Automatic x-ray image analysis for aviation security within limited computing resources

N A Andriyanov\textsuperscript{1}, Al K Volkov\textsuperscript{2}, An K Volkov\textsuperscript{2}, A A Gladkikh\textsuperscript{3,4} and S D Danilov\textsuperscript{3}

\textsuperscript{1}JSC "RPC "Istok" named after Shokin", Vokzalnaya Str., 2A, 141190, Moscow Region, Russia
\textsuperscript{2}Ulyanovsk Civil Aviation Institute, Mozhaisky Street, 8/8, 432071, Ulyanovsk, Russia
\textsuperscript{3}Ulyanovsk State Technical University, Severnyj venec, 32, 432027, Ulyanovsk, Russia
\textsuperscript{4}E-mail: a_gladkikh@mail.ru

Abstract. The objective of this work is to develop approaches to automating inspection procedure at airports. The article presents the deficiencies of the existing inspection system, concluding in the negative impact of the human factor. It is proposed to use convolutional neural networks for automatic x-ray image analysis of passenger baggage. The paper presents the results of the convolutional neural network with various input data and architecture within limited computing resources. In a view to further development, this study can contribute to the development of specialized software to help aviation security screeners through partial automation of their work.

1. Introduction
Currently, systems of technical vision are becoming more and more developed. Although until recently, clear and understandable detection and recognition algorithms were based on mathematical image models \cite{1, 2}, since computing tools were quite limited and machine learning methods provided sufficient efficiency for solving recognition problems only for a narrow range of tasks. The scope of currently solved vision system tasks has expanded significantly with the convolutional neural networks \cite{3-4}. The task of recognizing forbidden objects on X-ray images by aviation security screeners can be distinguished. The airport screening procedure is one of the key measures in aviation security. However, its effectiveness largely depends on the human factor, which is associated with the lack of aviation security training, the impact of the fatigue, the inability to process large volumes of data, and the distraction factor. To overcome these factors, it is necessary to implement automatic x-ray images recognition of passenger baggage and hand luggage.

In addition, it is still relevant to enter deep learning, which requires significant resources to ensure high efficiency of trained networks. Typically, computer clusters are used that perform parallel computing on multiple graphics processing units \cite{5}.

However, the question arises of what to do if resources are limited, and the initial solution needs to be obtained in the near future. This paper presents basic steps that can significantly increase the
recognition efficiency even for complex images using the example of X-ray images in aviation security.

2. Source data and network architecture
At the initial stage, the image database was pre-assembled. Its several features should be emphasized. Firstly, despite the fact that binary classification is performed, the task is more complex in fact. This is due to the fact that inside the classes "dangerous" and "safe", which are the outputs of the network, images of several completely different objects can be represented. Secondly, a significant area in the images can be of different sizes. In addition, with low target prevalence of prohibited items, the original x-ray images database of clean baggage will always predominate the similar database with prohibited items. In addition, for different subclasses within the division into "dangerous" and "safe", there is also an uneven distribution of images. Figure 1 shows examples of source images. In the first row – various images of dangerous objects, in the second – safe.

![Figure 1. Examples of source images.](image)

Preliminarily operations were carried out with the source dataset. They made it possible to increase the sample. Among them were various rotations and zoom operations.

The training was carried out on convolutional neural networks with the architecture shown in figure 2.

![Figure 2. Network architecture.](image)
It is important to note that the number of layers could vary. However, taking into account the characteristics of the computer technology (NVIDIA GeForce GTX1060 3GB GDDR5 1708 MHz), limitations on RAM did not allow research on more complex architectures. Nevertheless, due to some changes to the network architecture, it was possible to increase the percentage of correct recognition.

3. Training and results
For the adequacy of the analysis, the initial sample were divided into training, validation and test sample, according to the table 1.

| Table 1. Initial sample division. |
|----------------------------------|
| Initial sample | Dangerous | Safe |
| Training sample | 895       | 1814 |
| Validation sample | 62    | 107  |
| Test sample      | 59       | 106  |
| Total            | 1016     | 2027 |

According to table 1, if the network will always issue the "safe" object, the test sample will have 64 % of the correct answers. Let’s consider iterations that can increase this figure within the shortest time.

1. At the beginning, we will use more user-friendly MATLAB interface to track the learning process. First, we reveal how critical the source data are to the color features of the images. As image processing in grey levels can reduce computing resources, it is implemented a network with 3 levels of convolution: 40, 36 and 25 neurons. For the best speed, optimization by gradient descent was chosen and the duration of training is 8 epochs.

   The results of correct recognition on the test sample: black and white images – 62.34 %, color – 66.48 %. Therefore, it will take into account the color attribute.

2. We’ll check how much and if it’s possible to improve network efficiency when we move to ADAM optimization. The network architecture will remain the same. Correct recognition results on a test sample: 69.62 %.

3. Further, we complicate the network architecture a bit by making a network with 3 levels of convolution: 64, 48, and 32 neurons.

   Correct recognition results on a test sample: 75.32 %. Figure 3 shows the network learning process.

   Presented graphs analysis shows that convergence is not achieved for 8 epochs, while significant time is spent on training in MATLAB. Therefore, further steps were performed using Python.

4. The network from item 3 during 200 epochs of training worked for 1690 sec., which is much less than the time spent by MATLAB for training in 8 epochs. Correct recognition results on a test sample: 69.32 %. The analysis shows that such a network turned out to be retrained, and the results on the test sample decreased.

5. The next test was performed on 100 epochs of training.

   Correct recognition results on the test sample: 81.75 %.

6. Finally, the image resizing method was used. Instead of 112x112 images, the converted 128x128 image was initially input. This allowed to increase the percentage of correct recognition. Correct recognition results on a test sample: 85.71%.

7. The latest change was an increase in the number of neurons in layer 1 and 2 to 128, which also improved the results. At the same time, the number of training epochs, taking into account compression, was reduced to 50, since for 100 epochs, the results on the training sample are 100 %, which indicates retraining. In addition, it also allows to reduce computational efforts by 2 times.
Correct recognition results on a test sample: 90.47%. Figure 4 shows an example of processing some images with this network.

8. Further experiments with training parameters, input data and neural networks did not improve these results. However, it should be noted that due to the lack of time during the day optimization
methods were undertaken that ensured correct recognition of more than 90%. Table 2 summarizes all the results.

Table 2. Comparison of different approaches.

| Network         | Initial data | Epochs, amt. | Optimization | Correct recognition, % |
|-----------------|--------------|--------------|--------------|------------------------|
| CNN 40-36-25    | Black-and-white 112x112x1 | 8           | SGD         | 62.34                  |
| CNN 40-36-25    | Color        | 112x112x3    | 8           | SGD                    | 66.48                  |
| CNN64-48-32     | Color        | 112x112x3    | 200         | ADAM                   | 69.32                  |
| CNN 40-36-25    | Color        | 112x112x3    | 8           | ADAM                   | 69.62                  |
| CNN 64-48-32    | Color        | 112x112x3    | 8           | ADAM                   | 75.32                  |
| CNN 64-48-32    | Color        | 112x112x3    | 100         | ADAM                   | 81.75                  |
| CNN 64-48-32    | Color        | 128x128x3    | 100         | ADAM                   | 85.71                  |
| CNN 128-128-32  | Color        | 128x128x3    | 50          | ADAM                   | 90.47                  |

It is important to note that table 2 represents obtained solutions in increasing order. In general, for the original task, it was possible to increase the recognition efficiency from 64% to 90.47%. It is usually considered to be a successful solution in non-critical recognition problems.

4. Conclusion
A neural network was trained using primitive modifications of the input data and the architecture of the neural x-ray images network of the airport security service. The training provided a test result of more than 90%. Further improvement can be obtained by deep researching of other modifications related to an increase in the training sample or more complex networks organization. This study is expected to contribute to the development of specialized software to help aviation security screeners through partial automation of their work process.

Acknowledgments
The reported study was funded by RFBR and INSF, project number 19-47-730011.

Reference
[1] Vasil’ev K K, Andriyanov N A 2020 Image Representation and Processing Using Autoregressive Random Fields with Multiple Roots of Characteristic Equations Computer Vision in Advanced Control Systems-5 175 11-52 DOI: 10.1007/978-3-030-33795-7_2
[2] Andriyanov N A, Dementiev V E 2019 Developing and studying the algorithm for segmentation of simple images using detectors based on doubly stochastic random fields Pattern Recognition and Image Analysis 29(1) 1-9 DOI: 10.1134/S105466181901005X
[3] Tianyi L, ShuangSang F, Yuehui Z and Peng W 2006 Implementation of Training Convolutional Neural Networks https://arxiv.org/ftp/arxiv/papers/1506/1506.01195.pdf
[4] Sakshi I, Anil G, S. P. Mishra and Pooja A 2018 Conceptual Understanding of Convolutional Neural Network – A Deep Learning Approach Procedia Computer Science 132 679–88
[5] Campos V, Sastre F, Yagüés M and Bellver M, Giro-i-Nieto X and Torres J 2017 Distributed training strategies for a computer vision deep learning algorithm on a distributed GPU cluster Procedia Computer Science 108 315-24 DOI: 10.1016/j.procs.2017.05.074