

Voloshyn M.

DEVELOPMENT OF A COMBINED IMAGE RECOGNITION MODEL

Об'єктом дослідження є процеси ідентифікації та класифікації об'єктів у задачах комп'ютерного зору. В даний час для розпізнавання зображень найкращі результати демонструють штучні нейронні мережі. Однак, навчання нейронних мереж – це погано обумовлена задача. Погана обумовленість означає, що навіть великий набір даних може нести в собі малий обсяг інформації про задачу, що вирішується. Тому ключова роль в процесі синтезу параметрів конкретної математичної моделі нейронної мережі належить навчальним даним. Підбір репрезентативної навчальної вибірки є однією з найскладніших задач в машинному навчанні і не завжди можливий на практиці.

Запропонована у роботі нова комбінована модель розпізнавання зображень з використанням теорії несилової взаємодії володіє такими ключовими особливостями:

- призначена для обробки великої кількості даних;
- виділяє корисну інформацію з довільного потоку;
- дозволяє природно додавати нові об'єкти;
- терпима до помилок та дозволяє швидко перепрограмувати поведінку системи.

В порівнянні з існуючими аналогами, достовірність розпізнавання запропонованою моделлю в усіх експериментальних дослідженнях була вище відомих методів розпізнавання. Середня достовірність розпізнавання запропонованою моделлю склала 71,3 %; методом локальних бінарних шаблонів – 59,9 %; методом аналізу головних компонент – 65,2 %; методом лінійного дискримінантного аналізу – 65,6 %. Така достовірність розпізнавання в поєднанні з обчислювальною складністю робить даний метод прийнятним для застосування в системах, які працюють в умовах наближених до реального часу. Також такий підхід дозволяє керувати достовірністю розпізнавання. Це досягається завдяки налаштуванню кількості секторів гістограм локальних бінарних шаблонів, які використовуються при описі зображень, та кількості фрагментів зображення, які використовуються на етапі класифікації інтроформаційним підходом. Значною мірою на час класифікації впливає кількість фрагментів зображення, оскільки, в такому випадку, потрібно вираховувати попарно співставлення суміщення дій системи в кожному з можливих напрямків.

Ключові слова: системи комп'ютерного зору, аналіз зображень, розпізнавання та ідентифікація об'єктів.

Received date: 29.11.2018

Accepted date: 17.12.2018

Published date: 30.06.2019

Copyright © 2019, Voloshyn M.

This is an open access article under the CC BY license
(<http://creativecommons.org/licenses/by/4.0>)

1. Introduction

Every year the volumes of information grow from the formalization and subsequent algorithmic processes that were previously performed manually. 80 % of the information a person receives through vision, so any systems associated with automatic image processing are in demand. One of the key concepts in automatic processing is the concept of object recognition, which is an area of active work over the past twenty years. When algorithms perform recognition at the expert-person level, automation leads to acceleration of data processing systems and increasing their efficiency.

Despite significant success, image recognition only in some areas can match or exceed the cognitive function of human perception in terms of the quality of the result. In general, the problem of image recognition is still not completely solved.

There are various methods for image recognition potential functions, Bayesian networks, Markov networks, artificial neural networks, various types of associative memory, and so on. The study of the problem of image recognition has shown that recognition is carried out by methods that do not fully take into account the features of graphic objects, the main ones of which are a small

amount of a priori data regarding reference descriptions of recognition objects. Therefore, it is important to develop a universal model, which makes it possible to assign the object of recognition to a certain class of objects with a small training set with a high probability.

2. The object of research and its technological audit

The object of research is the processes of identification and classification of objects in computer vision tasks. Currently, for image recognition, stochastic models show the best results, namely: a subclass of artificial neural networks – convolutional neural networks. However, learning neural networks is a poorly conditioned task. Poor conditioning means that even a large data set can carry a small amount of information about a problem that is being solved. Therefore, a key role in the synthesis of parameters of a specific mathematical model of a neural network belongs to educational data. Selection of a representative training set is one of the most difficult tasks in machine learning and is not always possible in practice. So there is a need to develop a universal method, which allows with a high degree of probability to assign the object of recognition to a certain class of objects with a small training set.

3. The aim and objectives of research

The aim of research is increasing the efficiency of image processing in solving technical problems through the use of new approaches. To achieve this aim it is necessary to solve the following objectives:

1. Analyze the developed methods for computational complexity in solving the practical problem of image processing.
2. Develop a combined method for solving the problem of image processing.
3. Justify the need and advantages of using the new combined image processing method.

4. Research of existing solutions of the problem

In the past few years, significant progress has been made in the field of object recognition with small variations in lighting, stage, etc. [1], but reliable recognition methods with more extreme changes were unattainable.

The task of object recognition with the help of classifiers is traditionally divided into two parts: the selection of key features and the classification by these features [2]. Feature extraction is carried out due to a priori information, as a result of which structural invariance is achieved [3].

Statistical methods [4] are used for image analysis, but in cases where high recognition accuracy is not required. Not only image pixels, but also some low-level image representation, for example, wavelet coefficients, are input to the system. The support vector machine is a universal approximator that can reduce to zero the training error on a training set, which, in turn, allows to rely on a low generalization error on the test set. However, practice has shown that the error of generalization for networks with a large number of layers (deep architecture) is lower than for networks with a small number of layers (shallow architecture) [5]. Classical perceptrons, RBF networks and SVM, have a small number of layers, and the task of training the network when expanding the number of layers becomes quite complex.

In most cases, projection directions that are orthogonal over the entire range of their class are chosen for research. The principal component analysis method (PCA) is based on linear projection of image space in low-dimensional feature space. The basis of this algorithm is the use of fundamental statistical characteristics such as: average expectation and covariance matrix. For each object, its main components are calculated. The recognition process consists in comparing the principal components of an unknown object with the components of the remaining objects. However, this approach maximizes the overall range across all classes. PCA projection is well suited for reconstruction from a low-space basis, but can't be optimal from a discriminant point of view. The main disadvantage: high demands on the conditions for shooting images. Images should be obtained in similar lighting conditions, with the same angle and high-quality pre-processing should be carried out, leading the image to standard conditions (scale, rotation, centering, leveling brightness, background clipping). It is undesirable to have various kinds of distortions and other intra-class variations.

Linear discriminant is a «classical» method for pattern recognition, in which the characteristics of an object for recognition are obtained using polar quantization of the form.

Neural network structures give rather good results due to their non-linear structure. The disadvantages of networks include the rapid growth of their size when trying to tune to a large number of invariant transformations, the use of neurons with a threshold activation function, and special algorithms for setting [6].

Convolutional neural networks (CNS – Convolutional Neural Network) belong to multilayer neural networks, which are a modernized multilayer perceptron [7]. SNA consists of two parts: one part is responsible for the selection of signs, and the other is responsible for the classification. To highlight the signs using the alternation of layers of convolution (C-layer) and subsampling (S-layer). This is one of the variants of alternating layers of complex and simple cells, which appeared in the neocognitron. Neurons in these layers are organized into plates. Each neuron perceives information from the previous layer using its own receptive field, which is a group of neurons associated with the current neuron using settings. The double layer perceptron acts as a classifier.

The results of experimental studies of classification methods are described in [8, 9]. During the work, a process was proposed containing the following classification methods:

- naive Bayesian classifier (NB);
- method of k-nearest neighbors (k-NN);
- decision tree (DT);
- support vector machine (SVM);
- neural network (ANN);
- linear discriminant analysis (LDA).

The results of the study are summarized in Table 1.

Table 1

Recognition accuracy on different datasets

Number of set	NB	k-NN	DT	SVM	ANN	LDA
1	86.75 %	–	83.72 %	88.07 %	–	76.5 %
2	87.02 %	–	93.17 %	85.87 %	95.16 %	–
3	–	93.80 %	–	91.13 %	94.10 %	77.0 %

Note: built on the basis of data [8, 9]

Analysis of presented in Table 1 data shows that the neural network structure (ANN) has the lowest error rate, but neural networks have several disadvantages:

- most ANN design approaches are heuristic and often do not lead to unambiguous solutions;
- to build a model of an object based on ANN, it is necessary to perform the configuration of the internal elements and the connections between them;
- the problems of preparing a training set are associated with the difficulties of finding a sufficient number of training examples;
- significant time costs of training do not allow the use of ANN in real-time systems;
- behavior of a learned ANN can't be unambiguously predicted, which increases the risk of using ANN to control expensive technical objects.

One of the key tasks in the implementation of neural network structures is the search for the optimal ratio of parameters and their characteristics in each specific case. To effectively solve such a problem, a wide range of methods, algorithms and synthesis methods are needed, differing in

the amount of computations, quality of results, time to search for solutions and methods of data presentation.

Thus, the results of literary analysis allow to conclude that such approaches and methods still remain reliable means of recognition, but using only such approaches it is impossible to obtain the best results for various applied tasks [10]. Therefore, it is necessary to use combined approaches and methods that focus on minimizing hardware costs when implementing computer vision systems.

5. Methods of research

The recognition task is a relation:

$$Z \subseteq I^{\text{in}} \times I^{\text{out}},$$

where I^{in} – information about a certain set of objects, divided into classes (their number is finite), and the method of presenting information is precedent, and for Z only part of the information are known (training and control samples); I^{out} – class labels/indexes. Object recognition can be divided into stages:

- 1) selection of the observation area;
- 2) coding the found area for further recognition steps;
- 3) comparison and construction of classification models;
- 4) calculation of the probability of belonging to a particular class;
- 5) summary of the results.

The first three stages relate to the preparatory stages of image recognition, during which a model is formed suitable for recognition by the chosen method. In the framework of this work, such methods will not be considered. This paper discusses the combined model of image recognition based on the theory of non-force interactions. The fundamental concept of the theory of non-force interaction is the concept of information. In the generally accepted interpretation, information means information, data, knowledge obtained and transmitted in the process of interaction with other subjects. Another fundamental concept is the Vip-interpretation of motion, which is fed through a variety of displacements behind or against the direction of motion with a speed that is equal to the speed of light in a vacuum and the probability determined by the introformation in the content of moving objects. Elements of the theory of non-force interaction for the first time are used as an apparatus for image recognition. Basically, the theory of non-force interactions is used to solve the following problems:

- access to databases based on the analysis of natural language phonemes;
- assessment of investment proposals;
- assessment and prediction of the impact of harmful substances in water resources on the health of the population;
- forecasting the results of sporting events.

The image model for recognition by the introformation approach in the framework of this work is based on the voting approach and elements of the local binary pattern method. The ideological basis for voting in computer vision is Hough transformations, and the theoretical basis is set theory, methods for calculating estimates, and statistical analysis. An important feature of a voting system is the ability to obtain a resulting list of fragments or features by which a decision is made, as well as the ability to stop recognition at an early stage when certain characteristics are achieved, saving on the amount of computation.

Let feature classes be represented by reference descriptions in the form of finite sets of primitives, which are calculated for a classification data set. Let the classification object $Z = \{z_i\}_{i=1}^s$ be represented by a finite set of characteristic features (s – the number of elements), reference descriptions:

$$Z(j) = \{z_i^j\}_{i=1}^{s_j}, \quad j = 1, \dots, J,$$

where J – finite number of classes. Characteristic features are a numerical vector of fixed dimension $z_i, z_i^j \in R^n$, defined in a separate point of the image. The classification criterion is the number of distinctive features, assigned to $Z(j)$. The evaluation mechanism is associated with voting $z_i \in Z$. The description Z will be assigned to the class that received the most votes.

For a constant database of descriptions $Z(j)$, the probability $p(z, j)$ can be calculated in terms of the partial values of the two measures, which express the number of favorable cases to their total number:

$$p(z, j) = \frac{f(z, j)}{\sum_{i=1}^J f(z, i)}, \quad (1)$$

where $f(z, j) \geq 0$ – integral function that reflects the value of compliance z with the class j . For matching, it is necessary to calculate the value of the function $f(z, j)$:

$$f(z, j) = \exp[\lambda(z, Z(j))], \quad (2)$$

where $\lambda(z, Z(j)) = \min_{z_i^j \in Z(j)} \chi_z(z, z_i^j)$ – distance from z to $Z(j)$.

For the classification of images, the distance between the histograms of images is used, obtained by calculating the method of local binary patterns. Based on the data of distances and voting (1), a statistical sample (3), (4) is constructed, which allows to evaluate the events of the object (D_0), as well as the conditions ($b_j, j = \overline{1, n}$):

$$p_0 = p(D_0), \quad (3)$$

$$p_j = p\left(\frac{D_0}{b_j}\right), \quad j = \overline{1, n}. \quad (4)$$

From the known probabilities of the manifestation (action) of system (3), (4), its definiteness (5) is calculated with respect to these manifestations:

$$d_j = \begin{cases} +0.5 \sqrt{\frac{1-p}{p} + \frac{p}{1-p}} - 2, & \text{at } 0.5 \leq p < 1, \\ -0.5 \sqrt{\frac{1-p}{p} + \frac{p}{1-p}} - 2, & \text{at } 0 < p < 0.5. \end{cases} \quad (5)$$

From certainty of system (5), its awareness is calculated:

$$i_j = \sqrt{d^2 + 1}. \quad (6)$$

Then the total, for all actions on the system, increase in the certainty of the action of the system is calculated. At the same time, introformations are not used to represent the change in the impulses of objects:

$$\Delta d^P = i_0 \sum_{j=1}^n d_j - d_0 \sum_{j=1}^n i_j, \quad (7)$$

or introformation representation of the change in kinetic energy:

$$\Delta d^E = \begin{cases} +\sqrt{\frac{\alpha^2}{2} + \sqrt{\frac{\alpha^4}{4} + \alpha^2}}, & \text{at } \alpha \geq 0, \\ -\sqrt{\frac{\alpha^2}{2} + \sqrt{\frac{\alpha^4}{4} + \alpha^2}}, & \text{at } \alpha < 0, \end{cases} \quad (8)$$

where

$$\alpha = \sum_{j=1}^n \left[\text{sgn}(d_j i_0 - d_0 i_j) \cdot \frac{(d_j i_0 - d_0 i_j)^2}{\sqrt{(d_j i_0 - d_0 i_j)^2 + 1}} \right]. \quad (9)$$

The impulse corresponds to the magnitude of the influence of an image fragment on the reaction. Based on the introformation representation of the change in impulse of objects (7) or the introformation representation of the change in kinetic energy (8), an increase in the system awareness is calculated:

$$\Delta i = \sqrt{\Delta d^2 + 1}. \quad (10)$$

Then a new definition of the system action is calculated:

$$d_\Sigma = \Delta d \cdot i_0 + d_0 \cdot \Delta i, \quad (11)$$

and new awareness of the system action:

$$i_\Sigma = \sqrt{d_\Sigma^2 + 1}. \quad (12)$$

According to physical laws, and on the basis of the obtained new values of certainty (11) and awareness (12), the probability of the system is calculated:

$$p_\Sigma = 0.5 + \frac{d_\Sigma}{2i_\Sigma}. \quad (13)$$

The obtained probabilities (13) should be further generalized:

$$\begin{aligned} \overline{s(r_i)} &= \frac{\sum_{j \neq i} [(p(r_i) + p(r_j)) \cdot p_{r_j}(r_i)]}{2 \cdot \frac{C_m^2}{m}} = \\ &= \frac{\sum_{j \neq i} [(p(r_i) + p(r_j)) \cdot p_{r_j}(r_i)]}{m-1}, \end{aligned} \quad (14)$$

where m – the number of possible solutions; $\overline{s(r_i)}$ – evaluation of the probability of choosing a solution; $p(r_i)$ – unconditional probability of choosing a solution r_i ; $p(r_j)$ – unconditional probability of choosing a solution r_j ; $p_{r_j}(r_i)$ – compatible conditional probability of choosing a solution for an alternative solution r_j .

In the general case, the construction of a model for a specific object requires, according to the results of measurements of the input and output signals, the assignment of this object to a certain class of objects. This paper presents a generalized model of the combined model of image recognition, in which the image processing process

looks like a stream consistently passes through the stages of preprocessing, image description, mapping and classification. The combined model assumes that the transition from one phase of image processing to another occurs only after a full and successful completion of the previous phase, transitions back, forward or overlapping of the phases – does not occur.

6. Research results

Testing of the combined model of image recognition was carried out on four different data sets:

- a set of images of small objects ALOI (Amsterdam Library of Object Images) [11]. This collection contains 1000 classes of images of small objects of 24 copies per class. In each instance of the class systematically vary: angle of view, angle of illumination and color of the backlight. All images have a fixed size of 192×144 pixels in grayscale in PNG format;
- a set of images of handwritten characters formed by T de Campos in Microsoft Research India (The Chars 74k dataset) [12]. This set contains 62 classes of characters (0–9, A–Z, a–z) with 55 copies per class. Character images do not have fixed sizes and are not centered relative to the image. All images are strictly in a black and white palette, 1200×900 pixels in PNG format;
- a set of images of persons from the Cambridge AT & T Laboratory (ORL Database of Faces) [13]. In this set are images of 40 people in 10 copies per person. The pictures were taken at different times, with changes in lighting, facial expressions and details. All pictures were taken on a dark uniform background with subjects in a strictly vertical frontal position. All files have a fixed size of 92×112 pixels in grayscale in PGM format;
- a set of images of the base CDI Set (Celebrities Data Images) [14]. In this set 20 excellent people were used with 20 images per person. The pictures were taken at different times, with changes in lighting, background, angle of shooting, details, facial expressions and head position. All files have various sizes, made in polychrome coloring, the format of images is JPG.

The studies were conducted on Apple mobile devices, all devices were based on the iOS version 7.1.2 mobile operating system, using the SDK Xcode 5.1.1 stable release 5B1008. To study the reliability of image recognition of the proposed combined model, several experiments were conducted on the data sets described above. In these experiments, the dependences of the recognition accuracy on the method settings for the description of images and its key fragments, the dependence of recognition accuracy on the amount of input data, and the time dependencies on the method settings for the image description were determined. Respectively, sets of input data were prepared. The minimum size of the images used in the experiments was 92×112 pixels, the maximum 2000×3000 pixels.

During the research it was found that the proposed combined model of image recognition by the introformation method based on local binary templates with voting can be used in conditions close to real-time conditions on mobile devices. This approach showed high reliability of recognition (from 60 % to 95 % depending on the set of input data) with the ability to control the recognition accuracy and time losses by setting the description of objects.

Compared with existing analogues, the proposed approach showed mainly the best temporal indicators of learning and recognition on large volumes of input data (Table 2).

Table 2

Average recognition accuracy

No.	Input data	The average percentage classification by algorithms			
		LBP	PCA	LDA	Combined model
0	Item Images (ALOI)	70.861	98.154	98.702	99.429
1	ALOI+histogram alignment (h. e.)	44.607	90.948	98.769	98.722
2	Handwritten characters (numbers)	33.411	84.968	80.13	84.475
3	Handwritten characters (large letters)	29.85	84.754	68.157	80.126
4	Handwritten characters (lower case letters)	20.812	73.872	49.905	64.342
5	Handwritten characters (entire sample)	19.167	64.166	28.267	54.258
6	The face of the base of images AT & T	85.753	93.973	91.665	94.564
7	AT & T+h. e.	84.47	87.449	90.557	91.372
8	Faces of the CDI image base, the training sample size is 5 samples per class	34.675	26.918	29.916	34.833
9	CDI (5)+h. e.	37.3	26.018	31.633	33.463
10	CDI (5)+geometric alignment of images (g. e.)	39.341	28.959	42.085	45.425
11	CDI (5)+h. e.+g. e.	42.713	28.836	48.265	48.075
12	CDI, training sample size 10 samples per class	40.009	30.61	40.828	45.635
13	CDI (10)+h. e.	41.025	33.721	44.264	47.204
14	CDI (10)+g. e.	47.996	37.307	55.969	58.673
15	CDI (10)+h. e.+g. e.	46.467	41.11	61.01	60.175

The average classification accuracy by the proposed approach was 71.3 % (LBP – 59.9 %; PCA – 65.2 %; LDA – 65.6 %). This, combined with the temporary loss of learning and recognition, makes this approach acceptable for use in systems close to real-time conditions on mobile devices.

7. SWOT analysis of research results

Strengths. Studies have shown that the proposed combined model of image recognition allows not only image processing, but also training and retraining in near real-time mode with limited mobile device resources better than known approaches to image recognition. The proposed combined model also makes it possible to control the reliability and temporary loss of the recognition process due to the reduction of the feature space.

Weaknesses. The proposed model shows the highest reliability of recognition only in cases when the recognition object is «known» (handwritten symbol, faces, etc.). As part of this work, only the recognition model was studied and developed. Issues of classification and preprocessing in the framework of this work were not considered.

Opportunities. The promise of the proposed combined model is ensured by low computational complexity and high reliability of recognition. This allows not only to work on devices with limited hardware resources (such as mobile devices), but also to build complex computing systems for processing Big Data, which are a promising direction of the present. Since the elements of the theory of non-force interactions that lie at the basis of the proposed combined model, have proven themselves in other data processing tasks, the created combined model can be adapted for other areas of recognition: sound, behavior, etc. kind of events and consequences.

Threats. One of the main factors affecting the success of recognition with high confidence is the conditions for collecting information and methods for preparing and processing data. Moreover, the accuracy of recognition varies in a significant range depending on the conditions for collecting information. The research results were obtained from data sets obtained under conditions close to ideal. Therefore, when working with real data, it may be necessary to additionally apply methods and approaches to image preprocessing aimed at:

- compensation of distortions caused by camera vibration;
- reduction of noise introduced by the camera and depend on the type and size of its sensor, as well as shooting conditions;
- alignment of image brightness, etc.

8. Conclusions

1. Studies show that the theory of non-force interactions, in contrast to the existing classification methods, has the following key features:

- designed to handle large amounts of data;
- selects useful information from an arbitrary stream;
- allows to naturally add new objects;
- tolerant of errors and allows to quickly reprogram the behavior of the system.

2. Based on the theory of non-force interactions with the description of objects using local binary patterns and the voting approach, a new model for recognizing graphical objects in software systems was formed. The peculiarity of this model lies in the combination of non-linear structure, which is inherent in neural network structures (theory of non-force interaction) and low computational complexity. This allows software systems based on this model to operate in near real-time modes with limited resources of mobile devices.

3. A model built on the basis of the proposed image recognition method allows to control the recognition accuracy. This is achieved by adjusting the number of sectors of the histograms of local binary patterns that are used in the description of images and the number of image fragments used in the classification stage by the introduction approach. To a large extent, the number of image fragments affects the time of classification, since in this case, it is necessary to calculate the matching of the system actions in each of the possible directions in pairs.

References

1. Wagner, P. (2011). *Principal Component Analysis and Linear Discriminant Analysis with GNU Octave*. Available at: https://www.bytefish.de/blog/pca_lda_with_gnu_octave/

2. Samal, A., Iyengar, P. A. (1992). Automatic recognition and analysis of human faces and facial expressions. *Pattern Recognition*, 25 (1), 65–77. doi: [http://doi.org/10.1016/0031-3203\(92\)90007-6](http://doi.org/10.1016/0031-3203(92)90007-6)
3. Ojala, T., Pietikainen, M., Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 (7), 971–987. doi: <http://doi.org/10.1109/tpami.2002.1017623>
4. Tomashevich, N. S. (2007). Statisticheskie metody vydeleniia priznakov. *Neirokompiutery v prikladnykh zadachakh obrabotki izobrazhenii*, 121–128.
5. Volchenkov, M. P., Samonenko, I. Iu. (2005). Ob avtomaticheskoi raspoznavanii lits. *Intellektualnye sistemy*, 9 (1-4), 135–156.
6. Simard, P. Y., Steinkraus, D., Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. *12th International Conference on Document Analysis and Recognition*, 2, 958. doi: <http://doi.org/10.1109/icdar.2003.1227801>
7. LeCun, Y., Huang, F.-J., Bottou, L. (2004). Learning methods for generic objects recognition with invariance to pose and lighting. Los Alamitos. *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR'04)*, 2, 97–104. doi: <http://doi.org/10.1109/cvpr.2004.1315150>
8. Mitrofanov, C. A. (2015). Sravnenie effektivnosti razlichnykh metodov intellektualnogo analiza dannykh v zadachakh raspoznavaniia izobrazhenii. *Innovatsionnaia nauka*, 12 (2), 96–98.
9. Mamontov, D. Iu., Karaseva, T. S. (2015). Reshenie zadach finansovogo analiza s pomoshchiu intellektualnykh informatsionnykh tekhnologii. *ITSiT*. Available at: <https://studfiles.net/preview/5966499/>
10. Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1106–1114.
11. Amsterdam Library of Object Images (ALOI). Available at: <http://aloi.science.uva.nl/>
12. The Chars74K dataset. Available at: <http://www.ee.surrey.ac.uk/CVSSP/demos/chars74k/>
13. The Database of Faces. Available at: <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>
14. Celebrities Data Images Set for Computer Vision. Available at: <http://cdiset.blogspot.com/>

Voloshyn Mykola, Postgraduate Student, Department of Hybrid Modeling and Control Systems in Power Engineering, Pukhov Institute for Modelling in Energy Engineering of National Academy of Sciences of Ukraine, Kyiv, Ukraine, e-mail: info@dreams.global, ORCID: <http://orcid.org/0000-0003-1290-6152>