The problem of "brittleness" of convolutional neural networks in recognition of digit patterns with different writing styles

Nikolai Kuzmitsky[1], Stanislav Derechennik[2]
Brest State Technical University
Moskovskaja str. 267, Brest, 224017 Republic of Belarus
[1]knnbrest@yandex.ru, [2]cm@bstu.by

Abstract *- This paper describes the research on the problem of maintaining high accuracy of convolutional neural networks in recognition of digit patterns, writing style of which is different from style of training patterns. To overcome this problem we propose an approach based on the creation of experts that specialize in recognition of patterns of one of the basic types and integration of their knowledge in a committee. The used architecture of experts, methods of their training, techniques of pattern preprocessing and committee voting allow reaching higher average recognition accuracy on the used databases with the help of our approach rather than with similar ones.*

Keywords *– convolutional network, brittleness, regularization, integration, committee, generalization*

I. Introduction

The main goal of digits recognition tasks is the creation of machines, which would have technologies for reading visual numerical data with speed and accuracy not lower than the human ones [1]. How close are scientific results to the achievement of the given goal?

Technologies of isolated digit patterns analysis have received increased attention since the first years of research in the OCR field, because there are many areas of their potential application. Since then, the scientific community has accumulated a wealth of information in the form of classification methods, feature extraction techniques, databases, etc., which are described in numerous literatures [2], [3]. Neural networks have been developing especially intensively in the framework of the given subject matter. They have many advantages: automatic extraction of features, stability to noisy data, possibility of effective implementation on hardware, etc. [4].

So, at the end of the 1990s on the basis of multilayer perceptrons and back-propagation algorithm LeCun developed a convolutional model of neural network, which, in some author's opinion, is best suited for solving visual document analysis tasks [5], [6]. He also created the famous MNIST database, which is an important point of reference for comparing models of classifiers. The latest results in the recognition of the MNIST test part (27 errors per 10,000 examples [7]) showed a promising application of CNN for achieving the above mentioned main goal. However, there still remains an open question: what is the efficiency of a CNN classifier trained on one database, in recognizing patterns from other databases?

The analysis of the current literature revealed a significant lack of research aimed at solving this question. Among them we can highlight only Seewald's paper [8], in which on the example of three databases of handwriting patterns (MNIST, USPS, DIGITS) and different feature extraction methods he showed the vulnerability of classifiers based on k-NN, SVM and partly on CNN models in their ability to transfer knowledge from one control set to another. He called this problem "brittleness" (weakness in AI terminology) and in relation to CNN it is the main object of interest in the described research. At the same time great attention in this paper is paid to the ability of CNN to recognize digit patterns of different types and to increasing the efficiency of transferring CNN knowledge between different databases.

II. REGULARIZATION OF ISOLATED DIGIT IMAGES

Rapid development of OCR technology over the past 20 years has led to the emergence of intellectual products such as text recognition systems, postal and banking documents processing systems which satisfy practical needs [3], [9]. However, it should be noted that the narrowing of admissible variation of the input data has often been the key to effective implementations.

A. Types of images of isolated digits

Depending on the method of origin of isolated digit images there can be distinguished their three most essential types: machine-printed, handwritten and synthesized, examples of which are shown in Fig. 1.

The sources of the first group images can be scanned documents containing font text information. The main feature of these images is complying with the rules of symbols writing and their low variation, which made it possible to develop effective recognition methods.

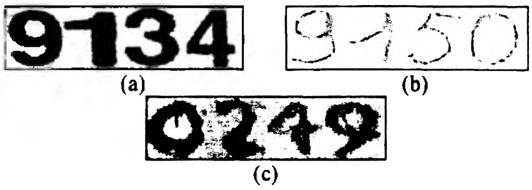

Fig. 1. Examples of isolated digit images of the singled out types: machine-printed (a), handwritten (b), synthesized (c).

In view of the individual writing style of every man (tilt, pen width, etc.), images of the second type cause difficulties in solving the OCR tasks, and it has not allowed creating a universal recognition technology of these images up to the present moment, although there are examples of accuracy similar to the human one on separate test sets [7]. Images of the third type are the result of applying spatial transformations to the images of the first two groups, which allows simulating distortions caused by adverse environmental conditions or by equipment defects. Thus, in this research KNI database was used, for the creation of which we used "wave distortion" formed with help of the following functions:

$$f_i(x) = (-1)^i \times a_i \times \cos(\pi / l_i \times x), \qquad (1)$$

where a_i is an amplitude, l_i is a length of the i-th function.

The selected types of images served as prototypes for the creation of three experts, by which we understand neural networks specializing in recognition of images of one of three types. Research on the problem of "brittleness" in relation to the CNN model on the basis of experts was conducted. At the same time for CNN training/testing, we used databases presented below.

B. Databases of digit patterns

MNIST [10] is a subset of the NIST [11] database that contains images of handwritten digits, divided into training (60,000 patterns) and test (10,000) parts. The images were received from respondents of the census bureau and students of education institutions of the USA; and patterns from different authors were put in different MNIST parts. The patterns from NIST were scaled to a rectangle of size 20x20, which was subsequently placed in an image of size 32x32 pixels, with coincidence of the centre of gravity of the symbol and the geometric centre of the image. It should be noted that on the NIST basis we created a new database called NIST_HSF4 that contains 55,000 patterns not included in MNIST.

In researches we also used the following databases:

- handwritten: USPS [12] – includes 7,291 training part patterns and 2,007 test part patterns of the database of same name, each with 16x16 pixels; OPTDIGTS [13] – contains all the 5,620 patterns of the database of the same name, each with 32x32 pixels;
- machine-printed: combined into one database under the name of FONT, consisting of a training part, which includes 5,008 patterns of computer fonts with normal and bold writing styles and 18,660 patterns from [14], and a test part, which contains 5,008 patterns, each with 128x128 pixels, of fonts with italic writing style;
- synthesized: KNI – includes training (60,000 patterns) and test (10,000) parts, which were obtained with the help of our own software tools on the basis of the wave distortions and at most ±20° rotation of computer font patterns of size 32 x32 pixels, with the control of patterns pairwise differences in each class.

Examples of databases images are shown in Fig. 2.

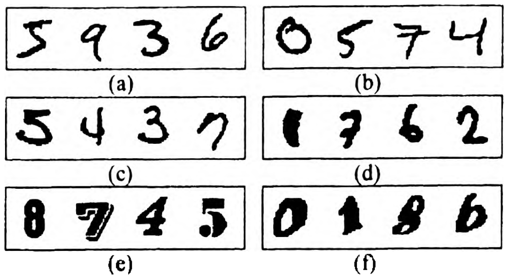
(a) (b) (c) (d) (e) (f)

Fig. 2. Images of patterns of used databases: MNIST (a), NIST_HSF4 (b), USPS (c), OPTDIGTS (d), FONT (e), KNI (f)

III. CREATION OF CNN EXPERTS

The architecture LeNet-5 [5] was selected as a base for the construction of CNN experts, because it is the most developed in this class of neural networks, its training can be performed on the equipment with standard computing resources. Each network has the input layer with 32x32 neurons. The first hidden layer is a convolutional layer with 6 feature maps of 28x28 neurons and size 5x5 filters. It is followed by the sub-sampling layer, which averages responses of neurons of the previous layer with the help of non-overlapping filters of size 2x2 and connects its neurons with 16 maps of 10x10 neurons from the next convolutional layer. The second sub-sampling layer reduces the size of the feature maps down to 5x5. The next layer has 120 neurons and performs full-connection interaction with the previous one. The output layer has one neuron per class, i.e. 10 for digits. The described architecture is shown in Fig. 3. Note that this architecture is more simple than LeNet-5, as in it there is no RBF layer, because it is better to use this layer with a large number of classes.

Neural networks were trained in a full online mode with help of modification of back-propagation algorithm, which bases on Levenberg-Marquardt method [15]. Learning rate evenly decreased from 0.001 down to 0.000001 within 68 epochs. Before each of them patterns from the training set had been deformed with elastic (parameters: σ = 8, α = 50 [6]) and affine distortions (at most ±15° rotation, for images of digits '1' and '7'– ±7°, and 15% scaling, for each dimension separately). Note that distortions are an important factor in improving network generalization ability. Besides they insure against overtraining, which can be caused by a violated Vapnik-Chervonenskis inequality relating the number of model parameters and the size of the training set [16].

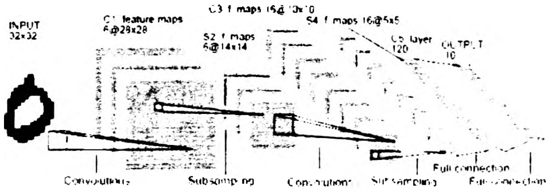

Fig. 3. The used CNN architecture.

In view of the fact that the patterns of the used databases differ not only in writing style, but also in geometric size and the distribution of brightness over the image, we decided to preprocess them according to the circuit shown in Fig. 4. Note significant importance of the patterns preprocessing in overcoming the problem of "brittleness", which was also mentioned in [8].

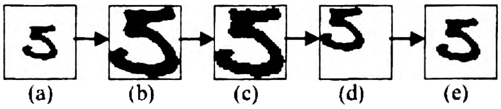

Fig. 4. Input pattern (a) after preprocessing steps: scaling to 32×32 (b), binarization (c), scaling to 20×20 (d), centering (e).

For the training of three CNN experts handwritten (MNIST), machine-printed (FONT) and synthesized (KNI) databases were used, which hereinafter are called the main databases. The results of testing the experts on training and test sets of these databases are presented in TABLE I and TABLE II correspondingly.

The analysis of the table's data shows the following:

1) The used CNN architecture and its training technique allowed the experts to reach pattern recognition accuracy above 99% on their own databases;

2) The highest average error rate (9.20%) was shown by the experts on the KNI_train set;

3) The best average accuracy (93.22%) on all the test sets belongs to the expert trained on the MNIST;

4) The average accuracy of the experts (92.58%) on all the test sets confirms the relevance of the problem of "brittleness" in relation to CNN.

TABLE I
ACCURACY (%) OF THE CNN EXPERTS ON TRAIN SETS

Expert	MNIST train	FONT train	KNI train	Aver
CNN MNIST	99.68	96.75	85.21	93.88
CNN FONT	86.68	99.88	88.34	91.63
CNN KNI	86.48	94.54	99.38	93.46
Aver	90.94	97.05	90.97	92.99

TABLE II
ACCURACY (%) OF THE CNN EXPERTS ON TEST SETS

Expert	MNIST test	FONT test	KNI test	Aver
CNN MNIST	99.39	95.23	85.06	93.22
CNN FONT	87.23	99.58	88.06	91.62
CNN KNI	86.52	92.97	99.29	92.92
Aver	91.04	95.92	90.80	92.58

IV. INTEGRATION OF EXPERT KNOWLEDGE

To overcome the problem of "brittleness" let's try to integrate knowledge of experts. There are two possible approaches: 1) creation of a CNN expert, training of which can be performed on a unified set of training parts of the three main databases; 2) creation of an expert in the form of a committee of three private CNN with combining their responses. In favor of the latter approach, we can give the following arguments:

1) Training three private CNN experts requires less time than training one expert on a unified set, while the accuracy of a committee, as a rule, is at the level not lower than that of one expert [17];

2) The distinction between types of training sets of experts suggests low correlation of their errors on other sets, and hence, there is a high probability of increasing the average accuracy of a committee on them;

3) Private experts increase the variety of the extracted features providing a broader analysis of the input data.

In addition, good preconditions for the formation of a committee are created by the same neural network architecture of experts, teaching methodology and pattern preprocessing, providing normalization of expert responses in a single numeric range.

An important efficiency factor of a committee is voting of its members [18]. There are different voting schemes in which votes of members in the form of predicted labels of classes, prediction probability, etc. can be applied. In this research, we used the following schemes:

1) Maximum voting – choosing the class with the maximum response of the CNN experts;

2) Average voting – choosing the class with the highest average response of the CNN experts;

3) Majority voting – choosing the class with most votes of the CNN experts.

On the basis of the CNN experts we made three committees, which were named in accordance with their voting schemes: Max_com, Aver_com, Major_com. The results of testing the committees on main databases and control databases are presented in TABLE III and TABLE IV correspondingly. It should be noted that in order to compare the efficiency of our committees the results of KADMOS system – character recognition software component are also presented in these tables [19]. The free download KADWOS MINI SDKs contains a classifier ('numbers_us.rec') of handwritten and machine-printed digit patterns and its accuracy is reflected in the tables.

As it can be seen from the tables, Max_com is the most effective among the created committees both on the main and on control databases. The analysis of Max_com results leads to the following conclusions:

TABLE III
ACCURACY (%) OF THE COMMITTEES ON MAIN DATABASES

Committee	MNIST test	FONT test	KNI test	Aver
Max com	98.23	98.62	98.56	98.47
Aver com	97.64	99.11	98.10	98.28
Major com	93.82	98.50	98.86	97.06
KADMOS	95.84	98.68	84.95	93.15

TABLE IV
ACCURACY (%) OF THE COMMITTEES ON CONTROL DATABASES

Committee	NIST HSF4	OPTDIGTS	USPS	Aver
Max com	97.19	94.51	97.80	96.50
Aver com	96.37	93.73	96.52	95.54
Major com	92.01	87.68	92.95	90.88
KADMOS	94.45	94.66	97.09	95.4

1) The committee allows increasing the average accuracy of experts by 5.89% on the test sets of main database, which confirms the efficiency of the chosen approach of knowledge integration;

2) The average accuracy on sets of control databases (96.50%) shows a good ability of the committee to generalization and its stability;

3) The average accuracy of the committee on sets of all the databases (97.48%) exceeds that of the KADMOS system (94.27%) by 3.21%, which proves better efficiency of our committee.

Thus it can be stated that integration of knowledge of the private CNN experts in the committee significantly improves recognition accuracy. For comparison, using the second approach to knowledge integration Seewald in [8] obtained the average accuracy for the CNN model on three handwritten databases 2.73% less than the committee on sets of patterns with different writing styles. However, it should be noted that our result is still not enough for the universality of the sphere of using the committee. So not high recognition accuracy on the OPTDIGTS database (94.51%) revealed the vulnerability of the created committee to thickness of digit patterns and the necessity of its further improvement.

Summing up the work carried out, we can say that CNN is a very promising, but not fully developed mechanism for recognition of digit patterns with different writing styles. Therefore the chosen approach needs further researches.

V. CONCLUSION

In this paper the actuality of the problem of "brittleness" in relation to the CNN model in recognition of digit patterns with different writing styles is shown. The proposed approach to overcoming this problem with the help of a committee of experts proved its perspectives. At the present time experiments are being conducted to improve the efficiency of committees. They are based on the increasing of regularization of digit patterns sets of the created types, recognition accuracy of private CNN experts. Preliminary results indicate that the chosen approach to achieving the main goal, i.e. creation of a universal classifier of digit patterns with different writing styles, has good potential.

REFERENCES

[1] C.-L. Liu, H. Fujisawa, "Classification and learning in character recognition: Advances and remaining problems", Springer, pp.139-161, 2008.

[2] C.-L. Liu, K. Nakashima, H. Sako, H. Fujisawa, "Handwritten digit recognition: Benchmarking of state-of-the-art techniques", Pattern Recognition, 36(10), pp. 2271-2285, 2003.

[3] H. Fujisawa, "Forty years of research in character and document recognition - an industrial perspective", Pattern Recognition 41(8), pp. 2435-2446, 2008.

[4] V.A. Golovko, "Neural networks: training, organization and application" [in Russian], Moscow: IPRZHR, 2001.

[5] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-Based Learning Applied to Document Recognition", Proceedings of the IEEE, 86(11), pp. 2278-2324, 1998.

[6] P. Simard, D. Steinkraus, and J. Platt, "Best practices for convolutional neural networks applied to visual document analysis", ICDAR, pp. 958–963, 2003.

[7] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Convolutional neural network committees for handwritten character classification". ICDAR, pp. 1250–1254, 2011.

[8] Alexander K. Seewald, "On the Brittleness of Handwritten Digit Recognition Models", ISRN Machine Vision, vol. 2012, Article ID 834127, 10 pages, 2012. doi:10.5402/2012/834127.

[9] R. Palacios, A. Gupta, "A system for processing handwritten bank checks automatically", Image and Vision Computing, vol. 26, no. 10, pp. 1297-1313, 2008.

[10] MNIST database, http://yann.lecun.com/exdb/mnist/index.html.

[11] Grother P.J., "Nist special database 19 – handprinted forms and characters database", National Institute of Standards and Thechnology (NIST), Tech. Rep., 1995.

[12] Hastie, T., Tibshirani, R., Friedman, J. "The Elements of Statistical Learning. Data Mining, Inference and Prediction", Springer, New York, 2001.

[13] Optdigits database, http://mlearn.ics.uci.edu/databases/optdigits/.

[14] J.J. Weinman, E. Learned-Miller, A. Hanson, "Scene text recognition using similarity and a lexicon with sparse belief propagation", IEEE Trans. on PAMI, 31 (10), pp. 1733–1746, 2009.

[15] Y. LeCun, L.Bottou, G. Orr,and K. Müller. "Efficient BackProp, in Neural Networks: Tricks of the Trade", Springer Lecture Notes in Computer Sciences, N 1524, pp. 5–50, 1998.

[16] C. M. Bishop, "Neural Networks for Pattern Recognition", Oxford University Press, 1995.

[17] K. Chellapilla, M. Shilman, and P. Simard, "Combining multiple classifiers for faster optical character recognition", Document Analysis Systems VII. Springer Berlin, pp. 358–367, 2006.

[18] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers,", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 3, pp. 226–239, 1998.

[19] KADMOS recognition software. http://www.rerecognition.com.