APPLICATION OF ALGORITHMS FOR OBJECT RECOGNITION BASED ON DEEP CONVOLUTIONAL NEURAL NETWORKS FOR VISUAL NAVIGATION OF A MOBILE ROBOT*

Z.S. Pershina
NSTU, Novosibirsk, Russia
pershina@tiger.cs.nstu.ru

S.Y. Kazdorf
NSTU, Novosibirsk, Russia
kazdorf.2014@stud.nstu.ru

V.A. Abrosimov
NSTU, Novosibirsk, Russia
abrosimov.2014@stud.nstu.ru

***Abstract*—In this paper, we present the results of an analysis of state-of-the-art approaches in the field of object recognition in an image based on deep convolutional neural networks. The purpose of the review is to assess the possibility of applying these approaches in solving problems of visual navigation a mobile robot.**

Keywords — object recognition; convolutional neural networks; visual navigation.

I. INTRODUCTION

One of the current trends in the development of ground mobile robotics is the improvement of the autonomous control of mobile robots. A promising way of such control is based on using of visual information about the external environment. This approach is called visual navigation.

With this method of navigation, the robot, in addition to obtaining information about passable sites, it is necessary to recognize objects of the observed scene, and then to evaluate their location, orientation and size. In this regard, in this paper, the application of algorithms for object recognition on the image of the observed scene on the basis of deep convolutional neural networks is considered.

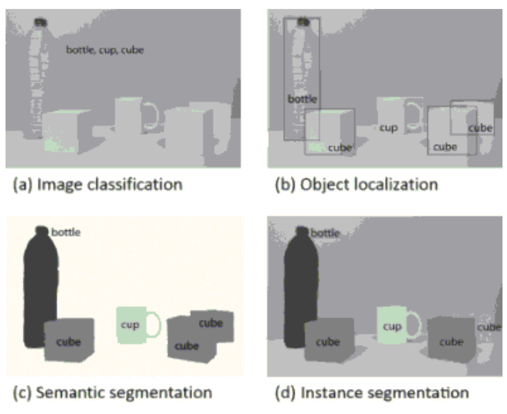

Fig. 1. Classification of object recognition task [12].

The reported study was funded by RFBR according to the research project №18-58-76003.

At the moment, there are four classes of tasks in the area of object recognition:

- image classification (Fig. 1a);
- object detection (Fig. 1b);
- semantic segmentation (Fig. 1c);
- instance segmentation (Fig. 1d).

II. IMAGE CLASSIFICATION

The task of classification involves determining which class the image belongs to, but the task of determining the position of objects on the image is not solved (Fig. 1a). And if there are signs of several classes in one image, the accuracy of the classification is significantly deteriorating.

The greatest success in the field of object classification in the image was achieved by solutions based on deep convolutional neural networks (CNN), in particular, on such CNN architectures as AlexNet [1], VGG [2], GoogLeNet [3], ResNet [4] and others.

While solving navigation problems, this approach can be successfully applied when the mobile robot needs to recognize a single object in the presented image or determine the type of the observed scene. To perform search and localization of objects in the image, it is required to use either detection algorithms or semantic segmentation.

III. OBJECT DETECTION

The detection task allows you to find several objects and determine their position in one image. The result of detection is a set of rectangular regions on the image, which contain the recognized objects of the corresponding classes (Fig. 1b). The most famous works based on convolutional neural networks in this area are Faster R-CNN [5], SSD [6], YOLO [7], which are successfully used in the development of navigation systems for autonomous mobile robots both indoors and urban environment. However, when navigation systems require more accurate algorithms for planning a safe and optimal trajectory of a mobile robot, for example, when solving tasks of capturing and moving objects, it is required to determine the contours of objects in the image as precisely as possible. This task can be solved using the semantic segmentation of objects in the image.

IV. Semantic Segmentation

Recently, in the field of object recognition in the image, a new approach, called semantic segmentation, has emerged. The key works in this area are FCN [8], U-Net [9], SegNet [10], DeepMask [11].

The application of this approach provides per-pixel localization and classification of objects in the image. Semantic segmentation can be divided into two types: semantic segmentation (Fig. 1c) and instance segmentation (Fig. 1d). Semantic segmentation allows you to per-pixel localize objects and classify each pixel, but does not split instances of the same class, while an instance segmentation solves this problem and, in addition to assigning a class label to the pixel, assigns an object ID. Class label allows you to define the entire class, and ID – all objects of this class.

V. 3D object Recognition

When solving the task of visual navigation, one of the models of the external environment representation can be a cloud of 3D points. This model can be obtained on the basis of data from laser range finders, stereo cameras or RGB-D cameras. However, the use of laser range finders allows you to obtain metric information about the observed scene and allows you to extract lesser semantic information.

Within the approaches discussed above, it is assumed that a stereo camera or a RGB-D camera is used as a data source. The data is a color (RGB) image, a depth map, as well as data on the rotation and displacement of the camera relative to the starting point (RT matrix), are information on the basis of which the formation of a 3D cloud points is performed.

Thus, the results of detection or semantic segmentation of objects on 2D images become possible, namely, the pixel coordinates of the recognized objects are correlated with the coordinates of the 3D space of the observed scene. Thus, with the use of detection algorithms, the result of the 3D point cloud mapping is parallelograms describing objects of a certain class, in the case of a semantic segmentation, each 3D point is assigned a class label that defining the entire class, and for instance segmentation – each 3D point is assigned 2 labels: a class label and an instance ID of the class. The class label will allow you to define the entire class, and the ID will identify all objects in this class.

VI. Conclusion

In this work the analysis of modern approaches in the field of recognition of objects on the image, based on deep convolutional neural networks, with the purpose of their application in solving the task of navigation of a mobile robot is performed.

It should be noted that these solutions require large computational resources. Accordingly, there is a need to carry out research in this direction with a view to finding a solution with an optimal ratio of the number of classes, accuracy, speed, taking into account limitations on computational resources.

References

[1] Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks //Advances in neural information processing systems. – 2012. – C. 1097-1105

[2] Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition //arXiv preprint arXiv:1409.1556. – 2014.

[3] Szegedy C. et al. Going deeper with convolutions. – Cvpr, 2015.

[4] He K. et al. Deep residual learning for image recognition //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – C. 770-778.

[5] Ren S. et al. Faster r-cnn: Towards real-time object detection with region proposal networks //Advances in neural information processing systems. – 2015. – C. 91-99.

[6] Liu W. et al. Ssd: Single shot multibox detector //European conference on computer vision. – Springer, Cham, 2016. – C. 21-37.

[7] Redmon J. et al. You only look once: Unified, real-time object detection //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – C. 779-788.

[8] Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2015. – C. 3431-3440.

[9] Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation //International Conference on Medical image computing and computer-assisted intervention. – Springer, Cham, 2015. – C. 234-241.

[10] Badrinarayanan V., Kendall A., Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation //IEEE transactions on pattern analysis and machine intelligence. – 2017. – T. 39. – №. 12. – C. 2481-2495.

[11] Pinheiro P. O., Collobert R., Dollár P. Learning to segment object candidates //Advances in Neural Information Processing Systems. – 2015. – C. 1990-1998.

[12] Garcia-Garcia A. et al. A review on deep learning techniques applied to semantic segmentation //arXiv preprint arXiv:1704.06857. – 2017.