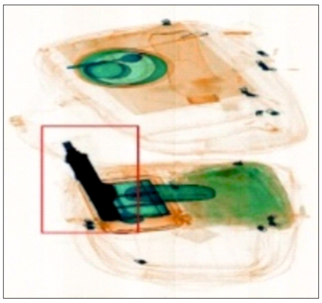
Homeland Security

16

AI based Object Detection in X-Ray Images for Homeland Security

Janhavi Deshpande* Padmini S., Preetha Nair and Sayan Banerjee

Electronics Division, Bhabha Atomic Research Centre, Mumbai 400085, INDIA



Auto detection of Guns

ABSTRACT

X-Ray baggage scanners and Cargo scanners are used at critical checkpoints to screen the item under inspection to reveal concealment of threats, contra-bands, illicit materials and prohibited items. Manual screening by human operators plays a vital role in their detection. Although experience and knowledge are the key factors for confident detection, external variables such as personnel fatigue, increase in traffic volume and cluttered items under inspection adversely impact manual screening. A more viable solution is automatic analysis of scanned images to assist the human operators to accurately identify anomalies. Recently a modern technique namely, Deep Learning has demonstrated remarkable success in automated object detection. Hence we have used AI based deep learning approach, with a particular focus on object classification, detection, segmentation and anomaly detection tasks for Auto-threat detection and Auto Content verification.

KEYWORDS: *Automated object detection, X-Ray image inspection, Baggage scanner, Cargo scanner automated content verification, Data augmentation for X-Ray images, Feature visualization*

Introduction

X-ray Radiography is a non-destructive imaging technique used to scan baggage and to inspect scanned images for anomalies, threats and prohibited items. Manual Inspection of X-ray images for high risk prohibited items is largely dependent on operator alertness and expertise. Automating the X-Ray image classification into threat and non-threat classes requires intelligence to be integrated into image processing. Due to variation in shape and size of threat objects, background and occlusion, it is difficult and impractical to create an exhaustive set of rules for threat object detection, rendering numerical models unsuitable to model threats. A better solution is to use machine learning techniques and train models to learn the most important rules by themselves. Convolutional Neural Network (CNN), is chosen for threat detection, as it scales well for images and considers spatial correlation. This article is divided in three parts:

- Automated threat recognition in X-Ray baggage scanner images
- Automated threat detection in X-Ray baggage scanner images
- Automated Content verification for cargo scanner

Automated Threat Recognition

Automated threat recognition discusses classification of dual energy X-Ray baggage image into threat or non-threat class based on presence of a threat item, using CNN.

Data Augmentation for Threat Recognition

CNN requires a large and balanced dataset to learn a generalizable model. The dataset obtained from X-Ray baggage scanner was imbalanced as the majority of the images do not contain a threat leading to a bias in the learning algorithms based on this dataset towards non-threat class. To balance the dataset, data augmentation based on Threat Image Projection technique[1] is utilized to artificially generate training images containing threat item. A dataset of 110 non-threat baggage images and 94 images of the featured threat item was split into (i) base training set consisting of 80 non-threat baggage images and 73 featured threat images and (ii) base test set consisting of 30 non-threat baggage images and 21 featured threat images.

The dataset was obtained from dual energy single view ECIL make X-Ray baggage scanner ECX6040:SVXBIS. Class preserving translation and rotation variations were applied to existing data to expose the CNN model to natural variations during training. Feasibility of projection was calculated to determine whether the featured threat image after transformation lies completely inside baggage region. Translated and rotated featured threat images were projected onto the clean image patches using Threat Image Projection technique. The intensity of a pixel in a threat scan was split into contributions from the threat and its background using Beer-Lambert law. Threat mask was calculated using the attenuation caused by the background in threat scan without any container. The threat mask was then projected into X-ray images by multiplication.

The training set and test set derived from the base set using data augmentation contained around 1.24 lakh images and around 18,000 images respectively. The validation set was formed by randomly choosing 5000 test set images from the above images. The training and testing dataset also included toy guns, which belonged to negative class. Toy guns have shape similar to real guns but made of different material. This introduced the material variation in the training set.

*Author for Correspondence: Janhavi Deshpande
E-mail: janhavid@barc.gov.in

Table 1: Model Performance using Adam Optimizer.

Model	Tr. A	VA	TA	FP	FN	Precision	Recall	F - measure	AUC - ROC	Accuracy on toy guns
Gradient Descent	0.9881	0.985	0.9846	85	121	0.9867	0.9812	0.9839	0.9845	0.896
Adam Optimizer	**0.9972**	**0.992**	**0.9905**	**32**	**95**	**0.9949**	**0.9852**	**0.9904**	**0.9924**	**0.972**

Table 2: Variation in VGG-19 layers.

Model	Tr . A	TA	FP	FN	Precision	Recall	F - measure	AUC - ROC
VGG - 19	0.9881	0.9846	85	121	0.9867	0.9812	0.9839	0.9845
VGG - 20	**0.9912**	**0.9864**	**90**	**93**	**0.9860**	**0.9855**	**0.9857**	**0.9836**
VGG - 21	0.9917	0.9847	25	180	0.9960	0.9669	0.9720	0.9842
VGG - 22	0.9908	0.9793	65	213	0.9896	0.9669	0.9781	0.9788
Google Net	0.9983	0.9386	27	226	0.9956	0.9649	0.9800	0.9803

CNN Training and Optimization

The Current state of the art CNN models such as AlexNet, VGGNet were considered for threat detection. Four channel input (-log HE, -log LE, HE + LE, HE - LE), derived from 2 channel (High Energy (HE) and Low Energy (LE)) was selected considering the occlusion characteristics and material discrimination in dual energy X-Ray baggage scanner images. AlexNet, VGG-16 and VGG-19 were trained on a Tesla K80 GPU, using mini batch gradient descent approach with batch size of 32. Dropout with probability of 0.5, in addition to weight decay (lambda=0.0001), was used in fully connected layers as a regularization technique. The performance of CNN models has been reported in terms of training accuracy (Tr. A.), testing accuracy (TA), false positives (FP), false negatives (FN), precision, recall, F-measure and AUC-ROC.

CNN Optimization

Among AlexNet and VGGNet models, VGG-19 achieved highest testing accuracy of 98.4% with four channel inputs. We further explored three different approaches for improving accuracy i) use of Adam optimizer[2] instead of vanilla gradient descent as CNN loss optimization algorithm, ii) Varying the number of layers in VGG-19 architecture and iii) ensemble of trained models.

Adam Optimizer: Adam Optimizer is a variant of stochastic gradient descent (SGD) which adapts the learning rate based on recent gradient magnitudes. Adam Optimizer improved the training accuracy from 98.81% to 99.72%. High testing accuracy (99.05%) which is close to training accuracy and improved toy gun classification (97.2%), also indicate good generalization achieved by trained model. (Table 1).

Variation in number of layers: VGG-20, VGG-21 and VGG-22 architectures were designed by varying number of layers in VGG-19 architecture with gradient descent. (Table 2). Considering the trade-off between higher learning capacity provided by deeper networks and overfitting due to large number of parameters in deeper networks, peak is found at 20 layers (VGG-20). Model starts overfitting and losing its generalization capacity beyond 20 layers.

Ensemble Network: VGG-19 and AlexNet achieved threat detection rate of 98.52% and 97.96% respectively using Adam Optimizer. Analysis of false negative images revealed that AlexNet was able to detect some of the guns among 1.48% guns, which were not discovered by VGG-19. Considering the importance of reducing false negatives for automated threat detection, AlexNet and VGG-19 models were combined to increase threat detection rate. The combined Ensemble Network reduced false negatives by 36% as compared to both the base models (Table 3). Out of all the tested models, Ensemble network of four channel AlexNet and VGG-19 had the highest accuracy (99.37%) and threat detection rate of 99.06%. Also, the model had F-measure value 0.9934 and 0.9936 AUC. Thus, the ensemble network of AlexNet and VGG-19 was found the most suitable for automated threat recognition.

Visualization of detected features

Visualization of CNN gives insight into the function of intermediate feature layers and the operation of the classifier. Deconvolutional Network[3] is one way to visualize the input pattern that caused given activation in feature maps by mapping the activation back to input space. Fig.1 shows the features of the threat item evolved over the layers of CNN.

Automated Threat Detection

The fixed input size (256x256) Ensemble network was unsuitable for real time prediction required in security applications as sliding window detection needs to be applied for variable sized X-Ray baggage scanner images. Also, the

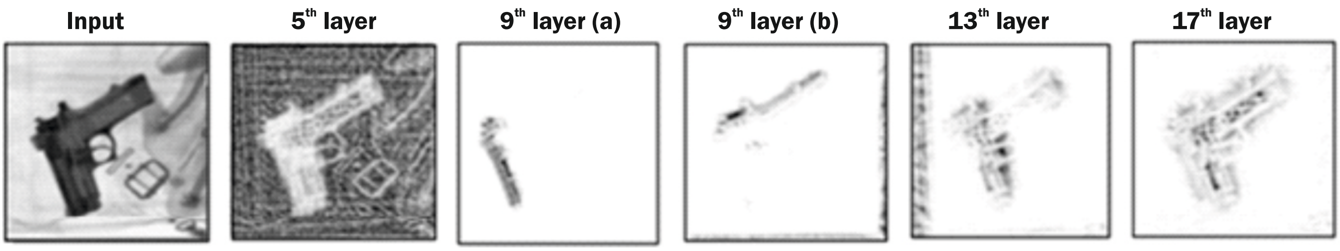


Fig.1: 5th Layer: Detection of edges, 9th Layer: parts of object, 17th Layer: Detection of object ignoring background.

Table 3: Ensemble Network Performance.

Model	VA	TA	FP	FN	Precision	Recall	F - measure	AUC - ROC	Accuracy on toy guns
VGG -19	0.992	0.9905	32	95	0.994	0.9852	0.99	0.9924	0.9720
AlexNet	0.986	0.9845	77	131	0.987	0.9796	0.9837	0.9851	0.9702
Ensemble Network	**0.995**	**0.9937**	**24**	**60**	**0.9962**	**0.9906**	**0.9934**	**0.9936**	**0.9825**

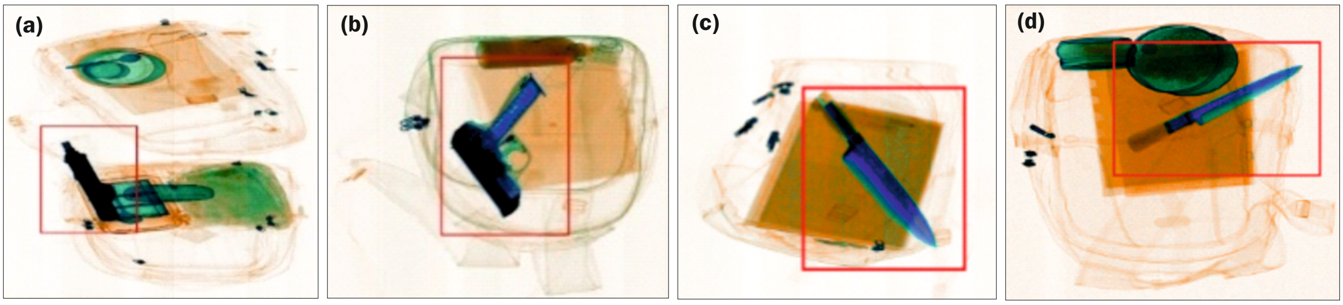


Fig.2: Auto detection of Guns (a, b) and Knives (c, d).

localization could only be as accurate as the bounding box provided by sliding window size (256x256). Thus, object detection (i.e. both, object recognition and object localization) model with variable input size, was required for automated threat detection in X-Ray images in real time. Such a model is also more effective in detecting multiple threats in a single image. Hence, YOLOv2, its improved version YOLOv3 and the smaller version Tiny-YOLOv3 for constrained environments, were considered for its applicability in threat detections in dual energy X-Ray baggage images.

Data Augmentation and Threat Projection as discussed in section 2.1 was used to obtain a balanced dataset. Ground truth labels for training object detection models were assigned to each threat object projected image. The low energy and high energy channels of clean bag images and threat object projected images were fused together to form pseudo coloured RGB images identifying material categories (Organic, Mixture, Inorganic and Dense). 'Spot the difference'[4] using threat and benign pairs was applied to the training set and testing set of 2280 images and 772 images respectively.

Results and Discussion

YOLOv2, YOLOv3 and Tiny-YOLOv3 models, initialised using pre-trained ImageNet weights, were trained on the training dataset created as explained above in 'Data augmentation for threat recognition' section. Table 4 shows the comparison of YOLOv2 model and YOLOv3 model trained with input resolution 416 x 416 pixels and YOLOv3 model trained with random input resolution, using multiscale training. In multiscale training, for every 10 batches, the network was trained on randomly chosen image sizes from range 320x320 pixels to 608x608 pixels as multiple of 32 pixels. The improvement in mAP (99.17%), accuracy (98.18%) and F-1 score (98.17%) shows that YOLOv3 model trained using multiscale option is most suitable for automated threat detection in material discriminated X-Ray baggage images.

While scanning a typical baggage in X-Ray inspection system, the baggage is displayed on the inspection screen for 2.5 seconds on an average. The best performing YOLOv3 model detects the threat in real time using GPUs (0.125 sec), but takes significantly more time without GPUs (around 3.6 sec). Tiny-YOLOv3 a compact variant model takes around 0.0149 sec on GPU and 0.369 sec on CPU which is well within the real time prediction limit for X-Ray baggage inspection system. Thus, trained YOLOv3 model with 99.17% mAP was found best performing for automated threat detection task, whereas Tiny-YOLOv3 was found best suitable for real time threat detection in non-GPU environments. A Trade-off was observed between model performance in terms of mAP and average prediction time. Table 5 shows average prediction time of YOLOv3 and Tiny-YOLOv3 trained models on: Intel 4th generation (Haswell) i-7 processor non-GPU system and NVidia Tesla K80 GPU with 4096 cores.

Table4: Performance of YOLO models.

Model	Resolution	mAP	Accuracy	F 1- score	IOU
YOLOv2	416 x 416	97.99%	96.89%	96.92%	73.98%
YOLOv3	416 x 416	98.63%	97.80%	97.79%	78.68%
YOLOv3	608 x 608 Multiscale	99.17%	98.18%	98.17%	78.01%

Table4: Average Prediction Time.

Model	mAP	Accuracy	F 1-score	IOU	Prediction Time sec (GPU)	Prediction Time sec (CPU)
YOLOv3	99.17%	98.18%	98.17%	78.01%	0.125	3.598
Tiny YOLOv3	96.34%	94.30%	96.72%	73.82%	0.0149	0.369

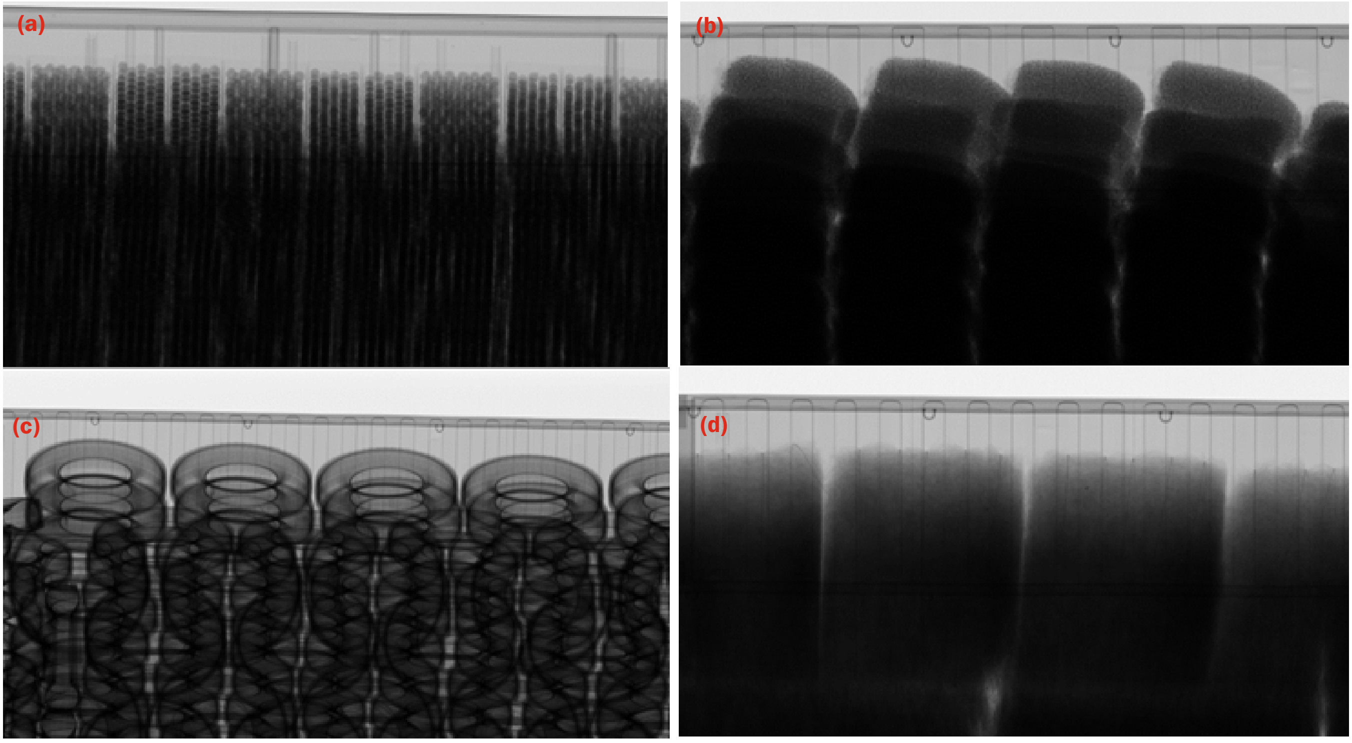


Fig.3: Content classification: A) Apple, B) Dates, C) Tyre, D) Paper.

Integration with XBIS Software

The model has been integrated as a pluggable add-on with the Single View X-ray Baggage Inspection System software(XBIS) developed by Electronics Division. Real scanned bag images obtained from XBIS machines deployed at site were superimposed with threat images and continuous scanning of a mixture of benign and threat-containing bags was done to evaluate performance during real scans. YOLO model could successfully detect guns and knives in real-time (Fig.2).

Auto Content Verification

Volume of cargo containers is several times larger than a regular luggage. Moreover, cargo images are cluttered and they maybe homogeneous or heterogeneous. Mutual overlap between objects and different stacking modes increase image complexities and makes manual inspection of cargo images a visually challenging task. Machine learning based object classification approach for autocontent verification of the Cargo under inspection has been used. A dataset containing over 3000 images per category of Apples, Dates, Tyre and Paper was subjected to automated image segmentation and classification using machine learning. Segmentation was done using texture, morphological and Statistical Region Merging algorithms. Classification of images using Bag of Words and Bayesian methods gave 86% accuracy. Second approach using deep learning (VGG-19 model) achieved accuracy of 97.8%. Fig.3 shows the classification.

Conclusions

In this article, following are presented i) Convolutional Neural Networks to classify dual energy X-ray images of baggage in threat and non-threat classes for threat object, ii) YOLO models to recognise as well as localize threat object in dual energy X-Ray baggage images and iii) Classical machine learning and deep learning models for automated content verification in cargo scanner X-ray images. It is found that, when trained on a large and balanced dataset, deep learning models prove to be a practical tool for automated image classification and object detection. In this study, it is concluded that, the use of ensemble network improved the accuracy over base models. We studied the trade-off between learning capacity and generalization capacity of a model by varying the number of layers of VGG-19 model and found that testing accuracy reached its peak at 20 layers.

References

[1] T. Rogers, J. Nicolas, J. Ollier and L. Griffin, “Threat Image Projection (TIP) into X-ray images of cargo containers for training humans and machines,” in Security Technology(ICCST), 2016 IEEE International Carnahan Conference, Orlando, FL, USA, 2016.

[2] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” in International Conference for Learning Representations, San Diego, 2015.

[3] M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional Networks,” in ECCV 2014, Zurich, 2014.

[4] T. Rogers, N. Jaccard and L. Griffin, “A Deep Learning Framework for the automated inspection of complex dual energy x-ray cargo imagery,” in Anomaly Detection and Imaging with X-Rays (ADIX) III, Orlando, 2017.