

Methods of the development of the architecture of the neural networks for identification and authentication of individuals

O. Y. Golikov* and M. A. Ramos

Laboratory of Low Temperatures, Department of Condensed Matter Physics,
Condensed Matter Physics Center (IFIMAC) and Nicolás Cabrera Institute,
Autonomous University of Madrid, 7, Francisco Tomás y Valiente, 28049 Madrid, Spain

*e-mail: golikov.ua@gmail.com

This paper deals with the neural network methods of the implementation of systems of identification of individuals based on videos and photographs. Over the last few decades, it has been considered to be one of the most powerful tools and has become very popular in the literature as it is able to handle a huge amount of data. The neural network architectures used in modern biometric identification systems have been reviewed. Based on the research conducted in this field, an approach was developed that can improve the accuracy of object recognition in photo and video images by increasing the quality of the attributes of the weights and reducing the number of the weights, as well as the number of the connections. The basis of the developed neural network model is a multilayer perceptron; the main system is a convolutional neural network. The neural network model has been implemented using the Python programming language with the most popular machine learning libraries Keras and TensorFlow. In addition, we will also enumerate the parameters that affect CNN efficiency.

Key words: biometric identification of individuals, three-dimensional object recognition, convolutional neural network, deep machine learning, subsampling.

PACS numbers: 84.35.+i, 87.18.Sn, 07.05.Mh

1 Introduction

Since the 1990s, systems based on machine learning methods and annotated database information have been widely used for automatic information recognition [1, 2, 3, 4]. By today, the field of “character and pattern recognition” has become very popular among researchers due to the possibility of increasingly frequent application of the technology in daily human activities. The most prominent examples of the use of information recognition technology are the recognition of faces, speech, images, handwritten documents, number plates, barcodes. The recognition of faces, speech and handwritten input is of particular interest among these examples as these parameters can solve one of the most topical problems of the modern digital society: identification and authentication of individuals [2, 3, 5]. Through the application of automatic recognition, it is possible to achieve digital “computer-based” identification of a person, which, in turn, would contribute to the replacement of the current, already obsolete methods.

The development of high-speed computing machinery, the increased performance of modern computers and improvements in audio and video capture systems indicate that further advancements in machine learning and autonomous object recognition technology are appropriate and that such technology should be introduced even more widely into everyday life.

Researchers have written a large number of publications on the problem of object recognition and the identification of individuals, but despite this, this topic is still relevant. This is primarily due to the large volume of information and the uncertainty that are inherent to the recognition technology and the objects, respectively, as well as to the improvement of computing technology capabilities.

Most modern biometric recognition systems based on the unique biological characteristics of an individual have two major drawbacks, which make the use of neural networks and video even more sensible [6, 7]:

– the use of special and/or expensive equipment;

– physical contact with the devices used for identification (fingerprinting, iris scanning).

Thus, the use of face recognition methods based on the neural networks for identity verification helps to create a fairly efficient, yet inexpensive and easy to implement solution to this problem [8].

A priority direction in the field of image detection and recognition in photo and video today is the use of Convolutional Neural Networks (CNN) [9, 10, 11, 12]. The main advantage of this method of neural network implementation is that the algorithm itself extracts information attributes, relative to which the weights are calculated [12].

The aim of this research is to investigate modern identification methods and to improve the quality of image recognition. The scientific value lies in the creation of a more efficient image analyser based on the already existing neural network organisation approaches, without using “heavy” methods. Such methods involve software tools that spend too many resources in the process of recognition (neural networks with pixel-by-pixel comparison) [13]. The practical value lies in the possibility of applying this method of neural network organisation as a working system for the identification of individuals (e.g., in the banking, government, law enforcement sectors).

2 Neural Network Model Architecture and Its Training.

Convolutional neural networks are a class of neural network models that originated from the studies of the feline visual apparatus conducted by Hubel and Wiesel in the 1960s [14]. The result of these studies was the discovery of two types of

neurons responsible for the visual perception of animals: the first type has the properties of local sensitivity and is responsible for detecting the simplest characteristics of objects (approximate silhouette, extreme points and angles of an object); the second type facilitates the recognition of higher-level features of objects, using combinations of neurons of the first type.

The first real-world model that implemented the detected behaviour of feline visual neurons was the neocognitron of Fukushima [15]. It was created using supervised training of a linear classifier and unsupervised adjustment of the filter bank. Subsequent research in this area made the complete transition to the supervised form of training possible and also greatly simplified the architecture of the neural network. Eventually, a new convolutional structure was developed, which is currently gaining more and more forms of practical application: navigation systems, image recognition, identification of individuals, recovery of noisy signals and much more [16].

There are several architectures of convolutional neural networks developed so far, which differ from each other in the topology of layers, the way the learning process is organised, activation function, etc. One of the classic architectures of convolutional neural networks is the LeNet-5 neural network model developed by Yan LeCun in the late 1990s [17].

Fig. 1 shows a schematic diagram of a classical, multilayer, fully connected neural network and how it works. Classic multilayer neural networks have two significant disadvantages that arise directly from the network architecture because the inter-neuronal connections between layers are organized by means of “each to each” connection.

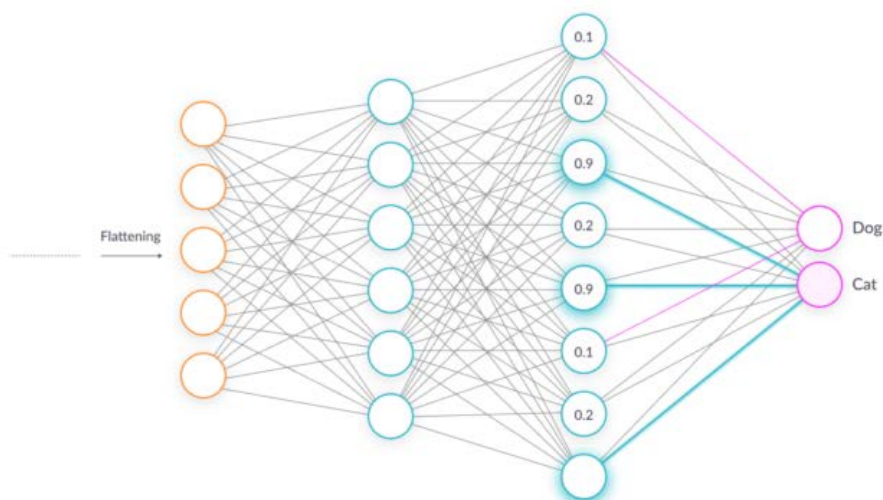


Figure 1 – Fully-connected layer

Because of this kind of connection, the image has the appearance of an n -dimensional vector, which does not take into account the local two-dimensional organisation of the pixels and the possibility of image deformation. At the same time, the convolutional neural network architecture (Fig.

2) provides an opportunity to correct the shortcomings of the classical multilayer neural network. It implements the principles of the neocognitron architecture, which is simplified and augmented with an error backpropagation algorithm used for training [18, 19].

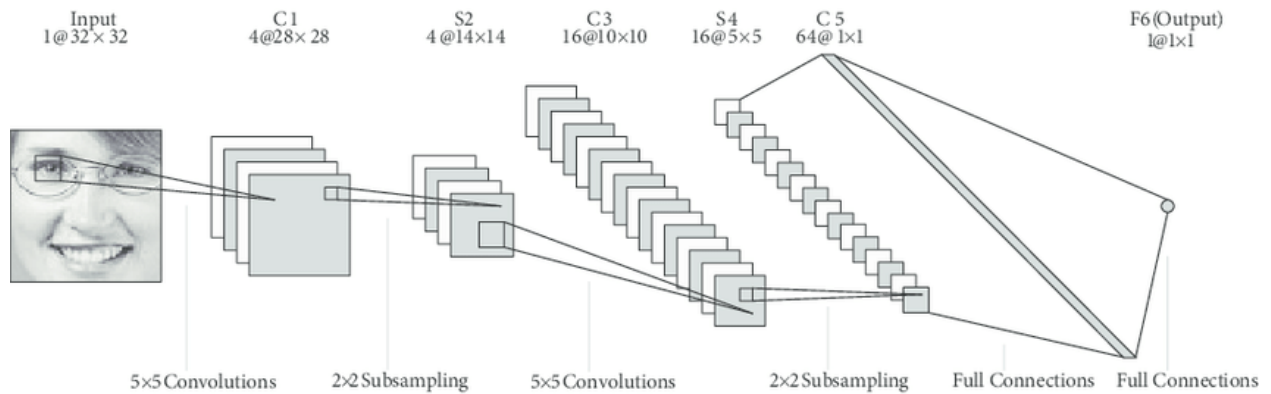


Figure 2 – LeNet introduced by Yan LeCun [17]

The use of local receptor fields, weights and hierarchical organisation with spatial subsamples in convolutional neural networks enables local two-dimensional connectivity of the neurons and detection of individual facial features that can be located anywhere in an image [20].

A convolutional neural network is able to provide partial robustness to changes in image scale, displacements, rotations, changes in viewing angle and other types of distortions. As already mentioned, a convolutional neural network architecture is multi-layered, but it can be divided into two main types: Convolutional and Subsampling, which alternate with each other in architectures. Each individual layer contains multiple planes. Neurons from any of the planes have identical weights that come to all the local fragments of the previous layer (identical to the human visual cortex). The incoming information is “scanned” by a window of a certain, often small, size; information is then “weighed” by a set of weights, and the result of these operations is transmitted to the corresponding neuron of the subsequent layer. Such planes are called feature maps. Each of the feature maps selects certain parts of the input image, sending the result to the next subsampling layer. The subsampling layer reduces the scale of the planes through the local averaging of the weights, thereby organising the hierarchical structure of the convolutional neural network. This algorithm corresponds to the first type of neurons in the feline

visual apparatus mentioned above. Subsequent layers corresponding to the second type of neuron extract more general characteristics from images, which are less dependent on distortion [18].

A comparison of multilayer neural networks with convolutional neural networks has shown a significant advantage of the latter both in terms of learning speed and quality of object recognition. At present, the parameters of learning speed, performance and percentage of accurate recognition play a key role in the development of the image recognition systems (including human identification systems).

3 Algorithm for Selecting a Convolutional Neural Network Architecture.

Before starting to implement the algorithm, it is necessary to introduce the concept of partial partitioning, which means the set of convolutional neural network layers $N = \{N_1, N_2, \dots, N_i\}$, where i is the number of neural network layers determined experimentally. The idea of introducing the following algorithm is to enumerate a combinatorial set of different configurations that are constrained by certain parameters and layout rules. In order to reduce the enumeration space and cut off the futile alternatives, we will rely on the following conclusion:

If from the architecture of the convolutional neural network $N^1 = N_1^1, N_2^1, \dots, N_i^1$ the recognition

accuracy P^1 was obtained after its training, then architecture $N^2 = N_1^2, N_2^2, \dots, N_i^2$ can be obtained by changing parameter a_j in one of the layers $N_x^1 = k_x^1 a_j$ to parameter a_{j+1} $N_x^2 = k_x^2 a_{j+1}$, where $x \in [1, \dots, i]$, k_x is the number of elements in the word with recognition accuracy $P^2 < P^1$, then for neural network architecture $N^3 = N_1^3, N_2^3, \dots, N_i^3$ $N_x^3 = (k_x^2 + 1) a_{j+1}$, the recognition accuracy will be even lower $P^3 < P^2 < P^1$ and further training of this branch is not needed.

The layout algorithm of the convolutional neural network architecture can be described as follows:

1. The input and the output layers of the desired architecture are created based on the data used as a basis:

$$- x_{i,j}^1 = m * n * ch;$$

$$- y_{i,j}^1 = r.$$

2. A randomly chosen number of layers i is set, and the layout counter is set to $j = 1$. A set A of convolutional neural networks is found.

3. $Q^* = 0$, $M^* = 0$, $N, N_1, \dots, Z_1 = \emptyset, j = j + 1$, $A = \emptyset$ is a set of futile architectures.

4. $n = 1$ is the layer counter.

5. $k = 1$ is the counter of neurons in the layer.

6. Check if the condition $n < i$ (the current layer is not the last one) is met. If the condition is true, proceed to the next step, otherwise, proceed to step 13.

7. Assign γ^5 as a rule of selecting the candidate a_i from the set of all neural network architectures A .

8. According to the rule γ^5 , element a_k is selected and placed in N_1 ,

9. The number of inter-neuronal connections is found:

$$Q_k^* = \sum W_{a_k} x_{i,j}^i, Q^* = Q^* + Q_k^*,$$

and the number of weights used:

$$M_k^* = W_{a_k} x_{i,j}^i, M^* = M^* + M_k^*.$$

10. Check if the condition $Q^* \leq \frac{Q}{i}$ is met. If the condition is true, increment $k = k + 1$ and return to

step 6. If the condition is false, increment $n = n + 1$ and return to step 5.

11. Insert i fully-connected output neurons in the subset $N_1, N = \{N_1, \dots, N_i\}$.

12. Check if the cut-off condition is met by the model. If the condition is true, then $N = \{N_1, \dots, N_i\}$ is placed in A . If the condition is false, proceed to the next step.

13. Model $N = \{N_1, \dots, N_i\}$ is excluded from the set of all convolutional neural networks A . After this model is trained on the test dataset, it is added to the database along with the result of the training P, P_1, P_2 under the name A^j .

14. Untested alternatives of the set of neural networks A are checked. If untested architectures exist, return to step 3, otherwise, proceed to the next step.

15. Selection of the architecture with the maximum $P - max$ value from the database A^j .

4 Results of the Experiment.

The training data set for the experiment consisted of photographs of real people. The architecture of the convolutional neural network was obtained by using the algorithm proposed above. The sizes of the filter windows were 3x3, 7x7, 11x11; the subsampling window size was 2x2; the activation function was adam (adaptive moment estimation) [21].

The important point was the comparison of the parameters such as epoch accuracy and epoch loss during the training of the neural networks. Figure 3 and Figure 4 show the comparison of these parameters for neural networks with different filter window and different volume of the training base.

The graphs show that with a 3x3 filter window size, high accuracy is achieved faster than with the 7x7 and 11x11 windows. At the same time, the training of the neural network with a 3x3 window is faster. A similar pattern can be observed in the case of epoch loss.

It is worth noting that after certain epochs of training in some cases epoch loss increases along with a decrease in epoch accuracy. Thus, there is a question of whether it is reasonable to use a large number of epochs, which can lead to the deterioration of the neural network parameters.

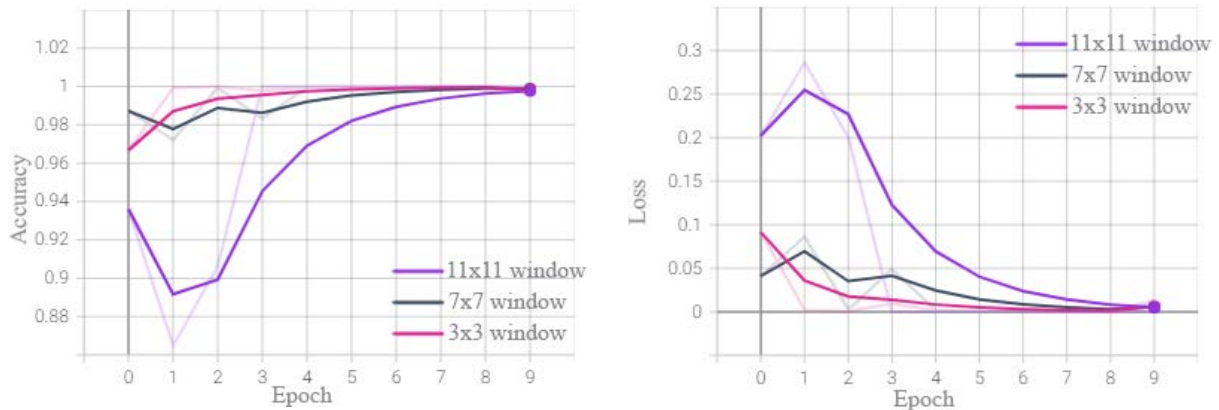


Figure 3 – The dependency of epoch accuracy and epoch loss on the size of the filter window when training data equals 20000

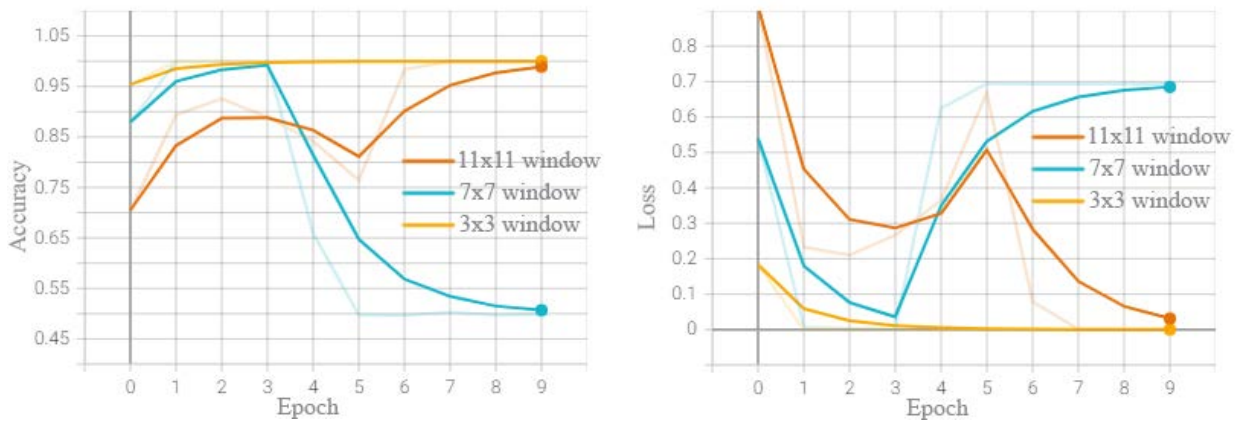


Figure 4 – The dependency of epoch accuracy and epoch loss on the size of the filter window when training data equals 4000

5 Conclusion

In this paper, methods of selection of the appropriate neural network architecture were reviewed, using the proposed algorithm. By varying the size of the filter window, the dependencies that affect the process of training, as well as the quality of the initial neural network, were obtained. The obtained data provides food for thought on how to properly constrain the process of training of a neural network in order to achieve an optimal solution with respect to the duration of the training and the quality of recognition.

The deterioration of the parameters after a certain epoch of training requires further explanation as well. It is possible that this phenomenon is caused by the activation function reaching one of the local minima.

Acknowledgements

These studies have been carried out with the financial support of the Ministry of Education and Science of the Republic of Kazakhstan under grant AP08855681.

References

- 1 Burshtein D., Robust parametric modeling of durations in hidden Markov models // IEEE Transactions on Speech and Audio Processing, May 1996. – 1996. – Vol. 1. – No. 3. – P. 240-242.
- 2 Gunawardana A., Mahajan M., Acero A., Platt J. C. Hidden conditional random fields for phone classification // In Proceedings Interspeech. – 2005. – P. 1117–1120.
- 3 Bourlard H., Wellekens C.J. Links between Markov models and multilayer perceptrons // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1990. – Vol. 12. – No. 12. – P. 1167-1178.
- 4 Nefian A.V., Liang L.H., Pi X., Xiaoxiang X., Mao C., Murphy K. A coupled HMM for audio-visual speech recognition // In Proceedings International Conference ICASSP-2002, Orlando, USA. – 2002. – P. 2013–2016.
- 5 Jain A.K., Bolle R., Pankanti S., eds., Biometrics: Personal Identification in networked society, Kluwer Academic, Norwell, Mass., 1999, 410 p.
- 6 Pankanti Sh., Bolle R.M. Antil Jain. Biometrics: The future of identification // IEEE Computer. – 2000. – P. 46-49.
- 7 Foltyniewicz R. Efficient high order neural network for rotation, translation and distance invariant recognition of gray scale images // Lecture Notes in Computer Science - Computer Analysis of Images and Patterns. – 1995. – P. 424-431.
- 8 Turner R. E. Lecture 14: Convolutional neural networks for computer vision. – 2014. – P.1-30.
- 9 Dumoulin V., Visin F. A guide to convolution arithmetic for deep learning. – 2016. – P. 1–28.
- 10 Kokkinos I., Paris E. C., Group G. Introduction to deep learning convolutional networks. Dropout, Maxout 1. – P. 1–70.
- 11 Andrew Lavin, Scott Gray; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2016. – P. 4013-4021.
- 12 Sarıgül M., Ozyildirim B.M., Avcı M., Differential convolutional neural network // Neural Networks. – 2019. – Vol. 116. – P. 279-287.
- 13 Salu Y., Tilton J. Classification of multispectral image data by the binary diamond neural network and by nonparametric, pixel-by-pixel methods // IEEE Transactions on Geoscience and Remote Sensing. – 1993. – Vol. 31. – No. 3. – P. 606-617.
- 14 Hubel D.H., Wiesel T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex // Journal of Physiology London. – 1962. – Vol. 15. – P 106–154.
- 15 Fukushima K., Miyake S., Ito T. Neocognitron: A neural network model for a mechanism of visual pattern recognition // IEEE Transactions on Systems, Man, and Cybernetics, – 1983. – Vol. SMC-13. – No. 5. – P. 826-834.
- 16 LeCun Y., Kavukcuoglu K., Farabet C. Convolutional networks and applications in vision // Proc. IEEE Int. Symposium on Circuits and Systems. – 2010. – P 253–256.
- 17 LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. Gradientbased learning applied to document recognition // Proceedings of the IEEE -1990. –Vol. 86(11). – P.2278-2324.
- 18 Lawrence S., Giles C. L., Tsoi A. C., Back A. D. Face recognition: A convolutional neural network approach // IEEE Transact. on Neural Networks, Special Issue on Neural Networks and Pattern Recognition. – 1997. – P. 1-24.
- 19 Santaji G., Jayshree G., Shamla M., Dhanaji G. Neural networks for facerecognition using SOM // IJCST. -2010. – Vol. 1, – No. 2. – P. 65—67.
- 20 Albawi S., Mohammed T. A., Al-Zawi S. Understanding of a convolutional neural network // 2017 International Conference on Engineering and Technology (ICET). – P. 1-6
- 21 Kingma D.P., Ba J.L., Adam: A method for stochastic optimization // ICLR. – 2015. – P 58–62.