Engineering Applications of Artificial Intelligence 157 (2025) 111421

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Research paper

Regressor driven incremental instance segmentation for contraband items identification

Ammara Nasim [a], Mahboob Anwar [b], Muhammad Owais [c], Samet Akcay [d], Divya Velayudhan [c], Muhammad Usman Akram [a], Mohammed Ghazal [b], Naoufel Werghi [c], Taimur Hassan [b],*

[a] Department of Computer and Software Engineering, National University of Sciences and Technology (NUST), Islamabad, 44000, Pakistan
[b] Department of Electrical, Computer and Biomedical Engineering, Abu Dhabi University, Abu Dhabi, 59911, United Arab Emirates
[c] Khalifa University Center for Autonomous Robotic Systems (KUCARS), Khalifa University, Abu Dhabi 127788, United Arab Emirates
[d] Research and Development Division, Intel Corporation, United Kingdom

ARTICLE INFO

Keywords:
X-ray security imagery
Baggage threat detection
Deep learning
Incremental instance segmentation

ABSTRACT

Detecting suspicious objects contained within passenger baggage is one of the most difficult tasks, even for the security experts. To address this problem, many researchers have developed computer-aided screening methods employing deep learning models to detect the presence of contraband items using X-ray imagery. However, all of these models have limitations or bottlenecks when it comes to detecting prohibited objects that are heavily obscured, cluttered, overlapping, and well concealed within the baggage. To overcome these challenges, we present a novel instance segmentation framework that transforms the conventional semantic segmentation models to perform instance-aware segmentation, via incremental learning, to automatically detect suspicious baggage items. This framework leverages knowledge distillation, enabling the model to iteratively learn and retain multiple instances of threat items. By continuously adapting, the model improves its ability to distinguish between different instances, showcasing significant advancements in security and threat detection. To overcome the under-segmentation and over-segmentation problem of the incremental instance segmentation, we introduced the use of a lightweight regressor model that can accurately identify the overlapping instances of the suspicious objects which enables the optimal selection of the segmentation model to perform the required task. Moreover, the proposed framework has been rigorously tested on two publicly available datasets on which it achieved the mask mean average precision scores of 0.54 and 0.51, respectively, at the inference stage. Similarly, the proposed method outperformed the state-of-the-art methods by 5.88%, and 8.51% in terms of mask mean average precision scores across both datasets, respectively. In addition to this, the proposed framework provides an optimal trade-off between performance and efficiency as compared to its competitors.

1. Introduction

The rapidly escalating demands in the aviation sector exacerbate the difficulties in security inspections due to the increased risk presented by suspicious individuals concealing contraband items in their baggage (Velayudhan et al., 2022b). Consequently, global authorities have enhanced their security protocols primarily through the utilization of security X-ray scanners to identify potential contraband items concealed within the baggage. Nonetheless, the manual inspection of baggage through security X-ray scans is a subjective procedure susceptible to human error, especially during peak hours. Consequently, it is necessary to develop autonomous systems that can aid security personnel in identifying suspicious objects within baggage X-ray scans. A number of researchers have worked on developing autonomous systems for detecting prohibited items. Nonetheless, most of these methods depend on traditional detectors, which are predominantly engineered for RGB images (Lin et al., 2014). When utilized for detecting suspicious baggage items from X-ray imagery, their efficacy significantly declines due to the inherent disparities between X-ray and RGB scanners (Velayudhan et al., 2022b). To address these challenges, researchers have created detection models tailored for X-ray imagery to identify concealed contraband items through contour and edge-based mechanisms (Hassan et al., 2023; Velayudhan et al., 2022a). However,

* Corresponding author.
E-mail address: taimur.hassan@adu.ac.ae (T. Hassan).

https://doi.org/10.1016/j.engappai.2025.111421
Received 24 July 2023; Received in revised form 20 April 2025; Accepted 4 June 2025
Available online 20 June 2025

these methods also fail to accurately localize the instances of suspicious items contained within the baggage X-ray scans.

Incremental learning enables models to adapt to new information over time without losing previously acquired knowledge. Various incremental learning strategies have been explored in the literature to address the issue of catastrophic forgetting, which occurs when a pre-trained AI model is adapted to new tasks. Among these, knowledge distillation has emerged as a popular technique for preserving previously learned information while enabling the model to adapt to new classes or domains. Khan et al. (2022), Sirshar et al. (2021), Michieli and Zanuttigh (2021), Tian et al. (2019), Dhar et al. (2019) and Cermelli et al. (2020). Incremental few-shot instance segmentation (iFSIS) has recently emerged as a solution for adapting models to novel classes with limited data and minimal computational resources. The Embedding Calibration Incremental Multi-Task Feature Aggregation technique by Gao et al. (2025) and the Fine-Tune-Free Incremental Segmentation method with Novel Weight Generation by Zhang et al. (2025) both address this challenge without relying on fine-tuning for newly added classes. The framework by Gao et al. (2025) enhances the representation of novel class prototypes through embedding diversification and calibration mechanisms, but it is sensitive to feature variability and often faces background confusions, especially in a cluttered baggage environment. while the method by Zhang et al. (2025) introduces a learnable weight generator along with a piecewise similarity function and probability-based confidence calibration, achieving superior stability and accuracy as compared to Gao et al. (2025).

Very limited research focuses on acquiring instance information of the suspicious baggage items, which could assist security personnel in differentiating between prohibited and non-prohibited items (e.g., *drugs* and *knives*). For example, Hassan et al. (2022a) introduced a framework that enables the model to incrementally learn and accurately differentiate between multiple instances of the same threat item.

To address the evolving challenges in the domain of baggage threat detection, we propose a new incremental learning-based approach that iteratively updates the semantic segmentation model to recognize individual instances of contraband items in baggage X-ray scans utilizing limited training samples. Semantic segmentation models are limited toward differentiating individual instances of the suspicious objects, as evident from Fig. 1. However, this problem is effectively addressed by the proposed scheme as it dynamically modifies the conventional segmentation model to perform instance segmentation through incremental learning.

Moreover, the prime motivation for designing the proposed incremental instance segmentation method is also to overcome the limitations of existing instance segmentation methods, which needed explicit re-training and fine-tuning rounds on each dataset in order to effectively recognize different instances of suspicious baggage items from the security X-ray imagery. The proposed scheme, because of its incremental learning nature, possesses the ability to segment new and evolving types of contraband object instances from different datasets without requiring exhaustive re-training and fine-tuning routines. Through iterative training, the model leverages incremental learning to continuously acquire multiple instances of the same threat class. By incorporating knowledge distillation, it not only learns new instances but also retains prior knowledge. By stage m, the model is capable of distinguishing between m distinct instances of threat items, demonstrating its robustness and adaptability. To overcome the under-segmentation and over-segmentation problems associated with incremental instance segmentation, we introduce a lightweight regressor model that accurately identifies overlapping instances of suspicious objects. This enables the optimal selection of the segmentation model to effectively perform the required task.

1.1. Related works

Many researchers have proposed baggage screening systems, from which early works leveraged traditional machine learning schemes, while subsequent works focused on deep learning. This section presents a comprehensive overview of the main techniques in both categories while we refer the readers to Velayudhan et al. (2022b) for an in-depth literature survey.

1.1.1. Conventional methods

The pioneering conventional methods were focussed on threat object classification using Bag-of-visual words in combination with Support Vector Machine (SVM) (Baştan et al., 2011; Hassan and Hassan, 2014; Akbar et al., 2017; Kundegorski et al., 2016), K-Nearest Neighbors (K-NN), and Random Forest (Jaccard et al., 2014). Moreover, researchers have also investigated the performance of several feature descriptors, such as SIFT (Mery et al., 2016a), SURF (Baştan et al., 2011), and FAST SURF (Kundegorski et al., 2016), after coupling them with the traditional classifiers. The clustering of descriptors further improved results for better discrimination between positive and negative categories (Ahmed et al., 2022a; Turcsany et al., 2013; Masood et al., 2022). Sparse representation (Mery et al., 2016b) and fuzzy-based image classification (Liu and Wang, 2008) were also explored to classify contraband items from X-ray images. In addition to this, earlier works also employed edge (Nercessian et al., 2008) and shape context descriptors (Belongie et al., 2002; Riesenhuber and Poggio, 1999) to localize and recognize contraband objects within the baggage X-ray imagery. Single and multi-view X-ray images were also evaluated in the literature for the detection of multiple contraband objects, including *guns*, *shurikens*, *knives*, and *laptops* (Rifo et al., 2019).

Moreover, image segmentation methods based on conventional machine learning were also proposed for contraband object recognition, where sections of the candidate scans are extracted, and then distinct features are computed for recognizing suspicious items (Wang et al., 2005b). Similarly, relational graph-based (Ding et al., 2006; Wang et al., 2005a), and probabilistic (Singh and Singh, 2004) approaches have also been proposed for segmenting threat objects from the baggage X-ray scans. However, such approaches are successful only in less cluttered scenarios. Region-growing approaches in combination with SURF features and nearest neighbor distance classifier have been observed to yield better segmentation results (Velayudhan et al., 2022b).

1.1.2. Deep learning methods

Similar to conventional methods, deep learning methods for identifying contraband objects are also built using classification, detection, and segmentation strategies (Chen et al., 2024; Gaus et al., 2024; Velayudhan et al., 2024).

Classification methods primarily used transfer learning for distinguishing between normal and contraband objects from the baggage scans. In addition, datasets like GDXray (Riffo et al., 2015), and SIXray (Miao et al., 2019) were also released publicly to aid researchers in developing robust threat detection models. Similarly, Griffin et al. (2019) utilized deep features extracted from the image using Inception-v2 (Szegedy et al., 2014) model to screen threatening and non-threatening items from the baggage X-ray scans.

Some studies explored conventional object detection models. For example, Akçay et al. (2018) utilized R-FCN (Dai et al., 2016) and F-RCNN (Girshick et al., 2014) in combination with VGG-16 (Simonyan and Zisserman, 2015) for detecting contraband items. In Liang et al. (2018), single and multiple-view X-ray scans were compared using detection models such as F-RCNN (Girshick et al., 2014), R-FCN (Dai et al., 2016), and SSD (Samet Akcay, 2018). Moreover, Hassan et al. (2023) proposed a two-step object detection approach to detect contraband items from the grayscale and colored X-ray scans. In the first step, a large number of region proposals were extracted offline which

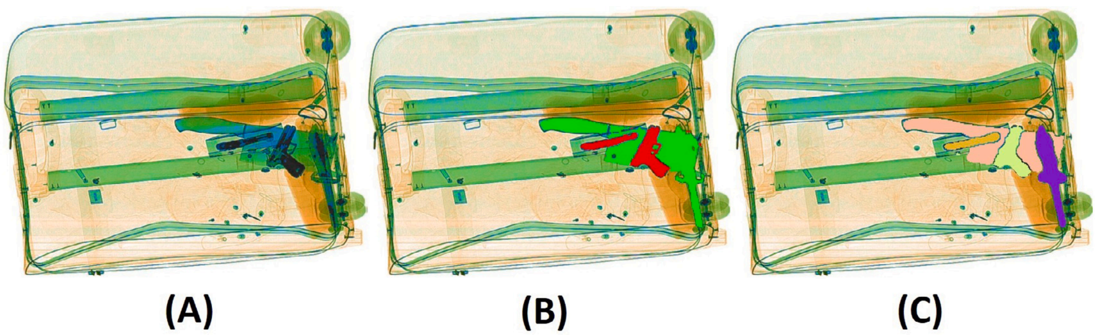

Fig. 1. An example showcasing (A) the original X-ray scan, (B) the extracted contraband objects through semantic segmentation, in which it is challenging to differentiate between multiple overlapping instances of the same threat category, and (C) overlapping instances of contraband objects extracted individually through the proposed scheme.

are then screened in the second-stage using a pre-trained backbone model. In another work, Hassan et al. (2020b) used intensity transitions to generate contours for identifying the concealed contraband objects. Similarly, Wang et al. (2021) proposed an attention-based framework to leverage the boundary and material-specific cues of the prohibited items in the baggage scans. Velayudhan et al. (2022a) coupled Broad Learning systems (Chen and Liu, 2018; Gao et al., 2020) with CNNs to detect hidden contraband items using few-shot learning. Furthermore, Ma et al. (2023) recently fused low-level and high-level features using an aggregation function and fed those features to object detection frameworks for the identification of highly concealed prohibited objects.

However, due to the limited availability of pixel-wise annotations, only a few segmentation schemes have been utilized in literature for baggage threat detection (Shafay et al., 2021b). A dual attention-based encoder–decoder architecture was employed by An et al. (2019) to classify threat objects from X-ray imagery. Chen et al. (2024) proposed a mix-paste method for prohibited item detection from the X-ray scans under noisy annotations. Gaus et al. (2024) used the Segment Anything Model (SAM) (Kirillov et al., 2023) to recognize threat objects from security X-ray scans. Velayudhan et al. (2024) presented an autonomous localization of suspicious objects using weakly supervised learning. Shafay et al. (2021a) exploited temporal information for semantic segmentation and assessed their model on the publicly accessible GDXray dataset (Riffo et al., 2015). Ahmed et al. (2022b) tackled class imbalance issues using a novel balanced affinity loss and evaluated their contour-based instance segmentation model across different datasets. In another work, Ahmed et al. (2024) developed a contour-driven classification and segmentation pipeline to recognize contraband objects from baggage X-ray scans. Bhowmik et al. (2021) extracted high discriminative features to be utilized in segmentation through high and low-energy scans. Chouai et al. (2020) proposed another semantic segmentation model based on auto-encoders for separating the contraband items within the baggage X-ray scans. Hassan et al. (2022a) proposed another instance-aware semantic segmentation scheme, which is capable of separating different suspicious baggage items and their instances via Bayesian inference.

1.2. Contributions

After a thorough literature review, we found out that there are limited instance-aware segmentation frameworks that can detect suspicious items concealed within the baggage. One of the pioneer works in this direction was done by Hassan et al. (2022a) in which they proposed an incremental learning-driven instance segmentation pipeline to detect cluttered contraband items. However, the detection results produced by their scheme were vulnerable to over-segmentation problems. For example, they used mth model instance ($\mathcal{M}_m$) to perform k-instance segmentation tasks, where $m > k$. Therefore, in cases where k value is much smaller than m, i.e., $k \lll m$, their proposed scheme produces over-segmentation results (Hassan et al., 2022a). To address these issues, we introduce the utilization of regressor R_g that optimally selects the kth model instance, dubbed $\mathcal{M}_k$, to perform k-instance segmentation, thus avoiding over-segmentation and under-segmentation issues that are typically produced during an incremental instance segmentation (Hassan et al., 2022a). Furthermore, we proposed a new L_t loss function that constrains the backbone model during incremental training to robustly detect the cluttered contraband items and their instances while avoiding catastrophic forgetting (Hassan et al., 2022a). To summarize, the main contributions of this paper are:

- We propose a regressor-driven optimal incremental instance segmentation framework capable of learning distinctive features of overlapping suspicious objects within baggage X-ray scans.
- The regressor $\mathcal{R}_g$ within the proposed framework identifies the optimal model $\mathcal{M}_k$ based on the number of overlapped instances k, reducing thus the false detections produced by $\mathcal{M}_m$ that can lead to over-segmentation and under-segmentation results (Hassan et al., 2022a).
- We also introduce the $\mathcal{L}_t$ loss function that is administered through the remembrance factor (β). $\mathcal{L}_t$ loss function allows the proposed model to resist catastrophic forgetting during incremental training and effectively retain its prior knowledge while learning new representations of suspicious baggage items and their instances.

The rest of the paper is organized as follows: Section 2 presents the details about the datasets used in this research and the proposed method. Section 3 presents the experimental setup and the results of the proposed method. Lastly, Section 4 discusses the strengths and weaknesses of the proposed method compared to the state-of-the-art (SOTA) works. Section 4 also concludes the paper and highlights the prospects of the proposed system.

2. Materials and methods

Despite their capacity to preserve categories pixel-by-pixel, semantic segmentation methods cannot differentiate overlapping instances of the same class (as illustrated in Fig. 1). To address this issue, we propose a novel regressor-driven incremental instance segmentation framework capable of learning distinctive feature representations of overlapping threat instances from highly cluttered and randomly stacked baggage scans. One of the uniqueness of the proposed system is its ability to resist catastrophic forgetting during incremental training and effectively retain its prior learned knowledge while learning new representations of the suspicious baggage items and their instances to screen them at the inference stage effectively. This remembrance ability makes the proposed system ideal to be deployed at airports for autonomous real-time baggage threat detection, especially during rush hours. Also, unlike competitive methods, the proposed system

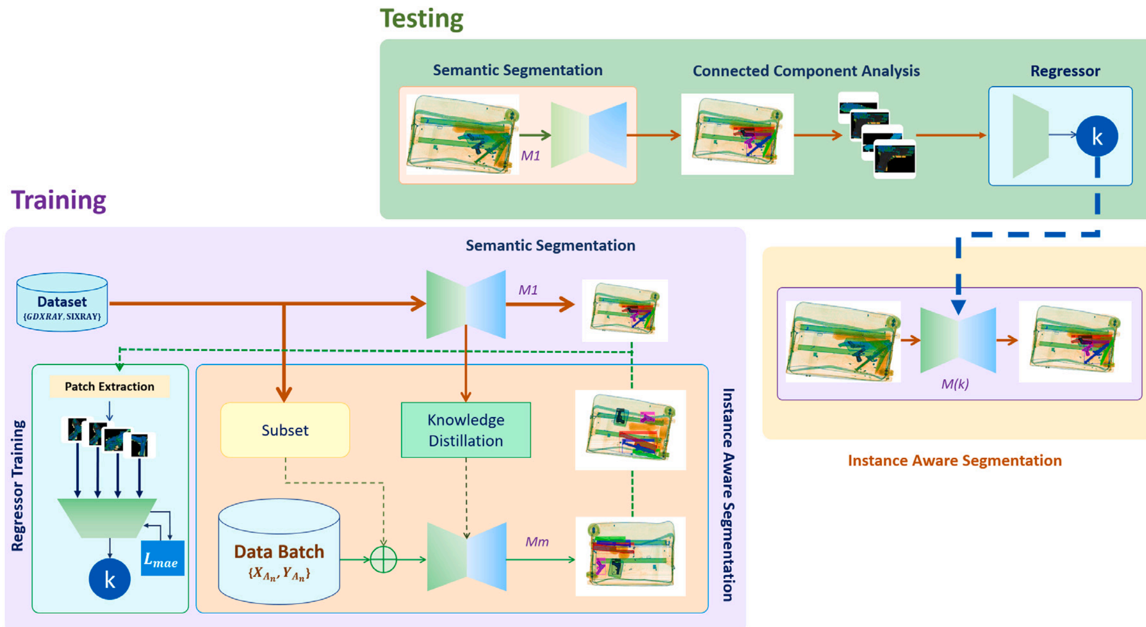

Fig. 2. Schematic diagram of the proposed framework. In the first training cycle, model $\mathcal{M}_1$ is trained to segment suspicious baggage items. Moreover, in lth cycle, the model $\mathcal{M}_l$ is able to identify up to l instances of the contraband objects. During the testing phase, model $\mathcal{M}_1$ first segments distinct objects (but not their individual instances). Afterward, a connected component analysis (CCA) is performed to identify the proposals of these contraband objects. Each proposal generates a fixed-size patch, which is then passed to $\mathcal{R}_g$, which outputs k, a number of merged item instances. The k value selects the appropriate model instance $\mathcal{M}_k$ to segment k occluded instances from the input scan.

does not require any superfluous object detector or exhaustive training overheads to perform instance segmentation (He et al., 2017; Li et al., 2017; Hariharan et al., 2015; Zagoruyko et al., 2016).

The incremental learning aspect of the proposed scheme focuses on learning the instances of overlapping threat items, where a distinct class represents each instance. Moreover, the proposed scheme employs a lightweight regressor model R_g to optimally select the kth model ($\mathcal{M}_k$) at the inference stage to perform the required task, thus avoiding over-segmentation issues as compared to its competitors. In the subsequent sections, we present the details about the datasets that are used in this research as well as a detailed description of the proposed scheme.

2.1. Datasets

The proposed system underwent training and testing on publicly available GDXray (Riffo et al., 2015) and SIXray (Miao et al., 2019) datasets. The GDXray dataset comprises five distinct data categories: *Baggage*, *Castings*, *Welds*, *Nature*, and *Settings*. However, our work is specifically focused only on the *Baggage* category, which consists of 8150 X-ray images containing three classes: *guns*, *shuriken*, and *razors*. Recognizing contraband items in these grayscale images is particularly challenging due to their texture-less nature. To expand the coverage of prohibited item detection beyond these three categories, we introduced two additional classes: *knives* and *chip* (that detect all the suspicious items like *laptops*, and *mobile phones*). Moreover, a total of 400 images were utilized for training the proposed method to recognize *guns*, *shuriken*, and *razors*, while 388 images were dedicated to training it for recognizing *chips* and *knives*. The rest of the 7632 images were allocated for testing purposes. Apart from this, SIXray (Miao et al., 2019) contains 1,059,231 colored X-ray images featuring instances of contraband items with significant occlusion and overlap. For our experiments, we utilized 80% of the SIXray images for training and kept the remaining 20% for testing (Miao et al., 2019). We also want to highlight that in order to test the generalization ability of the proposed model and compare it with the SOTA methods, we also trained it on SIXray (Miao et al., 2019) and tested it on an unseen PIDray (Zhang et al., 2023) dataset to extract contraband items, such as *scissors*, *knives*, *guns*, *wrenches*, and *pliers*. The experiments that assess the generalization ability of the proposed model and compare it with the SOTA methods are reported in Section 3.7 and Table 10.

2.2. Method overview

An overview of the proposed incremental instance segmentation framework is depicted in Fig. 2. It comprises an optimal encoder–decoder model, identified by the regressor R_g, to effectively discriminate between overlapping instances of the contraband objects. Assume m represents the maximum overlapping instances of suspicious baggage items within the respective dataset; the encoder–decoder network is trained incrementally, using m steps of incremental learning, enabling the proposed framework to learn distinctive representations of the individual threat instances and differentiate them effectively.

The training routine of the proposed scheme involves m models $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_m$. Then, at each incremental step l, ($l \leq m$), the model M_l learns to distinguish between l instances of contraband while preserving the learned knowledge of $\mathcal{M}_{l-1}$ by transferring the weights from $\mathcal{M}_{l-1}$. It is to be noted that $\mathcal{M}_l$ is similar to $\mathcal{M}_1$ except for the modifications in the final layer that enable it to distinguish up to l overlapping instances of the same threat category.

During the first iteration, model $\mathcal{M}_1$ extract suspicious baggage items. During the second iteration, the weights of $\mathcal{M}_1$ are transferred to model $\mathcal{M}_2$ that is capable of distinguishing up to two occluded instances of the same object. $\mathcal{M}_2$ is trained on a reduced batch of images that contains up to two overlapped items only. This same workflow is repeated till iteration m to train model $\mathcal{M}_m$, capable of identifying the maximum m instances of the same contraband object, using only a subset of the training data with m overlapping threat instances, while retaining the knowledge learned by $\mathcal{M}_{m-1}$. This is performed by constraining $\mathcal{M}_l$ (in lth iteration) using incremental loss function $\mathcal{L}_t$ to update its weights while making sure that the model retains its prior knowledge learned in $l-1$ iteration.

During inference, the patch extractor utilizes the segmentation mask generated by model $\mathcal{M}_1$ to extract the corresponding proposals from the input scan. These proposals are passed to $\mathcal{R}_g$ to identify the number of overlapping instances k contained within the scan, where $k \in 1, 2, \dots, m$. This value k then selects the optimal model $\mathcal{M}_k$ to perform the required task. It should also be noted that the proposed scheme utilizes only two segmentation models at the inference stage to detect contraband objects and their instances. These two segmentation models are $\mathcal{M}_1$ and $\mathcal{M}_k$.

2.3. Semantic segmentation

As detailed above, the proposed model $\mathcal{M}_1$, constitutes an encoder–decoder backbone that performs initial semantic segmentation to identify the prohibited items within the baggage imagery. The encoder head preserves the object semantics during image decomposition using a feature pyramid network (Zhao et al., 2017). Upsampling and convolution operations are utilized to reconstruct features during the decoding process. Models $\mathcal{M}_2$ to $\mathcal{M}_m$ are similar to $\mathcal{M}_1$ except for the modifications in the final layer to accommodate the different threat instances. Overall, there are 31.5M parameters in each model, of which 61k are non-trainable.

As a semantic segmentation framework, $\mathcal{M}_1$ is capable of identifying various contraband items such as a *gun* merged with a *knife*, but cannot differentiate between the same isolated or merged item within an image, such as two overlapping *guns*. During training, $\mathcal{M}_1$ is constrained via categorical cross-entropy loss function $\mathcal{L}_{ce}$, as expressed in Eq. (1):

$$\mathcal{L}_{ce} = -\frac{1}{b_{sz}} \sum_{i=0}^{B_{sz}-1} \sum_{j=0}^{c-1} s_{i,j} \log(p(l_{i,j})), \tag{1}$$

where b_{sz} is the batch size, c indicates the classes, while $s_{i,j}$ is ith pixel of the jth category and $p(l_{i,j})$ is the logit ($l_{i,j}$) probability distribution.

2.4. Incremental instance segmentation

Once the encoder–decoder model has been trained for semantic segmentation, the next step is to further train it through incremental learning to perform instance-aware segmentation, enabling it to differentiate highly cluttered similar contraband items. In order to enable the model to recognize the merged instances of the same category, new classes are added to the last layer while retaining the previous knowledge. Hence, model $\mathcal{M}_l$ (in lth iteration) learns distinctive feature representations to extract up to l instances of contraband items (i.e., l instances of a *knife*, *gun*, *wrench*, etc.) by retaining the knowledge of the previous model $\mathcal{M}_{l-1}$ through weight transfer. It should be emphasized, however, that $\mathcal{M}_l$ is trained on a reduced batch of scans that only contains up to l overlapping contraband items.

2.5. Regressor

Regressor module R_g intelligently identifies the maximum number of overlapping threats in an image patch and selects the optimal model to perform the required instance segmentation task. R_g is trained on a set of cropped sub-images containing overlapping threats to predict the number of such instances effectively. R_g consists of one input layer, four activations, two average pooling, four batch normalization layers, and one output layer.

Initially, model $\mathcal{M}_1$ performs semantic segmentation on the input scan to generate an output mask. The patch extractor utilizes the segmentation mask generated by model $\mathcal{M}_1$ to extract the corresponding proposals from the input scan. These proposals are passed to $\mathcal{R}_g$ to identify k, where $k \in 1, 2, \ldots, m$. This value k then selects the optimal model $\mathcal{M}_k$ to perform the required task.

2.6. Loss functions

Instance segmentation within the proposed scheme is performed by constraining $\mathcal{M}_1$ using $\mathcal{L}_t$ loss function. $\mathcal{L}_t$ is a combination of two loss functions, i.e., L_{pr} and L_{kld} that collectively constrain the model in order to learn new instance segmentation tasks incrementally while retaining its prior knowledge. L_{pr} allows the proposed model to retain its prior knowledge in each incremental iteration, while L_{kld} allows the model to learn new incremental instance segmentation tasks in each iteration. Additionally, L_{pr} is administered through the hyperparameter β that controls the ability of the proposed model to learn new tasks while remembering its prior knowledge during incremental training. The optimal value of β is determined through a set of ablation experiments reported in Table 3. Moreover, L_t is mathematically expressed in Eqs. (2)–(4):

$$\mathcal{L}_t = \mathcal{L}_{pr} + \mathcal{L}_{kld} \tag{2}$$

where

$$\mathcal{L}_{pr} = -\frac{\beta}{b_{sz}} \sum_{i=0}^{B_{sz}-1} \sum_{j=0}^{c-1} s^o_{i,j} \log(p(l^o_{i,j})) \tag{3}$$

and

$$\mathcal{L}_{kld} = \frac{1}{b_{sz}} \sum_{i=0}^{b_{sz}-1} \sum_{j=0}^{c-1} q(s^n_{i,j}) \log(q(s^n_{i,j})/p(l^n_{i,j})) \tag{4}$$

$\mathcal{L}_{pr}$ is the cross-entropy loss function where $p(l^o_{i,j})$ is the logit ($l^o_{i,j}$) probability obtained from the ith sample of jth class which model learns in iteration 1 to $k-1$.

β is a remembrance factor that controls the model's capability to retain its prior knowledge during incremental training. Through rigorous experimentation, we have found the most optimum value of β is 10 for effective learning in this application (see Section 3.3.2 and Table 3 for more details).

Furthermore, the proposed scheme also employs the Kullback Leibler Divergence (KLD) loss function ($\mathcal{L}_{kld}$) to learn new classes. $\mathcal{L}_{kld}$ is mathematically expressed in Eq. (4), where $p(l^n_{i,j})$ is the logit ($l^n_{i,j}$) probability, and $q(t^n_{i,j})$ is the ground truth ($t^n_{i,j}$) probability.

Apart from this, $\mathcal{R}_g$ is trained using mean absolute error loss function $\mathcal{L}_{mae}$, as expressed in Eq. (5):

$$\mathcal{L}_{mae} = \frac{1}{b_{sz}} \sum_{i=0}^{b_{sz}-1} (|y_i - y^p_i|) \tag{5}$$

where y^p_i and y_i are the predicted and true value of the ith sample. Another consideration to keep in mind here is that if there are two kinds of objects with a different number of overlapping instances (e.g., *gun* = 2, and *knife* = 3), $\mathcal{R}_g$ will look for the maximum number of suspicious baggage items' instances (i.e., 3). Thus, $\mathcal{M}_3$ will be chosen to perform up to 3-instance segmentation. Moreover, a detailed discussion on the conclusiveness of results and the strengths and weaknesses of the proposed scheme is presented in Section 4.

3. Results

This section presents a detailed evaluation of the proposed framework and its comparison with SOTA models across different public datasets. Before discussing the results, we also report the details about the training and implementation protocols, as well as the evaluation metrics.

3.1. Implementation and training process

The training process comprises two stages. Initially, models $\mathcal{M}_1$ through $\mathcal{M}_m$ undergo training, followed by the training of the regressor $\mathcal{R}_g$ in the subsequent step. Model $\mathcal{M}_1$ is trained according to the testing and training percentages of the datasets. In subsequent iterations, models $\mathcal{M}_2$ to $\mathcal{M}_m$ are tuned using only 15% of the images to perform instance segmentation tasks. $\mathcal{R}_g$ is trained on 10,564 proposals from SIXray and 143 proposals from the GDXray dataset. The proposed scheme is implemented on the Anaconda platform utilizing TensorFlow 2.2.0, executed on a Lambda Labs Tensorbook equipped with an Intel Core i7-9750H@2.6 GHz processor, 32 GB RAM, and a single NVIDIA RTX 2080 Max-Q GPU. The encoder–decoder underwent training for 10 epochs, followed by an incremental training cycle lasting five epochs,

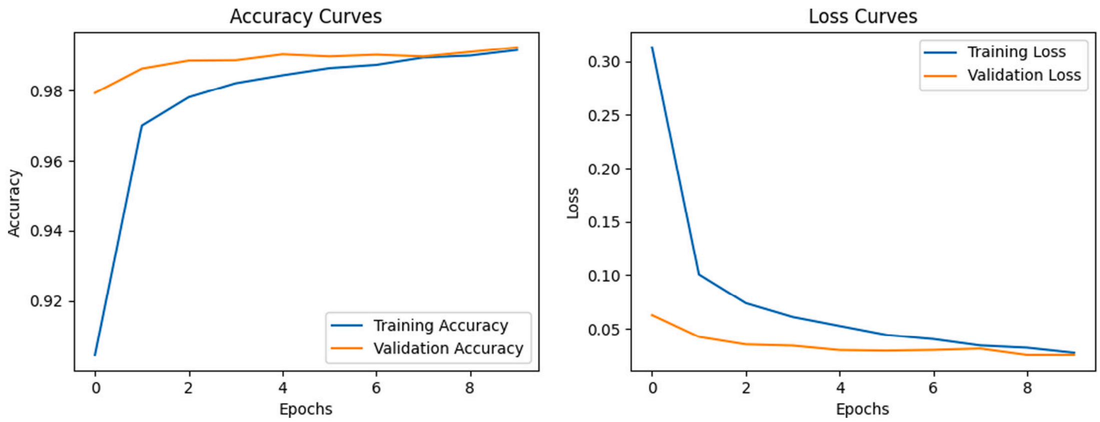

Fig. 3. Training and validation curves of $\mathcal{M}_1$ in terms of accuracy and $\mathcal{L}_{ce}$ loss function across the SIXray dataset. $\mathcal{M}_1$ achieves low model prediction error and low performance variance after training, which evidences its optimal training.

with optimizations applied in each epoch using the ADADELTA optimizer (Zeiler, 2012). The regressor $\mathcal{R}_g$ was trained for 120 epochs utilizing the ADAM optimizer (Kingma and Ba, 2014). The training and validation curves for $\mathcal{M}_1$ for SIXray (Miao et al., 2019) dataset is shown in Fig. 3. From Fig. 3, we can observe that the $\mathcal{M}_1$ achieves low model prediction error and low-performance variance across training and validation sets, which evidences its optimal training toward segmenting concealed and cluttered threat objects. Moreover, $\mathcal{M}_1$ has 31,516,059 parameters, of which 31,454,316 are trainable and 61,743 are non-trainable. The architectural details of $\mathcal{M}_1$ and the source code of the proposed framework along with demo videos are publicly released on GitHub.[1]

3.2. Evaluation metrics

The quantitative performance of the proposed scheme was evaluated using various performance metrics, following the guidelines set forth in the Microsoft COCO challenge of 2017 (Lin et al., 2014). The following five metrics were used to assess the semantic segmentation performance: (1) mAP_m: mean average precision (mAP) of mask at $50\% \leq$ IoU $\leq 95\%$ thresholds with 5% step size, (2) $\text{mAP}_\text{m}{}^{50}$: mAP of mask for IoU $\geq$ 50%, (3) $\text{mAP}_\text{m}{}^{75}$: mAP of mask for IoU $\geq$ 75%, (4) IoU_m: extracted mask quality using IoU scores, and (5) DC_m: extracted mask quality using the dice coefficient (DC). The metrics mAP_m, IoU_m and DC_m are mathematically expressed in Eqs. (6)–(11):

$$mAP_m = \sum_{i=0}^{c-1} \sum_{j=0}^{w-1} P_{i,j} . \Delta R_{i,j}, \tag{6}$$

$$P_{i,j} = \frac{TP_{i,j}}{TP_{i,j} + FP_{i,j}}, \tag{7}$$

$$R_{i,j} = \frac{TP_{i,j}}{TP_{i,j} + FN_{i,j}}, \tag{8}$$

$$\Delta R_{i,j} = R_{i,j} - R_{i,j-1}, \tag{9}$$

$$IoU_m = \frac{TP_m}{TP_m + FP_m + FN_m}, \tag{10}$$

$$DC_m = \frac{2TP_m}{2TP_m + FP_m + FN_m}, \tag{11}$$

where c denotes the number of classes, w denotes the number of predictions for the c classes, $P_{i,j}$ denotes precision calculated for jth prediction belonging to ith class of c. $R_{i,j}$ denotes recall calculated for jth prediction belonging to ith class of c. $TP_{i,j}$, $FP_{i,j}$, $FN_{i,j}$, respectively, denote true positives, false positives, and false negatives for jth prediction belonging to ith class in c. Also, TP_m, FP_m, and FN_m, respectively, denote true positives, false positives, and false negatives for the mask predictions and ground truth pairs. Apart from this, the detection performance of the proposed model was evaluated using the following three metrics: (1) mAP_b: mAP of the bounding box at $50\% \leq$ IoU $\leq 95\%$ thresholds with 5% step size, (2) $\text{mAP}_\text{b}{}^{50}$: mAP of the bounding box for IoU $\geq$ 50%, and (3) $\text{mAP}_\text{b}{}^{75}$: mAP of the bounding box for IoU $\geq$ 75%. It should be noted that mAP_m and mAP_b are calculated in the same manner. The only difference is that mAP_m is calculated from mask predictions and ground truths, whereas mAP_b is calculated from bounding box predictions and ground truths. Moreover, we also calculate root mean square error (RMSE), expressed in Eq. (12), to assess regressor $\mathcal{R}_g$ performance.

$$RMSE = \sqrt{\frac{1}{h} \sum_{i=0}^{h-1} (k_i^a - k_i)^2}, \tag{12}$$

where h denotes the total number of proposals passed to $\mathcal{R}_g$, k_i^a, and k_i represent the actual and predicted number of overlapping contraband instances contained within the ith proposal, respectively.

3.3. Ablation studies

We performed a series of ablation experiments to determine (1) optimal backbone model, (2) optimal β value, and (3) the regressor performance to detect the overlapping instances of the contraband objects.

3.3.1. Optimal backbone model

In the first ablation study, we determined the optimal backbone model that can be plugged into the proposed system to identify suspicious baggage items and their instances. For this purpose, we incrementally trained the proposed model $\mathcal{M}$ and the other popular segmentation models, such as, PSPNet (Zhao et al., 2017), UNet (Ronneberger et al., 2015), DSRL (Wang et al., 2020) and FCN-8 (Long et al., 2015). As reported in Table 1, all the models, except FCN-8, produced competitive performance, while the proposed model significantly outperformed all of these baseline models and is ranked as first. The UNet (Ronneberger et al., 2015) is ranked as the second-best performer, where the proposed model outperformed UNet (Ronneberger et al., 2015) specifically on GDXray dataset (Riffo et al., 2015) with an average improvement of 11.0%, 7.0%, and 4.0% in terms of IoU_m, DC_m, and mAP_m, respectively. Additionally, for SIXray dataset (Miao et al., 2019), our model exhibited a performance gain of 13.0% in IoU_m, 9.0% in DC_m, and 4.0% in mAP_m. The degraded performance of FCN-8 (Long et al., 2015) can be attributed to its inability to effectively incorporate high-resolution features during the up-sampling process. Since $\mathcal{M}$ had

[1] Source code: https://github.com/taimurhassan/regIncSeg.

Table 1
Performance evaluation of the proposed scheme using different backbone models across GDXray (Riffo et al., 2015) and SIXray (Miao et al., 2019). The best performance is indicated in bold, while the second-best performance is underlined.

Metric	Dataset	DSRL (Wang et al., 2020)	PSPNet (Zhao et al., 2017)	UNet (Ronneberger et al., 2015)	FCN-8 (Long et al., 2015)	Proposed
IoU_m	GDXray (Riffo et al., 2015)	0.76	0.74	<u>0.78</u>	0.62	**0.89**
IoU_m	SIXray (Miao et al., 2019)	0.66	0.65	<u>0.68</u>	0.50	**0.81**
DC_m	GDXray (Riffo et al., 2015)	0.86	0.85	<u>0.87</u>	0.76	**0.94**
DC_m	SIXray (Miao et al., 2019)	0.79	0.78	<u>0.80</u>	0.66	**0.89**
mAP_m	GDXray (Riffo et al., 2015)	0.50	0.47	<u>0.50</u>	0.43	**0.54**
mAP_m	SIXray (Miao et al., 2019)	0.46	0.44	<u>0.47</u>	0.39	**0.51**

Table 2
Ablation experiments for optimum loss function for incremental learning in terms of mAP_m for performing instance segmentation. Abbreviations used are CE: Cross Entropy, KLD: Kullback–Leibler divergence, LwM: Learning without Memorizing, CIEM: Contraband Items Extraction Model, CLR:Continual Learning Objective for analyzing complex knowledge representations.

Dataset	CE loss	KLD loss	CIEM loss (Hassan et al., 2022a)	LwM (Dhar et al., 2019)	CLR (Khan et al., 2022)	Proposed loss
GDXRAY (Riffo et al., 2015)	0.41	0.43	0.51	0.44	0.48	0.54
SIXRAY (Miao et al., 2019)	0.39	0.41	0.47	0.42	0.46	0.51

better performance compared to others toward performing the incremental baggage threat instance segmentation tasks, we chose it as a candidate backbone model in the rest of the experimentation.

3.3.2. Optimal loss function and β value selection

This ablation study is structured into two parts. In the first part, we evaluated the effectiveness of various loss functions from the literature that are commonly used for knowledge distillation segmentation tasks. This is followed by the selection of the optimal β value corresponding to the best-performing loss function. Specifically, we compared: (i) the standard cross-entropy loss, (ii) the Kullback–Leibler (KL) divergence loss, and (iii) a composite loss function combining cross-entropy, KL divergence, and mutual information, as introduced in Hassan et al. (2022a). Additionally, we evaluated the loss function proposed by Khan et al. (2022), which formulates a continual learning objective to preserve complex knowledge representations, although it was not originally designed for threat-specific instance segmentation. We also consider the approach by Dhar et al. (2019), which attempts to learn without relying on data memorization. However, this method exhibits limitations in maintaining robust segmentation performance across diverse datasets, particularly in scenarios involving evolving instance-level variations. The results of these experiments are discussed in Table 2. The results of this ablation study demonstrate that the proposed loss function outperforms other existing methods in terms of effectiveness. In second part of the experiment, we varied the β factor during training to determine its optimal value, which produces the best threat detection performance at the testing phase across both GDXray (Riffo et al., 2015) and SIXray (Miao et al., 2019) datasets. The results for this experiment are shown in Table 3 from which we can analyze that $\beta = 10$ yield the best performance of 0.54, and 0.68 in terms of mAP_m, and mAP_b, respectively across GDXray (Riffo et al., 2015) dataset. Similarly, across the SIXray (Miao et al., 2019) dataset, the proposed model also achieves the best detection performance of 0.51 and 0.63, in terms of mAP_m, and mAP_b, respectively at $\beta = 10$. Therefore, we chose the optimal value of β to be 10 in order to incrementally train the proposed model $\mathcal{M}$.

The results presented in Table 2 highlight the individual contributions of the Cross-Entropy and Kullback–Leibler Divergence loss functions, demonstrating their distinct roles in improving classification accuracy and knowledge retention. Additionally, Table 3 illustrates the effect of varying the β factor, showing how different weight distributions between these loss components influence the model's performance, particularly in balancing the trade-off between learning new instances and retaining previously learned knowledge.

Table 3
Varying β factor to analyze network segmentation and detection performance at the testing phase. Bold indicates the best performance.

β	SIXray (Miao et al., 2019)		GDXray (Riffo et al., 2015)	
	mAP_m	mAP_b	mAP_m	mAP_b
5	0.49	0.58	0.51	0.64
10	**0.51**	**0.63**	**0.54**	**0.68**
15	0.48	0.57	0.49	0.61
20	0.44	0.53	0.46	0.57

3.3.3. Regressor performance

In the last series of experiments, we evaluated the performance of the regression model R_g to accurately recognize suspicious baggage items and their instances via RMSE score. It can be observed that across both datasets, R_g performed reasonably well in recognizing the suspicious baggage items and their instances. For example, on GDXray (Riffo et al., 2015), R_g achieved the RMSE score of 0.2417 at the testing phase. Similarly, across the SIXray dataset (Miao et al., 2019), it achieved the RMSE score of 0.3152 at the testing phase. The performance difference of R_g between GDXray and SIXray is because of the fact that GDXray only contains at-most two overlapping instances of the contraband objects, whereas SIXray contains up to six suspicious baggage items and their instances within the single patch. So, accurately recognizing the instances of the threat objects from SIXray dataset is more challenging, compared to the ones on the GDXray dataset. However, despite the fact that the regressor has accurately recognized the overlapping instances of the contraband objects which significantly improved the threat detection performance of the proposed model across both datasets. We encourage its utilization in rest of the experimentation.

3.4. Assessments on GDXray

We first tested the proposed method on GDXray. It should be noted that GDXray only provides bounding box annotations; that is why to verify mask extraction performance, we used the annotations released by Hassan et al. (2022a). Moreover, we also used the original annotations to verify the bounding box identification performance of the proposed method, as reported in Table 4 in terms of mAP_m and mAP_b. Under fair comparison, the proposed method outperformed the segmentation performance of second-best contraband items extraction model (CE) (Hassan et al., 2022a) with an average improvement of 5.88% in mAP_m, 4.88% in mAP_m^{50}, and 7.27% in mAP_m^{75}. Furthermore, the proposed method also exhibited better detection performance compared to the second-best model (Hassan et al., 2022a), with an

Table 4
GDXray: Comparison of the proposed model with SOTA baggage threat detection frameworks. The best performance is in bold, while underlined is the second-best performance. Moreover, the abbreviations are: TST: Trainable Structure Tensors (Hassan et al., 2020a), CIEM: Contraband Items Extraction Model (Hassan et al., 2022a), CST: Cascaded Structure Tensors (Hassan et al., 2023), NPF-iFSIS: Novel Prototype and Function-based Incremental Few-Shot Instance Segmentation (Zhang et al., 2025), EC-iMTFA (Gao et al., 2025) and BLD: Broad Learning Detector (Velayudhan et al., 2022a).

Method	mAP_m	$mAP_m{}^{50}$	$mAP_m{}^{75}$	mAP_b	$mAP_b{}^{50}$	$mAP_b{}^{75}$
Mask R-CNN (He et al., 2017)	0.43	0.71	0.46	0.52	0.78	0.58
YOLACT (Bolya et al., 2019)	0.36	0.65	0.37	0.48	0.74	0.54
TST (Hassan et al., 2020a)	0.46	0.76	0.51	0.57	0.82	0.64
CIEM (Hassan et al., 2022a)	<u>0.51</u>	<u>0.82</u>	<u>0.55</u>	<u>0.64</u>	<u>0.88</u>	<u>0.71</u>
CST (Hassan et al., 2023)	0.49	0.79	0.53	0.61	0.85	0.68
BLD (Velayudhan et al., 2022a)	0.44	0.73	0.48	0.54	0.80	0.61
NPF-iFSIS (Zhang et al., 2025)	0.34	0.56	0.44	0.46	0.63	0.53
EC-iMTFA (Gao et al., 2025)	0.31	0.53	0.41	0.43	0.59	0.49
Proposed	**0.54**	**0.86**	**0.59**	**0.68**	**0.92**	**0.74**

Table 5
Performance evaluation of the proposed framework to extract the masks of the contraband items across GDXray (Riffo et al., 2015). Bold indicates the best performance, while underlined is the second-best performance. The abbreviations are: CIEM: Contraband Items Extraction Model (Hassan et al., 2022a), TST: Trainable Structure Tensors (Hassan et al., 2020a), and TP: Tensor Pooling (Hassan et al., 2022b).

Metric	CIEM (Hassan et al., 2022a)	TST (Hassan et al., 2020a)	Mask R-CNN (He et al., 2017)	TP (Hassan et al., 2022b)	Proposed
IoU_m	<u>0.85</u>	0.81	0.76	0.78	**0.89**
DC_m	<u>0.91</u>	0.89	0.86	0.87	**0.94**

average gain of 6.25% in mAP_b, 4.55% in $mAP_b{}^{50}$, and 4.23% in $mAP_b{}^{75}$, respectively. These significant performance improvements of the proposed model over SOTA is further substantiated through a t-test analysis. This analysis yielded a p-value of 0.0006 ($p < 0.01$) with a 99% confidence level, providing strong evidence for the superiority of the proposed model.

Also, these performance improvements can be attributed to the inherent capabilities of the proposed framework in achieving precise pixel-wise recognition of contraband objects, regardless of the presence of visual clutter. This capability of the proposed scheme is achieved through the integration of identity and pyramid pooling blocks, where the optimal model instance within the proposed scheme is determined at run-time by R_g. Also, the capacity of the proposed system to perform optimal incremental instance segmentation, via R_g, enables it to produce significantly better results as compared to Hassan et al. (2022a) (see Table 4), where Hassan et al. (2022a) has found to be vulnerable against over-segmentation and under-segmentation issues. Moreover, since the proposed model is specifically designed to learn incremental threat instance segmentation tasks, where its remembrance ability is administered effectively through the β factor, it outperformed the other SOTA models which cannot be tailored well for the incremental instance segmentation task due to their architectural constraints (Velayudhan et al., 2022a; Hassan et al., 2020a, 2023).

In another experimental analysis, we conducted an evaluation to assess the capability of the proposed framework in accurately extracting segmentation masks, utilizing the metrics IoU_m and DC_m. The results, presented in Table 5, demonstrate that the proposed method outperformed the second-best CE model (Hassan et al., 2022a) by 4.71% and 3.3% in terms of IoU_m and DC_m, respectively. To further evaluate the performance of the proposed method, a qualitative assessment was conducted on the GDXray dataset, and the results are depicted in Fig. 4. These findings highlight the potential of the proposed framework in accurately extracting occluded contraband items from X-ray images. Notably, Fig. 4(I) showcases the remarkable accuracy of the proposed method in detecting the *gun* object, even when it is completely concealed by the presence of the *laptop*. This demonstrates the ability of our method to handle challenging scenarios where contraband items are hidden within complex objects. Furthermore, Fig. 4(H), (I), (J), and (R) illustrate the effectiveness of our method in extracting extremely occluded contraband objects from grayscale X-ray scans within the GDXray dataset (Riffo et al., 2015). These results emphasize the robustness and efficacy of the proposed framework in detecting concealed contraband items, even in challenging imaging conditions.

3.5. Assessments on SIXray

Table 6 reports the results of the proposed method on SIXray (Miao et al., 2019). From Table 6, we can observe that the proposed method is leading the second-best model (Hassan et al., 2022a) by 8.51% in mAP_m, 9.86% in $mAP_m{}^{50}$, and 8.0% in $mAP_m{}^{75}$. Similarly, the detection performance of the proposed framework is also higher than the second-best model (Hassan et al., 2022a) by 14.55%, 7.89%, and 13.79% in terms of mAP_b, $mAP_b{}^{50}$, and $mAP_b{}^{75}$, respectively. A t-test analysis further confirms the superiority of the proposed model compared to the second-best model (Hassan et al., 2022a) by yielding a highly significant p-value of 0.0215 ($p < 0.05$, 95% confidence level). These performance improvements again stem from the in-built architecture of the proposed model that accurately extract the contraband items because of the inclusion of the identity and the pyramid pooling blocks. Furthermore, it is worth noting that the R_g enables the proposed framework to optimally extract the cluttered instances of the contraband objects, where such instances are incrementally learned by the proposed scheme using the $\mathcal{L}_t$ loss function. The $\mathcal{L}_t$ loss function, within the proposed scheme, prevents the backbone model from catastrophically forgetting the prior learned knowledge (while learning new representations of contraband items' instances), and the rate at which the model is constrained to either retain its prior knowledge or learn new representations is controlled through the β parameter.

In a series of additional experiments, we assessed the performance of the proposed framework with SOTA methods on various subsets of the SIXray dataset: SIXray10, SIXray100, and SIXray1000 (Miao et al., 2019). SIXray10 comprises all 8929 positive samples from the original SIXray dataset, accompanied by 10 times the number of negative samples. SIXray100 includes all the positive samples from SIXray, along with 100 times the number of negative samples. SIXray1000 consists of only 1000 positive samples, combined with the entire set of 1,050,302 negative samples from the SIXray dataset (Miao et al., 2019). Here, SIXray1000 is the subset that contains an extremely imbalanced ratio of normal and suspicious baggage items. From Table 7, we can observe that across all subsets, the proposed framework is leading the second-best CE model (Hassan et al., 2022a) by 2.25%, 3.53%, and 7.79% in terms of $mAP_b{}^{50}$. The margin of improvement here is significant on the extremely imbalanced subset, i.e., SIXray1000. This is because of the fact that the proposed framework recognizes the cluttered contraband items through pixel-wise recognition unlike SOTA methods (Miao et al., 2019; Velayudhan et al., 2022a). Furthermore, R_g, due to its robust

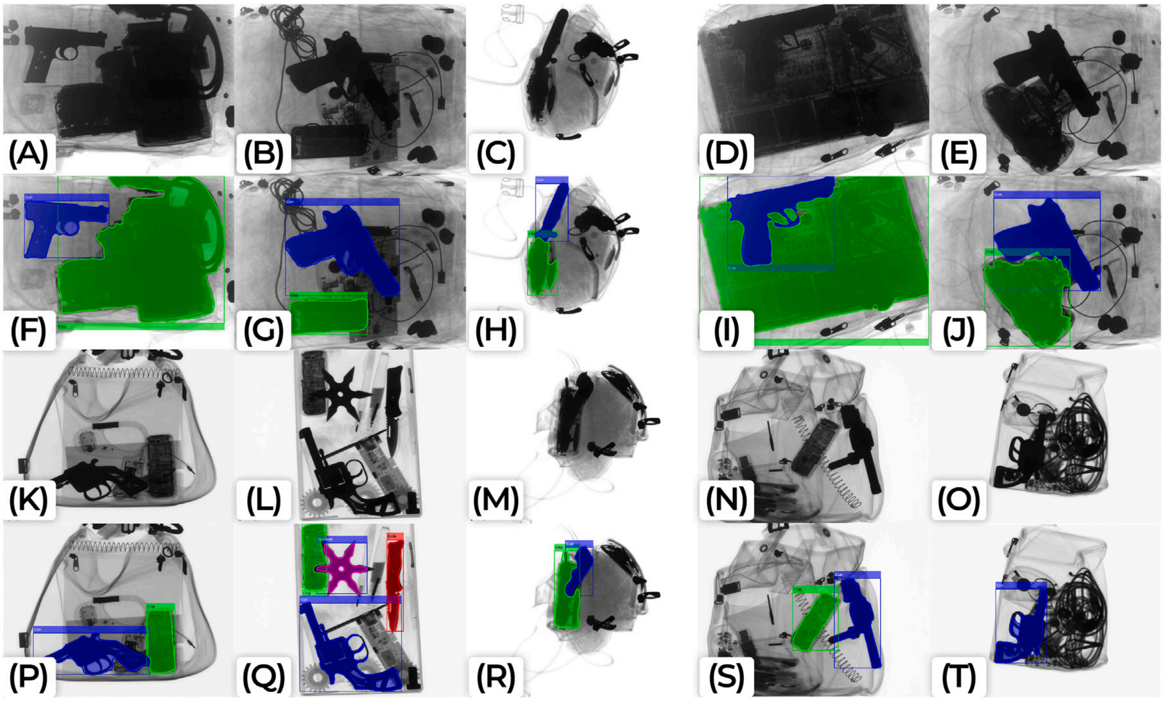

Fig. 4. The qualitative assessment of the proposed method on GDXray. The results highlight the significance of the proposed framework in accurately extracting occluded prohibited items from X-ray images. Please zoom-in for the best visualization.

Table 6
SIXray: Comparison of the proposed model with SOTA baggage threat detection frameworks. The best performance is in bold, while underlined is the second-best performance. The abbreviations are: MGAU: Model by Gaus et al. (2019), CHR: Class-Balanced Hierarchical Refinement (Miao et al., 2019), TST: Trainable Structure Tensors (Hassan et al., 2020a), CIEM: Contraband Items Extraction Model (Hassan et al., 2022a), CST: Cascaded Structure Tensors (Hassan et al., 2023), BLD: Broad Learning Detector (Velayudhan et al., 2022a), NPF-iFSIS: Novel Prototype and Function-based Incremental Few-Shot Instance Segmentation (Zhang et al., 2025) and EC-iMTFA:Embedding Calibration Incremental Multi-Task Feature Aggregation (Gao et al., 2025).

Method	mAP_m	$mAP_m{}^{50}$	$mAP_m{}^{75}$	mAP_b	$mAP_b{}^{50}$	$mAP_b{}^{75}$
Mask R-CNN (He et al., 2017)	0.34	0.59	0.36	0.40	0.63	0.42
YOLACT (Bolya et al., 2019)	0.31	0.56	0.33	0.36	0.62	0.38
MGAU (Gaus et al., 2019)	0.35	0.60	0.38	0.42	0.66	0.45
CHR (Miao et al., 2019)	0.27	0.52	0.30	0.31	0.59	0.34
TST (Hassan et al., 2020a)	0.38	0.61	0.42	0.45	0.70	0.48
CIEM (Hassan et al., 2022a)	0.47	0.71	0.50	0.55	0.76	0.58
CST (Hassan et al., 2023)	0.40	0.64	0.45	0.48	0.73	0.52
BLD (Velayudhan et al., 2022a)	0.37	0.59	0.41	0.42	0.67	0.44
NPF-iFSIS (Zhang et al., 2025)	0.32	0.52	0.39	0.40	0.58	0.47
EC-iMTFA (Gao et al., 2025)	0.29	0.49	0.37	0.39	0.56	0.44
Proposed	**0.51**	**0.78**	**0.54**	**0.63**	**0.82**	**0.66**

Table 7
Performance evaluation of the proposed model with SOTA methods across different subsets of SIXray dataset in terms of $mAP_b{}^{50}$. The best performance is in bold, while underlined is the second-best performance. For a fair comparison, all the methods are trained under the same experimental settings. The abbreviations are: TP: Tensor Pooling (Hassan et al., 2022b), CIEM: Contraband Items Extraction (Hassan et al., 2022a), BLD: Broad Learning Detector (Velayudhan et al., 2022a), CHR: Class-Balanced Hierarchical Refinement (Miao et al., 2019), and TST: Trainable Structure Tensors (Hassan et al., 2020a).

Subset	TP (Hassan et al., 2022b)	CIEM (Hassan et al., 2022a)	BLD (Velayudhan et al., 2022a)	CHR (Miao et al., 2019)	TST (Hassan et al., 2020a)	Proposed
SIXray10	0.87	0.89	0.84	0.77	0.81	**0.91**
SIXray100	0.82	0.85	0.78	0.63	0.74	**0.88**
SIXray1000	0.73	0.77	0.72	0.56	0.68	**0.83**

training on object patches, allows the proposed framework to accurately pick the optimal model instance in order to extract the cluttered contraband objects, regardless of the level of class imbalance. Thus, the inclusion of R_g within the proposed scheme significantly boosts its performance over other methods (Hassan et al., 2022a, 2020a, 2022b).

Table 8 presents the performance of the proposed framework in accurately extracting threat items mask, using IoU_m and DC_m. As shown in Table 8, the proposed framework outperforms the second-best model (Hassan et al., 2022a) by 6.58% and 3.49% in terms of IoU_m and DC_m, respectively. These improvements highlight the superiority of the proposed framework in effectively localizing cluttered threat objects within baggage X-ray scans compared to SOTA approaches. Additionally, Fig. 5 provides a visual representation of the qualitative results obtained by the proposed framework. The figure showcases the successful extraction of cluttered instances of contraband items, even in challenging scenarios. Specifically, in Fig. 5(F), (H), (P), (S), and (T), the proposed framework accurately identify and extract the cluttered instances of contraband items. The highlighted cases demonstrate the effectiveness of the proposed framework in handling complex scenarios where contraband items may be obscured by overlapping objects or

Table 8
Performance evaluation of the proposed method to extract the masks of the contraband items across the SIXray (Miao et al., 2019) dataset. The best performance is in bold, while underlined is the second-best performance. The abbreviations are: CIEM: Contraband Items Extraction Model (Hassan et al., 2022a), TST: Trainable Structure Tensors (Hassan et al., 2020a), and TP: Tensor Pooling (Hassan et al., 2022b).

Metric	TST (Hassan et al., 2020a)	CIEM (Hassan et al., 2022a)	Mask R-CNN (He et al., 2017)	TP (Hassan et al., 2022b)	Proposed
IoU_m	0.68	<u>0.76</u>	0.65	0.60	**0.81**
DC_m	0.80	<u>0.86</u>	0.78	0.75	**0.89**

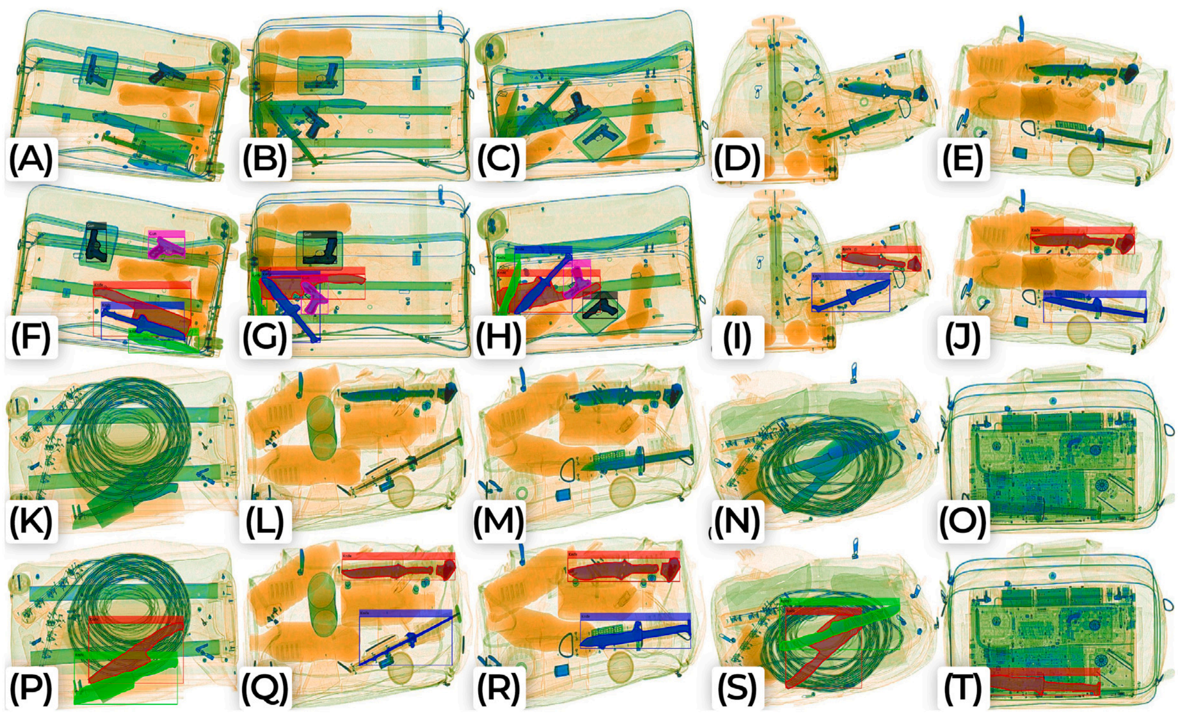

Fig. 5. The qualitative assessment of the proposed method on the SIXray dataset. The results highlight the significance of the proposed framework in accurately extracting the occluded and overlapping contraband items from the X-ray images. Please zoom-in for the best visualization.

exhibit irregular shapes. Despite these challenges, the proposed framework demonstrates its ability to precisely delineate and extract the threat item instances, ensuring reliable threat detection and localization. These qualitative results further support the robust performance of the proposed framework in accurately identifying and extracting cluttered instances of contraband items, highlighting its potential for enhancing security protection toward detecting baggage threats at the airports.

3.6. Comparison with SOTA incremental learning frameworks

To enable a fair comparison across different incremental learning strategies, we conducted instance segmentation using several established methods, including those proposed in Gao et al. (2025), Zhang et al. (2025), and Hassan et al. (2022a). It is worth noting that Gao et al. (2025) and Zhang et al. (2025) uses few shot incremental instance segmentation that are primarily designed for scenarios with scarce annotations and struggles with accuracy in cluttered environments. In contrast, Hassan et al. (2022a) presents an incremental approach tailored for instance segmentation of threat items. The results of this comparison are documented in Table 9. From the results presented in Table 9, it is evident that the proposed model consistently outperforms other incremental learning approaches. This performance gain can be attributed to the model's architecture, which is specifically tailored for incremental threat instance segmentation. Its capacity to retain previously acquired knowledge is effectively managed through the β factor, enabling more stable learning over time. In contrast, the competing state-of-the-art methods exhibit limited adaptability toward cluttered baggage imagery and face background confusion issues.

3.7. Assessments on unseen PIDray dataset

To validate the generalization ability of the proposed method, we have trained it on SIXray (Miao et al., 2019) and tested it on a new unseen dataset, i.e., PIDray (Zhang et al., 2023), to recognize threat items such as *guns*, *knives*, *scissors*, *wrenches*, and *pliers*. The generalization performance of the proposed method and its comparison with the SOTA techniques on PIDray (the unseen dataset) is reported in Table 10. To ensure fairness in the comparison between all models, we trained them on SIXray (Miao et al., 2019) and tested on PIDray (Zhang et al., 2023) to extract the same set of contraband items. From Table 10, we can observe that the proposed model was able to produce decent performance toward extracting these contraband items as compared to the SOTA methods. Such promising results are achieved by the proposed model because it was incrementally trained to extract the same threat object categories from the SIXray (Miao et al., 2019) dataset, where the optimal balance between learning new item instances and retaining the model's prior knowledge is achieved through the proposed L_t loss function. Such incremental training allowed the proposed model to generate discriminative features for all the contraband items' classes, ultimately leading to robust segmentation performance irrespective of the scanner characteristics. Apart from this, the performance of the proposed system in terms of IoU_m and DC_m across GDXray (Riffo et al., 2015), SIXray (Miao et al., 2019), and the unseen PIDray (Zhang et al., 2023) dataset is shown in Fig. 6. From Fig. 6, we can observe a slight decrease in the segmentation performance of the proposed system on PIDray (Zhang et al., 2023) dataset compared to its performance on GDXray (Riffo et al., 2015) and SIXray (Miao et al., 2019) datasets in terms of IoU_m and DC_m. The reason for this performance decrease is

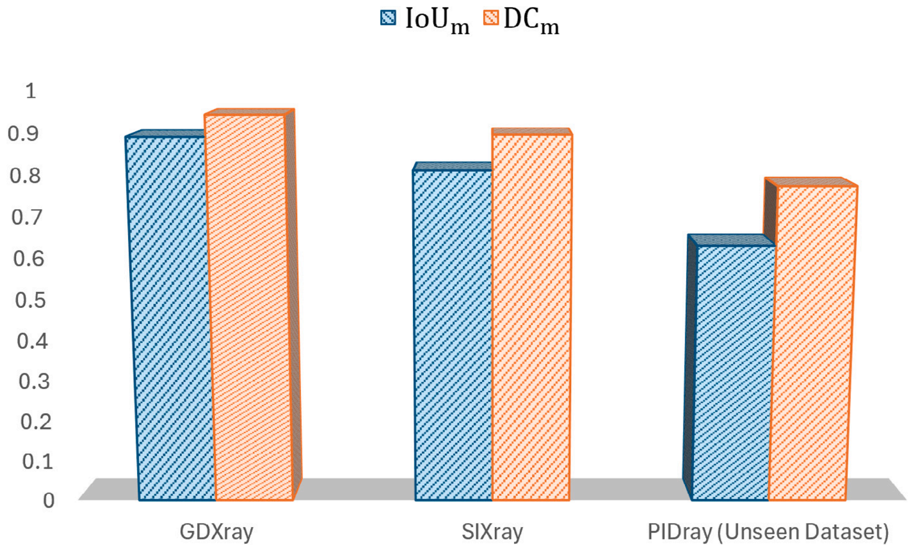

Fig. 6. Bar plot showing the segmentation performance of the proposed system in terms of IoU_m and DC_m across GDXray (Riffo et al., 2015) dataset, SIXray (Miao et al., 2019) dataset, and the unseen PIDray (Zhang et al., 2023) dataset.

Table 9
Comparison with SOTA incremental learning in terms of mAP_m and mAP_b for performing instance segmentation abbreviations used are NPF-iFSIS: Novel Prototype and Function-based Incremental Few-Shot Instance Segmentation, EC-iMTFA : Embedding Calibration Incremental Multi-Task Feature Aggregation and CIEM: Contraband Items Extraction Model.

Dataset	Parameter	NPF-iFSIS (Zhang et al., 2025)	EC-iMTFA (Gao et al., 2025)	CIEM (Hassan et al., 2022a)	Proposed
SIXray (Miao et al., 2019)	mAP_m	0.32	0.30	0.47	0.51
SIXray (Miao et al., 2019)	mAP_b	0.40	0.38	0.55	0.63
GDXray (Riffo et al., 2015)	mAP_m	0.39	0.36	0.51	0.54
GDXray (Riffo et al., 2015)	mAP_b	0.47	0.49	0.64	0.68

Table 10
Generalization comparison of the proposed model with SOTA methods on unseen PIDray dataset (Zhang et al., 2023). To ensure fairness in the comparison, all the models were trained on SIXray and tested on PIDray. The best performance is in bold, while underlined is the second-best performance. The abbreviations are: CIEM: Contraband Items Extraction Model (Hassan et al., 2022a), CDSM: Contour-Driven Segmentation Model (Ahmed et al., 2024), and BLD: Broad Learning Detector (Velayudhan et al., 2022a).

Method	mAP_m	mAP_m^{50}	mAP_m^{75}	mAP_b	mAP_b^{50}	mAP_b^{75}
Mask R-CNN (He et al., 2017)	0.12	0.35	0.17	0.19	0.41	0.22
YOLACT (Bolya et al., 2019)	0.11	0.28	0.14	0.20	0.33	0.24
CDSM (Ahmed et al., 2024)	<u>0.26</u>	<u>0.49</u>	<u>0.31</u>	<u>0.32</u>	<u>0.54</u>	<u>0.37</u>
BLD (Velayudhan et al., 2022a)	0.17	0.31	0.22	0.19	0.49	0.27
CIEM (Hassan et al., 2022a)	0.15	0.45	0.23	0.22	0.48	0.29
Proposed	**0.29**	**0.57**	**0.36**	**0.38**	**0.63**	**0.45**

because the proposed model was directly tested on the unseen PIDray dataset (Zhang et al., 2023) once it was trained on SIXray (Miao et al., 2019), where it is not trained or fine-tuned on PIDray (Zhang et al., 2023) dataset at all. Nevertheless, it is worth noting that the proposed model still manages to produce better results than SOTA methods on the PIDray (Zhang et al., 2023) dataset under fair comparison (see Table 10 for more details).

3.8. Run-time comparison

In addition to assessing the effectiveness of the proposed method in detecting prohibited items, we also conducted an evaluation of its run-time performance, specifically in terms of frames per second (FPS). To gain a comprehensive understanding of its run-time performance, we compared it with existing SOTA schemes. The results are presented in Table 11. We can observe here that the proposed framework lags behind the CE (Hassan et al., 2022a) model (the best model in terms of computation) by 6.31% in terms of computational efficiency, considering the average results obtained from both datasets. While there is a difference in the run-time performance between the proposed framework and the SOTA models, it is important to note that this difference is not particularly significant when compared to the substantial performance improvement achieved by the proposed framework, as evident from Tables 4 and 6. Therefore, it is worth acknowledging the trade-off presented by the proposed framework, which offers a balance between detection performance and computational efficiency in extracting cluttered threat object instances. Although there is room for enhancing the run-time performance of the proposed framework, we recognize it as one of the future prospects for further improvement. More detailed information regarding this aspect is provided in the subsequent section (Section 3.9).

3.9. Limitations and future works

The two limitations of the proposed scheme are its run-time performance, and the requirement to use at least two segmentation models to perform the threat detection tasks, where the first model performs the conventional semantic segmentation, and the second model performs the k-instance segmentation. The run-time performance of the proposed scheme can be remedied by either employing lightweight segmentation models or by reducing the number of convolutional blocks within the proposed network. Such alternation will affect the recognition performance of the proposed framework, but it will also improve its run-time performance. In order to examine the trade-off between

Table 11
Run-time performance comparison of the proposed scheme and the SOTA methods on both GDXray and SIXray datasets in terms of frames per second (FPS). Bold indicates the best performance. The abbreviations are: CIEM: Contraband Items Extraction Model (Hassan et al., 2022a), TST: Trainable Structure Tensors (Hassan et al., 2020a), and TP: Tensor Pooling (Hassan et al., 2022b).

Dataset	Mask R-CNN (He et al., 2017)	CIEM (Hassan et al., 2022a)	TP (Hassan et al., 2022b)	TST (Hassan et al., 2020a)	Proposed
GDXray (Riffo et al., 2015)	5.8	**7.1**	6.2	6.9	6.8
SIXray (Miao et al., 2019)	5.3	**6.6**	5.8	6.4	6.1

Table 12
Performance evaluation of the proposed scheme when different encoder backbones are employed within the proposed model. Bold indicates the best performance. The abbreviations are: FPS: Frame Per Second, No. of Params: Number of Parameters, PLWE: Proposed Lightweight Encoder, MNv2: MobileNetv2 (Sandler et al., 2018), ViT: Vision Transformer (Dosovitskiy et al., 2021), and ENetB4: EfficientNet-B4 (Tan and Le, 2019).

Dataset	Encoder	mAP_m	FPS	No. of Params
GDXray (Riffo et al., 2015)	Proposed	**0.54**	6.8	22.8M
	PLWE	0.50	7.2	10.4M
	MNv2	0.48	7.4	**5.6M**
	ViT	0.52	**5.7**	25.4M
	ENetB4	0.49	6.9	9.7M
SIXray (Miao et al., 2019)	Proposed	**0.51**	6.1	40.2M
	PLWE	0.47	6.8	22.8M
	MNv2	0.42	6.9	**11.3M**
	ViT	0.50	**4.8**	42.6M
	ENetB4	0.45	6.1	17.2M

Table 13
Performance comparison of R_g for selecting the kth model instance in order to perform k-instance segmentation tasks. The performance of R_g is measured through RMSE scores, and the overall performance of the proposed scheme is measured through mAP_m and mAP_b across each dataset. The abbreviations are: ROWS: R_g Operate on Whole Scan, and ROIP: R_g Operate on Instance Patches.

Dataset	R_g Variant	RMSE	mAP_m	mAP_b
GDXray	ROWS	0.4784	0.4586	0.6013
SIXray	ROWS	0.5931	0.4214	0.5527
GDXray	ROIP	0.2417	0.5482	0.6802
SIXray	ROIP	0.3152	0.5173	0.6354

recognition performance and computational efficiency, we conducted a comprehensive analysis by integrating various encoder backbones into the proposed model (check Table 12). It can be observed in Table 12 that utilizing a vision transformer as the backbone model yields optimal computational efficiency compared to other models. However, this advantage comes at the expense of a decrease in recognition performance when compared to the proposed encoder model. Apart from this, from Table 12, we can observe that the proposed framework, when coupled with MobileNetv2 (Sandler et al., 2018) provides the optimal model complexity in terms of number of parameters, i.e., 5.6M and 11.3M across GDXray (Riffo et al., 2015), and SIXray (Miao et al., 2019), respectively. Nevertheless, using a proposed encoder model provides the optimal trade-off between model complexity, recognition performance, and run-time efficiency. Therefore, we used the proposed encoder model as a backbone within the proposed framework to conduct all the experiments.

The second limitation of the proposed framework can be effectively addressed by training the R_g directly on the input scans, rather than focusing solely on instance patches. This modification allows R_g to identify the number of overlapping instances (k) without the need for a separate semantic segmentation model. Consequently, this alteration reduces the overall complexity of the proposed pipeline. However, it is important to note that predicting k directly from the input scans, instead of instance patches, compromises the performance of R_g. In Table 13, we present a quantitative comparison between these two scenarios using whole input scans versus instance patches. The results clearly demonstrate that the use of whole input scans leads to an increase in over-segmentation and under-segmentation errors, thus reducing significantly the detection performance when compared to the performance of the proposed scheme on instance patches. Nonetheless, the proposed scheme does provide flexibility to achieve an optimal trade-off between model complexity and recognition performance, which can be tweaked by the user to meet specific application requirements. Moreover, in the future, we aim to enhance the proposed framework's ability to identify newer threat categories through incremental learning. Currently, the model employs knowledge distillation to differentiate between various instances of the same threat class. By refining the loss function and making some modifications, we expect the proposed framework to learn new classes incrementally with each training increment.

4. Discussion and conclusion

This paper presents a novel regressor-driven incremental instance segmentation scheme that utilizes a single encoder–decoder network along with a regressor to screen concealed and cluttered instances of the threat items using security X-ray scans. The proposed scheme automates the recognition of concealed and cluttered contraband objects in baggage X-ray scans, and it was rigorously evaluated on two popular X-ray datasets, i.e., GDXray (Riffo et al., 2015) and SIXray (Miao et al., 2019), where across both datasets, the proposed scheme achieved an optimal segmentation performance as compared to SOTA methods. For example, across GDXray (Riffo et al., 2015), the proposed scheme outperformed the second-best model by 5.88% in terms of mAP_m. Similarly, across the SIXray dataset, the proposed scheme achieves 8.51% performance improvements compared to the second-best model. It should be noted here that most of the methods in the literature are based on detection instead of segmentation for screening prohibited items using X-ray scans. However, our choice of a segmentation approach is driven by the specific challenges posed by overlapping and cluttered threat items in baggage screening. In many cases, threat items are partially hidden behind other objects, making it difficult for traditional detection methods to identify and localize the threats accurately (Miao et al., 2019; Velayudhan et al., 2022a). Detection models often rely on bounding boxes, which can include non-threat items when objects overlap, leading to false positives and compromised detection performance. Segmentation models, on the other hand, allow for pixel-level recognition, making it more suitable for accurately delineating threat items, even when they exist in a cluttered environment. Apart from this, to analyze the generalization performance of the proposed scheme and compare it with SOTA methods, we trained the proposed model on SIXray (Miao et al., 2019) dataset and tested it on an unseen PIDray (Zhang et al., 2023) dataset. The results for these experiments are reported in Table 10 from which we can analyze that the proposed scheme produces better results than SOTA methods toward extracting contraband objects, such as *knives*, *scissors*, *pliers*, *guns*, and *wrenches*. These performance improvements are produced due to the inherent capabilities of the proposed scheme in achieving accurate pixel-wise recognition of contraband objects, regardless of the presence of concealment and clutter of the baggage items. Similarly, to overcome under-segmentation and over-segmentation challenges, we propose using a lightweight regressor model that accurately identifies overlapping instances of contraband objects. This enables the selection of the optimal model instance for precise instance-aware segmentation. Although, many researchers have designed segmentation models to recognize contraband objects from the security X-ray scans (Chen et al., 2024; Gaus et al., 2024; Velayudhan et al., 2024; Ahmed et al., 2024; Hassan et al., 2020a, 2022b,a). But all of these models require (1) exhaustive re-training or fine-tuning rounds to recognize evolving types of suspicious objects, which hinders their deployment in the real-world, and (2) they are limited toward incrementally adapting themselves to distinguish between different instances of the contraband objects while retaining their prior knowledge (Chen et al., 2024; Gaus et al., 2024; Velayudhan et al., 2024; Ahmed et al., 2024; Hassan et al., 2020a, 2022b,a). The proposed scheme eliminates all of these drawbacks and allows the robust recognition of contraband objects and their instances from the security X-ray scans while providing the benefit of incrementally learning new types of contraband object categories. However, there are two limitations of the proposed scheme related to its run-time performance and its need to use at least two segmentation models to recognize contraband objects, where the first model performs the conventional semantic segmentation and the second model performs the k-instance segmentation at the inference stage. The run-time performance of the proposed scheme can be remedied by either employing lightweight segmentation models or by reducing the number of convolutional blocks within the proposed model. Moreover, the second limitation of the proposed scheme can be effectively addressed by training the R_g directly on the input scans rather than focusing solely on instance patches. This modification allows R_g to identify the number of overlapping instances (k) without the need for a separate semantic segmentation model. Nevertheless, considering these limitations, the proposed scheme still provides an optimal trade-off between computational complexity and segmentation performance compared to SOTA works for recognizing concealed and cluttered instances of the contraband objects (see Tables 4, 6, 11). Furthermore, the incremental aspect of the proposed scheme makes it an ideal choice to be deployed in the real world to screen contraband objects, as it possesses the capability to learn new types of suspicious objects and their overlapping instances without requiring extensive ground truth supervision or exhaustive training rounds. In the future, we envisage testing the proposed framework's ability to detect the concealed and cluttered 3D contraband objects that showcase limited visual cues in the security X-ray scans (Velayudhan et al., 2024; Ahmed et al., 2024). Additionally, we will test the proposed scheme to detect dismantled and organic contraband objects. We also plan to validate the proposed scheme for different applications in order to perform incremental instance segmentation tasks.

CRediT authorship contribution statement

Ammara Nasim: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation. **Mahboob Anwar:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation. **Muhammad Owais:** Validation, Software. **Samet Akcay:** Writing – review & editing, Visualization, Conceptualization. **Divya Velayudhan:** Writing – original draft, Validation. **Muhammad Usman Akram:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Mohammed Ghazal:** Writing – review & editing, Supervision, Resources, Formal analysis. **Naoufel Werghi:** Writing – review & editing, Supervision, Project administration, Investigation, Formal analysis, Conceptualization. **Taimur Hassan:** Writing – review & editing, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is supported by funds from the Abu Dhabi University, UAE, Ref: 19300910. We also like to thank Anum Khan for her valuable suggestions related to this work.

Data availability

All the data used in this research is publicly available.

References

Ahmed, R., Chen, Y., Hassan, B., Du, L., Hassan, T., Dias, J., 2022a. Hybrid machine-learning-based spectrum sensing and allocation with adaptive congestion-aware modeling in CR-assisted IoV networks. In: IEEE Internet of Things Journal.

Ahmed, A., Obeid, A., Velayudhan, D., Hassan, T., Damiani, E., Werghi, N., 2022b. Balanced affinity loss for highly imbalanced baggage threat contour-driven instance segmentation. In: IEEE International Conference on Image Processing. ICIP.

Ahmed, A., Velayudhan, D., Hassan, T., Bennamoun, M., Damiani, E., Werghi, N., 2024. Enhancing security in X-ray baggage scans: A contour-driven learning approach for abnormality classification and instance segmentation. In: Engineering Applications of Artificial Intelligence.

Akbar, S., Hassan, T., Akram, M.U., Yasin, U.U., Basit, I., 2017. AVRDB: annotated dataset for vessel segmentation and calculation of arteriovenous ratio. In: Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV).

Akçay, S., Kundegorski, M.E., Willcocks, C.G., Breckon, T.P., 2018. Using deep convolutional neural network architectures for object classification and detection within X-Ray baggage security imagery. IEEE Trans. Inf. Forensics Secur. 13 (9), 2203–2215.

An, J., Zhang, H., Zhu, Y., Yang, J., 2019. Semantic segmentation for prohibited items in baggage inspection. In: International Conference on Intelligence Science and Big Data Engineering. Visual Data Engineering. pp. 495–505.

Baştan, M., Yousefi, M.R., Breuel, T.M., 2011. Visual words on baggage X-ray images. In: International Conference on Computer Analysis of Images and Patterns. pp. 360–368.

Belongie, S., Malik, J., Puzicha, J., 2002. Shape matching and object recognition using shape contexts. In: IEEE Transactions on Pattern Analysis and Machine Intelligence.

Bhowmik, N., Gaus, Y.F.A., Breckon, T.P., 2021. On the impact of using X-ray energy response imagery for object detection via convolutional neural networks. In: IEEE International Conference on Image Processing. ICIP.

Bolya, D., Zhou, C., Xiao, F., Lee, Y.J., 2019. YOLACT: Real-time instance segmentation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 9157–9166.

Cermelli, F., Mancini, M., Bulo, S.R., Ricci, E., Caputo, B., 2020. Modeling the background for incremental learning in semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, IEEE Computer Society, pp. 9233–9242. http://dx.doi.org/10.1109/CVPR42600.2020.00925.

Chen, C.L.P., Liu, Z., 2018. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. In: IEEE Transactions on Neural Networks and Learning Systems. pp. 10–24.

Chen, R., Yan, Y., Xue, J.-H., Lu, Y., Wang, H., 2024. Augmentation Matters: A Mix-Paste Method for X-Ray Prohibited Item Detection under Noisy Annotations. In: IEEE Transactions on Information Forensics and Security, December 2024.

Chouai, M., Merah, M., Mimi, M., 2020. Deep adversarial autoencoders for semantic segmentation in X-ray images of cabin baggage screening at airports. Journal of transportation security. In: Journal of Transportation Security.

Dai, J., Li, Y., He, K., Sun, J., 2016. R-FCN: Object detection via region-based fully convolutional networks. In: Advances in Neural Information Processing Systems. pp. 379–387.

Dhar, P., Singh, R.V., Peng, K.-C., Wu, Z., Chellappa, R., 2019. Learning without memorizing. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, IEEE.

Ding, J., Li, Y., Xu, X., Wang, L., 2006. X-ray image segmentation by attribute relational graph matching. In: IEEE International Conference on Signal Processing, ICSP. Vol. 2.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In: International Conference on Learning Representations. ICLR.

Gao, S., Guo, G., Huang, H., Chen, X., Chen, C.L.P., 2020. An endto end broad learning system for event-based object classification. In: IEEE Access. pp. 45974–45984.

Gao, F., Shi, C., Wang, R., Cai, A., Duan, C., Liu, M., 2025. Incremental few-shot instance segmentation via feature enhancement and prototype calibration. Comput. Vis. Image Underst. 253, 104317. http://dx.doi.org/10.1016/j.cviu.2025.104317.

Gaus, Y.F.A., Bhowmik, N., Isaac-Medina, B.K.S., Breckon, T.P., 2024. Performance Evaluation of Segment Anything Model with Variational Prompting for Application to Non-Visible Spectrum Imagery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops,.

Gaus, Y.F.A., et al., 2019. Evaluating the transferability and adversarial discrimination of convolutional neural networks for threat object detection and classification within X-Ray security imagery. arXiv preprint arXiv:1911.08966.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 580–587.

Griffin, L.D., Caldwell, M., Andrews, J.T.A., Bohler, H., 2019. ''Unexpected Item in the Bagging Area'': Anomaly detection in X-Ray security images.. pp. 1539–1553.

Hariharan, B., Arbeláez, P., Girshick, R., Malik, J., 2015. Hypercolumns for object segmentation and fine-grained localization. arXiv:1411.5752v2.

Hassan, T., Akçay, S., Bennamoun, M., Khan, S., Werghi, N., 2020a. Trainable structure tensors for autonomous baggage threat detection under extreme occlusion. In: Asian Conference on Computer Vision. ACCV.

Hassan, T., Akcay, S., Bennamoun, M., Khan, S., Werghi, N., 2022a. A novel incremental learning driven instance segmentation framework to recognize highly cluttered instances of the contraband items. In: IEEE Transactions on Systems, Man, and Cybernetics: Systems.

Hassan, T., Akçay, S., Bennamoun, M., Khan, S., Werghi, N., 2022b. Tensor Pooling Driven Instance Segmentation Framework for Baggage Threat Recognition. Neural Comput. Appl. 34, 1239– 1250.

Hassan, T., Akcay, S., Hassan, B., Bennamoun, M., Khan, S., Dias, J., Werghi, N., 2023. Cascaded structure tensor for robust baggage threat detection. In: Neural Computing and Applications.

Hassan, B., Hassan, T., 2014. Fully automated detection, grading and 3D modeling of maculopathy from OCT volumes. In: IEEE 2nd International Conference on Communication, Computing and Digital Systems (C-CODE).

Hassan, T., Shafay, M., Akçay, S., Khan, S., Bennamoun, M., Damiani, E., Werghi, N., 2020b. Meta-transfer learning driven tensor-shot detector for the autonomous localization and recognition of concealed baggage threats. In: MDPI Sensors.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2961–2969.

Jaccard, N., Rogers, T.W., Griffin, L.D., 2014. Automated detection of cars in transmission X-ray images of freight containers. In: IEEE International Conference on Advanced Video and Signal Based Surveillance. pp. 387–392.

Khan, A.M., Hassan, T., Akram, M.U., Alghamdi, N.S., Werghi, N., 2022. Continual learning objective for analyzing complex knowledge representations. Sensors 22 (4), 1667. http://dx.doi.org/10.3390/s22041667.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv:1412.6980.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., Girshick, R., 2023. Segment anything. arXiv:2304.02643.

Kundegorski, M.E., Akcay, S., Devereux, M., Mouton, A., Breckon, T.P., 2016. On using feature descriptors as visual words for object detection within X-ray baggage security screening. In: International Conference on Imaging for Crime Detection and Prevention.

Li, Y., Qi, H., Dai, J., Ji, X., Wei, Y., 2017. Fully convolutional instance-aware semantic segmentation. arXiv:1611.07709v2.

Liang, K., Gregory, C., Diallo, S.O., Roe, K., Heilmann, G., Carin, L., Carlson, D., Spell, G., Sigman, J., 2018. Automatic threat recognition of prohibited items at aviation checkpoint with X-ray imaging: a deep learning approach. In: Anomaly Detection and Imaging with X-Rays. ADIX.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollar, P., Zitnick, C.L., 2014. Microsoft COCO: Common objects in context. In: European Conference on Computer Vision. Springer, pp. 740–755.

Liu, D., Wang, Z., 2008. A united classification system of X-ray image based on fuzzy rule and neural networks. In: IEEE International Conference on Intelligent System and Knowledge Engineering.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440.

Ma, C., Zhuo, L., Li, J., Zhang, Y., Zhang, J., 2023. Occluded prohibited object detection in X-ray images with global context-aware multi-scale feature aggregation. Neurocomputing 519 (8), 1–16.

Masood, R.F., Taj, I.A., Khan, M.B., Qureshi, M.A., Hassan, T., 2022. Deep learning based vertebral body segmentation with extraction of spinal measurements and disorder disease classification. In: Biomedical Signal Processing and Control.

Mery, D., Svec, E., Arias, M., 2016a. Object recognition in baggage inspection using adaptive sparse representations of X-ray images. In: Pacific-Rim Symposium on Image and Video Technology. pp. 709–720.

Mery, D., Svec, E., Arias, M., 2016b. Object recognition in baggage inspection using adaptive sparse representations of X-ray images. In: Image and Video Technology.

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

Michieli, U., Zanuttigh, P., 2021. Knowledge distillation for incremental learning in semantic segmentation. Comput. Vis. Image Underst. 205, 103167.

Nercessian, S., Panetta, K., Agaian, S., 2008. Automatic detection of potential threat objects in X-ray luggage scan images. In: IEEE Conference on Technologies for Homeland Security.

Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. In: Nature Neuroscience.

Riffo, V., Lobel, H., Mery, D., 2015. GDXray: The database of X-ray images for nondestructive testing. J. Nondestruct. Eval. 34 (4), 42.

Rifo, V., Godoy, I., Mery, D., 2019. Handgun detection in single-spectrum multiple X-ray views based on 3D object recognition. In: Journal of Nondestructive Evaluation.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Samet Akcay, T.P.B., 2018. GANomaly: Semi-supervised anomaly detection via adversarial training. In: Asian Conference on Computer Vision. ACCV.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In: IEEE Conference on Computer Vision and Pattern Recognition. CVPR.

Shafay, M., Hassan, T., Damiani, E., Werghi, N., 2021a. Temporal fusion based mutli-scale semantic segmentation for detecting concealed baggage threats. In: IEEE International Conference on Systems, Man, and Cybernetics. SMC, pp. 232–237.

Shafay, M., Hassan, T., Velayudhan, D., Damiani, E., Werghi, N., 2021b. Deep fusion driven semantic segmentation for the automatic recognition of concealed contraband items. In: Proceedings of the 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020) 12. pp. 550–559.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556.

Singh, M., Singh, S., 2004. Image segmentation optimisation for X-ray images of airline luggage. In: IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety, CIHSPS 2004.

Sirshar, M., Hassan, T., Akram, M.U., Khan, S.A., 2021. An incremental learning approach to automatically recognize pulmonary diseases from the multi-vendor chest radiographs. Comput. Biol. Med. 134, 104435. http://dx.doi.org/10.1016/j.compbiomed.2021.104435.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2014. Going deeper with convolutions. arXiv:1409.4842v1.

Tan, M., Le, Q.V., 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: International Conference on Machine Learning. ICML.

Tian, Y., Krishnan, D., Isola, P., 2019. Contrastive representation distillation. arXiv preprint arXiv:1910.10699.

Turcsany, D., Mouton, A., Breckon, T.P., 2013. Improving feature-based object recognition for X-ray baggage security screening using primed visual words. In: IEEE International Conference on Industrial Technology. IEEE, pp. 1140–1145.

Velayudhan, D., Ahmed, A., Hassan, T., Gour, N., Owais, M., Bennamoun, M., 2024. Autonomous Localization of X-Ray Baggage Threats via Weakly Supervised Learning. In: IEEE Transactions on Industrial Informatics.

Velayudhan, D., Hassan, T., Ahmed, A.H., Damiani, E., Werghi, N., 2022a. Baggage threat recognition using deep low-rank broad learning detector. In: IEEE 21st Mediterranean Electrotechnical Conference. MELECON, pp. 966–971.

Velayudhan, D., Hassan, T., Damiani, E., Werghi, N., 2022b. Recent advances in baggage threat detection: A comprehensive and systematic survey. ACM Comput. Surv. 55 (8), 1–38.

Wang, L., Li, Y., Ding, J., Kangshun, 2005a. Structural X-ray image segmentation for threat detection by attribute relational graph matching. In: IEEE International Conference on Neural Networks and Brain.

Wang, L.-L., Li, Y.-X., Ding, J.-L., Ying, W.-Q., 2005b. ARG-based segmentation of overlapping objects in multi-energy X-ray image of passenger accompanied baggage. In: MIPPR 2005: Image Analysis Techniques.

Wang, L., Li, D., Zhu, Y., Tian, L., Shan, Y., 2020. Dual super-resolution learning for semantic segmentation. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 3773–3782.

Wang, B., Zhang, L., Wen, L., Liu, X., Wu, Y., 2021. Towards real-world prohibited item detection: A large-scale X-ray benchmark. In: IEEE/CVF International Conference on Computer Vision. ICCV.

Zagoruyko, S., Lerer, A., Lin, T.-Y., Pinheiro, P.O., Gross, S., Chintala, S., Dollar, P., 2016. A MultiPath Network for Object Detection. arXiv:1604.02135v2.

Zeiler, M.D., 2012. ADADELTA: An adaptive learning rate method. arXiv:1212.5701.

Zhang, L., Jiang, L., Ji, R., Fan, H., 2023. PIDray: A Large-Scale X-ray Benchmark for Real-World Prohibited Item Detection. In: International Journal of Computer Vision.

Zhang, L., Weng, L., Zhang, Y., Gao, F., 2025. Incremental few-shot instance segmentation without fine-tuning on novel classes. Comput. Vis. Image Underst. 254, 104323. http://dx.doi.org/10.1016/j.cviu.2025.104323.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2881–2890.