On the Use of Deep Learning for the Detection of Firearms in X-ray Baggage Security Imagery

Yona Falinie A. Gaus[1], Neelanjan Bhowmik[1], Toby P. Breckon[1,2]
Department of {Computer Science[1] | Engineering[2]}, Durham University, UK

***Abstract*—X-ray imagery security screening is essential to maintaining transport security against a varying profile of prohibited items. Particular interest lies in the automatic detection and classification of prohibited items such as firearms and firearm components within complex and cluttered X-ray security imagery. We address this problem by exploring various end-to-end object detection Convolutional Neural Network (CNN) architectures. We evaluate several leading variants spanning the Faster R-CNN, Mask R-CNN, and RetinaNet architectures. Overall, we achieve maximal detection performance using a Faster R-CNN architecture with a ResNet$_{101}$ classification network, obtaining 0.91 and 0.88 of mean Average Precision (mAP) for a two-class problem from varying X-ray imaging dataset. Our results offer very low false positive (FP) complimented by a high accuracy (A) (FP=0.00%, A=99.96%). This result illustrates the applicability and superiority of such integrated region based detection models within this X-ray security imagery context.**

***Index Terms*—CNN architecture, Object detection, X-ray security imagery, Prohibited item**

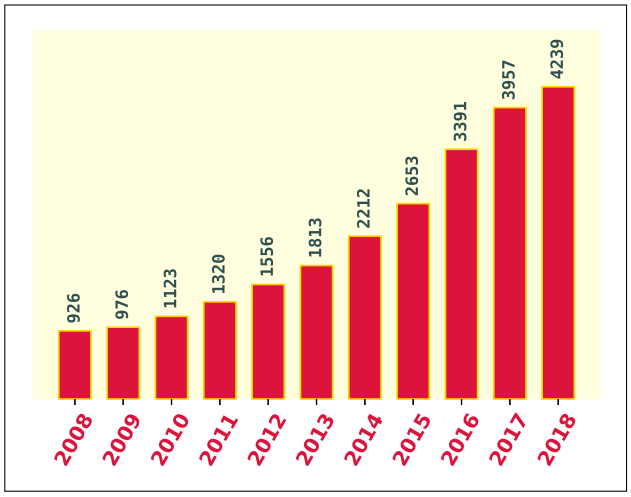

Fig. 1. Firearms discovered at TSA checkpoints, 2008-2018 [4].

I. Introduction

The Transportation Security Administration (TSA) is established to ensures the safety of the travelling public within the US, including the over 2.6 million aviation passengers daily across all US airports [1]. Subsequently, TSA screens approximately 4.9 million carry-on bags for explosives and other prohibited items daily [2]. Recently, TSA announced that it found 4,239 firearms in carry-on bags in the year 2018, of which 3,656 were loaded - the largest number in TSA history. Statistically, this is a 7% increase over 2017 and part of an increasing annual trend, as illustrated in the Figure 1. This implies an average of around 12 firearms were intercepted at security screening each day within the USA in 2018 [3].

When a firearm is discovered by TSA officers at a checkpoint, it is immediately reported to local law enforcement. While firearm possession laws vary by locale, TSA may impose civil penalties of up to $13,333 per incident. With increasing firearm occurrences at airport checkpoints, the security and safety of the passengers has become a prime concern. Within all the screening options available at the airports, multiple view, dual-energy X-ray screening is widely used as the primary means to maintain aviation and transport security. By using such X-ray screening, it will produce a colour-mapped image showing the material properties for the security officer to examine. This current current concept of operation presents a very challenging human image analysis task because this process of screening baggage in public places is highly complex, due to the nature of compact, cluttered and highly varying baggage contents to be examined within the limited available time. It can also be challenging due to inherent monotony leading to a varying concentration levels [5].

Considering the key challenge of firearm detection in compact cluttered X-ray imagery, we examine three different approaches to illustrate an automated pipeline for firearm detection in X-ray security imagery using recent advances in deep learning. Current scanners already implement algorithms that calculate material properties from dual-energy multiple view scans, automatically highlighting objects or regions that might contain prohibited material threat within the X-ray imagery. However, within baggage X-ray screening, the automated detection of prohibited items by their shape characteristics is still an open problem due to factors such as limited robustness, loss of generality and required optimization between false alarms and probabilty of detection [6].

Therefore, in this work we fully leverage the rapid advance of deep learning for image understanding [7], building on the original transfer learning based work of [8], for this X-ray imagery screening task. Building on our earlier work on this topic [8], [9], our goal is to detect and automatically highlight (i.e. localize) prohibited items within X-ray baggage security imagery whilst determining the varying performance of these approaches on different datasets collected for this study. In this approach, we fully utilize large-scale X-ray security imagery provided by [9] and [10]. Denoted as *Dbf2* [9]

978-1-7281-5092-5/19/$31.00 ©2019 IEEE

and *SIXray* [10] respectively, we aim to provide an insight into baseline performance for three CNN architectural variants:- Faster R-CNN [11], Mask R-CNN [12] and RetinaNet [13]. Although our prior work has considered a broader definition of prohibited items [8], [9], our discussion will be limited to the detection of firearms and firearm components.

II. Related Work

There has been a steady increase in research work considering prohibited item detection based CNN architecture variants in X-ray baggage security imagery. Preliminary work on [8] classifies baggage objects by type, by optimizing CNN structure designed by Krizhevsky *et al.* [14] to the X-ray baggage screening problem. The work of [8] shows that by leveraging the use of transfer learning, CNN architectures outperformed hand-crafted features, by achieving 98.92% detection accuracy in firearm classification. The comparison of CNN architecture and hand-crafted features were further evaluated in [15], where hand-crafted features such as bag-of-visual-words (BOVW), sparse representation and codebooks is compared with deep features. Consistent with results in [8], CNN architecture achieve superior results with more than 95% of recognition rate in classification of three prohibited items $\{$*Gun, Shuriken, Razor blade*$\}$. The work on [9] exhaustively compares various CNN architecture and the impact of its overall performance in classifying six prohibited items $\{$*Firearm, Firearm Component, Ceramic Knife, Laptop, Camera, Knife*$\}$. Fine tuning the entire network architecture for this problem domain yields 0.996% true positive, 0.011% false positive and 0.994% accuracy for prohibited item classification tasks.

Techniques such as transfer learning for CNN architectures have shown themselves capable on classifying prohibited item even when the X-ray security imagery dataset is very limited [16]. However, in order to detect and automatically highlight the region that might contain prohibited item, such as abnormalities in an object or any explosive devices [17], an advanced CNN architecture are needed. With the recent development of object-based CNN architecture detection approaches, [9] examines the relative performance of contemporary region-based CNN variants in X-ray security imagery [18] [19] [11]. Methods such as Faster R-CNN [11], R-FCN [18], and YOLOv2 [19], achieves maximal 0.885 and 0.974 mean Average Precision (mAP) over six-class object detection $\{$*Firearm, Firearm Component, Ceramic Knife, Laptop, Camera, Knife*$\}$ and and two-class firearm detection $\{$*Firearm, Firearm Component*$\}$ problems respectively. Here, we further investigate this performance with our evaluation on X-ray security imagery on larger dataset taken from different source, namely Durham University Full Two-Class and Security Inspection X-ray images [10]. Denoted as *Dbf2* [9] and *SIXray* [10] respectively, it provide more inter-occluding X-ray security imagery examples with large variations in pose, scale and item construction.

III. Proposed Approach

The task on automatic prohibited item detection and localisation follows the recent work of [9], where we employ the convolutional neural network based object detection on X-ray security imagery. While most of the work on CNN architectures is based on natural photographic colour images, X-ray security imagery posses certain unique aspects. X-ray security scan images are produced by transmission, where photons pass completely through the baggage item. This means that individual items can overlap on top of each other, in any pose, orientation and scale. Deep learning has already demonstrated its capability for x-ray imagery security screening [20] [9] [17]. Contemporary region based CNN variants such as Faster R-CNN [11], R-FCN [18] and YOLOv2 [19] were adapted in [9] for detection and classification of prohibited item in X-ray security imagery. Here we extend the capability of region based CNN variants by incorporating Faster R-CNN [11], Mask R-CNN [12] and RetinaNet [13] for prohibited item detection.

A. Faster R-CNN

Faster R-CNN [11] is an object detection algorithm which combine Fast R-CNN [21] and Region Proposal Network (RPN). This architecture has its own region proposal network, which is consists of convolutional layers that generate object proposals and two fully connected layers that predict coordinates of object bounding boxes. The corresponding locations and bounding boxes are then fed into objectness classification and bounding box regression layers. Finally the objectness classification layer classify whether a given region proposal is an object or a background region while a bounding box regression layer predicts object localisation, at the end of the overall detection process.

B. Mask R-CNN

Mask R-CNN [12] is build upon Faster R-CNN [11], where Faster R-CNN is augmented by adding convolutional layers to construct an object boundary segmentation mask. This is done by adding an additional branch to Faster R-CNN that outputs an image mask indicating pixel membership of a given detected object. Mask R-CNN also address feature map misalignment in Faster R-CNN [11] by applying bi-linear boundary interpolation onto feature map boundaries.

C. RetinaNet

RetinaNet is another object detection approach [13] where the key idea is to solve the extreme class imbalance between foreground and background class. To solve the imbalance classes, RetinaNet employs Focal loss as loss function, where it modifies the cross-entropy loss such that it down-weights the loss in easy negative samples so that the loss is instead focused on the sparse set of more challenging samples.

978-1-7281-5092-5/19/$31.00 ©2019 IEEE

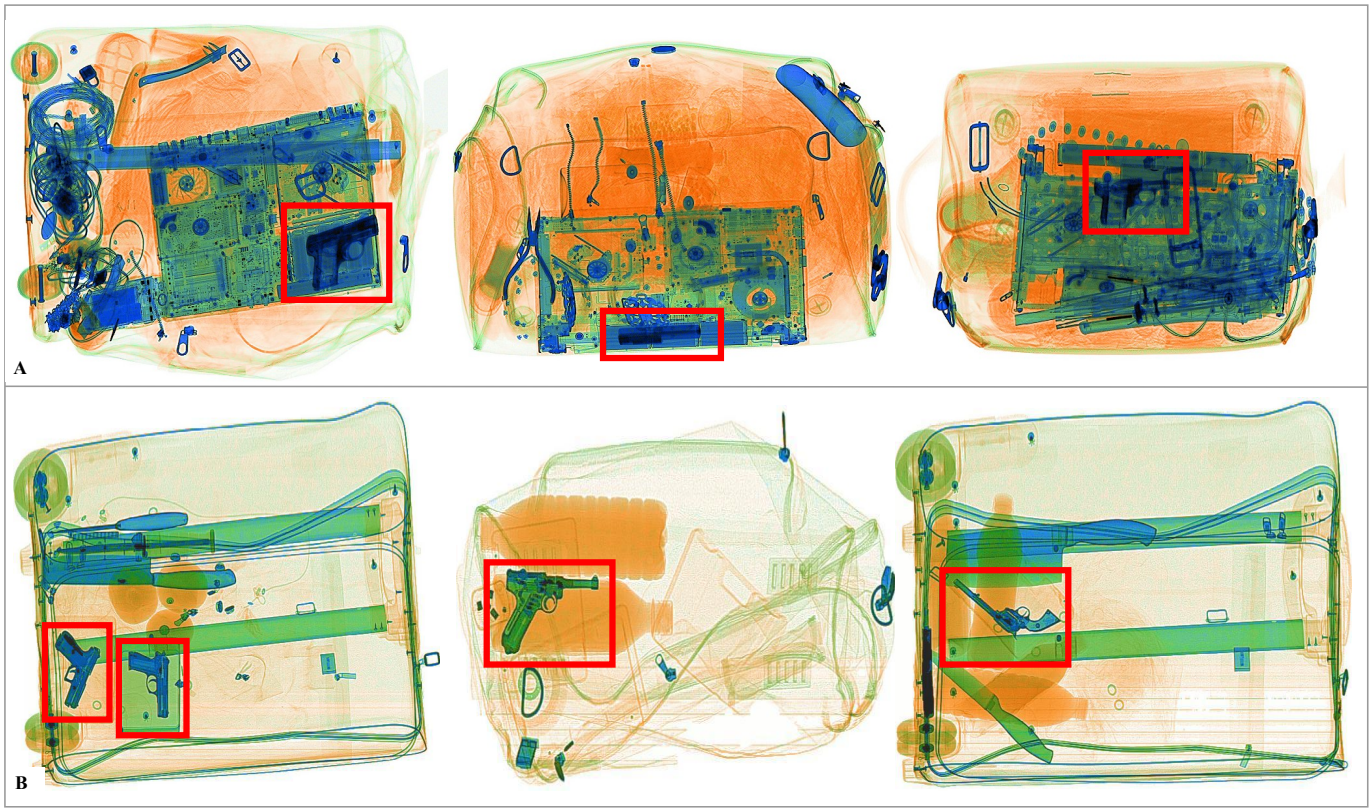

Fig. 2. X-ray baggage security imagery with threat objects (Firearm, Firearm Components) highlighted from dataset (A) *Dbf2* and (B) SIXRay10.

IV. Evaluation

The prohibited item detection task is carried out across the X-ray security imagery datasets to analyse the performance of different CNN detection architectures as presented in Section III.

A. Experimental Setup

Our evaluation comprises on two varying X-ray security imagery datasets.

Dbf2 is the X-ray security imagery from Durham Dataset Full Two-class, consists of of two classes, {*Firearm, Firearm Components*}. To generate the dataset, a systematic process was performed to acquire the dataset whereby metallic prohibited items such as firearms and firearm components are placed inside various styles of bags, which cover the full dimensions of airline cabin luggage. It consists of 3,192 images of *firearms* and 1,204 images of *firearm components*. Figure 2A depicts exemplary images from this dataset.

SIXray is public dataset [10] for prohibited item detection in security inspection X-ray images. These images are collected from several metro sub-way stations, using a Nuctech dual-energy X-ray scanner, where the distribution of the general baggage/parcel items corresponds to stream-of-commerce occurrence. We use a subset of SIXray dataset {*SIXray10*} where we incorporate 3,130 images of *firearms*, as depicted in in Figure 2B.

In our experiments, we use the CNN implementation of [22], where we trained the models on GTX 1080 Ti GPU, optimised by Stochastic Gradient Descent (SGD) with a weight decay of 0.0001, learning rate of 0.01 and termination at 180k iterations. The ResNet_{101} [23] is used as a network backbone. We split the datasets into training (60%), validation (20%) and test sets (20%) such that each split has similar class distribution. All experiments are initialised with ImageNet pre-trained weights for their respective model [14].

B. Results and Discussions

The performance is evaluated in terms of mean Average Precision (mAP) as well as by the comparison of True Positive Rate (TP) (%), False Positive Rate (FP) (%) together with Precision (P), Accuracy (A) and F-score (F) (harmonic mean of precision and true positive rate), following the statistical evaluation of [9].

Table I shows prohibited items detection in X-ray security imagery using the CNN architecture set out in Section III with network configuration of ResNet_{101}. The AP/mAP in bold in the tables signifies the best results. In the first set of experiments (Table I, upper), we observe maximal mAP performance (mAP = 0.89) is achieved on *Dbf2* by Faster R-CNN with ResNet_{101} configuration, in two classes, {*Firearm, Firearm Components*}. In the second set of experiments (Table I, lower), RetinaNet architecture with ResNet_{101} configuration outperforms other CNN architectures. As reported in the

978-1-7281-5092-5/19/$31.00 ©2019 IEEE

TABLE I

DETECTION RESULTS OF FASTER R-CNN [11], MASK R-CNN [12] AND RETINANET [13] ON TWO DATASETS. CLASS NAME REFLECTS CORRESPONDING AVERAGE PRECISION (AP) FOR THE INDIVIDUAL OBJECT CLASS AND MAP IS THE MEAN AVERAGE PRECISION ACROSS ALL OBJECT CLASSES.

Datasets	Model	Network configuration	Average precision		mAP
			Firearm	Firearm Components	
Dbf2	Faster R-CNN [11]	ResNet_{101}	**0.91**	**0.88**	**0.89**
	Mask R-CNN [12]	ResNet_{101}	0.89	0.86	0.88
	RetinaNet [13]	ResNet_{101}	0.89	0.86	0.88
SIXray10	Faster R-CNN [11]	ResNet_{101}	0.91	–	0.91
	Mask R-CNN [12]	ResNet_{101}	0.89	–	0.89
	RetinaNet [13]	ResNet_{101}	**0.92**	–	**0.92**

TABLE II

CLASSIFICATION RESULTS OF FASTER R-CNN [11], MASK R-CNN [12] AND RETINANET [13] ON TWO DATASETS.

Datasets	Model	Network configuration	A	P	F1	TP	FP
Dbf2	Faster R-CNN [11]	ResNet_{101}	**99.96**	**100**	**99.93**	99.87	**0.00**
	Mask R-CNN [12]	ResNet_{101}	99.93	99.78	99.89	100	0.11
	RetinaNet [13]	ResNet_{101}	97.25	100	95.69	91.74	0.00
SIXray10	Faster R-CNN [11]	ResNet_{101}	**99.83**	99.68	**99.84**	**100**	0.36
	Mask R-CNN [12]	ResNet_{101}	99.66	99.68	99.68	99.68	0.36
	RetinaNet [13]	ResNet_{101}	90.96	99.81	90.74	83.17	0.18

prior work of [10], the highest AP achieved for the class {*Firearm*} is 90.64% with ResNet_{50} whereby our proposed CNN architecture produces a marginally superior AP of 0.92.

Table II shows the binary threat classification results for prohibited item detection in X-ray security imagery. In the first set of experiments (Table II, upper), we classify whether prohibited items, {*Firearm, Firearm Components*}, are present in the X-ray security imagery *Dbf2*. We observe performance such that all of the CNN architecture consistently offer very low false positive (FP) complimented by a high true positive (TP) classification across both problems. With a *Dbf2* trained model, Faster R-CNN with ResNet_{101} achieves maximal performance with the lowest FP (0.00%) and accuracy of (99.96%). In the second sets of experiments, (Table II, lower), we classify whether prohibited item of {*Firearm*} is present in the X-ray security imagery *SIXray*, or it just benign X-ray imagery security. With *SIXray* as trained model, RetinaNet offers low FP (0.18%) while Faster R-CNN gives the highest TP (100.00%). These results shows a clear impact on CNN architecture in detecting prohibited item under X-ray images. It also clearly illustrates the applicability and performance capability of deep learning within cluttered X-ray security imagery context.

V. CONCLUSION

This paper assesses the performance of contemporary region based CNN object detection variants within the context of prohibited item detection in X-ray security imagery. Experimentation demonstrates that Faster R-CNN achieves superior performance (0.91/0.88 mAP) over a two class prohibited item detection problem {*Firearm, Firearm Components*}. On the other hand, RetinaNet achieves superior results (0.92 mAP) over one class prohibited item detection problem {*Firearm*} evaluated on two disparate datasets. Furthermore, we directly evaluate the prohibited item detection models in classification mode, where it gives maximal performance with the lowest FP (0.00). This provides strong insights that artificial intelligence techniques such as CNN models can detect prohibited items with high accuracy and at the same time with minimal false alarm rates. Ultimately, the goal of this approach is to lessen the burden on the security office by automating threat detection system in aviation security checkpoint. While in this paper we only focusing on firearm and firearm components, given the promise shown by artificial intelligence in this initial study, we are well positioned to bring artificial intelligence security system in the near future.

Acknowledgements: The authors would like to thank the UK Home Office for partially funding this work. Views contained within this paper are not necessarily those of the UK Home Office.

978-1-7281-5092-5/19/$31.00 ©2019 IEEE

Fig. 3. Threat item (Firearm, Firearm Components) detection in X-ray security imagery by using CNN architecture such as Faster R-CNN [11], Mask R-CNN [12] and RetinaNet [13] on *Dbf2*.

978-1-7281-5092-5/19/$31.00 ©2019 IEEE

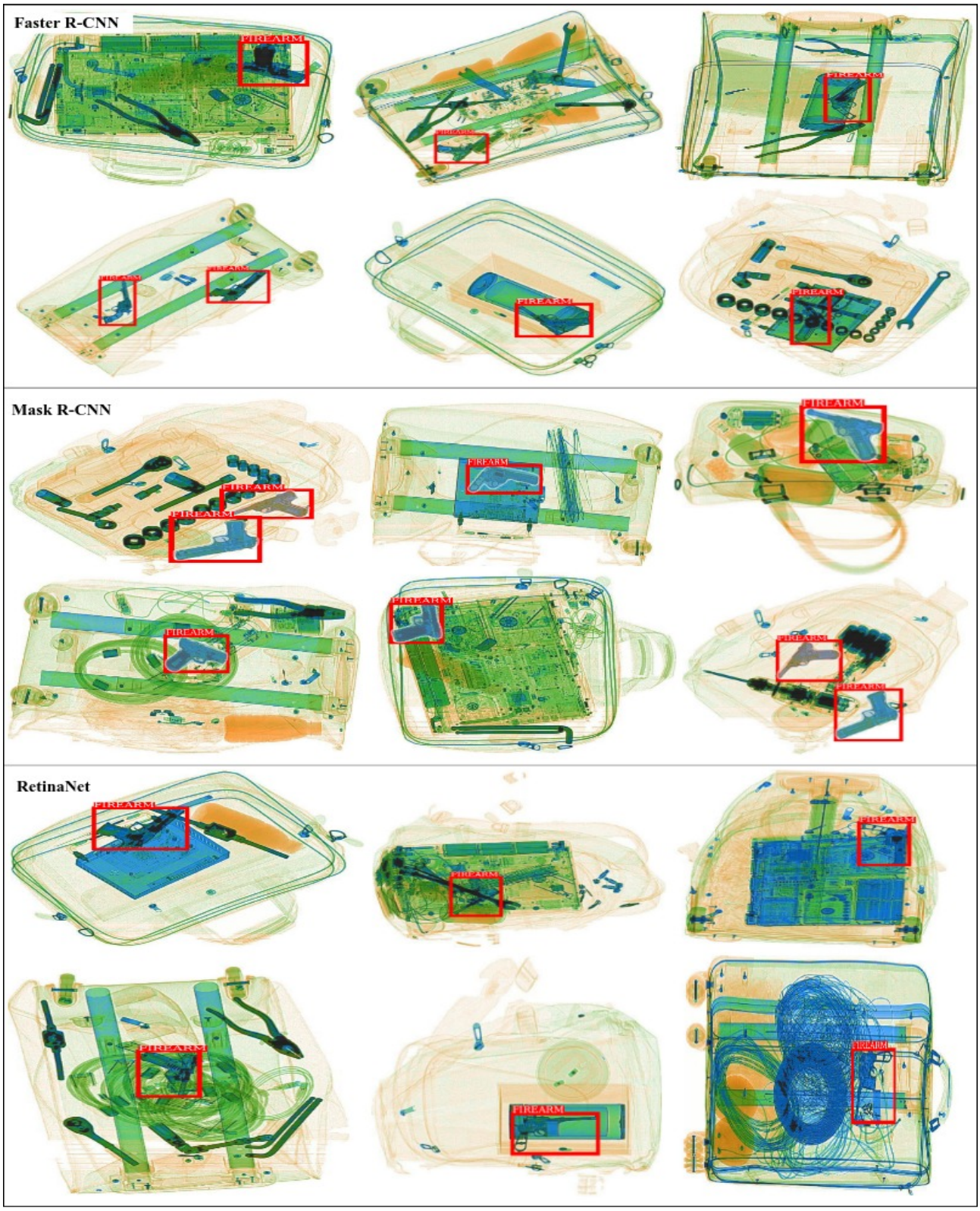

Fig. 4. Threat item (Firearm) detection in X-ray security imagery by using CNN architecture such as Faster R-CNN [11], Mask R-CNN [12] and RetinaNet [13] on *SIXray10*.

978-1-7281-5092-5/19/$31.00 ©2019 IEEE

REFERENCES

[1] "Air traffic by the numbers," https://www.faa.gov/air_traffic/by_the_numbers/, accessed: 2019-08-20.

[2] "Security screening," https://www.tsa.gov/travel/security-screening, accessed: 2019-08-20.

[3] "Tsa finds record number of guns carried by travelers at airports," https://www.forbes.com/sites/garystoller/2019/02/07/tsa-finds-record-number-of-travelers-carrying-firearms-at-airports/#6c22c6065588, accessed: 2019-08-20.

[4] "Tsa year in review: A record setting 2018," https://www.tsa.gov/blog/2019/02/07/tsa-year-review-record-setting-2018, accessed: 2019-08-20.

[5] A. Bolfing, T. Halbherr, and A. Schwaninger, "How image based factors and human factors contribute to threat detection performance in x-ray aviation security screening," in *HCI and Usability for Education and Work*, 2008, pp. 419–438.

[6] D. Mery and A. K. Katsaggelos, "A logarithmic x-ray imaging model for baggage inspection: Simulation and object detection," in *Conference on Computer Vision and Pattern Recognition Workshops*, July 2017, pp. 251–259.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[8] S. Akçay, M. E. Kundegorski, M. Devereux, and T. P. Breckon, "Using convolutional neural networks for object classification within x-ray baggage security imagery," in *International Conference on Image Processing*, 2016, pp. 1057–1061.

[9] S. Akcay, M. E. Kundegorski, C. G. Willcocks, and T. P. Breckon, "Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security imagery," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2203–2215, 2018.

[10] C. Miao, L. Xie, F. Wan, C. Su, H. Liu, J. Jiao, and Q. Ye, "Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images," in *Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2119–2128.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems 28*, 2015, pp. 91–99.

[12] K. He, G. Gkioxari, P. Dollr, and R. Girshick, "Mask r-cnn," in *International Conference on Computer Vision*, Oct 2017, pp. 2980–2988.

[13] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollr, "Focal loss for dense object detection," in *International Conference on Computer Vision*, Oct 2017, pp. 2999–3007.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[15] D. Mery, E. Svec, M. Arias, V. Riffo, J. M. Saavedra, and S. Banerjee, "Modern computer vision techniques for x-ray testing in baggage inspection," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 4, pp. 682–692, April 2017.

[16] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 1717–1724.

[17] Y. F. A. Gaus, N. Bhowmik, S. Akçay, P. M. Guillen-Garcia, J. W. Barker, and T. P. Breckon, "Evaluation of a dual convolutional neural network architecture for object-wise anomaly detection in cluttered x-ray security imagery," in *International Joint Conference on Neural Networks*, July, 2019.

[18] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," in *Advances in Neural Information Processing Systems 29*, 2016, pp. 379–387.

[19] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *Conference on Computer Vision and Pattern Recognition*, July 2017, pp. 6517–6525.

[20] S. Akcay and T. P. Breckon, "An evaluation of region based object detection strategies within x-ray baggage security imagery," in *International Conference on Image Processing*, Sep. 2017, pp. 1337–1341.

[21] R. Girshick, "Fast r-cnn," in *International Conference on Computer Vision*, Dec 2015, pp. 1440–1448.

[22] R. Girshick, I. Radosavovic, G. Gkioxari, P. Dollár, and K. He, "Detectron," https://github.com/facebookresearch/detectron, 2018.

[23] S. Xie, R. Girshick, P. Dollr, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Conference on Computer Vision and Pattern Recognition*, July 2017, pp. 5987–5995.

978-1-7281-5092-5/19/$31.00 ©2019 IEEE