

An Approach for Adaptive Automatic Threat Recognition Within 3D Computed Tomography Images for Baggage Security Screening

Qian Wang ^{*1}, Khalid N. Ismail^{1,3}, and Toby P. Breckon^{1,2}

¹Department of Computer Science, Durham University, United Kingdom

²Department of Engineering, Durham University, United Kingdom

³Information Technology Department, Faculty of Computers and Information, Menoufia University, Egypt

Abstract

BACKGROUND: The screening of baggage using X-ray scanners is now routine in aviation security with automatic threat detection approaches, based on 3D X-ray computed tomography (CT) images, known as Automatic Threat Recognition (ATR) within the aviation security industry. These current strategies use pre-defined threat material signatures in contrast to adaptability towards new and emerging threat signatures. To address this issue, the concept of adaptive automatic threat recognition (AATR) was proposed in previous work.

OBJECTIVE: In this paper, we present a solution to AATR based on such X-ray CT baggage scan imagery. This aims to address the issues of rapidly evolving threat signatures within the screening requirements. Ideally, the detection algorithms deployed within the security scanners should be readily adaptable to different situations with varying requirements of threat characteristics (e.g., threat material, physical properties of objects).

METHODS: We tackle this issue using a novel adaptive machine learning methodology with our solution consisting of a multi-scale 3D CT image segmentation algorithm, a multi-class support vector machine (SVM) classifier for object material recognition and a strategy to enable the adaptability of our approach. Experiments are conducted on both open and sequestered 3D CT baggage image datasets specifically collected for the AATR study.

RESULTS: Our proposed approach performs well on both recognition and adaptation. Overall our approach can achieve the probability of detection around 90% with a probability of false alarm below 20%.

CONCLUSIONS: Our AATR shows the capabilities of adapting to varying types of materials, even the unknown materials which are not available in the training data, adapting to varying required probability of detection and adapting to varying scales of the threat object.

Keywords: Adaptive automatic threat recognition, X-ray computed tomography, Image segmentation, Baggage security screening, Automatic threat detection

*Corresponding: Qian Wang, Email: qian.wang@durham.ac.uk/qian.wang173@hotmail.com

This paper is based upon work supported by the U.S. Department of Homeland security under Award Number 2013-ST-061-E001. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

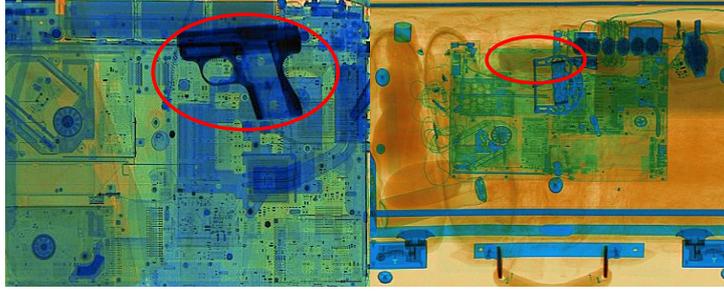


Figure 1: Threat objects in 2D X-ray images [2]. Left: threat object signatures of specific appearance (e.g., firearms) are easy to identify. Right: threat materials are more challenging to recognize in cluttered X-ray images.

1 Introduction

X-ray computed tomography (CT) scanners have been deployed at airports across the world to detect threat materials hidden in baggage. By inspecting the content in the 3D reconstructed image from the scanner, operators are able to identify potential threat objects without the need for manual search [27, 14]. Some threat objects such as firearms and knives can be recognized by their appearance, whilst threat (explosive) materials which can be made into threat devices could have varying physical appearances (e.g., different shapes and volumes, as illustrated for related work on 2D X-ray in Figure 1) [28, 12, 14]. Current approaches for material based threat detection, denoted as automatic threat recognition (ATR) or explosive detection systems (EDS) within the aviation security industry, may use dual energy computed tomography (DECT) [21] to perform this detection task based on a range of standardised threat material signatures. In addition, operators may perform a secondary manual inspection of the reconstructed 3D CT image to resolve materials highlighted by the ATR/EDS process.

Automatic threat recognition (ATR) intends to enable baggage screening more efficient and effective with only limited human intervention. Attempts have been made in recent works to address the ATR problem in 2D X-ray images [3, 1, 23, 6, 2] and 3D CT images [11, 40, 9, 22, 28, 27, 26, 7, 20]. Most existing works, however, focus on the recognition of threat objects having specific shape-based appearances (e.g., firearms, bottles, knives, etc.), which can not be directly applied to the recognition of threat materials. In this paper, we aim to detect threat materials in baggage based on 3D CT images. It is usually formulated to be a visual recognition problem [7] in the X-ray CT imagery domain. A typical approach to addressing this problem is to use machine learning techniques [27, 7]. Specifically, the potential object volumes are firstly extracted from the CT images using a 3D segmentation algorithm, then the objects are classified as threats or non-threats using a binary classifier [7].

One limitation of conventional ATR is the lack of adaptability. Once the ATR models are deployed to the scanners, they work for the detection of threat materials pre-defined during the model training. In real scenarios, however, the definition of threat materials evolves with the changing nature of the international aviation security landscape and the threat actors involved. As a result, material signatures not originally defined as a threat can emerge as new threat materials [17]. Although such materials may be detected with current ATR, specific adaptations are required to improve both sensitivity and specificity of screening. Adapting the deployed ATR algorithms to these new material signatures is a non-trivial task. First, scanner manufacturers need to adjust their algorithms to tackle the new definitions which might involve new training data collection, annotation and algorithm re-training. In addition, it may take considerable time to have the new algorithm variants certified for deployment [7].

To address this issue, adaptive automatic threat recognition (AATR) is proposed [8, 18, 31] by the US Department of Homeland Security and the ALERT (Awareness and Localization

of Explosives-Related Threats) Center of Excellence [4]. The problem AATR aims to address differs from conventional binary classification problems in that the definition of threats and non-threats could vary from situation to situation, which raises the requirement for the algorithm to be adaptive to varying definitions of threat and non-threat material signatures. Conventional automatic threat recognition aims to classify an object within baggage as a threat or a non-threat deterministically without the capability of adapting to the change of threat definitions. However, an adaptive automatic threat recognition algorithm could take the definition of threats as an additional argument and classify an object as a threat or not accordingly. As a result, the output of AATR is determined not only by features of the object but also by the definition of threats. To formulate the AATR problem, ALERT defines a series of adaptability metric (AM). Each adaptability metric consists of multiple objects requirement specifications (ORS). For example, the AM of varying materials aims to evaluate how well an AATR algorithm can adapt to varying materials. Further conceptual details of the AM and ORS defined by ALERT are described in [8] and related to our work in Section 4.5.

In this paper, we present an approach to resolving the AATR problem originally posed by [8]. In the initial stage, we propose a multi-scale 3D image segmentation algorithm to segment objects of different scales and shapes in the baggage. Subsequently, we utilize a multi-class support vector machine (SVM) [15] to classify the segmented object signatures into different classes of material. Once the materials of segmented objects are known, our AATR approach can recognize potential threat objects according to the material signatures specified by the requirements. Instead of assigning a class label to a given segmented object, we calculate its probability of belonging to each class. By adaptively manipulating these probabilities, we are able to make the classification results bias to some specific classes intentionally to satisfy varying probability of detection requirements.

Contributions of this paper are summarized as follows:

- A novel 3D segmentation algorithm is developed to separate multi-scale material objects within the CT baggage image so that each segmented object can be further classified as a threat or non-threat material signature;
- An approach to the AATR problem, as originally posed by [8], is proposed with the adaptabilities to varying specifications of threat material signatures, varying *probability of detection* (PD) requirements and new unseen threat materials for which there are no examples available during algorithm training;
- An evaluation is conducted to validate the feasibility and effectiveness of the proposed solution to the AATR problem.

2 Related Work

In recent years, 2D X-ray image based threat object recognition has been extensively studied [1, 2, 24], although even the use of multiple view X-ray suffers from the challenges of object recognition under varying orientation and inter-object occlusion. This can be addressed by 3D X-ray CT imaging which provides abundant information as a 3D volume comprised of successive, parallel X-ray image slices [14]. 3D X-ray CT imaging can be categorized into single-, dual- and multi-energy CT based on how many separate energy levels are used for scanning. For the purpose of material differentiation, it has been shown that dual- and multi-energy CT has an advantage over single-energy [11, 26]. Since the main focus of this paper lies in the adaptability of an automatic threat recognition algorithm, we particularly focus on 3D object segmentation and material classification for automatic threat recognition in 3D baggage CT images. For a broader view on the background knowledge and techniques of automatic CT image processing for baggage screening, we refer readers to [35, 38] and [27].

2.1 CT Image Segmentation

Although CT image segmentation has been extensively studied in literature [22, 16, 27], most approaches are specially designed for medical CT images of organisms and thus not directly applicable to our task in which the threat signatures could have varying shapes and materials [27]. To promote the automatic threat recognition performance in baggage screening, the first initiative regarding 3D CT image segmentation was proposed via [7, 9]. The initiative resulted in effective CT image segmentation approaches which can be used for automatic threat recognition [9]. Among the outcomes of this initiative, Grady et al. [13] proposed a 3D CT image segmentation algorithm featured by a novel Automatic Quality Measure (AQUA) model that measures the segmentation quality for the segmented object so that good segmentation results can be achieved by learning segmentation parameters to optimize this measurement based objective. The model is learned based on 92 “good quality” object segments by a data-driven machine learning approach. As a result, the effectiveness of the learned AQUA model might rely on the features (e.g., shape, geometric features) of the training data and is limited when there exist new objects of arbitrary shapes to segment. Wiley et al. [40] adapted a medical segmentation technique called Stratovan Tumbler [39] to the baggage CT image segmentation problem. Kernels of different parameter values are used to deal with the boundaries of objects having different shapes and sizes. Our segmentation algorithm employs a similar idea that multi-scale structuring elements (Section 3.1.1) are used to segment objects of different sizes recursively.

Mouton et al. [26] proposed a novel technique for the 3D segmentation of unknown objects in the baggage CT images. The proposed algorithm takes advantage of the appearance information of objects to be segmented (e.g., handguns and bottles) which, however, is not applicable to our problem since the concerned threat materials could appear in very different forms. Martin et al. [20] proposed a learning based framework for joint object segmentation and threat recognition from volumetric CT images. Their work was based on the dual-energy X-ray computed tomography which can provide more essential information to differentiate materials [21] while our work focuses on single-energy CT images.

2.2 CT Image Based Material Classification

One key element in automatic threat recognition is materials classification. CT image based material classification is used to distinguish the candidate signatures of threat materials from those of non-threat materials. Recently, a new initiative was proposed towards automatic threat materials recognition based on baggage CT images [7]. Threat materials recognition in baggage CT images is usually formulated as a material classification problem based on the segmentation results. Ye et al. [7] proposed a new 3D CT image segmentation approach followed by classifying the segmented candidate threat material signatures to be threats or non-threats using multiple SVM classifiers. These classifiers are trained on different clusters of training object signatures based on their shapes so that the threats and non-threats falling into different clusters can be recognized using different classifiers respectively. Evaluated on the particular dataset for this initiative, they achieved 95% probability of detection with a probability of false alarms less than 10%. Resulting from the same initiative, Zhang et al. [7] used a pixel classification approach for CT image segmentation which can be implemented by an expectation-maximization (EM) algorithm and enhanced by using a Markov random field (MRF) to impose some spatial smoothness-constraints on the iterative two-step EM process. As for the classification of threats and non-threats, a collective classification algorithm was employed, where, one SVM classifier was trained for each type of material. Evaluated on the same data as Ye et al. [7], the performance is slightly worse yet still competitive with 89.2% probability of detection and 9.7% probability of false alarms. SVM classifier was also employed for CT image classification by Flitton et al. [12] in which, however, the classification focuses on

specific objects rather than materials.

The AATR problem to be addressed in this paper aims to improve the adaptability of conventional ATR algorithms. On one hand, we take advantage of the framework used in many works that consist of a 3D segmentation algorithm followed by classification. On the other hand, our proposed segmentation and classification methods make the least use of supervision information (especially the shapes of objects) to promote the generalization capability. More importantly, our approach has the adaptability to varying requirements in practical applications. In parallel to this work, Paglieroni et al. [31] proposed a spatial consensus relaxation method to determine the most likely material composition for each CT image voxel, followed by threat signature classification to address the same AATR problem.

3 Method

Our AATR approach consists mainly of three components as shown by the diagram in Figure 2 (i.e. 3D segmentation, material classification and adaptation to varying requirements). These three components form the pipeline of our approach to threat recognition with the adaptability to varying requirements (e.g., objects of which properties are required to be recognized as threats). As the first step, 3D segmentation is used to separate all the object signatures in baggage based on the raw CT volumetric data. Given a 3D CT image as input, it is firstly binarised by voxel thresholding and we define these voxels as background. Within this context, the non-zero voxels are now defined as object volumes. The segmentation algorithm attempts to find out the 3D volume of each individual object signature which is characterized by a set of connected voxels of the same material intensity (i.e. voxel value within a CT image) range. Voxels of each segment are labelled with a unique integer label, while the background voxels are labelled with zeros. Once segmented objects have been obtained, we need to decide if the segment object is a threat or non-threat according to the definition of threat objects. Since the threat objects are defined in terms of their materials and physical properties such as the density (i.e. defined as the mean intensity of voxels of an object signature), a classification step is employed to recognize the material types of the segmented objects (i.e. *saline*, *rubber*, *clay* and *others*). As a result, any segmented object in the image can be classified as a threat or non-threat according to its material type derived from the classification result and density which can be calculated. Due to the inaccuracy of segmentation and classification results, the detection performance is not perfect with the *probability of detection* (PD) less than 100% along with a non-zero *probability of false alarm* (PFA). To satisfy the particular requirements in varying practical scenarios, we employ adaptation techniques to trade off the probability of detection and false alarm. For instance, we take measures to have more segmented objects classified as threats, which will inevitably mis-classify some non-threats as threats (increasing the probability of false alarm) though more true threats are expected to be classified correctly (increasing probability of detection) at the same time. As a result, the probability of detection of a specific material can be adaptively increased or decreased to satisfy the requirements. In our approach, the adaptation can be implemented by manipulating the output probabilities of different classes (i.e. material types) from a support vector machine (SVM) classifier.

3.1 3D Segmentation

Image segmentation is an important pre-process in many computer vision tasks. By segmenting objects in images, we are able to focus on the visual information of the individual object without interference from the background and other objects. Image segmentation has been studied extensively in recent decades, achieving good performance by learning from a large amount of training data, using the heterogeneous features of the objects such as color, shape, texture and semantics. However, the techniques commonly used for image segmentation are not readily

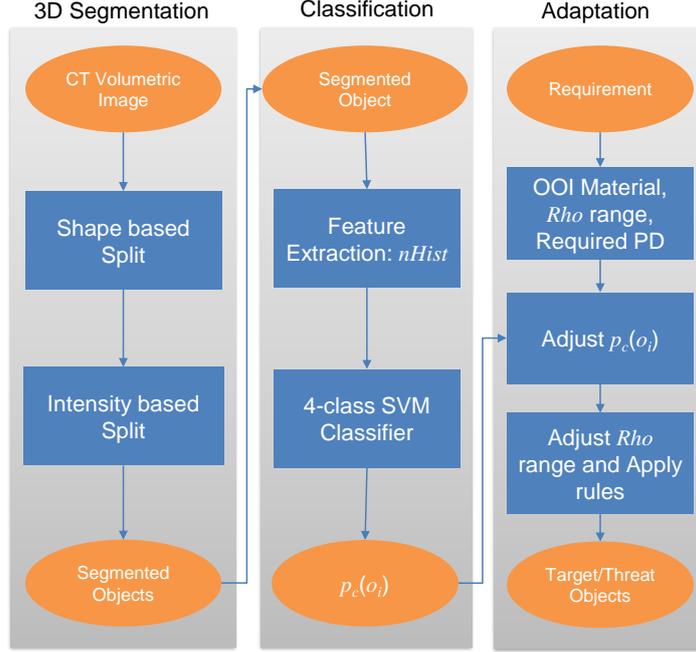


Figure 2: The framework of our proposed adaptive automatic threat recognition (AATR) approach.

applicable to the 3D CT images in our problem. First, the CT images are 3D volumetric data while most state-of-the-art image segmentation algorithms are designed for 2D images. Second, the objects in the CT images are characterized only by the voxel intensities, lacking of the key information such as color and texture in the natural images. Finally, the objects concerned in our problem could have arbitrary shapes without the prior semantic knowledge. Therefore we propose a novel 3D segmentation algorithm to segment objects of varying materials in 3D CT images. Our segmentation algorithm consists of two steps: shape based split and intensity based split. In the shape based split, the algorithm takes advantage of the morphological operations and the connected-component labelling (CCL) [33] method to split the objects which are not or slightly touching one another. Subsequently, the intensity based split is employed to handle the merged objects on which the first step has failed but the merged objects are of different materials.

3.1.1 Shape Based Split

The basic idea of the shape based split is to use the classical connected-component labelling (CCL) method to label objects which are not connected. The objects not connected to any other objects can be directly segmented from others. Objects in baggage, however, are usually compactly packed, touching each other thus posing a big challenge to the original CCL method. To address this issue, we apply morphological opening operations to split touching objects. We first apply erosion to remove the connection voxels among multiple objects, then the CCL method can be applied to label the individual objects. Finally, the dilation operation is applied to the segmented individual objects to compensate the voxel loss caused by erosion operations by using the same structuring element as used for erosion.

One key factor in the morphological opening operation is the selection of the structuring element (or the kernel). The parameterisation of the structuring element (e.g., the radius of a sphere) used for erosion determine what kind of connections between two objects can be broken down. A small-scale structuring element can only split objects connected by a small number of voxels, while a large-scale structuring element has the risk of completely removing small

objects. To solve this problem, we use a novel multi-scale morphological opening approach, using multiple structuring elements of different parameter values recursively. Our multi-scale morphological opening approach is shown in Figure 3. It can split objects connected by a large number of voxels without mistakenly removing small objects.

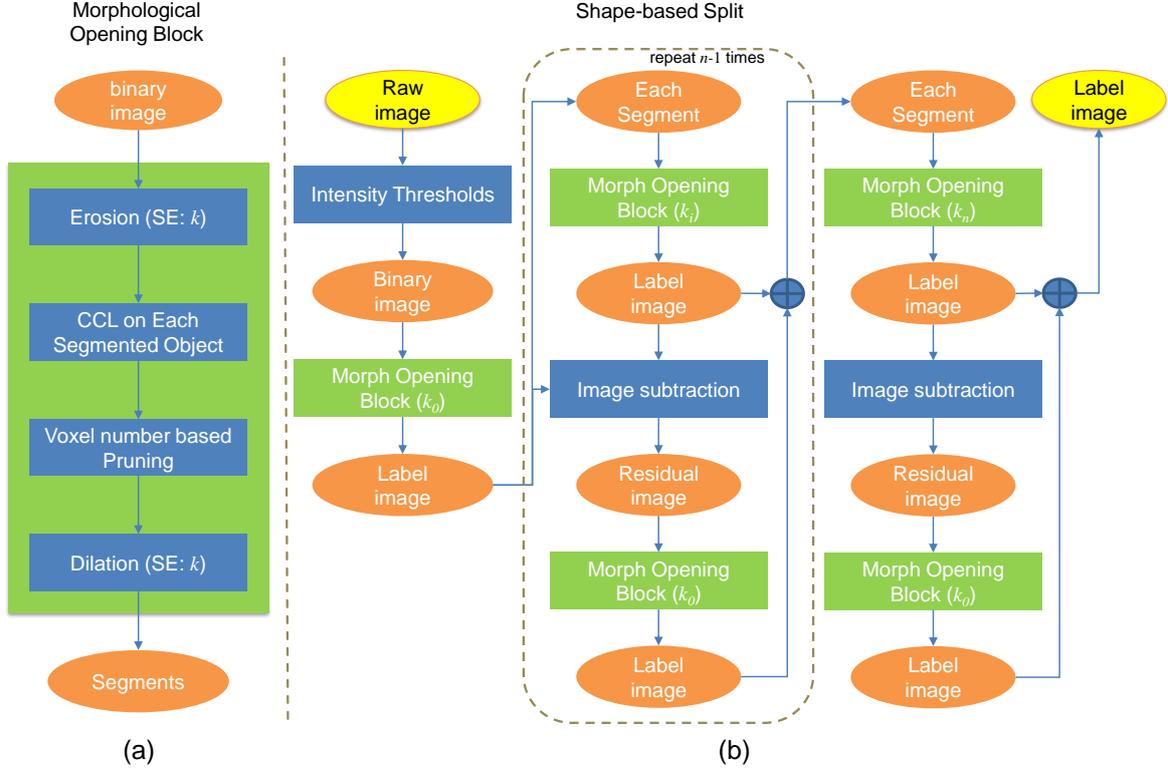


Figure 3: The proposed shape-based split using multi-scale morphological opening operations. (a) The morphological opening operation with the structuring element parameter k . (b) Multi-scale morphological opening work-flow.

Figure 3(a) shows the basic morphological opening operation used in the multi-scale morphological opening approach. The workflow of the multi-scale morphological opening is illustrated by the diagram in Figure 3(b). To facilitate our discussion, we describe some notations and terms used in the following content. A *raw image* is defined as a three-dimensional CT image having continuous voxel values. A *binary image* is defined as a three-dimensional volumetric image with binary voxel values (i.e. 0 indicates the background voxels and 1 indicates the foreground voxels). A *label image* I^L is defined as a three-dimensional volumetric image with voxel values $v \in \{0, 1, \dots, n\}$, where n is the number of labelled/segmented objects, 0 indicates the background voxels and the values of i indicates the voxels belonging to the i -th segmented object. Therefore a binary image is actually a label image with only one labelled object in this sense.

The morphological opening block takes a label image $I^L, v \in \{0, 1, 2, \dots, n\}$ as input and output another label image $I^{L'}, v' \in \{0, 1, 2, \dots, n, n+1, \dots, n+m\}$, where $m \geq 0$. The morphological opening operation is applied to each object specified in the input label image. Different from the classical morphological opening with only erosion and dilation operations, we add the voxel number based pruning process to remove the segmented objects having too few voxels and use the CCL method to identify segmented objects in the image. For any object $o_i = \{v | v = i\}$ in the original label image, if it can be further split into more than one objects after the morphological opening, the voxels of these segmented objects will be relabelled using different values respectively, otherwise the object will keep unchanged in the output label image. Particularly we

keep the object unchanged if there is no segmented object left after the morphological opening operation.

Now we describe the multi-scale morphological opening approach for the shape-based split in detail. Given a raw image I and its voxel value $v \in \mathbb{R}$, our goal is to obtain a label image $I^L, v \in \{0, 1, 2, \dots, n\}$, specifying n segmented objects, and the segmented object o_i is specified by a set of connected voxels whose values are equal to i . First, the raw image I is transformed to a binary image by voxel-wisely comparing its voxel values with a lower threshold t_{lower} and an upper threshold t_{upper} which are determined by the prior knowledge, so that the objects of unconcerned materials can be directly removed. The binary image, which is a special case of the label image defined above, is then fed into the morphological opening block where the parameter of the structuring element used for erosion is set as k_1 . The resultant label image I_1^L contains n_1 segmented objects. Second, we apply the morphological opening block again on the label image $I_1^L, v \in \{0, 1, 2, \dots, n_1\}$ with a greater structuring element parameter $k_2 > k_1$. As a result, a label image I_2^L containing $n_2 \geq n_1$ segmented objects can be obtained. However, there exists a risk of mistakenly removing small/thin objects by applying the morphological opening with a relatively large k_2 . For example, assuming there is an object o_i labelled in the label image I_1^L which is merged by two objects, and one of them happens to be small or thin such that it is completely removed after being split. To avoid this erroneous occurrence, we reinspect the voxels removed from the label image I_1^L to recover the mistakenly removed objects if there are any. Specifically, we apply an *image subtraction* operation to get the *residual image* I^R , a binary image characterizing the background difference between label images I_1^L and I_2^L ,

$$I^R = I_1^L - I_2^L, v^R \in \{0, 1\} \quad (1)$$

where any voxel value v^R in the residual image I^R is computed by:

$$v^R = \begin{cases} 0 & \text{if } v_1^L > 0, v_2^L > 0 \text{ or } v_1^L = v_2^L = 0, \\ 1 & \text{if } v_1^L > 0, v_2^L = 0. \end{cases} \quad (2)$$

The residual image I^R is re-inspected by applying the morphological opening operation with the same structuring element parameter k_1 as that used in the first step. The value of k_1 is selected so that none of the threat material signatures can be mistakenly removed by the erosion operation with such a small structural element. The resultant label image is denoted as $I_2^{LR}, v \in \{0, 1, \dots, n_R\}, n_R \geq 0$. We update the label image I_2^L by adding the residual label image I_2^{LR} :

$$I_2^{L'} = I_2^L + I_2^{LR}, v_2^{L'} \in \{0, 1, 2, \dots, n_2 + n_R\}, \quad (3)$$

where any voxel value $v_2^{L'}$ in the label image $I_2^{L'}$ is computed by:

$$v_2^{L'} = \begin{cases} v_2^L & \text{if } v^{LR} = 0, \\ n_2 + v_2^{LR} & \text{otherwise.} \end{cases} \quad (4)$$

Once the label image $I_2^{L'}$ is updated, it can be fed into another morphological opening block with a greater structuring element parameter $k_3 > k_2$ and repeat the steps again to further split the objects specified in the label image $I_2^{L'}$ as shown in Figure 3(b).

Technically, the multi-scale morphological opening operations can be repeated many times with gradually increased structuring element parameters k . The depth of such recursion is subjective to the scales of objects-of-interest in the CT volumes and the constraint of computation time. On one hand, more rounds of recursion are needed if the scales of the objects-of-interest vary significantly for good segmentation performance. On the other hand, more rounds of recursion will require more computation. As a result, the depth of recursion can be decided by trading-off required segmentation performance against computation. In our experiments, we recursively apply the morphological operations for three rounds and empirically set $k_1 = 2$, $k_2 = 3$ and $k_3 = 8$.

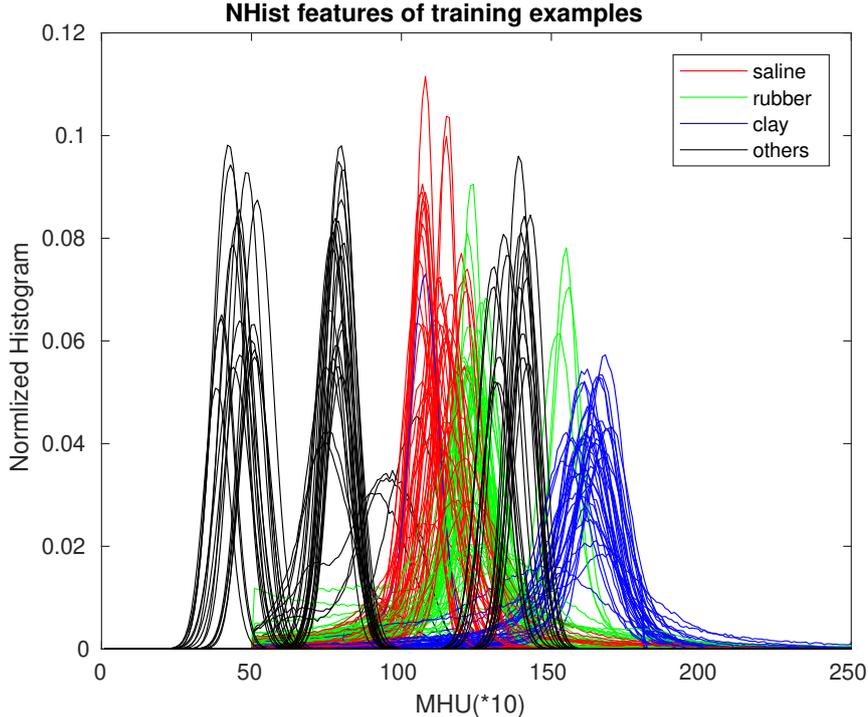


Figure 4: The normalised histogram features of different materials.

3.1.2 Intensity Based Split

The shape based split method can fail when the merged objects are compactly touching each other. The intensity based split method provides a complementary way to address this issue when the merged objects happen to have distinguishable intensity distributions. Objects of different materials have different intensity distributions as shown in Figure 4. When the merged objects are of different materials, the voxel intensity will be distributed in multiple ranges corresponding to the intensity range of each material. Based on this idea, the intensity based split aims to discover the potential materials in a merged object and assign respective labels to voxels of the merged object.

A variety of enabling techniques could be used for our purpose of intensity based split. For example, a probabilistic clustering algorithm using Expectation-Maximization (EM) can be employed to model the intensity distributions of different materials which are then used to label each voxel of a given volume as one of the materials [34]. Due to the variations of intensity distributions for the same material (c.f., Fig. 4), the estimated models using EM might not generalize well to some specific new volumes. Instead of modelling the intensity distributions for all materials, we employ a peak-detection method that explores the volume-specific intensity distribution information and does the segmentation.

Specifically, given a segmented object o_i in the label image I^L generated by the shape based split, we extract the intensity values of all the voxels belonging to this object from the raw image I . The histogram of the intensity can be computed and gives an intensity distribution curve as shown in Figure 5. Subsequently, we use a peak detection algorithm to search the number of objects residing in the segmented object o_i and the intensity range for each of them. The peak detection algorithm uses a sliding window to locate a peak whenever the value in the middle of the sliding window is the maximum. Once the peaks are detected, we search the lowest value between two peaks as the intensity boundary of the two objects. If only one peak is detected, it means there is only one type of material in this segmented object, and we leave it unchanged.

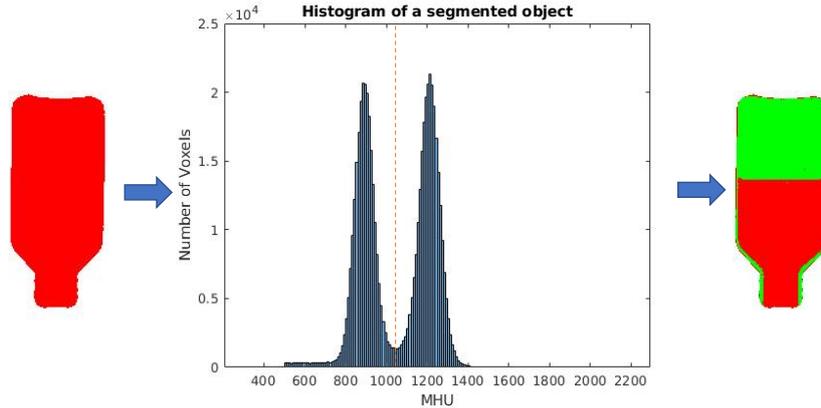


Figure 5: An illustration of intensity based split.

Otherwise, we assign the voxels whose intensity values are in the same intensity range to a new object label. As a result, the merged object o_i can be split into multiple objects of different materials. The whole procedure of the intensity based split is illustrated in Figure 5.

3.2 Material Classification

The 3D segmentation algorithm creates a label image I^L for each baggage specifying a collection of object signatures, among which there are both material signatures of interest (threats) and others (non-threats). The existence of non-threats in the label image will incur false alarms. To get rid of the false alarms, we need to distinguish the threat signatures from the non-threat items. In this section, we first describe how the classification is employed in the traditional ATR and then extend it to the AATR problem.

In the conventional automatic threat recognition problem, the threat materials are pre-defined and fixed, we only need to annotate the signatures as threats or non-threats by a binary classifier. The binary classifier can be trained using training data in which all the segmented objects are annotated as threats or non-threats according to the definition. For example, the materials of *saline*, *rubber* and *clay* are defined as threat materials in our experiments, we annotate all the signatures of these three materials in the training data as threats and other objects as non-threats. To train a classifier for this purpose, we need to extract useful features from the segmented objects. Based on the definition of threats (i.e. materials of *saline*, *rubber* and *clay*), we use the normalised histogram of intensities as the feature for classification since it is able to characterize the material type of an object.

In the adaptive automatic threat recognition problem, it is required to recognize the material type of objects so that the detection decision can be made adaptively to varying requirements of threat material recognition. To this end, we train a four-class classifier, $\{“saline”, “rubber”, “clay”, “others”\}$, where the class “others” denotes a collection of unknown materials other than the three known materials. Again, we use the normalised histogram of intensities as features for classification. We use support vector machine (SVM) with a linear kernel as the classifier in our experiments, though other types of classifiers such as random forest [5] and neural networks are also applicable.

The SVM classifier is trained using features of ground truth object signatures in the training data for known material classes (i.e. *saline*, *rubber* and *clay*). For the “others” class, we synthesize a collection of features as training data belonging to this class. Specifically, a Gaussian function $g(x; a, \mu, \sigma)$ is used to synthesize a feature vector \mathbf{v} :

$$\mathbf{v} = g(x; a, \mu, \sigma) = a \cdot e^{-\frac{(x-\mu)^2}{\sigma^2}}, \quad (5)$$

where μ is randomly selected only if it is out of the density ranges of three known materials, σ is a random number chosen from an empirical range derived from the intensity distribution of the ground truth material signatures in the training data (see Figure 4).

3.3 Adaptation to Varying Requirements

The adaptation step aims to enable our approach to adapt itself to varying requirements in practical scenarios. Specifically, we focus on the adaptability metrics defined by [8] in terms of: (a) varying threat materials; (b) varying probability of detection (PD) requirements.

Given a segmented object, the classification result allows us to know which class it belongs to. With this information, it is straightforward to classify the segmented objects whose material types are not included in the requirement specification as non-threats. To enable the adaptability, we also need to know the probability of belonging to each class for a given segmented object. The probability estimation method for SVM proposed in [41] is used in our experiments. We trade off the probability of detection and the probability of false alarm by manipulating the classification results thus adapting our approach to satisfying the requirements of PD/PFA. Specifically, we adjust the probability of classes $p_c(o_i)$ for a given segmented object o_i as follows:

$$p'_c(o_i) = p_c(o_i) + f_c(t_{PD}), c \in \{saline, rubber, clay, others\} \quad (6)$$

where $f_c(t_{PD})$ is an *offset* used to adjust the output probability of the classifier for class c , and t_{PD} is the required probability of detection for a specific material class. Intuitively, when the offset for class c is positive, the objects will have higher probabilities to be classified as class c . On the other hand, if the offset is negative, chances are less to classify the objects as this class.

The key issue is how to get the offset functions $f_c(t_{PD})$ for different classes. We repeat experiments using training data by setting the offset to be varying values and calculate the corresponding probability of detection. With pairs of probability of detection and offset values, we can fit the data using smoothing splines [10] to obtain the offset functions $f_c(t_{PD})$. Subsequently, these functions can be used to compute the offset values given the required probability of detection t_{PD} for a specific class.

Other than the manipulation of output probabilities by Eq. (6), we can also trade off the probability of detection and the probability of false alarm by adjusting the specified density range. In real scenarios, threat material signatures are defined with additional information of the density range, i.e., a threat signature of a specific material should have a density within the pre-defined density range in the requirement specification. The density, Rho , is the mean intensity over all voxels of an object. Due to the inaccuracy of segmentation and noises in the raw CT data, the computed densities of segmented objects could have biases. With this regard, we adjust the density range by a factor α in the following way:

$$Rho'_{min} = Rho_{min} * \alpha, Rho'_{max} = Rho_{max} / \alpha, \quad (7)$$

where a larger α leads to a smaller density range and thus a lower probability of detection and a lower probability of false alarm. Since the parameter α in Eq. (7) plays a similar role as the *offset* in Eq. (6), we will find and fix the optimal values of α and adjust the *offset* values in our experiments to enable the adaptability to varying probability of detection requirements.

4 Experiments and Results

To validate the effectiveness of our proposed approach, we conduct a series of experiments on real baggage CT image data. In this section, we first describe the CT image data and evaluation metrics used in our experiments. Subsequently, we describe the experiments designed to evaluate different components (i.e. segmentation, classification and adaptation) of the proposed approach. Experimental results are presented in each subsection following the description of experiment settings.

4.1 Experimental Data

We use data provided by ALERT [8] in all the experiments. The baggage are scanned with a medical CT scanner Imatron C-300. The image size is 512×512 corresponding to the $475 \text{ mm} \times 475 \text{ mm}$ field of view. The slice spacing is 1.5 mm , leading the voxel volume to be $(475/512)^2 * 1.5 = 1.291 \text{ mm}^3$. The range of the voxel values is $0 \rightarrow 32,767$ MHU, where the value of air is 0, and the value of water is 1024. A public dataset (ALERT TO4 [7]) and a sequestered dataset (ALERT TO7 [8], unseen) collected by ALERT are employed in our experiments. The public dataset consists of 188 baggage CT images in which there are 446 object signatures of three materials (i.e. *saline*, *rubber* and *clay*). These object signatures could be threats or non-threats according to the definitions of threat in different situations. For example, in one of our experiments, *saline* is the only material defined as threat, as a result, all the *rubber* and *clay* signatures will be non-threats. Apart from the material, other physical properties of the objects such as *mass* and *density* are also used to define the threat. The definition of threat will be clarified in each experiment. The ground truth voxels are labelled by ALERT for all the objects which could potentially be threats. These ground truth annotations of object signatures in CT images can be used for the evaluation of our segmentation and classification algorithms described in Section 3. The sequestered dataset is used for testing the adaptability of AATR approaches. The CT images in the sequestered dataset consist of not only three known materials in the public dataset, but also some new materials not seen in the training stage. Since the sequestered dataset is not openly available to the authors, most of the experiments presented in this paper are conducted based on the public dataset, though some experimental results on the sequestered data are also provided with testing on this unseen dataset by the ALERT team, independent from the authors. We divide the 188 baggage CT images in the public dataset into two subsets (i.e. the odd set and the even set) by the baggage scanning serial number. This leads to 94 images in the odd set and 94 images in the even set. We use one subset for training and test our approach on the other subset. Our use of both a public and sequestered dataset in this manner directly mimics the unknown nature of the AATR task under deployment conditions against an unknown threat adversary.

4.2 Evaluation Metrics

We employ the evaluation metrics defined by ALERT [8] in our experiments. Segmentation performance is evaluated by measuring how many ground truth object signatures in the test data can be matched with the segmented objects. Here the concept of *match* is defined by the precision P and the recall R . Suppose we use G and S to denote the voxel sets of a ground truth signature and a segmented object respectively, the precision and recall can be obtained by:

$$P = \frac{\text{volume}(G \cap S)}{\text{volume}(S)}, R = \frac{\text{volume}(G \cap S)}{\text{volume}(G)}. \quad (8)$$

A ground truth object signature G is thought of as correctly segmented if any segmented object S achieves $P \geq 0.5$ and $R \geq 0.5$ for bulks or $P \geq 0.2$ and $R \geq 0.2$ for sheets.

Once the matching results are obtained for all segmented objects, probability of detection (PD) and the probability of false alarms (PFA) are employed to evaluate the performance. Probability of detection is the ratio of the number of detected signatures to the number of ground truth signatures in all the CT images, whilst probability of false alarm is the ratio of the number of falsely detected non-threat signatures to the total number of non-threats in the CT images:

$$PD = \frac{N_{\text{detections}}}{N_{\text{threats}}}, PFA = \frac{N_{\text{false alarms}}}{N_{\text{non-threats}}} \quad (9)$$

Table 1: Segmentation performance of our algorithm and baselines.

Method	PD						PFA
	saline	rubber	clay	bulk	sheet	overall	overall
baseline 1 ($k=2$)	0.55	0.69	0.55	0.55	0.71	0.60	0.31
baseline 2 ($k=3$)	0.60	0.64	0.51	0.59	0.58	0.62	0.30
baseline 3 ($k=8$)	0.58	0.13	0.12	0.38	0.10	0.29	0.21
baseline 4 (shape-split)	0.80	0.92	0.73	0.78	0.91	0.83	0.49
ours	0.90	0.96	0.97	0.93	0.95	0.94	0.56

4.3 Evaluation of 3D CT Segmentation

To thoroughly evaluate the performance of our proposed segmentation algorithm, we conduct an ablation study to investigate the contribution of each component (e.g., shape based split, intensity based split) in the segmentation algorithm. In addition, we also investigate how different structuring element parameter values affect the segmentation performance. Specifically, we conduct a comparative experiment by comparing different baseline methods with our full model. The baseline methods used in our experiment are as follows:

- **Baseline 1** uses only the morphological opening ($k = 2$) with CCL.
- **Baseline 2** uses only the morphological opening ($k = 3$) with CCL.
- **Baseline 3** uses only the morphological opening ($k = 8$) with CCL.
- **Baseline 4** uses only the shape based split.
- **Our approach** uses shape based split followed by intensity based split.

The segmentation performance is evaluated by matching the segmented object signatures with the ground truth. In this experiment, we define all of 446 *saline*, *rubber* and *clay* objects as threats regardless of their physical properties. A good segmentation algorithm is expected to successfully segment all these objects from the 3D CT images, which will lead to a high probability of detection. Inevitably, the segmentation algorithm will also segment objects other than the 446 objects of interest, which will contribute to the probability of false alarm. However, in this stage, we focus more on the probability of detection as an evaluation of segmentation performance and leave the false alarm issue to the next step of classification. The results of our approach and different baseline methods are shown in Table 1.

The segmentation performance shown in Table 1 clearly indicates the effectiveness of our proposed 3D segmentation algorithm. The best overall probability of detection (0.94) is achieved by our full method. Comparing the performance of our full method with that of different baseline methods, we know that each component in our method plays an important role in the segmentation. Baseline methods with only one uniform structuring element parameter value perform poorly as shown by baseline 1-3 with overall probability of detection as 0.60, 0.62 and 0.29 respectively. Baseline 1 and 2 using a small structuring element ($k = 2, 3$) fail to split objects touching each other, while the baseline 3 with a big structuring element ($k = 8$) mistakenly removes small threat objects. Finally, a comparison between baseline 4 (0.83 probability of detection) using only shape-based split and our approach (0.94 probability of detection) using both shape and intensity based splits provides evidence that the intensity based split is complementary to the shape based split and able to further improve the segmentation performance.

To give an intuitive insight of how the segmentation works for different cases, we present some exemplars in Figure 6. The ground truth object signatures, segmented results with some baseline and segmented results with our approach are shown in the left, middle and right side of each sub-figure respectively. Each segmented object signature is indicated by one color for better visualization. Figure 6 (a) presents an example of three merged objects touching each other, which fails the shape based split when a small structuring element ($k = 3$) is used, but can be successfully segmented when using a bigger structuring element ($k = 8$). This lends

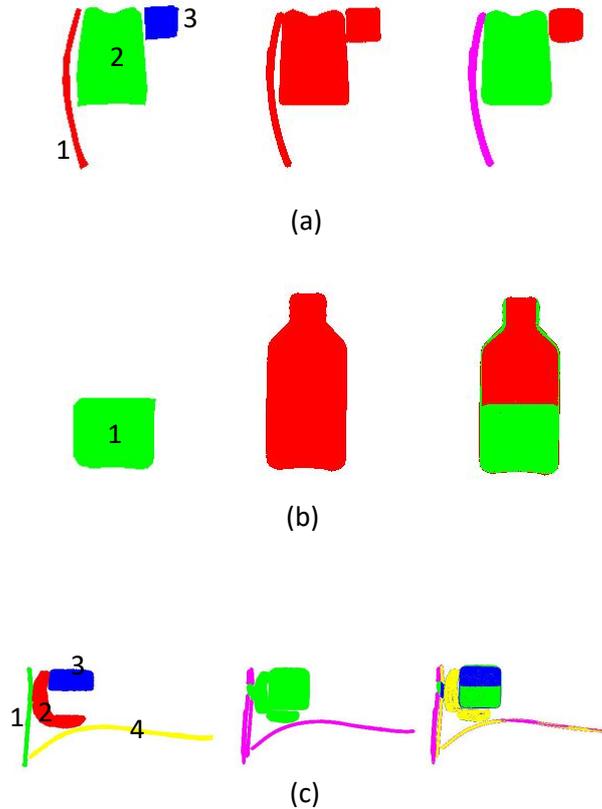


Figure 6: Exemplar segmentation results. Each sub-figure represents the segmented object signatures in one slice of the 3D CT image. From left to right are the ground truth object signatures (only *saline*, *rubber* and *clay* materials are annotated), segmented results with one of the baselines and segmented results with our approach respectively. Different colours are used to specify the separated object signatures. (a) The three merged objects cannot be split using a small structuring element $k = 3$ (middle) but segmented well when using our multi-scale segmentation algorithm (right). (b) A half bottle of *saline* can only be segmented by the intensity-based split (right). (c) The intensity-based split could cause inaccuracy when the merged objects have similar voxel intensities (objects 2 and 4 are not well separated even by our approach).

us evidence that the multi-scale segmentation is effective and necessary. Figure 6 (b) and (c) illustrate two cases where the intensity based split is helpful. Figure 6 (b) presents an example of *saline* in a bottle, for which shape base split successfully segments the bottle whilst the intensity based split can further identify the *saline* signature. Figure 6 (c) shows the effectiveness as well as the limitation of our intensity based split when merged objects have similar intensities. One can see that the figure on the right side exemplifies the inaccurate segmentation of the object colorized as *yellow* and *pink*.

Overall, our proposed algorithm is able to segment objects of varying scales and shapes. Using only morphological operations and intensity-based split, our segmentation performance is comparable with other more complex approaches in [7, 18, 31]. In addition, our segmentation method is not based on training hence less likely to suffer from overfitting issues than other learning based methods. It is also noteworthy that a high probability of false alarm is expected when obtaining a high probability of detection. Therefore a further classification is necessary

Table 2: A comparison of different classifiers for material classification.

Method	k NN ($k=3$)	k NN($k=10$)	SVM(linear)	SVM(RBF)	NN
Accuracy	92.9 ± 1.9	90.3 ± 1.6	91.2 ± 1.4	86.7 ± 2.2	93.5 ± 1.1

to reduce the false alarms resulted in the segmentation process.

4.4 Evaluation of Material Classification

We first investigate the performance of different classifiers in our material classification task. Considering the algorithm complexity and the capability of calculating conditional probability, we choose k Nearest Neighbour (k NN), Neural Networks (NN) and kernel SVM in the comparative experiment. We use a radial basis function (RBF) in the kernel SVM method. The experiment is conducted using both true and synthetic object features, as a result, a multi-class classification problem with four classes (i.e. *saline*, *rubber*, *clay* and *others*) is formulated. We use cross validation and report the mean classification accuracy and standard deviation in Table 2. The Neural Network model performs the best, however, the class conditional probability output by this classifier cannot be adjusted to fulfill our purpose of Eq.(6) since the calculated class conditional probability is usually very close to either 0 or 1. As for k NN model, it performs the second best when $k = 3$ which, however, makes it difficult to calculate the class conditional probability required by Eq.(6). As a trade-off, we use the linear SVM to classify materials in our approach.

To evaluate the classification performance for conventional ATR and our adaptive ATR (AATR) respectively, we design two more experiments. The first experiment aims to evaluate conventional ATR, in which we assume the threat materials are pre-defined (i.e. all of the 446 *saline*, *rubber* and *clay* objects are threats) and the training data for all types of materials are available. A binary SVM classifier is trained and used to classify the segmented objects to be either threats or non-threats. In the second experiment, we use a multi-class classifier to recognize different types of materials (i.e. *saline*, *rubber*, *clay* and *others*) where all the unknown materials are collectively treated as one class “*others*”. To compare with the binary classification, the same definition of threats has been used in this experiment, i.e., all of the 446 *saline*, *rubber* and *clay* objects are treated as threat to detect. It is noteworthy that the multi-class classification results can be directly applied to varying tasks of different threat definitions (e.g., only one of the three materials is defined as threat), while the binary classification cannot unless the classifier is re-trained hence limiting its adaptability.

As described in Section 4.1, we divide the public dataset into two subsets and use one of them for training and the other for testing. During training, we use ground truth voxel annotations to extract the volumes of objects in the training data, and extract the normalised histogram features from the object volumes for the classifier training. Again, we use probability of detection and probability of false alarm as the evaluation metrics for classification performance. Good classification will lead to a reduced probability of false alarm without affecting the probability of detection too much when comparing to those before applying the classification.

The experimental results are shown in Table 3 including the threat detection performance without classification, with the binary classification and with the multi-class classification. The use of binary classification significantly reduces the overall probability of false alarm from 0.56 down to 0.29, and keeps high probabilities of detection at the same time. When the multi-class classification is used, the probabilities of false alarm can be further reduced to 0.24 whilst the probabilities of detection only decrease slightly. These results validate the effectiveness of our multi-class classifier which is essential for the adaptability of ATR.

Table 3: Material classification performance on the public data.

Method	Test Set	PD						PFA
		saline	rubber	clay	bulk	sheet	overall	overall
w/o classification	all	0.90	0.96	0.97	0.93	0.95	0.94	0.56
binary classification	odd	0.90	0.99	0.98	0.94	0.99	0.96	0.28
	even	0.90	0.93	0.93	0.92	0.91	0.92	0.31
	all	0.90	0.96	0.96	0.93	0.95	0.94	0.29
multi-class classification	odd	0.86	0.98	0.98	0.92	0.97	0.94	0.22
	even	0.87	0.92	0.93	0.91	0.89	0.91	0.26
	all	0.87	0.95	0.96	0.92	0.93	0.92	0.24

Table 4: Results of the adaptation to varying probability of detection requirements.

Threat Definition	Required PD	Test Set	PD	PFA
saline (minMass:137, Rho:1050-1215)	0.70	odd	0.72	0.08
		even	0.72	0.08
		all	0.72	0.08
	0.80	odd	0.76	0.10
		even	0.80	0.11
		all	0.78	0.10
	0.85	odd	0.81	0.12
		even	0.86	0.13
		all	0.83	0.12
	0.90	odd	0.83	0.16
		even	0.87	0.19
		all	0.85	0.17
	0.95	odd	0.85	0.17
		even	0.87	0.21
		all	0.86	0.19

4.5 Evaluation of Adaptation

We conduct extensive experiments to evaluate the adaptability of the proposed approach in terms of adaptability metrics defined by ALERT [8] to simulate varying practical scenarios. All experiments in this section are carried out by employing the proposed segmentation algorithm in Section 3.1 followed by the multi-class classification. Again, the performance is evaluated in terms of probability of detection and probability of false alarm. Apart from the experiment on unknown materials which is trained on the public dataset and tested on sequestered dataset, all other experiments in the following sub-sections follow the same training/testing scheme as before, i.e., training on one subset (odd/even set) and test on the other (even/odd set) of the public dataset.

4.5.1 Varying Probability of Detection Requirements

An ideal detection algorithm is expected to have 100% probability of detection and 0% probability of false alarm. Improving the probability of detection usually causes an increase of probability of false alarm, and reducing the probability of false alarm leads to a lower probability of detection.

Firstly, we try to find optimal values of α in Eq. (7) for each known material. To this end, the values of α are selected from a candidate set of $\{1.0, 0.9, 0.8, 0.7\}$ exhaustively. Starting from $\alpha = 1$, by decreasing the value of α , we may achieve a higher probability of detection and a higher probability of false alarm. The optimal value of α is selected when the probability of detection comes to its maximum and an even smaller α will only lead to the increase of false alarms but no further improvement of detection. In our experiments, we find the optimal α to be 0.9 for *saline* and 0.8 for *rubber* and *clay*.

Our AATR approach enables the trade-off of probability of detection and probability of

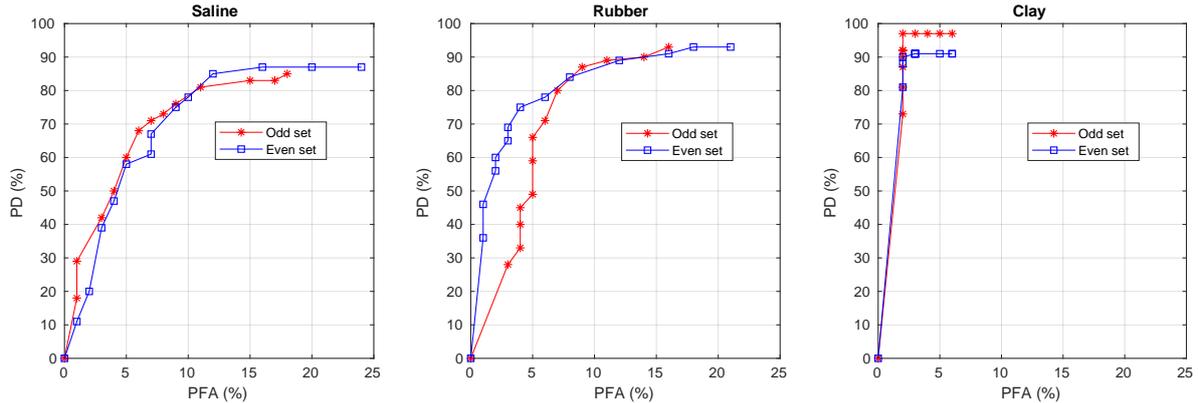


Figure 7: Probability of detection and probability of false alarm achieved by our approach for three different materials.

false alarm by manipulating the output probabilities of different threats and non-threats (see Section 3.3). To learn the relation (i.e. $f_c(t_{PD})$ in Eq.(6)) between probability of detection and the offset value for each material from training data, we set the offset value in Eq.(6) as $\{-0.5, -0.4, \dots, 0.4, 0.5\}$ respectively to get the corresponding probabilities of detection, assuming only one type of material is the threat in each experiment. Based on the pairs of probability of detection and offset values, we can have the offset as a function of the target probability of detection, i.e., $f_c(t_{PD})$ for each material class c . In the testing phase, the functions are used to calculate the offset values for the given required probabilities of detection to enable our AATR adaptive to the requirements. We use odd/even set to learn the functions and test them on the even/odd set.

Figure 7 shows the relations between probability of detection and probability of false alarm for each type of material. Our algorithm achieves the highest probability of detection of 85% and 87% on the odd and even set with the probability of false alarm of 18% and 16% respectively, when *saline* is the only threat material. The highest probability of detection our algorithm can achieve for *rubber* is 93% for both odd and even set with the probability of false alarm of 16% and 18% respectively. For the *clay* material, the highest probability of detection of 97% and 91% can be achieved with a very low probability of false alarm of 2% on the odd and even set respectively. For both *saline* and *rubber*, we can see from Figure 7 that the probability of detection and probability of false alarm gradually change with different offset values in Eq. (6). The behaviour of *clay* is quite different in that the change of offset in Eq.(6) does not affect the probability of detection and probability of false alarm very much. The reason is *clay* objects are separated from others with a large margin, thus it is easier for the classifier to make the decision with high confidence, i.e., the probability is close to one for a positive object and zero for a negative one. Therefore changing the probabilities slightly does not affect the classification results very much.

Figure 8 shows the relations between offset values and the probability of detection for different materials. In each sub-graph, we have pairs of probability of detection and offset values obtained from the experiments on training data (i.e. odd set in red and even set in blue), based on which we can infer the offset functions of target probability of detection for different materials. In our experiments, we use the smoothing spline approach to fit the data points and get the offset functions. The offset function learned from the odd set will be used for evaluation experiments on the even set, and vice versa.

To evaluate the adaptability of our approach to varying probability of detection, we set the target probability of detection for *saline* as $\{0.7, 0.8, 0.85, 0.9, 0.95\}$ respectively, and estimate the offset value for output probability manipulation in Eq.(6) using the learned offset functions. The experimental results are shown in Table 4. We can see our approach achieves 0.72 and

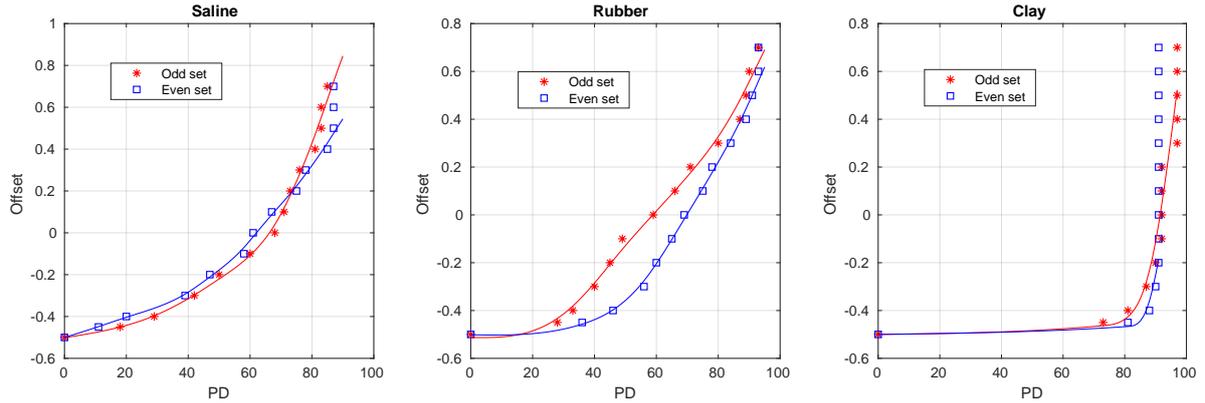


Figure 8: The relations between offset and probability of detection for three different materials.

Table 5: Results of the adaptation to varying threat material signatures.

Threat Definition			Test Set	PD						PFA
Material	minMass	Rho range		saline	rubber	clay	bulk	sheet	overall	overall
saline	250	1000-2000	odd	0.87	0.99	0.98	0.93	0.99	0.95	0.22
rubber	250	1000-2000	even	0.90	0.97	0.94	0.93	0.96	0.94	0.26
clay	250	1000-2000	all	0.88	0.98	0.96	0.93	0.97	0.94	0.24
saline	137	1050-1215	odd	0.86	0.98	0.92	0.90	0.97	0.92	0.21
rubber	180	1170-1290	even	0.87	0.92	0.91	0.91	0.88	0.90	0.25
clay	80	1530-1715	all	0.87	0.95	0.92	0.90	0.93	0.91	0.23
saline	137	1050-1215	odd	0.86	0.98	-	0.89	0.97	0.92	0.18
rubber	180	1170-1290	even	0.87	0.92	-	0.90	0.89	0.90	0.20
			all	0.87	0.95	-	0.95	0.89	0.93	0.19
saline	137	1050-1215	odd	0.83	-	-	0.83	1.00	0.83	0.16
			even	0.87	-	-	0.87	1.00	0.87	0.19
			all	0.85	-	-	0.85	1.00	0.85	0.17
rubber	180	1170-1290	odd	-	0.89	-	1.00	0.85	0.89	0.11
			even	-	0.91	-	1.00	0.88	0.91	0.15
			all	-	0.90	-	1.00	0.86	0.90	0.13
clay	80	1530-1715	odd	-	-	0.87	0.85	1.00	0.87	0.02
			even	-	-	0.90	0.92	0.80	0.90	0.02
			all	-	-	0.88	0.88	0.89	0.88	0.02

0.78 probability of detection when the requirements are 0.7 and 0.8 respectively. When the requirement increases to 0.9 and 0.95, our approach achieves 0.85 and 0.86 respectively which has been the best performance of this approach. In summary, our approach is able to adapt itself to satisfy varying probability of detection requirements unless the required probability of detection is beyond the upper performance bound of the approach.

4.5.2 Varying Threat Materials

This experiment aims to evaluate how our approach adapts to varying threat materials. We simulate this scenario by defining the threats in different ways which are shown in the first column of Table 5. Each definition specifies the materials that could potentially be threats as well as the requirements of physical properties (i.e., minimum mass ($minMass$) and density (Rho) range) that make an object be a threat for each material. In all cases, the required probability of detection is set as 0.9 and the learned offset functions are used to estimate the offset values in Eq.(6). There are totally six different definitions of threats in Table 5 to simulate six scenarios of threat recognition.

In the first two scenarios, three materials are defined as threats. The differences between these two scenarios lie in the specified minimum masses ($minMass$) and density (Rho) ranges.

The results in Table 5 show that our approach performs well in both scenarios. The performance for the second scenario is slightly worse than that for the first scenario in terms of probability of detection (PD). The reason is that smaller objects are required to be detected in the second scenario, which poses more challenges. On the other hand, the probabilities of false alarm achieved in the second scenario are reduced due to the additional restrictions of density ranges in this scenario so that many non-threat signatures can be correctly excluded from the final detection results.

In the third scenario, two materials of “*saline*” and “*rubber*” are defined as threats with the material of “*clay*” defined as non-threat. Benefiting from the material classification employed in our approach, the threat materials can be detected with a comparable probability of detection to that in the second scenario and the probabilities of false alarm are also significantly reduced since non-threat object signatures with densities close to that of “*clay*” can be easily excluded.

Finally, in the last three scenarios only one material is defined as threat. Significant drops of probability of false alarm can be observed from Table 5 when comparing to that in the second scenario where all three materials are defined as threats. The probabilities of detection also decrease by considerable margins. This is due to the overlapping intensity distributions of different materials and classification errors. For instance, in the second scenario, a true “*saline*” signature can still have a chance to be reserved as a positive detection even if it is misclassified into “*rubber*” or “*saline*” since all three materials are defined as threats, in the last three scenarios, however, it will definitely be mistakenly missed once it is misclassified hence leading to lower probability of detection.

Overall, the experimental results in Table 5 clearly indicate our proposed AATR approach has the adaptability to varying material requirements.

4.5.3 Varying Minimum Mass/Thickness

This experiment aims to evaluate how our approach adapts to varying physical properties (i.e. mass and thickness) of the threats. The proposed multi-scale segmentation algorithm (see Section 3.1.1) has the capability of detecting objects of varying scales, thus enabling our AATR adaptive to threats of varying mass/thickness.

To investigate how the approach performs in terms of varying masses in detail, we define three threat definitions with varying masses. We consider “*saline*” as the only threat material and set the mass range as $400 - Inf$, $300 - 400$ and $100 - 300$ as the requirements for being a threat respectively. The experimental results are shown in Table 6. Our approach is able to achieve the probability of detection of 0.94 and 0.93 when the threats have masses within the range of $400 - Inf$ and $300 - 400$, respectively. Apparently, threats with lower masses are more difficult to detect, but our approach can still achieve reasonable good performance with a probability of detection of 0.80 for the threats within the mass range of $100 - 300$.

A similar investigation is also made on how our approach performs in terms of varying thickness. In this experiment, we consider “*rubber*” as the only threat material since there are a reasonable amount of *rubber* sheets in the dataset. The thickness ranges of $10 - Inf$, $6.5 - 10$ and $0 - 6.5$ are selected as the requirements for being a threat respectively. Experimental results are shown in Table 7. Our approach is able to achieve the probability of detection of 0.90 and 0.89 when the threats have thickness within the range of $10 - Inf$ and $6.5 - 10$, respectively. Again, thinner threats within the thickness range of $0 - 6.5$ are more difficult to detect in which case our approach achieves reasonable well performance with a probability of detection of 0.84.

4.5.4 Unknown Threat Materials

One important requirement of the next generation AATR algorithms for baggage screening is to improve the detection performance of unknown threat materials which are not seen during training phase. The proposed approach in this paper has the ability to recognize unknown threat

Table 6: Results of the adaptation to varying Mass.

Threat Definition			Test Set	PD	PFA
Material	Rho range	Mass range			
saline	1050-1215	400-Inf	odd	0.85	0.16
			even	1.00	0.19
			all	0.94	0.17
saline	1050-1215	300-400	odd	0.92	0.16
			even	0.94	0.19
			all	0.93	0.17
saline	1050-1215	100-300	odd	0.82	0.16
			even	0.77	0.19
			all	0.80	0.17

Table 7: Results of the adaptation to varying Thickness.

Threat Definition			Test Set	PD	PFA
Material	Rho range	Thickness			
rubber	1170-1290	10-Inf	odd	0.83	0.11
			even	1.00	0.15
			all	0.90	0.13
rubber	1170-1290	6.5-10	odd	0.80	0.11
			even	1.00	0.15
			all	0.89	0.13
rubber	1170-1290	0-6.5	odd	0.76	0.11
			even	0.96	0.15
			all	0.84	0.13

material signatures due to its unsupervised segmentation and flexible classification algorithms. In this experiment, we test our algorithms on the sequestered dataset which consists of four different types of unknown materials named $m1$, $m2$, $m3$ and $m4$. We train our approach on the public dataset and test it on sequestered dataset in terms of the detection of unknown material signatures. The experimental results are shown in Table 8. Our AATR achieves high probability of detection of 76%, 100%, 92%, 100% for four types of unknown materials with the probability of false alarm of 12%, 46%, 15% and 11% respectively. It can be seen the performance on $m3$ and $m4$ are better than that on $m1$ and $m2$. The probability of detection for $m1$ is 76% which is lower than average, probably due to the small scales of those object signatures in the data. On the other hand, the probability of false alarm for $m2$ is as high as 0.46 although the probability of detection is perfect. One reasonable explanation is that many non-threat objects have overlapped intensity distribution with this material, thus making it difficult to distinguish between these threat and non-threat signatures.

Table 8: Results of the adaptation to unknown materials.

Threat Definition			PD	PFA
Material	minMass	Rho range		
m1	42	380-525	0.76	0.12
m2	67	770-810	1.00	0.46
m3	174	1300-1375	0.92	0.15
m4	183	1350-1430	1.00	0.11

We also compare the performance of our approach with those of [18] and [31] under the same setting. [18] achieved the probabilities of detection of 27%, 57%, 38% and 55% for $m1$ - $m4$ respectively [19] which are significantly lower than ours, while [31] achieved the probabilities of detection of 94%, 84%, 85% and 80% [32] which are lower than ours for three out of four materials but also with lower probabilities of false alarms than ours. Overall, it is indicated that our proposed AATR is promising in the adaptability to unknown materials which are not

available in training data.

5 Discussion

In this section, we discuss the potential issues underlying the proposed approach. One issue within X-ray CT image segmentation and classification are the metal artefacts caused by metal material in the baggage [25]. In our experiment, we focus more on the adaptability to varying requirements than the segmentation performance. Since the data used in our experiments were captured by a medical CT machine which suffers less from the metal artefacts, there is a risk of performance drop when the segmentation algorithm is applied to baggage data with significant artefacts. A potential solution to addressing this issue is to apply the metal artefact reduction algorithms [30, 29] as a pre-process, though this could be marginal and yet require considerable additional computation [25].

The proposed approach involves some parameters which are empirically selected based on the training data, leading to a risk of over-fitting to the data we currently have. In the segmentation algorithm, there is no training process but some very important parameters whose values are empirically selected. First, the parameter values of the structuring element used in the multi-scale morphological opening operation are set as 2, 3 and 8 in our experiments. These values work under the assumption that all the threat signatures should have volumes bigger than the smallest-scale structuring element so that they cannot be removed when applying the erosion operation. If the material of interest has smaller volumes, poorer performance can be expected. The other parameter in the segmentation algorithm, voxel number threshold, has a similar effect on the performance. Currently, the value of this parameter is set as 80×10^3 empirically, assuming that any segmented object having less than this volume will be non-threats. The potential over-fitting issues raised by these parameter values could be partly eliminated with prior knowledge of the task at hand. For example, if we know, in prior, the minimum volume of threat materials, proper values of these parameters can be decided.

In the classification algorithm, we need to train a four-class SVM classifier to recognize the material of a given segmented object. The classification accuracy is important for a low probability of false alarm by distinguishing non-threats from threats. It is also important for a high probability of detection by not mistakenly classifying a threat segment to be the non-threat. Our SVM classifier is trained on the ground truth threat objects in the training data plus synthesized features for unknown materials based on their density ranges. We use only the normalised histogram as the features for classification. As we can see from Figure 4, there are overlapping of features between different materials, hence learning a large margin between classes would be favourable to fight off the over-fitting issue.

Evaluations in this work are based on a relatively small dataset due to the challenge of collecting and annotating a large-scale CT baggage dataset. As a proof of concept, the proposed approach can achieve superior performance compared with other works [7, 18, 31] using the same dataset. However, more cluttered CT images and more types of threat materials will be involved in the real-world applications. To address these issues we will consider improving the 3D segmentation algorithms and employing more powerful material classification methods within the proposed AATR framework in our future work.

6 Conclusion

In this paper, we propose a solution to the adaptive automatic threat (explosive materials) recognition (AATR) problem. We propose an unsupervised segmentation algorithm for 3D CT images which works well on objects of different materials. Based on the segmentation results, our approach is able to adapt to the requirements of varying threat materials, varying probability

of detection and the capability of detecting new materials, without the need of re-training on the labelled data.

In our future work, the classification performance could be improved by using more features than only the normalised histogram or using dual-/multi-energy CT images. In our current approach, we have not explicitly handled the CT image noises such as metal artefacts which have been reported affecting the segmentation [20]. To enhance the adaptability to new materials, the techniques of zero-shot learning [37] and few-shot learning [36] can also be employed in future work.

References

- [1] Samet Akcay, Mikolaj E Kundegorski, Michael Devereux, and Toby P Breckon. Transfer learning using convolutional neural networks for object classification within x-ray baggage security imagery. In *IEEE International Conference on Image Processing*, pages 1057–1061. IEEE, 2016.
- [2] Samet Akcay, Mikolaj E Kundegorski, Chris G Willcocks, and Toby 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*, 13(9):2203–2215, 2018.
- [3] Muhammet Baştan. Multi-view object detection in dual-energy x-ray images. *Machine Vision and Applications*, 26(7-8):1045–1060, 2015.
- [4] Yudhijit Bhattacharjee. New efforts to detect explosives require advances on many fronts, 2008.
- [5] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [6] Matthew Caldwell, M Ransley, Thomas W Rogers, and Lewis D Griffin. Transferring x-ray based automated threat detection between scanners with different energies and resolution. In *Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies*, volume 10441, page 104410F. International Society for Optics and Photonics, 2017.
- [7] Carl Crawford. Advances in automatic target recognition (ATR) for CT-based object detection systems. https://myfiles.neu.edu/groups/ALERT/strategic_studies/T04_FinalReport.pdf, 2015.
- [8] Carl Crawford, Harry Martz, and David Castanon. Adaptive automated threat recognition (AATR). https://myfiles.neu.edu/groups/ALERT/strategic_studies/TaskOrder07/presentations/crawford%20aatr%20v07.pdf, 2018.
- [9] Carl Crawford, Harry Martz, and Homer Piena. Segmentation of objects from volumetric CT data. https://myfiles.neu.edu/groups/ALERT/strategic_studies/SegmentationInitiativeFinalReport.pdf, 2013.
- [10] Carl De Boor. *Spline toolbox for use with MATLAB: user’s guide, version 3*. MathWorks, 2005.
- [11] Limor Eger, Synho Do, Prakash Ishwar, W Clem Karl, and Homer Pien. A learning-based approach to explosives detection using multi-energy x-ray computed tomography. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2004–2007. IEEE, 2011.

- [12] Greg Flitton, Andre Mouton, and Toby P Breckon. Object classification in 3d baggage security computed tomography imagery using visual codebooks. *Pattern Recognition*, 48(8):2489–2499, 2015.
- [13] Leo Grady, Vivek Singh, Timo Kohlberger, Christopher Alvino, and Claus Bahlmann. Automatic segmentation of unknown objects, with application to baggage security. In *European Conference Computer Vision*, pages 430–444. Springer, 2012.
- [14] Nicole Hättenschwiler, Marcia Mendes, and Adrian Schwaninger. Detecting bombs in X-ray images of hold baggage: 2D versus 3D imaging. *Human factors*, 61(2):305–321, 2019.
- [15] Chih-Wei Hsu and Chih-Jen Lin. A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, 13(2):415–425, 2002.
- [16] Lay Khoon Lee, Siau Chuin Liew, and Weng Jie Thong. A review of image segmentation methodologies in medical image. In *Advanced computer and communication engineering technology*, pages 1069–1080. Springer, 2015.
- [17] Robert Liscouski and William McGann. The evolving challenges for explosive detection in the aviation sector and beyond. *CTC Sentinel*, 9(5):1–6, 2016.
- [18] Ankit Manerikar, Tanmay Prakash, and Avinash C Kak. Adaptive target recognition: A case study involving airport baggage screening. *arXiv preprint arXiv:1811.04772*, 2018.
- [19] Ankit Manerikar, Tanmay Prakash, and Avinash C Kak. A cascaded classification approach to ct-based adaptive target recognition. https://myfiles.neu.edu/groups/ALERT/strategic_studies/TaskOrder07/presentations/AATR_T07_May_Program_Review_Purdue_RVL_v04.pdf, 2018.
- [20] Limor Martin, Ahmet Tuysuzoglu, W Clem Karl, and Prakash Ishwar. Learning-based object identification and segmentation using dual-energy CT images for security. *IEEE Transactions on Image Processing*, 24(11):4069–4081, 2015.
- [21] Cynthia H McCollough, Shuai Leng, Lifeng Yu, and Joel G Fletcher. Dual-and multi-energy CT: principles, technical approaches, and clinical applications. *Radiology*, 276(3):637–653, 2015.
- [22] Najla Megherbi, Toby P Breckon, and Greg T Flitton. Investigating existing medical CT segmentation techniques within automated baggage and package inspection. In *Optics and Photonics for Counterterrorism, Crime Fighting and Defence IX; and Optical Materials and Biomaterials in Security and Defence Systems Technology X*, volume 8901, page 89010L. International Society for Optics and Photonics, 2013.
- [23] Domingo Mery, Erick Svec, Marco Arias, Vladimir Rizzo, Jose M Saavedra, and Sandipan Banerjee. Modern computer vision techniques for x-ray testing in baggage inspection. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(4):682–692, 2017.
- [24] Caijing Miao, Lingxi Xie, Fang Wan, Chi Su, Hongye Liu, Jianbin Jiao, and Qixiang Ye. Sixray: A large-scale security inspection X-ray benchmark for prohibited item discovery in overlapping images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2119–2128, 2019.
- [25] Andre Mouton and P Toby Breckon. On the relevance of denoising and artefact reduction in 3d segmentation and classification within complex computed tomography imagery. *Journal of X-ray science and technology*, 27(1):51–72, 2019.

- [26] Andre Mouton and Toby P Breckon. Materials-based 3D segmentation of unknown objects from dual-energy computed tomography imagery in baggage security screening. *Pattern Recognition*, 48(6):1961–1978, 2015.
- [27] Andre Mouton and Toby P Breckon. A review of automated image understanding within 3d baggage computed tomography security screening. *Journal of X-ray science and technology*, 23(5):531–555, 2015.
- [28] Andre Mouton, Toby P Breckon, Greg T Flitton, and Najla Megherbi. 3D object classification in baggage computed tomography imagery using randomised clustering forests. In *IEEE International Conference on Image Processing*, pages 5202–5206. IEEE, 2014.
- [29] Andre Mouton, Najla Megherbi, Katrien Van Slambrouck, Johan Nuyts, and Toby P Breckon. An experimental survey of metal artefact reduction in computed tomography. *Journal of X-ray Science and Technology*, 21(2):193–226, 2013.
- [30] Ram Naidu, Ibrahim Bechwati, Seemee S Karimi, Sergey Simanovsky, and Carl R Crawford. Method of and system for reducing metal artifacts in images generated by x-ray scanning devices, April 13 2004. US Patent 6,721,387.
- [31] David W Paglieroni, Hema Chandrasekaran, Christian Pechard, and Harry E Martz. Consensus relaxation on materials of interest for adaptive ATR in CT images of baggage. In *Anomaly Detection and Imaging with X-Rays (ADIX) III*, volume 10632, page 106320E. International Society for Optics and Photonics, 2018.
- [32] David W Paglieroni, Hema Chandrasekaran, Christian Pechard, and Harry E Martz. Consensus relaxation on materials of interest for adaptive ATR in CT images of baggage. https://myfiles.neu.edu/groups/ALERT/strategic_studies/TaskOrder07/presentations/LLNL_AATR_2018_ADSA_ProgramReview_V2.pdf, 2018.
- [33] Azriel Rosenfeld and John L Pfaltz. Sequential operations in digital picture processing. *Journal of the ACM*, 13(4):471–494, 1966.
- [34] Sergey Simanovsky, Zhengrong Ying, and Carl R Crawford. Method of and system for splitting compound objects in multi-energy computed tomography images, May 26 2009. US Patent 7,539,337.
- [35] Sameer Singh and Maneesha Singh. Explosives detection systems (EDS) for aviation security. *Signal processing*, 83(1):31–55, 2003.
- [36] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, pages 4080–4090, 2017.
- [37] Qian Wang and Ke Chen. Zero-shot visual recognition via bidirectional latent embedding. *International Journal of Computer Vision*, 124(3):356–383, 2017.
- [38] K Wells and DA Bradley. A review of x-ray explosives detection techniques for checked baggage. *Applied Radiation and Isotopes*, 70(8):1729–1746, 2012.
- [39] David Wiley. Analysis of anatomic regions delineated from image data, June 5 2012. US Patent 8,194,964.
- [40] David F Wiley, Deboshmita Ghosh, and Christian Woodhouse. Automatic segmentation of CT scans of checked baggage. In *Proceedings of the 2nd International Meeting on Image Formation in X-ray CT*, pages 310–313, 2012.

- [41] Ting-Fan Wu, Chih-Jen Lin, and Ruby C Weng. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 5(Aug):975–1005, 2004.