On the Evaluation of Semi-Supervised 2D Segmentation for Volumetric 3D Computed Tomography Baggage Security Screening

Qian Wang
Department of Computer Science
Durham University
Durham, UK

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

***Abstract*—We address the automatic contraband material detection problem within volumetric 3D Computed Tomography (CT) data for baggage security screening. Distinct from the prohibited item detection using object detection techniques, contraband material detection is usually formulated as a segmentation problem due to the variations of their potential appearances and shapes. Previous studies have employed either morphological operation based traditional methods or 3D Convolutional Neural Networks (CNN) for 3D segmentation towards target material detection within volumetric 3D CT baggage security screening imagery. In this work, we investigate the effectiveness of 2D semantic segmentation techniques in this 3D CT segmentation problem. Specifically, we extract 2D slices from three planes of the 3D CT volumes and train a 2D segmentation model which is subsequently used to predict segmentation results for all the slices from a given test CT volume. Moreover, we also evaluate how the performance is affected when using a reduced number of annotated slices for training. As a result, it is demonstrated reasonable performance can be achieved with very limited annotated slices (1-2) per CT volume during training. Finally, we propose a semi-supervised learning framework for 3D CT segmentation. Using only 1/128 of the total number of annotated slices, our framework can achieve comparable performance with full supervision.**

***Index Terms*—3D volumetric data, X-ray computed tomography, baggage security screening, 3D segmentation, material discrimination.**

I. Introduction

Automatic threat detection in baggage security screening enables more efficient and safer transportation. Recent advances in deep learning and image processing make it possible for automatic threat detection in 2D X-ray and 3D Computed Tomography (CT) baggage security imagery [1]–[6]. State-of-the-art deep learning models for image classification and object detection trained on ImageNet [7] and MS-COCO [8] can be transferred to X-ray imagery by fine-tuning the pre-trained models on relatively smaller X-ray datasets. However, existing studies on X-ray and 3D CT imagery focus on the detection of prohibited items such as firearms, knives and electronics whilst overlooking contraband materials which do not have specific shape and appearance characteristics.

To provide enhanced security screening capabilities, dual energy 3D CT scanners are now our commonplace for both hold and carry-on baggage screening within aviation security. 3D volumetric CT imagery provides more information for baggage security as well as the possibility of distinguishing different types of materials based on their density and effective-Z characteristics [9]. 3D Convolutional Neural Network (CNN) based methods have been used for prohibited item detection within 3D CT imagery and achieved promising results [4], [5], [10]–[13]. Similar to the situation in 2D X-ray imagery [1], these approaches to prohibited item detection rely on the specific shapes and appearances hence unsuitable for contraband material detection.

Efforts were made towards material classification within 3D CT imagery in [6] and [14] using a hand-engineered framework and 3D CNN based deep learning techniques, respectively. One limitation of these existing works is the expensive computational cost introduced by the use of 3D CNN. In addition, annotating the 3D CT volumes is also a laborious and time-consuming task. Earlier work on the segmentation of 3D CT imagery for baggage security [6], [9], [15] had similarly suffered from high computational cost. In this paper, to address these problems, we explore the possibility of replacing the 3D segmentation with its 2D counterparts for contraband material segmentation and detection within 3D CT imagery for baggage security screening.

Specifically, we apply the leading contemporary 2D semantic segmentation methods to 2D slices extracted from 3D CT volumes and compare the performance with other methods. Moreover, we investigate the effect of reducing the number of annotated CT slices used for training. To ensure the performance of segmentation and detection with a small number of annotated slices, we propose a semi-supervised learning framework based on pseudo-labeling and evaluate its effectiveness via extensive experiments.

To summarize, the contributions of this paper are as follows:

- the first attempt to address contraband material detection within volumetric 3D CT baggage security screening imagery using deep learning based 2D semantic segmentation methods.
- a semi-supervised learning framework is proposed to enable comparable segmentation performance with a sig-

nificantly reduced number of annotated slices for training.

- extended experimentation is performed to investigate how the number of annotated slices affects the performance of 3D CT segmentation. Using only a couple of annotated slices per CT volume for training harms the performance whilst the proposed semi-supervised learning framework can boost the performance. On a 3D CT segmentation dataset, our methods based on 2D CNN outperform existing approaches based on 3D CNN [14] or hand-engineered morphological operations [6].

II. Related Work

In this section, we review existing works related to ours from the perspective of *object detection for baggage security screening*, *material discrimination in 2D X-ray imagery*, *material discrimination in 3D CT volumes* and *Weakly Supervised 3D Segmentation*.

A. Object Detection for Baggage Security Screening

Automatic threat detection has been studied within 2D X-ray imagery [1]–[3] and 3D CT imagery [4], [5], [10]–[13]. State-of-the-art object detection frameworks such as Faster R-CNN [16] has been used to detect firearms, knives and other prohibited items in 2D baggage screening imagery. Pre-trained on large-scale ImageNet and MS-COCO, these object detection models can be easily fine-tuned for X-ray imagery. In [4], [5], the object detectors were adapted from 2D to 3D and used for firearm, bottle detection within volumetric 3D CT imagery. Promising results have been reported whilst the scale of dataset in terms of the numbers of category and training CT samples remains moderate compared with 2D X-ray datasets due to the difficulties of collating, annotating and storing large 3D CT datasets.

B. Material Discrimination in 2D X-ray Imagery

There exist some limited work on material classification based on X-ray imagery. For example, Chen et al. [17] proposed a curve-based material recognition method by theoretic analysis of X-ray imaging processing using high-energy dual energy X-ray (6/3 MeV). Specifically, they consider the ratio of two X-ray energies after penetrating materials, resulting in a standard curve for a specific material which can be used to discriminate the material from others.

Li et al. [18] proposed a dynamic material discrimination algorithm to tackle the material overlapping problem. A dual-energy radiograph database of both pure basis materials and pair combinations was established. This method can only handle the overlapping of two known materials.

Instead of using standard *classification curves*, Chang et al. [19] investigate the use of machine learning methods for X-ray imagery based material classification. In this work, different numbers of energies are compared in terms of the classification of metal, organic and inorganic materials. It is concluded the use of as least as four energies could achieve reasonable classification performance.

The X-ray technology based Explosive Detection System (EDS) [20] used in aviation security screening is also based on material discrimination within X-ray imagery [21]. According to [22], each object in baggage is examined for a match to a specific effective atomic number Z_{eff} [23], density, and mass threshold so that the material components of the object can be identified. Prior knowledge is required on the typical components of explosive and non-explosive materials (e.g. metal, organic and inorganic materials). On the other hand, the X-ray intensity is not only related to the material but also related to the thickness of the materials [21] hence leading to more complexity in cluttered baggage.

Although achievements have been made in these works, there exist an inherent limitation of 2D X-ray imagery in material discrimination. The overlap of different materials in 2D X-ray images can pose a significant additional challenge in real-world applications of baggage security screening. Our work takes the advantage of 3D CT imagery and expects to better distinguish different types of materials especially the target contraband ones within volumetric 3D baggage security screening imagery.

C. Material Discrimination in 3D CT Volumes

Material discrimination within 3D baggage CT imagery has been studied in literature [6], [14], [24]–[26]. Existing works usually formulate it as a 3D segmentation problem followed by classification. Early works take advantage of the multi-energy CT data and extract discriminative hand-crafted features from the CT images and Z_{eff} images [24]. CT images contain density information whilst Z_{eff} images are the measurements of atomic numbers. Mouton et al. [25] proposed a two-stage approach for object segmentation within dual-energy CT imagery based on the voxel intensity ranges of pre-defined materials followed by a classifier. Wang et al. [6] and others [27], [28] studied the problem of adaptive automatic threat recognition problem for baggage security screening. The proposed solutions were also based on the segmentation and classification of material characteristics by extracting hand-crafted features from single-energy CT data.

More recently, 3D CNN based U-Net and its variations were investigated for 3D CT segmentation and contraband material detection in [14]. In addition, the volumetric 3D CT volumes were converted into point clouds in order to offer a potentially computationally efficient alternative processing representation. The point clouds were subsequently processed by PointNet [29] and PointNet++ [30]. However, these existing methods suffer from high computational cost and have to down-sample the CT volumes to fit within the memory constraints of a typical GPU for model training. To solve this problem, our work in this paper employ 2D CNN models to solve the 3D CT segmentation problem.

D. Weakly Supervised 3D Segmentation

Annotating 3D volumetric data such as CT imagery is extremely time-consuming hence weakly supervised approaches have been proposed for 3D segmentation without the need of exhaustive voxel-wise annotations of training data. Weak supervision can have varying meanings in different contexts, such as incomplete annotations and inexact annotations. Kervadec et al. [31] incorporate inequality constraints into the loss function of CNN training so that the model trained with a fraction of annotations achieves comparable segmentation performance in medical image analysis with those trained with full supervision. The constraints used in [31] are based on the prior knowledge of target size which is reasonable for medical image segmentation but not applicable for our case of baggage security screening where contraband materials can be of arbitrary sizes. Xu et al. [32] addressed the incomplete supervision problems in point cloud segmentation. Several complementary components were combined in their framework including an incomplete supervision branch, an inexact supervision branch, Siamese self-supervision, spatial and color smoothness constraints. Li et al. [33] addressed the labelled data scarcity issue from the perspective of semi-supervised learning. In semi-supervised learning setting for 3D CT segmentation, training data are composed of both labelled and unlabelled CT volumes. By contrast, our study assumes the availability of partial annotations (e.g., only a few slices within a CT volume are annotated) of volumetric CT data.

Pseudo-labeling is a popular technique for semi-supervised learning and unsupervised domain adaptation problems [34]. We take advantage of this technique in our study to handle the partially annotated training data in a semi-supervised learning framework.

III. Method

In this section, we describe our proposed framework for 3D CT volume segmentation using 2D CNN models as well as the framework of semi-supervised learning for weakly supervised scenarios when only a small number of annotated slices are available per CT volume. In addition, we also describe how to perform the contraband material detection based on the segmentation results in the real-world application of baggage security screening.

A. 3D CT Segmentation Using 2D Segmentation Frameworks

In this work, we leverage leading contemporary architectures for semantic segmentation to solve our problem in 3D CT segmentation. We use Fully Convolutional Networks (FCN) [35], [36], one of the most popular frameworks for semantic segmentation in our work with ResNet101 [37] as the backbone. The classifier layer and the fully-connected layers in the original ResNet101 architecture are replaced with convolutional layers to output segmentation maps. The final segmentation map contains the predicted category index for each pixel. For more details of FCN and ResNet101 we refer the readers to [35] and [37] respectively. Here we focus more on the framework of applying the segmentation model to 3D CT segmentation.

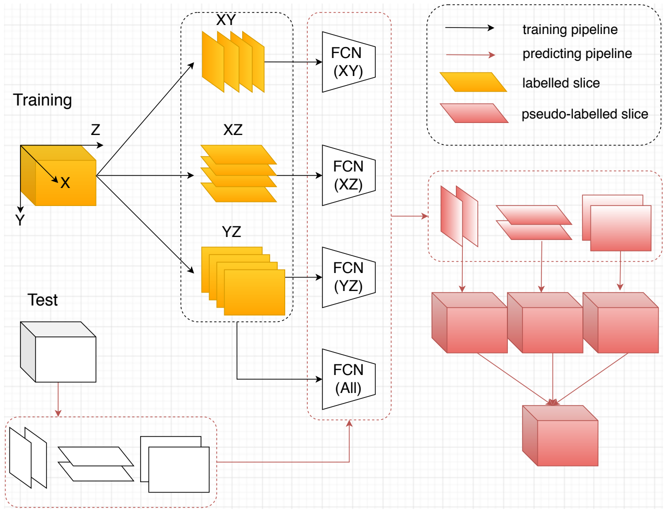

Fig. 1. The framework of 3D CT segmentation using 2D segmentation networks FCN.

Figure 1 shows the schematic pipelines of training and testing the segmentation model. As depicted in the figure, a 3D CT volume can be represented within a 3D coordinate system XYZ where Z axis is the direction of scanner tunnel and the XY plane is the scanning plane. Given a volumetric 3D CT volume of size $W \times H \times N$, we can extract 2D slices from three planes. As a result, the numbers of extracted 2D slices from XY, XZ and YZ planes will be N, W and H, respectively. For training data, the corresponding label volumes can be sliced into 2D labels associated with the 2D slices in the same way. As a result, four different models can be trained by using different sets of training slices (i.e. XY, XZ, YZ and All).

As indicated by the "red" pipeline in Figure 1, these four trained models can be subsequently used for prediction. The slices from one single plane are enough to form the final segmentation results for the 3D volume. We can have three options by using slices extracted from three different planes to do the prediction and form the 3D segmentation results. In addition, the three different segmentation results can be combined using majority voting towards more accurate results.

B. Semi-supervised Learning with Pseudo-labeling

In previous section, we assume the existence of annotated 3D volumes. However, the annotation of 3D CT volumes are laborious and time-consuming. In this section, we employ semi-supervised learning for segmentation so that the amount of annotated slices required for training can be significantly reduced.

The semi-supervised learning framework is illustrated in Figure 2. Assuming there are only a couple of annotated slices in a given 3D CT volume (indicated by the yellow color) in a specific plane, we use these annotated slices to train a segmentation model, i.e. FCN (XZ) in our case. The model can be used to predict the segmentation results of other unlabeled slices in the training data. The prediction (indicated by the red color) can be noisy but provide additional supervision based on which we train a second segmentation model. The second model is trained based on all pseudo-labeled slices from three planes, hence denoted as FCN (All) in Figure 2.

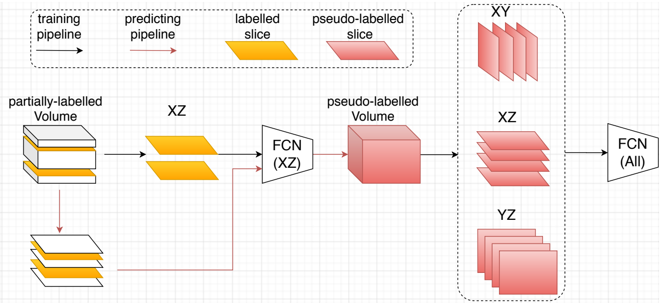

Fig. 2. The framework of semi-supervised learning approach for weakly supervised 3D CT segmentation using pseudo-labeling.

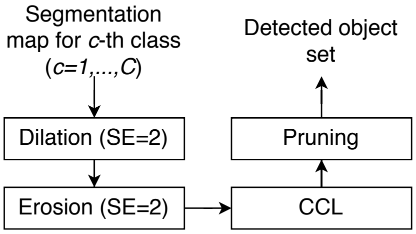

Fig. 3. The pipeline of post-processing using morphological operations to convert segmentation results to detection results.

C. Contraband Material Detection by Segmentation

In the baggage security screening, we expect to detect contraband materials based on the segmentation results. To this end, we follow [14] and convert the segmentation results to detection results in the post-processing stage. Specifically, we group the connected voxels which are labelled as the same class as a detected object. To these ends, we use morphological operations to correct the mislabeling information in the segmentation results.

The pipeline of post-processing is shown in Figure 3. For each foreground class, we apply *dilation* and *erosion* operations sequentially to the binary segmentation map to correct the missing voxel labels within the detected objects. A sphere structural element is used for 3D dilation and erosion. Subsequently, the connected component labeling (CCL) algorithm is employed to group the labelled voxels into a set of potential detected objects. We prune the detection results by removing the objects whose volumes are smaller than a pre-defined threshold. In our experiment, the threshold is empirically set, as $V/10e4$ where $V = WHN$ is the number of voxels in the 3D CT volume. After the post-processing, we obtain the locations and material types of a list of detected objects in a given CT volume.

IV. Experiments and Results

In this section, we present details of datasets and evaluation metrics used in the experiments, followed by experimental settings and results. Three experiments are conducted to evaluate the effectiveness of the proposed methods. The first experiment is designed to compare the performance of 2D segmentation and its 3D counterparts including different variations of 3D U-Net and PointNet. In the second experiment, we investigate the effect of number of annotated slices used for training. Finally, we evaluate the proposed semi-supervised framework for 3D CT segmentation.

A. Dataset

We follow [6], [14] and use the Northestern University Automated Threat Recognition (NEU ATR) dataset [38] collected and annotated by NEU ALERT [27] throughout our experiments in this study. Baggage CT volumes were collected by a medical CT scanner (Imatron C-300). The slice size is 512×512 corresponding to the field view of 475 mm×475 mm hence the in-plane pixel size is 0.928 mm. The number of slices varies in different volumes and the slice spacing is 1.5 mm. Pixel values are represented by the Modified Hounsfield Unit (MHU) ranging from 0 to 32,767 MHU in which air and water are 0 and 1024 respectively.

The ATR dataset consists of 188 CT volumes in which there are 446 object signatures of three target materials (i.e. *saline*, *rubber* and *clay*) and other non-target materials as cluttered background of typical packed baggage. The ground truth voxels are labelled by NEU ALERT for all the objects of three target materials. We follow [6] to split the whole dataset into two subsets evenly: *odd* set and *even* set containing 94 odd and even indexed volumes respectively. In our experiments, we use one subset for training and the other for testing.

B. Evaluation Metrics

We follow previous works [6], [14] and use three groups of evaluation metrics in our experiments. The first one is the mean Intersection over Union (IoU) which is a typical evaluation metric for semantic segmentation as the contraband materials detection has been formulated as a semantic segmentation problem. The IoU is computed for each class and the mean IoU is the mean of each IoU for all classes. Although the proposed method works on 2D slices, the IoUs are computed based on 3D volumes.

In addition, we evaluate the performance of contraband materials detection. The detection results are obtained after post-processing the segmentation results (c.f. Section III-C). To this end, we use the second group of evaluation metrics precision and recall which are computed based on detection results.

The third group of evaluation metrics are similar to typical ones for object detection (i.e. precision and recall in the second group) but concern the Probability of Detection (PD) and the Probability of False Alarms (PFA) which have been used in [6]. PD is similar to recall and the main difference is that PD is computed over all detections regardless of their classes whilst recall is computed class-wisely. PD is defined in this way so that the detection model focuses more on the difference between contraband materials and benign ones rather than the difference between different types of contraband materials. PFA is defined as the ratio of the number of falsely detected non-threat signatures to the total number of non-threats in the CT images.

C. Implementation Details

We use the open-source 2D segmentation tool MMSegmentation developed by [39]. Fully Convolutional Networks (FCN) [36] is selected as the segmentation model and ResNet101 [37] is used as the backbone in our experiments. The SGD optimizer is used for training with the learning rate of 0.01, the momentum of 0.9 and the weight decay of 0.0005. A polynomial learning rate decay [40] is used with the minimum learning rate of 1e-4. Different numbers of training iterations [1] (e.g., 20k, 40k, 80k and 160k) are investigated in our preliminary experiments and the number of 80k is chosen as the trade-off between performance and efficiency in the following experiments.

TABLE I
THE EFFECT OF TRAINING/PREDICTING WITH SLICES EXTRACTED FROM DIFFERENT PLANES IN TERMS OF IOU.

Training	Predict	Background	Saline	Rubber	Clay	mIoU
XY	XY	99.5	49.4	60.0	67.8	69.2
	XZ	99.5	41.4	50.9	65.9	64.4
	YZ	99.3	13.9	28.7	54.7	49.1
	All	99.6	42.6	59.9	72.1	68.5
XZ	XY	99.2	37.1	49.0	62.8	62.0
	XZ	99.6	57.4	60.4	71.7	72.3
	YZ	99.3	22.8	35.5	61.5	54.8
	All	99.6	49.5	62.2	73.2	71.1
YZ	XY	99.1	22.4	41.2	46.3	52.2
	XZ	99.2	21.5	45.7	58.1	56.1
	YZ	99.5	44.0	57.6	68.0	67.3
	All	99.4	28.4	57.0	67.1	63.0
All	XY	99.4	52.1	57.6	70.0	69.8
	XZ	99.6	57.6	65.2	75.8	74.5
	YZ	99.4	45.5	58.5	72.9	69.1
	All	**99.6**	**59.0**	**68.9**	**78.3**	**76.5**

D. On the Effect of Slice Planes

We conduct experiments to investigate the effect of training/predicting with slices extracted from different planes (i.e. XY, XZ or YZ). On one hand, we use annotated slices extracted different planes for training and report the segmentation performance in terms of IoU. On the other hand, we predict the final 3D segmentation results by stacking the segmented slices extracted from different planes. In addition, we consider using slices from all planes by combining all slices during training and via the majority voting method during prediction.

The results are reported in Table I from which the following conclusions can be drawn. Firstly, when training on one plane, prediction within the same plane gives better results than prediction within the other planes. Secondly, among the three planes, training and prediction within the XZ plane gives the best performance. Given that XZ plane is not the plane in which CT slices are internally reconstructed by the scanner, such a finding is interesting and worth a further investigation across larger dataset availability. Finally, the combination of slices from all three plane orientations benefits the training of segmentation models, but is not always helpful for prediction except the case of last row in Table I where all the slices are also used for training.

TABLE II
SEGMENTATION PERFORMANCE (IOU) OF THE PROPOSED SEMI-SUPERVISED LEARNING METHOD.

Method	Predict	Background	Saline	Rubber	Clay	mIoU
Supervised	XY	99.3	32.2	47.2	48.8	56.9
	XZ	99.4	40.5	49.8	58.4	62.0
	YZ	99.3	18.8	22.6	50.8	47.9
	All	99.5	32.2	52.8	60.8	61.3
Semi-Supervised	XY	99.5	49.8	59.2	68.4	69.2
	XZ	99.6	57.6	65.2	74.5	74.2
	YZ	99.5	44.7	57.8	72.5	68.6
	All	**99.6**	**59.4**	**69.1**	**77.6**	**76.4**

TABLE III
IOU RESULTS OF MATERIAL SEGMENTATION WITHIN 3D CT VOLUMES ON NEU ATR DATASET (* DENOTES THE NUMBER OF ANNOTATED SLICES IS SIGNIFICANTLY REDUCED TO 1/128 OF THE FULL ANNOTATIONS).

Method	Background	Saline	Rubber	Clay	mIoU
PointNet [14]	99.0	15.7	15.5	31.0	40.3
PointNet++ [14]	98.9	39.8	28.2	61.9	57.2
3D U-Net [14]	**99.6**	64.9	63.0	72.5	75.0
Resisual 3D U-Net [14]	**99.6**	**67.4**	64.6	67.9	74.9
Ours 2D FCN	**99.6**	59.0	68.9	**78.3**	**76.5**
Ours 2D FCN *	99.5	32.2	52.8	60.8	61.3
Ours 2D FCN *(semi-supervised)	**99.6**	59.4	**69.1**	77.6	76.4

E. On the Number of Annotated Slices

In this experiment, we aim to investigate how the segmentation performance will be affected by reducing the number of the annotated slices. We extract annotated slices from the XZ plane since it has demonstrated XZ plane gives the best performance in the previous subsection. To simulate the situation when we only annotate a fraction of slices for training, we sample annotated slices evenly by a factor of $f \in \{2, 4, 6, 8, 16, 32, 64, 128\}$ and use the sampled slices for training. Once the models are trained, the slice-wise segmentation results are stacked to form the final 3D volume segmentation results. Similar to the previous experiments, the prediction can be done in either of three planes or by the combination of all planes.

The experimental results are shown in Figure 4. The segmentation performance for three target materials *saline*, *rubber*, *clay* and the mean IoU over these three materials are presented in four plots respectively. In each plot, the segmentation performances based on slices within three different planes and their combination are compared. From Figure 4 we can see that segmentation performance drops when less annotated slices (i.e. greater down-sampling factors) are used for training. The performance drop is not significant when the down-sampling factor f increases from 1 to 16, which means the existence of redundancy in the fully annotated training data. However, when the down-sampling factor f increases further to 32, 64 and 128, clear performance drops can be observed for most cases shown in Figure 4. In particular, when the down-sampling factor f is as high as 128, equivalent to using an average of ~1-2 annotated slices per CT volume for training, the segmentation performance in terms of mIoU is 61.3% which is not a drastically reduced level of performance when compared with the mIoU of 71.1% achieved with a full set of training annotations (i.e. no down sampling, factor f=1).

[1] One iteration processes one batch of data and updates the parameters.

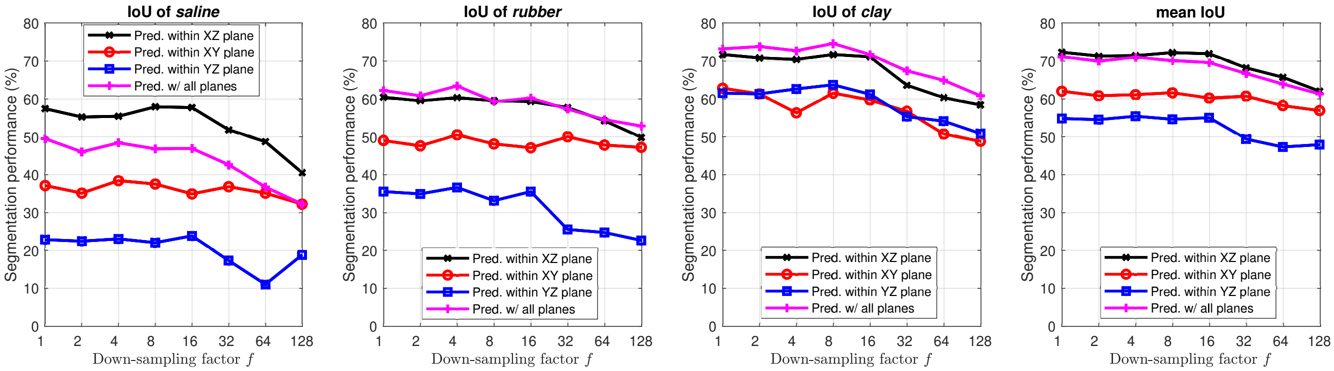

Fig. 4. Segmentation performance with different slice sampling factors f (i.e. $1/f$ of the annotated CT slices are evenly sampled from the CT volumes for training).

TABLE IV

PRECISION AND RECALL RESULTS OF MATERIAL SEGMENTATION WITHIN 3D CT VOLUMES ON NEU ATR DATASET (* DENOTES THE NUMBER OF ANNOTATED SLICES IS SIGNIFICANTLY REDUCED TO 1/128 OF THE FULL ANNOTATIONS).

Method	Saline		Rubber		Clay		Overall	
	P (%)	R (%)	P (%)	R (%)	P (%)	R (%)	P (%)	R (%)
PointNet [14]	35.0	62.8	40.4	56.8	58.2	73.6	44.5	64.4
PointNet++ [14]	41.9	84.1	59.7	58.9	60.1	76.9	53.9	73.3
3D U-Net [14]	71.8	**91.0**	87.4	92.5	94.6	85.8	84.6	**89.8**
3D Residual U-Net [14]	78.4	82.5	87.8	87.3	90.4	91.3	85.5	87.1
Ours 2D FCN	83.1	77.6	87.2	94.9	98.1	89.7	89.5	87.4
Ours 2D FCN *	82.1	52.0	86.7	87.8	95.3	83.5	88.0	74.4
Ours 2D FCN * (semi-supervised)	**84.1**	76.3	**91.5**	**94.9**	**98.1**	**91.4**	**91.2**	87.5

TABLE V

PD AND PFA RESULTS OF MATERIAL SEGMENTATION WITHIN 3D CT VOLUMES ON NEU ATR DATASET (* DENOTES THE NUMBER OF ANNOTATED SLICES IS SIGNIFICANTLY REDUCED TO 1/128 OF THE FULL ANNOTATIONS).

Method	PD (%)				PFA (%)
	Saline	Rubber	Clay	Overall	Overall
SVM [6]	87	95	**96**	92	24
PointNet [14]	81	84	88	84	29
PointNet++ [14]	**97**	87	94	**92**	24
3D U-Net [14]	91	97	83	91	6
3D Residual U-Net [14]	86	92	**92**	90	4
Ours 2D FCN	72	87	88	82	**3**
Ours 2D FCN *	46	86	86	72	**3**
Ours 2D FCN * (semi-supervised)	78	**100**	92	90	**3**

F. Semi-supervised Learning with Pseudo-labeling

In the previous experiment, it has been demonstrated using only ~1-2 annotated slices per CT volume for training can achieve reasonably good segmentation performance. In this experiment, we aim to investigate how the proposed semi-supervised learning with pseudo-labeling can further improve the performance. We employ the method described in Section III-B and use down-sampled ($f = 128$) annotated slices from plane XZ to train the initial segmentation model. This model is then used to segment the slices which are not used for training (i.e. the simulation of unlabeled slices in real-world applications). As a result, the training CT volumes are fully pseudo-labelled which are then subsequently used for training a second segmentation model with extracted pseudo-labelled slices from all three planes. The performance of the final segmentation model trained in such a semi-supervised learning way is shown in Table II. Compared with the results of supervised learning, the proposed semi-supervised learning method significantly improves the segmentation performance in all cases. In particular, the best performance of the semi-supervised learning method is achieved by using an ensemble of three planes in prediction. The best mIoU of 76.4% is comparable with the performance when all annotated slices are used (mIoU=76.5% in Table I), which is impressive given the fact only ~1-2 annotated slices are needed per CT volume.

G. Comparison with 3D Segmentation

Prior work has investigated to use 3D CNN for material segmentation and detection in [14]. In addition, volumetric 3D CT volumes are converted to point clouds on which PointNet [29] and PointNet++ [30] are employed for 3D segmentation to save GPU memory during training. In this experiment, we compare our proposed method with those based on 3D segmentation in [14].

Following [14], we evaluate the performance using three metrics (i.e. IoU, precision/recall and PD/PFA) and report the comparison results in Tables III-V. We compare against the best reported results of 3D segmentation based methods (i.e. PointNet [29], PointNet++ [30], 3D U-Net [41] and Residual 3D U-Net [42]) in [14]. It is clearly shown that our 2D FCN method achieves superior performance than its 3D counterparts in terms of mIoU and precision/recall. In addition, when the number of annotated slices for training is significantly reduced to 1/128 of the full annotation (equivalent to ~1-2 slices per CT volume), the proposed semi-supervised learning method can still achieve state-of-the-art performance with overall mIoU

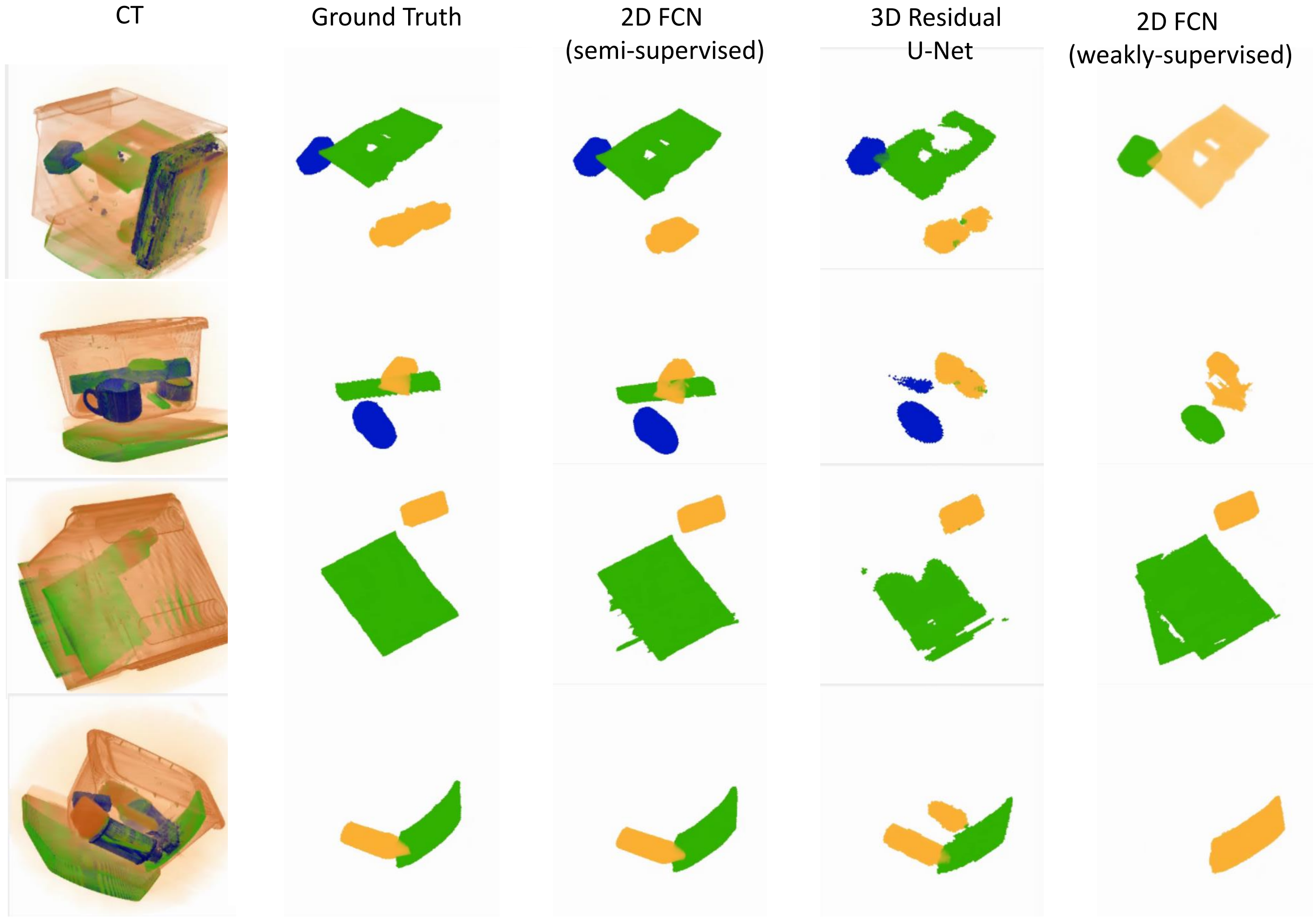

Fig. 5. Qualitative evaluation of detection results of different approaches (from left to right: CT volumes, ground truth labels, our proposed 2D FCN using semi-supervised learning, 3D Residual U-Net [14]), 2D FCN without semi-supervised learning).

of 76.4%, overall precision/recall of 91.2%/87.5% and overall PD/PFA of 90%/3%.

Figure 5 presents a qualitative comparison between 2D and 3D CNN based approaches. From the leftmost column to the right, we present the CT volumes, ground truth labels, segmentation results of our proposed 2D FCN with pseudo-labeling based semi-supervised learning using reduced annotated slices, results of 3D Residual U-Net with full annotation, results of 2D FCN without semi-supervised learning using reduced annotated slices, respectively. Four examples are shown in Figure 5 from which we can clearly see that our proposed 2D FCN with semi-supervised learning achieves the best segmentation results than other two approaches. 3D Residual U-Net suffers from missing labeling whilst 2D FCN without semi-supervised learning is prone to voxel misclassification.

V. Conclusion

In this paper, we investigate the possibility of using 2D CNN semantic segmentation methods for 3D CT segmentation and achieved comparable performance with its 3D counterparts in [14]. The use of 2D segmentation relaxes the requirement of high-level computational resources during training. More importantly, the success of semi-supervised 2D slice segmentation framework for 3D CT segmentation enables us to annotate a small number of slices per CT volume hence to more readily scale future baggage screening datasets for contraband material detection.

One limitation of the this work is that only three target materials are considered in our experiments due to the constraints of dataset availability. In the future we will investigate further experimentation using a larger and more varied dataset. Although we have demonstrated promising results in our experiments, the effectiveness of the proposed approach in real-world applications is yet to be validated.

Acknowledgement [27], [38] - This material is based upon work supported by the U.S. Department of Homeland Security, Science and Technology Directorate, Office of University Programs, under Grant Award 18STEXP00001-03-02, Formerly 2013-ST-061-ED0001. 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.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] ——, "Multi-class 3D object detection within volumetric 3D computed tomography baggage security screening imagery," in *Proc. IEEE Int. Conf. on Machine Learning and Applications*, 2020, in press.

[6] 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.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. Ieee, 2009, pp. 248–255.

[8] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European Conf. on Computer Vision*. Springer, 2014, pp. 740–755.

[9] A. Mouton and T. Breckon, "Materials-based 3d segmentation of unknown objects from dual-energy computed tomography imagery in baggage security screening," *Pattern Recognition*, vol. 48, no. 6, p. 1961–1978, June 2015.

[10] G. Flitton, T. Breckon, and N. Megherbi, "A 3D extension to cortex like mechanisms for 3D object class recognition," in *Proc. Computer Vision and Pattern Recognition*. IEEE, June 2012, pp. 3634–3641.

[11] N. Megherbi, J. Han, G. Flitton, and T. Breckon, "A comparison of classification approaches for threat detection in CT based baggage screening," in *Proc. Int. Conf. on Image Processing*. IEEE, September 2012, pp. 3109–3112.

[12] G. Flitton, T. 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, September 2013.

[13] G. Flitton, A. Mouton, and T. Breckon, "Object classification in 3d baggage security computed tomography imagery using visual codebooks," *Pattern Recognition*, vol. 48, no. 8, p. 2489–2499, August 2015.

[14] Q. Wang and T. P. Breckon, "Contraband materials detection within volumetric 3D computed tomography baggage security screening imagery," *arXiv preprint arXiv:2012.11753*, 2020.

[15] A. Mouton and T. 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*, vol. 27, no. 1, pp. 51–72, April 2019.

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

[17] Z.-Q. Chen, T. Zhao, and L. Li, "A curve-based material recognition method in mev dual-energy x-ray imaging system," *Nuclear Science and Techniques*, vol. 27, no. 1, p. 25, 2016.

[18] L. Li, R. Li, S. Zhang, T. Zhao, and Z. Chen, "A dynamic material discrimination algorithm for dual mv energy x-ray digital radiography," *Applied Radiation and Isotopes*, vol. 114, pp. 188–195, 2016.

[19] Q. Chang, W. Li, and J. Chen, "Application of machine learning methods for material classification with multi-energy x-ray transmission images," in *International Conference on Artificial Intelligence and Security*. Springer, 2019, pp. 194–204.

[20] K. Wells and D. Bradley, "A review of X-ray explosives detection techniques for checked baggage," *Applied Radiation and Isotopes*, vol. 70, no. 8, pp. 1729–1746, 2012.

[21] S. Singh and M. Singh, "Explosives detection systems (eds) for aviation security," *Signal processing*, vol. 83, no. 1, pp. 31–55, 2003.

[22] R. F. Eilbert and K. D. Krug, "Aspects of image recognition in Vivid Technologies' dual-energy x-ray system for explosives detection," in *Applications of Signal and Image Processing in Explosives Detection Systems*, J. M. Connelly and S. M. Cheung, Eds., vol. 1824, International Society for Optics and Photonics. SPIE, 1993, pp. 127 – 143.

[23] R. Rutherford, B. Pullan, and I. Isherwood, "X-ray energies for effective atomic number determination," *Neuroradiology*, vol. 11, no. 1, pp. 23–28, 1976.

[24] Z. Ying, R. Naidu, S. Simanovsky, M. Hirsch, and C. R. Crawford, "Method of and system for classifying objects using local distributions of multi-energy computed tomography images," Sep. 2010, US Patent 7,801,348.

[25] 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.

[26] ——, "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.

[27] C. Crawford, "Advances in automatic threat recognition (ATR) for CT-based object detection systems," https://myfiles.neu.edu/groups/ALERT/strategic_studies/TO4_FinalReport.pdf, Retrieved 11 April 2018.

[28] D. W. Paglieroni, H. Chandrasekaran, C. Pechard, and H. E. Martz, "Consensus relaxation on materials of interest for adaptive ATR in CT images of baggage," in *Anomaly Detection and Imaging with X-Rays III*, vol. 10632. International Society for Optics and Photonics, 2018, p. 106320E.

[29] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3D classification and segmentation," in *Proc. Computer Vision and Pattern Recognition*, 2017, pp. 652–660.

[30] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Advances in Neural Information Processing Systems*, 2017, pp. 5099–5108.

[31] H. Kervadec, J. Dolz, M. Tang, E. Granger, Y. Boykov, and I. B. Ayed, "Constrained-cnn losses for weakly supervised segmentation," *Medical image analysis*, vol. 54, pp. 88–99, 2019.

[32] X. Xu and G. H. Lee, "Weakly supervised semantic point cloud segmentation: Towards 10x fewer labels," in *Proc. Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13 706–13 715.

[33] X. Li, L. Yu, H. Chen, C.-W. Fu, L. Xing, and P.-A. Heng, "Transformation-consistent self-ensembling model for semisupervised medical image segmentation," *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[34] Q. Wang and T. Breckon, "Unsupervised domain adaptation via structured prediction based selective pseudo-labeling," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 6243–6250.

[35] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[36] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017.

[37] 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.

[38] ALERT, "Automated threat recognition (ATR) initiative dataset," http://www.northeastern.edu/alert/transitioning-technology/automated-threat-recognition-atr-initiative/, 2015.

[39] OpenMMLab, "Mmsegmentation is an open source semantic segmentation toolbox based on pytorch," https://github.com/open-mmlab/mmsegmentation, Retrieved 11 August 2020.

[40] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[41] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 424–432.

[42] K. Lee, J. Zung, P. Li, V. Jain, and H. S. Seung, "Superhuman accuracy on the SNEMI3D connectomics challenge," *arXiv preprint arXiv:1706.00120*, 2017.