IEEE Second International Conference on Data Stream Mining & Processing
August 21-25, 2018, Lviv, Ukraine

Deep Neural Network for Image Recognition Based on the Caffe Framework

Myroslav Komar
Research Institute for Intelligent Computer Systems
Ternopil National Economic University
Ternopil, Ukraine
mko@tneu.edu.ua

Pavlo Yakobchuk
Research Institute for Intelligent Computer Systems
Ternopil National Economic University
Ternopil, Ukraine

Vladimir Golovko
Department of Intelligent Information Technologies
Brest State Technical University
Brest, Republic of Belarus
gva@bstu.by

Vitaliy Dorosh
Research Institute for Intelligent Computer Systems
Ternopil National Economic University
Ternopil, Ukraine

Anatoliy Sachenko
Kazimierz Pulaski University of Technology and Humanities in Radom
Radom, Poland
sachenkoa@yahoo.com

***Abstract*— Deep Leaning of the Neural Networks has become one of the most demanded areas of Information Technology and it has been successfully applied to solving many issues of Artificial Intelligence, for example, speech recognition, computer vision, natural language processing, data visualization. This paper describes the developing the deep neural network model for image recognition and a corresponding experimental research on an example of the MNIST data set. Some practical details for creating the Deep Neural Network and image recognition in the Caffe Framework are given as well.**

Keywords*— *Deep Neural Network, Information Technology, Image Recognition, Artificial Intelligence, Caffe Framework

I. INTRODUCTION

In order to proceed efficiently with large amounts of data at the acceptable time, special information technologies are needed. Nowadays such information technologies can be represented by Deep Neural Networks [1-8], which have the greater efficiency of the non-linear transformation and data representation in comparison with traditional neural networks. A Deep Neural Network performs a deep hierarchical transformation of images in the input space. Moreover the Deep Neural Networks, thanks to the multi-layer architecture, enable to process and analyze the large amount of data, as well as modeling the cognitive processes in various fields. Currently, most high-tech companies in the US (Microsoft, Google, Facebook, Baidu, etc.) use deep neural networks to design the various intelligent systems. According to the scientists of the Massachusetts Institute of Technology, deep neural networks are on the list of the 10 most promising high technologies capable in the near future to largely transform the everyday life of most people on our planet and solve many problems of artificial intelligence, for example, speech recognition, computer vision, natural language processing, data visualization, etc. [9-13, 26].

II. RELATED WORKS

In 2006, Hinton proposed a greedy layer-wise algorithm [1], which became an effective tool for teaching deep neural networks. It was shown that a deep neural network has a greater efficiency of the non-linear transformation and data representation in comparison with a traditional perceptron. This network performs a deep hierarchical transformation of the input space. As a result, the first hidden layer separates the low-level space of attributes of the input data, the second layer detects the space for attributes for a higher level of abstraction, etc. [14]. Currently there are many works devoted to the recognition of images by means of deep neural networks [15-19]. Authors [15] present the Maxout network in Network architecture. Their approach is based on the convolutional layer and a two layer maxout MLP and it's used to convolve the input and average pooling in all pooling layers. In [16], the architecture of the deep neural network is applied in biology domain. The small receptive fields of convolutional winner-take-all neurons yield large network depth are resulting in roughly as many sparsely connected neural layers. Ikuro Sato in [17] offers an optimal decision rule for a given data sample using classifiers that are trained on extended data. A paper [18] reports to introducing the DropConnect, a generalization of Hinton's Dropout for regularizing large fully-connected layers within neural networks. As a result authors derive a bound on the generalization performance of both Dropout and DropConnect. A simple and effective stochastic pooling strategy is developed [19] to secure over-fitting during the training deep convolutional networks. According to the MNIST [21, 22] the best generalized recognition accuracy was 99.79% [18]. So, we propose below how to improve this value.

III. STRUCTURE OF DEEP NEURAL NETWORK

For the implementation of the above-mentioned architecture, we used Caffe deep learning library [20]. The main advantage of Caffe is the speed of operation. The framework supports CUDA and, if necessary, can switch the processing flow between the processor and the graphics card. The process of training the Deep Neural Network in framework Caffe has been lasted 30 epochs and finished with achieving the given accuracy of learning (Fig. 1). The deep neural network consists of the following layers (Fig. 2): 1st layer – Convolution (out filters: 24, size: 5x5, stride: 1x1); 2nd layer – Convolution (out filters: 12, size: 5x5, stride: 1x1); 3rd layer – Pooling (size: 2x2, stride: 2x2); 4th layer – Convolution (out filters: 8, size: 5x5, stride: 1x1); 5th layer – Convolution (out filters: 4, size: 5x5, stride: 1x1); 6th layer – Pooling (size: 2x2, stride: 2x2); 7th layer – InnerProduct (out:

500, filter: xavier); 8th layer – InnerProduct (out: 0, filter: xavier); 9th layer – Softmax (out: 10, activation: softmax).

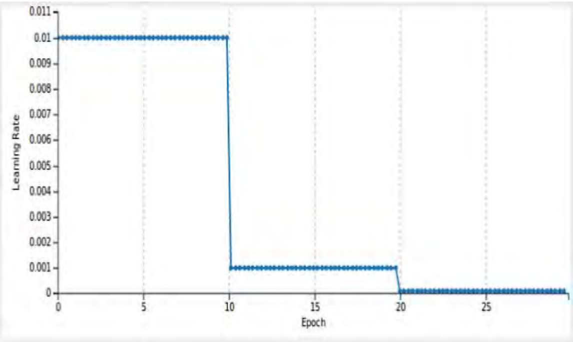

Fig. 1. The process of learning the deep neural network

IV. EXPERIMENTAL RESULTS

The image from the test sample of the MNIST data set [21, 22] was used for case study. The MNIST data set consists of 28x28 pixel handwritten digital images organized in 10 classes (0 to 9) with both 60,000 training and 10,000 test samples. Testing on this data set has performed without increasing the data. Results, of the image recognition for the number 0 to 9 from the test sample are show in Fig. 3, the generalized recognition accuracy was 99.93%. The visualization of the image, which is applied to the input of the Deep Neural Network is illustrated by Fig. 4. Fig. 5 shows the visualization of the image processing on the first convolution layer, 24x24x24 (24 functional maps with the element 24x24).

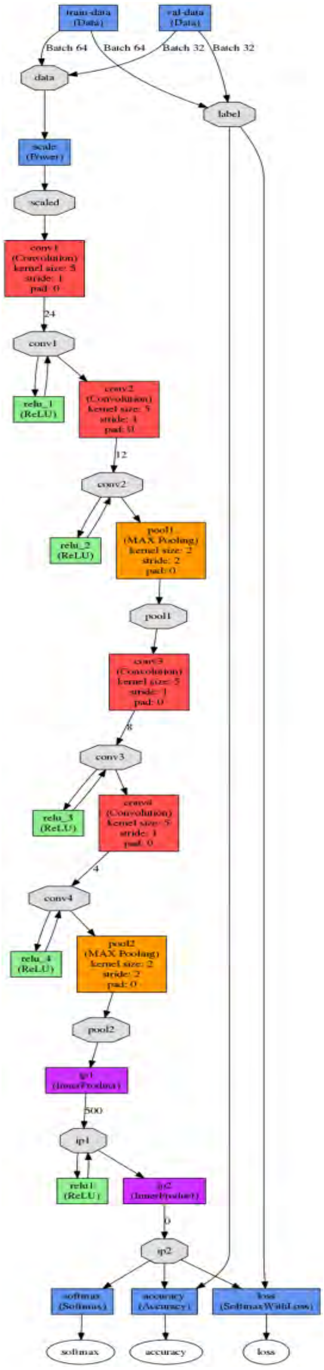

Fig. 2. The structure of the deep neural network

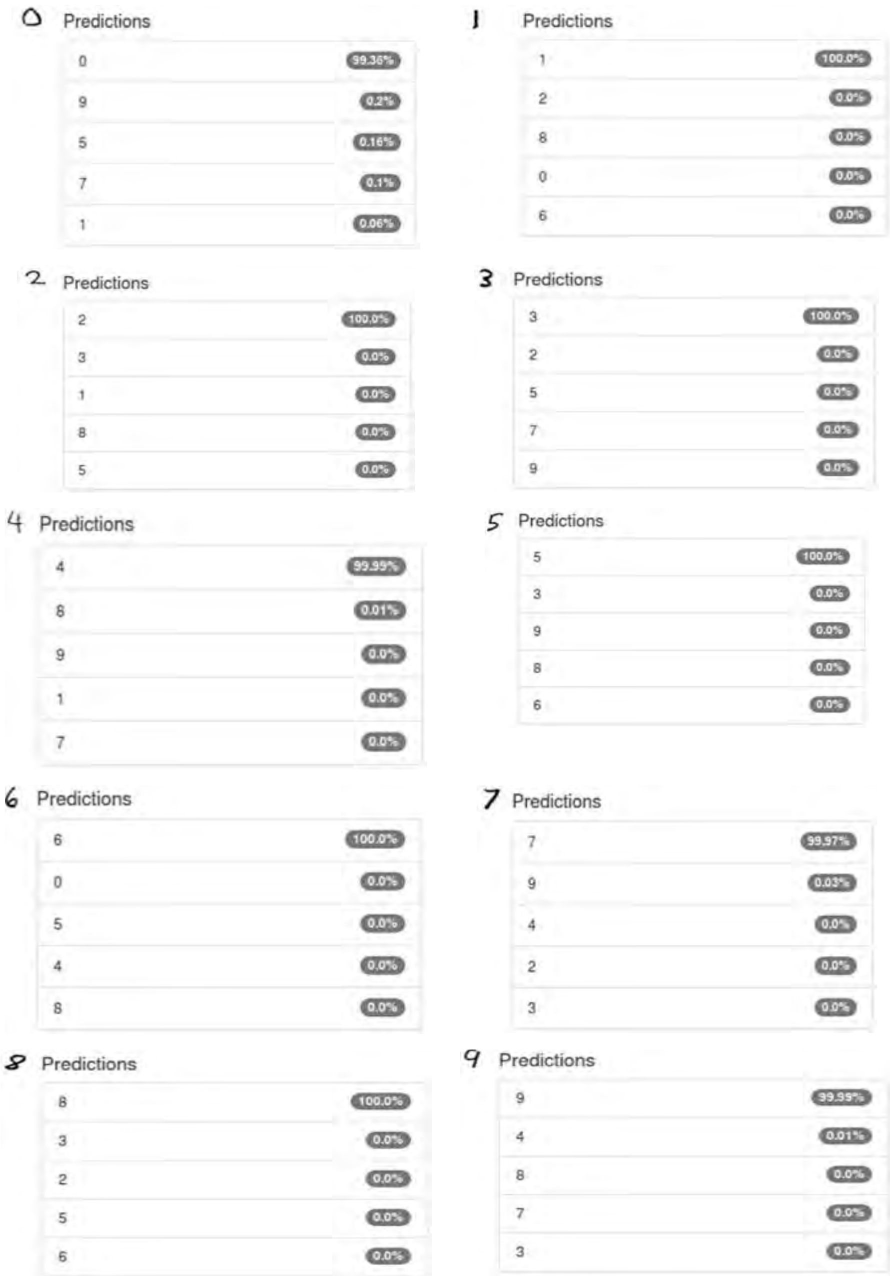

Fig. 3. Results of image recognition from the test sample

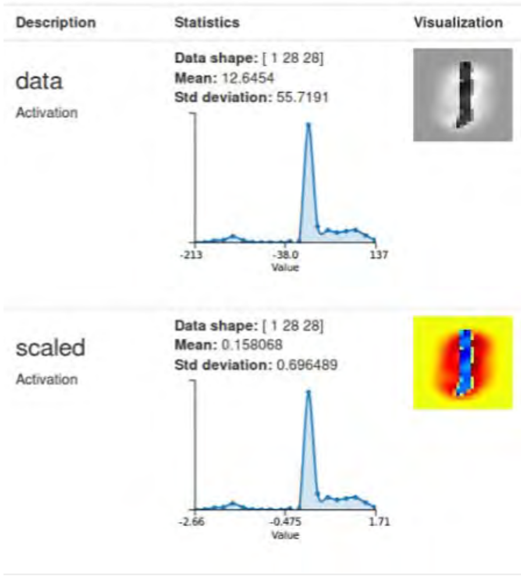

Fig. 4. Visualization of the input image

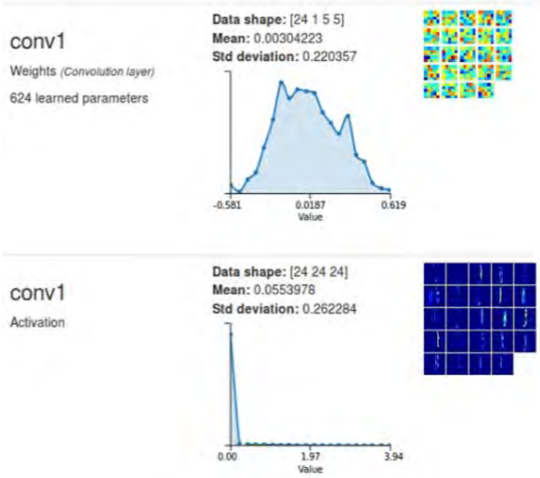

Fig. 5. Image processing on the first convolutional layer

The visualization of image processing on the second convolution layer, 12x20x20 (12 functional maps with the element 20x20) is illustrated by Fig. 6. Fig. 7 shows the visualization of image processing on the first layer of the spatial association.

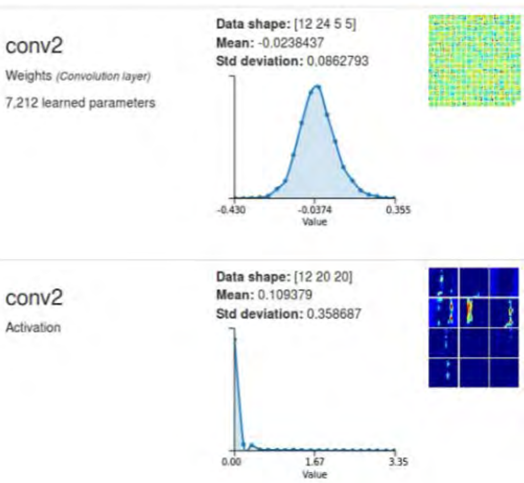

Fig. 6. Image processing on the second convolutional layer

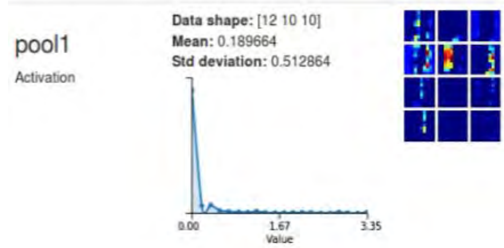

Fig. 7. Image processing on the first layer of the spatial association

The visualization of the image processing on the third convolution layer, 8x6x6 (8 functional maps with the 6x6 element) is illustrated by Fig. 8. Fig. 9 shows the image processing of the image on the fourth convolution layer, 4x2x2 (4 functional maps with the element 2x2).

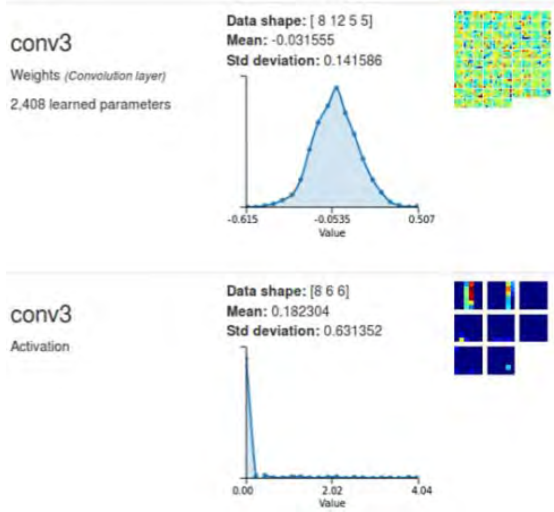

Fig. 8. Image processing on the third convolutional layer

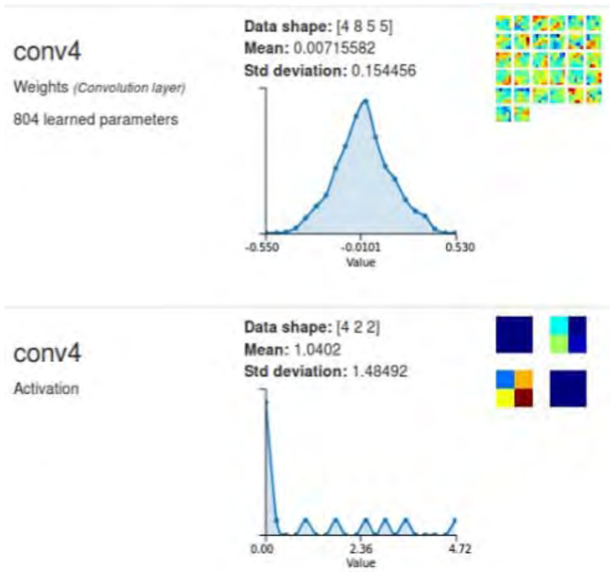

Fig. 9. Image processing on the fourth convolutional layer

In Fig. 10 the visualization of image processing on the second layer of spatial association is illustrated.

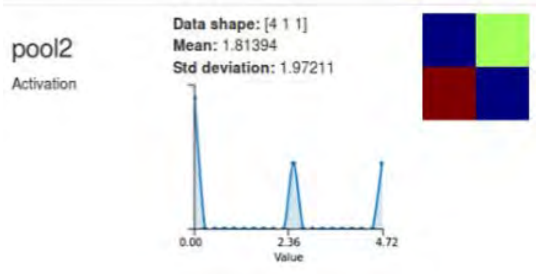

Fig. 10. Image processing on the second layer of the spatial association

The visualization of image processing on the first, second and Softmax full layer is illustrated be Fig. 11, 12 and 13 correspondingly.

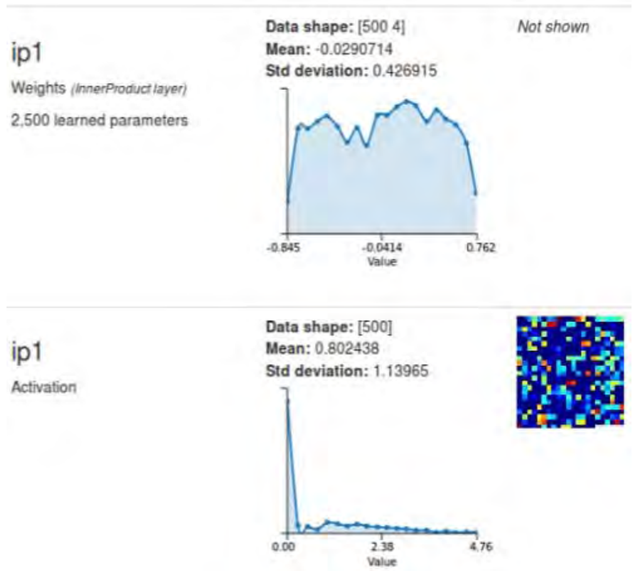

Fig. 11. Image processing on the first full layer

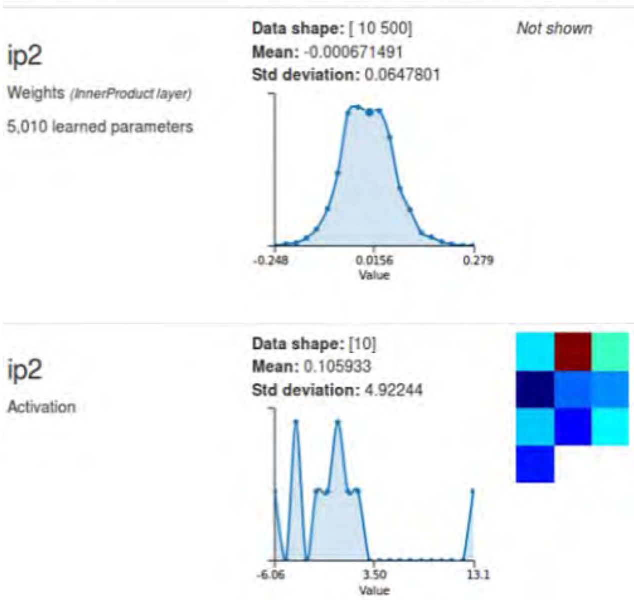

Fig. 12. Image processing on the second full layer

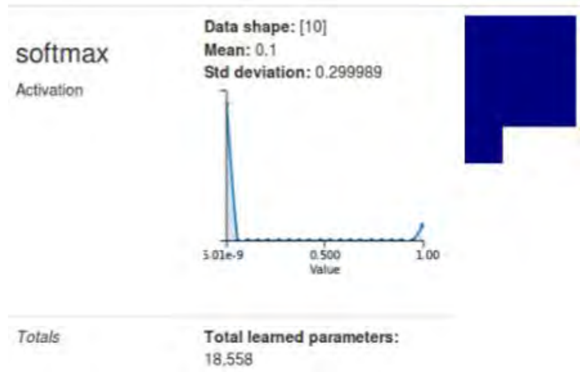

Fig. 13. Image processing on the Softmax layer

As it comes from Fig. 3 above the reached generalized recognition accuracy is equal 99.93%, that is in 0,14% better in comparison with a work [18].

V. Conclusion and Future Work

Authors proposer a model of the Deep Neural Network for the recognizing the images of handwritten digits, using the structure of the neural network in the Caffe Framework. Experimental results have been carried out on an example of the MNIST data set and the generalized recognition accuracy was 99.93%.

Employing the deep neural network in Big Data processing is one of perspective direction for a future research. Moreover, it is planned to conduct experimental research on the following data sets CIFAR-10/100 [23], SVHN [24], and ImageNet [25].

Acknowledgment

This work is carried out within the framework of the project of the Ministry of Education and Science of Ukraine «Methods of intellectual processing and analysis of big data based on deep neural networks», 2018-2019 years.

References

[1] G. E. Hinton, S. Osindero, and Y. Teh, “A fast learning algorithm for deep belief nets,” Neural Computation, vol. 18, pp. 1527–1554, 2006.

[2] G. E. Hinton, A practical guide to training restricted Boltzmann machines, Department of Computer Science, University of Toronto, 2010.

[3] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521 (7553), pp. 436–444, 2015.

[4] Y. Bengio, “Learning deep architectures for AI,” Foundations and Trends in Machine Learning, vol. 2(1), pp. 1–127, 2009.

[5] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier networks,” In Proc. of the 14th Int. Conf. on Artificial Intelligence and Statistics (AISTATS), vol. 15, pp. 315–323, 2011.

[6] V. Golovko, A. Kroshchanka, U. Rubanau, and S. Jankowski, “A Learning Technique for Deep Belief Neural Networks,” Communication in Computer and Information Science, vol. 440, pp. 136–146, 2014.

[7] V. Golovko, A. Kroshchanka, V. Turchenko, S. Jankowski, and D. Treadwell, “A New Technique for Restricted Boltzmann Machine Learning,” 8th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS’2015), Warsaw, Poland, pp. 182–186, 24–26 September, 2015.

[8] V. Golovko, A. Kroshchanka, and D. Treadwell, “The Nature of Unsupervised Learning in Deep Neural Networks: A New

Understanding and Novel Approach," Optical Memory and Neural Networks, vol. 25(3), pp. 127–141, 2016.

[9] S. Jankowski, Z. Szymański, U. Dziomin, V. Golovko, and A. Barcz, "Deep learning classifier based on NPCA and orthogonal feature selection," International Conference on Photonics Applications in Astronomy, Communications, Industry, and High–Energy Physics Experiments, Wilga, Poland, pp. 5–9, May 29, 2016.

[10] G. Hinton, at al., "Deep neural network for acoustic modeling in speech recognition," IEEE Signal Processing Magazine, vol. 29, pp. 82–97, 2012.

[11] T. Mikolov, A. Deoras, D. Povey, L. Burget, and J. Černocký, "Strategies for training large scale neural network language models," in Automatic Speech Recognition and Understanding, pp. 195–201, 2011.

[12] A. Krizhevsky, L. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in Neural information Processing Systems, vol. 25, pp. 1090–1098, 2012.

[13] V. Golovko, S. Bezobrazov, A. Kroshchanka, A. Sachenko, M. Komar, and A. Karachka, "Convolutional Neural Network Based Solar Photovoltaic Panel Detection in Satellite Photos," 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2017), Bucharest, Romania, pp. 14-19, September 21-23, 2017.

[14] G. Hinton, and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science, vol. 313 (5786), pp. 504–507, 2006.

[15] Jia-Ren Chang, and Yong-Sheng Chen, "Batch-normalized Maxout Network in Network," arXiv:1511.02583, 2015.

[16] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," 25th IEEE conference on computer vision and pattern recognition (CVPR), New York, pp. 3642-3649, 2012. DOI: 10.1109/CVPR.2012.6248110, 2012.

[17] I. Sato, H. Nishimura, and K Yokoi, "APAC: Augmented PAttern Classification with Neural Networks," arXiv:1505.03229v1, 2015.

[18] L. Wan, M. Zeiler, S. Zhang, Y. Le Cun, and R. Fergus, "Regularization of Neural Networks using DropConnect," Proceedings of the 30th International Conference on Machine Learning, PMLR, vol. 28(3), pp. 1058-1066, 2013.

[19] M. D. Zeiler and R. Fergus. "Stochastic pooling for regularization of deep convolutional neural networks," ArXiv:1301.3557, 2013.

[20] Caffe Deep Learning Framework, http://caffe.berkeleyvision.org, last accessed 15.03.2018.

[21] The MNIST database, http://yann.lecun.com/exdb/mnist, last accessed 15.03.2018.

[22] Y. Le Cun, L. Bottou, Y. Bengio, and P. Haffner. "Gradientbased learning applied to document recognition," Proceedings of the IEEE, vol. 86(11), pp. 2278–2324, 1998.

[23] A. Krizhevsky, and G. Hinton, Learning multiple layers of features from tiny images. Technical report, University of Toronto, 1 (4), 7, 2009. https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf, last accessed 15.03.2018.

[24] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," In NIPS workshop on deep learning and unsupervised feature learning, Granada, Spain, vol. 2011, pp. 5. 2011.

[25] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. FeiFei, "Imagenet: A large-scale hierarchical image database," In CVPR09, pp. 248–255, 2009.

[26] D. T. V. Dharmajee Rao, and K. V. Ramana, "Winograd's Inequality: Effectiveness for Efficient Training of Deep Neural Networks," Intelligent Systems and Applications, vol. 6, pp. 49-58, 2018..