

APPLICATION OF KOHONEN SELF-ORGANIZING MAP TO SEARCH FOR REGION OF INTEREST IN THE DETECTION OF OBJECTS

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
Fundamentals of Radio Engineering Department
National Research University "Moscow Power Engineering Institute"
14 Krasnokazarmennaya str., Moscow, Russian Federation, 111250
msedankin@yandex.ru*

Abstract

Today, there is a serious need to improve the performance of algorithms for detecting objects in images. This process can be accelerated with the help of preliminary processing, having found areas of interest on the images where the probability of object detection is high. To this end, it is proposed to use the algorithm for distinguishing the boundaries of objects using the Sobel operator and Kohonen self-organizing maps, described in this paper and shown by the example of determining zones of interest when searching and recognizing objects in satellite images. The presented algorithm allows 15–100 times reduction in the amount of data arriving at the convolutional neural network, which provides the final recognition. Also, the algorithm can significantly reduce the number of training images, since the size of the parts of the input image supplied to the convolution network is tied to the image scale and equal to the size of the largest recognizable object, and the object is centered in the frame. This allows to accelerate network learning by more than 5 times and increase recognition accuracy by at least 10 %, as well as halve the required minimum number of layers and neurons of the convolutional network, thereby increasing its speed.

Keywords: pattern recognition, Kohonen self-organizing maps, search and recognition of objects in the image, satellite and radar images, region of interest.

DOI: 10.21303/2461-4262.2020.001133

1. Introduction

Currently, the low speed of detecting objects in images is one of the main problems of modern systems for processing various visual data. An increase in the speed of search and recognition of objects can lead to a significant increase in the performance of systems for analyzing satellite and radar images, medical images, data from robotic and military systems, unmanned vehicles, etc., having both technological and serious economic effects.

One of the ways to increase the accuracy and speed of recognition algorithms is the use of neural networks (NN), and the most modern type of NN used in pattern recognition is convolutional neural network (CNN) [1]. For the high-quality operation of this type of network, it is necessary that the size of the recognized object and the sizes of the objects in the training set be comparable. If, in addition to the various types of desired objects, there is also a variation in the scale of the object in the image, the creation of a training sample and training of such a network will take considerable time and computational resources, the number of errors in recognition will increase, and the speed will be low. The presented algorithm allows to reduce the amount of data analyzed when searching and recognizing objects in the image by 15–100 times, increase accuracy and save computing resources. This is achieved through the use of Kohonen maps to determine areas of interest in the input image, in which there is a high probability of finding the desired object.

2. Literature review and problem statement

Today, there are a number of works where the use of the self-organizing map of Kohonen (Self-organizing map, SOM) for image recognition has been investigated. Some authors use Kohonen maps to prepare input data with subsequent analysis by other NNs, while others use it to directly recognize objects in images. In [2], image segmentation by Kohonen maps is first applied, and then analysis in a hybrid NN. Kohonen maps were used to reduce the training set for hybrid NN and to speed up the analysis. The disadvantage of [2] is that the input image is classified as a whole, without the ability to search for an object. Therefore, when zooming in on the desired fragment in the input image, recognition will not be performed. In [3], Kohonen maps were used to solve clustering problems with a large number of objects and to distinguish among them those that have unusual characteristics.

Articles [4–8] describe the evolution of the R-CNN (Region-based Convolutional Network) algorithm for searching and recognizing objects in images. In [4, 5], the initial version of the R-CNN algorithm is described. The algorithm extracts about 2000 regions on the input image, each of which is scaled using the affine transformation, and fed to the input of the convolution network, which extracts the feature vector (map). The article [5] describes the linear regression training to refine the coordinates of the object window. R-CNN has several drawbacks, mainly due to the high time spent on training the network, as well as on direct image processing by the algorithm, so that processing one image takes about 47 seconds. The following describes the development of this algorithm (Fast R-CNN [6] and Faster R-CNN [7]) up to Mask R-CNN [8] in which the ability to predict the position of the mask covering the found object is added.

The article [9] describes the YOLO algorithm (You Only Look Once), which allows searching and recognizing objects in images 103 times faster than R-CNN and 102 times faster than Fast R-CNN, but with lower accuracy. This algorithm superimposes a grid on the input image and divides it into cells. Around each cell, the algorithm determines the bounding box of the zone of the possible location of objects with an assessment of the accuracy of detection and the probability of belonging to classes. Then, the accuracy estimate for each zone is multiplied by the probability of the class and the final value of the probability of detection is obtained. In [10], the SSD: Single Shot MultiBox Detector algorithm is presented, which is comparable in accuracy and speed to YOLO. The algorithm overlaps the entire area of the input image with bounding frames, the size of which varies within the set limits, allowing the detection and recognition of objects of various sizes. Both algorithms perform analysis from several thousand to several tens of thousands of parts of the input image.

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 NN. To save time and resources, the applied algorithm should reduce the amount of data needed for further analysis. Such an algorithm can be implemented in the following ways:

- 1) sequentially splitting the input image into frames of the required size with a frame shift of a certain number of pixels relative to the previous one;
- 2) using the algorithms R-CNN, Fast and Faster R-CNN, Mask R-CNN, YOLO, as well as SSD.

Both of the presented methods require significant time and computational resources. In one case, several thousand frames are submitted to the subsequent convolutional NN for analysis, into which each input image is divided [4, 5], and in the other case, the image is analyzed as a whole, with enumeration of the possible boundaries of regions of interest and without preliminary screening of areas of no interest [6–10].

Thus, there is an obvious need for algorithms that can reduce the amount of data being analyzed, and thereby increase the speed of search and recognition systems for objects in images.

3. The aim and objectives of research

Interesting objects and the possibility of recognizing its type in a convolutional NN.

To achieve the aim, the following tasks are set:

- highlight the boundaries of the objects present in the image;
- perform a search for centers of objects;
- check the selected algorithms on real input data.

4. The algorithm for distinguishing the boundaries of objects located on the underlying surface

This scientific work is based and is a continuation of the previous study [11], in which the possibility and prospect of using neural networks and Kohonen maps to determine the centers of objects in images are investigated, and the performance of these two types of neural networks is compared.

Before transferring the input image for analysis to the convolutional NN, it is necessary to carry out preliminary processing.

Between the object in the image 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, for example, airplanes is usually higher than that of the underlying surface. Also, in most cases, the bodies of objects such as tanks or planes cast a shadow on the underlying surface, which also forms the interface. If to select these boundaries around the objects with dots in the input image, then the center of the cluster will actually coincide with the center of the object.

The algorithm for extracting the boundaries of objects located on the underlying surface, given in [11], has been finalized and now looks like this:

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

5. Methodology for determining centers of interest

To increase the speed of analysis and reduce the necessary computing resources when objects of interest are detected in the input image, it is proposed to identify the areas of their possible location with further sequential analysis of the found areas in the convolutional NN.

To highlight the centers of zones of interest in the input image, it is necessary to train NN without a teacher, i. e. lack of training sample, as processing of each input image is carried out independently of the previous ones. The main types of Kohonen networks that use teaching without a teacher are [12, 13]:

- Kohonen network for vector quantization of signals;
- self-organizing maps of Kohonen.

In this paper, let's describe the SOM application to solve the problem of determining the centers of zones of interest, which is a development [11].

The main goal of SOM is to convert input vectors of arbitrary dimension into a one- or two-dimensional discrete map with a topologically ordered shape [13]. Kohonen maps are based on competitive learning. The neurons of the output layer compete for the right to activate, as a result of which only one output neuron, the winner neuron, is active. One way to organize this kind of competition between neurons is to use negative feedbacks between them. In the general case, neurons in SOM are located at the nodes of a two-dimensional grid with rectangular or hexagonal cells (**Fig. 1**). The magnitude of the interaction between neurons in the network is determined by the distance r_n between them. The distance between the individual neurons is more consistent with the Euclidean distance for the hexagonal grid. The more neurons in the grid, the higher the degree of detail of the SOM result.

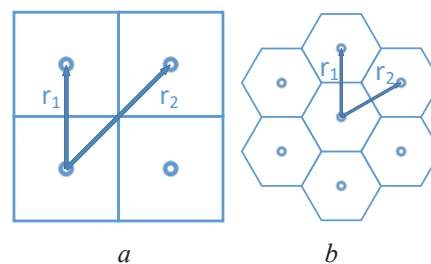


Fig. 1. The initial arrangement of neurons in the nodes of a two-dimensional grid with cells:
a – quadrangular; *b* – hexagonal

The location of the winning neurons determined during the competitive process is ordered in relation to each other, then ask the ones in the grid a significant coordinate system in which the coordinates of the neurons are an indicator of the statistical features contained in the input images.

The work of the SOM algorithm usually begins with the initialization of the synaptic network weights and, after the correct initialization, three main processes are launched to form the map: competition, cooperation and synaptic adaptation.

The principle of operation of the SOM algorithm is considered in more detail in [11].

6. Description of the area of interest search algorithm

A satellite or radar image of the underlying surface with a known scale is fed to the input. Next, pre-processing is carried out, which is carried out in several stages:

1. Highlighting the boundaries of objects present in the image. For this, the algorithm described in paragraph 4 is used, based on the use of the Sobel operator.
2. Search for centers of objects. SOM is used for this.
3. Identification of areas of interest. Around the found centers, an expanded zone of interest is formed, within which parts of the original image are extracted with a certain “window”. The selected parts of the image should contain the point of the center of the cluster and overlap the vicinity of the zone of interest. The dimensions of the “window” are selected based on the given dimensions of the largest detectable object and change when the image scale is changed.
4. In the presence of closely located relative to the scale of the desired object network nodes, the zone of interest expands so as to overlap the common area for these nodes, instead of looking for the region of interest for each node separately.
5. 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 on and its type is determined.

The algorithm for the search for zones of interest is shown in **Fig. 2**.

This algorithm can be used to search and recognize in the input image not only various types of armored vehicles and aircraft, but also other objects of interest, for example, when analyzing the visual data of unmanned vehicles or when testing robotic systems. For this, it is necessary

to set the maximum and minimum sizes of objects of interest for the algorithm to search for zones of interest, compose a training sample for these objects and train the convolutional NN.

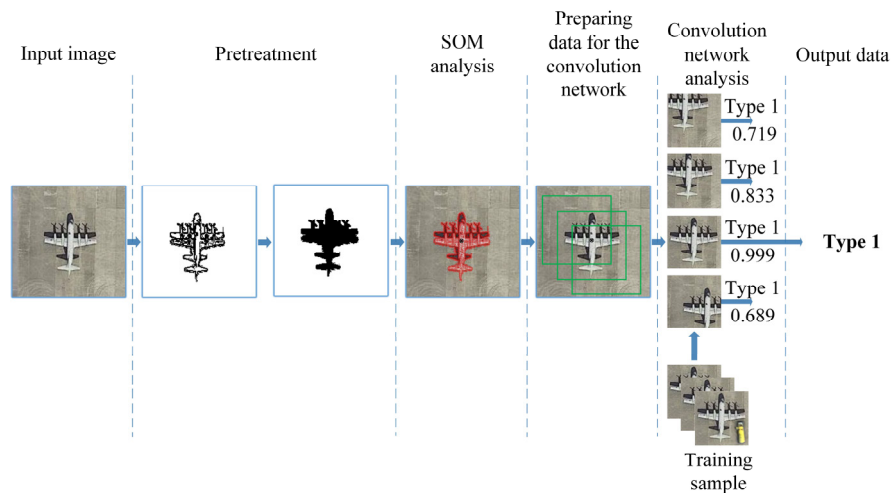


Fig. 2. The algorithm of the search for zones of interest and determining the type of object

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 real scale of the image and their relative dimensions. In order 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 on the input image with a known scale. **Fig. 3** shows an example of the algorithm for determining zones of interest in a real color satellite image measuring 671×493 pixels, which contains several types of aircraft, various underlying surfaces and structures. After preliminary processing, the input image is submitted for analysis to SOM with the initial arrangement of neurons at the nodes of a two-dimensional grid with hexagonal cells. SOM needs about 150 learning eras. Based on the scale of the input image, the number of smallest aircraft that can fit on the image is 6×4. Therefore, at the output of the neural network, 24 cluster centers must be defined. If the centers of the clusters are located relatively close to each other, then their combination into one located in the center between them is possible. The same applies to “windows” around neighboring cluster centers if they overlap more than 90 % of the area. It is also possible to exclude from the analysis the centers located in the immediate vicinity of the edge of the image, since they can’t be the center of the aircraft, which the convolutional network could recognize during further processing.

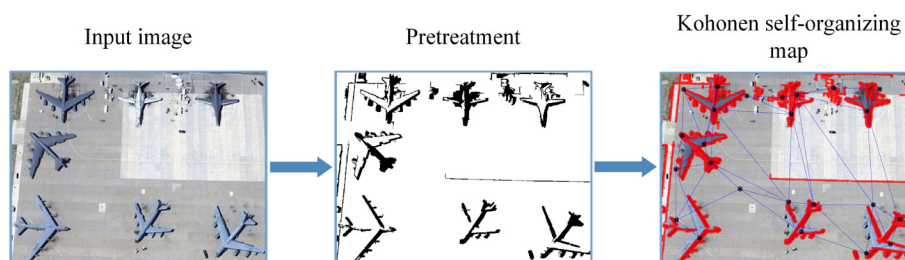


Fig. 3. Input image processing with SOM analysis

As can be seen from **Fig. 3**, the Kohonen map identified the centers of the clusters in such a way that all aircraft will be highlighted with a “window” and will be used for further analysis in the convolution network. The minus of the map is the dependence of the position of the centers of the clusters on neighboring neurons, which leads to the presence of centers in areas with no differences in brightness. In this case, after the “window” identifies zones of interest around the centers of the clusters, about 100 cut frames the size of the maximum aircraft are sent for

further analysis. This is 1000 times less than when using sequential splitting of the input image into frames or 15–100 times less than when 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. This means that there is no need to vary the size of the aircraft in training images.

Table 1 shows the performance of the Kohonen map 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 with the parameters specified in paragraph 4.

Table 1
Kohonen map performance

The size of the input image, pix.	Number of search objects	SOM processing, sec.
120×230	3	0,1
200×400	10	0,3
671×493	24	3,3

As can be seen from the **Table 1**, the image processing speed of the Kohonen map has a non-linear dependence on the image size and the number of search objects, so the size of the input image should be selected based on the required speed of the algorithm.

According to the results of [11], the main drawback of the SOM algorithm was the frequent location of network nodes between objects, and not inside them. This drawback is eliminated using the algorithm for initializing the network weights according to random examples and changing the communication function between neurons in such a way that with increasing distance between them, the interaction quickly decreased. This allows to reduce the number of network nodes outside the facilities by about 4.7 times.

8. Discussion of the results of the created algorithm for the search for zones of interest on the underlying surface

Existing algorithms for searching and recognizing objects select several thousand or even hundreds of thousands of parts on the input image and submit them to the convolution network for analysis. The algorithm proposed in this article allows one to reduce the number of analyzed portions of the input image and to reduce the training sample for the convolution network, which can significantly reduce the time for searching and recognizing objects. This is achieved through a preliminary search for areas of interest in the image with further analysis of only the selected areas. The size of the “window” scanning certain areas of interest is determined based on the image scale and the given dimensions of the largest detectable object. Highlighting the boundaries of objects and applying SOM allow to determine the centers of objects in the image and create zones of interest around them. These solutions make it possible to eliminate variations in scale and center the desired object, due to which the training sample size of the terminal convolution network is reduced several times, its training is accelerated, and recognition accuracy is increased.

The disadvantage of the created algorithm is a significant increase in the recognition speed only when analyzing images on which the boundaries of objects do not occupy most of the space. For example, in satellite and radar images, when the desired objects are located on a homogeneous underlying surface of great length. When analyzing images occupied by small objects, the algorithm will determine the entire image as the zone of interest, while still reducing the number of analyzed portions of the input image. The analyzed parts will be 300–500 times less than when using sequential partitioning or 3–10 times less than when using algorithms similar to RCNN, YOLO and SSD. In the future, it is planned to identify unique features of the shape of the boundaries of the de-

sired objects, which will help to distinguish these objects against the background of the underlying surface and the boundaries of objects of no interest.

The pre-processing algorithm shows a high speed of extracting the boundaries of the objects of the input image. The search for cluster centers using a Kohonen self-organizing map allows to process images with sizes up to 500×500 pixels at high speed. and the number of search objects about 40 pcs. In the future, it is planned to use Kohonen map acceleration and optimization algorithms for GPU calculations, as well as consider other options for searching for cluster centers, for example, the Kohonen neural network.

Compared with the previous work [11], the following is done:

- the algorithm for highlighting the boundaries of objects by performing image contrasting, filling the space inside closed borders, and filtering selected objects in the image which geometric shape does not match the shape of the boundaries of the desired objects has been improved;
- the algorithm for determining the zone of interest for closely located, relative to the scale of the desired object, network nodes has been changed. Now the zone of interest is expanded in such a way as to cover the common area for these nodes, instead of looking for the region of interest for each node separately;
- improving the initialization algorithm of the initial network weights and the function of reducing the connections between neurons with increasing distance between them, allows to significantly reduce the number of network nodes that are between objects, and not inside them.

In further work, it is planned to improve the presented algorithm due to a better allocation of the boundaries of objects and the search for their centers. It is also planned to create and optimize a convolutional neural network for object recognition, which will create a complete system for searching and recognizing objects on radar or satellite images of the underlying surface.

9. Conclusions

1. To highlight the boundaries of objects present in the image, an algorithm based on the use of the Sobel operator is used. The algorithm has high speed, highlighting the boundaries of the objects of the input image in 0.008–0.03 sec.

2. SOM is used to search for centers of objects, as this is one of the fastest algorithms for determining the centers of input data clusters.

3. The developed algorithm is tested on real satellite images. Its application allows to reduce the amount of data analyzed by the convolution network by 15–100 times, which accordingly reduces the time of searching and recognition of necessary objects. Also, the use of this algorithm reduces the required number of training images for the convolution network, since the size of the “window” is related to the image scale and corresponds to the size of the largest detected object. This fact and the centering of the object on training images can accelerate network learning by more than 5 times and increase recognition accuracy by at least 10 %, as well as halve the required minimum number of layers and neurons of the convolutional network, thereby increasing its speed.

References

- [1] Simard, P. Y., Steinkraus, D., Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. Proceedings of the Seventh International Conference on Document Analysis and Recognition. doi: <https://doi.org/10.1109/icdar.2003.1227801>
- [2] Novikova, N. M., Dudenkov, V. M. (2015). Modelirovanie neyronnoy seti dlya raspoznavaniya izobrazheniy na osnove gibridnoy seti i samoorganizuyuschihsy kart Kohonena. Aspirant, 2, 31–34.
- [3] Narushev, I. R. (2018). Neural network on the basis of the self-organizing kochonen card as a means of detecting anomalous behavior. Ohrana, bezopasnost', svyaz', 2 (3 (3)), 194–197.
- [4] 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>
- [5] 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>

- [6] Girshick, R. (2015). Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV). doi: <https://doi.org/10.1109/iccv.2015.169>
- [7] Ren, S. et. al. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 91–99.
- [8] 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>
- [9] 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>
- [10] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. *Computer Vision – ECCV 2016*, 21–37. doi: https://doi.org/10.1007/978-3-319-46448-0_2
- [11] 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>
- [12] Haykin, S. (2008). *Neyronnye seti: polniy kurs*. Moscow: Izdatel'skiy dom Vil'yams.
- [13] Kohonen, T. (2001). *Self-organizing maps*. Vol. 30. Springer Science & Business Media, 502. doi: <https://doi.org/10.1007/978-3-642-56927-2>

Received date 25.12.2019

Accepted date 24.01.2020

Published date 31.01.2020

© The Author(s) 2020

This is an open access article under the CC BY license

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