

- [13] ISO/IEC 31010:2009 Risk management – Risk assessment techniques (IDT) (2009). Available at: <https://www.iso.org/standard/51073.html>
- [14] Henley, E. J., Kumamoto, H. (1981). Reliability Engineering and Risk Assessment. By Englewood Cliffs, Prentice-Hall, New Jersey.
- [15] Baldwin, J., Lin, Z. (2002). Impediments to advanced technology adoption for Canadian manufacturers. Research Policy, 31 (1), 1–18. doi: [https://doi.org/10.1016/s0048-7333\(01\)00110-x](https://doi.org/10.1016/s0048-7333(01)00110-x)
- [16] Chapman, C., Ward, S. (1996). Project risk management: processes, techniques and insights. John Wiley, 322.
- [17] DeMarco, T., Lister, T. (2003). Waltzing with Bears: Managing Risk on Software Projects. Dorset House, 144.

Received date 05.11.2019

Accepted date 30.01.2020

Published date 31.07.2020

© The Author(s) 2020

This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0>).

APPLICATION OF A CONVOLUTIONAL NEURAL NETWORK AND A KOHONEN NETWORK FOR ACCELERATED DETECTION AND RECOGNITION OF OBJECTS IN IMAGES

Victor Skuratov

*All-Russian Research Institute of Radio Engineering
22 Bol'shaya Pochtovaya str., Moscow, Russian Federation, 105082
viktor.skuratov@gmail.com*

Konstantin Kuzmin

*Department of Mathematical and Instrumental Methods in Economics
University of Russian Innovation Education
10 Krasnobogatyrskaya str., Moscow, Russia, 107061
konstantin.alexandrovich@yahoo.com*

Igor Nelin

*Department of Radiolocation, Radio Navigation and On-Board Radio Electronic Equipment
Moscow Aviation Institute
4 Volokolamskoe highway, Moscow, Russian Federation, 125993
nelin.iv@yandex.ru*

Mikhail Sedankin

*State Research Center – Burnasyan Federal Medical Biophysical Center of
Federal Medical Biological Agency
23 Marshala Novikova str., Moscow, Russia, 123098
Department of Fundamentals of Radio Engineering
National Research University "Moscow Power Engineering Institute"
14 Krasnokazarmennaya str., Moscow, Russian Federation, 111250
msedankin@yandex.ru*

Abstract

One of the most effective ways to improve the accuracy and speed of algorithms for searching and recognizing objects in images is to pre-select areas of interest in which it is likely to detect objects of interest. To determine areas of interest in a pre-processed radar or satellite image of the underlying surface, the Kohonen network was used. The found areas of interest are sent to the convolutional neural network, which provides the final detection and recognition of objects. The combination of the above methods allows to speed up the process of searching and recognizing objects in images, which is becoming more expedient due to the constantly growing volume of data for analysis. The process of preliminary processing of input data is described, the process of searching and recognizing patterns of aircraft against the underlying surface is presented, and the analysis of the results is carried out. The use of

the Kohonen neural network makes it possible to reduce the amount of data analyzed by the convolutional network by 18–125 times, which accordingly accelerates the process of detection and recognition of the object of interest. The size of the parts of the input image fed into the convolutional network, into which the zones of interest are divided, is tied to the image scale and is equal to the size of the largest detectable object, which can significantly reduce the training sample. Application of the presented methods and centering of the object on training images allows accelerating the convolutional network training by more than 5 times and increasing the recognition accuracy by at least 10 %, as well as minimizing the required minimum number of layers and neurons of the network by at least halving, respectively increasing its speed.

Keywords: pattern recognition, object search, Kohonen neural network, convolutional neural network, radar and satellite images.

DOI: 10.21303/2461-4262.2020.001360

1. Introduction

Currently, the low speed of searching and recognizing objects in images is one of the main problems of the used visual data processing systems. An increase in the speed of image processing can lead to a significant increase in the performance of systems for analyzing satellite and radar images, medical images, data from robotic systems, and unmanned vehicles. Obviously, this fact is capable of exerting both serious technological and obvious economic effects.

One of the most modern ways to increase the accuracy and speed of recognition algorithms is to use a convolutional neural network (CNN) [1]. For the high-quality work of this type of neural networks (NS), it is necessary that the size of the recognized object and the sizes of objects of the same type in the training set be comparable. If the image has a variation in the scale of the object, then the creation of a training sample and training of such a network will take considerable time and computational resources, the number of errors will increase, and the analysis speed will be low. The presented algorithm allows one to reduce the amount of data being analyzed when searching and recognizing objects in an image by 15–100 times, increase accuracy and save computing resources. This is achieved through the application of the algorithm for object allocation and the determination of the centers of interest zones in the input image by the Kohonen network, in which the probability of finding the desired object is high.

2. Literature review and problem statement

Today, there are a number of works [2–4], where the authors describe the application of the Kohonen neural network to the problems of image analysis and recognition. The Kohonen network is used both for preparing input data with subsequent analysis by other types of NS, and for direct pattern recognition. In [2], the Kohonen network for recognizing color images is presented, and a significant drawback is that the image is classified as a whole without detecting the object, and when the object is scaled in the input image, the probability of correct recognition will be low. In [3], an algorithm for finding an object in color images by the Kohonen network is shown. A “window” is passed pixel by pixel in the image, the size of which corresponds to the size of the desired object. Search for an object of 211×169 pixels in an image of 621×497 pixels is performed in 122 seconds. During this time, 135,219 histograms are analyzed, which requires serious computational resources and time costs, which do not allow using the proposed algorithm for working with big data. In [4], a classification algorithm for images of a hybrid NS is described, in which the Kohonen network first analyzes the data and then the data are fed to the perceptron network, which allows reducing the required number of network neurons by more than 2 times and reducing the classification error by 4–5 times. In [5], the Kohonen network is used to search for cluster centers by finding the minimum Euclidean distance between cluster points.

One of the most modern types of NS used in pattern recognition is the convolutional NS, which got its name due to the presence of a convolution operation. In [6], the structure of the convolutional NS and the algorithm for its application in pattern recognition are presented, and in [7] for the analysis of video sequences.

Articles [8–12] describe the development of the R-CNN (Region-based Convolutional Network) algorithm for searching and recognizing objects in images. In [8, 9], the initial version of the

algorithm is described, where about 2000 regions are selected on the input image, each of which is scaled by the affine conversion and fed to the input of a convolutional NS that performs recognition. The article [9] additionally describes the training of linear regression to clarify the location of an object in the image, which improves the quality of detection by 3–4 %. R-CNN has several drawbacks, mainly due to the high time spent on network training, as well as on direct image processing by the algorithm. Processing one medium-sized image takes about 47 seconds. The article [10] Fast R-CNN is a continuation of [8, 9], and in it the authors propose to submit a complete image to the input of the NS, generate a feature map for it, and then determine the maps for each region, which significantly reduced the detection time. [11] Faster R-CNN proposes replacing the procedure for generating regional maps with a separate convolution network that extracts features from the image and determines the boundaries of regions of interest. The sign map of the input image is traversed by a moving “window”, highlighting the vector of low-dimensional signs, which are then submitted for analysis to the NS. Mask R-CNN [12] develops the early architecture and adds the ability to predict the position of the mask covering the found object.

The article [13] describes the YOLO algorithm (You Only Look Once), which superimposes a grid on the input image and divides it into cells, determining the zones of possible location of objects, assessing the accuracy of detection and the probability of belonging to classes. The algorithm analyzes from several thousand to several tens of thousands of image parts with bounding frames of different sizes, allowing the detection and recognition of objects 10^3 times faster than R-CNN and 10^2 times faster than Fast R-CNN, but with a lower accuracy.

In [14], the SSD: Single Shot MultiBox Detector algorithm is presented, in which the entire area of the input image is overlapped by bounding frames. The size of the frames varies within the specified limits, which allows to detect and recognize objects of various sizes. For each part of the image highlighted by the frame, the probability of belonging to the classes is estimated and the frame sizes are corrected. The algorithm performs analysis from several thousand to several tens of thousands of parts of the image, and the accuracy and speed are comparable to YOLO.

When detecting and recognizing objects in images, it is necessary to apply algorithms that divide the input image into a set of images suitable in size for analysis in a convolutional NS. To save time and resources, it is also necessary to minimize the amount of data being analyzed. An algorithm that satisfies the above requirements can be implemented in the following ways:

- 1) the sequential splitting of the input image into frames, the size of which is determined by the architecture of the convolutional NS;
- 2) the use of the algorithms R-CNN, Fast and Faster R-CNN, Mask R-CNN, YOLO, as well as SSD and the like.

Both of the presented methods require serious expenditures of time and computing resources. In the first case, several thousand frames are submitted to the convolutional NS for analysis [3, 8, 9], and in the other, the image is analyzed as a whole with enumeration of the boundaries of regions of interest even in areas of no interest [10–14].

Thus, there is a need for the use of algorithms that reduce the amount of data analyzed, and thereby increase the performance of search and recognition systems based on NS.

3. The aim and objectives of research

The aim of research is to create an algorithm for detecting objects in images, accelerated by searching for zones of interest, with further recognition of the object in a convolutional NS.

To achieve the aim, the following objectives are set:

- highlight the objects present in the image;
- perform and speed up the search for centers of objects;
- check the algorithm on real input data.

4. The algorithm for selecting objects located on the underlying surface

This scientific work is based on previous studies [17, 18] and is a continuation of them. They investigated the possibility and prospect of using neural networks and self-organizing maps of Kohonen (Self-organizing map, SOM) to determine the centers of objects in images. In addition, let's

compare the performance of these two types of neural networks and described earlier versions of the algorithm for searching for zones of interest in images. Detailed information on the Kohonen network and SOM is presented in [15, 16].

When analyzing radar and satellite images between the target object and the underlying surface, there is always an interface, which is formed due to the difference in brightness or light levels, because the reflectivity of objects of interest, for example, airplanes, is usually higher than that of the underlying surface. Also, in most cases, the bodies of objects such as cars or airplanes cast a shadow, which also forms the interface. When to select these boundaries around the objects with dots, the center of the cluster will be close to the center of the object.

Early versions of the algorithm for selecting objects located on the underlying surface are given in [17, 18]. In addition, it is proposed to introduce a decrease in resolution obtained after preliminary processing of the image to accelerate the determination of cluster centers by the Kohonen network and additional filtering from noise.

The algorithm for selecting objects is as follows:

- 1) color image is input;
- 2) input image is converted to shades of gray;
- 3) image is contrasted;
- 4) Sobel operator is applied to the image – a differential operator that calculates the approximate value of the brightness gradient;
- 5) conversion to a binary image, using clipping by the brightness threshold. The boundaries of the section take the value 1, all the rest – 0;
- 6) removal of small objects;
- 7) removal of the boundaries of the desired object of the selected parts of the image that are not characteristic of geometric shapes;
- 8) filling inside closed borders;
- 9) reducing the resolution of the resulting image.

5. Application of the Kohonen network to identify centers of interest zones

In earlier published works [17, 18], the results of determining the centers of zones of interest on images by the network and the Kohonen map were compared, and an algorithm for selecting objects was presented. According to the results of the research, the Kohonen network proved to be the most suitable for solving this problem, since certain centers of the clusters in it are independent of each other, but inferior to SOM in terms of the processing speed of large images. In order to speed up processing by the Kohonen network, the size of the image obtained after searching for objects can be reduced by a number of times. This will not only speed up the search for cluster centers by the network, but also additionally clear the image of noise.

The Kohonen network consists of a number of parallel acting linear elements having the same number of inputs and receive the same vector of input signals $\mathbf{x}=[x_1, x_2, \dots, x_n]$ at their inputs. At the output of the j -th linear element, let's obtain a signal

$$y_j = \omega_{j0} + \sum_{i=1}^m w_{ji} x_i, \quad (1)$$

where w_{ji} – synaptic weight coefficient of the i -th input of the j -th neuron, i – input number, j – the number of the neuron, ω_{j0} – threshold coefficient.

After passing through the layer of linear elements, the signals are processed according to the rule “The winner receives everything”, when the maximum signal is equal to one, the rest to zero. All vectors from a certain region of the input space are replaced by one vector, which is their common, nearest neighbor in Euclidean distance, which is calculated as follows:

$$d_E(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} = \|\mathbf{x} - \mathbf{y}\|. \quad (2)$$

The Euclidean distance between the vectors x and y is the Euclidean norm of the difference between the vectors, or the length of the segment connecting the points x and y .

Considering the properties of the Kohonen network described above, it is obvious that when an image with only the boundaries of objects is input, the network will place the output vectors in the center of the objects. This is due to the fact that the centers of the objects correspond to the minimization points of the Euclidean distance between groups of input vectors.

6. Description of the algorithm for accelerated detection and recognition of objects in images

The algorithm for accelerated detection and recognition of objects in images is shown in Fig. 1.

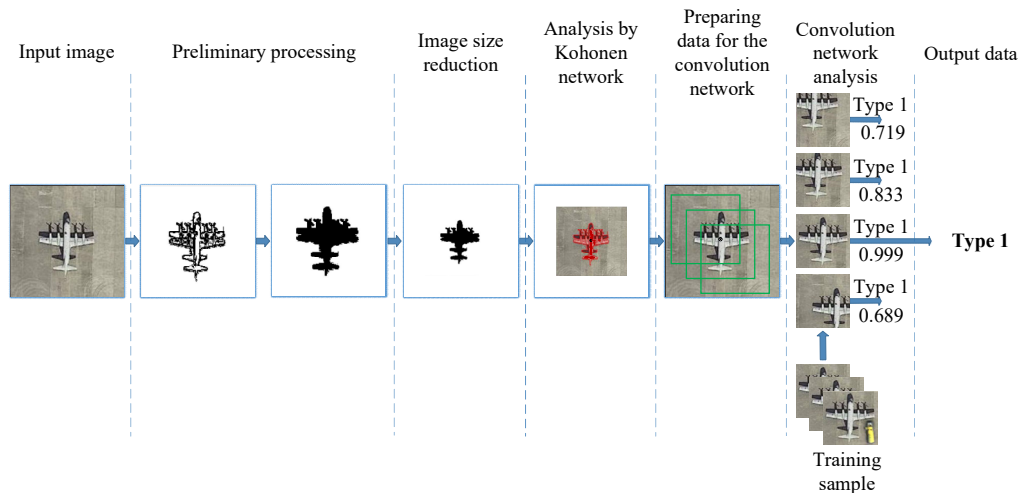


Fig. 1. The scheme of the algorithm for accelerated detection and recognition of objects in images

The algorithm consists of the following operations:

1. A radar or satellite image with a known scale is fed to the input.
2. An algorithm for selecting objects is applied to the input image.
3. Reducing the size of the resulting image.
4. Search for centers of objects by the Kohonen network.

5. Formation around the found centers of an expanded zone of interest, along which a window of a certain size is run, which is selected based on the dimensions of the largest object currently being detected in accordance with the image scale. The portions of the input image highlighted by the “window” must contain the center point of the cluster and overlap the vicinity of the zone of interest. If the points found by the Kohonen network are located close to the scale of the desired object, the zone of interest expands to cover the common area for these nodes, instead of analyzing two separate areas of interest.

6. Analysis in the convolution network. The obtained parts of the input image are submitted for analysis to the convolution network, where the presence of the object and its type are determined. The structure of the convolutional neural network is shown in Fig. 2.

A convolutional NS with two layers of 2D convolution, using the BatchNormalization, Max-Pooling, Dropout, and RELULayer layers in front of the fully connected layer and the classification layer, was trained on images of various types of aircraft. Scaling all the “windows” to a single size made it possible to avoid variations in the overall dimensions of an aircraft of the same type in the training set, which significantly reduced its required size.

This algorithm can be used to search and recognize in the input image not only various types of aircraft, but also other objects of interest, for example, in the detection of armored vehicles and ships or in the analysis of visual data of unmanned vehicles. For this, it is necessary to set the maximum and minimum overall dimensions of the objects of interest and the image scale, compose a training sample for these objects and train the convolutional NS.

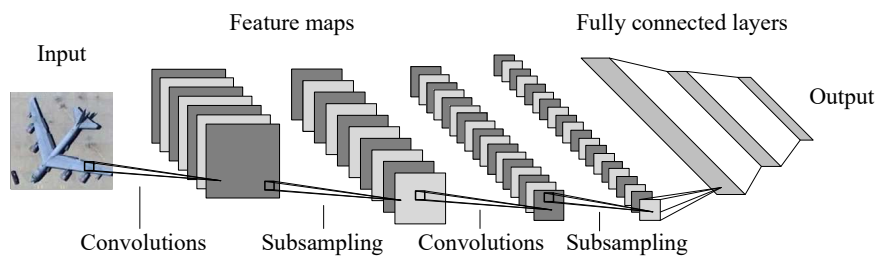


Fig. 2. The structure of the convolutional neural network

7. Demonstration of the algorithm on real input data

To search for several objects in the input image, it is necessary to take into account the image scale, as well as the maximum and minimum overall dimensions of the objects of interest. This is necessary to avoid missing objects. It is necessary that the determined number of cluster centers in the image coincides with the number of smallest objects that can fit in the input image. Fig. 3 shows an example of determining areas of interest in a real color satellite image from Google Maps measuring 671×493 pixels with several types of aircraft, various underlying surfaces and structures.

After applying the object extraction algorithm to the input image, it was reduced several times and submitted for analysis to the Kohonen network. At the output of the neural network, 24 center of the clusters should be determined, since it is precisely so many aircrafts of the minimum size that can fit in this image, based on its scale. If the found points are located relatively close to each other, then it is possible to combine them into one located in the center between them. The same applies to “windows” that overlap with each other by more than 90 % of their area. It is also possible to exclude from the analysis points located in the immediate vicinity of the image edge, since they can’t be the center of the aircraft located in the whole image.

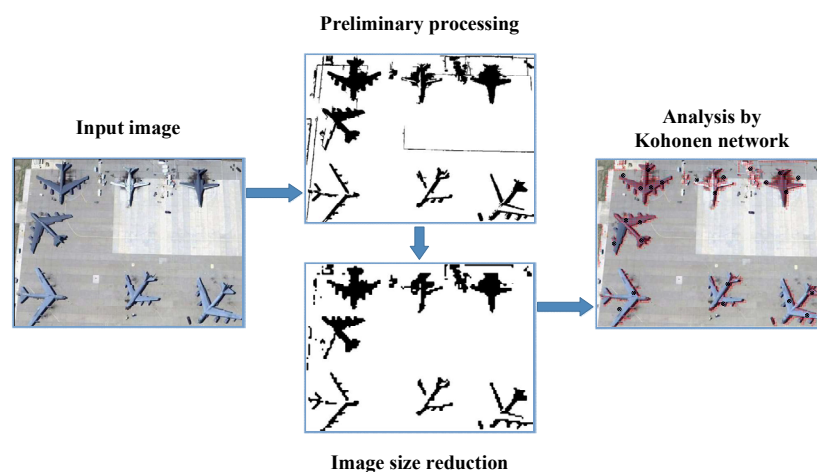


Fig. 3. Input image processing

As can be seen from Fig. 3, the Kohonen network determined the centers of the clusters in such a way that all aircraft will be highlighted with “windows” and will be further analyzed in the convolution network, without having to locate points in empty areas, as SOM often does, where the location of the found points depends on the position of neighboring ones. In this case, after the “window” marks out zones of interest around the centers of the clusters, about 80 cut frames the size of the maximum aircraft are sent for further analysis. This is 1000 times smaller than using sequential splitting of the input image into frames or 18–125 times smaller than using algorithms similar to RCNN, YOLO and SSD. Also, the application of this algorithm significantly reduces the required number of training images of the aircraft for the convolution network, because the size of the “window” is tied to the image scale and is equal to the size of the largest detectable aircraft, since there is no need to vary the size of the aircraft in training images.

Table 1 shows the speed of the developed algorithm for searching and recognizing objects depending on the resolution of the input image and the number of possible objects of minimum size. The calculations were performed in a program of our own design without the use of a GPU (graphics processor), acceleration and optimization algorithms on a personal computer in the configuration of Core i7-8700, 32 GB of RAM, NVIDIA GeForce GTX 1080.

Table 1
The speed of the search algorithm and object recognition

The size of the input image, pixels	Number of search objects, items	The speed of the algorithm, s
120×230	3	0,05
200×400	10	0,1
671×493	24	0,3

Compared to previously published results [17, 18], the speed of the algorithm has increased by more than 10 times. And subject to the use of the GPU, multithreading and optimization of calculations, acceleration of more than 30 times is possible. This is extremely relevant in the context of an ever-increasing number of satellites and an increase in the resolution of the images they receive, reducing the processing of high-definition data from a few minutes to seconds.

8. Discussion of the results of the created algorithm for accelerated detection and recognition of objects in images

The existing algorithms for searching and recognizing objects are divided into several thousand or even hundreds of thousands of parts for the input image and submit them to the convolution network for analysis. The algorithm proposed in this article allows, due to preliminary allocation of zones of interest, to reduce the number of analyzed parts of the input image. Also, by reducing the training sample for the convolutional network, the detection time of objects is significantly reduced and the recognition accuracy is increased. The size of the “window” of the scanning zone of interest is selected based on the image scale and dimensions of the largest detectable object, and the number of possible cluster centers is determined by the number of smallest objects that can fit on the image. Due to the selection of the centers of objects by the Kohonen network, when compiling the training sample for the convolutional NS, it becomes possible to locate the desired object in the center of the images, eliminating the variation in the shift of the object. Also, sizing the “window” sizes to an object of maximum size eliminates scale variation in the training set, which significantly improves recognition accuracy.

The disadvantage of the created algorithm is the drop in recognition speed when analyzing images in which objects occupy most of the space. In this case, it is possible that the whole image will become a zone of interest, since the found centers of the clusters will be evenly distributed throughout the image. But even in this case, the parts of the image analyzed by the convolutional NS will be 400–600 times smaller than when using sequential splitting or 5–15 times smaller than when using algorithms similar to RCNN, YOLO and SSD.

In the future, it is planned to embed another convolutional neural network in the algorithm, which would make it possible to determine the angle of rotation of the aircraft around the axis in order to eliminate this variation when recognizing the type of aircraft in the final convolutional NS.

9. Conclusions

To select the objects present in the input image, an algorithm based on the use of the Sobel operator is used.

To search for centers of objects, the Kohonen network was used, since it very rarely locates the found centers between objects, in contrast to SOM, where there is a relationship between the locations of neighboring nodes. By reducing the resolution of the image obtained after preliminary processing and sent to the Kohonen network, it was possible to accelerate the search for cluster centers by more than 10 times.

The developed algorithm was tested on real satellite images and its application allows to reduce the amount of data sent to the convolution network by 18–125 times, which, accordingly, speeds up the search and recognition. Application of the algorithm reduces the required number of training images for the final convolutional NS, because the size of the “window” corresponds to the size of the largest detectable object. This fact, coupled with the centering of the object on the training images, allows to accelerate the training of convolutional NS by more than 5 times and increase the recognition accuracy by at least 10 %, as well as halve the required minimum number of layers and neurons of the network, thereby accelerating data analysis.

References

- [1] Simard, P. Y., Steinkraus, D., Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. doi: <https://doi.org/10.1109/icdar.2003.1227801>
- [2] Agayan, K. Yu., Hanzhin, V. G. (2018). Neyronnaya set' s arhitekturoy Kohonena dlya raspoznavaniya izobrazheniy. Sb. trudov IX-oy Evraziyskoy nauchno-prakticheskoy konferentsii «Prochnost' neodnorodnyh struktur» - PROST 2018. Moscow: OOO «Studio-Print», 153.
- [3] Gerasimova, N. I., Verkhoturova, A. E. (2014). Earch of the image fragment with application of Kohonen neural network. Informatsionnye tekhnologii v nauke, upravlenii, sotsial'noy sfere i meditsine: sbornik nauchnyh trudov Mezhdunarodnoy konferentsii s mezhdunarodnym uchastiem. Ch. 1. Tomsk: Izd-vo TPU, 68–70.
- [4] Soldatova, O. P., Chayka, P. D. (2015). Efficiency analysis of solution of classification using hybrid kohonen neural networks. Izvestiya Samarskogo nauchnogo tsentra Rossiyskoy akademii nauk, 17 (2 (5)), 1147–1152.
- [5] Kajan, S., Sekaj, I., Lajtman, M. (2011). Cluster Analysis Applications in Matlab Using Kohonen. Network learning, 4 (6), 7.
- [6] Le Cun, Y., Bengio, Y. (1995). Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, 3361 (10).
- [7] LeCun, Y., Kavukcuoglu, K., Farabet, C. (2010). Convolutional networks and applications in vision. Proceedings of 2010 IEEE International Symposium on Circuits and Systems. doi: <https://doi.org/10.1109/iscas.2010.5537907>
- [8] Girshick, R., Donahue, J., Darrell, T., Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition. doi: <https://doi.org/10.1109/cvpr.2014.81>
- [9] Girshick, R., Donahue, J., Darrell, T., Malik, J. (2016). Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38 (1), 142–158. doi: <https://doi.org/10.1109/tpami.2015.2437384>
- [10] Girshick, R. (2015). Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV). doi: <https://doi.org/10.1109/iccv.2015.169>
- [11] Ren, S., He, K., Girshick, R., Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39 (6), 1137–1149. doi: <https://doi.org/10.1109/tpami.2016.2577031>
- [12] He, K., Gkioxari, G., Dollar, P., Girshick, R. (2017). Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV). doi: <https://doi.org/10.1109/iccv.2017.322>
- [13] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: <https://doi.org/10.1109/cvpr.2016.91>
- [14] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. Lecture Notes in Computer Science, 21–37. doi: https://doi.org/10.1007/978-3-319-46448-0_2
- [15] Haykin, S. (2008). Neural Networks: A Comprehensive Foundation. Izdatel'skiy dom Vil'yams, 1104.
- [16] Kohonen, T. (2001). Self-organizing maps. Vol. 30. Springer Science & Business Media, 501. doi: <https://doi.org/10.1007/978-3-642-56927-2>
- [17] Skuratov, V., Kuzmin, K., Nelin, I., Sedankin, M. (2019). Application of Kohonen neural networks to search for regions of interest in the detection and recognition of objects. Eastern-European Journal of Enterprise Technologies, 3 (9 (99)), 41–48. doi: <https://doi.org/10.15587/1729-4061.2019.166887>
- [18] Skuratov, V., Kuzmin, K., Nelin, I., Sedankin, M. (2020). Application of kohonen self-organizing map to search for region of interest in the detection of objects. EUREKA: Physics and Engineering, 1, 62–69. doi: <https://doi.org/10.21303/2461-4262.2020.001133>

Received date 20.05.2020

Accepted date 10.07.2020

Published date 31.07.2020

© The Author(s) 2020

This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0>).