International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Resolute Airport Screening of Baggage and Classifying Threats Using Deep Learning and Computer Vision

Aravind Karrothu [a], Chakrapanda Suraj [b], Godaba Amukthamalyada [c], Harsha Vardhan [d], Rama Bhagavan [e], Burada Anjali [f]

[a] *Assistant Professor, Department of CSE, GMR Institute of Technology, Rajam, AP-532127, India*
[b,c,d,e,f] *Department of CSE, GMR Institute of Technology, Rajam, AP-532127, India*

ABSTRACT

The baggage checking is a regular process in airports. Generally, the baggage of the passengers goes through a panel for scanning. A security guard or an agent is assigned to check for the threat items. This is an overwhelming process. Sometimes mistakes may happen with naïvely. Because humans may cause mistakes. Machines are the resources which can be trusted while the given parameters are correct. The scanner in the panel produces x-ray images. The x-ray images consist only black and white intensities. While the baggage in the panel goes through scanner, the scanner generates x-rays and present the inside out of the baggage. Also, the scanner black and white intensities may create a dilemma to a security guard or an agent. Therefore, machines can verify and detect the threat without any error. To build a machine, a software feature is required. The software feature must be trained to detect the threat items in baggage screening. To train the machine we propose a supervised and combined state-of-art Deep Learning (DL) and Computer Vision (CV) techniques to predict the location and classify the threat in baggage. In this work, we are using combined models because the state-of-art implementation of the problem statement is fewer.

Keywords: Baggage, Scanning, Machines, Supervised, Deep Learning, Computer Vision

1. Introduction

Transportation is a very large process of transporting various goods, people along with their baggage. There are various types of transportation services such as by road, by water, and by air. Every transportation is helping a passenger to reach their destination properly. But in recent days, the transportation is very crucial for all the passengers because some of the people were carrying threat items such as blade, knife, gun, electronic devices, and other conducting materials. But in some situations, there is a chance of getting attacks from unauthorized people due lack of security screening of the passengers and their baggage's at entry points. At one particular time, there is an analysis of around 1.5 million passengers travelled with dangerous threat items [1]. Earlier, the baggage was checked at entry points in a manual procedure. But it will take so much time, also could not able to identify such threat items due to lack of overwhelming process to the security guard or agent at the entry point. Later, this system was enhanced by various updated techniques like x-ray screening panels.

a

b

Fig. 1 - (a) Sample x-ray images of gun class; (b) Sample x-ray images of shuriken class.

In this there may be a chance of getting dilemma creation of black and white intensities of x-ray images to the security guard or an agent. Some of the sample X-ray images represented in Fig. 1. Day by day, the usage of transportation was vastly increasing with huge number of passenger count along their baggage. In such cases, we require an efficient mechanism for identifying threat items at entry points. For that, Deep Learning techniques plays a vital role in baggage screening. This work describes a resolute screening of baggage and classifying threat items using Deep Learning (DL) and Computer Vision (CV) at airports. In section 2, an analysis of earlier mechanisms will be discussed. The next section describes about methodology, architecture, and evaluation criteria. Section 4 explains tabular along with graphical results, and concludes with future scope.

2. Related works

This section describes various contributions of deep learning along with computer vision and other existing works. In recent days, DL algorithms (CNN, DNN, Energy-Based models, Deep Residual Learning models, and others) has been successfully used for object detection and it can replace the traditional manual procedures. The CV techniques have used for video or image preprocessing.

Taimur Hassan et al., proposed novel anomaly segmentation method which recognizes the ground truth of the object in an image. They have followed an unsupervised approach i.e., without labels of the threat items. They tested their model on various challenging datasets and finally achieved best results [1].

The authors have proposed a technique to detect multi-view of object in an image by using structured-learning architecture and branch & bound search. The x-ray images were calibrated to convert into pseudo images for better object detection. They have also used Sparse interest Points (SPIN) feature extraction for extended extraction features. Finally, they have concluded multi-view imaging with SPIN extractor improves the performance of detection [2].

Domingo Mery et al., have proposed a new method adaptive sparse representation consists two phases (learning along with testing). In the first phase, every object is going to be trained and form into an appropriate dictionary template. At other side, from this template, the author can classify set of images by using sparse representation classification methodology. This work compared with four different algorithms to shows that adaptive sparse representation is better than others. But this algorithm is only applicable for very less parameters like image size, image style, image contrast and other related conditions [3]. The authors made transfer learning using deep convolution neural networks. They employed learning technique to multi classify the input image. The algorithms such as AlexNet, and GoogleNet are pre-trained on ImageNet which is a large database. The weights of these algorithms are used to classify the threat item in the x-ray scan [4].

Domingo Mery et al., have contributed for an approach to identifying an object through x-ray testing by using various CV techniques. For that, they have created a new dataset and it can be given as input for those CV techniques. Also, they proposed a pair of CV techniques by using a combination of Bag of Words (BoW) and sparse K-Nearest Neighbors (KNN) [5]. Later another algorithm was developed Vladimir et al., based on x-ray images by using active vision concept. It could highlight three view-points such as ‘good view-point’, ‘bad view-point’ along with ‘no view-point’. Also, it can easily detect a threat object by single view-point with Adaptive Implicit Shape Model (AISM). Finally, they have concluded that this improved methodology gives better performance [6].

The authors made transfer learning for classification and detection of threat items from the given input image. The methodology was built with combination of trained support vector machines on classification and AlexNet object detection. They have also transfer learned various strategies in object detection. Finally, they have concluded with a comparative result that states object detection strategies are better than classification techniques [7]. Samet Akcay et al., semi-supervised detection technique in Generative Adversarial Network (GAN). Generally, GAN consists of two subnetworks namely encoder (generator) and decoder (discriminator). The novel model gives better result when compared to traditional GAN model and autoencoder model [8].

The authors introduced an unsupervised approach anomaly detection in the baggage screening. The model is an encoder – decoder CNN which also have skip connections in between the CNN layers. In the training, the image reconstruction error metric is minimized to learn distribution of features via normality. The experimentation is done on various real world challenging datasets and achieves better results than previous works [9]. D. Castro Piñol et al., made a selective search approach in CV. They have used an extensive of BoW CV algorithm and it is Bag of Visual Words (BoVW) to detect the threat items. They have annotated their data with bounding boxes in order to inference their proposed model. The model shows effective results on predictions [10]. Neelanjan Bhowmik et al., proposed a segmentation strategy for anomaly detection in a supervised manner. They have used two automated segmentation strategies such as object segmentation and sub-component segmentation. The object segmentation is used to detect the baggage on the panel whereas, the sub-component segmentation detects the items inside the object. Finally, they have concluded binary classification of anomaly and benign of threat items of sub-component segmentation achieved best performance [11]. Yona Falinie et al., made a two-class classification of threat items from the input images. They have used some the leading CNN architectures such mask Region-Based CNN, faster Region-Based CNN, and RetinaNet. Instead of checking a baggage in manual process, they proposed an automation system for complex & overlapped images. Finally, they have obtained effective results of their experimentation [12].

The authors made a combination of Generative Adversarial network (GAN) & faster R-CNN detection model. The GAN is used for data augmentation. They have considered the dataset which has single class (only one object in an image) images and augmented the data for high iterations. The generated data is injected into detection model i.e., faster R-CNN. Later testing was done with multi class (more than one object in an image) images. Finally, they have achieved better results of their work [13]. Rohit Gupta et al., proposed a theory for medical imaging. GAN networks have best encoder-decoder networks. They were used for data manipulation such as augmentation, resizing, enhancing, and resolution. They have concluded that GAN outperforms the DL models on MRI scanning images resolution enhancement [14].

The authors proposed a new search method which makes the architecture backbone lightweight of the proposed model. They have referenced Neural Architecture Search (NAS) method to inference their proposed model which has multi-branch and gathers the features of the images in way more efficient. Finally, the proposed find the optimal regions way faster than the existing works [15].

Table 1 – Techniques, number of categories, and performance metric with result of related works.

	Technique	Number of categories	Performance metrics used	Result
[1]	Unsupervised encoder-decoder network	4	Mean Average Precision	**85.9%**
[2]	Multi-view structured learning framework	3	Precision	**72.6%,**
[3]	Extension of Adaptive Sparse Recognition	4	Accuracy	**97%**
[4]	AlexNet, GoogleNet	4	Mean Average Precision	**95.26%, 98.40%**
[5]	Sparse K-Nearest Neighbors	3	Precision	**97.7%**
[6]	Active vision approach	2	Precision	**89.81%**
[7]	State-of-art CNN object detection models	4	Mean Average Precision	**97.4%**
[8]	Generative Adversarial Networks: Semi supervised	2	Area under curve	**88.2%**
[9]	Skip: GAN Networks	2	Area under curve	**90.3%**
[10]	Bag of visual words	1	True positive rate	**92%**
[11]	Supervised anomaly segmentation along with sub level	2	True positive rate	**99.1%**
[12]	CNN based object detection	1	Mean average precision	**92%**
[13]	GAN + Faster R-CNN	3	Mean average precision	**91.41%**
[14]	GAN networks	0	Peak signal noise ratio Structural similarity index	**38.13 dB, 94.3%**
[15]	Multi-branch CNN	0	PSNR, SSIM	**34.3 dB, 92.5%**

Table 2 – Analysis of class labels of existing works.

Year of Publication	**Reference Number**	**Classes**								
		Blade	**Bottle**	**Camera**	**Clip**	**Firearm**	**Gun**	**Knife**	**Laptop**	**Shuriken**
2021	**[1]**	✓	×	×	×	×	✓	✓	×	✓
2015	**[2]**	×	✓	✓	×	×	✓	×	✓	×
2016	**[3]**	✓	×	×	✓	×	✓	×	×	✓
2016	**[4]**	×	×	✓	×	✓	×	✓	✓	×
2017	**[5]**	✓	×	×	×	×	✓	×	×	✓
2017	**[6]**	✓	×	×	×	×	✓	×	×	×
2018	**[7]**	×	×	✓	×	✓	×	✓	✓	×
2018	**[8]**	×	×	×	×	×	✓	✓	×	×
2019	**[9]**	×	×	×	×	×	✓	✓	×	×
2019	**[10]**	×	×	×	×	×	✓	×	×	×
2019	**[11]**	×	×	×	×	×	✓	×	×	×
2019	**[12]**	×	×	×	×	✓	×	×	×	×
2020	**[13]**	✓	×	×	×	×	✓	×	×	✓
2020	**[14]**	–	–	–	–	–	–	–	–	–
2021	**[15]**	–	–	–	–	–	–	–	–	–

3. Materials and methods

The existing state-of-art works describes the object detection through annotations, mask generation, and classifications. The mask generation may predict the wrong structure or region of the threat items. The classifications are made for only finite classes of threat items. Therefore, we want to proceed with annotations for training of object detection models on various classes. We want to label and detect the threat items into four class labels. The labels are blade, gun, knife, and shuriken.

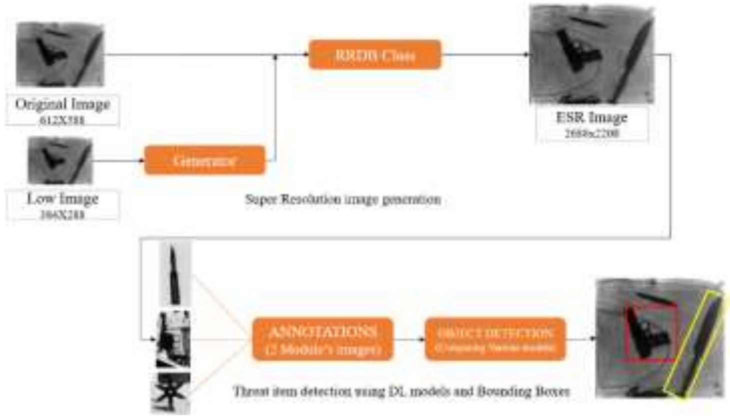

Fig. 2 - Architecture of combined proposed models.

The Fig. 2 represents an outline structure of our proposed model. The proposed model is a combination of CV and various DL object detection algorithms. This works consists of various models to detect the threat items. The model development depends upon input images. The entire project is divided is 2 modules. The first module is a CV algorithm that is ESRGAN. The annotations are drawn for all the two types of images i.e., normal images and super resulted images. The last module consists of various state-of-arts models to test the theory of our proposed work.

3.1 Identified openings and yet to enhance of existing works

- The detection rate of the threat items will be challenging for overlapping objects.
- Some X-ray images consists of noisy areas which makes any model to predict the bounding box of the threat items.
- A model should detect any threat item with regarding to its dimensions, regions, and angles. But if a model has less annotated dataset, then it will be challenging.
- Some X-ray images contains void spaces which creates high density around the objects (threat items). The model may find difficult in localizing the objects.
- The bounding box is used to locate the threat items. The bounding box for threat items should be in a refined manner.

3.2 Dataset details

The dataset which is used in this work is known as GDXray. The dataset is a publicly available dataset. The dataset consists x-ray baggage images which are scanned through the panel in the airport screening. The dataset has various threat materials. The main aim of this work is too resolute the baggage images and classify the threat. If we have to classify all the threat items in the baggage, the false detections may occur because the major threat items in the baggage are Blade, Gun, Knife, and Shuriken.

Fig. 3 – (a) Blade class image; (b) Gun class image (c) Gun, shuriken, and knife classes image; (d) Shuriken, and blade classes image.

3.3 Dataset splitting

The problem statement requires some great number of training samples because the detection categories are four. Generally, 1000 samples are great enough for good predictions. The dataset GDXray also contains single object images of threat items. But we want to detect the threat items which have multiple objects in an image. The total number of samples that we have used for experimentation is 900.

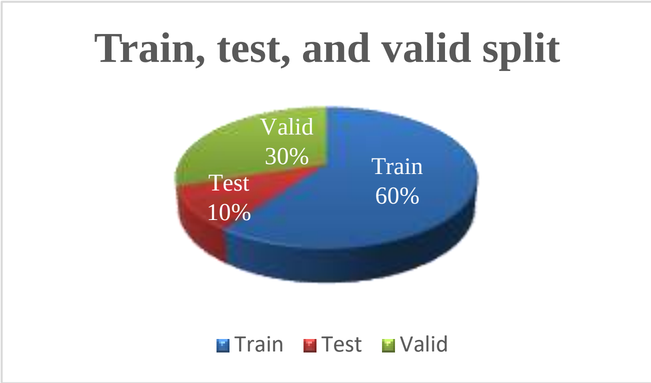

Fig. 4 – Dataset splitting into 60% (540) train, 30% (270) valid, and 10% (90) test.

3.4 Super resolution

The super-resolution module is a Computer Vision (CV) algorithm which is used to get high-resolution images from low resolution using Generative Adversarial Network (ESRGAN). When compared with standard definition (SD) images, High definition (HD) images give more accurate results. The HD images are also helpful for clear object detection. GAN is used to generates the SR image from fake image (Low Resolution image) produced by the generator. PSNR is used to measure the performance of the model with regarding to the dataset.

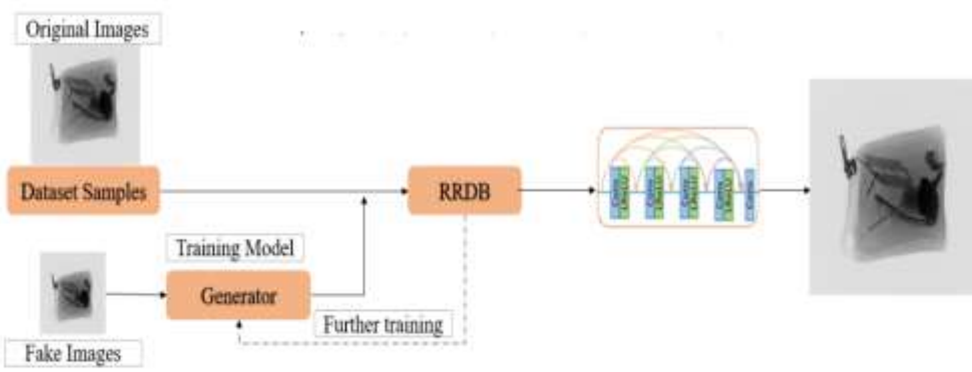

Fig. 5 – workflow of super resolution module.

3.5 Object detection models

In this project, there are 6 object detection models. They are Faster R-CNN, AI Detectron2, YOLOv7.

Faster R-CNN: (Tfrecords annotations) The regions of the images are extracted, and classification of the objects are done. The training of this model takes lot of time.

AI Detectron2: (JSON annotations) The backbone of this algorithm is Mask R-CNN. It supports large databases and enhances pixeled images.

YOLOv7: (Xml annotations) The speed of this model is very high. It has good stability and good prediction rate of bounding boxes.

3.6 Evaluation criteria

The major evaluations are done to inference the object detection of proposed model is mean average precision with 0.5 as threshold. To evaluate the super resulted images peak signal noise ratio is used.

- Peak Signal Noise Ratio: It is a metric which is used to compare the noise reduction of the original and the generated or compressed image. The PSNR ranges depends upon the bit depth of the images. 30 to 50 dB is 8 bits, 60 and >60 dB is 12 bits, and 60 to 80 dB is 16 bits. But if the SR image (8 bits) is way qualitied than input image (8 bits) then the PSNR will be higher.
- Mean Average Precision over IOU: The average precision over the union of ground truth label and predicted bounding boxes is known as the AP over IOU.

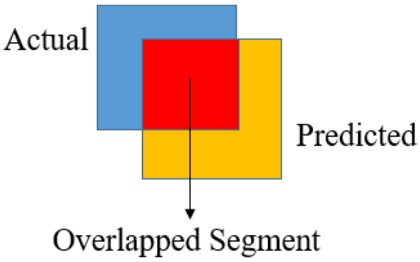

Fig. 6 – Sample actual & predicted bounding box overlapping.

4. Results and discussions

Each image in the dataset is preprocessed by module 1 CV algorithms. The normal images are resized into 640 x 640 whereas, the generated SR images have high length and width. So, the SR images are resized into 1024 x 1024. Because, the bigger the image, the better the detections. The entire images are classifying as four labels. They are blade, gun, knife, and shuriken.

4.1. Super resolution results

Generally, a CNN layers reduce the input image size. But, to increase the size of the image, we must increase the pixel features in the image. The RRDB is used to cover the loss the features in next layers. The upsamling rate of the input images is 4x. The PSNR value achieved from the original and SR images is 91.1 dB

a

b

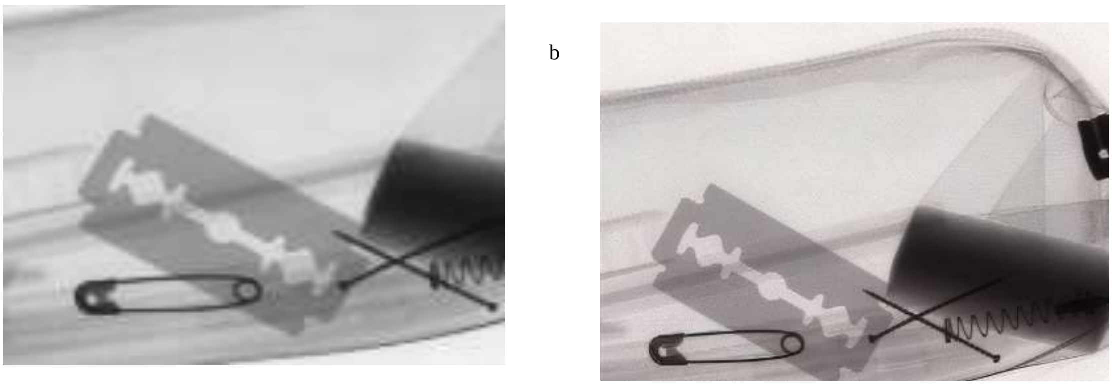

Fig. 7 – (a) Sample normal image; (b) Super resulted image of normal image.

4.2. Object detection results

The results of the normal images and SR images are tabulated.

Table 1 - mAP of object detection models.

Network	**Normal (*AP*)**	**ESRGAN (*AP*)**
Faster R-CNN	75.9%	78.3%
AI Detectron2	**87.4%**	**92.3%**
YOLOv7	90.7%	87.8%

From Table 1, the best model is AI Detectron2. Not only with AP but also confidence level of detecting the threat items with test images.

a

b

Fig. 8 – (a) Random normal image of gun class; (b) Random SR image of blade, and gun class.

4.3. Graphical results

The average precision of four classes individually and combined with threshold as 0.5 is inferenced. The **x-axis** is number of epochs and **y-axis** is average precision.

The following are the graphs of the training of AI Detectron2:

- Mean average precision at 0.5 as threshold.
- Average precision of blade class.
- Average precision of gun class.
- Average precision of knife class.
- Average precision of shuriken class.

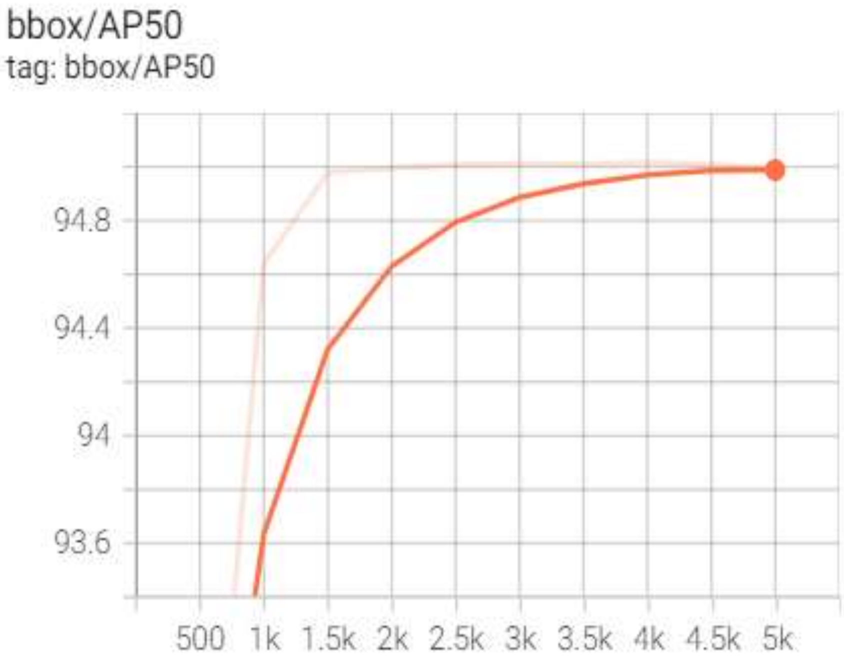

Fig. 9 – Mean average precision of bounding box prediction with 0.5 as threshold.

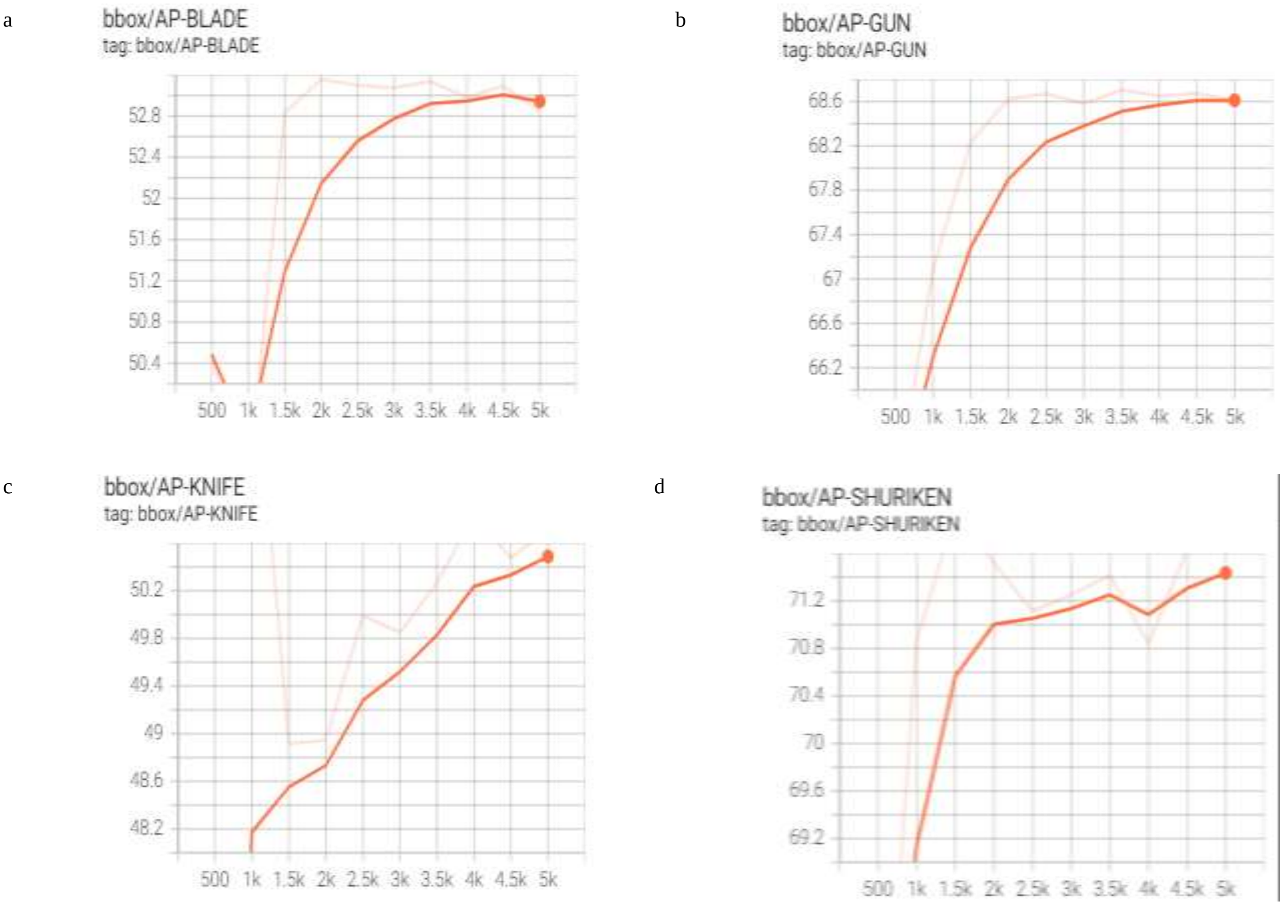

Fig. 10 – (a) Average precision of bounding box prediction of blade class; (b) Average precision of bounding box prediction of blade class; (c) Average precision of bounding box prediction of blade class; (d) Average precision of bounding box prediction of blade class.

5. Conclusion and future scope

5.1. Conclusion

In this project, the object detection models are used for all three types of datasets. Some models are very good at predicting the confidence percent of threat items, but the overall mean AP is not so good. For example, faster R-CNN is good at confidence predicting but mean AP for normal images is less. When compared to normal and ESRGAN images, ESRGAN images are predicted best by using faster R-CNN. The AI detectron2 is used with the backbone of mask R-CNN and it also good at mean AP and confidence prediction. YOLO versions are straight forward and single staged models. YOLOv7 is the best in training time, mean AP percentage, confidence percentage but lack in stability of training. To detect the threat item with good confidence, AI detectron2 is best among remaining models.

5.2. Future scope

The ESRGAN image generation is good task as pre-processing model. But the image generation takes lot time. We have used Colab Notebooks 16 GB GPU and 4x up sampling rate even though the GPU memory is getting full. Further we want to generate the images with 32 GB GPU. We have faced a lot of pressure and time taking at annotation creation. To detect the threat, we must refine the bounding box efficiently. The bounding box should be defined for all images in the dataset. It is difficult task for all the images to draw the bounding boxes. In the object detection models, faster R-CNN not performed as good as others. Further, we increase the steps or epochs included in the faster R-CNN to improve the detection rate.

Acknowledgements

It gives us an immense pleasure to express deep sense of gratitude to my guide Dr. Aravind Karrothu, Assistant Professor, Department of Computer Science and Engineering for his whole hearted and invaluable guidance throughout the project work. Without his sustained and sincere effort, this project work would not have taken this shape. He encouraged and helped us to overcome various difficulties that we have faced at various stages of our project work.

References

Hassan, T., Akçay, S., Bennamoun, M., Khan, S., & Werghi, N. (2021). Unsupervised anomaly instance segmentation for baggage threat recognition. Journal of Ambient Intelligence and Humanized Computing, 1-12.

Baştan, M. (2015). Multi-view object detection in dual-energy X-ray images. Machine Vision and Applications, 26(7), 1045-1060.

Mery, D., Svec, E., & Arias, M. (2015, November). Object recognition in baggage inspection using adaptive sparse representations of X-ray images. In Image and Video Technology (pp. 709-720). Springer, Cham.

Akçay, S., Kundegorski, M. E., Devereux, M., & Breckon, T. P. (2016, September). Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery. In 2016 IEEE International Conference on Image Processing (ICIP) (pp. 1057-1061). IEEE.

Mery, D., Svec, E., Arias, M., Riffo, V., Saavedra, J. M., & Banerjee, S. (2016). Modern computer vision techniques for x-ray testing in baggage inspection. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 47(4), 682-692.

Riffo, V., Flores, S., & Mery, D. (2017). Threat objects detection in x-ray images using an active vision approach. Journal of Nondestructive Evaluation, 36(3), 1-13.

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.

Akcay, S., Atapour-Abarghouei, A., & Breckon, T. P. (2018, December). Ganomaly: Semi-supervised anomaly detection via adversarial training. In Asian conference on computer vision (pp. 622-637). Springer, Cham.

Akçay, S., Atapour-Abarghouei, A., & Breckon, T. P. (2019, July). Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

Piñol, D. C., & Reyes, E. J. M. (2019). Automatic Handgun Detection in X-ray Images using Bag of Words Model with Selective Search. arXiv preprint arXiv:1903.01322.

Bhowmik, N., Gaus, Y. F. A., Akçay, S., Barker, J. W., & Breckon, T. P. (2019, December). On the impact of object and sub-component level segmentation strategies for supervised anomaly detection within x-ray security imagery. In 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA) (pp. 986-991). IEEE.

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.

Kim, J., Kim, J., & Ri, J. (2020). Generative adversarial networks and faster-region convolutional neural networks based object detection in X-ray baggage security imagery. OSA Continuum, 3(12), 3604-3614.

Gupta, R., Sharma, A., & Kumar, A. (2020). Super-resolution using gans for medical imaging. Procedia Computer Science, 173, 28-35.

Ahn, J. Y., & Cho, N. I. (2021). Multi-Branch Neural Architecture Search for Lightweight Image Super-Resolution. IEEE Access, 9, 153633-153646.