Object Detection in Images Using Deep Neural Networks for Agricultural Machinery

N V Andreyanov¹, A S Sytnik¹, M P Shleymovich¹

¹Department for Automated Systems for Information Processing and Control, Kazan National Research Technical University named after A.N.Tupolev-KAI, 10, K. Marks, Kazan, 420111, Russia

E-mail: assytnik@kai.ru

Abstract. The article deals with the creation of intelligent tractor driver support systems based on computer vision technologies for analyzing the direction of movement and detecting obstacles when performing specified operations, such as plowing, harrowing, weeding, and fertilizing. Electric power poles, trees, rocks, bird nests, animals, people, and field roads are identified as obstacles. The solution of functional problems in the system is based on the extraction of information from images using methods for detecting and recognizing objects in images. The analysis of existing approaches to solving the problems under consideration is carried out and it is shown that the use of deep neural networks is effective. The practical use of the methods based on the chosen approach is based on the performance of the computing system, the availability of sufficient training data and the optimality of the training method. It is shown that these factors are important when implementing an intelligent tractor driver support system.

1. Introduction

The current stage of economic development is determined by the processes of digital transformation [1, 2]. They cover almost all sectors, including the agro-industrial complex, for which the "smart agriculture" has been developed. One of the directions here is the "smart field" [3]. The main goal of the work in this direction is to provide special support for the work of drivers of agricultural vehicles - tractor drivers and combine harvesters. This is due to the fact that their shift can last up to fourteen hours, and their work itself is quite stressful and monotonous at the same time.

The implementation of the concept of "smart agriculture" in general and "smart field" in particular involves the use of digital technologies, such as high-performance computing technologies, geoinformation technologies, automatic and automated control technologies, network technologies, etc. Computer vision and machine learning technologies are also important here. These technologies are already beginning to be actively used in the practice of agriculture. For example, precision agriculture is based on image processing and analysis using computer vision methods and tools for monitoring, measuring and responding to crop variability [4]. The research here is aimed at creating decision support systems for the management of agricultural enterprises. Computer vision technologies are widely used for automated solution of tasks that are traditionally performed manually. At the same time, machine learning algorithms are used that allow you to quickly and accurately...
analyze huge amounts of data, providing the means for implementing computer vision applications in agriculture.

2. Intelligent transport systems for agricultural machinery

One of the modern developing scientific and technical directions is the creation of intelligent transport systems [5, 6]. Within the framework of this direction, automated complexes for vehicles are being developed that are able to provide: long-term operation, high performance, the specified accuracy and safe operation. Intelligent transport systems are of particular importance for urban and intercity communications. Therefore, they are actively developing within the framework of the "smart city" concept. In the field of intelligent transport, two directions can be distinguished. Within the first direction, systems of unmanned vehicles are considered, and within the second – support systems for drivers of manned vehicles.

The concept of "smart city" assumes a developed infrastructure that ensures the functioning of unmanned transport, which is not always possible to implement for agriculture. Therefore, fully automated unmanned agricultural vehicles are not yet widely used, despite the existing positive experience. This is primarily due to the fact that they are based on the use of satellite navigation technologies. However, the latter can not always provide a sufficiently powerful signal at the locations of equipment on the cultivated fields. Its instability cannot provide the necessary quality of agricultural machinery management. In addition, a large number of foreign objects may occur in the path of its movement. Stones, pillars, trees and other obstacles cannot be determined in this case with the necessary accuracy. An equally important indicator is the relatively high cost of unmanned vehicles – the benefits of their use will not pay off the installation and maintenance costs.

Support systems for vehicle drivers are of greater importance for agriculture. However, an overview of the market of such systems, which many developers are currently engaged in, shows that most of them also focus on the global positioning of objects and use navigation calculations and machine vision in a single information and analytical complex. The main goals of such systems are to improve the safety of vehicle operation and prevent road accidents, which contribute to a global reduction in their level on the roads. Similar systems are proposed to be used for agricultural machinery – for tractors and combines. Various models of navigators and course indicators are designed to solve such problems as insufficient yield and poor quality of crop production. Due to the uniform application of mineral fertilizers, irrigation and sowing, it is assumed that the most effective use of the area of agricultural land is expected.

Agronavigator for agriculture is widely advertised and experimentally used for accurate orientation of the machine movement during soil treatment, sowing of grain and row crops, row-to-row cultivation, spraying and spreading fertilizers. The main task of using course indicators is the possibility of passing equipment with a mounted or trailed unit across the field so that each subsequent lane is exactly along the edge of the previous one, with the exception of overlaps and gaps. It should be noted that the accuracy of their positioning also largely depends on the signal strength of navigation satellite systems.

Taking into account the above, it can be concluded that it is advisable to use agricultural vehicles equipped with driver support systems based on computer vision technologies. This work is devoted to the issues of creating such systems.

3. Tractor driver support system based on computer vision technologies

This article discusses a tractor driver support system based on the use of computer vision technologies to determine the direction of movement and detect obstacles during field work.

The main tasks of the system under consideration are to inform the tractor driver about the deviation of the tractor movement from the required direction of movement and to warn him about obstacles encountered on the way through the field. The direction of movement is chosen by the tractor driver based on the processing of video information, taking into account the operations of
plowing, harrowing, weeding, fertilizing, etc. performed by the tractor. In addition, the system identifies obstacles – electricity poles, trees, rocks, birds' nests, animals, people, field roads.

The system operates in offline mode and does not use satellite navigation tools to determine the location of the vehicle. It is installed on a tractor and consists of a video camera, a microprocessor image processing unit, a monitor and a power supply. The process of forming a solution in the proposed system is shown in Figure 1.

Image recording is provided by cameras installed on board the tractor. Preprocessing is designed to bring the input images to the optimal form for subsequent analysis. At this stage, geometric, brightness and color transformations are carried out that emphasize the characteristics of objects. In the next two stages, the procedures for extracting and analyzing features are performed. Based on the results of the analysis, a decision is made regarding the direction of movement of the tractor or the obstacle.

The solution of problems in the system is based on methods of detection and recognition of objects in images, which are conditionally divided into two classes – methods of classical machine learning and methods of deep learning.

Classical machine learning considers object models in the form of feature vectors and classifiers that are trained to recognize objects corresponding to these models. Most often, object models are based on features of color, texture and shape [8-10]. Color attributes are intended to represent images in terms of their color content. Among the various color features, a color histogram, a color connectivity vector, a color correlogram, color moments, and a dominant color descriptor are popular. Texture features determine the spatial distribution of colors (or brightness) of pixels in images. Statistical texture features, local binary patterns, spectral features, Tamura features, etc. are often used. The shape features refer to the image areas. Such features are, for example, the roundness of the area or its squareness, the perimeter, the area, the orientation of the main axes, etc.

The most commonly used methods for detecting and recognizing objects in images based on classical machine learning include the Viola-Jones cascade classifier, the HOG-SVM classifier, the bag of words algorithm and correlation methods [14-19].

**Figure 1.** The process of forming a solution.
The cascade Viola-Jones classifier, originally developed for face detection, is based on the use of haar-like features and local binary patterns. It has a two-level scheme. The first level provides the calculation of features as a result of processing a grayscale image, and the second level provides a cascade classification based on the use of sums and differences of calculated features over rectangular areas. At the same time, these levels are independent.

The HOG-SVM method is based on the use of histograms of oriented gradients. In this case, HOG descriptors are formed, which were originally developed to solve the problem of detecting pedestrians in images. When forming HOG-features, the gradient directions of the intensity function are calculated in the areas of the image previously divided into blocks. The use of these features in combination with a classifier based on a support vector machine made it possible to achieve a degree of pedestrian detection that was not available to earlier algorithms. Currently, detectors based on HOG descriptors are used to detect other objects as well. In the modification of the method, a latent SVM (or a method for detecting an object in parts) is used. The latent SVM-based algorithm is well suited for recognizing deformable objects, because it explicitly implements the idea of several components connected together with a deformable structure.

The bag of words algorithm, also called the bag of keypoints algorithm, – is a method of visual classification, i.e. identification of objects on the scene. The algorithm under consideration provides the selection of the most significant features, their identification in the test image and their comparison with the database for classification. At the first stage, the classifier is trained, as a result of which a dictionary of visual words is formed. After the dictionary is created, presence vectors are formed – vectors of binary values representing the presence (1) or absence (0) of each word in the dictionary. Note that the dimension of this vector is very large – hundreds or thousands of dimensions. Then the process of learning the classification algorithm takes place. In this case, the bag of words algorithm can transform each image from the input vector into a presence vector, and then use these vectors to train the algorithm to generate the correct class labels. Any method of teaching with a teacher can be used to build a classifier. Usually, a naive Bayesian classifier or a support vector machine is used.

One of the quite old and still used approaches to detecting and recognizing objects in images is based on comparison with standards (templates). The traditional technique of comparing a test image with a reference is based on considering images as two-dimensional brightness functions (discrete two-dimensional intensity matrices). In this case, a correlation metric based on a normalized correlation coefficient is often used. Correlation methods are sensitive to radiometric (brightness) and geometric (rotation, shift, scaling) distortions of the original image compared to the reference one. In addition, correlation methods are quite slow. Therefore, work is constantly being carried out to reduce the impact of these shortcomings. For example, methods using a pyramid of images are used, which significantly reduce the processing time. Also, to reduce the search time for reference objects in the image and reduce the influence of distortion when determining the similarity measure, it is often used not the images themselves, but their characteristics, considered earlier.

An alternative approach to the detection and recognition of objects in images based on classical machine learning methods is an approach based on deep learning using deep neural networks [20-22]. One of the reasons for the successful use of deep neural networks is that the network automatically selects important features from the data that are necessary for solving the problem, whereas in machine learning algorithms, the features should be highlighted by people. To date, a large number of variants of deep neural networks have been proposed. At the same time, the term "deep learning" has been widely used only since 2006. The growth in the popularity of deep neural networks, which has been occurring in the last few years, can be explained by three factors:

1) there was a significant increase in the performance of computers, including GPU computing accelerators, which made it possible to train deep neural networks much faster and with higher accuracy;

2) a large amount of data has been accumulated, which is necessary for training deep neural networks;
3) methods of training neural networks have been developed that allow fast and high-quality training of networks consisting of a hundred or more layers, which was previously impossible due to the problem of disappearing gradient and retraining.

The combination of these three factors has led to significant progress in the training of deep neural networks and their practical use, including for solving problems of detecting and recognizing objects in images. Examples of widely used deep neural network architectures for the detection and recognition task are SSD, YOLO, Faster R-CNN, R-FCN, AlexNet, VGGNet, GoogLeNet, ResNet, MobileNet.

The development of deep neural networks is associated with the support of manufacturers of hardware and software for parallel computing. For example, one of the most well-known manufacturers in this area is NVIDIA. Among various products, NVIDIA offers the cudna software library with support for GPU primitives for deep neural networks, which provides performance, ease of use and low memory load. Based on NVIDIA's cuDNN, the DIGITS package has been developed, which provides tools for solving the most common tasks in the field of deep learning, such as: data preparation, definition of a convolutional network, parallel training of several models, monitoring the learning process in real time, as well as choosing the best model. A successful example of the use of DIGITS is the DetectNet network, designed to detect objects in images.

Progress in the development of computing technologies has led to a wide spread of energy-efficient mobile computing architectures, on the basis of which autonomous intelligent systems are built. Specialized deep learning software libraries are being developed for such systems, including those that allow solving the problems of object detection and recognition in video sequences. For example, there are TensorFlow Mobile and TensorFlow Lite platforms designed for use in Android and iOS mobile devices.

It is believed that deep learning methods provide faster and more accurate detection and recognition of objects in images. But at the same time, the task of choosing a specific architecture for specific application conditions is an independent scientific task that needs to be solved, especially in the case of creating new systems and considering new objects. It should be noted that there are no large sets of training data for the systems under consideration. Their creation is an important practical task. It is based on the procedures for obtaining, processing, analyzing, marking and synthesizing images with specified objects. In addition, it is necessary to solve the problem of choosing a computer system for use in the working conditions of agricultural machinery (real time, atmospheric interference, vibration, electromagnetic influences, etc.), taking into account the need to ensure an optimal ratio between cost and performance requirements.

Based on the analysis, a deep neural network YOLO using the DarkFlow framework was used to detect and recognize objects in the tractor driver support system proposed by the authors. This solution allows you to ensure high efficiency of solving functional problems in combination with ease of implementation. An example of the operation of a neural network is shown in figure 2. At the moment, the accuracy of detecting specified objects on agricultural images is about 85%.
Figure 2. Example of a neural network operation.

4. Conclusion
Detecting obstacles to the movement of a tractor is one of the primary tasks, the most useful for a tractor driver working in the field. Paying attention to the emerging obstacle in time, analyzing it, telling a person the degree of its danger and giving options for the proposed further actions allows to increase the efficiency of agricultural work by reducing their labor intensity and losses. The advantages of the proposed approach are the implementation of the most effective method of processing video information using deep neural networks at the moment. The autonomy of the system implementation, independence from navigation calculations of global positioning makes it possible for its wide application in any fields. At the same time, its relatively low cost is available even for small farms.

The implementation of the considered approach makes it possible to ensure the creation and introduction of intelligent technologies in agriculture, which corresponds to the Development Strategy of the agro-industrial complex of the Russian Federation for the period up to 2030.

5. References
[1] Pavlov B P, Garifullin R F, Mingaleev G F and Babushkin V M 2019 Key technologies of digital economy in the Russian Federation Proceedings of the 33rd International Business Information Management Association Conference, IBIMA 2019: Education Excellence and Innovation Management through Vision 2020 pp 3401–3407
[2] Pavlov B P, Garifullin R F, Babushkin V M and Mingaleev G F 2019 Digital transformation of the economy Proceedings of the 33rd International Business Information Management Association Conference, IBIMA 2019: Education Excellence and Innovation Management through Vision 2020 pp 3359–3364
[3] 2019 Digital transformation of agriculture in Russia: official Ed. Ministry of Agriculture of Russia (Moscow: Rosinformagrotech) p 80
[4] Mavridou E, Vrochidou E, Papakostas G A, Pachidis T and Kaburlasos V G 2019 Machine Vision Systems in Precision Agriculture for Crop Farming Journal of Imaging 5(12) 89
[5] Makhmutova A Z, Anikin I V and Dagaeva M 2020 Object Tracking Method for Videomonitoring in Intelligent Transport Systems 2020 International Russian Automation Conference RusAutoCon 2020 p 535–540
[6] Cheng H 2011 Autonomous Intelligent Vehicles (London: Springer) p 164
[7] Gonzalez R C and Woods R. E. 2018 Digital Image Processing (London: Pearson) p 1168
[8] Shapiro L G and Stockmann G C 2001 Computer Vision (London: Pearson) p 608
[9] Davies E R 2017 Computer Vision: Principles, Algorithms, Applications, Learning (London:
[10] Sonka M, Hlavac V and Boyle R 2015 *Image Processing, Analysis, and Machine Vision* (Boston: Cengage Learning) p 920

[11] Li S Z 2009 *Markov Random Field Modeling in Image Analysis* (London: Springer) p 384

[12] Nixon M and Aguado A 2019 *Feature Extraction and Image Processing for Computer Vision* (London: Academic Press) p 650

[13] Flusser J, Suk T and Zitova B 2009 *Moments and Moment Invariants in Pattern Recognition* (New York: Wiley) p 312

[14] Viola P and Jones M J 2001 Rapid Object Detection using a Boosted Cascade of Simple Features *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2001)* 1 pp 511–518

[15] Dalal N, Triggs B and Schimid C 2006 Human detection using oriented histograms of flow and appearance *European Conference on Computer Vision (ECCV)* pp 428–441

[16] Zhu Q, Avidan S, Yeh M C and Cheng K T 2006 Fast human detection using a cascade of histograms of oriented gradients *Conference on Computer Vision and Pattern Recognition (CVPR)* pp 1491–1498

[17] Felzenszwalb P F et al 2010 Object detection with discriminatively trained part-based models *IEEE transactions on pattern analysis and machine intelligence* 32(9) pp 1627–1645

[18] Kaehler A and Bradski G 2017 Learning OpenCV 3 (Sebastopol: O’Reilly) p 1024

[19] Visilter Y V, Zheltov S Y, Bondarenko A V, Ososkov M V and Morzhin A V 2010 Image processing and analysis in machine vision applications: Course of lectures and practical classes (Moscow: Fizmatkniga) p 672

[20] Goodfellow I, Bengio Y and Courville A 2016 Deep learning (Cambridge: The MIT Press) p 800

[21] Sozykin A V 2017 An Overview of Methods for Deep Learning in Neural Networks *Bulletin of the South Ural State University Series: Computational Mathematics and Software Engineering* 6 3 pp 28–59

[22] Redmon J, Divvala S, Girshick R and Farhadi A 2016 You only look once: Unified, real-time object detection *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* pp 779–788