Research of recognition accuracy of dangerous and safe x-ray baggage images using neural network transfer learning

N A Andriyanov1,2, Al K Volkov3, An K Volkov3, A A Gladkikh2,4
1 JSC "RPC "Istok" named after Shokin", Vokzalnaya Str., 2A, 141190, Moscow Region, Russia
2 Ulyanovsk State Technical University, Severnyj venec, 32, 432027, Ulyanovsk, Russia
3 Ulyanovsk Civil Aviation Institute, Mozhaisky Street, 8/8, 432071, Ulyanovsk, Russia
4 E-mail: a_gladkikh@mail.ru

Abstract. The article considers the use of neural networks to solve the problem of recognizing dangerous and safe objects carried in the luggage of airport passengers. A comparative analysis is performed to define the accuracy achieved on the test sample for different convolutional neural networks. It also explores the influence of various regularizations on the accuracy of a two-class classification. The increased probability of correct recognition is achieved due to augmentation, reset weights and saturation of the network. The method of transfer training is used to increase the efficiency of the recognizer. In this case, a study was carried out for the transfer of various neural networks.

1. Introduction
Among relevant tasks on computer vision, one can single out the task of recognizing dangerous objects that can be used to commit an act of unlawful interference, according to the results of an X-ray baggage or hand luggage scan by the aviation security service [1-3]. Moreover, automation of such a process is possible using convolutional neural networks (CNN) [4,5], which are currently replacing algorithms focused on mathematical image modeling [6, 7]. Now there are many approaches to the learning process of neural networks, and their application to solve this problem is advisable. At the same time, the special interest in solving a narrow specialized problem, such as transfer learning and the use of regularizations for X-ray data, has not been fully studied. Therefore, in this paper, special attention is paid specifically to improving recognition accuracy due to transfer learning and regularization.

2. A brief analysis of the current state of the problem
Currently, screening at airport checkpoints is carried out by security officers who visually check x-ray images of baggage and hand luggage of passengers in order to determine whether they are harmless or contain prohibited items and, therefore, ask additional screening using explosives detection or manual inspection of baggage [2]. Examples of prohibited items are pistols, knives, improvised explosive devices, self-defense gas cans or devices for electric shock [8]. Inspection of x-ray images by airport security officers (inspectors) includes visual search and decision-making [9,10]. The problems of
visual search include low target prevalence, changing target visibility, finding an unknown set of goals, and the possible presence of multiple targets [11,12].

However, in the considered works [2,8–12], computer vision methods are considered only as secondary. The main attention is paid to aspects of the operator’s perception of the visual flow in order to classify the recorded images as accurately as possible. In [13], a neural network was developed that could effectively separate dangerous and safe objects in X-ray images. However, with the training used from scratch, a sufficiently large error remains.

At the same time, one of the popular techniques for preliminary adjustment of neural networks to solve a narrow task in computer vision is to use the experience of a neural network that has proven itself in a wider class of problems, or, in other words, transfer learning [14]. With this approach, a trained neural network is used as a layer (in fact, many layers) of feature extraction. At the same time, a narrow task leaning occurs only in the output fully connected layer, since the coefficients of the internal network do not change. It provides quite high results in accuracy with much lower computational work.

In addition, it is necessary to develop neural networks in such a way that they have the greatest generalizing ability for a specific narrow task. To do this, the widest range of options needs to be considered. Augmentation algorithms are often used due to the lack of data [15]. Finally, great versatility can be achieved by using a weight reset [16], since this avoids neural network overfitting.

In this article, transfer learning will be considered for the following neural networks:

- VGG-19: is a convolutional neural network that contains 19 layers in depth. It is possible to download a pre-trained version of the network, trained on more than a million images from the ImageNet database. A pre-trained network can classify images into 1,000 categories of objects, such as a keyboard, mouse, pencil, and many animals. The network has an input image size of 224x224.
- U-Net: is a convolutional neural network that was created in 2015 to segment biomedical images in the Computer Science division of the University of Freiburg. The network architecture is a fully-connected convolutional network modified so that it can work with fewer examples (images of learning) and make more accurate segmentation.
- GoogleNet: is a convolutional neural network that contains 22 layers in depth. It is possible to download a pre-trained version of the network, trained on either ImageNet or Places365 databases. An ImageNet network categorizes images into 1,000 categories of objects, such as keyboards, mice, pencils, and many animals. A Places365 network is similar to ImageNet network, but classifies images into 365 different place categories, such as field, park, runway and lobby. An ImageNet network will be used. Pre-trained networks have an input image size of 224x224.

3. Description of the original data sample

The database of x-ray images included both dangerous and safe objects. The following groups belong to dangerous objects: improvised explosive devices, cold weapons and firearms, as well as other objects (stun guns, gas sprays, etc.). Safe items include various household items, clothing, food, etc. It should be noted that the database of x-ray images of safe objects will objectively always surpass the similar base with dangerous objects due to their low target prevalence. The task of binary classification is also complicated by the following facts: inside the classes “dangerous” and “safe”, which are the outputs of the network, x-ray images of several object groups are presented with their uneven distribution, moreover, a significant area in the images can have different sizes. Fig. 1 shows examples of x-ray images. The first row of X-ray images includes dangerous objects, the second row includes safe objects.
Before studying recognition efficiency under various training methods, all available images were divided as follows: 1894 safe and 855 dangerous in the learning sample, 106 safe and 52 dangerous in the test sample.

4. The study of weight regularization and augmentation impact on a neural network

First, we will perform learning and inference of a neural network trained on the basis of a conventional convolutional architecture with zero coefficients. We’ll set up 50 learning epochs with ADAM optimization, and each layer will contain 64 neurons. The fully connected layer contains 128 neurons. All image sizes are reduced to 224x224.

Table 1 presents the recognition results on a test sample for neural networks with different depth (number of layers). Table 2 presents the recognition characteristics of a 7-layer network for various weight loss probability, that is a change in the weight coefficient for a neuron with a certain probability set to 0 at each epoch. Finally, Table 3 shows results when applying augmentation to a learning image in a given epoch with some given probability for each image. It should be noted that the number of augmentations considered shifts, turns, scaling and adding white noise, as well as blurring images. In this case, the data augmentation algorithm works so that it is possible to overlay different augmentations to each other.

Increased network depth and augmentation results in better recognition accuracy.

| Number of layers | Recognized, % |
|------------------|---------------|
| 1                | 67.722        |
| 3                | 74.684        |
| 5                | 77.848        |
| 7                | 82.278        |

**Figure 1. X-ray image examples**

| Image Example 1 | Image Example 2 | Image Example 3 | Image Example 4 |
|-----------------|-----------------|-----------------|-----------------|
| ![Image 1]      | ![Image 2]      | ![Image 3]      | ![Image 4]      |
Table 2. Accuracy from weight loss

| Probability of weight loss | Recognized, % |
|----------------------------|---------------|
| 0.1                        | 79.114        |
| 0.2                        | 77.215        |
| 0.3                        | 62.025        |
| 0.4                        | 33.544        |
| 0.5                        | 32.911        |

Table 3. Accuracy from data augmentation

| Augmentation probability | Recognized, % |
|--------------------------|---------------|
| 0.1                      | 82.911        |
| 0.2                      | 86.076        |
| 0.3                      | 84.177        |
| 0.4                      | 82.911        |
| 0.5                      | 81.013        |

Obviously, the accuracy in this case is directly proportional to the depth of the network. However, over time, a further increase in the number of layers will lead to a much larger increase in computational work than a gain in recognition accuracy. It is interesting that weight loss in this problem led to a significant decrease in recognition accuracy, therefore, the parameter characterizing the probability of weight loss was chosen to be 0.1 for future studies. At the same time, for augmentations, the maximum accuracy was achieved with a probability of 0.2. At the same time, for the considered example, resetting the scales did not lead to an improvement in the quality of the network, and a slight augmentation made it possible to increase the accuracy by about 3.8%.

5. The study of transfer learning impact on a neural network

Now consider the results of the neural network, which learning took place with pre-trained weights. There were 128 neurons in the fully connected learning layer. Given the transfer learning, a fully connected layer has been trained for 10 epochs. When submitting test images, there were used augmentations with probability \( p = 0.2 \) and the probability of weight regularization on a fully connected layer was set at \( p = 0.05 \).

Table 4 presents the results for various neural networks.

Table 4. Transfer learning

| Neural network                      | Accuracy, % |
|-------------------------------------|-------------|
| Trained from scratch (2nd line of the 3rd table) | 86.076      |
| U-Net                               | 91.244      |
| GoogleNet                           | 91.365      |
| VGG-19                              | 93.018      |

The results show that transfer learning is extremely advisable. The accuracy was increased by 7\% using VGG-19 network weights compared to the best network trained from scratch.

6. Conclusion

Thus, with a relatively small initial image sample, sufficiently high recognition percentage was achieved when augmentation and transfer learning were used. In particular, the best results were obtained for the VGG-19 network, which can be recommended for further research in the field of X-ray image processing.
7. Acknowledgments
The reported study was funded by RFBR, project number 19-47-730011.

Reference
[1] Chavaillaz A, Schwaninger A, Michel S, Sauer J 2019 Expertise, Automation and Trust in X-Ray Screening of Cabin Baggage Frontiers in Psychology 10(256) pp 1-11 DOI: 10.3389/fpsyg.2019.00256
[2] Sterchi Y, Schwaninger A 2015 A first simulation on optimizing EDS for cabin baggage screening regarding throughput International Carnahan Conference on Security Technology (ICCST) pp 55–60 DOI: 10.1109/CCST.2015.7389657
[3] D’Arcy G, Márquez-Grant N, Lane D W 2020 Baggage scanners and their use as an imaging resource in mass fatality incidents International Journal of Legal Medicine 134 pp 1419–1429 DOI: https://doi.org/10.1007/s00414-019-02132-y
[4] Indolias A, Kumar G A, Mishrab S P, Asopaa P 2018 Conceptual Understanding of Convolutional Neural Network – A Deep Learning Approach Procedia Computer Science 132 pp 679–688
[5] Boukaye Boubacar Traore, Bernard Kamsu-Foguem, Fana Tangara 2018 Deep convolution neural network for image recognition Ecological Informatics 48 pp 257–268
[6] 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) pp 1–9 DOI: 10.1134/S105466181901005X
[7] Smith J, Best E, Sum E, Guzel Y, Saville M A, LoMonte L, Wicks M 2015 Physical-model-based image processing for feature aided analysis 2015 International Conference on Electromagnetics in Advanced Applications (ICEAA) pp 565–568 DOI: 10.1109/ICEAA.2015.7297179
[8] Schwaninger A 2005 Increasing efficiency in airport security screening WIT Transactions on the Built Environment 82 pp 407–416
[9] Koller S M, Drury C G, Schwaninger A 2009 Change of search time and non-search time in X-ray baggage screening due to training. Ergonomics 52(6) pp 644–56
[10] Wolfe J M, Van Wert M J 2010 Varying target prevalence reveals two dissociable decision criteria in visual search Current Biology 20(2) pp 121–124 DOI 10.1016/j.cub.2009.11.066
[11] Mitroff S R, Biggs A T, Cain M S 2015 Multiple-target visual search errors Policy Insights the Behavioral Brain Sciences 2 pp 121-128 DOI: 10.1177/2372732215601111
[12] Biggs A T, Kramer M R, Mitroff S R 2018 Using cognitive psychology research to inform professional visual search operations Journal of applied research in memory and Cognition 7 pp 189–198 DOI: 10.1016/j.jarmac.2018.04.001
[13] Andriyanov N A, Volkov Al K, Volkov An K, Gladkikh A A, Danilov S D 2020 Automatic x-ray image analysis for aviation security within limited computing resources IOP Conference Series: Materials Science and Engineering 862(1) pp 1–5 DOI:10.1088/1757-899X/862/5/052009
[14] Renu Khandelwal 2019 Deep Learning using Transfer Learning Towards Data Science https://towardsdatascience.com/deep-learning-using-transfer-learning-cfbce1578659
[15] Buslaev A, Parinov A, Khvedchenya E, Iglovikov V, Kalinin A 2018 Albumentations: and flexible image augmentations https://arxiv.org/abs/1809.06839
[16] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R 2014 Dropout: A simple way to prevent neural networks from overfitting Journal of Machine Learning Research 15 pp 1929–1958