Computer Vision-Based Hand Baggage Inspection System

Marcelo Albergaria Paulino Fernandes Ferreira[1*], Oberdan Pinheiro Rocha[1]
[1]SENAI CIMATEC University Center; Salvador, Bahia, Brazil

Object detection is a crucial task in computer vision. It brings two essential tasks, classifying and locating the object in the image or video. This task is performed by specialized human operators, which requires the knowledge and experience for reliable detection. However, the high volume of work often drives the operators to exhaustion, causing a low prevalence of targets. This work aims to build a model using a Deep Neural Network (Darknet) and a single pass object detection method (YOLOv4) for the detection of handguns, shurikens, and blade shavings in hand luggage, using the data set GDXray. The YOLO algorithm model identified different items that could impact security, which could be a tool that makes passenger boarding faster and safer.
Keywords: Object Detection. YOLO. Neural Network. Computer Vision. Deep Learning.

Introduction

Air transport is essential for the economic and social development of a country, and due to the possibility of fast connections, it facilitates the movement of people and goods and boosts commercial activities and tourism [1].

The Brazilian National Civil Aviation Agency (ANAC) was created to oversee civil aviation activities and aeronautical and airport infrastructure. Established in 2005, it began operating in 2006 and is a federal autarchy with a special regime linked to the Ministry of Infrastructure.

In 2021, according to ANAC [2], approximately 62,583,158 passengers were transported. Therefore, searching for processes and tools that help make air travel faster and safer is essential, mainly due to the increased operations in the Brazilian air market. Therefore, it is possible to highlight the large amount of flow of passengers and luggage that has dramatically lengthened the response time of security agents at boarding. Furthermore, due to the importance of this transport, the concerns about air security are severe. Therefore, one of the measures adopted at airports is the X-ray screening of passengers' bags at security checkpoints, being a key component to ensure that prohibited items do not get into airplanes.Currently, this procedure is carried out by inspectors who are specialists in baggage tracking, who, during screening at checkpoints, visually inspect the X-ray images of passengers' bags to decide whether they are harmless or contain prohibited items.

Visual search challenges include:

- Low prevalence of targets,
- Variation in target visibility,
- Searching for a set of unknown targets,
- The possible presence of multiple targets,
- External variables such as job satisfaction.

Computer vision techniques can minimize this problem automatically, quickly, and reliably, enabling the detection and classification of prohibited and potentially dangerous objects for security in passengers' hand luggage. Such a solution can maintain alertness and improve human operators' detection and response time, thus ensuring air transport safety.

According to Butler [3], the ideal system would have a high processing rate, low initial and operational costs, and low rates of false readings (false-positive, false-negative). A false positive reading occurs when the system identifies an object as being dangerous when it is not. This point becomes a problem because these false readings lead to other decision-making that slows down processes, annoys passengers, and

Received on 26 March 2023; revised 18 May 2023.
Address for correspondence: Marcelo Albergaria Paulino Fernandes Ferreira. Rua Professor José Andrea, 81, Vila Laura. Zipcode: 40270-390. Salvador, Bahia, Brazil. E-mail: Marcelo_albergaria@hotmail.com. DOI 10.34178/jbth.v6i2.295.

J Bioeng. Tech. Health 2023;6(2):143-150

costs money. On the other hand, the false negative is just as significant as it occurs when the dangerous object is present, but the system cannot detect it. This work compares five primary means of baggage inspection, favoring X-ray systems for their great processing capacity but indicating their high cost for implementation and high false negative rates.

Computer vision techniques can minimize this problem automatically, quickly, and reliably, enabling the detection and classification of prohibited and potentially dangerous objects for security in passengers' hand luggage. Such a solution can maintain alertness and improve human operators' detection and response time, thus ensuring air transport safety.

According to Butler [3], the ideal system would have a high processing rate, low initial and operational costs, and low rates of false readings (false-positive, false-negative). A false positive reading occurs when the system identifies an object as being dangerous when it is not. This point becomes a problem because these false readings lead to other decision-making that slows down processes, annoys passengers, and costs money. On the other hand, the false negative is just as significant as it occurs when the dangerous object is present, but the system cannot detect it. This work compares five primary means of baggage inspection, favoring X-ray systems for their great processing capacity but indicating their high cost for implementation and high false negative rates.

This work aims to build a model using Darknet and the YOLOv4 algorithm to detect possible threat objects through X-Ray images in the initial screening of passenger luggage. The detection of objects of interest will use the GDXray dataset, a single-pass object detection method to locate and classify revolvers, shurikens, and blade shavings. These objects pose a possible threat to safety when boarding passengers.

Related Literature

Hassan and colleagues [11] present a deep multiscale structure tensor-based framework that can automatically extract and recognize regular and suspicious items, independent of their position and orientation, from multivendor X-ray scans. The proposed framework is unique as it intelligently extracts each object by iteratively picking transactional information based on contours of different orientations. Furthermore, it uses only a single Feedforward Convolutional Neural Network for recognition. The proposed structure was tested on two publicly available datasets containing 1,067,381 X-ray scans and uses the pre-trained ResNet architecture for object recognition.

Akcay and colleagues [12] considered using deep convolutional neural networks with transfer learning for image classification and detection problems in the context of X-ray luggage security images. A pre-trained CNN was used in generalized image classification tasks, aiming to overcome the problem of the limited availability of examples. Using YOLOv2, using images of size 544x544 as input, a mAP of 88.5% was achieved for a six-class detection problem. The same approach, with an input size of 416x416, produced a 97.4% mAP for the two-class firearm detection problem. Akcay and Breckon [13] aim to review computerized X-ray security imaging algorithms by taxonomically fielding conventional machine learning and contemporary deep learning algorithms. The proposed taxonomy subcategorizes deep learning approaches into supervised and unsupervised learning, focusing on object classification, detection, segmentation, and anomaly detection tasks. It also explores well-established X-ray datasets and provides a benchmark of performance.

This project is an alternative to the object detection problem, as we apply a single Deep Neural Network (Darknet) to the complete image. This network divides the image into regions and predicts bounding boxes and probabilities for each region, using the YOLOv4 algorithm that proved to be faster and more accurate than its previous versions.

Materials and Methods

Detection Algorithms

Several methods seek to solve the object detection problem. The Harr Cascade technique was considered the first approach to achieve satisfactory results that could be implemented in real applications. With the rise of Deep Learning, methods based on Convolutional Neural Networks (CNN) were developed, such as R-CNN, Fast R-CNN, SPP-net, and YOLO. Convolutional Neural Networks recognize patterns in data, usually organized in layers, each with weights that are adjusted during the training phase, allowing the network to adapt to the problem and the given dataset. They emerged from studying the brain's visual cortex and have been used in image recognition since the 1980s. In recent years, thanks to the increased computational power, and the amount of training data available, CNNs have achieved superhuman performance in some complex visual tasks.

You Only Look Once: YOLO

YOLO [4] uses Deep Learning and Convolutional Neural Networks for object detection, one of the fastest algorithms being the object detection problem in images. The image is divided into an SxS grid of cells to perform detection. Each of the Cells will predict "N" possible bounding boxes. Most of these bounding boxes will have a very low probability, so the algorithm deletes the boxes below the minimum probability threshold. Bounding boxes with a probability above a threshold are selected to locate the object within the image.

In YOLO, a single convolutional network predicts each detected object's bounding boxes and the class membership probabilities. Unlike classification, which only seeks to predict the class present in the image, object detection, in addition to predicting the class, also needs to identify the object's location in the image, using a Convolutional Neural Network as a feature extractor.

Darknet

YOLO uses a deep neural network called Darknet to implement object detection. This framework is open source, written in the C language, and has GPU support.

Dataset

The data set used in the work is the GDXray [5], provided by the Department of Computer Science of the Catholic University of Chile. The images are organized in a public database that can be used free of charge but for educational purposes only. The database includes five groups of X-ray images: castings, welds, luggage, natural objects, and configurations (Figure 1). In addition, we use 793 X-ray images that can be used for the automatic detection of guns, shurikens, and steel blades to shave.

Proposed Model

The Darknet neural network was implemented to generate embeddings in the Google Collaboration (Colab) environment using a free account with access to the GPU resource. In the accessible version of Colab, the access to GPUs is very limited, and the notebooks run for a maximum of 12 hours.

The pre-processing step involves data standardization and is fundamental for applying Artificial Intelligence techniques. So in this project, we standardized the images in PNG (Portable Network Graphic) format and separated the images into two main sets "Train" and "Test". The Train set contains 713 images that will be used exclusively for network training, while the Test file contains 80 images that will be used exclusively for network testing. After properly separating the dataset, we move to

Figure 1. Sample GDX-ray baggage.

the image annotation stage with their respective classes. We used LabelImg [6], an open-source graphic image annotation tool that supports the task's YOLO, CREATEML, and PASCAL VOC formats. This step aims to help models learn patterns from the data set and use them for future predictions, the work was dedicated to detecting three classes (revolver, shuriken, laminaBarbear), and we used LabelImg to create bounding boxes in the objects of interest, these boxes contain the characteristics of the class, coordinates, height, and width. After annotating the images, we trained and tested the object detector using YOLOv4 and Darknet, performing three training rounds varying the learning rate and momentum of the model. The algorithm was configured with three values for the learning rate (0.001, 0.00261, 0.1), for the momentum (0.9 and 0.949), and the number of epochs in 4700. These parameters have already been used in the literature for training object detection networks using YOLO. The validation metrics are used to analyze the model's quality, bringing the performance information of the experiments [7]. We use a set of these metrics to gauge how far the model is from perfect detection. In this article, we used the metrics Precision, Recall, F1-score, Confusion matrix, Intersection over union (IoU), and Mean Average Precision (mAP). The Confusion Matrix indicates how many examples there are in the true positive (TP), false negative (FN), false positive (FP), and true negative (TN) groups. Allowing one to easily view how many examples were classified correctly and wrongly in each class. Precision is the metric that evaluates the number of true positives over the sum of all positive values.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (1)$$

It is common to consider the combination of Precision and Recall, also known as sensitivity, as it measures what fraction of the positives our model identifies [8].

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive + False Negative}} \quad (2)$$

The F1-score measure is the harmonic mean of accuracy and sensitivity. This metric evaluates the method's ability to detect positive results successfully.

$$f1 = 2 \text{ x } \frac{\text{precision x sensitivity}}{\text{precision + sensitivity}} \quad (3)$$

Mean Average Precision (mAP) and Intersection over Union (IoU) metrics are popular for checking the performance of object detectors. The IoU compares the bounding boxes with the detected boxes, returning a normalized score. For example, we can calculate the mAP by taking the average of all the average Precision of each class, that is, the average of the area of the Precision-Recall curve of each class.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k \quad (4)$$

AP_k = the AP of class k
n= the number of classes

Result and Discussion

Our study has three possible scenarios that differ in the learning rate and momentum. We have YOLO configured with a learning rate of 0.001 and the momentum term 0.949, providing the confusion matrix in the first scenario (Table 1).

Table 1. Confusion matrix (Scenario 1).

Learning Rate = 0.001	True Positive	False Negative
Handgun	44	4
Shuriken	27	2
Blade shaving	32	3

We modified the training parameters in the second scenario, 0.00261 and 0.9, respectively. Finally, Table 2 presents the confusion matrix.

Table 2. Confusion matrix (Scenario 2).

Learning Rate = 0.00261	True Positive	False Negative
Handgun	44	4
Shuriken	27	2
Blade shaving	32	3

In the third scenario, YOLO failed to converge the training and provided no performance metrics. Training parameters were 0.1 for learning rate and 0.5 for momentum. Table 3 provides metrics for all scenarios. Given the metrics presented by the models, we noticed that the algorithm of the first scenario performs better in the object detection task.

Figure 2 presents a graph demonstrating the evolution of training concerning the number of periods. We also submitted images of luggage that were separated before training to validate the model. Finally, we present the results in Figures 3 to 6.

Conclusion

This work demonstrated the applicability and viability of using YOLOv4 to detect objects in hand luggage at airports, which could be a tool that makes passenger boarding faster and safer, having as a driver the growing demand of the air market that indicates an average growth in an optimistic scenario of 6.14% until 2037. Applying the Deep Learning model provided mAP results more significant than 95% in the two main scenarios (Scenario 1 = 98.79%%, Scenario 2 = 97.62%) (Table 3).

From the analyses, YOLOv4 is a good algorithm for object detection, mainly if we consider the challenges of the visual search, which include a low prevalence of targets, the variation in the target's visibility, the presence of multiple targets, and external variables. Furthermore, despite the need for more significant computational resources, the technique had a performance gain in inference and assertiveness (Figure 2), in addition to the option of applying more effective techniques to the processing on GPUs.

Among the techniques that could not be applied is using other versions of the YOLO algorithm, such as version 5, created in 2020, and version 7 of the same algorithm. These techniques could increase the performance and predictability of the model analyzed in the work. However, studying these algorithms remains a suggestion for exploring future work.

Table 3. Scenario performance metrics.

Taxa de Aprendizado	Momentum	Epocas	Precision	Recall	F1-score	Average loU	mAP
0.001	0.949	4700	0.92	0.99	0.95	73.67%	98.79%
0.00261	0.9	4700	0.94	0.96	0.95	75.24%	97.62%

Figure 2. Training evolution chart.

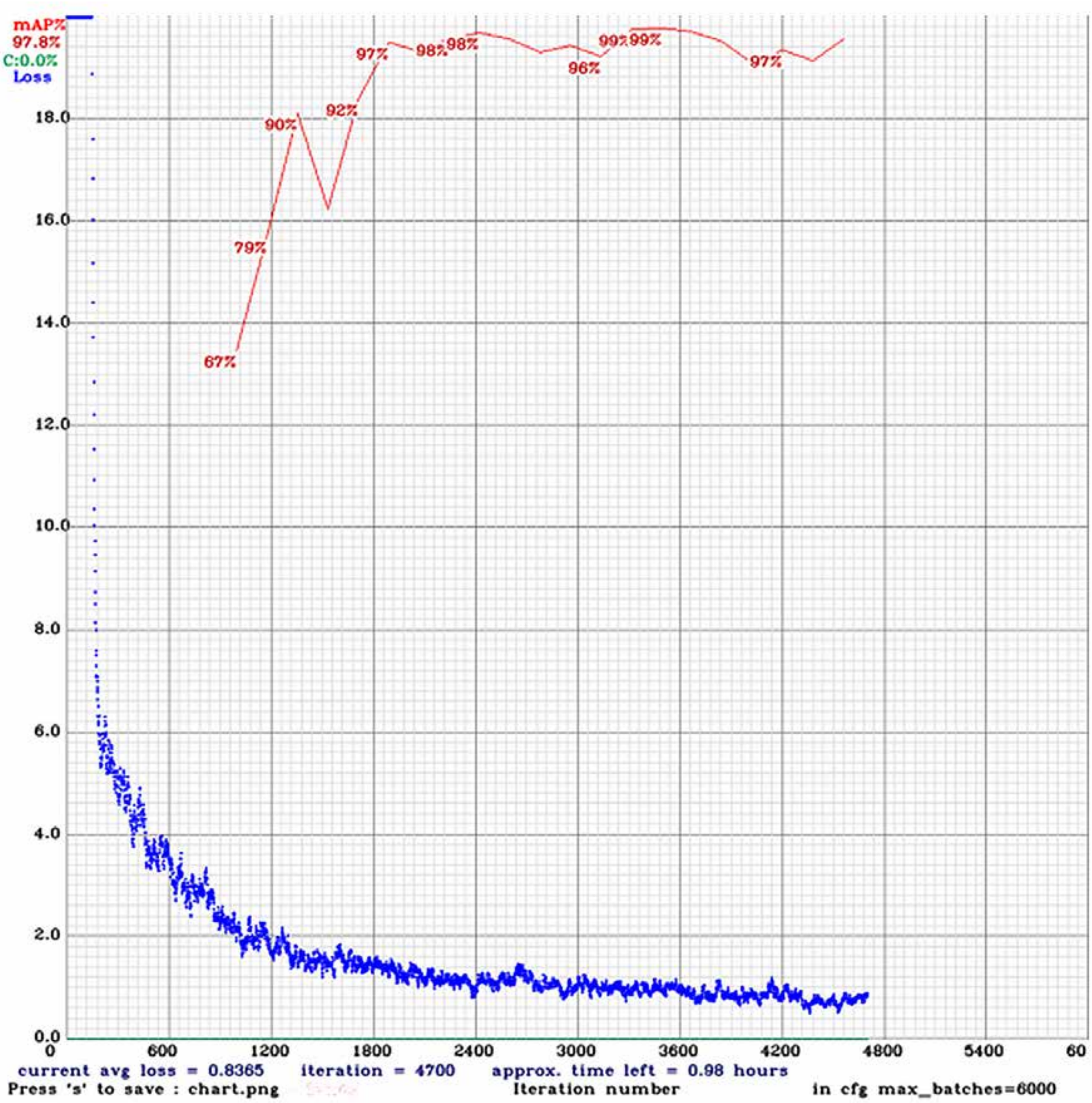


Figure 3. Handgun and shuriken detection.

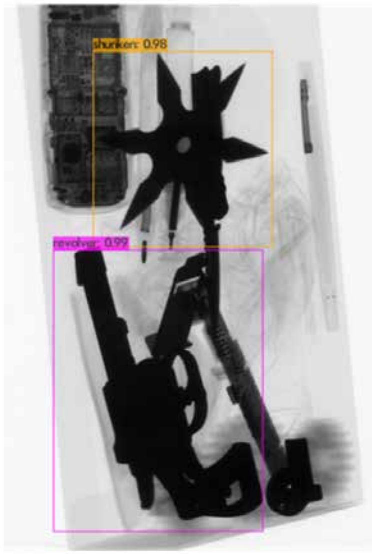


Figure 4. Shuriken and blade shaving detection.

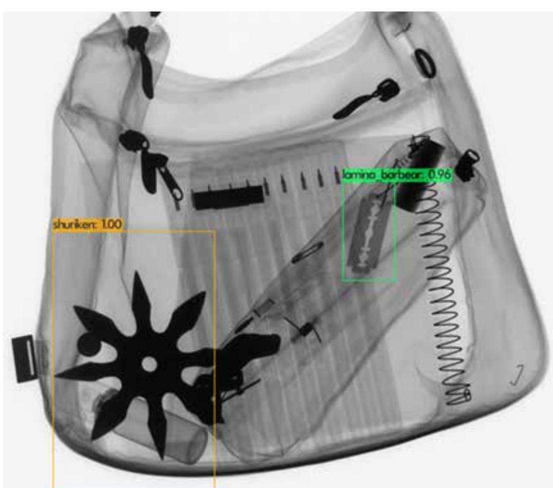


Figure 5. Detection of a blade shaving in a wallet.

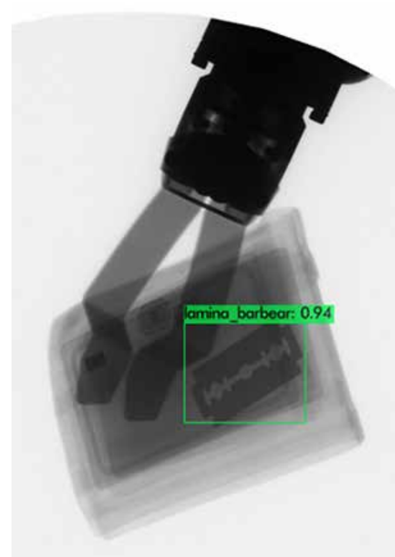
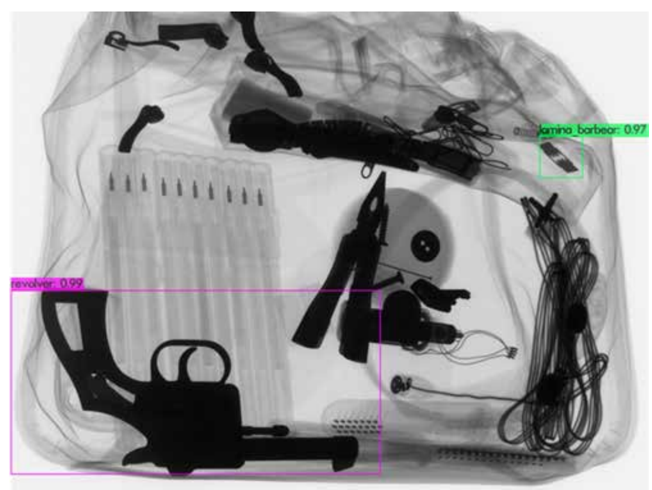


Figure 6. Handgun and blade shaving detection.



References

1. Ontl ON de Transporte e L. A importância do transporte aéreo para o Brasil. 2022.
2. ANAC. Anuário do transporte aéreo. Sumário Executivo 2021. 2022.
3. Butler RWPV. Rethinking Checked-baggage screening. Reason Public Policy Institute 2002.
4. Bochkovskiy A, Wang C-Y, Liao H-Y M. YOLOv4: Optimal speed and accuracy of object detection. 2004 arXiv:2004.10934 [cs.CV]. https://doi.org/10.48550/arXiv.2004.10934.
5. Mery D, Riffo V, Zscherpel U et al. GDXray: The database of X-ray images for nondestructive testing. J Nondestruct Eval 2015;34:42. https://doi.org/10.1007/s10921-015-0315-7.
6. Tzutalin. LabelImg. Git code (2015). https://github.com/tzutalin/labelImg.
7. Géron,Aurélien. Mãos à Obra Aprendizado de Maquina com Scikit-Learn & TensorFlow. Altas Books, 2019.
8. Grus J. Data Science From Scratch: First Principles with Python. Altas Books, 2016.
9. Russel SJ. Inteligência Artificial. Elsevier, 2013.

10. Haykin S. Redes Neurais: Princípios e Pratica. Bookman Editora, 2007.
11. Hassan TK, Salman & Akcay, Samet & Bennamoun, Mohammed & Werghi. Deep CMST framework for the autonomous recognition of heavily occluded and cluttered baggage items from multivendor security radiographs. Naoufel 2019. Available: https://deepai.org/publication/deep-cmst-framework-for-the-autonomous-recognition-of-heavily-occluded-and-cluttered-baggage-items-from-multivendor-security-radiographs.
12. Akcay S, Kundegorski ME, Willcocks CG, Breckon TP. Using deep convolutional neural network architectures for object classification and detection within X-ray baggage security imagery. IEEE Transactions on Information Forensics and Security 2018;13(9): 2203-2215.
13. Akcay S, Breckon T. Towards automatic threat detection: A survey of advances of deep learning within X-ray security imaging. Pattern Recognition 2021. https://doi.org/10.1016/j.patcog.2021.