Using Convolutional Neural Networks to Monitor Security at an Industrial Facility

V.V. Mokshin, A.I. Sultanova, L.M. Sharnin

YOLOv3 neural network

The main feature of the YOLO neural network compared to others is that most systems apply the CNN several times to different regions of the image, while YOLO applies the CNN once to the entire image at once. Therefore, the floating window method is not suitable for this network. Instead, the entire image is divided into S × S cells using a grid. After that, two indicators are predicted for each cell: bounding boxes and the probability of finding the desired object in them.

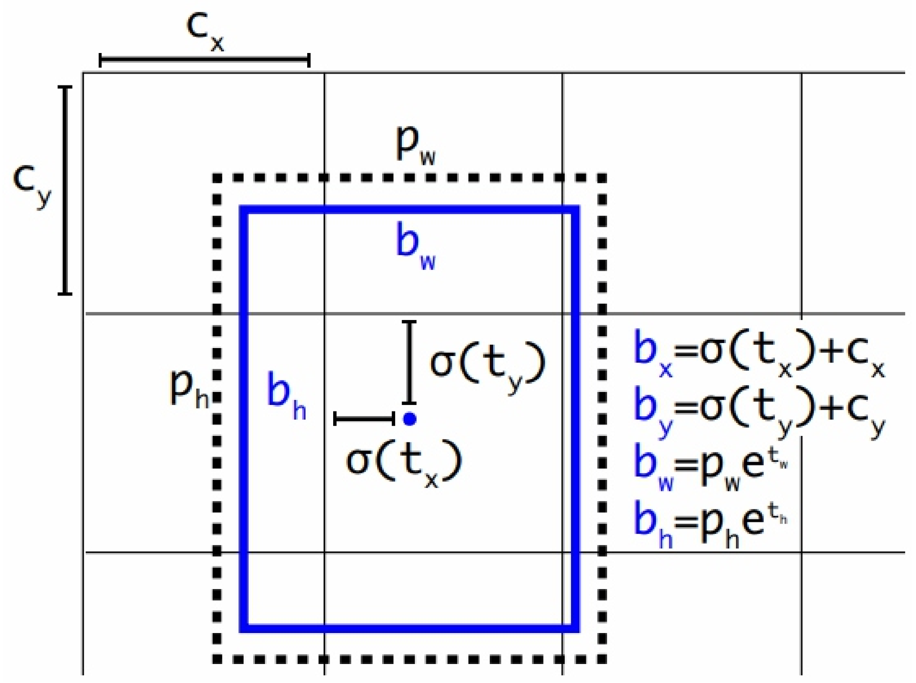

Fig. 1. Determining the location of a leak in the estimated section of the water supply

The dimension and position of the bounding boxes around the object in the image are calculated by the equation:

$$b_x=\sigma(t_x)+c_x \qquad b_y=\sigma(t_y)+c_y \qquad b_h=p_h e^{th} \qquad b_w=p_w e^{tw}$$

The probability of hitting the desired object inside the predicted frame is calculated by the equation:

$$\sigma(t_o)=\Pr(object)*IOU(b,object)$$

To get rid of duplicate bounding boxes for each object, a Non-maximum suppression algorithm is used. It searches for the bounding boxes containing the recognition object and calculates for them YOU according to the equation:

$$IOU=\frac{A\cap B}{A\cup B}$$

Algorithm	Scores	
	Average accuracy (mAP), %	*Processing speed, ms*
R-CNN	43.56	2991
Fast R-CNN	49.87	2271
Faster R-CNN	58.78	122
YOLOv2 (S x S)	81.64	59.1
YOLOv3 (S x S)	87.42	24.8

Table 1. Comparison of object recognition methods

According to the table 1, we can conclude that the YOLOv3 model has the highest recognition accuracy and the fastest input data processing speed compared to other algorithms

Stage of training

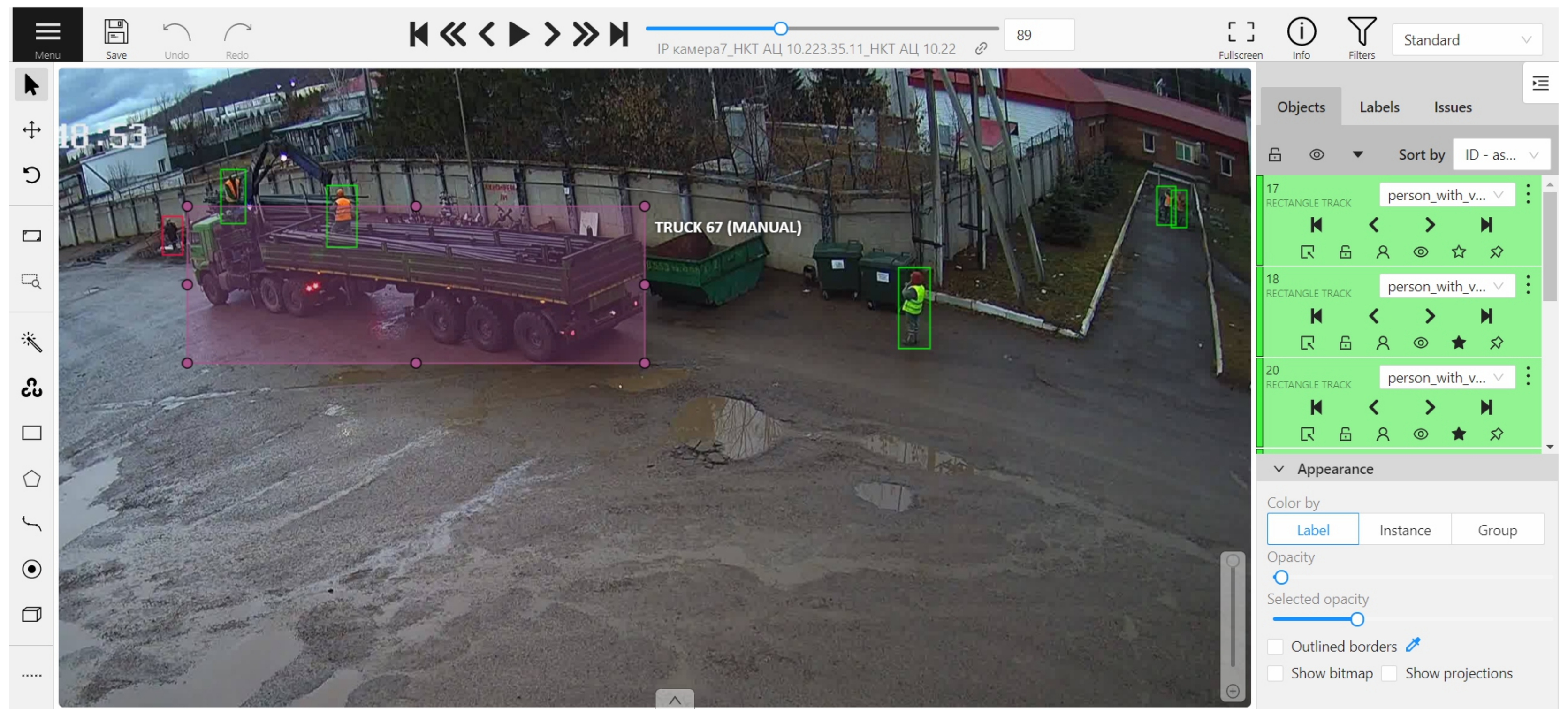

Fig. 2. Marking of workers in vests, a worker without a vest and a truck

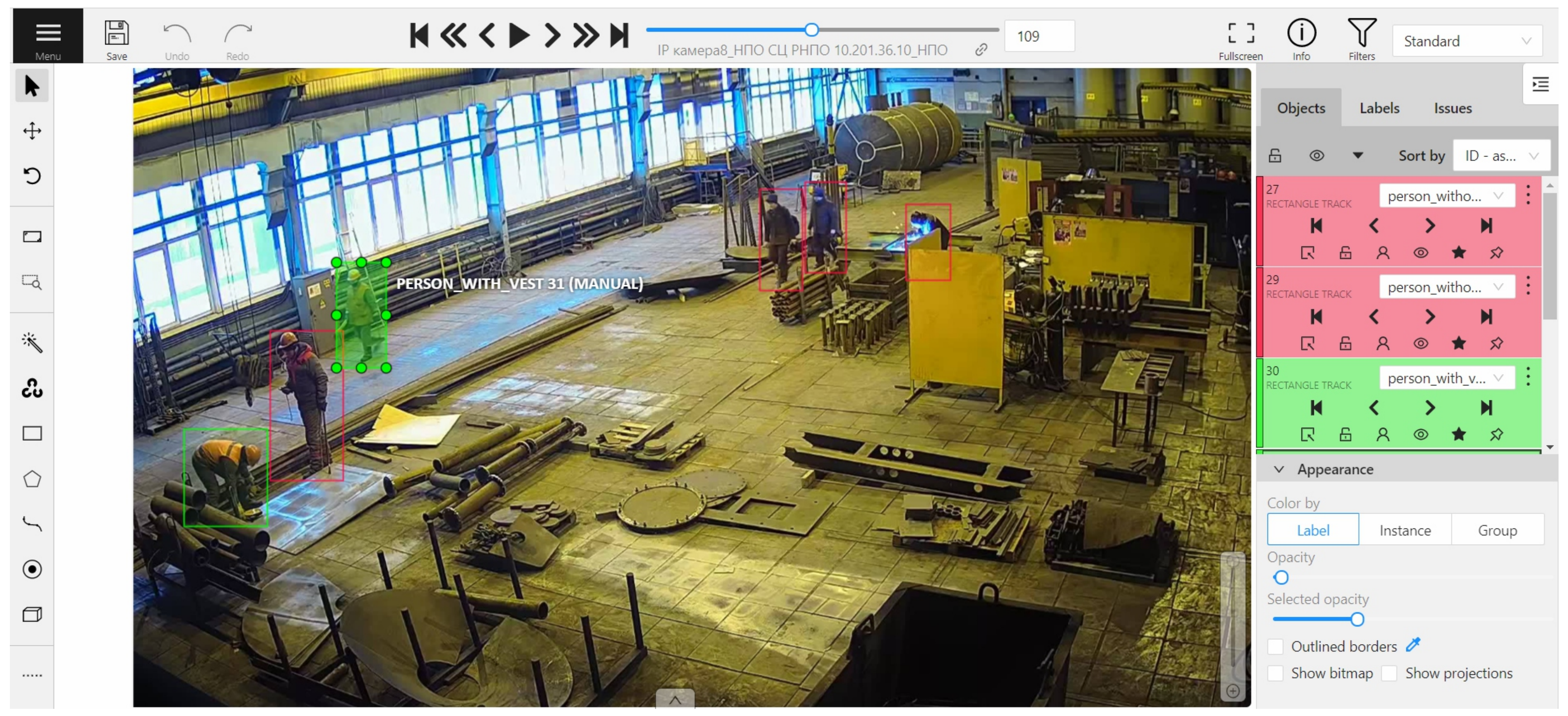

Fig. 3. Marking of workers in vests and workers without vests

On fig. 4 shows the training graph (the loss function approaches 0 with each new iteration), the average value of the current loss is 0.2789 and the number of iterations is 6000.

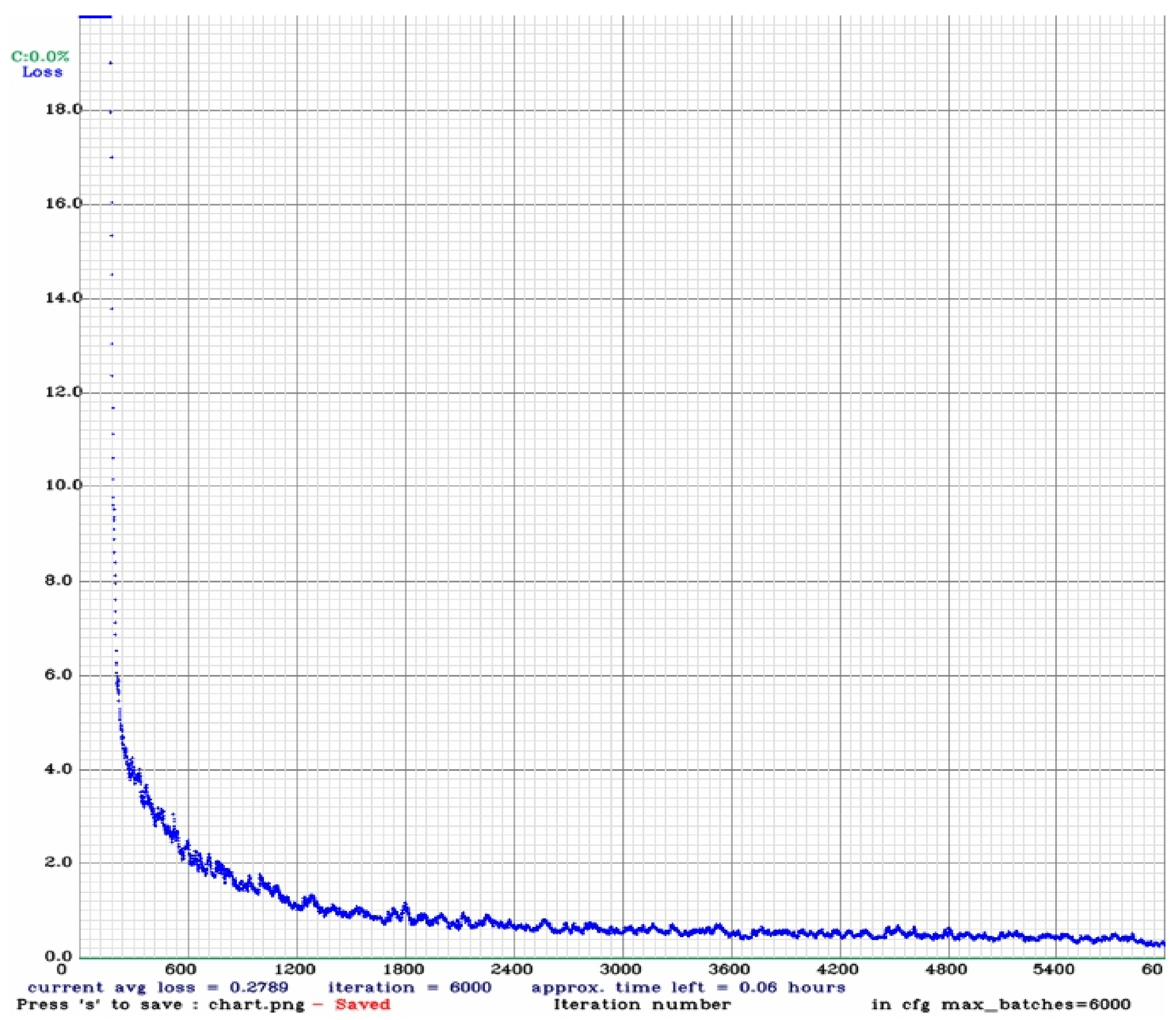

Fig. 4. Total training loss

Stage testing

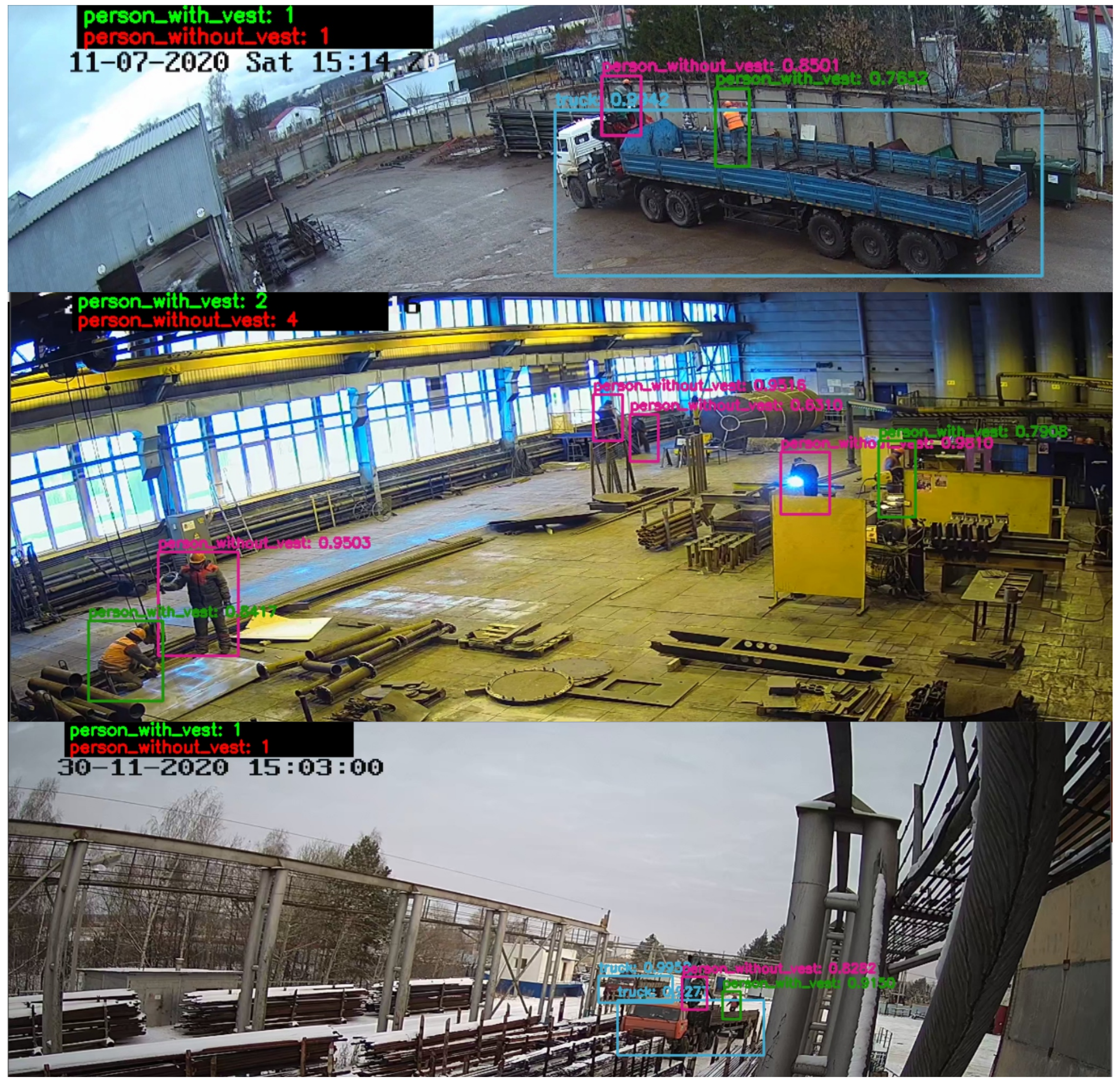

Fig. 5. The result of the YOLOv3 convolutional neural network trained on the production dataset

Conclusion

As a result of scientific work, the YOLOv3 neural network was implemented in the Python programming language to monitor security at a production facility. For training, a dataset was formed, consisting of 300 images and annotations to them.

Also during the training, key points were identified that significantly affect the quality of the results obtained and the efficiency of the recognition system:

– The selection of images for the dataset should be as complete and of high quality as possible;

– Annotation of objects should be performed as accurately as possible;

– At the beginning of training, it is worth using the weights of already trained models.

Based on the results obtained, it is possible to build higher-level models for use in research purposes, as well as to create systems for analyzing and monitoring safety at production facilities.

[1] **KAZAN NATIONAL RESEARCH TECHNICAL UNIVERSITY NAMED AFTER A. N. TUPOLEV - KAI**
10 K.Marx St., Kazan,
Tatarstan 420111, Russia (www.kai.ru)

[2] **SAMARA NATIONAL RESEARCH UNIVERSITY**
34, Moskovskoye shosse, Samara, Russia (www.ssau.ru)