UDC 004.[8+93+94]

INCREASING THE EFFICIENCY OF NEURAL NETWORKS IN RECOGNITION PROBLEMS

Y. A. Bury

Researcher in technical science, Minsk, Belarus, e-mail: edidici@tut.by

Abstract

The article describes the issues of increasing the efficiency of neural networks in terms of their design and coding of input and output signals. The application of multiple signal coding using extrapolation of the input parameters is described on the example of a system of recognition character sequences on images of arbitrary size with a complex background.

An effective combination of multiple positional and configuration-competitive coding for various types of signals makes it possible to achieve performance rates of the building number recognition algorithm of up to 74 images per second in the adaptive learning mode and 218 images per second in the recognition only mode.

The paper also outlines general recommendations for signal coding in artificial intelligence systems based on neural networks.

Keywords: neural networks, convolution neural network, neuroevolutionary learning, image recognition, character recognition, neural network learning, deep learning, reinforcement learning, extrapolation learning, positional coding, configuration coding, single coding, multiple coding, input coding, output coding.

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧАХ РАСПОЗНАВАНИЯ

Я. А. Бурый

Аннотация

В статье рассмотрены вопросы повышения эффективности нейронных сетей с точки зрения их проектирования и кодирования входных и выходных сигналов. Описано применение множественного кодирования сигналов за счёт экстраполяции входного сигнала на примере системы распознавания цепочек символов на изображениях произвольного размера со сложным фоном. Эффективное сочетание множественного позиционного и конфигурационно-конкурентного кодирования для различных типов сигналов позволяет добиться показателей скорости работы алгоритма распознавания номеров зданий до 74 изображений в секунду в режиме адаптивного обучения и 218 изображений в секунду в режиме только распознавания.

Также в работе изложены общие рекомендации по кодированию сигналов в системах искусственного интеллекта, основанных на нейронных сетях.

Ключевые слова: нейронные сети, нейронная сеть свёртки, нейроэволюционное обучение, распознавание изображений, распознавание символов, обучение нейронных сетей, глубокое обучение, обучение с подкреплением, экстраполирующее обучение, позиционное кодирование, конфигурационное кодирование, одиночное кодирование, множественное кодирование, кодирование входных сигналов, кодирование выходных сигналов.

Introduction

The choice of a specific architecture of a neural network and the structure of its interface is an important part of the development of a computer vision system, since the accuracy and speed of the system depends on this choice.

Multilayer neural networks trained with gradient descent are capable to build complex multidimensional regions based on a large number of training parameters. Therefore, they can be used as classifiers in pattern recognition [1].

The highest invariance to distortions of the input signal has been achieved in neural networks of the convolutional type, such as the cognitron [2], neocognitron [3], and the convolution neural network. Convolutional neural networks evolved from a simpler type of neural networks called perceptrons. These are direct signal propagation networks with one hidden layer and a threshold transfer function [4] [5].

Convolutional neural networks have special convolutional layers. The neurons in these layers calculate the weighted sum (convolution) of the signals of their local receptive fields. The *receptive field* of a neuron is a region of the output signal of the previous layer with given sizes and with a center that, as a rule, geometrically corresponds to this neuron in the previous layer.

The *convolution neural network* also has additional pooling layers that perform local averaging of the signal, which greatly simplifies its architecture.

The architecture of convolutional neural networks mimics the work of the visual cortex as closely as possible. In view of this, they are widely studied from the point of view of building computer vision systems. They are successfully used in agriculture for segmenting aerial photographs of agricultural fields [6], for segmenting satellite images [7], in medicine for segmenting images of examined tissues [8], for recognizing emotions from images, for removing noise from images, and for solving many other problems.

However, the scope of application of convolutional neural networks is not limited to image processing tasks. This architecture is also widely used in other areas, such as speech recognition [9] and computer attack detection [10].

Neural networks are characterized by a large consumption of computing resources, especially in cases where high accuracy is required. This problem also exists for other algorithms from the field of artificial intelligence.

Therefore, it is necessary to take into account the limitations on the cost of computing resources, despite the widespread use and progress made in the field of neural networks. This is due to the fact that the maximum decision-making accuracy achieved and the amount of computing resources required for this are related to each other. This problem is especially relevant in embedded and real-time systems.

The purpose of the present research is to reduce the computing resources consumption of recognition systems and improve their performance while maintaining the given accuracy of decision making.

Maximizing of accuracy

Various improvements and sophistications of the system can be applied to improve recognition accuracy. These can be adding new layers and signal paths, applying teams of decision rules, as well as involving additional calculations for errors of the first and second kind with a corresponding slowdown in comparison with the primary version of the algorithm.

The idea of methods based on teams of decision rules is to combine the advantages of various algorithms in a single recognition system using multilevel analysis. First, it is necessary to apply separate algorithms to the recognizable image to obtain intermediate results, which are then used to do the final conclusion about the image belonging to a particular class.

In addition, different algorithms may behave differently depending on the external conditions in which the neural network operates. Different algorithms can be applied depending on these initial conditions, or areas of expertise. So, the entire original feature space is divided into areas of competence depending on the values of certain features defined in the image.

The maximum accuracy of recognition on a given neural network architecture is achieved with positional coding of variables contained in the output signal [11]. Fig. 1 shows an example of the positional and a configuration coding scheme for the number 3, respectively, for the output and input signals of a neural network.

It has also been observed that increasing the complexity of a neural network can have a significant effect on the overall recognition accuracy achievable.

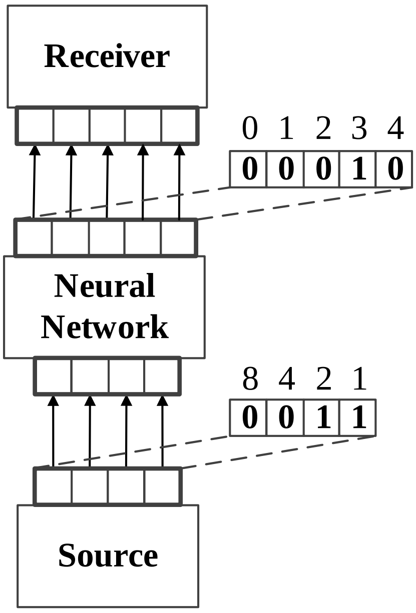

Figure 1 – An example of positional (top) and configuration (bottom) coding of input and output signals of a neural network

Table 1 – Description of the parameters of the convolution neural network 1

Layer number	Layer type	Feature map size	Number of feature maps	Receptive field size	Total number of neurons	Total number of connections
0	Input	28 x 28	1	0	784	0
1	Convolution, TanH	24 x 24	6	5 x 5 x 1	3 456	86 400
2	Pooling	12 x 12	6	2 x 2	864	3 456
3	Convolution, TanH	8 x 8	16	5 x 5 x 6	1 024	153 600
4	Pooling	4 x 4	16	2 x 2	256	1 024
5	Fully connected, output	1 x 1	10	4 x 4 x 16	10	2 560
Total:					6 394	247 040

Table 2 – Description of the parameters of the convolution neural network 2

Layer number	Layer type	Feature map size	Number of feature maps	Receptive field size	Total number of neurons	Total number of connections
0	Input	28 x 28	1	0	784	0
1	Convolution, ReLU	24 x 24	20	5 x 5 x 1	11 520	288 000
2	Pooling	12 x 12	20	2 x 2	2 880	11 520
3	Convolution, ReLU	8 x 8	50	5 x 5 x 20	3 200	1 600 000
4	Pooling	4 x 4	50	2 x 2	800	3 200
5	Fully connected, ReLU	1 x 1	500	4 x 4 x 50	500	400 000
6	Fully connected, output	1 x 1	10	500	10	5 000
Total:					19 694	2 307 720

Optimization experiments were carried out on a character recognition system based on a convolution neural network. The handwritten character base MNIST [12] and the building number base SVHN [13] were chosen as training sets. When the convolution neural network 1 [11] (Table 1) was changed to the convolution neural network 2 (Table 2), the maximum achievable accuracy of the recognition system based on it was increased from 0.9821 on the training and 0.9713 on the test set to 0.9996 on the training and 0.9933 on the test set for the MNIST handwritten character base. The maximum achievable accuracy of the recognition system for the digits of the SVHN building number base was similarly increased from 0.9131 on the training and 0.8902 on the test set to 0.9549 on the training and 0.9184 on the test set.

So, the number of convolution neural network connections increased from 247 040 to 2 307 720, i.e., by a factor of 9.34, while the number of incorrectly recognized characters in the MNIST base fell from 2.87% to 0.67%, i.e., by a factor of 4.28. Similarly, the number of incorrectly recognized characters in the SVHN base dropped from 10.98% to 08.69%, i.e. by 1.26 times.

Minimization of computing resources

Some types of tasks may require low consumption of computing power. For example, this may be important in embedded and real-time systems. These systems must respond in a timely manner to events occurring in the external environment. The main design goal in this case will be to increase the efficiency of the developed algorithms. However, this efficiency depends inversely on the complexity of the system and, accordingly, on the number of its connections.

A slight drop in the accuracy of the subsystems for determining the location, size and other numerical parameters of recognizable images has little effect on the final recognition accuracy. This is due to some invariance of convolutional neural networks to input signal deformations. In addition, the using of positional coding to represent numerical parameters will greatly influence the increase in the complexity of the neural network. This allows to recommend configuration coding of input and output signals in such cases.

There is configuration-competitive coding shows a higher accuracy by 3.41%, compared with configuration-threshold coding, on the same architecture of the convolution neural network [11].

Configuration coding makes it possible to reduce the complexity of the developed systems by reducing the number of their binary inputs and outputs. However, when configuration coding was applied directly to codes of recognizable classes, the decrease in the accuracy of the recognition system in [11] was 4.11%. Therefore, if high recognition accuracy is required, it is recommended to use positional coding to encode object types whenever possible. Configuration coding can be allowed here only in a limited range of tasks. It can, for example, significantly increase the efficiency of algorithms when the number of recognized types is very large.

Additional intermediate transcoding of signals from one representation to another may be appropriate in some cases.

Single and multiple coding of input and output signals of a neural network can be determined depending on the number of sources and receivers of its signals. A common practice is to build separate subsystems for different sources and receivers. Each subsystem can determine one certain characteristic of the image in case of recognition of multiple objects on it. These can be the number of characters in the image, their coordinates, sizes, and other parameters.

Schematic examples of such separate subsystems for recognition of various features are shown in Fig. 2 (a, b and c). They show a single output coding for all subsystems. Information from the outputs of each subsystem is intended for a separate receiver. The difference between them is as follows: Fig. 2a also shows single coding everywhere for the inputs of all subsystems, i.e., all subsystems receive data from different sources; neural networks 2 and 3 in Fig. 2b receive data simultaneously from two sources, i.e. multiple coding is used for their inputs; and Fig. 1c shows the limiting case of multiple coding of inputs of subsystems, when all of them receive signals from all sources at once. The maximum accuracy of decision-making can be achieved, presumably, precisely with such coding, since all neural network subsystems have maximum awareness, i.e., they receive data from all sources of the system. However, all subsystems also spend computing resources on receiving, storing and processing the same information, which is redundant and increases resource consumption.

There is it may be appropriate to generalize and move to the architecture shown in Fig. 2d. It presents the application of multiple coding of all inputs and outputs of a single neural network in the system.

Civil and environmental engineering
doi.org/10.36773/1818-1112-2022-129-3-5-8

All information transmitted to the system from the sources, in this case, will be located in only one place. Resource costs will not be duplicated. The signals for all receivers will contain the result of signal processing from all sources at the same time.

The described neural network, which has a minimal internal structure, will contain a subsystem containing knowledge common to all receivers, and subsystems for processing and storing knowledge specific to each signal receivers.

In [14], to create this architecture, extrapolating training of neural networks was used in [15]. The diagram of the resulting system is shown in Fig. 3.

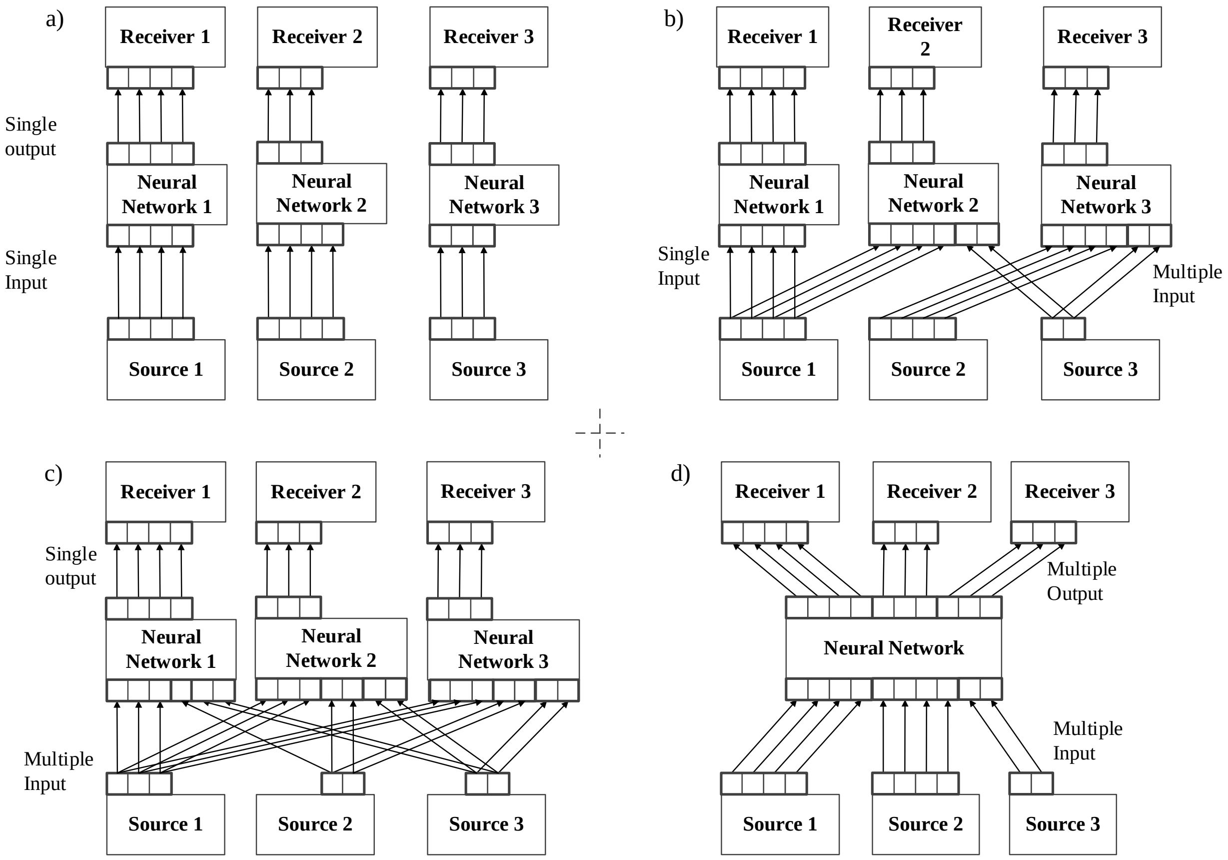

Figure 2 – Single and multiple coding of inputs and outputs of neural network subsystems

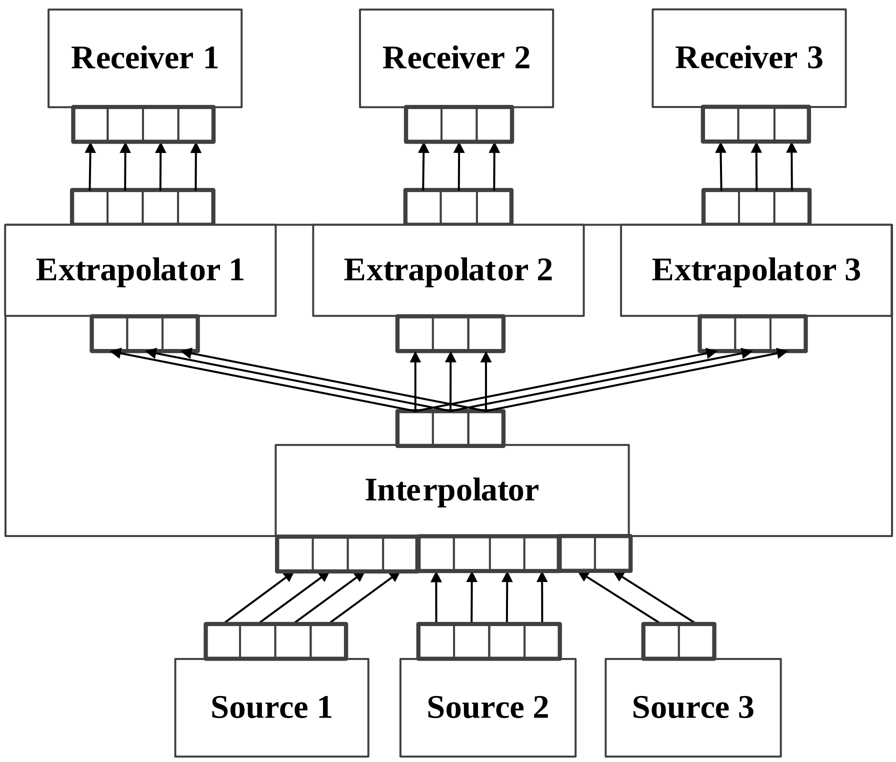

Figure 3 – The internal structure of a neural network with multiple coding of input and output signals

Different learning approaches can be applied to different types of subsystems. So, the interpolator is trained by the method of error back propagation, but extrapolators is trained with using evolutionary methods in [14]. The solution to the problem of determining image characteristics and recognizable characters using multiple coding of inputs and outputs is integrated here into a single recognition system trained using neuro-evolutionary reinforcement learning.

The following speed parameters were obtained with comparable accuracy indicators: the average training iteration duration is 1.7 sec., the population size is 2000 chromosomes, the average processing speed of chromosomes is 1176 per sec., the average processing speed of images with text sequences is 74 per sec. during the evolutionary process and 218 per sec. for an accelerated version of the algorithm. These results were obtained on a workstation CPU Intel Core i7 2.4 GHz, RAM 12 GB, GPU NVIDIA GeForce GT 650M.

Combining several recognition subsystems into one neural network may be associated with some loss of accuracy, as it is described in [15]. However, this effect can be reduced by adding additional neurons to the network. The accuracy achieved when training the basic recognition system was even exceeded by using an additional layer of neurons in the mentioned experiment.

Conclusion

Various optimizations of the neural network underlying the recognition system were tested to increase its accuracy and performance in the present research. Different architectures show different results on the same number of neurons, but despite this, in general, an increase in complexity has a positive effect on the achievable recognition accuracy. Also, the better informative and accuracy of the representation of the parameters encoded in the input signal of the neural network increase the quality of its work. But if its complexity, the number of sources, or the accuracy of data representation are decreased, then the final recognition accuracy also is decreased. Some optimization options show greater efficiency in terms of the ratio of computational cost gained to the loss of recognition accuracy, similar to the considered scheme for applying input signal extrapolation. But some options do not give a gain in efficiency at all. Anyway it is required to find a balance between the performance and work accuracy of the developed systems when implementing algorithms in the general case.

References

1. Kashirina, L. Issledovanie i sravnitel'nyj analiz metodov optimizacii, ispol'zuemyh pri obuchenii nejronnyh setej. Research and comparative analysis of optimization methods used in training neural networks / L. Kashirina, M. Demchenko. – Voronezh: Bulletin of the Voronezh State University, System Analysis and Information Technologies No. 4, 2018. – P. 123–132.
2. Fukushima, K. Cognitron: A self-organizing multilayered neural network/K. Fukushima. Biological Cybernetics 20: 1975. – P. 121–36.
3. Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition uneffected by shift in position / K. Fukushima. – Biological Cybernetics, 36(4), 1980. – P. 193–202.
4. Minsky, M. Perseptrony. Perceptrons / M. Minsky, S. Papert. – M.: Mir, 2007. – 261 s. (in Russian)
5. Rosenblatt, F. The Perceptron: a probabilistic model for information storage and organization in the brain / F. Rosenblatt. – Cornell Aeronautical Laboratory, Psychological Review, vol.65, No. 6, 1958. – P. 386–408.
6. Ganchenko, V. Raspoznavanie sel'skohozyajstvennoj rastitel'nosti na izobrazheniyah zemnoj poverhnosti na osnove svyortochnoj nejronnoj seti U-NET. Recognition of agricultural vegetation on images of the earth's surface based on the convolutional neural network U-NET / V. Ganchenko, A. Dudkin, S. Sheleg. - Joint Institute for Informatics Problems of the National Academy of Sciences of Belarus, proceedings of the conference No. 7-1, 2021. – P. 110–116. (in Russian)
7. Zhuk, A. Ispol'zovanie augumentacij pri segmentacii zdanij na sputnikovyh snimkah s pomoshch'yu nejronnoj seti U-NET. The use of augmentations in the segmentation of buildings on satellite images using the U-NET neural network / A. Zhuk, S. Ablameyko. - Minsk: Belarusian State University, Web programming and Internet technologies WebConf2021, Proceedings of the 5th International Scientific and Practical Conference, 2021. – P. 140–142. (in Russian)
8. Kiselev, V. Segmentaciya pazuh nosa na nabore KT-izobrazhenij. Segmentation of the sinuses on a set of CT images / V. Kiselev, I. Kosik. - Minsk: Belarusian State University, Web programming and Internet technologies WebConf2021, Proceedings of the 5th International Scientific and Practical Conference, 2021. – P. 146–148. (in Russian)
9. Trafimau, A. Metady glybokaga navuchannya dlya raspaznavannya belaruskaga maulennya. Deep learning methods for Belarusian speech recognition / A. Trafimau. - Minsk: Joint Institute for Informatics Problems of the National Academy of Sciences of Belarus, Abstracts of the XI International Scientific and Technical Conference "Information Technologies in Industry, Logistics and the Social Sphere" (ITI), 2021. – P. 148–149. (in Russian)
10. Saskevich, A. Opredelenie MITM-atak (MAN-IN-THE-MIDDLE) posredstvom mashinnogo obucheniya. Determination of MITM attacks (MAN-IN-THE-MIDDLE) through machine learning / A. Saskevich, M. Sterzhanov. – Minsk: Educational Institution "Belarusian State Academy of Communications", Materials of the XXVI International Scientific and Technical Conference "Modern Communications" 2021. – P. 71–73. (in Russian)
11. Bury, Y. Primenenie konfiguracionnogo kodirovaniya vhodnogo signala v nejronnyh setyah svyortki dlya raspoznavaniya rukopisnyh simvolov. Application of configuration input signal coding in convolution neural networks for handwriting digits recognition / Y. Bury, D. Samal. – M.: Belarusian State University of Informatics and Radioelectronics, BigDATA, 2019. – P. 366–371. (in Russian)
12. Images dataset «The MNIST database of handwritten digits». – Available at: http:// yann.lecun.com/exdb/mnist/ (accessed 13.11.2022).
13. Images dataset «The Street View House Numbers (SVHN) Dataset» – Available at: http://ufldl.stanford.edu/housenumbers (accessed 13.11.2022).
14. Bury, Y. Nejroevolyucionnoe podkreplyayushchee obuchenie nejronnyh setej. Neuroevolutionary reinforcement learning of neural networks / Y. Bury, D. Samal. - M.: Belarusian National Technical University, System Analysis and Applied Informatics, No. 3, 2021. – P.16–24. (in Russian)
15. Bury Y. Extrapolating training method for artificial neural networks / Y. Bury, D. Samal, I. Kuharchuck. – M.: UIIP NASB, PRIP, 2019. – P. 387–391.

Accepted 17.11.2022

doi.org/10.36773/1818-1112-2022-129-3-5-8