IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Revue d'Intelligence Artificielle
Vol. 35, No. 5, October, 2021, pp. 431-435
Journal homepage: http://iieta.org/journals/ria

Deep Learning-Based X-Ray Baggage Hazardous Object Detection – An FPGA Implementation

Vijayakumar Ponnusamy[1*], Diwakar R. Marur[1], Deepa Dhanaskodi[2], Thangavel Palaniappan[3]

[1] ECE Department SRM IST-KTR, Chennai 603202, India
[2] Bannari Amman Institute of Technology, Erode 638401, Tamil Nadu, India
[3] Kongu Engineering College, Perundurai 638060, Tamil Nadu, India

Corresponding Author Email: vijayakp@srmist.edu.in

https://doi.org/10.18280/ria.350510

Received: 26 July 2021
Accepted: 18 August 2021

Keywords:
deep learning, X-ray image, YOLO, hazardous object detection, FPGA, CNN architectures, image processing

ABSTRACT

This work proposes deep learning neural network-based X-ray image classification. The X-ray baggage scanning machinery plays an essential role in the safeguard of customs, airports, and other systematically very important landmarks and infrastructures. The technology at present of baggage scanning machines is designed on X-ray attenuation. The detection of threatful objects is built on how different objects attenuate the X-ray beams going through them. In this paper, the deep convolutional neural network of YOLO is utilized in classifying baggage images. Real-time performance of the baggage image classification is an essential one for security scanning. There are many computationally intensive operations in the You Only Look Once (YOLO) architecture. The computational intensive operations are implemented in the Field Programmable Gate Array (FPGA) platform to optimize process delays. The critical issues involved in those implementations include data representation, inner products computation and implementation of activation function and resolving these issues will also be a significant task. The FPGA implementation results show that with less resource occupancy, the YOLO implementation provides maximum accuracy of 98.9% in classifying X-ray baggage images and identifying hazardous materials. This result proves that the proposed implementation is best suited for practical system deployments for real-time Baggage scanning.

1. INTRODUCTION

Security screening of X-ray baggage is majorly used to maintain security in transport, primarily in aviation and provides valuable image-based investigating operation for human admins reviewing immense, disorganized and immensely varying contents in baggage within a limited period [1]. Result of elevated consumers turnout in the universal traveling network and the increased attention on wider extension principles on national security, resulting in a challenged automated timeous classification of images task.

The database images are the inputs that have been provided for checking it where it goes on to the pre-processing stage to analyze its width, height and many other features. After that, the next process that comes on to the stage is Convolutional Neural Network (CNN), where the image classification process occurs that is split up into class I and class II, where the first-class identifies the hazardous objects and the other class identifies the non-hazardous objects. For this methodology, YOLO v3 (which is one type of CNN architecture) is applied for the X-ray image classification process.

The problems that arise in automatic visualization X-ray baggage imagery can be treated as typical image classification issues [2]. In the following paragraph use of CNN in image classification is discussed. Further, the outline of a wide generalized background for CNN and an explanation of our approach in the application of these processes to object classification of X-ray baggage are described.

The modern CNN architectures are trained on a large number of datasets which consist of approximately a million data samples and many more varied class labels [3]. Yet, limitations in the application process of such immense training and optimization of parameter techniques to issues where such immensely sized datasets are unavailable. This leads to transfer learning and illustration of the work that every hidden layer in a CNN has various feature representation-related characteristics in which the below layers present the general feature extraction capabilities and the upper layers carry information that is positively more specific to the original classification task. Lower layers perform as edge detectors; on the other hand, higher layers give us more coordinated representations that belong to the input.

This hidden layer concept facilitates the re-usage of the general extraction of features. Unlike the lower layers of CNN, the higher layers are finely tuned across the secondary problem domains with every related characteristic. By this, one can advance a prior CNN parametrizing quality of a currently trained network on a general class of object problem, as an initial point for optimization process towards the particular problem domain of limited class of object detection [4] (ex. images in X-ray). Instead of designing a new CNN from scratch, the adoption of pre-trained CNN is done, pre-optimization for general object finding process and tuning the weights in a finer way on our specific classification domain. Deep CNN has been used widely in many more challenging

computerized visioning processes such as classification of images, object detection and process of segmentation. This kind of high-level parametrization, with its capacity to represent, makes the network suitable to fit much in those traditional deep learning areas. Using the dropouts, where neurons that are hidden are removed at random during the training. These are proposed for overfitting avoidance problems in the course of that performance dependency on individual network elements are diluted in favor of reducing error and responsibility for the problematic space. Adding to that dropout which also increases the size of the networks so that it does not overfit, ReLu, another novelty approach which of course in this work is being introduced as an activation function for its non-linearities. The major success of this particular work led to the design of a very similar architecture plan with all those smaller fields. Further, the work also paves the way for a new approach to visualize features of representations within networks.

And its inspiration by the outcome, the width of the network is entirely explored using the comparable area of the networks with dynamic width. By this, the study of the importance of networking core on classification accuracy by stacking convolutional layered parts with minor receptive fields is possible. The use of smaller receptive filters not only increases the non-linearity but also lowers down entire parameters. It has been shown that layers of convolutional kernels inside a network with dynamic depth can significantly improve the process. Duo stacked convolutional parts are used to feed the input. Later they are combined with the output. At last, the application of non-linearity is made. This process is repeated multiple times. For networks that are deep, filter factorization is carried out.

As the input dataset is messy in nature, there are possibilities where CNN-based classification might fail to classify threats. CNN labels the example images as laptops as the major signature object present, while failure to detecting the foreground objects of interest which leads to a minor boost in wrong negative occurrences. We are considering this mainly as a problem of object detection, and hence leading to the exploration of the object detection strategies.

2. LITERATURE SURVEY

Those past work is majorly focused on the Bag of Visual Words (BoVW) model, though there is certainly minor research that uses some other processes such as using single X-ray sourced automated checking [5]. Now, Convolutional Neural Networks (CNN) are considered as ultra-modern pattern recognition techniques for current computer optical issues. It is established into the area of X-ray baggage imagery [6]. In the paper [6], Subramani et al. have used Fully Conventional Network. This network, even though able to provide good classification accuracy, computational complexity is high. In comparison, the BoVW method employs conventional procedures of manual specifications, which are processed with a Support Vector Machine (SVM). Classification of objects with CNN equally compares it against all traditional classifiers.

Automatic analysis and recognition of hazardous objects in baggage using X-ray images produced by X-ray screeners at airports is a computation-intensive process. A novel way is proposed to reduce the workload by Baştan et al. [7]. The paper investigates the applicability of Bag-of-Visual-Words (BoVW) methods for classifying and retrieving X-ray images. Fusion of the low energy and high energy images into a single-color image facilitates the interpretation of the contents in the baggage. Finally, they conclude by stating that the improvement in performance can be made by the usage of extra information available in X-ray Images. The visual-word approach for classification is very intuitive and used extensively. But this approach is based on an image processing algorithm requiring little manual involvement to detect the object and find a visual word. Moreover, this visual word-based mechanism not having generalization capability like machine learning algorithms. The arrival of machine learning algorithms offers more robust classification without manual involvement.

At first development of appearance-based object detection methods were developed for photographic images. Later they were adapted to suit the properties of X-ray image data. Franzel et al. [8] extended the analysis by taking into account the variations in different sources of object appearance that make detecting objects. They addressed in those variations by adaptation of standard appearance-based object detection approach to the species of dual-energy X-ray data and the inspection. Finally, they gave us a detailed multi-view detection approach that combines single view detections from multiple views. They also evaluated the method of detecting handguns in luggage. Franzel et al. [8] also uses a multi-view approach using multi-camera to improve detection accuracy using the image processing approach. This multicamera system increases the cost of the system and computational complexity.

Ackay et al. [9] provide the details for understanding various methods of CNN and their involvement in decoding the architectures of object detection and classification. They also provide us with clear information on different layers of CNN and their major application in object detection and image classification of X-ray images. In works [9], Automatic hazardous baggage detection systems employed algorithms that employed BoVW with SVM classifier, CNN networks and YOLO v2. During the evolution of techniques, classification accuracy increased, and computational complexity decreased, which is the work's uniqueness. But we need to implement the system in an embedded system to deploy the design practical. The study [9] was carried out in computer implementation, not FPGA or embedded system implementation. Ample information on the importance of kernels CNN was provided by Sun et al. [10]. In this paper, the authors debated on employing various types of kernels for different types of inputs provided and for multiple applications. In the end, a procedure for designing the kernels to employ it for multiple applications was provided.

Wu et al. [11] provided the major principles of wavelet transform and with the use of that, its application on image fusion was constructed. Their paper also dealt with various types of images; the process of fusion was executed using the theory of wavelet transform and the result was achieved. Finally, the authors ended with the further possibility in its development of fusing multiple ranges of images.

Gang et al. [12] explained the importance of deep learning in medical fields and its application in reducing the dimensions for easier and clearer analysis. They also insisted on the importance of employing dual-energy X-ray over low energy X-ray analysis, as dual-energy gave the clearer output so that those can be examined for various purposes.

3. METHODOLOGY

The hazardous and non-hazardous classification system is developed in the following way. Figure 1 provides the block diagram of the proposed system. The main computation is carried out in the FPGA-based PYNQ board. The raw X-ray files are sent by the computer as input to the PYNQ board. It processes the X-ray images to draw boxes on the objects and the objects are later labeled as hazardous or non-hazardous. CNN is used to carry out the classification. The CNN architecture employs YOLO v3 model for efficient classification operation. The resultant labelled X-ray images are sent back to the computer.

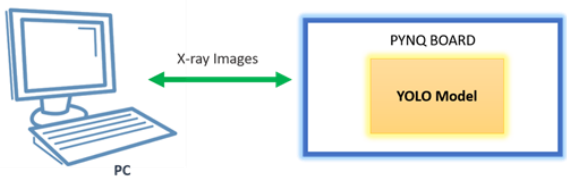

Figure 1. Block diagram of proposed hazardous and Non-hazardous objects detector in X-ray images

3.1 CNN architecture – YOLO v3

Motivated, we perform substantial experiments to examine the strength of CNN and its traditional manual features. As, the performance of layer examining by fixation of parameters from the source without any continued escalation to analyse how finalizing the layer features at differing points in these networks influences the overall operation of the transfer learning-based modulation of the CNN. Moreover, comparison of classification with conventional feature-driven pipelines results in analyzing the advantage in one particular feature of YOLO where the bounding boxes procedure in this process helps us to analyze classification of images with further care and with brighter efficiency.

YOLO v3 is a fully connected CNN achieving absolute fruition for the object detection process [13]. It also uses precise techniques for the improvement of its function against the former works. Detection is carried out by not more than one forward pass. The regional approach makes use of sub-network for region generation and uses anchors as in forward-RCNN. In the YOLO v3, anchors feature learning is based on k-means clustering that operates on the input data. Additionally, to anchor, batch normalization is performed after every layer, resulting in betterment in the overall performance. Unlike other classification networks, high resolution images are permitted to be used as input images with the resolution that varies between 350×350 to 600×600. Additionally, the model randomly varies the size of input images, allowing the network to work with objects with a variety of scales, handling scaling issues. The above-mentioned methods yield vivid improvement in performance and the approach to achieving the perfection.

In YOLO v3, the grid cell is formed to detect an object with its central location coordinate. It will be more appropriate to have an odd size of rectangular grid cells to make the center point. From the center point, we can have an equal size of the pixel in all four directions, which makes it more convenient for the further analysis of the image like bounding box fixing, Confidence score calculation etc. Moreover, the size of rectangular grid cells can be decided based on the size of the small object we are interested in detecting. The 13×13 rectangular grid cells size is fixed in this study because in most image analysis applications, the object size is within 13×13 rectangular grid size. The path YOLO v3 functions by dividing the input image into 13×13 rectangular grid cells as shown in Figure 2. Each cell in turn guesses 5 bounding box coordinating points for each anchor. Additionally, in the individually predicted bounding boxes, the output of the network improves the score delivering the commonalities between the ground truth and bounding boxes. So finally, the output to accepts the PD of the classes that the predicted bounding boxes belong to. Regression and classification in a single forward pass network make YOLOv3 vividly quicker and achieving real time performance. The vast majority of bounding boxes will have low probability. It is very common to have bounding boxes with low probability. Thus, the algorithm compares the score of bounding boxes probability with the set threshold. The bounding boxes that are less than the threshold is deleted by the algorithm. In the next pass, the bounding boxes are subjected to a "non-max suppression" operator. This ensures the elimination of any possible duplicate objects. These two operations make sure only fit bounding boxes are available.

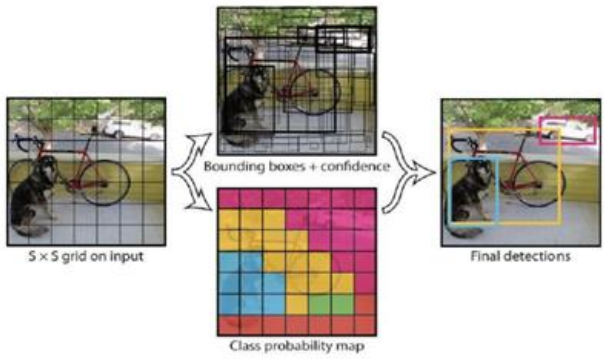

Figure 2. Stages in YOLO v3 processing Image Courtesy [14]

The process by which a raw X-ray image is converted into hazardous and non-hazardous labeled X-ray images is broken down into steps and the steps are given below.

Step 1: Setting up the threshold (0.5).
Step 2: Defining input Width and Height.
Step 3: Defining the layers.
Step 4: Positioning of the image.
Step 5: Identifying class Ids, confidences and boxes.
Step 6: Verification of confidence and comparison to the threshold.
Step 7: Input of test image.
Step 8: Defining the training class.
Step 9: Defining windows and weights.
Step 10: Output Generation.

3.2 FPGA implementation – PYNQ board

Xilinx has released an open-source project titled 'PYNQ' to design embedded systems easily. With this productivity platform, one can configure Xilinx Zynq System-on-chip (SoC) without using Application-Specific Integrated Circuit (ASIC)-style design tools. It enables the use of Python language, and its libraries to design programmable logic and

microprocessors in Zynq. PYNQ is an apt platform for high-performance applications that focus on high frame-rate video processing and real-time signal processing.

4. RESULTS AND DISCUSSION

Google collab tool was used to classify X-ray images with hazardous and non-hazardous objects. The sample output images are shown in Figure 3. In the output images, hazardous objects are enclosed by violet box and non-hazardous objects are enclosed in the green box. The training is done through the darknet using 500 sample data. The threshold point plays a great influence on the result; if the threshold is set high, it takes a longer time to compute the result but has greater accuracy. When the threshold is set low, then the computation is faster but the accuracy is low. An experiment is carried out with 200 iterations for fixing the optimal threshold to have better performance. In the experiment, the threshold value that can provide 98% of accuracy is taken as optimal. This optimal threshold value is used in the implementation.

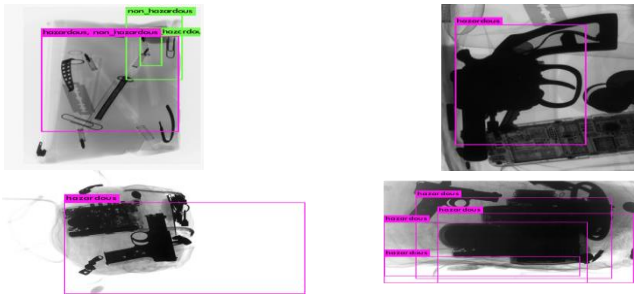

Figure 3. Detection of hazardous and Non-hazardous objects in X-ray images using proposed method

The object detection capability of the YOLO model is evaluated using the Mean Average Precision. This means Average Precision compares the ground-truth bounding box to the detected box by YOLO and computes the score. The higher value of Mean Average Precision shows that capability of the model to detect the object is higher.

The mean Average Precision (mAP) of the hazardous objects present in an X- ray image is provided in the Table 1. The hazardous objects that are taken for testing are gun, blade and knife. Each object has its own shape and size When the parameter mAP is compared for all given hazardous objects, one can safely deduce that mAP is independent of object shape and size. This is a boon for automatic image classification.

Table 1. Detection accuracy among various hazardous objects

Hazardous Object	Detection accuracy mAP
Gun	98.5 %
Blade	98.0 %
Knife	98.9 %

The basic building block of FPGA is called a logical block used to program and implement the required algorithm and functionality. The flip-flop and Look-Up Table (LUT) are combined to form the logical block. LUT is a customized/programmable Truth Table to implement combinational logic. Apart from, that FPGA has RAM memory to store the required data for the computation of the implemented logic/algorithm. The RAM that is embedded in the FPGA is called Block RAM (BRAM). As a third element, Digital Signal Processor (DSP) in core form is embedded in FPGA for computationally intensive signal processing operations. The Size of LUT, the number of flip-flops and DSP units occupied by the proposed implemented system decide area occupancy on FPGA chip and power consumption. Lesser the occupancy of them, power consumption will be lesser.

The proposed mechanism is implemented on PYNQ embedded system board FPGA chip. The PYNQ embedded system board has FPGA with limited resources like LUT, flip flops and DSP core .so we need to analyze the occupancy of such resources when we implement any algorithm, especially the image processing algorithm, which occupies more resources. Table 2 provides such an occupancy analysis. In the PYNQ embedded system with FPGA chip, three types of image processing operations are implemented viz pre-processing, 3x3 kernel operations and classification (using YOLO v3). Table 2 provides the details of the operations as well as units used in that operation. The element used are Lookup Table (LUT), Flip-Flop (FF), Block Random Access Memory (BRAM) and Digital Signal Processing (DSP) core. Out of the three image processing operations, classification based on YOLO v3 architecture requires a lot of FPGA resources (LUT, FF, BRAM and so on). This is the small price that is paid for accurate hazardous object detection.

Table 2. Type of operation vs. resource requirement

Type of Operation	Name of Element and Quantity Required			
	LUT	FF	BRAM	DSP
Pre-Processing	25	56	4	1
3x3 kernel operation	1096	4046	6	2
Classification (using YOLO Architecture)	82150	130 k	365	194

Through this work, exploration of the necessity of CNN in classification and detection tasks within X-ray baggage imagery is done exhaustively. For the process of classification, a comparison between CNN and the traditional way of approach, which is based on handcrafted features, has been made. In this work, every convolution process is carried out in multiple layers. In addition to this, the model is trained using the X-ray image database in Python to bring out the efficiency and improvement in performance. YOLO v3 architecture which is a part of CNN, is explored here. The extent of complexity in the network and overall performance is studied. The simulation results demonstrate YOLO v3 features achieve finer performance compared to manual detection or other automatic detection techniques.

5. CONCLUSION

It is important to verify the baggage of people in sensitive locations such as airports, railway stations, etc. Automatic Identification of baggage with hazardous objects within the shortest possible time is the need of hour. The solution is proposed using CNN. The implementation is carried out on FPGA through the PYNQ-Z1 board. The board has multiple features in it. Complications arise from porting of Python through FPGA, such as data representation. By implementing this method in the places where large gatherings of people occur can be secured from any form of disastrous events.

REFERENCES

[1] Hättenschwiler, N., Sterchi, Y., Mendes, M., Schwaninger, A. (2018). Automation in airport security X-ray screening of cabin baggage: Examining benefits and possible implementations of automated explosives detection. Applied Ergonomics, 72: 58-68. https://doi.org/10.1016/j.apergo.2018.05.003

[2] Dhiraj, 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. https://doi.org/10.1016/j.patrec.2019.01.014

[3] Clement, J.C., Indira, N., Vijayakumar, P., Nandakumar. R. (2021). Deep learning based modulation classification for 5G and beyond wireless systems. Peer-to-Peer Networking and Applications, 14: 319-332. https://doi.org/10.1007/s12083-020-01003-3

[4] Gaus, Y.F.A., Bhowmik, N., Breckon, T.P. (2019). 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. https://doi.org/10.1109/HST47167.2019.9032917

[5] Wells, K., Bradley, D.A. (2012). A review of X-ray explosives detection techniques for checked baggage. Applied Radiation and Isotopes, 70(8): 1729-1746. https://doi.org/10.1016/j.apradiso.2012.01.011

[6] Subramani, M., Rajaduari, K., Choudhury, S.D., Topkar, A., Ponnusamy, V. (2020). Evaluating one stage detector architecture of convolutional neural network for threat object detection using X-ray baggage security imaging. Revue d'Intelligence Artificielle, 34(4): 495-500. https://doi.org/10.18280/ria.340415

[7] Baştan, M., Yousefi, M.R., Breuel, T.M. (2011). Visual words on baggage X-ray images. In Proc. of International Conference on Computer Analysis of Images and Patterns, 6854: 360-368. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-23672-3_44

[8] Franzel, T., Schmidt, U., Roth, S. (2012). Object detection in multi-view X-ray images. In: Pinz, A., Pock, T., Bischof, H., Leberl, F. (eds) Pattern Recognition. DAGM/OAGM 2012. Lecture Notes in Computer Science, 7476: 144-154. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-32717-9_15

[9] 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. https://doi.org/10.1109/TIFS.2018.2812196

[10] Sun, Z., Ozay, M., Okatani, T. (2016). Design of kernels in convolutional neural networks for image classification. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. Lecture Notes in Computer Science, 9911: 51-66. Springer, Cham. https://doi.org/10.1007/978-3-319-46478-7_4

[11] Wu, J.P., Yang, Z.X., Su, Y.T., Chen, Y., Wang, Z.M. (2007). Wavelet transform and fuzzy reasoning based image fusion algorithm. In Proc. of 2007 International Conference on Wavelet Analysis and Pattern Recognition, pp. 73-77. https://doi.org/10.1109/ICWAPR.2007.4420639

[12] Gang, P., Zhen, W., Zeng, W. et al. (2018). Dimensionality reduction in deep learning for chest X-ray analysis of lung cancer. In Proc. of 2018 Tenth International Conference on Advanced Computational Intelligence (ICACI), pp. 878-883. https://doi.org/10.1109/ICACI.2018.8377579

[13] Ponnusamy, V., Coumaran, A., Shunmugam, A.S., Rajaram, K., Senthilvelavan, S. (2020). Smart glass: Real-time leaf disease detection using YOLO transfer learning. In 2020 International Conference on Communication and Signal Processing (ICCSP), pp. 1150-1154. https://doi.org/10.1109/ICCSP48568.2020.9182146

[14] Rosebrock, A. (2018). YOLO object detection with OpenCV. PyImageSearch. (Available Online) https://www.pyimagesearch.com/2018/11/12/yolo-object-detection-with-opencv/