IN PROCEEDINGS OF IEEE INT. CONFERENCE ON IMAGE PROCESSING (ICIP'03), VOL. I, PP. 549-552 – BARCELONA, SPAIN, 2003.

A TWO-STAGE SYSTEM FOR METER VALUE RECOGNITION

Sven Behnke

International Computer Science Institute
1947 Center St., Berkeley, CA, 94704, USA
behnke@icsi.berkely.edu, www.icsi.berkely.edu/∼behnke

ABSTRACT

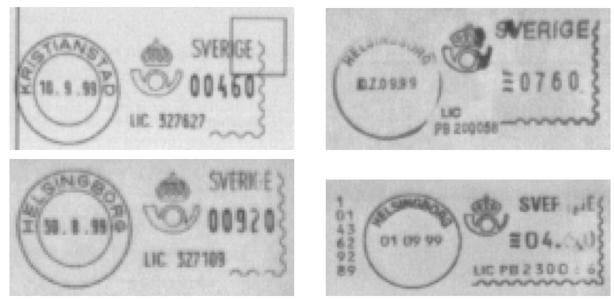
This paper describes a two-stage system for the recognition of postage meter values. A feed-forward Neural Abstraction Pyramid is initialized in an unsupervised manner and trained in a supervised fashion to classify an entire digit block. It does not need prior digit segmentation. If the block recognition is not confident enough, a second stage tries to recognize single digits, taking into account the block classifier output for a neighboring digit as context. The system is evaluated on a large database. It can recognize meter values that are hard to read for humans.

1. INTRODUCTION

Meter stamps are commonly used to mark letters. Postage meters print stamps usually with red ink in the letter's upper right corner. During automatic mail processing the stamps are recognized in order to compare the postage value to the letter's weight. Meter value reading consists of three steps: meter stamp detection, value localization, and value recognition. In the following, a system is described that covers only the last step, the recognition of isolated meter values. A block classifier is trained to recognize the entire meter value, without prior digit segmentation. It is based on the Neural Abstraction Pyramid architecture [1, 2, 3]. If this classifier cannot make a confident decision, single digit classifiers are queried and combined with its results.

A database of Swedish Post meter marks is used that has been collected by Siemens ElectroCom. It contains 5,471 examples, randomly assigned to a training set of size 4,372 and a test set of 1,099. Fig. 2 shows some sample images. As can be seen, the recognition of the meter value is not an easy task. The images have low resolution and low contrast. High variance of the print, the lighting, and the background complicate recognition further. Frequently, the meter values are difficult to read even for humans.

On the other hand, the meter values are not arbitrary combinations of digits, but come from a set of standard postage values. The 16 most frequent values, accounting for 99.2% of the data, are also shown in the figure. One can observe that the top 5 values cover almost 90% and that it suffices to read the two digits next to the point separator to uniquely identify a meter value.



	4.60	4.40	9.20	3.90	4.10	3.80	8.80	7.00
#	3176	534	522	341	319	98	91	82
%	58.1	9.8	9.5	6.2	5.8	1.8	1.7	1.5

	3.50	6.00	8.20	5.00	13.00	12.00	7.80	8.00
#	78	64	42	19	18	16	14	14
%	1.4	1.2	.77	.35	.33	.29	.26	.26

Fig. 2. Swedish Post meter stamps and most frequent labels.

An automatically determined rectangular region is given for each example that should contain the digits belonging to the meter value, but nothing else. Before it can be presented to the block classifier, some preprocessing is needed to make its task easier, as shown in Fig. 3. Goal is to reduce the variance of the examples by color filtering and by increasing the image contrast, such that the print becomes black and the background becomes white, and by normalizing slant and position of the meter value. The two digits of interest are centered in a 32×16 window. A detailed description of the procedure can be found in [3].

2. BLOCK CLASSIFICATION

The task of the block classifier is to recognize a meter value from a preprocessed image. Although preprocessing discarded some of the variances, this is still challenging. The digits come in different sizes and fonts, with varying spaces between them. Only some examples contain a delimiting point and significant noise is still present. One could now

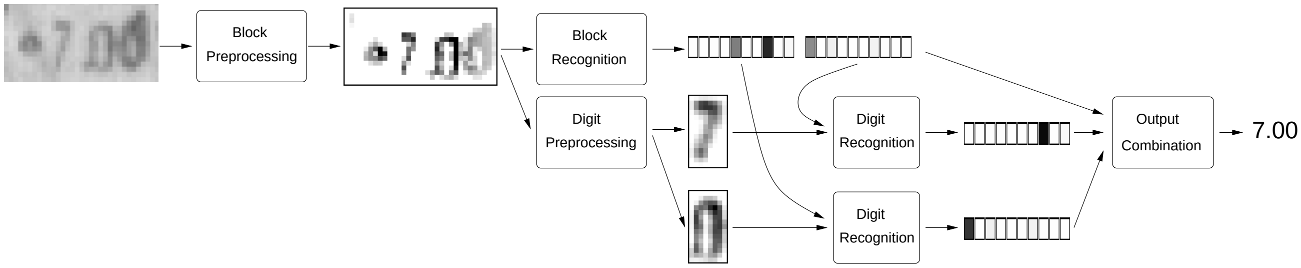


Fig. 1. Combined meter value recognition system.

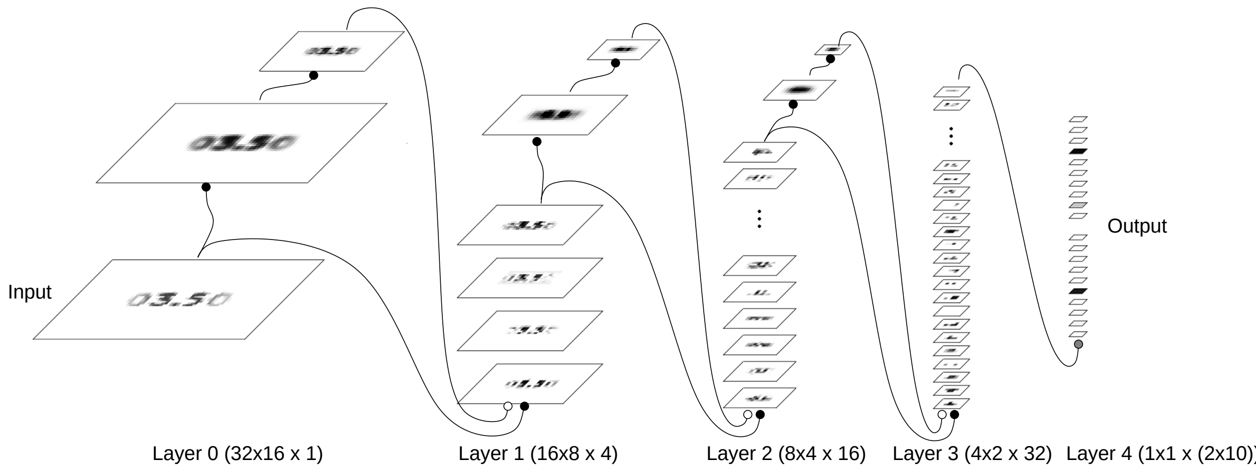


Fig. 4. Network architecture for recognition of entire meter value blocks.

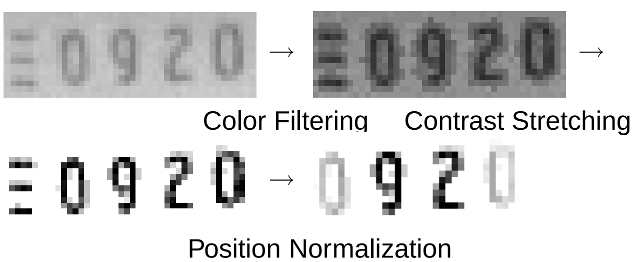


Fig. 3. Preprocessing of meter value blocks.

try to segment the digit block into single digits, recognize them, and combine the digit classifier outputs to a meter value. This approach would require reliable digit segmentation and reliable digit classification. Both requirements are not easy to meet. It is fairly hard to segment the digits and it is also difficult to read isolated digits reliably.

For these reasons, a block classifier was developed that recognizes the two digits of interest simultaneously within the context of neighboring digits. Unlike a digit classifier that could only use the a-priory digit distribution, this classifier is able to take advantage of the non-uniform meter value distribution.

The architecture of the network used is sketched in Fig. 4. It is a feed-forward Neural Abstraction Pyramid [1, 2, 3] consisting of five layers. This architecture is characterized by its hierarchical structure and its local connectivity with weight sharing. It has similarities to the Neocognitron [4] and to convolutional networks [5], like LeNet 5 or SDNNs. In contrast to classifiers sliding over an input line, it processes the entire meter value block in parallel.

The bottom Layer 0 has a resolution of 32×16. It contains only the input array. Resolution decreases from layer to layer by a factor of two in both dimensions, until Layer 3 reaches a size of only 4×2. At the same time, the number of excitatory features rises from four in Layer 1, to 16 in Layer 2, to 32 in Layer 3. The network contains 20 output units in the topmost layer which indicate the identity of the two digits of interest in a $2\times$(1-out-of-10) code.

The output units receive their inputs directly from all positions of all feature arrays of Layer 3. Their weights are allowed to change sign. They compute a weighted sum that is limited by a sigmoidal transfer function to $[0, 1]$. In contrast, the units computing features in layers 1-3 are driven by specific excitation and unspecific inhibition. Their specific excitatory weights originate from overlapping 4×4 windows of the feature arrays in the layer below them. Unspecific inhibition comes via a single weight from the subsampled sum of these features. The rectifying transfer function used here saturates at one. This ensures that the network learns sparse representations of the digit block, since the activity becomes exactly zero if inhibition exceeds excitation.

The network is initialized using the unsupervised learning of sparse features, described in [2]. Supervised training

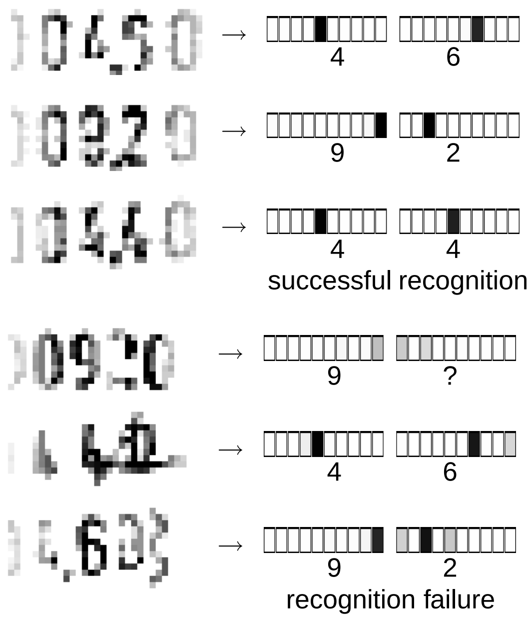


Fig. 5. Difficult block recognition test set examples.

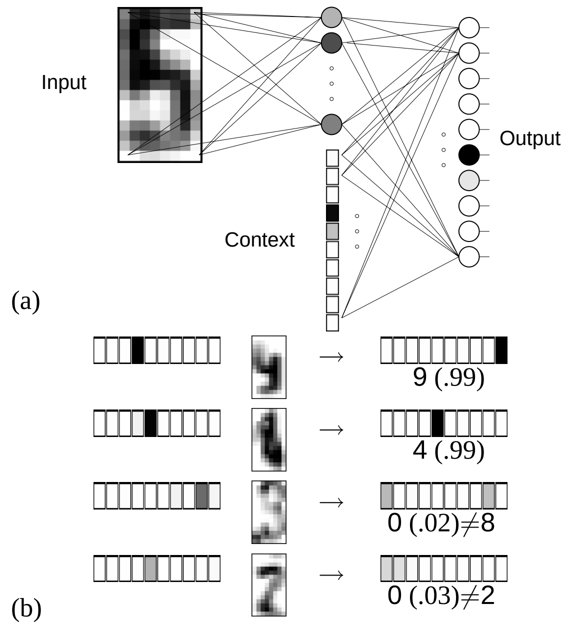


Fig. 6. Digit classification: (a) network architecture; (b) some input/output examples from the test set.

is done with gradient descent on the squared output error until the performance on the test set does not improve any more. The training enforces the desired weight signs. If a specific excitatory weight would become negative, it is set to zero and the unspecific inhibitory weight is changed instead. This leads to sparse excitatory weights.

The trained network is able to perform the recognition task almost perfectly for the training set. Test performance is illustrated in Fig. 5. Some successfully recognized difficult examples are shown along with some examples for which recognition failed. One can observe that for most ambiguous images, the network is able to indicate its uncertainty by producing outputs that deviate from the desired 1-out-of-10 pattern. This makes it possible to compute a meaningful classification confidence based on the difference between the two most active units for the two digits. It can be used to reject ambiguous examples.

In Fig. 7 the test performance of the hierarchical block classifier is shown. About 2% of the examples are substituted when all outputs are accepted. Half of the substitutions can be avoided by rejecting 2.4%. To reduce the substitution rate further, a larger fraction of the examples must be rejected.

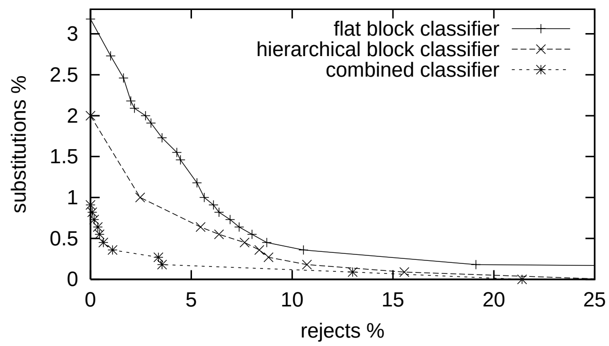
For comparison, several fully connected three-layered feed-forward neural networks with sigmoidal activation functions were trained on the same data. The networks had 16, 32, 64, 128, or 256 hidden units. Recognition performance of the best flat network, which had 32 hidden units, is also shown in Fig. 7. It substitutes 35 (3.18%) of the 1,099 test examples in the zero-reject case and is outperformed by the hierarchical network for higher reject rates as well.

3. DIGIT RECOGNITION

The block classifier is complemented by a digit recognition system as illustrated in Fig. 1. A separate classifier is used for the left and the right digit of interest, since they have different a-priori class distributions and are embedded in different context. Both receive the output of the block classifier for the other digit as contextual input, in addition to the preprocessed digit. They are queried only if the block classifier is not confident enough and rejects an example. Digit recognition consists of three steps: digit preprocessing, digit classification, and combination of the digit outputs with the results of the block classifier.

The preprocessing step, described in detail in [3], segments the digit from the rest of the meter value and normalizes its size to 8×15 pixels.

A three-layered feed-forward network, sketched in Fig. 6, is used for digit classification. It contains in the second layer 10 context units, in addition to 32 hidden units. The network is trained using gradient descent. Afterwards, the performance on the training set is almost perfect. From the test set 4 (.36%) of the left digits and 15 (1.36%) of the right digits are substituted when none are rejected. Fig. 6(b) shows some example inputs and outputs of the right-digit classifier. Analysis of the problematic cases reveals that segmentation problems, missing digit parts, and unusual context are the most frequent reasons for substitutions. As control experiment, the same network was trained without access to the context information and classification performance degraded to 1.91% zero-reject substitutions for the left digit and 7.25% for the right digit.



flat block classifier			hier. block classifier			combined classifier		
% ok	% sub	% rej	% ok	% sub	% rej	% ok	% sub	% rej
96.8	3.2	0	98.0	2.0	0	99.1	.91	0
95.3	2.0	2.73	96.5	1.0	2.4	98.9	.45	.64
93.4	1.0	5.64	93.1	.54	6.4	96.4	.27	3.37
90.8	.45	8.74	89.1	.18	10.7	96.3	.18	3.55
89.1	.36	10.6	84.3	.09	15.6	86.9	.09	13.0
80.7	.18	19.1	74.2	0	25.8	78.6	0	21.4

Fig. 7. Test set performance of the different classifiers.

The outputs of the digit classifiers are combined with the ones of the block classifier by computing the average of the corresponding output vector sections.

The test set performance of the combined classifier is shown in Fig. 7. If all outputs are accepted, only 10 (.91%) examples are substituted. The substitution rate can be lowered to .45% if only .64% of the examples are rejected and to .18% for 3.55% rejects. Thus, adding the digit classification stage significantly improved performance. Fig. 8 illustrates the combined recognition for some problematic examples. The left part of the figure contains some examples for which the person labeling them did not assign a valid value. Although fairly hard to read, they were successfully recognized by the combined classifier.

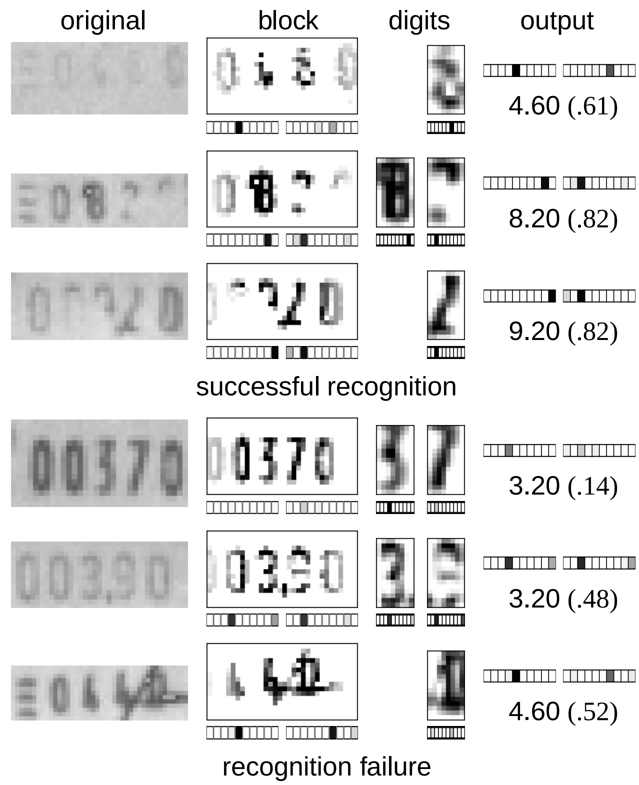


Fig. 8. Problematic test examples for combined classifier.

4. CONCLUSIONS

This paper described a meter value recognition system consisting of two stages: block recognition and digit recognition. The hierarchical block classifier with local connectivity recognizes two digits of interest simultaneously within their context. It performs significantly better than a flat neural classifier.

If block recognition cannot make a confident decision, the system focuses its attention on the ambiguous digits by presenting them to a digit classifier that has access to the block classification output for the other digit as context.

The system was evaluated using a database of Swedish Post meter values. It performs well. If the given region of interest contains a readable meter value, it can be recognized with high accuracy. Even meter values challenging for trained humans can be read. Analysis of problematic test examples revealed that the recognition performance could be improved further if the training set were larger, such that rare labels were better represented, and if the region of interest contained only the digits of the meter value and no additional objects.

5. REFERENCES

[1] S. Behnke and R. Rojas, “Neural Abstraction Pyramid: A hierarchical image understanding architecture,” in *Proceedings IJCNN’98*, 1998, pp. 820–825.

[2] Sven Behnke, “Hebbian learning and competition in the Neural Abstraction Pyramid,” in *Proceedings of IJCNN’99 – Washington, DC, paper #491*, 1999.

[3] Sven Behnke, “Hierarchical neural networks for image interpretation,” Dissertation thesis, Freie Universität Berlin, 2002.

[4] K. Fukushima, S. Miyake, and T. Ito, “Neocognitron: A neural network model for visual pattern recognition,” *IEEE Trans. SMC*, vol. 13, pp. 826–834, 1983.

[5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. of IEEE*, vol. 86(11), pp. 2278–2324, 1998.