SEBHA UNIVERSITY JOURNAL VOL.03 NO. 1 2024
DOI: 10.51984/SUCP.V3I1.3817

مجلة جامعة سبها
Sebha University Journal

Journal homepage: https://sebhau.edu.ly/journal/index.php/CAS

Improving Content Recognition of X-ray Images of Baggage Using Deep CNN in Customs Administration

*Mabroukah Amarif[1] and Abraheem Jabur [2]

[1]Department of Information Systems, Faculty of Information Technology, Sebha University, Libya
[2]Department of Computer Sciences, Faculty of Sciences, Sebha University, Libya

Keyword:
content recognition
X-ray images
deep convolutional neural network

A B S T R A C T

With the increasing security threats and the need for ensuring the safety of the movement of passengers and goods through countries, customs inspections at airports and border crossings are more important than ever. Customs departments currently rely mainly on visual inspection to detect dangerous and prohibited items in X-ray images of baggage. However, this traditional approach faces many challenges such as the possibility of human errors and the long time spent in the inspection process, in addition to the possibility of human errors in analyzing and interpreting complex X-ray images. Therefore, the need has emerged to use advanced artificial intelligence techniques to improve the speed and accuracy of detecting prohibited materials. This research seeks to improve the recognition of the content of X-ray images of baggage in customs administration using deep convolutional neural network models of artificial intelligence. A number of experiments have been carried out to improve the accuracy and detection of dangerous materials, especially firearms, sharp materials and knives, through automated X-ray images of baggage using three models of convolutional neural networks which are VGG16, ResNet50, and InceptionV3. The aim of using different models is to obtain the highest accuracy among them. The three models have been trained and tested using a huge SIXray dataset, which specializes in X-ray images of baggage. The results show that the VGG16 model has outperformed the others with a high accuracy exceeding 96%. The contribution of this research is to enhance the efficiency of customs inspection operations and improving security and safety levels through accurate and rapid automated detection of dangers materials to prevent them from being entered to the country.

تحسين التعرف على محتوى صور الأشعة السينية للأمتعة باستخدام الشبكات العصبية التلافيفية العميقة في إدارة الجمارك

*مبروكه معيوف[1] و ابراهيم جبر[2]

[1]قسم علوم الحاسب، كلية تقنية المعلومات، جامعة سبها، ليبيا
[2]قسم علوم الحاسب، كلية العلوم، جامعة سبها، ليبيا

الكلمات المفتاحية
التعرف على المحتوى
صور الأشعة السينية
الشبكة العصبية التلافيفية العميقة

الملخص

مع تزايد التهديدات الأمنية والحاجة إلى ضمان سلامة حركة الركاب والبضائع عبر الدول، أصبحت عمليات التفتيش الجمركي في المطارات والمعابر الحدودية أكثر أهمية من أي وقت مضى. تعتمد إدارات الجمارك حالياً بشكل أساسي على الفحص البصري للكشف عن المواد الخطرة والمحظورة في صور الأشعة السينية للأمتعة. إلا أن هذا النهج التقليدي يواجه العديد من التحديات مثل احتمال حدوث أخطاء بشرية والوقت الطويل الذي تستغرقه عملية التفتيش، بالإضافة إلى احتمال حدوث أخطاء بشرية في تحليل وتفسير صور الأشعة السينية المعقدة. ولذلك ظهرت الحاجة إلى استخدام تقنيات الذكاء الاصطناعي المتقدمة لتحسين سرعة ودقة الكشف عن المواد المحظورة. يسعى هذا البحث إلى تحسين التعرف على محتوى صور الأشعة السينية للأمتعة في الإدارة الجمركية باستخدام نماذج الشبكة العصبية التلافيفية العميقة للذكاء الاصطناعي. تم إجراء عدد من التجارب لتحسين دقة وكشف المواد الخطرة، وخاصة الأسلحة النارية والمواد الحادة والسكاكين، من خلال صور الأشعة السينية الآلية للأمتعة باستخدام ثلاثة نماذج من الشبكات العصبية التلافيفية وهيVGG16 ، وResNet50، و InceptionV3. الهدف من

*Corresponding author:
E-mail addresses: mab.imaref@sebhau.edu.ly ,(A. Jabur) abra.jabur@sebhau.edu.ly
Article History : Received 19 February 2024 - Received in revised form 08 May 2024 - Accepted 25 May 2024

استخدام النماذج المختلفة هو الحصول على أعلى دقة فيما بينها. تم تدريب النماذج الثلاثة واختبارها باستخدام مجموعة بيانات SIXray الضخمة، والمتخصصة في صور الأشعة السينية للأمتعة. وأظهرت النتائج تفوق نموذج VGG16على النماذج الأخرى بدقة عالية تجاوزت 96%. وتتمثل مساهمة هذا البحث في تعزيز كفاءة عمليات التفتيش الجمركي وتحسين مستويات الأمن والسلامة من خلال الكشف الآلي الدقيق والسريع للمواد الخطرة لمنع دخولها إلى الدولة.

1. Introduction

Security screening operations at airports and border crossings are a vital element for ensuring the safety and security of passenger and cargo movement. Customs administrations play a pivotal role in these operations, as they are responsible for monitoring and controlling the movement of goods and detecting hazardous and prohibited materials that may pose a threat to public health and the environment [1, 2], including explosives, weapons, and sharp objects. The current detection of these materials primarily relies on the use of X-ray imaging devices, followed by visual inspection by customs officers. However, this traditional approach faces numerous challenges and limitations [3, 1].

The possibility of human errors in analyzing and interpreting complex X-ray images is one of the most prominent challenges facing inspectors. Additionally, in cases of uncertainty about the contents of bags, inspectors are required to conduct thorough manual inspection, which is time-consuming and labor-intensive. These factors lead to slowdown of the inspection process an decrease of efficiency, which may jeopardize the safety of travelers and the security of facilities [1, 4].

According to these challenges, it has become necessary to leverage advanced artificial intelligence techniques to accelerate and improve the accuracy of hazardous material detection. Artificial intelligence techniques, particularly Convolutional Neural Networks (CNNs), are distinguished by their exceptional ability to process data and analyze images rapidly and with high accuracy, by extracting key patterns and features and efficiently recognizing objects within images [4, 5, 6].

This study aims to employ the capabilities of Convolutional Neural Networks in analyzing X-ray images of bags to automatically detect hazardous materials with high accuracy. The research focuses specifically on designing a specialized CNN model for recognizing two main risk categories: firearms and knives, based on the specialized SIXray dataset.

The research seeks to design and develop an innovative model based on Convolutional Neural Networks to process and analyze X-ray images of bags and cargo, with the aim of improving customs screening operations and enhancing accuracy and efficiency in detecting hazardous and prohibited materials, by leveraging advanced artificial intelligence and machine learning techniques.

2. Related Works

To achieve the study's objective of developing an effective Convolutional Neural Network (CNN) model for recognizing knives and firearms in X-ray images of bags, it is crucial to explore relevant previous research and studies in this field. Understanding prior work and the techniques employed helps identify gaps and challenges that have not yet been addressed, and sheds light on best practices and promising models that can be built upon. To this end, several important studies that addressed the detection of prohibited materials in X-ray images of bags using artificial intelligence and deep learning techniques were reviewed.

The study [8] presents a large-scale X-ray security inspection benchmark called SIXray, which aims to provide a large dataset of X-ray security images. These images are crucial for detecting prohibited items inside bags and belongings. This study allows researchers and developers to use this benchmark to test and develop effective techniques for detecting prohibited items. SIXray represents an important source of X-ray images that can be used for training and testing models to improve the performance of X-ray inspection systems. It can be utilized in airport, port, and other sensitive facility security applications. This study represents a significant contribution to the field of security and the development of techniques for detecting prohibited items using X-ray technology.

The study [5] addressed the problem of automatic detection of firearms and their components in X-ray images of luggage, which is one of the major challenges in airport security. The researchers used three advanced deep learning object detection and localization models: Faster R-CNN, Mask R-CNN, and RetinaNet. These models were trained and tested on two datasets of X-ray luggage images: the first (Dbf2) contained images of firearms and their components, and the second (SIXRay) contained images of firearms only. The results showed that the Faster R-CNN model achieved the highest average accuracy of 91% and 88% on the Dbf2 dataset for detecting firearms and their components, respectively. Meanwhile, the RetinaNet model achieved the highest accuracy of 0.92% on the SIXRay dataset for detecting firearms. The false alarm rate was very low in all three models, indicating their efficiency in distinguishing firearms. The study concluded that intelligent techniques and deep learning can be effectively used for detecting firearms in X-ray images of luggage and recommends further research to develop an intelligent system for use in airports.

The study [4] focused on using Deep Convolutional Neural Networks (DCNNs) for object classification and detection within X-ray security images of bags. This study is based on using deep networks to extract distinctive features from images and improve classification and detection accuracy. Different designs of DCNNs are applied to achieve better performance in object classification and detection within X-ray images. The study aims to develop effective models that can be used in security applications and X-ray bag inspection. This study reflects the importance of using deep techniques to improve the efficiency of inspection and detection systems using X-ray technology in fields such as airport and port security, contributing to the development of performance enhancement techniques in security domains.

The study [7] provided a review of the progress made in the field of automatic threat detection using deep learning in X-ray security imaging techniques. The main objective of the study was to review the techniques and developments applied in this field that rely on deep neural networks. The study highlighted how deep learning has advanced in security applications and threat detection using X-ray technology, providing an overview of recent innovations and research, and how detection techniques have been improved over time using deep learning.

Through reviewing these studies, it is evident that many researchers have made significant efforts to develop automated techniques for detecting prohibited materials in X-ray images of bags using artificial intelligence and deep learning techniques. However, there is still an opportunity to further improve classification accuracy and detection speed. The current study aims to bridge this gap by designing a specialized CNN model that achieves superior performance in recognizing knives and firearms in particular.

2.1 Convolutional Neural Network Architectures

Convolutional Neural Networks (CNNs) have been at the forefront of deep learning advancements for image analysis tasks. Several influential CNN architectures have been developed over the years, each introducing innovative concepts and techniques to improve performance and efficiency. In this study, three prominent CNN models were explored and evaluated: VGG16, ResNet50, and InceptionV3. These models have been widely adopted in various computer vision applications and have contributed significantly to the field of deep learning for image recognition and classification [9].

VGG16: The VGG16 model was developed by researchers at the University of Oxford in 2014 [9] It is a deep convolutional neural network that consists of 16 weight layers. The network is characterized by its simplicity, using only 3x3 convolutional layers stacked on top of each other in increasing depth. Despite being simple, the VGG16 model achieved excellent results on the ImageNet dataset. The 16 layers are divided into five groups, with some groups having multiple convolutional layers and others containing a max-pooling layer to progressively reduce the spatial dimensions. The final layers are fully connected and perform the actual classification. The VGG16 model is known for its uniform architecture, where all convolutional layers use the same filter size and the same padding.

ResNet50: The ResNet50 model, introduced by Microsoft researchers in 2015 [10], is part of the ResNet (Residual Network) family of models. It addresses the vanishing gradient problem in deep neural networks by introducing skip connections or shortcut connections that bypass one or more layers. These skip connections allow gradients to flow more easily across layers during backpropagation, enabling the training of much deeper networks. The ResNet50 model consists of 50 layers and utilizes bottleneck blocks, where the input is projected into a lower-dimensional space, processed, and then projected back to the original dimension. This architecture helps reduce computational complexity and improves model performance [12].

InceptionV3: The InceptionV3 model is part of Google's Inception family of models, which were introduced in 2014 [11]. It is a deep convolutional neural network that incorporates several distinct architectural innovations. The most notable feature of InceptionV3 is the use of Inception modules, which apply multiple convolutional filters of different sizes on the same input to capture different levels of detail. These modules are repeated multiple times throughout the network, with intermittent max-pooling layers to reduce spatial dimensions. The Inception modules are designed to be computationally efficient and allow for increased depth and wider features without significantly increasing the number of parameters. InceptionV3 consists of 48 layers and incorporates techniques like batch normalization and label smoothing to improve training and generalization [12].

These models have been widely used in various computer vision tasks, including object recognition, image classification, and object detection. Their architectures and innovations have contributed to significant advancements in the field of deep learning for image analysis.

3. Dataset

In this study, we primarily relied on the SIXray benchmark dataset [8] as the source for training and testing data. The SIXray dataset is considered one of the largest and most comprehensive datasets available in the field of detecting prohibited materials using X-ray images of bags.

This dataset comprises over one million X-ray images of travelers' bags, captured using advanced multi-energy X-ray imaging systems. These images include realistic samples of a diverse range of hazardous and prohibited materials, such as firearms, knives, blades, explosives, and hazardous chemicals, as illustrated in Figure (1).

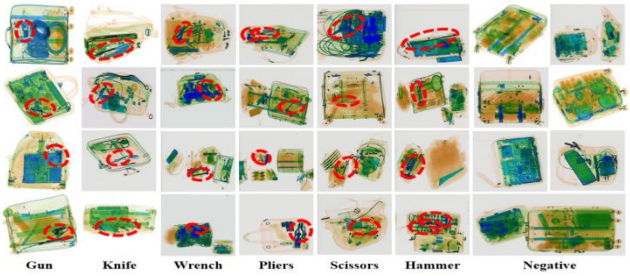

Figure 1: Categories of the SIXray dataset [8]

4. Methodology

An applied research methodology has been adopted by analyzing previous studies and identifying a CNN model for recognizing image contents, with the possibility of making appropriate modifications to achieve the highest level of accuracy. Then, the enhanced model will be evaluated through a case study, and the steps of the data mining methodology will be applied, including data collection, processing, model training, and testing. Finally, the results will be analyzed and interpreted using appropriate methods. The following steps illustrate the research approach, as shown in Figure 2.

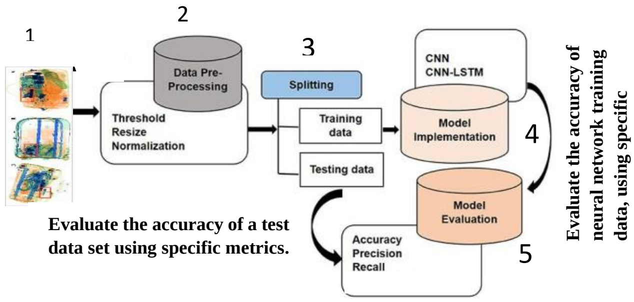

Figure 2: Research Methodology

The SIXray dataset [8], containing over one million X-ray images of travelers' bags, was loaded. The integrity of the loaded data and its alignment with the study's objectives were verified.

- **Data Processing and Preparation**

After loading the massive SIXray dataset containing over one million X-ray images of travelers' bags, extensive data processing was necessary to ensure the quality and suitability of the data for the model to be trained. The processing began with the application of a set of image enhancement techniques, such as contrast enhancement and noise removal, to improve the clarity of the images and highlight important details, aiding the model in better feature extraction.

Next, data reduction techniques like image resizing and color encoding were applied to reduce the data size and computational requirements for training and inference, without sacrificing important information in the images. The images were then automatically labeled using machine learning algorithms, categorizing them into two main classes: knives and firearms, which were the target classes in this study.

To ensure accurate labeling, further human verification was performed on the automatically labeled data, and any incorrect or anomalous data was removed.

- **Dataset Splitting**

The dataset was split into a training set (80%) and a test set (20%) using the hold-out method to ensure an accurate and fair evaluation of the model's performance on new data.

- **Model Building and Training**

A Convolutional Neural Network (CNN) model was designed and trained using the well-known VGG16 model for effective feature extraction [9]. The model was trained and tested on the processed SIXray dataset, focusing solely on knives and firearms. The performance of the VGG16 model was compared with other models such as ResNet and Inception to determine the most suitable model for achieving the study's objectives.

- **Model Performance Evaluation**

The model's performance will be evaluated using an independent test dataset, relying on metrics such as accuracy, precision, recall, and F1-score to measure the model's efficiency in accurately classifying X-ray images of knives and firearms.

5. Results

After following the methodology outlined above, three different Convolutional Neural Network (CNN) models were trained and tested on the SIXray dataset. The complete dataset contained 1,059,231 X-ray images of travelers' bags, out of which 8,929 images contained prohibited items. Among these prohibited items, there were 1,212 firearms and 1,366 knives, totaling 2,578 images of firearms and knives.

This subset of 2,578 images of firearms and knives was used exclusively for training the models. The subset was split into 2,062 images (80%) for the training set and 516 images (20%) for the test set.

The three models (VGG16, ResNet50, and InceptionV3) were trained on the training set for 50 epochs using a learning rate of 0.001 and a batch size of 32. Data processing techniques such as augmentation and contrast enhancement were applied to increase data diversity and improve model performance.

After completing the training, the performance of the three models was evaluated on the test dataset using the following metrics, as shown in Table (1).

Table (1) presents the model results.

Model	Accuracy	Precision	Recall	F1-score
VGG16	96.1%	95.2%	95.1%	95.6%
ResNet50	95.3%	93.8%	94.6%	94.2%
InceptionV3	95.7%	94.9%	94.8%	95.2%

The results show that the VGG16 model emerged as the top performer, outshining the other models across all evaluation metrics. It achieved remarkable results, boasting the highest accuracy of 96.1%, the highest precision of 95.2%, the highest recall of 95.1%, and the highest F1-score of 95.6%. The InceptionV3 model followed closely in performance, while the ResNet50 model lagged behind, exhibiting the lowest scores among the three models.

The superior performance of the VGG16 model in recognizing knives and firearms in X-ray bag images can be attributed to several key factors. Its architectural design is relatively simple and straightforward compared to the more complex ResNet and Inception models. The VGG16 model consists of a sequential stack of convolutional layers with increasing depth and small 3x3 filters. This design allows the model to learn robust and translation-invariant features from the input images, which is particularly beneficial for detecting and recognizing specific objects like knives and firearms.

The pre-trained weights of VGG16 on the ImageNet dataset, which contains a diverse range of object categories, provided a strong foundation for fine-tuning on the SIXray dataset. The lower convolutional layers learned general features like edges, textures, and shapes, which could be effectively transferred and adapted to the target domain of X-ray images, enhancing the model's performance.

With a depth of 16 convolutional layers, VGG16 strikes a balance between model complexity and the ability to capture intricate feature representations. While deeper models like ResNet and Inception may have the potential to learn more complex patterns, they also risk overfitting on the relatively small dataset of firearms and knives, leading to poorer generalization.

Moreover, the data augmentation techniques and careful hyperparameter tuning applied during the training of the VGG16 model likely played a crucial role in preventing overfitting and improving generalization performance. The simpler architecture of VGG16 may have responded better to these regularization and optimization strategies compared to the more complex ResNet and Inception models.

Despite being a deep model, the VGG16 architecture is computationally more efficient than ResNet and Inception models due to its sequential design and smaller filter sizes. This efficiency may have allowed for more effective training and fine-tuning on the available computational resources, leading to better convergence and performance.

While VGG16 outperformed the other models in this specific study, it is worth noting that the relative performance of different CNN architectures can vary depending on factors such as the dataset size, complexity, and domain. However, the simplicity and effectiveness of the VGG16 architecture, combined with the appropriate transfer learning and regularization techniques, likely contributed to its superior accuracy in recognizing knives and firearms in X-ray bag images.

6. Conclusion and Future Work

This study demonstrated the effectiveness of employing advanced Convolutional Neural Network (CNN) models for the automatic recognition of hazardous items, specifically knives and firearms, in X-ray images of travelers' bags. Three prominent CNN architectures, VGG16, ResNet50, and InceptionV3, were thoroughly evaluated on the SIXray dataset, a large-scale benchmark comprising over a million X-ray images.

The results clearly indicated the superior performance of the VGG16 model, outperforming the other two models across all evaluation metrics, including accuracy, precision, recall, and F1-score. The VGG16 model achieved remarkable scores, with an accuracy of 96.1%, precision of 95.2%, recall of 95.1%, and F1-score of 95.6%. These outstanding results can be attributed to the model's relatively simple yet effective architecture, which consists of a sequential stack of convolutional layers with increasing depth and small 3x3 filters. This design allows the model to learn robust and translation-invariant features from the input images, making it well-suited for detecting and recognizing specific objects like knives and firearms.

Additionally, the pre-trained weights of VGG16 on the ImageNet dataset, combined with carefully applied data augmentation techniques and hyperparameter tuning, contributed to the model's exceptional performance. The transfer learning approach and regularization strategies helped prevent overfitting and improved generalization, further enhancing the model's accuracy.

While the VGG16 model excelled in this study, it is important to note that the relative performance of CNN architectures can vary depending on factors such as dataset size, complexity, and domain. However, the simplicity and effectiveness of the VGG16 architecture, coupled with appropriate transfer learning and regularization techniques, make it a compelling choice for X-ray image analysis tasks.

For future work, it would be beneficial to extend this research to a broader range of prohibited items beyond knives and firearms. Exploring the performance of CNN models on other categories of hazardous materials, such as explosives or chemical substances, could further enhance the capabilities of automated X-ray screening systems. Additionally, investigating the integration of these models into real-world security screening operations and evaluating their performance in practical scenarios would be valuable, Furthermore, as the field of deep learning continues to advance, exploring the potential of more recent and cutting-edge CNN architectures or ensemble models could lead to further improvements in accuracy and efficiency. Continual research and development in this domain are crucial for ensuring the safety and security of passengers and cargo in various transportation and security settings.

References

[1] Singh, S., & Singh, M. (2003). Explosives detection systems (EDS) for aviation security. *Signal processing*, *83*(1), 31-55.

[2] Mery, D., Saavedra, D., & Prasad, M. (2020). X-ray baggage inspection with computer vision: A survey. Ieee Access, 8, 145620-145633.

[3] Zhang, Y., Kong, W., Li, D., & Liu, X. (2020). On using XMC R-CNN model for contraband detection within X-ray baggage security images. Mathematical Problems in Engineering, 2020, 1-14.

[4] Akcay, S., Kundegorski, M. E., Willcocks, C. G., & Breckon, T. P. (2018). Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security imagery. *IEEE transactions on information forensics and security*, *13*(9), 2203-2215.

[5] Gaus, Y. F. A., Bhowmik, N., & Breckon, T. P. (2019, November). On the use of deep learning for the detection of firearms in x-ray baggage security imagery. In *2019 IEEE International Symposium on Technologies for Homeland Security (HST)* (pp. 1-7). IEEE.

[6] Mery, D., & Katsaggelos, A. K. (2017). A logarithmic X-ray imaging model for baggage inspection: Simulation and object

detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 57-65).

[7] Akcay, S., & Breckon, T. (2022). Towards automatic threat detection: A survey of advances of deep learning within X-ray security imaging. Pattern Recognition, 122, 108245.

[8] Miao, C., Xie, L., Wan, F., Su, C., Liu, H., Jiao, J., & Ye, Q. (2019). Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2119-2128).

[9] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[10] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[11] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[12] Kayalvizhi, R., Malarvizhi, S., Choudhury, S. D., Topkar, A., & Vijayakumar, P. (2020, September). Detection of sharp objects using deep neural network based object detection algorithm. In 2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP) (pp. 1-5). IEEE.

[13] Wang, Q., Bhowmik, N., & Breckon, T. P. (2020, July). On the evaluation of prohibited item classification and detection in volumetric 3d computed tomography baggage security screening imagery. In 2020 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

[14] Hassan, T., Akçay, S., Bennamoun, M., Khan, S., & Werghi, N. (2022). Tensor pooling-driven instance segmentation framework for baggage threat recognition. Neural Computing and Applications, 1-12.

[15] Bhowmik, N., Gaus, Y. F. A., & Breckon, T. P. (2021, September). On the impact of using x-ray energy response imagery for object detection via convolutional neural networks. In 2021 IEEE International Conference on Image Processing (ICIP) (pp. 1224-1228). IEEE.

[16] Jain, D. K. (2019). An evaluation of deep learning based object detection strategies for threat object detection in baggage security imagery. pattern recognition letters, 120, 112-119.

[17] Koçi, J., Topal, A. O., & Ali, M. (2020, December). Threat object detection in X-ray images using SSD, R-FCN and faster R-CNN. In 2020 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA) (pp. 10-15). IEEE.