UDC 629.735.05 (045)
DOI: 10.18372/1990-5548.55.12771

[1]V. M. Sineglazov,
[2]S. V. Kostiuchenko

DIAGNOSTIC SYSTEM BASED ON AUTOENCODERS

[1,2]Aviation Computer-Integrated Complexes Department, Educational & Research Institute of Information and Diagnostic Systems, National Aviation University, Kyiv, Ukraine
E-mails: [1]svm@nau.edu.ua, [2]sv9toslavkost@gmail.com

***Abstract—**It`s considered the problem of image processing which is used in diagnostic systems when it is necessary to process the results of ultrasound, computed tomography and magnetic resonance imaging. For the solution of this problem it`s often used artificial neural networks especially convolution neural networks. It`s considered the structure of convolutional neural networks especially types of layers. In the paper it is analyzed the creation of convolutional neural networks based on autoencoders. It`s considered the features of such neural network, the algorithm of learning and the important parameters which determine the function quality of image processing. The possible improvement of such topology is possible with help of restricted Boltzmann machine which can be used for pre-learning.*

Index Terms—Image; recognition; autoencoders; neural network; layer; convolution; perceptron; training.

I. INTRODUCTION

Computer Vision problems are among the better studied problems of machine learning. The most common dilemma of these methods is the choice of features that is used to train the classifier. The best realization of classifier is use of neural networks (NN). The classifiers are tested on a very popular digit recognition benchmark Modified National Institute of Standards and Technology (MNIST).

The aim of this work was to implement deep-learning algorithms to supply a stable learning of basic NN, which solves the problem of classification.

Autoencoders are simple learning NN which aim to transform inputs into outputs with the least possible amount of distortion. While conceptually simple, they play an important role in machine learning.

More recently, autoencoders have taken center stage in the “deep architecture" approach [1] where autoencoders, particularly in the form of Restricted Boltzmann Machines (RBMS), are stacked and trained bottom up in unsupervised fashion, followed by a supervised learning phase to train the top layer and ne-tune the entire architecture.

A digital image can be defined as the numerical representation of a real image. This representation can be coded as a vector or a bitmap (raster). In the first case it describes the primitive elements (lines or polygons) which compose the image, in the second case the image is composed of a matrix of points, called pixels. Their color is defined by one or more numerical values. In colored bitmap images the color is stored as level of intensity of the basic colors, for example in the RGB model there are three colors: red, green and blue.

The algorithm can be lossy (JPEG) or lossless, i.e. without loss (GIF, PNG). This type of images is generated by a wide variety of acquisition devices, such as scanners, digital cameras, webcams, but also by radar and electronic microscopes.

The convolution [4] is an operation that is performed on a mono-color bitmap image for emphasizing some of its features. The convolution matrix is also called Filter.

A Filter can be thought as a sliding window moving across the original image. At every shift it produces a new value, this value is obtained by summing all the products between the filter elements and the corresponding pixels. The values obtained from all the possible placements of the filter above the image are inserted in an orderly fashion in a new image.

II. PROBLEM STATEMENT

Let X be set of described objects, Y the set of numbers (or names) of classes. There is a known target dependency is reflection y^*: $X \to Y$, the meaning of which is known only on the objects of the final training sample:

$$X^m = \{(x_1, y_1), \ldots, (x_m, y_m)\},,$$

where X^m set of elements of the training sample by dimension m.

It is necessary to construct an algorithm capable of determining the affiliation of an arbitrary object $x \in X$ to the class $y \in Y$.

The mathematical model of a curved layer in a simplified form can be described by the following formula $x^l = f\left(x^{l-1} * k^l + b^l\right)$, where x^l is the output layer l; $f(\cdot)$ is activation function; b^l is bias ratio, the symbol * denotes the operation of the convolution of the input x^{l-1} with the core k^l. While fill out the formula in the standard form, will get:

$$x_j^l = f\left(\sum_{i=0}^{n}\sum_{j=0}^{m} x_i^{l-1} \cdot k_j^l + b_j^l\right),$$

where x_i^l map of features j (output of layer l; $f(\cdot)$ is activation function; b_j^l is the offset coefficient for the characteristics map j; k_j^l convolution core j; b_i^{l-1} x_i^{l-1} is feature map of previous layer; n, m the image size. Scheme of convolutional layer shown on Fig. 1.

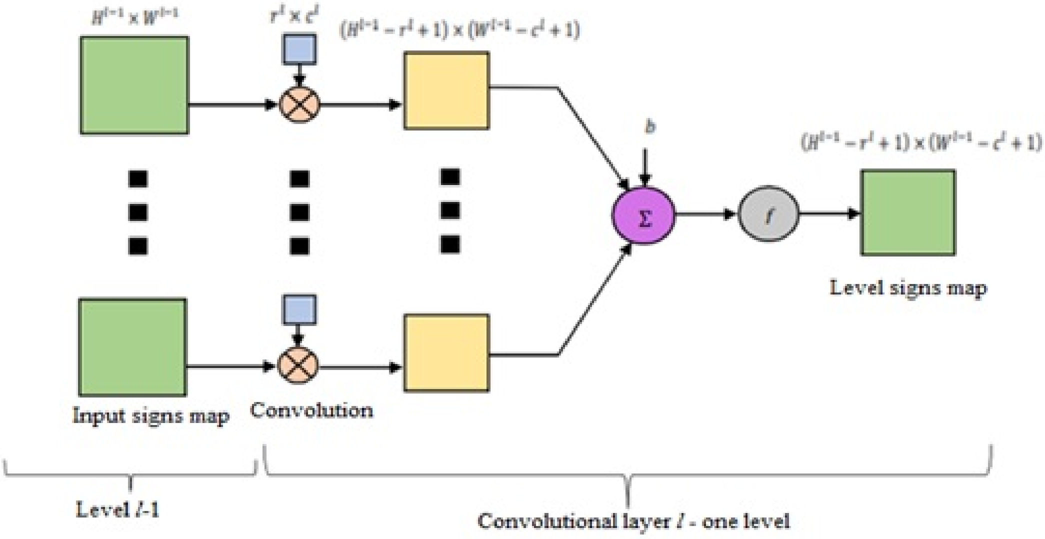

Fig. 1. Scheme of convolutional layer

Aggregating layer acts independently on each section depth entry, and reduces its spatial dimensions. Formally, the layer can be described as follows $x^l = f\left(a^l \cdot \text{subsample}\left(x^{l-1}\right) + b^l\right)$,

where x^l is the layer output l; $f(\cdot)$ is activation function a^l, b^l are coefficients; subsample$(\cdot)$ is sampling operation of local maximum values.

Scheme of the aggregate layer is shown on Fig. 2.

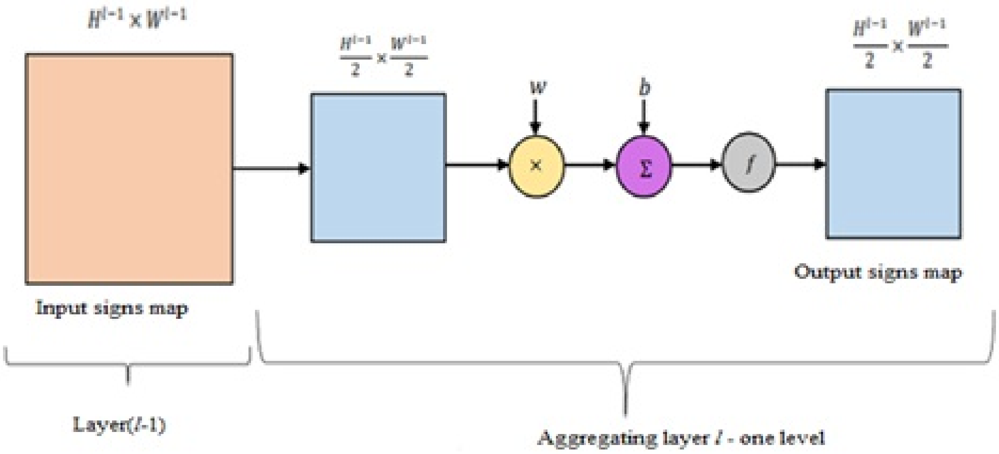

Fig. 2. Scheme of aggregation layer l

After several convolutional and maximizing aggregation layers that are used to select a special area, placed a classifier, which is implemented with fully connected layers, and solves the problem of classification of the selected area. As a classifier in this paper it's considered: auto-encoder, Softmax, fuzzy deep learning classifier. The fully coated layer can be described by the equation

$$x_j^l = f\left(\sum_i^n x_j^{l-1} w_{ij}^{l-1} + b_j^{l-1}\right),$$

where x_j^l is jth output of layer l; f (.) the activation function; b_j^{l-1} is jth layer displacement ratio l; w_{ij}^{l-1} is ith, jth element of the matrix of weight coefficients $l-1$ layer; n is the number of neurons in the full-fledged layer.

III. CONVOLUTIONAL NEURAL NETWORKS

Convolutional Neural Networks (CNN) [6] are powerful classifiers which can be used in many tasks. They are extremely suitable when the number of training pattern and their dimensionality are particularly high. This is why they are often used in computer vision dealing successfully with digital images.

To train a CNN the classic supervised approach based on back-propagation is used. This approach allows the network to learn how to discern a pattern from another based on the training set. In CNNs the sigmoidal function is commonly used as the activation function of all the layers (including the neural network). In this section a general definition of a convolutional neural network has been provided. However, there are many variants depending on the task CNN:

- without final neural network;
- with a final neural network with one or more hidden layers;
- with an even number of layers in their features module.
- with only convolutional layers.

In the dissertation will continue to talk about CNNs referring to their general definition, even if it is important to understand that there are many researchers who can call with the same right convolutional neural networks slightly different architectures (Fig. 3).

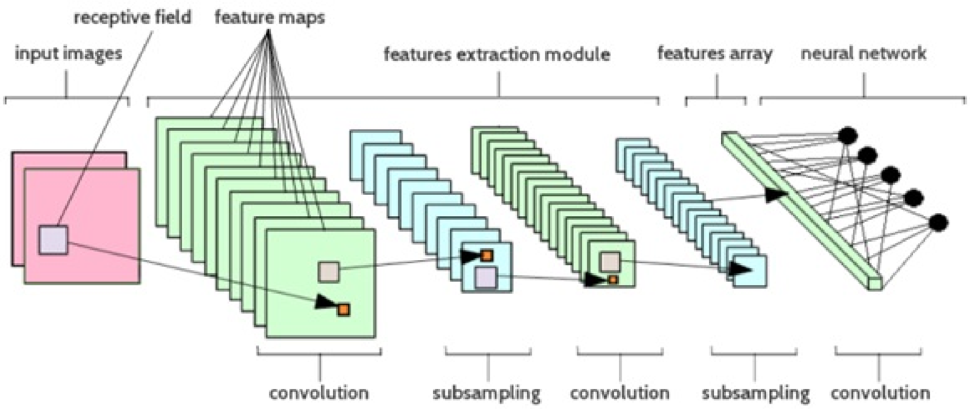

Fig. 3. A complete example of a CNN architecture with seven layer, or commonly called LeNet7

A Convolutional Neural Networks receives in input p mono-color bitmap (for example, can receive the three channels R, G and B of a colored image); then the input is given to a feature extraction module which releases an array of features consisting of m elements; finally this array is delivered to a full connected neural network that generates the results.

Fully connected AEs and DAEs both ignore the 2D image structure. This is not only a problem when dealing with realistically sized inputs, but also introduces redundancy in the parameters, forcing each feature to be global (i.e., to span the entire visual field). However, the trend in vision and object recognition adopted by the most successful models [2], [3] is to discover localized features that repeat themselves all over the input. CAEs differs from conventional AEs as their weights are shared among all locations in the input, preserving spatial locality. The reconstruction is hence due to a linear combination of basic image patches based on the latent code.

For a mono-channel input x the latent representation of the kth feature map is given by

$$hk = \sigma(x \cdot Wk + bk), \tag{1}$$

where the bias is broadcasted to the whole map, σ is an activation function (used the scaled hyperbolic tangent in all our experiments), and denotes the 2D convolution. A single bias per latent map is used, as required each filter to specialize on features of the whole input (one bias per pixel would introduce too many degrees of freedom).

The reconstruction is obtained using

$$y = \sigma\Big(\sum_{k\in H} hk \cdot \tilde{W}k + c\Big), \tag{2}$$

where again there is one bias c per input channel H identifies the group of latent feature maps $\tilde{W}$; identifies the flip operation over both dimensions of the weights. The 2D convolution in equation (1) and (2) is determined by context. The convolution of an $m \times m$ matrix with an $n \times n$ matrix may in fact result in an $(m + n - 1) \times (m + n - 1)$ matrix (full convolution) or in an $(m - n + 1) \times (m - n + 1)$ (valid convolution). The cost function to minimize is the mean squared error (MSE):

$$E(\theta)=\frac{1}{2n}\sum_{i=1}^{n}(x_i-y_i)^2.$$

Just as for standard networks the backpropagation algorithm is applied to compute the gradient of the error function with respect to the parameters. This can be easily obtained by convolution operations using the following equation:

$$\frac{\partial E(\theta)}{\partial W^k}=x\cdot\delta h^k+\tilde{h}^k\cdot\delta y$$

where δh and δy are the deltas of the hidden states and the reconstruction, respectively. The weights are then updated using stochastic gradient descent.

IV. AUTOENCODERS

Today, the most innovative method of reducing the dimensionality of the analyzed vector space is the application of an autoencoder**.** Autoencoder (autoassocator) – a special architecture of neural networks allowing to apply learning without a teacher when using the method of errors back propagation.

The simplest architecture of the auto-encoder is shown on Fig. 4 – direct propagation network without feedbacks, most similar to a perceptron and containing an input layer, an intermediate layer and an output layer.

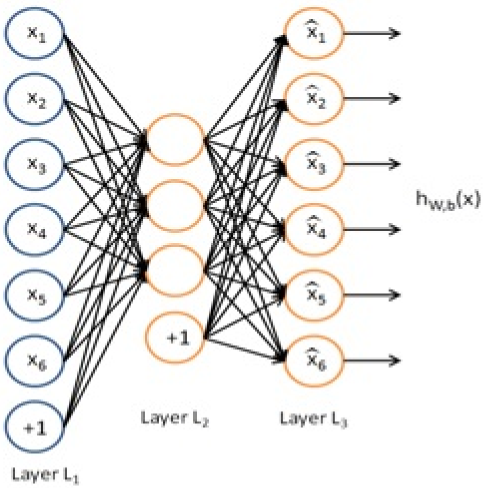

Fig. 4. Autoencoder architecture

The main goal of training an autoencoder is to ensure that the input vector of the characteristics causes a network response equal to the input vector. That is, the problem of the functioning of the auto-encoder is reduced to finding the approximation of such a function so that the response of the neural network is equal to the value of the input characteristics to within a given error value.

Auto-encoders can be stacked to form a deep neural network. A good way to get good parameters for the stack of auto-encoders is to use greedy layer-wise training. In order to do this, first train the first layer on the input sample in order to get the parameters $W^{(1.1)},W^{(1.2)},b^{(1.1)},b^{(1.2)}$. Then it`s necessary to use the first layer to convert the input sample into a vector consisting of data from the hidden layer neurons. Then needed will train the second layer with this vector to obtain the parameters $W^{(2.1)},W^{(2.2)},b^{(2.1)},b^{(2.2)}$.

These steps are repeated for successive layers using the output of each preceding layer as an input to the next.

This method trains the parameters of each layer individually, without affecting the parameters of the rest of the model. To get better results after this learning phase, fine-tune the error propagation method to improve the results of the model settings work of all layers of the model together. When using the stack of autoencoders to solve the classification problem, autoencoders are discarded with the last layer (decoder) and send the output of the last layer to the softmax classifier. The error gradients of the softmax classifier will then be extended back to the coding layers of the model.

Formulation of the research task. The analysis of the scientific and methodical apparatus made it possible to identify certain contradictions in theory and practice.

To solve practical problems, it is necessary to improve the quality of object recognition on images. At the same time, in order for practical problems to be solved, existing methods and neural network models of object recognition on images need modification.

Thus, the purpose of the study is the development of a new intelligent imaging system based on autoassociators. The scientific task of the study is the development of a method for recognizing objects based on the use of the deep learning network model. It can be concluded that, along with the advantages in neural networks has its disadvantages, as any approach. First of all, this concerns the need to create a training sample, without which the network cannot be used. First of all, this concerns the need to create a training sample, without which the network cannot be used.

Today, the issue of creating effective training samples is under-researched. For most scientific studies, large samples are used (NORB, Caltech 101, Pascal). But if it is a question of solving a specific applied problem, there is uncertainty about how certain parameters of the training sample affect the learning quality and network recognition, what automation methods are used to create the sample, and what algorithms and methods to use.

For neural network algorithms, as for any other, speed of execution is critical, but, unfortunately, most modern models are not designed for fast processing. Therefore, the critical issue is the acceleration of the work of neural networks.

Also, most models based on neural networks are not capable of independently isolating individual objects in the image and require preliminary segmentation and preliminary processing of the input image at the input of the network

For the searching optimal values of structural parameters (number of layers and neurons in every layer) and weighting coefficients it is used genetic algorithm.

V. MODIFIED NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY

The Modified National Institute of Standards and Technology [5] data set of handwritten digits is a very popular benchmark for image recognition tasks. It has a training set of 60.000 examples, and a test set of 10.000 examples. Considered the quality of digit classification on the MNIST data set using the provided test set. The classification accuracy will be expressed as a ratio of correctly classified digits to the size of the test set. Because of its popularity the MNIST data set is a de facto standard benchmark for learning methods.

The first stage of training convolutional network is the initiation of weight coefficients. In the general case, if the convolutional neural network uses the layers of convolutions, aggregates, and fully-coherent layers, then only the nuclei of the convolution of the coil layers and the full-coated layers should be initiated. If the network is also used in the convolutional layers, it is also necessary to initiate their weighting factors. To train the convolutional neural network in this work, a normalized initialization is used, which is called the initialization of Glorot $W \cong U\left[-\frac{\sqrt{6}}{\sqrt{n_j+n_j}}, \frac{\sqrt{6}}{\sqrt{n_j+n_{j+1}}}\right]$, where U is uniform distribution on a segment; n_j the number of neurons in the current layer of the network; n_{j+1} is the number of neurons in the next layer of the network. The use of normalized initialization leads to a decrease in the saturation of the neurons and the error signal extends much better.

The greedy layerwise training protocol can be summarized as follows: to construct a deep pretrained network of n layers divide the learning into n stages. In each stage exploited an unsupervised training algorithm: in the first stage train an autoencoder on the provided training data sans labels. After training the resulting autoencoder will learn the features of this data. Next necessary map the training data to the feature space (in practice by cutting out the last layer of autoencoder so that inputs are mapped to hidden layer activations – features). The mapped data is then used to train the next stage autoencoder. The training follows layer by layer until the last one. The last layer is trained as a classifier (not as an autoencoder) using supervised learning.

VI. CONCLUSION

It`s considered the solution of image processing problem based on stacked autoincoders and determined the approach of structural-parametric synthesis of this neural network. It`s proposed to use genetic algorithm. It permits to improve the efficiency of image processing problem solution.

REFERENCES

[1] Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent, and SamyBengio, “Why does unsupervised pre-training help deep learning,” *Journal of Machine Learning Research*, 11:625-660, February 2010.

[2] D. Lowe, “Object recognition from local scale-invariant features,” in *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, 1999, pp. 1150–1157.

[3] T. Serre, L. Wolf, , and T. Poggio, “Object recognition with features inspired by visual cortex,” in *Proc. of Computer Vision and Pattern Recognition Conference*, 2007.

[4] Tim Morris. Computer vision and image processing. Palgrave Macmillan, 2004.

[5] Yann Lecun and Corinna Cortes, The MNIST database of handwritten digits.

[6] Yann Lecun, Leon Bottou, Yoshua Bengio, and Patrick Haffner, “Gradient-based learning applied to document recognition,” in *Proceedings of the IEEE*, 1998, pp. 2278–2324.

Received December 02, 2017.

Sineglazov Victor. Doctor of Engineering Science. Professor.
Aviation Computer-Integrated Complexes Department, Education&Scientific Institute of Information-Diagnostics Systems, National Aviation University, Kyiv, Ukraine.
Education: Kyiv Polytechnic Institute, Kyiv, Ukraine, (1973).
Research area: Air Navigation, Air Traffic Control, Identification of Complex Systems, Wind/Solar power plant.
Publications: more than 600 papers.
E-mail: svm@nau.edu.ua

Kostiuchenko Sviatoslav. Bachelor.
Aviation Computer-Integrated Complexes Department, Education & Scientific Institute of Information-Diagnostics Systems, National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine, (2016).
Research interests: neural networks, image processing, computer vision, autoencoders
Publications: 1.
E-mail: sv9toslavkost@gmail.com

В. М. Синєглазов, С. В. Костюченко. Діагностична система на базі автоенкодерів

Розглянуто проблему обробки зображень, яка використовується в діагностичних системах, коли необхідно обробляти результати ультразвуку, комп'ютерна томографія та магнітно-резонансна томографія. Для вирішення цієї проблеми часто використовуються штучні нейронні мережі, особливо згорткові нейронні мережі. Розглянуто структуру згорткових нейронних мереж, особливо типи шарів. У роботі проаналізовано створення згорткових нейронних мереж на базі автоенкодерів. Розглянуто особливості такої нейронної мережі, алгоритм навчання та важливі параметри, що визначають якість функції обробки зображень. Можливе поліпшення такої топології можливе за допомогою обмеженої машини Больцмана, яку можна використовувати для попереднього навчання.

Ключові слова: зображення; розпізнавання; автоенкодер; нейронна мережа; шар; згорткова; персептрон; навчання.

Синєглазов Віктор Михайлович. Доктор технічних наук. Професор.
Кафедра авіаційних комп'ютерно-інтегрованих комплексів, Навчально-науковий інститут інформаційно-діагностичних систем, Національний авіаційний університет, Київ, Україна.
Освіта: Київський політехнічний інститут, Київ, Україна (1973).
Напрям наукової діяльності: аеронавігація, управління повітряним рухом, ідентифікація складних систем, вітроенергетичні установки.
Кількість публікацій: більше 600 наукових робіт.
E-mail: svm@nau.edu.ua

Костюченко Святослав Віталійович. Бакалавр.
Кафедра автоматизації та комп'ютерно-інтегрованих технологій, Навчально-науковий інститут інформаційно-діагностичних систем, Національний авіаційний університет, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна (2016).
Напрямок наукової діяльності: нейронні мережі, обробка зображень; комп'ютерний зір; автоекнодери.
Публікації: 1.
E-mail: sv9toslavkost@gmail.com

В. М. Синеглазов, С. В. Костюченко. Диагностическая система на базе автоэнкодеров

Рассмотрена проблема обработки изображений, которая используется в диагностических системах, когда необходимо обрабатывать результаты ультразвука, компьютерная томография и магнитно-резонансная томография. Для решения этой проблемы часто используются искусственные нейронные сети, особенно сверточные нейронные сети. Рассмотрена структура сверточных нейронных сетей, особенно типы слоев. В работе проанализировано создание сверточных нейронных сетей на базе автоэнкодеров. Рассмотрены особенности такой нейронной сети, алгоритм обучения и важные параметры, определяющие качество функции обработки изображений. Возможное улучшение такой топологии возможно с помощью ограниченной машины Больцмана, которую можно использовать для предварительного обучения.

Ключевые слова: изображение; распознавание; автоэнкодер; нейронная сеть; слой; сверточная; персептрон; обучение.

Синеглазов Виктор Михайлович. Доктор технических наук. Профессор.
Кафедра авиационных компьютерно-интегрированных комплексов, Учебно-научный институт информационно-диагностических систем, Национальный авиационный университет, Киев, Украина.
Образование: Киевский политехнический институт, Киев, Украина (1973).
Направление научной деятельности: аэронавигация, управление воздушным движением, идентификация сложных систем, ветроэнергетические установки.
Количество публикаций: более 600 научных работ.
E-mail: svm@nau.edu.ua

Костюченко Святослав Витальевич. Бакалавр.
Кафедра автоматизации и компьютерно-интегрированных технологий, Учебно-научный институт информационно-диагностических систем, Национальный авиационный университет, Киев, Украина.
Образование: Национальный авиационный университет, Киев, Украина(2016).
Направление научной деятельности: нейросети, обработка изображений, компьютерное зрение, автоэнкодеры.
Публикации: 1.
E-mail: sv9toslavkost@gmail.com