Multi-Class 3D Object Detection Within Volumetric 3D Computed Tomography Baggage Security Screening Imagery

Qian Wang Durham University Durham, UK

Neelanjan Bhowmik Durham University Durham, UK

Toby P. Breckon Department of Computer Science Department of Computer Science Department of {Computer Science | Engineering} Durham University Durham, UK

Abstract-Automatic detection of prohibited objects within passenger baggage is important for aviation security. X-ray Computed Tomography (CT) based 3D imaging is widely used in airports for aviation security screening whilst prior work on automatic prohibited item detection focus primarily on 2D Xray imagery. Whilst some prior work has proven the possibility of extending deep convolutional neural networks (CNN) based automatic prohibited item detection from 2D X-ray imagery to volumetric 3D CT baggage security screening imagery, it focuses on the detection of one specific type of objects (e.g., either bottles or handguns). As a result, multiple models are needed if more than one type of prohibited item is required to be detected in practice. In this paper, we consider the detection of multiple object categories of interest using one unified framework. To this end, we formulate a more challenging multi-class 3D object detection problem within 3D CT imagery and propose a viable solution (3D RetinaNet) to tackle this problem. To enhance the performance of detection we investigate a variety of strategies including data augmentation and varying backbone networks. Experimentation carried out to provide both quantitative and qualitative evaluations of the proposed approach to multi-class 3D object detection within 3D CT baggage security screening imagery. Experimental results demonstrate the combination of the 3D RetinaNet and a series of favorable strategies can achieve a mean Average Precision (mAP) of 65.3% over five object classes (i.e. bottles, handguns, binoculars, glock frames, iPods). The overall performance is affected by the poor performance on glock frames and *iPods* due to the lack of data and their resemblance with the baggage clutter.

Index Terms—3D volumetric data, deep convolutional neural network, X-ray computed tomography, baggage data, multi-class 3D object detection.

I. INTRODUCTION

X-ray baggage security screening is widely used to maintain aviation security. Currently, multi-view X-ray is predominantly used in aviation security for cabin baggage screening. This traditional baggage screening process, using 2D X-ray scanners, has the disadvantage of both inter-object occlusion and clutter within any given image projection of the scanned baggage item. As a result, it poses a considerably challenging visual search task for the human operators to discover prohibited items (e.g., liquids, firearms, knives, etc.) overlapped with other benign items (e.g., electronic devices) within a constrained time frame. For this reason, passengers are currently

required to divest large electronic devices and liquids which decreases checkpoint throughput significantly. Furthermore, human operator performance can be subjective and is heavily affected by many factors such as the experience, fatigue, monotony and concentration, although many successful measures have been taken to alleviate the problem in practice (e.g., Threat Image Projection (TIP) [1], [2] and shorter shift rotations [3]).

By leveraging recent advances in object classification and detection, significant progress has been made in automatic prohibited item detection within 2D X-ray imagery [4]. The use of deep learning techniques allows real-time and accurate detection of prohibited items even in cluttered X-ray images [5]-[7]. However, performance can be affected when the baggage contains significant clutter and inter-object occlusion due to the fundamental limitation of projected 2D X-ray imagery. To improve the detection rate without affecting the checkpoint throughput, airports are currently increasing the use of 3D CT screening which does not require the removal of electronic devices and liquids during baggage screening. The reconstructed 3D CT images provide more information and make it possible for the human operators to inspect the 3D CT images from differing views. However, current technology does not facilitate the automatic detection of (non-explosive) prohibited items such as prohibited items and liquid containers. Prior work [8] has shown the possibility of using deep 3D CNN models for object classification and detection within baggage security imagery. However, this study was limited to the detection of only one specific object category (i.e. either bottles or handguns) by one model. It is unknown how the detection performance will be affected for multi-class object detection in a unified framework and which strategies are beneficial to the enhancement of detection performance in 3D CT baggage security screening imagery?

To answer the above questions, in this paper we extend the prior work [8] in single-class object detection to a more challenging multi-class object detection problem in a unified framework and propose a viable solution to this problem. The proposed approach is evaluated in real volumetric 3D CT baggage security screening imagery to get insightful observations and conclusions for this emerging research topic. Specifically, we investigate different CNN architectures (i.e. ResNet [9]) with variable depths under the RetinaNet object detection framework [10]. We also evaluate the effectiveness of data augmentation techniques including 3D volume flipping and rotation.

The contributions of this work are summarized as follows:

- a unified framework using deep CNN models for multiclass prohibited item detection within volumetric 3D CT baggage imagery;
- an evaluation of different 3D CNN models in the detection of prohibited items within volumetric 3D CT baggage imagery and the effect of data/feature augmentation.

II. RELATED WORK

Automatic object detection and recognition algorithms have been proposed and evaluated for baggage aviation security screening based on 2D X-ray images [1], [4]. The use of CNN architectures and object detection frameworks boosts the performance with a high detection rate and a low false positive rate. For instance, Gaus et al. [11] evaluate the effectiveness of Faster R-CNN [12], Mask R-CNN [13] and RetinaNet [10] in detecting six different objects (i.e. bottle, hairdryer, iron, toaster mobile and laptop) in 2D X-ray baggage images.

To enable automatic baggage screening using 3D CT imagery, a variety of studies have been carried out in recent years [2], [14]–[21].

One research direction is object segmentation based on the material and morphological structure [14], [19], [21]. Specifically, Mouton et al. [19] propose a two-stage approach for object segmentation within 3D CT imagery. A CT volume is firstly coarsely segmented based on the voxel intensity ranges of pre-defined materials. Subsequently, a variety of shape descriptors are computed as features for the random forest classifier to determine a segment resulted from the first stage is good (containing only one object) or bad (containing multiple objects and hence need further segmentation). Wang et al. [21] studied the issue of object segmentation and classification in 3D CT imagery and focused mainly on the material characteristics without considering any specific prohibited item (e.g., firearm, knife, etc.). An approach to 3D segmentation is proposed based on recursive morphological operations and the Support Vector Machines (SVM) were employed for the classification of three types of materials.

3D object detection within 3D CT baggage security screening imagery has been studied in [8], [22], [23]. Flitton et al. [23] evaluate the effectiveness of different 3D descriptors in a search-based detection approach. Their approach is limited to detect known objects for which the reference data are assumed to be available. Such an assumption hinders its application in practice when the reference data are usually unavailable. Wang et al. [8] use contemporary object detection frameworks based on 3D CNN and evaluate its performance on individual object detection independently. Based on this work, we present a unified framework for multi-class object detection within 3D CT imagery for baggage security screening.



Fig. 1. 3D RetinaNet object detection framework with ResNet [9] as the backbone model.

III. METHOD

RetinaNet is employed in this study since it has proved better than its counterpart Faster R-CNN in the prior work [8]. We extend the RetinaNet framework to the 3D version used for 3D object detection in our study. Subsequently, data augmentation is described as a favourable technique to boost detection performance.

A. 3D RetinaNet

RetinaNet [10] is one of the most successful object detection frameworks designed for 2D natural images. Follow the same spirit, we extend it for 3D object detection within 3D CT imagery. As shown in Figure 1, our 3D RetinaNet consists of a feature pyramid network (FPN) implemented by ResNet [9] and a 3D bounding box regression and classification module which is implemented by a few 3D convolutional layers.

The ResNet based FPN is formed by four ResNet blocks and the 3D feature volumes (corresponding to the feature maps in 2D) output from these four blocks are considered as {C2, C3, C4, C5} which have strides of {4,8,16,32} voxels with respect to the input volume. The top-down pathway and lateral connections are used to enhance the features generated in the bottom-up pathway (i.e. C2-5). The top-down feature volumes corresponding to C2-C5 are denoted as P2-P5. The highestlevel feature volumes P5 are generated by a 3D convolutional layer with the stride of 1 and the kernel size of 3 from input C5. The feature volumes P4 are the summation of upsampled P5 and the output of a 3D convolutional layer with C4 as the input. Similarly, the feature volumes P3 and P2 are calculated.

Multi-scale feature volumes {P2, P3, P4, P5} are fed into the bounding box regression and classification module. The module consists of a branch for 3D bounding box regression and a branch for classification. These two branches have the same architecture with four 3D convolutional layers and an output layer. The output of the regression branch is a 3D



Fig. 2. An illustration of 3D threat image projection for synthetic CT image generation (a signnature of *binocular* is inserted into the bottom part of the baggage CT volume).

volume with $6 \times n_a$ channels corresponding to the bounding box biases with respect to the pre-defined anchor in a specific location. n_a denotes the number of anchors pre-defined in each location. The output of the classification is a 3D volume with $6 \times n_a$ channels corresponding to one background class and five foreground object classes under consideration in this study for each pre-defined anchor in a specific location.

The cross-entropy loss and smooth L_1 loss are used for the classification and regression respectively. The positive targets are calculated by comparing the pre-defined anchors against the ground truth bounding boxes with the Intersection Over Union (IOU) threshold of 0.1 which is also used as the threshold for detection during testing.

B. Data Augmentation

We investigate data augmentation for 3D volumetric CT data to enhance the object detection performance in our study. The 3D data augmentation strategies considered in this study are 3D threat image projection (TIP), data flipping and data rotation.

1) Threat Image Projection: Threat image projection is a technique used in baggage security screening for training human screeners and automatic threat recognition algorithms [1]. Specifically, TIP approaches superimpose a threat item signature onto a benign baggage image to generate a realistic synthetic baggage image containing threat objects. Recently, the technique has been extended to 3D volumetric CT imagery [2]. We employ the approach presented in [2] to generate synthetic 3D volumes containing objects of interest. The isolated objects are first extracted from a CT volume and then inserted to other target CT volumes to generate more volumes with the objects of interest. As illustrated in Figure 2, we use 3D TIP techniques to insert a signature of *binocular* into a baggage CT volume. We use this technique to address the issue of training data sparsity.

2) 3D Volume Flipping and Rotation: To alleviate the overfitting issue in training, we employ volume flipping and rotation to augment training data randomly. For 3D volumetric CT data, the flipping can be conducted in three planes (i.e. x-y, y-z and x-z). The rotation of a 3D volume is limited

to 90 degrees around a specific axis which can be easily implemented by swapping axes. Each type of flipping and rotation (6+6) has a probability of p to be activated during training so that the training data can be diversified significantly to alleviate the overfitting issue. We will investigate how the use of data augmentation and the value of p can affect detection performance.

IV. EXPERIMENTAL SETUP

In this section, we describe the experimental setup for the evaluation of multi-class object detection within baggage CT volumes. We describe the dataset used in our experiments and implementation details of the detection methods.

A. Dataset

We create a dataset for experimental evaluation with data collected from a CT80-DR dual-energy baggage-CT scanner manufactured by Reveal Imaging Inc. Five object categories (i.e. bottle, handgun, binocular, glock frame and iPod) are considered in our experiments to simulate a multi-class 3D object detection problem. Due to the limited number of instances of binocular, glock frame and iPod (i.e. 16, 29 and 12 respectively) in the original CT volume data, we use the 3D Threat Image Projection (TIP) technique proposed in [2] to generate synthetically composited 3D CT volumes containing these object signatures of interest. As a result, the dataset is a combination of 478 real CT volumes and 287 synthetically composited ones generated by the TIP algorithm. The dataset is randomly divided into two subsets for training (70%) and testing (30%) respectively. Three random splits are used throughout our experiments. The detailed numbers of different object signatures within the dataset and three splits are shown in Table I.

 TABLE I

 STATISTICS OF THE DATASET AND DATA SPLITS.

Object	Split 1 (train/test)	Split 2 (train/test)	Split 3 (train/test)	Total
Bottle	483/223	501/205	498/208	706
Handgun	269/108	263/114	267/110	377
Binocular	86/33	83/36	86/33	119
Glock frame	81/40	82/39	80/41	121
iPod	77/39	80/36	81/35	116

B. Implementation Detail

The detection models evaluated in this work are implemented in PyTorch [24] based on the work in [25]. In the experiments, we use the Adam [26] optimiser with an initial learning rate of 1e-3 for the first 100 epochs followed by a decreased learning rate of 1e-4 for 100 epochs and 1e-5 for last 100 epochs. This learning rate scheduler has been used throughout our experiments if not otherwise specified since it has been proved effective empirically in most cases. All experiments are conducted on a GTX 1080Ti GPU.

 TABLE II

 Multi-Class Prohibited Object Detection Results (P: precision; R: recall; mAP: mean Average Precision).

Model	Bo	ttle	Han	dgun	Bino	cular	Glock	Frame	iP	od	$m \Delta P (%)$
Widdei	P (%)	R (%)	P (%)	R (%)	P (%)	R (%)	P (%)	R (%)	P (%)	R (%)	
ResNet-10	80.2 ± 1.4	70.9 ± 3.6	77.5 ± 3.7	83.1 ± 1.9	75.4 ± 5.0	84.3 ± 5.0	72.3 ± 3.3	44.1 ± 7.0	29.4 ± 1.8	37.6 ± 7.2	58.2 ± 4.2
ResNet-18	81.6 ± 3.6	68.0 ± 4.0	78.4 ± 0.6	81.9 ± 1.1	81.4 ± 3.4	86.3 ± 8.7	69.7 ± 8.9	46.7 ± 7.0	28.3 ± 3.8	49.3 ± 6.1	57.3 ± 4.4
ResNet-34	84.8 ± 2.8	68.2 ± 2.9	80.8 ± 7.4	81.3 ± 1.2	80.9 ± 2.0	86.5 ± 7.9	78.5 ± 9.6	36.5 ± 10.5	33.1 ± 4.4	47.3 ± 1.9	57.7 ± 3.8
ResNet-50	81.0 ± 2.2	74.0 ± 3.1	74.8 ± 2.4	91.0 ± 1.1	78.4 ± 4.8	$\textbf{87.3} \pm \textbf{5.0}$	63.9 ± 7.7	53.5 ± 7.9	32.8 ± 2.3	$56.6~\pm~6.2$	65.3 ± 3.6
ResNet-101	80.6 ± 1.6	71.1 ± 2.8	71.3 ± 4.7	89.8 ± 0.5	75.5 ± 3.5	85.1 ± 6.8	70.6 ± 9.7	51.8 ± 4.9	30.2 ± 4.1	53.6 ± 1.7	62.8 ± 4.3

TABLE III EXPERIMENTAL RESULTS OF MULTI-CLASS OBJECT DETECTION WITH DIFFERENT DATA AUGMENTATION STRATEGIES.

Data Augmentation Average Precision (%)							
Flipping	Rotation	Bottle	Handgun	Binocular	GlockFrame	iPod	
X	×	65.3 ± 1.5	79.7 ± 1.5	69.2 ± 3.1	32.5 ± 3.8	28.2 ± 10.0	55.0 ± 1.2
0.2	×	67.5 ± 3.0	84.5 ± 0.3	81.5 ± 1.9	48.3 ± 4.8	37.6 ± 4.6	63.9 ± 2.6
×	0.2	70.1 ± 2.2	77.8 ± 1.4	74.9 ± 8.3	43.0 ± 10.2	28.1 ± 6.1	58.8 ± 5.0
0.2	0.2	70.0 ± 3.6	84.9 ± 0.3	83.7 ± 4.5	48.9 ± 7.1	38.9 ± 4.3	65.3 ± 3.6
0.5	0.5	62.8 ± 2.1	$85.6~\pm~2.2$	$\textbf{84.6} \pm \textbf{3.0}$	41.4 ± 8.7	27.0 ± 8.5	60.3 ± 2.6

V. EXPERIMENTAL RESULTS

Thorough experiments are conducted to evaluate the effectiveness of the proposed approach to multi-class object detection in 3D CT baggage security screening imagery. Specifically, we evaluate varying ResNet [9] architectures as the backbone FPN models, the effectiveness of data augmentation strategies, varying anchor sizes and scaling factors.

A. On the Backbone Networks

ResNet architectures [9] with variant depths are employed as the backbones for FPN in the RetinaNet. We investigate the effect of different backbone models (i.e. ResNet10, ResNet18, ResNet34, ResNet50 and ResNet101) in this experiment. The experimental results are shown in Table II. The precision and recall are reported for each object category with the mean and standard deviation over three splits. In addition, we also report the mean Average Precision (mAP) as the overall evaluation metric in Table II. We can see ResNet50 performs the best overall with a mAP of 65.3% over five object categories, followed by ResNet101 with a slightly lower mAP of 62.8%. The other three architectures with less depth perform comparably with one another with the mAP around 57-58%. Although ResNet50 achieves the best overall performance and the best recall rates, ResNet34 always results in better precision for all five object classes. By comparing the results of different object classes, we can see that Glock Frames and *iPods* have lower precision and recall than other three objects. This is due to the fact that glock frames are plastic hence more challenging to distinguish from background clutter within the baggage CT imagery whilst iPods, as a piece of electronic device, have less salient features to detect. In conclusion, the proposed approach, an extension of RetinaNet to 3D CT imagery, has the capability of detecting different objects within 3D CT baggage screening imagery but the performance varies across different object categories.

B. On the Data Augmentation

This experiment aims to investigate the effect of data augmentation strategies (i.e. volume flipping and rotation). It has shown using ResNet50 as the backbone network for FPN gives the best overall performance in the previous experiments, we use ResNet50 in this experiment. We compare the detection performance when no data augmentation is used and the performance when data augmentation is applied with different values of p (i.e. 0.5 and 0.2).

The experimental results are presented in Table III. When the data augmentation strategies are not used, a mAP of 55% is achieved which can be boosted by the use of either random flipping or random rotation of the training data during training. The combination of two data augmentation strategies generates the best performance with the mAP of 65.3% over five object categories. By increasing the probability of random flipping and rotation from 0.2 to 0.5, the overall performance degrades by a significant margin (as shown in the last row of Table III). These results provide evidence that data augmentation is beneficial to the performance when properly used.

C. On the Anchor Size and Scaling Factor

It is observed during the experiments that input CT volume scaling and anchor sizes have a significant effect on the performance of detection. We investigate how these two factors affect the detection results in this experiment. The scaling factor s is a parameter used to down-sample the input CT volumes so that the down-sampled volumes will have 1/s of the original sizes in all three dimensions. The anchor size is the other crucial factor affecting the detection performance. Proper anchor sizes should be comparable with the ground truth bounding box sizes. It is easy to understand the anchor sizes should be adaptive to values of scaling factor s for optimal detection performance. To these ends, in this experiment, we investigate different values of scaling factor s as well as the anchor sizes.

The experimental results are displayed in Table IV. The best performance is achieved when the input CT volumes are down-scaled by a factor of 3 with an appropriate combination of anchor sizes for features in different pyramid levels. When the input CT volumes have higher or lower resolutions (i.e. scaling factor of 2 or 4), performance can be degraded

	TABLE IV	
EXPERIMENTAL RESULTS	ON VOLUME SCALING AND	ANCHOR SIZE.

Scaling factor	Anchor size		mAP(%)				
Scaling factor	Allelior size	Bottle	Handgun	Binocular	GlockFrame	iPod	
2	8-16-32-64	52.0 ± 1.5	72.0 ± 4.8	64.0 ± 4.3	22.1 ± 7.0	30.2 ± 6.2	48.1 ± 4.3
3	4-8-16-32	61.0 ± 2.6	81.7 ± 0.7	77.0 ± 8.7	32.8 ± 4.5	19.3 ± 7.5	54.4 ± 4.1
3	8-16-32-64	70.0 ± 3.6	84.9 ± 0.3	$83.7~\pm~4.5$	48.9 ± 7.1	38.9 ± 4.3	65.3 ± 3.6
4	4-8-16-32	61.2 ± 0.4	74.1 ± 2.3	57.3 ± 8.2	28.4 ± 2.6	18.5 ± 7.8	47.9 ± 1.9
4	8-16-32-64	64.2 ± 3.0	81.7 ± 1.5	76.1 ± 3.8	36.3 ± 7.2	10.6 ± 4.4	53.8 ± 3.0

TABLE V Experimental Results on High and Low Energy Data

Data	Bottle				$m\Delta P(\%)$		
Data	Precision (%)	Recall (%)	AP (%)	Precision (%)	Recall (%)	AP (%)	IIIAI (<i>1</i> 0)
Low	89.7 ± 2.1	70.8 ± 2.6	68.7 ± 2.7	82.5 ± 1.5	94.5 ± 2.4	85.7 ± 3.7	77.2 ± 3.0
High	90.4 ± 1.0	68.5 ± 0.8	66.3 ± 1.3	83.6 ± 3.9	92.6 ± 1.7	85.7 ± 3.0	76.0 ± 2.0
High+Low	90.2 ± 2.1	71.0 ± 2.7	68.9 ± 2.7	82.9 ± 2.0	92.9 ± 3.1	84.0 ± 2.0	76.4 ± 2.3



Fig. 3. Detection results using the 3D RetinaNet with ResNet50 as the backbone network (the 3D bounding boxes for *bottles, handguns, binoculars, glockframes and iPods* are represented by blue, red, magenta, yellow and black colours respectively).



Fig. 4. Exemplar false positive and false negative detection results (the false detection are emphasized with red arrows).

significantly as shown in Table IV even the anchor sizes are adaptively adjusted. This observation may be caused by the fixed convolution kernel sizes and network architectures of the FPN which is worth further investigating in our future work.

D. On the Raw Data

We investigate how low- and high-energy raw data affect the object detection performance in this experiment. The CT volumes are reconstructed from raw CT slices generated by low or high energy X-ray. We also combine low and high energy data as two channels before feeding them into the 3D object detection networks. We use the optimal experimental settings derived from previous experimental findings and the results are shown in Table V. We use Bottles and Handguns as two representative prohibited items in this experiment since there are the most numbers of instances of them in the dataset (Table I). It is demonstrated the low-energy and high-energy data lead to comparable object detection performance and the combination of them does not improve the performance. As a result, either low- or high-energy data from a dual-energy machine is necessary for the purpose of automatic object detection.

E. Qualitative Evaluation

To give qualitative evaluations of the proposed approach to 3D object detection within 3D CT baggage imagery, we present exemplar detection results in Figures 3 - 4. Figure 3 list the detection results of eight typical CT volumes containing bottles, handguns or binoculars. The detected 3D bounding boxes are shown in different colours (i.e. blue, red, magenta, yellow and black for *bottles, handguns, binoculars, glock frames and iPods* respectively). The visualization in Figure 3 demonstrates that the proposed approach can detect objects in varying orientations with relatively high localization accuracy. On the other hand, the approach also suffers from false positives can be caused by misclassification of the target objects or non-target objects (i.e. background clutter having similar appearance characteristics to the target objects). False negatives can be caused by cluttered background or overcompact objects in the real baggage images. Overall, the poor detection performance of *glock frames* and *iPods* are caused by the high number of false positives and hence a low precision rate. It is also frequently observed that two predicted bounding boxes corresponding two different object categories (especially for binocular and iPods) overlap with each other with a high IOU value. This phenomenon can be caused by the artefacts introduced by 3D TIP in the synthetic CT volumes. Specifically, the artefacts rather than the real characteristics of the objects have been learned by the model for classification. This needs to be further investigated in future work with more real data available.

VI. CONCLUSION

We address the multi-class object detection problem within volumetric 3D baggage security screening CT imagery. 3D RetinaNet is employed as the detector incorporated with different FPN architectures. 3D TIP and data augmentation techniques are employed to generate a synthetic dataset to alleviate the data sparsity issue. Experimental results validate the effectiveness of the proposed approach to the detection of five object categories in baggage CT volumes and also disclose the limitations of the current study (e.g., the lack of real data).

A few research directions will be considered in our future work. Firstly, it is essential to scale up the dataset used for experiments in terms of both CT volumes and prohibited item types. Secondly, it is interesting to compare the effectiveness of 3D and 2D (slice based) CNN models in object detection within CT imagery. Finally, it is of great value to complement current approach by enabling the detection of material based prohibited items without specific shapes and appearances (e.g., explosive materials).

REFERENCES

- N. Bhowmik, Q. Wang, Y. F. A. Gaus, M. Szarek, and T. P. Breckon, "The good, the bad and the ugly: Evaluating convolutional neural networks for prohibited item detection using real and synthetically composited X-ray imagery," in *British Machine Vision Conference Workshops*, 2019.
- [2] Q. Wang, N. Megherbi, and T. P. Breckon, "A reference architecture for plausible threat image projection (TIP) within 3D X-ray computed tomography volumes," *Journal of X-Ray Science and Technology*, vol. 28, no. 3, pp. 507–526, 2020.
- [3] R. F. Meuter and P. F. Lacherez, "When and why threats go undetected: Impacts of event rate and shift length on threat detection accuracy during airport baggage screening," *Human factors*, vol. 58, no. 2, pp. 218–228, 2016.
- [4] 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.
- [5] Y. Gaus, N. Bhowmik, S. Akcay, and T. Breckon, "Evaluating the transferability and adversarial discrimination of convolutional neural networks for threat object detection and classification within x-ray security imagery," in *Proc. Int. Conf. on Machine Learning Applications*. IEEE, December 2019.
- [6] N. Bhowmik, Y. Gaus, S. Akcay, J. Barker, and T. Breckon, "On the impact of object and sub-component level segmentation strategies for supervised anomaly detection within X-ray security imagery," in *Proc. Int. Conf. on Machine Learning Applications*. IEEE, December 2019, pp. 986–991.

- [7] N. Bhowmik, Y. Gaus, and T. Breckon, "Using deep neural networks to address the evolving challenges of concealed threat detection within complex electronic items," in *Proc. Conf. on Homeland Security*. IEEE, November 2019, pp. 1–6.
- [8] Q. Wang, N. Bhowmik, and T. P. Breckon, "On the evaluation of prohibited item classification and detection in volumetric 3d computed tomography baggage security screening imagery," in *International Joint Conference on Neural Networks*, 2020, to appear.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. computer vision and pattern recognition*, 2016, pp. 770–778.
- [10] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. Int. conf. on computer vision*, 2017, pp. 2980–2988.
- [11] Y. F. A. Gaus, N. Bhowmik, S. Akçay, P. M. Guillén-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*. IEEE, 2019, pp. 1–8.
- [12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards realtime object detection with region proposal networks," in Advances in Neural Information Processing Systems, 2015, pp. 91–99.
- [13] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *IEEE International Conference on Computer Vision*. IEEE, Oct 2017, pp. 2980–2988.
- [14] D. F. Wiley, D. Ghosh, and C. Woodhouse, "Automatic segmentation of CT scans of checked baggage," in *Proc. Int. Meeting on Image Formation in X-ray CT*, 2012, pp. 310–313.
- [15] G. Flitton, T. P. Breckon, and N. Megherbi, "A 3D extension to cortex like mechanisms for 3D object class recognition," in *Proc. Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 3634–3641.
- [16] A. Mouton, T. P. Breckon, G. T. Flitton, and N. Megherbi, "3D object classification in baggage computed tomography imagery using randomised clustering forests," in *Proc. International Conference on Image Processing*, 2014, pp. 5202–5206.
- [17] P. Jin, D. H. Ye, and C. A. Bouman, "Joint metal artifact reduction and segmentation of CT images using dictionary-based image prior and continuous-relaxed potts model," in *Proceedings of International Conference on Image Processing*. IEEE, 2015, pp. 798–802.
- [18] G. Flitton, A. Mouton, and T. P. Breckon, "Object classification in 3D baggage security computed tomography imagery using visual codebooks," *Pattern Recognition*, vol. 48, no. 8, pp. 2489–2499, 2015.
- [19] A. Mouton and T. P. Breckon, "Materials-based 3D segmentation of unknown objects from dual-energy computed tomography imagery in baggage security screening," *Pattern Recognition*, vol. 48, no. 6, pp. 1961–1978, 2015.
- [20] —, "A review of automated image understanding within 3D baggage computed tomography security screening," *Journal of X-ray Science and Technology*, vol. 23, no. 5, pp. 531–555, 2015.
- [21] Q. Wang, K. N. Ismail, and T. P. Breckon, "An approach for adaptive automatic threat recognition within 3D computed tomography images for baggage security screening," *Journal of X-ray Science and Technology*, vol. 28, no. 1, pp. 35–58, 2020.
- [22] N. Megherbi, G. T. Flitton, and T. P. Breckon, "A classifier based approach for the detection of potential threats in ct based baggage screening," in *International Conference on Image Processing*. IEEE, 2010, pp. 1833–1836.
- [23] G. Flitton, T. P. Breckon, and N. Megherbi, "A comparison of 3d interest point descriptors with application to airport baggage object detection in complex ct imagery," *Pattern Recognition*, vol. 46, no. 9, pp. 2420–2436, 2013.
- [24] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 2019, pp. 8024–8035.
- [25] P. F. Jaeger, S. A. Kohl, S. Bickelhaupt, F. Isensee, T. A. Kuder, H.-P. Schlemmer, and K. H. Maier-Hein, "Retina u-net: Embarrassingly simple exploitation of segmentation supervision for medical object detection," in *Proc. Neural Information Processing Systems Workshops*, 2019.
- [26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. Int. conf. on Learning Representations, 2015.