Proceeding Paper

Deep Learning for Detecting Dangerous Objects in X-rays of Luggage [†]

Nikita Andriyanov [iD]

Data Analysis and Machine Learning Department, Financial University under the Government of the Russian Federation, Leningradsky pr-t 49, Moscow 125167, Russia; naandriyanov@fa.ru; Tel.: +7-927-818-8416

[†] Presented at the 15th International Conference "Intelligent Systems" (INTELS'22), Moscow, Russia, 14–16 December 2022.

Abstract: The investigation presented in this text is the study of object detection algorithms in the task of analyzing images of baggage and hand luggage. A modified version of the YOLOv5 convolutional neural network with additional rechecking based on the VGG-19 network is proposed. The modification is based on transfer learning from the available images. A comparison is made with other known algorithms. The article shows that the application of the proposed model made it possible to achieve the value of the mean average recall (mAR) at the level of 87% for dangerous objects of five classes.

Keywords: object detection; convolutional neural networks; aviation security; pattern recognition; computer vision

1. Introduction

At present, the issue of using various intelligent systems to ensure security in crowded places is quite acute [1–4]. Of course, optical cameras are an important source of information for preventing and detecting suspicious situations. At the same time, sufficient progress has been made in the processing of optical images, including the task of detecting objects [5,6].

Nevertheless, in the overwhelming majority of works, the marked problem of object detection is considered for optical images. In the case of developing a computer vision system for analyzing luggage images, such a system must be able to perform efficient processing of X-ray images. There is one important thing about such data. Unlike optical images, X-ray images have the property of "transparency", i.e., some objects can be seen against the background of others, and not overlapped by them. This complicates the task of analyzing such images. In the literature, there are works on the recognition of objects in X-ray images of luggage [7–9], but they consider that the image contains an object of any single class in the full image. Some solution to the problem of prohibited item detection is presented in [10]. At the same time, the value of the mean average precision (mAP) metric is about 64%. The main weakness of the work is the proposal of its own architecture of a convolutional neural network, which is insufficient for high-quality training.

In view of the foregoing, the task of detecting prohibited objects in X-ray images of baggage and hand luggage remains an urgent task. The solution to this problem will automate the routine activities of inspection operators and avoid errors caused by human factors. The solution to this problem will be described in the following sections so that the Section 2 will consider the data for which training and testing are carried out. Then, in the Section 3, special attention is paid to the proposed approach for detecting dangerous objects in X-ray images. Finally, the Section 4 is a description of the processing results for test images, accompanied by analysis and comparison with known algorithms. In conclusion, the main conclusions of the study are formulated and options for further improvement of the results obtained are proposed.

2. Description of the Baggage and Hand Luggage Images Set

Training data plays a critical role in deep learning and machine learning. For the purposes of the presented study, X-ray photographs of luggage were used, provided by the Ulyanovsk Civil Aviation Institute (UCAI) named after the Chief Marshal of Aviation Boris Pavlovich Bugaev (Ulyanovsk, Russia). In total, there were 1500 images in the database. Figure 1 shows some X-ray images to be processed by developed algorithms. It should be noted that the images of various prohibited items were taken for the purposes of the study and do not correspond to actual cases of attempted smuggling of prohibited items on board the aircraft.
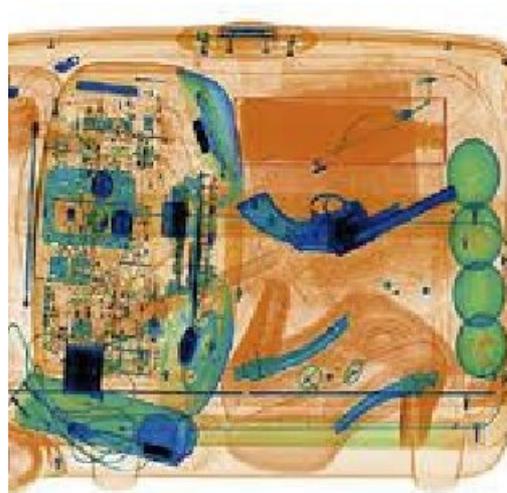

Figure 1. An example of an X-ray image of baggage with a dangerous object.

In Figure 1, it is easy to spot firearms. However, detailed thinking about the objects in Figure 1 provides the following conclusions. First of all, the definition of the image is quite small, and, secondly, the image itself has a specific nature when perceived by a person.

For further convenience and in order to train the verification network for pattern recognition, image preprocessing procedures were carried out, including the selection of only areas with dangerous objects. As part of this selection, five classes of hazardous objects were identified and 40 to 50 images were collected for each. Figure 2 shows examples of images for each selected class.

Figure 2. Examples of images for dangerous objects.

In Figure 2, from left to right, shockers, ammunition, grenades, firearms, and steel arms are presented.

Figure 3 shows the distribution of training images by class for the recognition task.

From Figure 3, it follows that a balanced dataset was prepared with a fairly small number of objects. As for processed images, the truth is that the good and bad pattern recognition in the form in which dangerous objects are presented in Figure 1 is possible only at the post-processing stages, where it is planned to use the recognizer. In addition, the objects are visually different.
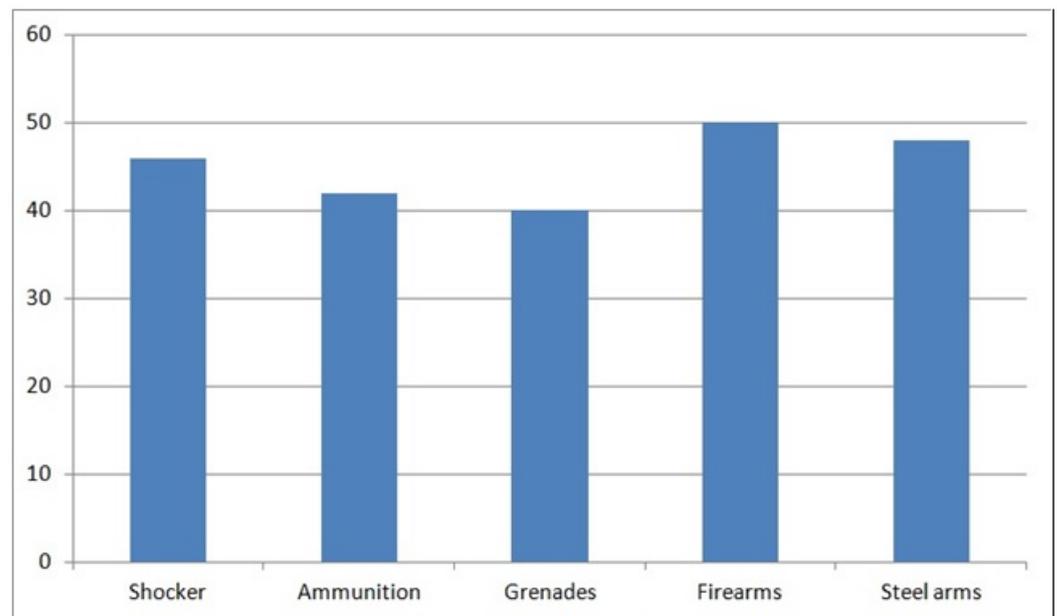
Figure 3. Distribution of objects of interest.

For images such as the one shown in Figure 1, manual labeling was performed using the CVAT tool [11] and RoboFlow [12]. Image markup example for the image from Figure 1 is shown in Figure 4. The accompanying files store the coordinates of the bounding box needed to train the neural networks.

Figure 4. Markup example.

The number of images in the training set was 1200. The remaining 300 images were left for testing. Table 1 shows the distribution of objects in the images of the training and test sets.

Despite the fact that the analysis of the data in Table 1 shows that the set is not balanced enough, it was decided to train on such data.

Next, let us consider the following approaches. They are very useful and applied to detect dangerous objects in images.

Table 1. Number of dangerous objects of each class.

| Class	Train, N = 1200 Samples	Test, N = 300 Samples
Shocker	122	34
Ammunition	109	28
Grenades	135	31
Firearms	180	46
Steel arms	166	37

3. Methods for Detecting Dangerous Objects in Images

As noted earlier, when describing the original images, the level of training was allocated for the detection task and the recognition task. This is due to the fact that more convenient images are formed when solving the recognition problem.

The convolutional neural network VGG-19 [13] was chosen as a model for recognition. The architecture of such a network is a convolutional network of 19 convolution blocks, pooling, trained under the ImageNet dataset. The transfer learning method was used with the following hyperparameters: batch size is 16, optimizer is ADAM, learning rate is 0.00001, activation function is softmax, loss function is cross-entropy.

The YOLOv5 architecture network [14] was used to train the solution to the detection problem. This network belongs to single-pass detectors, i.e., sets the class and bounding box for objects in one image processing pass. One more attractive thing about architecture is that the neural network is also convolutional. At the same time, the weights are adjusted in accordance with the COCO dataset, which includes 80 classes. The default hyperparameters are used.

Figure 5 shows the final scheme for processing new images.
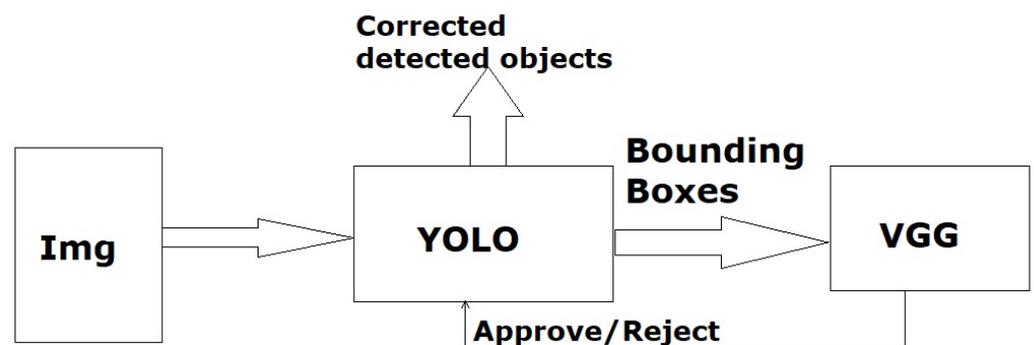
Figure 5. Baggage image processing scheme.

From Figure 5, it is clear that the transfer-trained YOLO model is used as the base detector. However, after the initial analysis, the images inside the bounding boxes are transferred to the transfer-trained VGG network. In this network, predictions are adjusted for one of the five classes according to the original dataset mentioned above.

Such an approach also has some drawbacks. For example, a special pre-trained model such as the VGG network in the diagram of Figure 5 is not capable of detecting non-hazardous items. However, this approach, taking into account the fact that the detection recall indicator is quite important in such a task, helps to detect more objects.

Next, let us consider the results obtained.

4. Results and Discussion

Since the problem of detecting dangerous objects has a higher priority than the error of missing the target, it is necessary to use the mAR metric as a comparison characteristic. The mean average recall is calculated as the average recall for each class in each image.

One of the properties of the developed approach is that training and inference took place on an NVIDIA GeForce GTX 1060 video card.

Figure 6 shows an example of image processing using the trained model.
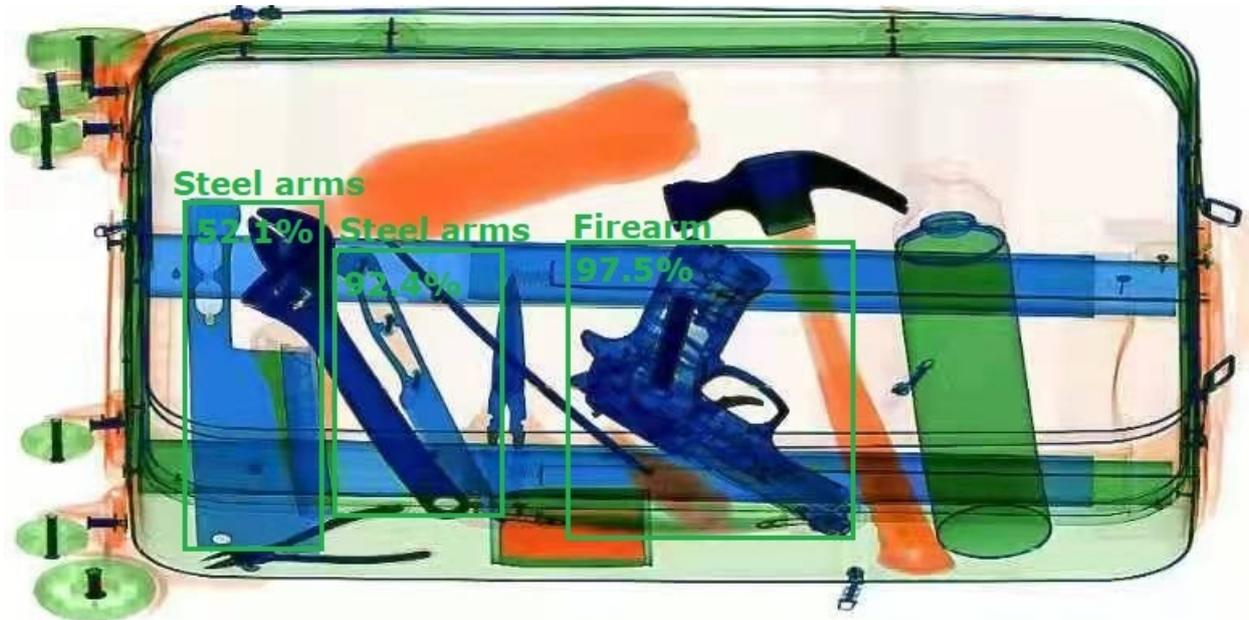
Figure 6. Processing results.

From Figure 6, it is clear that two steel arms and one firearm were found with class probabilities of 52.1%, 92.4%, and 97.5%, respectively.

Let us make a detailed comparison of research in terms of the proposed algorithm with other optical image-detecting methods transferred to X-ray images. The results for the mean average recall metric are presented in Table 2.

Table 2. Recall comparison by models.

Model	mAR
YOLOv5	0.843
SSD	0.672
DETR	0.697
Ours	0.871

Let us look at Table 2. It is interesting because the results show that the proposed model, because of using an additional level of verification based on the VGG network, allows us to increase the result compared to YOLO by 3%.

Finally, let us consider the completeness of each of their objects, as presented in Table 3.

Table 3. Recall comparison by models.

Class	Average Recall
Shocker	0.912
Ammunition	0.714
Grenades	0.839
Firearms	1
Steel arms	0.892

It can be seen from the presented results that firearms are detected in all cases. The worst of these things are those with the definition of ammunition.

5. Conclusions

This paper proposes a new algorithm for detecting dangerous objects on X-ray images of luggage, combining YOLO, and VGG. This approach provided an increase in the mean

average recall of 3%. In the future, it is planned to use augmentation and tuning of model hyperparameters to obtain higher recall values.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Janssen, S.; van der Sommen, R.; Dilweg, A.; Sharpanskykh, A. Data-Driven Analysis of Airport Security Checkpoint Operations. *Aerospace* **2020**, *7*, 69–74. [CrossRef]
2. Kim, M.H.; Park, J.W.; Choi, Y.J. A Study on the Effects of Waiting Time for Airport Security Screening Service on Passengers' Emotional Responses and Airport Image. *Sustainability* **2020**, *12*, 10634. [CrossRef]
3. Asmer, L.; Popa, A.; Koch, T.; Deutschmann, A.; Hellman, M. Secure rail station—Research on the effect of security checks on passenger flow. *J. Rail Transp. Plan. Manag.* **2019**, *10*, 9–22. [CrossRef]
4. Janssen, S.; Berg, A.; Sharpanskykh, A. Agent-based vulnerability assessment at airport security checkpoints: A case study on security operator behavior. *Transp. Res. Interdiscip. Perspect.* **2020**, *5*, 1–14. [CrossRef]
5. Andriyanov, N.A.; Dementiev, V.E.; Tashlinskii, A.G. Detection of objects in the images: From likelihood relationships towards scalable and efficient neural networks. *Comput. Opt.* **2022**, 46, 139–159. [CrossRef]
6. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers. *arXiv* **2020**, arXiv:2005.12872. Available online: https://arxiv.org/abs/2005.12872 (accessed on 27 July 2022).
7. Andriyanov, N.A.; Volkov, A.K.; Volkov, A.K.; Gladkikh, A.A. Research of recognition accuracy of dangerous and safe X-ray baggage images using neural network transfer learning. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *1061*, 012002. [CrossRef]
8. Hättenschwiler, N.; Michel, S.; Kuhn, M. A first exploratory study on the relevance of everyday object knowledge and training for increasing efficiency in airport security X-ray screening. *IEEE ICCSTAt* **2015**, *49*, 12–25.
9. Andriyanov, N.A.; Volkov, A.K.; Volkov, A.K.; Gladkikh, A.A.; Danilov, S.D. Automatic X-ray image analysis for aviation security within limited computing resources. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *862*, 052009. [CrossRef]
10. Hassan, T.; Shafay, M.; Akçay, S.; Khan, S.; Bennamoun, M.; Damiani, E.; Werghi, N. Meta-Transfer Learning Driven Tensor-Shot Detector for the Autonomous Localization and Recognition of Concealed Baggage Threats. *Sensors* **2020**, *20*, 6450. [CrossRef] [PubMed]
11. Computer Vision Annotation Toolbox. Available online: https://cvat.ai/ (accessed on 28 July 2022).
12. Roboflow. Available online: https://app.roboflow.com/ (accessed on 28 July 2022).
13. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:1409.1556. Available online: https://arxiv.org/abs/1409.1556 (accessed on 29 July 2022).
14. YOLOv5. Available online: https://github.com/ultralytics/yolov5 (accessed on 29 July 2022).