

# Holographic method of neural network design

Natalya Yakasova<sup>1,2,\*</sup>

<sup>1</sup>Siberian Federal University, 660041 Krasnoyarsk, Russia

<sup>2</sup>Khakassia state University named after N. F. Katanov, 655017 Abakan, Russia

**Abstract.** The paper presents a method for solving the problem of changing a full-connected neural network of direct propagation with a sigmoid activation function in the conditions of the emergence of a new class, with the ability to preserve the network to recognize already known classes and to classify objects of a new class. Such conditions generate a new neural network, which is trained on examples of all classes, including the new class. The learning process takes a long time and requires the selection of several parameters. The emergence of a new class in natural neural networks does not cause a transformation of the network structure, only the strength of connections between neurons changes. The network shows the properties of stability and plasticity at the same time. The authors draw attention to the analogy between neural networks and holograms in their ability to store information and form an image of a class in response to an input signal. Following the holographic analogy, the paper proposes a model of the wave nature of neural networks, which treats the network weights as a hologram and the input signal as a wave passing through a hologram. The construction of a new network is created with two neural networks, which are a combination of two holograms. The first hologram represents the original network, and the second is a new network with a similar structure, but it is trained to recognize one new class. The addition of the holograms of these neural networks implements the mechanisms of plasticity and stability in the model.

## 1 Introduction

Artificial neural networks have become a popular tool for solving many problems. Examples of neural networks show how they perform well in classification tasks on predefined sets of classes. Information systems face the emergence of new classes because of external factors or internal reasons. A neural network becomes incapable of differentiating a new class under such conditions. We face creating a new network capable of recognizing both already known classes and a new class.

The task is to create and train a new neural network, while training remains a very time-consuming process with the selection of many empirical parameters. Parameters in full-link multilayer neural networks affect the learning time and quality. It is impossible to be sure in advance to achieve the expected result in a certain time.

---

\* Corresponding author: [ice@khsu.ru](mailto:ice@khsu.ru)

Forming a network by constructive methods looks like a solution to the problem. They model the mechanisms that are involved in the formation of natural neural networks, providing plasticity and stability.

Existing constructive algorithms offer ways that, through the addition of neurons, create changed networks that can classify new images. Here, network training is required after each addition of neurons or neuron complexes. The disadvantage of such methods is the necessity of strictly controlling the appearance of an excessive number of neurons [1, 2].

Constructive methods are suitable at the beginning of neural network formation when each new class can generate one or more new neurons in the network structure, which at the initial stages does not affect significantly the total number of neurons and training time. As the number of classes grows, the number of neurons and the number of connections between them increases. With an excessive number of neurons and many connections, the neural network loses its ability to generalize and reacts to noise [3]. While the existing number of neurons and network layers is enough to recognize more classes.

Observation of natural neural networks shows that the network structure changes not so often and dramatically. Only in early ontogenesis, when the active formation of a neuronal network is in progress, there is an addition of neurons to the network, formation of neuronal connections, and even their restructuring by breaking existing connections [4].

Functional plasticity appears in the mature neural network, associated with increased reactivity of neurons exposed to stimulation and functional change of connections [5]. Prolonged exposure of neurons during the learning process leads to changes in their metabolic activity and synaptic membrane permeability, actually changing the strength of connections.

So, learning of an already formed natural neural network does not lead to the addition of new neurons. There is not a quantitative, but a qualitative transformation of a neural network because of changes in connections between neurons [6, 7].

According to neurophysiological observations of D. Hebb, if neurons on both sides of a synapse are activated simultaneously and regularly, the strength of synaptic communication increases [8]. An important feature of this rule is that the change in synaptic weight depends only on the activity of the neurons connected by the synapse.

Connections between neurons do not exist in isolation. O.V. Kuznetsov noted holographic properties of neural networks in the example of the human brain: distribution of information in the network and its preservation [17]. G. Shepherd noted that natural neural networks show stability through the replacement of damaged nerve pathways by competing for the possession of synaptic regions [18]. The systemic properties of the entire neural network compensate for the poor reliability of individual connections and neurons, ensuring stability.

The properties of plasticity and stability of neural networks lead to the analogy with holographic images. Works [9-14] put forward a hypothesis about the similarity of information processes in optical holography with information processes in the brain.

Nowadays, physics deals not only with optical holograms. The common feature of holograms is their wave nature. Any hologram is a wave pattern recorded in some medium, corresponding to the interference of coherent waves having a common source. Modern studies admit the existence of more complex holograms formed by sources of non-monochromatic and incoherent waves.

Works [15, 16] show that we can achieve switching between spontaneous activity patterns on neuronal cultures, showing long-term stability of activity under external electrical stimulation. Without changing the structure of the network, we can achieve a new behavior with new knowledge without traditional training procedures. You can transfer new information into the neural network, despite its holographic stability when it is externally stimulated, to achieve a new behavior with new knowledge without traditional training procedures.

The problem is to add a new class to the network with minimal training, so as not to train all classes again. The optimal solution is to create and train a new neural network with one class, and then combine it with the existing neural network.

## 2 Materials and methods

It follows from the holographic nature of neural networks: an effect on only one connection will lead to a change in the properties of the entire network to some extent. As in holograms, only significant interventions in connections will cause qualitative changes.

We can consider a neural network (its weights and structure) as a hologram, having made some simplifications. We assume coherent waves form a hologram, and the thickness of the hologram registration medium will be negligibly small. Then the input signal of the neural network passing through it will generate an image of the class corresponding to it, like a wave restoring an image from the holographic plate.

Having two neural networks  $N_1$  and  $N_2$  with the same structure, differing only by weight coefficients of connections and trained to recognize different classes, we will consider them as holograms.

The first neural network is trained on examples of the set  $P_1$ , partitioned into a set of classes  $C_1$ , the second one is trained on  $P_2$ , partitioned into classes  $C_2$ . In the original problem statement, the power  $m(C_2)=1$ .

If you combine two holograms in space and apply different signals to them, the classes corresponding to the signals will be reconstructed. The signals  $p_1 \in P_1$  on hologram  $N_1$  will recreate the corresponding  $c_1 \subset C_1$ ,  $p_2 \in P_2$  on  $N_2$  will recreate  $c_2 \subset C_2$ . Since the signals will hit both holograms, they will distort the result: either amplifying it or reducing it, i.e. intervening connections of the two neural networks.

Since the conditions of hologram formation and its orientation in space are unknown, we can only suppose that refine transformations should apply to one hologram for better alignment. Let us choose one network as the base one -  $N_1$ , the second one as a changing one -  $N_2$ .

Let us assume that space is Euclidean and the rotation in space does not exceed  $\pi$ . Thus the hologram will change: the phase difference fixed on the plate will change by  $\Delta\varphi$  at each point of the plate (in general each class will have its  $\Delta\varphi$ , but since  $m(C_2)=1$ , we can assume that  $\Delta\varphi$  for  $N_2$  is one).

To perform the hologram addition, it is necessary to change from the amplitude-frequency representation to the frequency-phase representation. We use the fast Fourier transform for the discrete signal, which allows us to avoid increasing the complexity of the algorithm to  $O(N^2)$  and keep it at  $O(N \log N)$ .

For the  $N_1$  neural network with weight coefficients  $w_{ij}^1 (i = 1..m, j = 1..n)$  we obtain the weight coefficients:

$$w_{ij}^{1'} = F(w_{11}^1, \dots, w_{ij}^1, \dots, w_{mn}^1) (i = 1..m, j = 1..n) \quad (1)$$

where  $F$  is the fast Fourier transform.

Neural network  $N_2$  after the transformation will have weighting coefficients:

$$w_{ij}^{2'} = F(w_{11}^2, \dots, w_{ij}^2, \dots, w_{mn}^2) + \Delta\varphi, (i = 1..m, j = 1..n) \quad (2)$$

By adding the transformed weights of neural networks  $N_1$  and  $N_2$ , we get the weights of the new neural network  $N_3$ :

$$w_{ij}^{3'} = w_{ij}^{1'} + w_{ij}^{2'}, (i = 1..m, j = 1..n) \quad (3)$$

The weight coefficients of  $w_{ij}^{3'}$  in the inverse Fourier transform are complex numbers, so we can apply the Hartley transform to get the real coefficients of the  $N_3$  neural network

$$w_{ij}^3 = H(w_{ij}^{3'}) \tag{4}$$

Then function  $G_1 = G_1(w_{11}^3, \dots, w_{ij}^3, \dots, w_{mn}^3)$  – is the percentage of correctly distributed examples  $p_k \in P_1, c_s \subset C_1, p_k \in c_s$  neural network  $N_3$ . Similarly, we define  $G_2 = G_2(w_{11}^3, \dots, w_{ij}^3, \dots, w_{mn}^3)$  – the percentage of correctly distributed examples  $p_t \in P_2, c_u \subset C_2, p_t \in c_u$  neural network  $N_3$ .

We can represent the target function for the new neural network as

$$G_3 = G_3(w_{11}^3, \dots, w_{ij}^3, \dots, w_{mn}^3) \rightarrow \max, |G_1 - G_2| \rightarrow 0 \tag{5}$$

Thus, we achieve not only the maximum  $G_3$ , but also guarantee the balance of the neural network in the recognition of classes from the sets  $C_1$  and  $C_2$ . Depending on the problem, we can redefine the target function by relaxing the condition  $|G_1 - G_2| \rightarrow 0$ . Maximizing  $G_3$  without this condition does not guarantee that  $N_3$  will recognize classes  $C_1 \cup C_2$  equally well.

It is possible, for example, to require that the network keeps the ability to recognize  $C_1$  classes in full and in part, but with a high percentage of new classes. Or on the contrary, we can simulate the process of "forgetting" by reducing the requirements for  $G_1$  and requiring a maximum for  $G_2$ . Thus, not only the maximum  $G_3$  is achieved, but also the balance of the neural network in the recognition of classes from the sets  $C_1$  and  $C_2$  is guaranteed. Of course, depending on the problem, it is possible to redefine the target function by relaxing the condition  $|G_1 - G_2| \rightarrow 0$ . Maximizing  $G_3$  without this condition does not guarantee that  $N_3$  will recognize classes  $C_1 \cup C_2$  equally well.

Networks are trained to recognize different classes have different weights with the same structure. Networks trained on only a class have weakly expressed connections, approximately from the same range. Neural networks trained on more classes have more pronounced connections than those trained on one class, Table 1.

**Table 1.** Average weights of networks trained in grades 1-4 of handwritten digits.

Number of classes	1	2	3	4
Minimum value	-0,04255	-3,25154	-5,54	-72,71
Maximum value	0,046264	2,893664	4,94	71,52
Range	0,088815	6,145204	10,48	144,23
Arithmetic mean.	0,002732	-0,00956	-0,06	0,02
Mean square deviation	0,021148	0,350034	0,64	6,20

To stress the force of the neural network's connections, the forces  $N_2$  in its connections can be scaled by the connections  $N_1$ , replacing (2) with

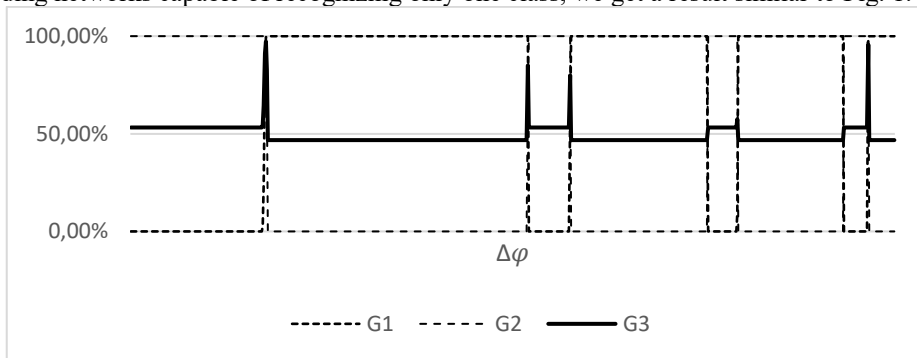
$$w_{ij}^{2'} = F(kw_{11}^2, \dots, kw_{ij}^2, \dots, kw_{mn}^2) + \Delta\varphi, (i = 1..m, j = 1..n) \tag{6}$$

We experimentally tested the method on a set of handwritten digits from the MNIST database. Neural networks were trained on a different number of classes of handwritten digits, and then a new digit class was added to them. Both neural networks  $N_1$  and  $N_2$  were created with the same structures, and the number of outputs in both networks was set equal to  $m(C_1 \cup C_2)$ . Then  $N_1$  was trained to recognize  $C_1$  classes, and  $N_2$  a new class, after which we applied a holographic method to the networks.

### 3 Results and discussion

The proposed method depends on two parameters: the scaling factor  $k$  and the rotation angle  $\Delta\varphi$ .

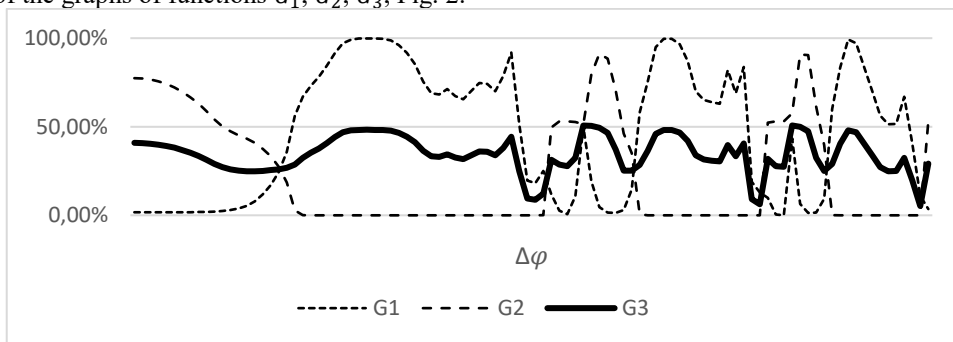
Different values of  $G_1$  and  $G_2$  can be achieved with different  $\Delta\varphi$ , for example, when adding networks capable of recognizing only one class, we get a result similar to Fig. 1.



**Fig. 1.** Graphs of the function  $G_1, G_2, G_3$ :  $G_1$  - recognize digit 1,  $G_2$  - recognize digit 2,  $G_3$  - recognize digits 1 and 2.

The graph shows that for the case  $m(C_1)=1$ , there is a  $\Delta\varphi$  at which  $G_3$  reaches a maximum close to 100%. The experiment showed that if the original networks can recognize their classes with an accuracy close to 100%, the resulting network also shows an accuracy close to 100%.

The more classes the neural networks can recognize, the more complicated for analysis is the graphs of functions  $G_1, G_2, G_3$ , Fig. 2.



**Fig. 2.** Function graphs  $G_1, G_2, G_3$ :  $G_1$  - handwritten digit recognition 0 and 1,  $G_2$  - handwritten digit recognition 2,  $G_3$  - handwritten digit recognition 0, 1, and 2.

We note that the functions  $G_1, G_2, G_3$  have a fractal form because of the wave nature of the model (Figures 1 and 2). The optimal value of  $G_3$  in fragments may vary, but there is a limiting value of  $G_3$  under optimality conditions (5).

Therefore, we can search for  $\Delta\varphi$  within a single fragment of self-similar parts of the graph. The difficulty of the search is related to the fractality of the target function. If you use a brute-force algorithm, choose the step change of  $\Delta\varphi$  as small as possible. This does not guarantee that even with a small step of  $\Delta\varphi$ , you will hit the local maximum  $G_3$  on the fragment.

To get the best result, we can scale connections of the neural network  $N_2$  by connections of the neural network  $N_1$  with the coefficient  $k$ , chosen empirically.

Because of the holographic nature of the neural network, scaling in some limits does not affect the ability of the neural network to recognize classes known to it. If all weight coefficients are scaled simultaneously with some coefficient  $k$ , it will change the slope of the neuron sigmoid. It will not significantly affect the network with small values of weights, but when weights are added to weights of the network with more significant coefficients, it will help to strengthen the effect on the result. The scaling factor has no significant effect on the optimal value in the sense of (5), only changing the conditions for finding this value for the set  $\Delta\varphi_i$  ( $i \in N$ ).

The choice of the basic and changing neural network is important, if  $m(C_1) = m(C_2) = 1$ , then we choose the basic function experimentally. An incorrectly chosen basis function gives  $G_1 = const = 0$  or  $G_2 = const = 0$ . In this case, we should swap the basic and modifying neural networks.

Searching for the optimum depending on the problem (the number of classes  $N_1$  and their peculiarities). The method demonstrates the ability of the neural network to store the old knowledge and accept the new at the same time.

We can apply this method to speed up the learning process by training a neural network quickly on a single class, and then using it to change an existing network trained on multiple classes. The resulting network can be pre-trained, which is faster than creating and training a new network.

The question of the fractal nature of the target function and determination of conditions for finding its maximum by a minimum enumeration of  $\Delta\varphi$  values, the possibility of introducing additional parameters into the model that increases the efficiency of the method requires further research.

## 4 Conclusion

The holographic method, unlike other constructive methods, does not require complete retraining of the network in conditions of new classes. It is enough to create a network with the same structure as the old network, train the new network with one class and stack it with the existing network. It is much faster to train a network in a single class than to train the entire network in all classes (old and new in aggregate).

By relaxing the optimality requirements (5), we can vary the gained knowledge of the neural network, emphasizing more plasticity or stability. As in natural neural networks, acquiring new experience influences existing knowledge.

We can use the method of learning. We can retrain the constructed network to improve accuracy. The time to construct and retrain in aggregate would be considerably less than the time to create and train a new network based on all classes.

The perspective of the holographic method idea is the possibility of transferring knowledge from one neural network to another, bypassing the long training procedures.

## References

1. S. E. Fahlman, C. Lebiere, *The Cascade-Correlation Learning Architecture* (School of Computer Science Carnegie Mellon University, Pittsburgh, 1991)
2. L. Prechelt, *Neural Networks* **10(5)**, 885-896 (1997)
3. J. Hertz, A. Krogh, R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Reading, Mass, 1991).
4. E. A. Agrba, *Selective properties of neural networks of primary hippocampal cultures under low-frequency electrical stimulation*, Dissertation for the degree of Candidate of

- Biological Sciences. Nizhny Novgorod (2018), <https://diss.unn.ru/files/2018/842/diss-Agrba-842.pdf>
5. D. A. Farber, N.V. Dubrovinskaya, *Human physiology* **17(5)**, (1991)
  6. V. Menon, *Trends in Cognitive Sciences* **17(12)**, 627-640
  7. L. V. Kolobashkina, *Biologically Inspired Algorithm for Increasing the Number of Artificial Neurons*, 225–231 (BICA, 2019)
  8. D. O. Hebb, *The Organization of Behavior* (John Wiley & Sons, 1949)
  9. K. Pribram, *Scientific American* **220**, 75 (1969)
  10. K. Pribram, *Languages of the Brain* (Wads worth Publishing, 1977)
  11. D. Goleman, *Psychology Today* **12(9)**, 72 (1979)
  12. J. Collier, C. B. Burckhardt, L. H. Lin, *Optical Holography*, (Academic Press, 1971)
  13. P. V. Heerden, *The Foundation of Empirical Knowledge* (N.V. Uitgeverij Wistik-Wassenaar, Netherland, 1968)
  14. D. Gabor, *IBM J. Research and Development* **13(2)**, 156-159 (1969)
  15. R. E. Hampson, D. Song, R. H. Chan, A. J. Sweatt, M. R. Riley, G. A. Gerhardt et al., *IEEE. Trans Neural Syst Rehabil Eng*, **20(2)**, 184-97 (2012)
  16. I. S. Sokolov, M. K. Tatarintsev, R. Y. Khasanov, A. M. Azieva, E. Y. Makarenko, and M. S. Burtsev, *Bulletin of the Russian State Medical University* **2**, 45-49 (2016)
  17. O. P. Kuznetsov, *Theory and Control Systems* **5**, 3-23 (1995)
  18. G. Shepherd, *Neurobiology* (Mir, Moscow, 1987)