappa\sum_\nu c_l(\nu)(u_{cl-1}(n+\nu,\kappa)^2\right)^{1/2} \\ \phi(x) &= \begin{cases} x \; if \; x\geq 0 \\ 0 \; if \; x<0 \end{cases} \end{aligned} \tag{1}$$

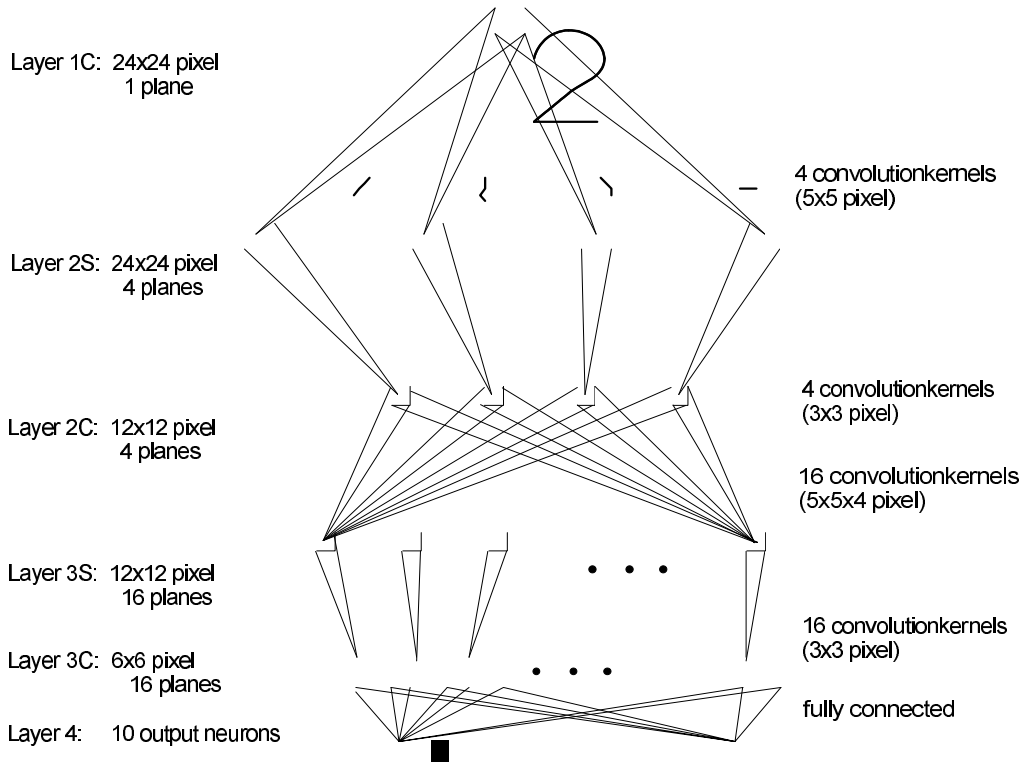

Figure 1: The topology of Neocognitron and Neoperceptron used for face and digit recognition consists of four layers with four convolutional planes in the first hidden layer and sixteen convolutional planes in the second hidden layer.

$u_{vl}(n)$	V-neuron, layer l, position n;
$u_{cl}(n+\nu,\kappa)$	C-neuron, layer $l-1$, plane κ, position $n+\nu$;
$u_{sl}(n,k)$	S-neuron, layer l, plane k, position n;
$u_{cl-1}(n+\nu,\kappa)$	C-neuron, layer $l-1$, plane κ, postion $n+\nu$;
$a_l(\nu,\kappa,k), b_l(k)$	modifiable weights;
$c_l(\nu)$	positive fixed weights;
$r_l(k)$	selectivity;
$\phi(x)$	nonlinearity;

In contrast to the Neocognitron the Neoperceptron is based on McCulloch-Pitts neurons with sigmoid nonlinearity Thus Neoperceptron is a combination of a neuron function based on Perceptron with the convolutional network structure of Neocognitron (2). The Neoperceptron performs a convolution without normalization resulting in a simpler function which is equivalent to a weighted sum between inputvector (within the receptive field) and weightvector of the neuron. In the Neoperceptron a sigmoid function is used as nonlinearity.

The main difference to the concept used in [16] is based on a sequential learning strategy.

$$\begin{aligned} u_{sl}(n,k) &= \phi\left(\sum_{\kappa}\sum_{\nu} a_l(\nu,\kappa,k)u_{cl-1}(n+\nu,\kappa)\right) \\ \phi(x) &= 1/(1+exp(-x)) \end{aligned} \tag{2}$$

2.3 Learning Rules

In this approach the convolutional neural networks are trained layer by layer starting from the first hidden layer. After training of the first hidden layer with images containing simple features the training set for the next layer containing more complex patterns is propagated through the first layer. This procedure is repeated until the output layer is reached. Thus higher layers represent features of increasing complexity. One advantage of this approach is that the first hidden layer does not have to be retrained for each classification problem since for typical visual recognition tasks usually edges have to be extracted at the first level. For the Neocognitron the reinforcement learning rule proposed by Fukushima is used here both for supervised and unsupervised training (3). For the Neoperceptron the LMS-rule is used for supervised learning. This is feasible for the digit recognition problem where it is intuitively clear that curved lines and endpoints are important higher level features. For a problem like face recognition on the other hand it is difficult to train intermediate layers supervised since it is not known which higher level features are most important. In this case unsupervised algorithms for feature extraction are applied like principle component analysis or autoencoding learning [6], [23]. Both unsupervised algorithms give similar results since it has been shown that the weights in the hidden layer of a threelayered autoencoding network converge to principal components [3], [5].

$$\begin{aligned} \Delta a_l(\nu,\kappa,\hat{k}) &= q_l c_l(\nu) u_{cl-1}(\hat{n}+\nu,\kappa) \\ \Delta b_l(\hat{k}) &= q_l u_{\nu l}(\hat{n}) \end{aligned} \tag{3}$$

3 Classification of Handwritten Digits

First the performance of convolutional neural networks and fully connected networks is compared for recognition of handwritten digits. In the following experiments no specific feature extraction takes place but the raw images are used directly for classification. The digit dataset consists of 10000 digits (1000 per class) for learning and 10000 digits for validation. In the original dataset the digits are normalized with respect to size, position, orientation and contrast and the resolution is 16 by 16 pixels with 8 bit greyvalues. Examples of the dataset are shown in Fig. 2.

Two experiments are described in detail. In experiment I normalized digits are classified. Variations in the human style of writing require a strong tolerance to deformations. In experiment II (Fig. 3) additionally the digits are shifted, scaled and rotated by affine transformations in order to measure the tolerance of the classifiers with respect to these transformations separately and in combination.

Figure 2: Examples of the digit dataset, containing digits normalized with respect to size, position and orientation used in experiment I (16x16pixels).

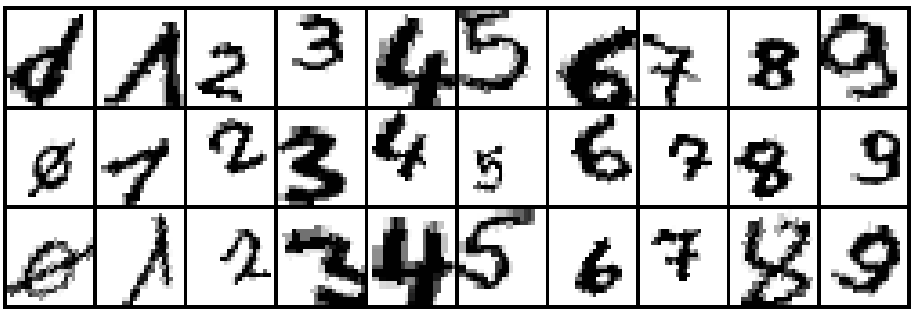

Figure 3: Examples from the digit dataset with additional variation of digit size, position and orientation, which is used in experiment II (24x24pixel).

3.1 Experiment I: Handwritten Digits with Constant Size, Position and Orientation

Neoperceptron and Neocognitron are compared with three fully connected classifiers. Fully connected classifiers are represented by the two layered Perceptron, the threelayered Multilayerperceptron and the Nearest Neighbor Classifier with the appropriate topology (Table 1). In this context the Nearest Neighbor Classifier can be regarded as a neural network where the number of hidden units is equal to the number of training samples and a winner takes all rule is applied at the output layer.

Table 1: Classifiers based on local and global feedforward connections, which are compared in experiment I on normalized digits.

Local connections		Global connections	
	planes		neurons
Neoperceptron	1-4-12-10	Nearest Neighbor classifier	256-x-10
Neocognitron	1-4-12-10	Two layer Perceptron	256-10
		Three layer Multilayerperceptron	256-50-10

While the topology of Nearest Neighbor Classifier and Perceptron is determined by the

problem, for the Multilayerperceptron the number of hidden neurons has to be optimized in advance by multiple training runs with different numbers of hidden neurons. Fifty neurons turned out to be a good trade off between classification performance and model complexity. All classifiers except Nearest Neighbor Classifier have to be trained iteratively, whereby the convergence of the validation error is observed simultaneously on an independent test set in order to avoid overfitting. Training is stopped when the validation error is not further decreasing. The training sample size is an important key parameter for generalization. Therefore the learning procedure is repeated several times for each classifier with different training sample sizes. Nine training repetitions are carried out with training sample sizes ranging from 10 to 1000 patterns per class. For example Fig. 4 shows for the Multilayerperceptron the convergence of training and corresponding validation error for different training sample sizes.

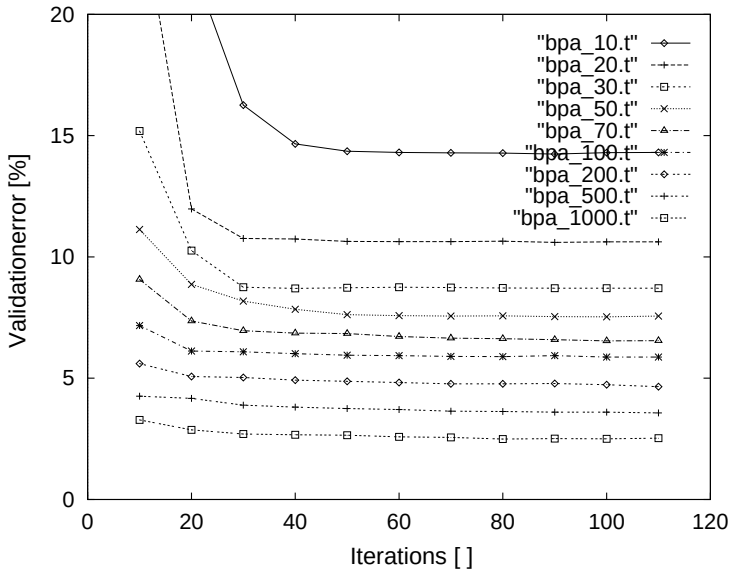

Figure 4: Convergence of training and validation error of Multilayerperceptron for nine different training sample sizes between 10 and 1000 samples per class.

Fig. 5 summarizes the results of experiment I. The validation error is plotted as a function of training sample size for the five different classifiers. It is clearly shown that both convolutional networks generalize significantly better than the fully connected networks. The performance of Perceptron is worst since a validation error of 5.52% remains after training with 1000 samples per class. This shows that a linear separation of the ten digit classes is not sufficiently accurate on raw images. The Multilayerperceptron has got 2.53% error. On the other hand for the convolutional networks - Neoperceptron and Neocognitron - only 1.23% and 1.59% misclassifications have been observed respectively. This indicates that the Neoperceptron is even slightly better than the Neocognitron with more complicated neurons. For small training sample sizes (10 patterns per class) the Nearest Neighbor Classifier has got the worst error rate (18.93%) but with increasing training sample size it gradually improves relatively to the other algorithms and achieves 2.31% test error for 1000 learning patterns per class. This might be due to the fact that the Nearest Neighbor

Classifier approximates the optimal Bayes classifier for large sample sizes. However we are interested primarily in good generalization for small training sample size and computational efficiency so that the Nearest Neighbor classifier based on 1000 samples per class is not feasible for practical applications.

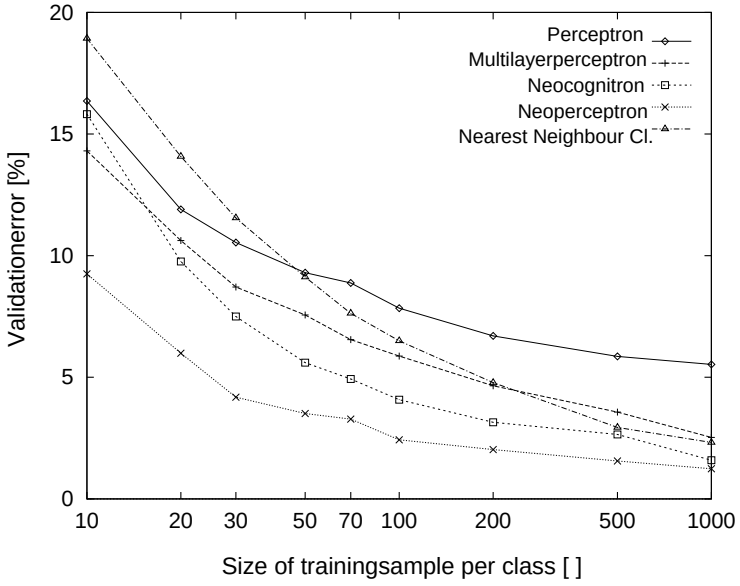

Figure 5: Validation error as a function of training sample size for five classifiers. Classifiers based on local connectivity (Neoperceptron, Neocognitron) are superior to classifiers with full feedforward connectivity in particular for small training sample size.

3.2 Experiment II: Digits with Variable Size, Position and Orientation

After experiment I based on normalized digits, here the performance of convolutional networks is determined for digits varying with respect to position, orientation and size within the image frame. For this experiment four datasets with 5000 patterns for training and 5000 patterns for testing are generated from the original normalized dataset by random affine transformations. The spatial resolution is increased to 24x24 pixels. In sample 1 digits are transformed with a random combination of rotation, shift and scaling (see Fig. 3). Sample 2 contains randomly shifted patterns with shifts of maximal 50% with respect to digit size in each direction. In sample 3 rotations smaller or equal 22.5° are applied and in sample 4 digits are scaled by a factor between 0.5 and 2. This comparison is focusing on Neoperceptron and Multilayerperceptron, since they performed best in their class (local and global topology) in the previous experiment. Learning curves are generated by variation of the training sample size between 10 and 500 digits per class in eight steps for the four different datasets. The classification results on the four datasets are summarized in Fig. 6 and Fig. 7 for Multilayerperceptron and Neocognitron respectively.

Once more the Neoperceptron achieves significantly better results than the Multilayerperceptron. For example the validation error of Neoperceptron for sample 1 which was subject to all transformations is 14.20% compared to 33.74% for the Multilayerperceptron

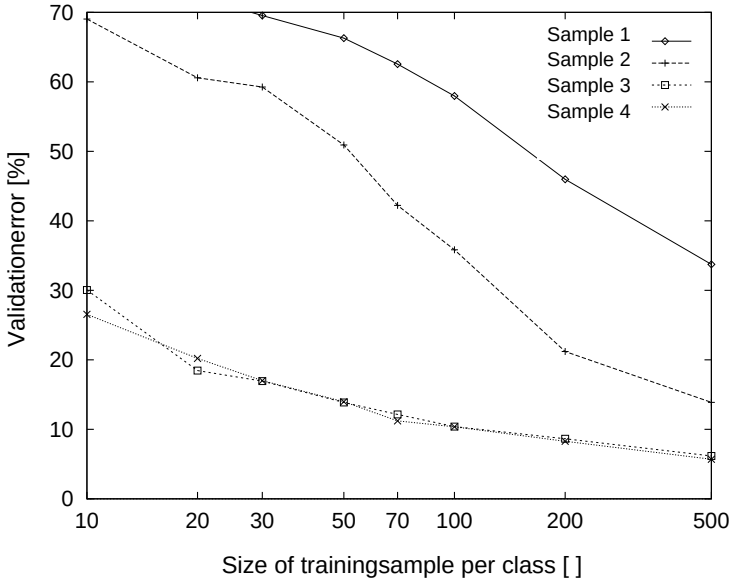

Figure 6: Validation error of Multilayerperceptron as a function of learning sample size for samples 1-4 subject to affine transformations. 1: shift, rotation and scaling; 2: shift; 3: rotation 4: scaling.

if 500 patterns per class are used for training. However compared to the classification results in experiment I with normalized digits a decreasing accuracy can be observed for both classifiers which indicates that there is only limited tolerance to affine transformations even for the Neoperceptron. In Fig. 7 and Fig. 6 the influence of shift, scaling and rotation is also shown separately, which is important for the choice of appropriate preprocessing steps. While position variations had a high impact on the performance (MLP (Multilayerperceptron) 13.88%, NEO (Neoperceptron) 5.06%), scaling (MLP 6.16%, NEO 2.20%) and rotation (MLP 5.70%, NEO 3.08%) did not affect the classification accuracy as much. Thus it can be concluded that the convolutional architecture generalizes significantly better than the Multilayerperceptron but the results are still depending on proper normalization, in particular if only small training sample sizes are used.

4 Recognition of Human Faces

Human face recognition is another challenging visual classification task. There exist several reliable methods for identification of persons like iris diagnosis [7] or fingerprint analysis but they are quite uncomfortable so that identification by face images is preferred if the accuracy is sufficient. For face recognition small deviations form a three-dimensional shape have to be detected, while the recognizer has to be able to cope with a large variety of appearances due to possible illumination and pose variations. In order to simplify this task it is necessary, first to locate the face within the scene and then to normalize the face with respect to size and orientation. Therefore in this approach a hierarchical face recognition is realized. A separate network is trained to localize a face within a scene and afterwards the

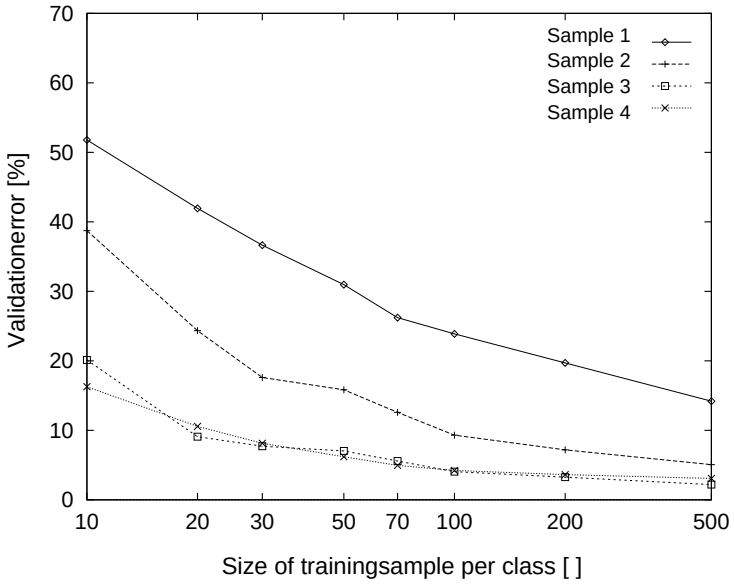

Figure 7: Validation error of Neoperceptron as a function of learning sample size for samples 1-4 subject to affine transformations. 1: shift, rotation and scaling; 2: shift; 3: rotation 4: scaling.

subimage containing the face is analyzed by the identification network. Three experiments are carried out under different conditions as described below. Two datasets have been gathered: one with constant illumination and homogeneous background (Fig. 8) and one with variations of illumination and background which is more similar to real live pictures (Fig. 9).

4.1 Localization of Faces

For face localization a threelayered Multilayerperceptron with ten hidden neurons is trained to detect faces of standard size in a window of fixed size. A resolution of 32x32 pixel turns out to be sufficient for this task since a face is primarily characterized by existence of eyes, nose and mouth and their geometrical relationship which can easily be recognized at low spatial resolution. Network training with 1000 unlabeled faces and the same amount of arbitrary background images results in a sufficiently precise face detector with an accuracy summarized in Table 2. For localization a window is shifted over the whole image. Several possible sizes and orientations of the face are taken into account by a multiresolution technique and by variation of the window orientation within the scene image. These subimages are presented to the face detector so that it is possible to locate arbitrary faces in cluttered scenes if nearly frontal views are provided. From subimages with significantly high response of the localization network the background outside the central circle is cut off and the subimages are feed into the identification network as described in Fig. 10.

Figure 8: Examples from the face dataset with constrained illumination and homogenous background

Figure 9: Examples from the face dataset with varying illumination and background

4.2 Experiment I: Constrained Training Sample

Similarly to the digit recognition experiments here the convolutional network is compared with several classifiers based on full connectivity. The Nearest Neighbor Classifier and the threelayered Multilayerperceptron are used for this purpose as before. Alternatively the threelayered fully connected feedforward architecture is trained with two other training strategies. Here instead of overall supervised learning the hidden layer is first trained unsupervised and afterwards the output layer is trained supervised by least mean square error minimization. For this approach the Selforganizing Feature Map [1], Autoencoding network [6] and Eigenfaces [24] have been proposed for unsupervised training of the first hidden layer. Beside pure feedforward classifiers there exist also iterative approaches like dynamic link architecture [15], which perform a graph matching between image and model, using simulated annealing for optimization. By this concept it is also possible to compensate for small deformations, but this method is quite slow due to the simulated annealing optimization necessary during classification. Feature extraction by a selforganizing map after training with unlabeled faces is illustrated in Fig. 11, where the prototypes corresponding to the 100 neurons are shown. Surprisingly the prototypes of the 10 by 10 map represent

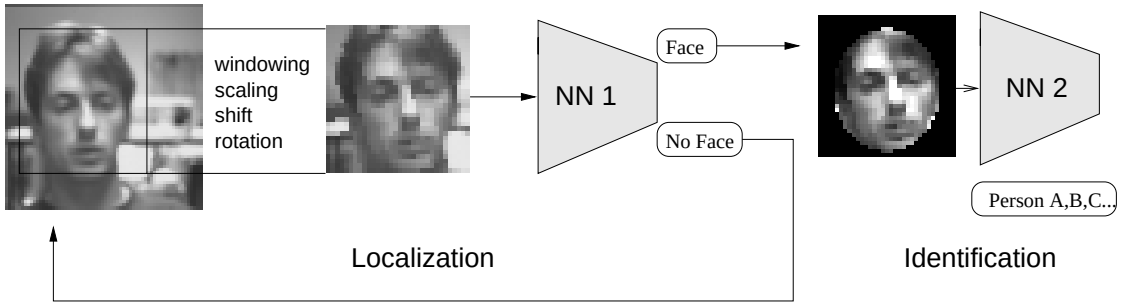

Figure 10: Concept for localization of faces in cluttered scenes.

Table 2: Performance of the face-background classifier.

[%]	Training		Validation	
classified from/to	Face	Background	Face	Background
Face	95.1	1.5	86.8	2.5
Background	4.9	98.5	13.2	97.5

typical faces from the training set in a ordered way. Similar faces are close together within the map. For example women are concentrated in the upper left and men in the lower right area.

In this experiment the dataset used for training is acquired under fairly constant, diffuse and frontal illumination as shown in Fig. 8. The head pose has been varied randomly within approximately 45° in each direction. 1080 images from 18 persons (60 images per person) have been selected randomly from a video covering different face poses from each person. The first validation set has been grabbed independently but under fairly the same illumination and with homogenous background. The classifiers achieved very different results in this comparison. In Tab. 3 in the first column the type of classifier is shown, in the second column the image resolution and in the third column the results for the constrained validation set are shown. The Nearest Neighbor Classifier had a misclassification rate of 4.9% for images with 64x64 pixel resolution and 12.9% for images with 32x32 pixel. The Multilayer-perceptron correctly identifies only 6.4%, which indicates that no reasonable learning took place in this case. Since the weightvectors are very highdimensional the system is underdetermined, if trained with a face training set of only 1080 images. If the first layer is trained as a Selforganizing map the threelayered network resulted in 17.7% test error, while with Autoencoding learning 78.7% are misclassified. In contrast the Neoperceptron has got 4.5% validation error on 32x32 pixel images, which is similar to the performance of the Nearest Neighbor Classifier at higher inputresolution (64x64 pixels). However the Neoperceptron requires less computations. In this example the Neoperceptron is approximately eight times faster than the Nearest Neighbor Classifier.

In order to determine, if the classifiers, trained with the constrained dataset, are capable to classify more unconstrained images a second validation set has been grabbed from a real life video as illustrated in Fig. 9. Hereby it turns out that all classifiers have significant

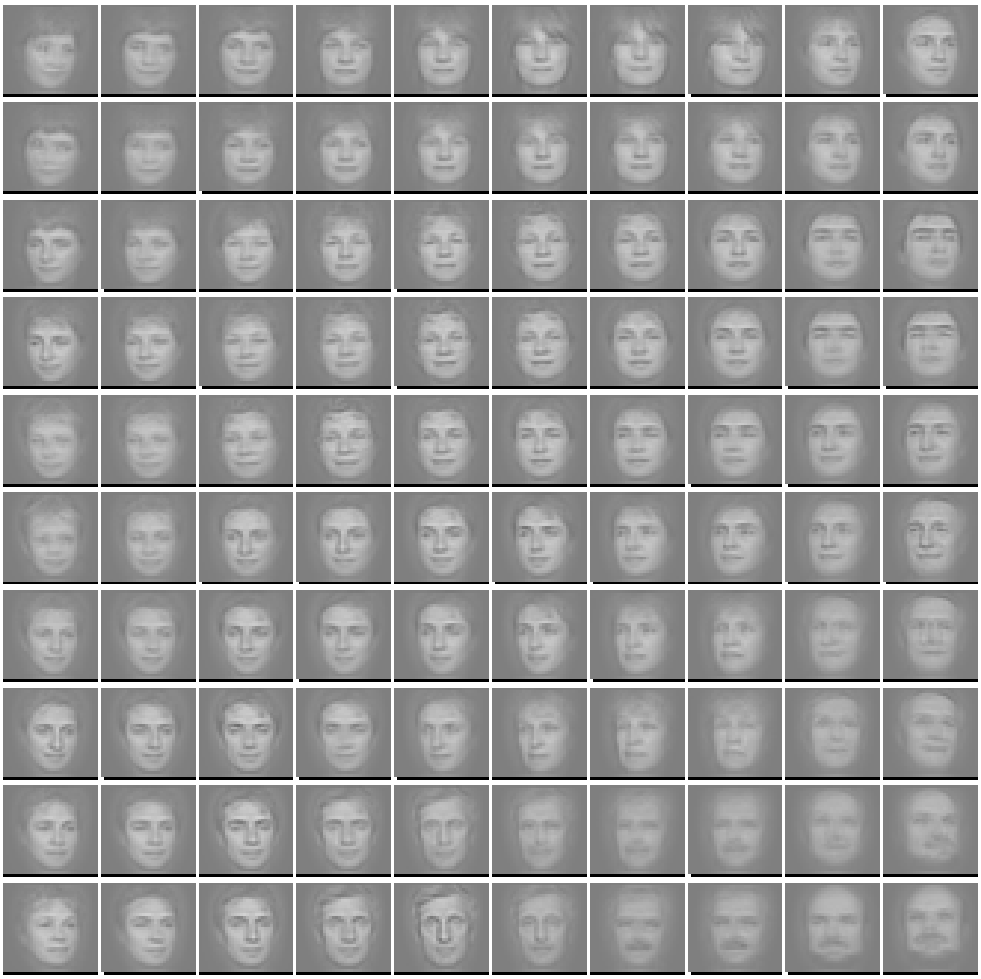

Figure 11: Prototypes, which evolve by training a twodimensional selforganizing map (10x10 neurons) with faces. These prototypes are used for feature extraction in the hidden layer of a three layered fully connected network for face recognition.

problems to identify persons correctly. As shown in Tab. 3, forth column, the misclassification lies above 60% for all classifiers. The convolutional architecture is still relatively better compared to the other classifiers but is not able to generalize from constrained to unconstrained images.

4.3 Experiment II: Unconstrained Training

In the previous experiment the face recognition performance has been determined for training with faces under constrained conditions. If the recognition is to be improved for unconstrained images it is necessary to take these situations into account already during learning. On the other hand it is relatively time consuming and exhaustive to collect various images under different pose and illumination from all persons to be recognized. Therefore finally a restricted task is considered, where only one particular person is to be recognized and

Table 3: Validationerror for the constrained (Fig. 8) and unconstrained face sample (Fig. 9).

		constrained face sample	unconstrained face sample
	input resolution [pixel]	validation error [%]	validation error [%]
Nearest Neighbor Classifier	64 x 64	4.9	68.5
Nearest Neighbor Classifier	32 x 32	12.9	69.7
Multilayerperceptron	32 x 32	93.6	95.2
Selforganizing Map + LMS	32 x 32	17.7	71.3
Autoencoding network + LMS	32 x 32	78.7	85.8
Neoperceptron	32 x 32	4.5	59.7

all other persons are to be rejected. Such a classifier is for example necessary to verify the correspondence of a person with a personal authorization (passport, credit card etc.). For this test 800 images of one person have been grabbed randomly from a video under unconstrained conditions and the Neoperceptron has been trained to recognize this person. 1200 images containing eighteen other persons have been used as examples for the rejection class. This classifier has a reasonable accuracy as shown in Tab. 4.

Table 4: Confusionmatrix for classification between person A and all other persons (not A).

[%]	Training		Validation	
classified from/to	A	not A	A	not A
A	97.9	1.9	87.6	2.7
not A	2.1	98.1	12.4	97.3

5 Summary and Conclusions

In this article two types of convolutional neural networks - Neocognitron and Neoperceptron - are examined. The Neoperceptron is introduced as a combination of neurons based on the Perceptron with the weightsharing architecture of the Neocognitron. Convolutional networks show several advantages compared to fully connected networks: they are more similar to biological neural networks, they can be easily migrated to parallel hardware and they are tolerant to small deformations of input patterns. Both networks are evaluated quantitatively based on visual recognition tasks. Handwritten digit classification and human face recognition are chosen to compare convolutional networks with fully connected classifiers.

Classification of handwritten digits showed that both Neoperceptron and Neocognitron

are superior to fully connected classifiers. For example after training with 1000 patterns per class (16x16 pixels) the Neoperceptron has got 1.23% and the Neocognitron 1.59% misclassifications, while the Nearest Neighbor Classifier has got 2.31%, the Perceptron 5.52% and Multilayerperceptron 2.53% misclassifications. The Neoperceptron performs slightly better than the Neocognitron even though its S-neurons are simplified. Both types of convolutional networks outperform the fully connected classifiers. If smaller learning sets are considered the performance difference is even larger. Further experiments on digit datasets which where subject to affine transformations (shift, rotation, scaling) within a 24x24 window confirmed the advantage of the convolutional architecture. For example the Neoperceptron reached 3.08% error on digits with varying orientation while the Multilayerperceptron has got 5.70% misclassifications on the same set (500 training samples per class). However for digits with varying size, orientation and position the performance of both types of classifiers is significantly worse than for normalized digits. Apparently convolutional networks are tolerant to affine transformations only to some degree.

In addition to the classification performance the amount of computation required by the different classifiers has to be considered. On a serial computer Neocognitron and Neoperceptron require ten times more computations for the digit classification example described above than the Multilayerperceptron but they are about the same factor faster than the Nearest Neighbor Classifier. On the other hand the convolutional networks have got the best classification accuracy and they can be accelerated much easier by parallel computers with local communication than fully connected networks.

The experiments on face recognition indicate that for reliable recognition good alignment of the faces is essential for each type of classifier. For accurate localization of faces a threelayered Multilayerperceptron is trained for detection of faces with fixed size. A window with varying size and orientation is moved over the whole scene while the normalized subimages are evaluated by the face detection network. The face identification experiments with eighteen persons proved again, that the convolutional neural network (Neoperceptron) performs better than fully connected networks. The Neoperceptron has got 4.5% misclassifications while the Nearest Neighbor Classifier has got 12.9%, Selforganizing Map 17.7%, Autoencoding Network 78.7% and Multilayerperceptron 93.6% error rate. These good recognition rates for Nearest Neighbor Classifier and Neoperceptron drop down however when the classifiers are tested under more unconstrained conditions with respect to pose and illumination than they were trained on.

In contrast to digit recognition the influence of out of the plane rotation and illumination variation require a sophisticated training set which covers the spectrum of possible face appearances as good as possible. Therefore accuracy under more unconstrained conditions can be improved by a more versatile training sample covering more aspects of pose, illumination and background. A further test which takes into account more general training samples from real live video sequences, showed significant improvements but this approach is on the other hand quit uncomfortable for practical problems since a lot of instances are required per person.

In future attention will be focused on the representation of highlevel face features for improved generalization. These features have to be trained by a large amount of image data from various persons in order to cover most aspects of pose and illumination. Based on these features it should be possible to learn a new person by a few images only and nevertheless

good generalization on unknown instances and poses should be obtained. In this case for example the generation of virtual views proposed by [2] can help to further improve identification accuracy. Beside the applications discussed here convolutional networks will be integrated as classifiers for automatic x-ray inspection of solder joints in electronic production, where three-dimensional datasets have to be evaluated. Neural networks have already been successfully applied for defect detection and classification of solder joints [21] and convolutional networks combined with computer tomography will help to further improve quality control of Printed Circuit Boards.

References

[1] N.M. Allinson, A.W. Ellis, B. Flude, A. Luckman, "A Connectionist Model of Familiar Face Recognition", in IEE Colloquium on Machine Storage and Recognition of Faces, pp. 5.1-5.9, 1992

[2] D. Beymer, T. Poggio, "Image Representations for Visual Learning", Science, vol. 272, pp. 1905-1909, 1996

[3] C. Bishop, "Neural Networks for Pattern Recognition", Oxford Press, 1995.

[4] H. Bouattour, F. Fogelman Soulie, E. Viennet, "Solving the Human Face Recognition Task using Neural Nets", in Artificial Neural Networks II, I. Alexander, J. Taylor, eds, North- Holland, Amsterdam, pp. 1595-1598, 1992

[5] H. Bourlard, Y. Kamp, "Autoassoziation by Multilayerperceptrons and Singular Value Decomposition", Biological Cybernetics, vol. 59, pp. 291-294, 1988

[6] G.W. Cottrell, "EMPATH: Face, Emotion, and Gender Recognition Using Holons", in Advances in Neural Information Processing Systems, Morgan Kaufmann, vol. 3, pp. 564-571, 1991

[7] J.G. Daugman, "High Confidence Visual Recognition of Persons by a Test of Statistical Independence", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 15, pp. 1148-1161, 1993

[8] K. Fukushima, "Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position", Biological Cybernetics, vol. 36, pp. 193-202, 1980

[9] K. Fukushima, "A Neural Network Model for Selective Attention in Visual Pattern recognition", Biological Cybernetics, vol. 55, pp. 5-15, 1986

[10] K. Fukushima, "Analysis of the Process of Visual Pattern Recognition by the Neocognitron", Neural Networks, vol. 2, pp. 413-421, 1989

[11] K. Fukushima, T. Imagawa, "Recognition and Segmentation of Connected Characters with Selective Attention", Neural Networks, vol. 6, pp. 33-41, 1993

[12] D.H. Hubel, T.N. Wiesel, "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex", J. Physiology, vol. 160, pp. 106-154, 1962

[13] D.H. Hubel, T.N. Wiesel, "Receptive Fields and Functional Architecture in Two Nonstriate Visual Areas (18 und 19) of the Cat", J. Neurophysiology, vol. 28, pp. 229-289, 1965

[14] T. Ito, K. Fukushima, "Realization of a Neural Network Model Neocognitron on a Hypercube Parallel Computer", Int. J. of High Speed Computing, vol. 2, pp. 1-16, 1990

[15] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Würtz, W. Konen, "Distortion Invariant Object Recognition in the Dynamic Link Architecture", IEEE Trans. on Computers, vol. 42, pp. 300-311, 1993

[16] Y. Le Cun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition", Neural Computation, vol. 1, pp. 541-551, 1989

[17] Y. LeCun, Y. Bengio, "Convolutional Networks for Images, Speech, and Time Series" in The Handbook of Brain Science and Neural Networks, MIT Press, M. Arbib ed., pp. 255-258, 1995

[18] C. Neubauer, "Fast Detection and Classification of Defects on Treated Metal Surfaces using a Backpropagation Neural Network", in Proc. of IJCNN, Singapore, pp. 1148-1153, 1991

[19] C. Neubauer, "Shape, Position and Size Invariant Visual Pattern Recognition Based on Principles of Neocognitron and Perceptron", in Artificial Neural Networks, I. Alexander, J. Taylor, eds, North-Holland, Amsterdam, vol. 2, pp. 833-837, 1992

[20] C. Neubauer, "Segmentation of Defects in Textile Fabric", in Proc. of ICPR, Den Haag, pp. 688-691, 1992

[21] C. Neubauer, R. Hanke, "Improving X-Ray Inspection of Printed Circuit Boards by Integration of Neural Network Classifiers", in Proc. of IEMT, Santa Clara, pp. 14-18, 1993

[22] C. Neubauer, "Modellierung visueller Erkennungsvorgänge mit neuronalen Netzen", Ph.D. dissertation, University Erlangen-Nürnberg, Dept. Technische Elektronik, 1995

[23] D.Rumelhart, J. McClelland, "Parallel Distributed Processing: Exploration in the Microstructure of Cognition", MIT-Press, 1986

[24] M. Turk, A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, vol. 3, pp. 71-86, 1991

[25] U. Schramm, T. Wagner, S. Schmölz, K. Spinnler, F. Böbel, R. Haas, H. Haken, "A Practical Comparison of Synergetic Computer, Restricted Coulomb Energy Networks and Multilayer Perceptron", in Proc. WCNN 93, Portland, vol.3, pp. 657-660, 1993

[26] A.S. Weigend, D. Rumelhart, B. Huberman , "Generalization by Weight-Elimination with Application to Forecasting", in Advances in Neural Information Processing, Morgan Kaufmann, vol. 3, pp. 875-882, 1991

[27] P.J. Werbos, "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences", in Ph.D. Dept. of Applied Mathematics, Havard University, Cambridge, 1974