ORIGINAL RESEARCH

Unsupervised anomaly instance segmentation for baggage threat recognition

Taimur Hassan[1] · Samet Akçay[2] · Mohammed Bennamoun[3] · Salman Khan[4] · Naoufel Werghi[1]

Abstract

Identifying potential threats concealed within the baggage is of prime concern for the security staff. Many researchers have developed frameworks that can automatically detect baggage threats from security X-ray scans. However, to the best of our knowledge, all of these frameworks require extensive training efforts on large-scale and well-annotated datasets, which are hard to procure in the real world, especially for the rarely seen contraband items. This paper presents a novel unsupervised anomaly instance segmentation framework that recognizes baggage threats, in X-ray scans, as anomalies without requiring any ground truth labels. Furthermore, thanks to its stylization capacity, the framework is trained only once, and at the inference stage, it detects and extracts contraband items regardless of their scanner specifications. Our one-staged approach initially learns to reconstruct normal baggage content via an encoder–decoder network utilizing a proposed stylization loss function. The model subsequently identifies the abnormal regions by analyzing the disparities within the original and the reconstructed scans. The anomalous regions are then clustered and post-processed to fit a bounding box for their localization. In addition, an optional classifier can also be appended with the proposed framework to recognize the categories of these extracted anomalies. A thorough evaluation of the proposed system on four public baggage X-ray datasets, without any re-training, demonstrates that it achieves competitive performance as compared to the conventional fully supervised methods (i.e., the mean average precision score of 0.7941 on SIXray, 0.8591 on GDXray, 0.7483 on OPIXray, and 0.5439 on COMPASS-XP dataset) while outperforming state-of-the-art semi-supervised and unsupervised baggage threat detection frameworks by 67.37%, 32.32%, 47.19%, and 45.81% in terms of F1 score across SIXray, GDXray, OPIXray, and COMPASS-XP datasets, respectively.

Keywords Anomaly instance segmentation · Fast Fourier transform · X-rays · Baggage threat detection

1 Introduction

Recognizing contraband items concealed within baggage is a prime security concern as it endangers public safety. According to a recent report, approximately 1.5 million passengers in the United States are searched every day for weapons and other dangerous items (Council 1996). Manual detection of these items is a tiring task and also subject to human errors caused due to fatigued work schedules, amount of baggage clutter, aviation traffic load, or simply because of less experience towards screening the contraband data. To overcome this, many researchers have developed automated frameworks (Gaus et al. 2019a; Akçay et al. 2018a; Turcsany et al. 2013) to screen baggage at airports, malls, and cargoes. However, the majority of these frameworks are developed using conventional RGB detectors, which have limited performance towards localizing the occluded suspicious objects due to their region based proposal generation mechanisms (Hassan et al. 2020a), and because of the inherent differences between the RGB and the X-ray imagery (Akçay et al. 2018b). To handle this, researchers have recently proposed clutter-aware solutions possessing

✉ Taimur Hassan
  taimur.hassan@ku.ac.ae

1   Center for Cyber-Physical Systems (C2PS), Department
    of Electrical Engineering and Computer Sciences, Khalifa
    University, Abu Dhabi, United Arab Emirates

2   Department of Computer Sciences, Durham University,
    Durham, UK

3   Department of Computer Science and Software Engineering,
    The University of Western Australia, Perth, Australia

4   Mohamed bin Zayed University of Artificial Intelligence,
    Abu Dhabi, United Arab Emirates

the capacity to recognize concealed and occluded baggage threats regardless of the scanner specifications (Hassan and Werghi 2020; Hassan et al. 2020a, b), acquisition noise (Tao et al. 2021), and the imbalanced nature of the contraband data (Miao et al. 2019). In addition to this, the majority of these methods have been quantitatively evaluated to detect threatening items against different levels of occlusion on the publicly available datasets (Wei et al. 2020; Hassan et al. 2020b; Hassan and Werghi 2020). Also, the researchers have utilized 3D detectors to get rid of occlusion while screening the baggage threats from the volumetric computed tomography (CT) imagery (Wang and Breckon 2020; Wang et al. 2020a). Despite these recent advancements, many state-of-the-art baggage threat detection frameworks are still based on conventional supervised learning schemes which require extensive ground truth labels to ensure robust detection performance. Some researchers have also presented semi-supervised (Akçay et al. 2018a) and unsupervised anomaly detection (Akçay et al. 2019) via adversarial learning. However, such schemes require an explicit re-training process for recognizing baggage threats from different datasets. Also, they are driven by scan-level analysis for recognizing the anomalous baggage threats and do not possess the capacity to extract and localize the threatening items within the baggage X-ray scans (Akçay et al. 2018a, 2019).

To address the above limitations, we present in this work a novel unsupervised anomaly instance segmentation. The proposed approach requires only one-time training, without any ground truth labels, to recognize the baggage threats.

2 Related work

Early baggage threat detection frameworks were based on conventional machine learning employing hand-engineered features (Bastan et al. 2011; Riffo and Mery 2015). Then deep learning methods took over, proposing supervised and unsupervised strategies for recognizing the suspicious baggage content. Here, the recent approaches have also addressed the imbalanced (Miao et al. 2019) and cluttered (Hassan and Werghi 2020; Wei et al. 2020) nature of the threatening items in X-ray scans, which are often observed in the real world at airports, malls, and transmission cargoes. This section first gives a brief overview of some of the conventional baggage threat detection schemes, and then it sheds light on the state-of-the-art deep learning-based approaches. For an exhaustive survey, we refer the reader to the work of Akçay and Breckon (2020) and Mery et al. (2017, 2020).

2.1 Conventional machine learning methods

Earlier methods for screening baggage threats are based on handcrafted features (Megherbi et al. 2012; Wanget al. 2020b) and descriptors such as SIFT (Mery et al. 2016; Zhang et al. 2014), SURF (Bastan et al. 2011), and FAST-SURF (Kundegorski et al. 2016), employed with Support Vector Machines (SVM) (Turcsany et al. 2013; Kundegorski et al. 2016), Bag of Words (BoW), K-Nearest Neighbors (Riffo and Mery 2015) and Random Forest (Jaccard et al. 2014) classifiers. Many researchers have also proposed supervised segmentation (Heitz and Chechik 2010) and detection (Bastan 2015) schemes for recognizing prohibited items via high, low and multi-view X-ray imagery (Bastan 2015). Similarly, Riffo and Mery (2015) proposed an Adapted Implicit Shape Model (AISM) for recognizing different contraband items from the publicly available GDXray (Mery et al. 2015) dataset. In another approach, they developed structure-from-motion-based 3D feature descriptors for recognizing the threatening items (Mery et al. 2016).

Although, traditional machine learning methods can mass-screen the baggage content using security X-ray scans. However, they are only applicable to limited experimental settings and cannot be well-generalized to multiple scanner specifications.

2.2 Deep learning methods

Deep learning has greatly enhanced the recognition capabilities of the threat screening frameworks such that they can now identify suspicious objects within grayscale or colored baggage X-ray scans regardless of their scanner properties. Here, we categorized all the deep learning-based threat detection frameworks as supervised and unsupervised approaches.

2.2.1 Supervised approaches

Supervised approaches for recognizing baggage threats employed classification (Akçay et al. 2016; Jaccard et al. 2017; Zhao et al. 2018; Miao et al. 2019), detection (Liu et al. 2018; Xu et al. 2018; Hassan et al. 2020a, b) and segmentation (Hassan and Werghi 2020; An et al. 2019; Gaus et al. 2019b) strategies. Akçay et al. (2016) introduced GoogleNet (Szegedy et al. 2014) (in a transfer learning mode) to detect threatening objects within baggage X-ray imagery. Jaccard et al. (2017) used VGG-19 (Simonyan and Zisserman 2015) on log-transformed scans to detect suspicious objects. Zhao et al. (2018) initiated the use of GANs to enhance the classification performance of the customized networks towards baggage threat detection. Apart

from this, researchers have also used two-staged (Liu et al. 2018) and one-staged (Gaus et al. 2019b) detectors along with attention mechanisms (Xu et al. 2018) to recognize and localize threatening objects. Moreover, Gaus et al. (2019a) measured the transferability of Faster R-CNN (Ren et al. 2016) Mask R-CNN (He et al. 2017) and RetinaNet (Lin et al. 2017) between various X-ray scanners to detect the contraband data. Motivated by the class imbalance between normal and suspicious objects, Miao et al. (Miao et al. 2019) presented the class-balanced hierarchical refinement (CHR) model, proposing architecture-oriented mitigation of the class imbalance problem. Other approaches proposed contour-driven object detectors such as Cascaded Structure Tensors (CST) (Hassan et al. 2020a), and Dual-Tensor Shot Detector (DTSD) (Hassan et al. 2020b). Similarly, Wei et al. (2020) developed De-occlusion Attention Module (DOAM), a plug-and-play module that can be integrated with conventional object detectors to increase their capacity in screening occluded baggage threats. For the segmentation approaches, An et al. (2019) employed encoder–decoder models leveraging dual attention mechanisms, while (Hassan and Werghi 2020) proposed a first-ever contour instance segmentation framework exclusively designed to extract cluttered contraband data from the security X-ray scans.

2.2.2 Unsupervised approaches

Researchers have also developed semi-supervised and unsupervised methods for recognizing suspicious items. Akçay et al. pioneered this by developing GANomaly (Akçay et al. 2018a), an encoder–decoder–encoder-driven adversarial framework trained on the normal security X-ray scans. After training, GANomaly (Akçay et al. 2018a) recognizes the baggage threats, as anomalies, from the abnormal test scans through its in-built discriminator. In another approach, Skip-GANomaly is proposed (Akçay et al. 2019) as an improved version of GANomaly utilizing encoder–decoders with skip-connections and adversarial learning to detect anomalous baggage threats with a significantly lesser amount of computational resources.

To the best of our knowledge, the majority of the existing frameworks are based on supervised learning, requiring an extensive amount of well-annotated training data to perform well at the inference stage (Hassan and Werghi 2020; Miao et al. 2019; Wei et al. 2020; Gaus et al. 2019b). However, procuring a large-scale and well-annotated dataset is often impractical and infeasible, especially for recognizing those items which are rarely observed during the aviation screening. Furthermore, re-training or even fine-tuning the deployed framework to identify a new type of threat is an inefficient process and could lead to compromised performance (Gaus et al. 2019a; Hassan et al. 2020b). Despite recent efforts leveraging meta-transfer-learning (Sun et al. 2019) to alleviate the scanner

differences and increase the generalizability of baggage threat detectors (Hassan et al. 2020b), these frameworks still require fine-tuning on different datasets for achieving good performance. Although researchers have utilized semi-supervised and unsupervised adversarial learning to recognize suspicious baggage items (Akçay et al. 2018a, 2019). These frameworks still require explicit re-training on the normal data of each dataset to identify baggage threats. Also, these methods can recognize suspicious items but are unable to localize them through bounding boxes or masks. Hence, this paper presents the first unsupervised anomaly instance segmentation framework exclusively designed to recognize and localize illegal baggage items from the security X-ray scans to the best of our knowledge.

3 Contributions

This paper presents a novel unsupervised anomaly instance segmentation framework to detect and extract baggage threats as anomalies. To the best of our knowledge, this is the first approach towards unsupervised anomaly instance segmentation, in a baggage threat detection territory, exhibiting the following distinctive features:

- The proposed framework is the first of its kind that is trained only once on the normal baggage X-ray scans. Afterward, it does not require re-training to eliminate the scanner differences to extract anomalous regions (across different datasets).
- The proposed framework is built upon a novel Gaussian-Weighted Fourier Stylization (GW-FS) scheme that drastically removes the scanner variations to achieve high generalizability towards extracting the suspicious baggage items as anomalies from the baggage X-ray scans.
- A thorough validation on four public X-ray datasets showcases that the proposed framework outperforms its unsupervised and semi-supervised competitors while achieving a competitive performance with the other fully supervised frameworks.

The rest of the paper is organized as follows: Sect. 4 presents the proposed framework, Sect. 5 showcases the experimental setup, Sect. 6 presents the detailed evaluation results, and Sect. 7 discusses the prospects of the proposed framework and concludes the paper.

4 Proposed approach

The block diagram of the proposed framework is shown in Fig. 1. We can see here that the encoder–decoder network is trained only once on the X-ray scans (containing the normal baggage data). During this one-time training, the
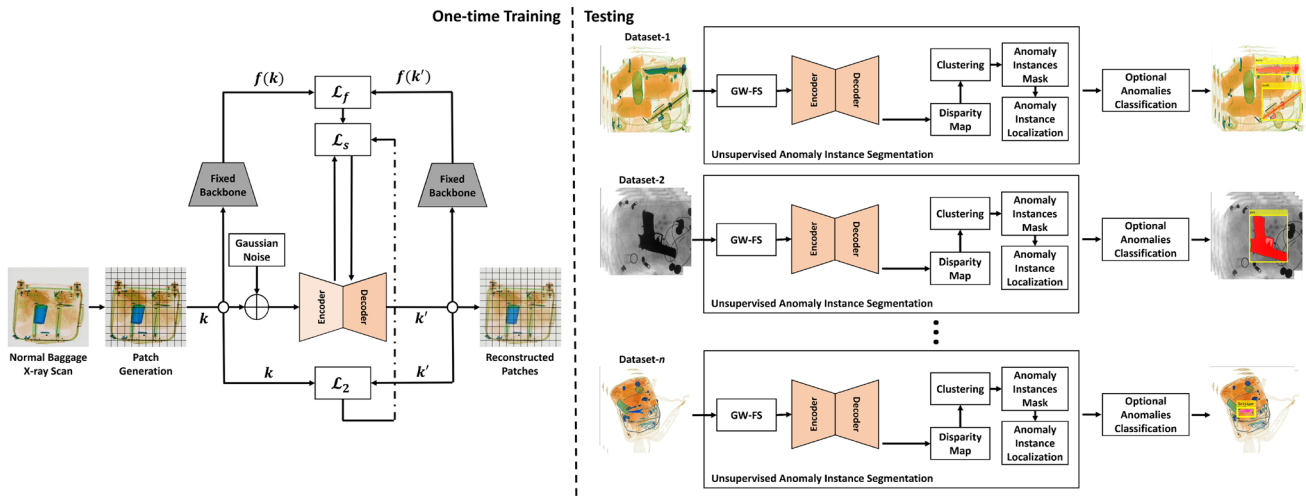
Fig. 1 Block diagram of the proposed framework. In the one-time training stage, the X-ray scans (containing the normal baggage data) are decomposed into fixed-size non-overlapping patches, which are passed to the proposed encoder–decoder network that learns to reconstruct them. Afterward, during the inference stage, the trained model is fed with the abnormal scans. After reconstructing them, the proposed framework exploits the original and reconstructed scans' disparities to recognize anomalous regions. Furthermore, the proposed GW-FS scheme removes the scanner-specific appearance enabling the proposed framework to identify baggage threats irrespective of the scanner specifications or the dataset.

network is constrained via custom stylization loss function ($\mathcal{L}_s$) to reconstruct the baggage X-ray scans accurately. The reconstruction is performed patch-wise, and to ensure that the network maintains the spatial characteristics of the original input scan, we perturb it, in each patch, with randomized zero-mean Gaussian noise. We empirically found that the addition of Gaussian noise within each patch puts more constraint to $\mathcal{L}_s$ in chastising the encoder–decoder network towards producing accurate reconstruction.

Moreover, $\mathcal{L}_s$ minimizes not only the $\mathcal{L}_2$ loss but also the differences in the feature representations obtained from the fixed backbone model. These two mechanisms enhance the encoder–decoder model's capacity for reconstructing the normal scans, leading to the generation of distinct disparity maps (for the abnormal scans) at the inference stage. The disparity maps are then clustered together and are post-processed to detect and locate the anomalous regions. Moreover, an optional lightweight classifier can be mounted at the back of the proposed framework to recognize the localized anomalous items' categories. It should be noted here that before feeding the test scan into the proposed model, we stylize it first based upon the proposed Gaussian-Weighted Fourier Stylization (GW-FS). This stylization removes the scanner-specific appearances (even the drastic ones), enabling the encoder–decoder network to reconstruct the test scan patches accurately regardless of the scanner model. The detailed description of each module is presented below:

4.1 Gaussian-weighted Fourier stylization

To perform stylization, we propose a Gaussian-Weighted Fourier Stylization (GW-FS) scheme. The GW-FS is inspired from Fourier Domain Adaptation (FDA) (Yang and Soatto 2020) that computes the Fast Fourier Transform (FFT) (Cooley and Tukey 1965) of the reference and target scans and copy the frequency samples within the magnitude spectrum of the reference scans to the target scan spectrum (defined by the rectangular window $\mathcal{R}$) without altering the phase spectrum (Yang and Soatto 2020). However, in our approach, we perform stylization by mixing the low-frequency components within the candidate scan's magnitude response with the reference scan's magnitude spectrum by fitting a Gaussian window (parameterized by the variance σ). Let $x \in \mathbb{R}^{M \times N}$ be the input scan (such that M denotes its height and N denotes its width), and $y \in \mathbb{R}^{M \times N}$ be the reference image. Taking FFT yields:

$$X_{u,v} = \frac{1}{MN} \sum_{s=0}^{M-1} \sum_{t=0}^{N-1} (-1)^{s+t} x_{s,t} e^{-j2\pi\left(u\frac{s}{M} + v\frac{t}{N}\right)}, \quad (1)$$

where $X = \mathcal{F}\{x\}$ represents the complex frequency spectrum of x, $\mathcal{F}$ denotes the FFT operator, and the factor $(-1)^{s+t}$ shifts the image spectrum by $M/2$ and $N/2$ to center-align it. We apply the same transformation to reference scan y, yielding $Y = \mathcal{F}\{y\}$. Afterward, the magnitude spectra of Y, i.e., $|Y|$ is multiplied by Gaussian window $\mathcal{G}$ to extract the low ranging spectral components for stylization:
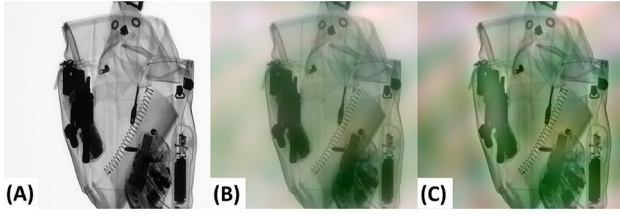
Fig. 2 A Original grayscale scan, **B** stylization through GW-FS scheme with $\sigma = 5$, **C** stylization through FDA (Yang and Soatto 2020) with $\beta = 5$. For fairness, the value of scaling factor (σ and β) in both schemes are chosen the same

$$S_{u,v} = |Y_{u,v}| \times \mathcal{G}_{u,v} = \frac{1}{2\pi\sigma^2} |Y_{u,v}| \times e^{-(u^2+v^2)/2\sigma^2}, \quad (2)$$

$S \in \mathbb{R}^{M \times N}$ denotes the obtained stylization mask that is added to $|X|$ to produce $|X'| = |X| + S$. Afterward, we apply the inverse FFT ($\mathcal{F}^{-1}$) of X' to obtain the stylized scan $x' = \mathcal{F}^{-1}\{X'\}$, as expressed below:

$$x'_{s,t} = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} (-1)^{u+v} X'_{u,v} e^{j2\pi\left(s\frac{u}{M}+t\frac{v}{N}\right)}, \quad (3)$$

We can observe here that using a Gaussian window $\mathcal{G}$ instead of the rectangular window $\mathcal{R}$ (employed in FDA) results in fewer ripples within the stop-band, ensuring thus a much better stylization as evidenced in Fig. 2.

Also, $\mathcal{R}$ (in FDA) does not blend the frequency spectrum between $|X|$ and $|Y|$. It just replaces samples within $|X|$ with that of $|Y|$ (constraint by $\mathcal{R}$), where the length of $\mathcal{R}$ is administered by the β factor (Yang and Soatto 2020). Therefore, the stylization through FDA (Yang and Soatto 2020) produces additional noisy artifacts (as shown in Fig. 2), and optimizing them for each training-testing combination is a haggling job. GW-FS scheme addresses this by first weighting the frequency samples of $|Y|$ by G (through Eq. 2) before merging them with the target spectra $|X|$.

4.2 Scans reconstruction

After stylizing the test scan, it is passed to the asymmetric one-time trained encoder–decoder model, which generates the reconstructed images (x'') patch-wise. Afterward, x'' is utilized in developing the disparity maps for anomaly instance segmentation. The proposed encoder–decoder network is a lightweight model containing one input layer, seven convolution layers with ReLU activations, three maxpooling, and three up-sampling layers. Furthermore, it has around 4,923 trainable parameters. For more architectural

details about the proposed encoder–decoder architecture, we refer the reader to the source code repository[1].

Moreover, to train the proposed encoder–decoder network, we used the proposed stylization loss function ($\mathcal{L}_s$) to constrain it, at the training time, to recognize shape, context, and edge feature appearances from the latent space vectors. The $\mathcal{L}_s$ loss function is further discussed in the subsequent section below.

4.2.1 The $\mathcal{L}_s$ loss function

To train the proposed encoder–decoder model, we propose a novel stylization loss ($\mathcal{L}_s$) which is is a linear combination of feature reconstruction loss function ($\mathcal{L}_f$) (Johnson et al. 2016) and the conventional $\mathcal{L}_2$ loss function.

$$\mathcal{L}_s = \alpha_1 \mathcal{L}_f + \alpha_2 \mathcal{L}_2, \quad (4)$$

where $\alpha_{1,2}$ represent the loss weights. $\mathcal{L}_f$ is generated from the feature representations obtained from the frozen pretrained backbones, and $\mathcal{L}_2$ is obtained using the pixel-wise difference between the training scan (k) and the reconstructed version (k'), as expressed in Eq. 5 and 6:

$$\mathcal{L}_f = \frac{1}{b_s} \sum_{i=0}^{b_s} |f(k_i) - f(k'_i)|, \quad (5)$$

$$\mathcal{L}_2 = \frac{1}{b_s} \sum_{i=0}^{b_s} |k_i - k'_i|^2, \quad (6)$$

where b_s is the batch size, and $f(.)$ denotes the feature representations obtained from the frozen backbone model. The loss weights α_1 and α_2 are empirically determined to be 0.7 and 0.3, respectively.

4.2.2 Unsupervised anomaly instance segmentation

The proposed unsupervised anomaly instance segmentation scheme is shown in Fig. 3. At the inference stage, after stylizing the input scan, it is patch-wise reconstructed through the trained encoder–decoder model. Then, the reconstructed scan (x'') is subtracted from the stylized scan (x') to produce the disparity map. The disparity map is utilized in extracting the suspicious items' instances by clustering the color distribution between anomalous and normal baggage content. The detailed description of disparity maps and the color distribution-based clustering scheme is presented in the subsequent sections.

---

[1] The source code of the proposed framework, along with the complete documentation is available at https://github.com/taimurhassan/anomaly.
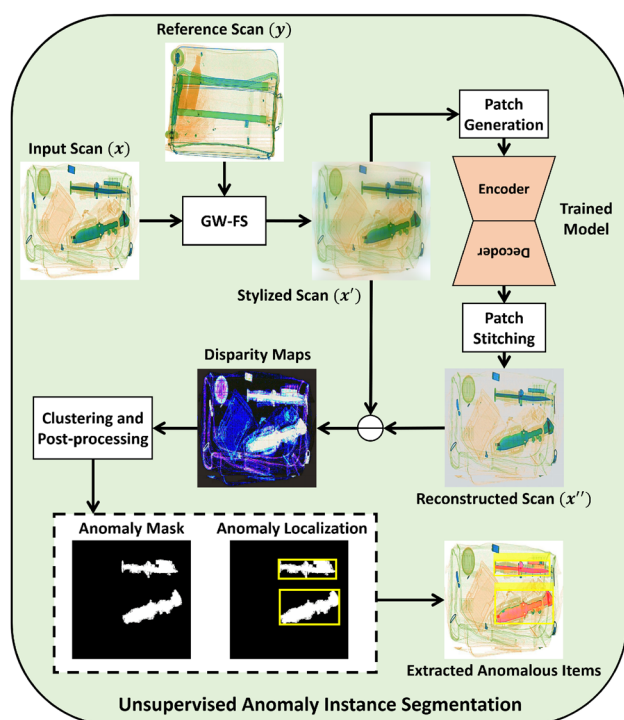
Fig. 3 Unsupervised anomaly instance segmentation framework. Here, we train an encoder–decoder model only once to reconstruct normal baggage X-ray scans. During the inference stage, the trained model recognizes the anomalous regions by exploiting the actual and reconstructed scans' disparities. To eliminate the scanner variations, we propose a GW-FS scheme that mixes the frequency representations within the reference scan and the input scans.

Disparity maps The disparity maps reveal the deviation of the anomalous items w.r.t the normal baggage content by subtracting x'' from x'. It should be noted here that the meaningful interpretation from these disparity maps is subject to how accurately the encoder–decoder model reconstructs the normal areas within the abnormal scans. For example, in Fig. 3, we can see how effectively the encoder–decoder has reconstructed the abnormal scan (containing *knives*). However, there are still some noticeable intensity variations between x' and x'', which results in the blue, pink, and cyan color noisy regions within the disparity maps. Having a three-channeled representation here allows better discrimination between anomalous regions (corresponding to suspicious items) and the rest of the baggage content as compared to the single-channeled representations. This aspect is further evidenced in Fig. 4. Here, for each dataset, red points indicate anomalies, whereas blue color showcases normal pixels. We can observe that the distributions of anomalous and normal baggage content are well-separated. Therefore using an adequate clustering scheme, the anomalous region representing the suspicious items can be extracted.

Color clustering In order to extract suspicious (anomalous) items' instances, the disparity maps are clustered
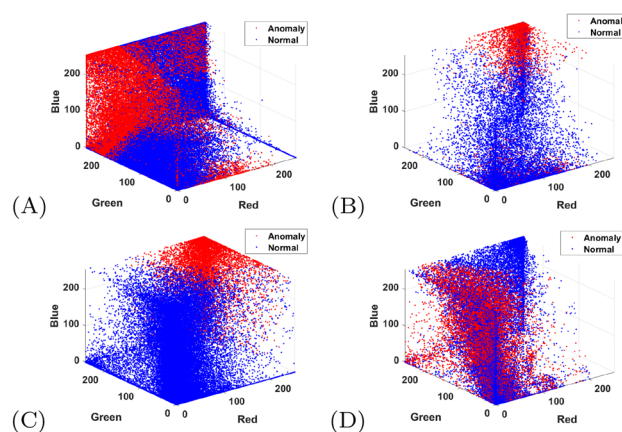

Fig. 4 Color distributions between anomalous and non-anomalous regions in **A** SIXray (Miao et al. 2019) dataset, **B** GDXray (Mery et al. 2015) dataset, **C** OPIXray (Wei et al. 2020) dataset, and (D) COMPASS-XP (Griffin et al. 2019) dataset

through K-means Clustering (parameterized by the number of clusters $\mathcal{C}$). Here, $\mathcal{C}$ for each dataset varies depending upon normal and anomalous items contained within the respective scans. Through empirical analysis, we determined the optimal choice of $\mathcal{C}$ for SIXray (Miao et al. 2019) and OPIXray (Miao et al. 2019) dataset is four (i.e., $\mathcal{C} = 4$). Similarly, for GDXray (Mery et al. 2015) and COMPASS-XP (Griffin et al. 2019) dataset, $\mathcal{C} = 3$.

Moreover, the noisy clusters (obtained after K-means clustering) are automatically removed through morphological post-processing. Afterward, each isolated instance of the anomalous region is identified through the connected-component analysis, and then each item instance is localized by fitting the bounding box generated using the minimum and maximum of their masks in both image dimensions (see Fig. 3).

4.3 Recognition of anomalous items

After extracting the anomalous items we identify their categories (such as *gun*, *knife*, *razor*, *shuriken*, *wrench*, *pliers*, *scissor*, *hammer*, and *axe* etc.) using a proposed lightweight classification model (see Fig. 5). We want to highlight here that the recognition of anomalies items is just an optional module. It neither relates to our actual unsupervised anomaly instance segmentation approach nor it is mandatory in the proposed framework.

The patches of the anomalous items are cropped from x' using their bounding boxes, and the classifier is trained only once to recognize their categories. The data used to train this model is based on patches (representing each suspicious item category), and it is taken from all four datasets. Architecturally, the classification model contains three convolutions, three ReLUs, three max-pooling, one fully
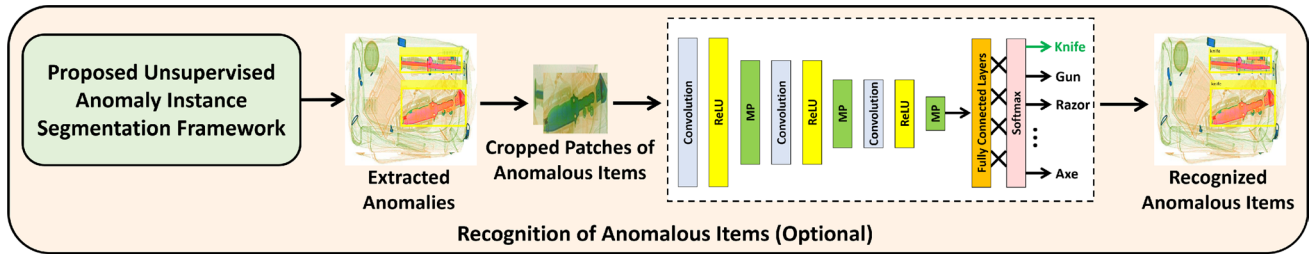
Fig. 5 Recognition of anomalous items' categories using the proposed classification model. The model is trained only once on the suspicious items patches obtained from all four datasets

connected, and one softmax layer as depicted in Fig. 5. The total number of parameters within the proposed model is 3.2 M. Note that any pre-trained model can be used here for classifying the suspicious object categories. However, our proposed model exhibits two advantages (1) it is lightweight compared to the popular pre-trained models, and (2) it achieves a good trade-off between accuracy and the number of hyper-parameters (as evidenced from Table 3). The classification model is optimized via the cross-entropy loss function ($\mathcal{L}_c$), as expressed below:

$$\mathcal{L}_c = -\frac{1}{b_s} \sum_{i=0}^{b_s-1} \sum_{j=0}^{n_c-1} t_{i,j} \log(p(l_{i,j})) \qquad (7)$$

where n_c denotes the number of suspicious item categories, $t_{i,j}$ denotes the i th training example for the j th class, and $p_{i,j}$ denotes the softmax probability for the i th training sample belonging to j th class. Moreover, the training details of this optional classification model is presented in Sect. 5.2.

5 Experimental setup

This section presents the details about the datasets, the training protocol, as well as the evaluation metrics:

5.1 Datasets

The proposed framework is thoroughly evaluated on all the four public X-ray datasets used in baggage threat recognition benchmarking. The detailed description of these datasets is as follows:

5.1.1 SIXray

SIXray (Miao et al. 2019) is the largest and most challenging baggage X-ray dataset to date. It contains 1,050,302 negative X-ray scans containing only the normal baggage content, and 8,929 positive scans having one or more suspicious items in it such as *guns*, *knives*, *wrenches*, *pliers*, *scissors*, and *hammers*. Furthermore, the dataset also contains detailed

box-level annotations to train and evaluate the baggage threat detection frameworks. Moreover, SIXray (Miao et al. 2019) is primarily designed to test the capacity of autonomous frameworks to screen contraband items in a highly imbalanced scenario.

5.1.2 GDXray

The GDXray (Mery et al. 2015) dataset contains 19,407 high resolution grayscale X-ray scans divided into five groups, namely, *welds*, *casting*, *baggage*, *nature*, and *settings*. The only relevant category for the proposed study is the baggage category that contains 8150 X-ray scans along with detailed markings. Moreover, the scans within the GDXray (Mery et al. 2015) contains suspicious items such as *razors*, *handguns*, *knives*, and *shuriken* (Mery et al. 2015).

5.1.3 OPIXray

The OPIXray (Wei et al. 2020) is the most recent publicly released baggage X-ray dataset. It contains a total of 8,885 colored X-ray scans containing suspicious items such as *folding knives*, *straight knives*, *utility knives*, *multi-tool knives*, and *scissors* (Wei et al. 2020). Furthermore, OPIXray (Wei et al. 2020) also contain the detailed box-level annotations for these items which can be used in order to evaluate the autonomous baggage threat detection frameworks.

5.1.4 COMPASS-XP

Different from the previous datasets, COMPASS-XP (Griffin et al. 2019) is mainly designed to assess classification (rather than detection) frameworks, i.e., it contains scan-level markings to recognize baggage threats without ground truths masks for the localization. However, the novel aspect of the COMPASS-XP dataset (Griffin et al. 2019) is that it contains different scanner images for each case. For example, for a single scene in which a baggage contains a *gun*, the COMPASS-XP gives six different scanner representations (i.e., the high-energy X-ray, low-energy X-ray, high density, colored X-ray, grayscale X-ray, and the normal RGB scan).

So, in total, the complete dataset contains 11,568 = 6 x 1928 X-ray scans (Griffin et al. 2019).

5.2 Training details

The proposed framework has been implemented using Python 3.7.8 with TensorFlow 2.3.0 and the MATLAB R2020a on a machine with Intel Core i9-10940@3.3 GHz processor and 132 GB RAM with a single NVIDIA Quadro RTX 6600, cuDNN v7.5, and a CUDA Toolkit v11.0.221. The training was conducted for 200 epochs using 80% of the normal baggage X-ray scans (i.e., 840,241 normal scans) from the SIXray dataset. The choice of the SIXray dataset for one-time training is driven from an extensive ablation study (presented in Sect. 6.1.4). Moreover, the total number of test scans from all four datasets which we used for evaluating the proposed framework are 238,664 (8,150 scans are taken from GDXray, 210,061 scans are taken from SIXray, 8,885 scans are taken from OPIXray, and 11,568 scans are taken from COMPASS-XP dataset). Apart from this, we used 8,929 scans from the SIXray dataset for validation purposes.

Furthermore, the optimizer used during the training was ADAM (Kingma and Ba 2015) with default learning and decay rates. For recognition of anomalous items, we trained the modular classification model for 100 epochs with ADAM (Kingma and Ba 2015) having an initial learning rate of 0.0001. The total training patches we used to train this classifier are around 5,000, obtained from all four X-ray datasets for each suspicious item category. The source code of the proposed framework is also released publicly for the research community[1].

5.3 Evaluation metrics

We used standard object detection, instance segmentation, and classification metrics such as mAP, MSE, accuracy, recall, precision, F1 to test the proposed framework's performance and compare it with the state-of-the-art solutions.

6 Results

6.1 Ablation study

The ablation study for the proposed framework includes

1. The choice of σ for GW-FS stylization;
2. The optimal backbone network for computing $\mathcal{L}_f$;
3. The optimal classification backbone model, and
4. the choice of training dataset which is used for performing one-time training.

Table 1 Effects of varying σ on scan reconstruction

σ	SIXray	GDXray	OPIXray	COMPASS-XP
5	**72.92**	**16.29**	**21.12**	**21.94**
10	89.16	24.83	41.96	43.62
25	159.53	73.92	119.65	136.42
50	235.98	134.76	216.94	218.16

Bold indicates the best MSE scores

Table 2 Performance of different pre-trained models for computing $\mathcal{L}_f$ during one-time training

Model	SIXray	GDXray	OPIXray	COMP
VGG-16	**72.92**	**16.29**	**21.12**	**21.94**
ResNet-50	80.16	20.54	27.82	32.11

Bold indicates the best MSE scores. Moreover, the abbreviation 'COMP' represents the COMPASS-XP dataset (Griffin et al. 2019)

6.1.1 Choice of σ

The σ is a hyper-parameter within the GW-FS scheme to determine the Gaussian window's cut-off frequency. Increasing the value of σ increases the pass-band region and allows more frequencies to pass through, whereas decreasing the value of σ only allows the lowest frequencies (among all) to remain while the rest are clipped. Table 1 reports the performance of GW-FS for reconstructing three-channeled scans in terms of MSE scores. Here, we can observe that for $\sigma = 5$, the proposed framework achieves the minimum reconstruction error for all datasets. However, when σ increases, the reconstruction performance starts to deteriorate because higher frequencies are being allowed to pass through the window (defined by σ), which generates more noisy edges.

6.1.2 Backbone network for computing $\mathcal{L}_f$

This ablation study reports the backbone's choice for computing the feature reconstruction loss function ($\mathcal{L}_f$). Here, we compared the performance of pre-trained VGG-16 (Simonyan and Zisserman 2015) and ResNet-50 (He et al. 2016) that produces $\mathcal{L}_f$ to penalize the encoder–decoder model towards reconstructing the abnormal scans robustly. It should be noted here that these pre-trained models were fixed, i.e., the weights of these networks were not trained explicitly for minimizing $\mathcal{L}_f$. The results for this experiment are reported in Table 2. We can see here that $\mathcal{L}_f$ with VGG-16 (Simonyan and Zisserman 2015) achieves 9.03% better performance on SIXray (Miao et al. 2019) dataset. Similarly, on GDXray (Mery et al. 2015), OPIXray (Wei et al. 2020), and COMPASS-XP (Griffin et al. 2019) dataset, VGG-16 (Simonyan and Zisserman 2015) driven $\mathcal{L}_f$ achieved 20.69%,

Table 3 Comparison of classification performance for recognizing anomalous items patches

Model	ACC	TPR	PPV	F1	NP
PB	**0.9669**	**0.8683**	**0.5083**	**0.6412**	**3.2M**
V-16	0.9630	0.8538	0.4759	0.6111	14.7M
R-50	0.9740	0.9172	0.5745	0.7064	23.5M
R-101	0.9786	0.9341	0.6242	0.7483	42.6M
R-152	0.9877	0.9446	0.7568	0.8403	58.3M
D-121	0.9754	0.9053	0.5909	0.7150	7.03M
MNv2	0.9527	0.8247	0.4049	0.5431	2.2M

The good trade-off between classification performance and the computational requirements is highlighted in bold. Moreover, the abbreviations are *ACC* accuracy, *TPR* true positive rate, *PPV* positive predicted value, *F1* F1 Score, *NP* number of parameters, *PB* proposed backbone, *V-16* VGG-16 (Simonyan and Zisserman 2015), *R-50* ResNet-50 (He et al. 2016), *R-101* ResNet-101 (He et al. 2016), *R-152* ResNet-152 (He et al. 2016), *D-121* DenseNet-121 (Huang et al. 2017), *MNv2* MobileNetv2 (Howard et al. 2017)

24.08%, and 31.67% better reconstruction performance, respectively. However, if we increase the training epochs for ResNet-50 (He et al. 2016), we can achieve similar performance. But since VGG-16 (Simonyan and Zisserman 2015) produced better results with lesser training, we opted for it in the proposed framework to compute $\mathcal{L}_f$ during one-time training.

6.1.3 The optimal classification backbone

Recognition of anomalous item (after unsupervised anomaly instance segmentation) is an optional step required to identify the type of anomaly contained within the localized anomalous region. To perform this, we exploited different pre-trained models (fine-tuned on suspicious items patches). We also compared the performance of a proposed classification model with these pre-trained models to see how well it recognizes the suspicious items (contained within the patches). The comparison is reported in Table 3. Here, we can observe that with few training examples (i.e., the suspicious items patches) from all the datasets, the proposed model achieves competitive classification performance compared to other pre-trained networks. Furthermore, considering that it has 54.48% fewer parameters than best performing DenseNet-121 (Huang et al. 2017) model. We believe that it provides a good trade-off between performance and computational requirements, especially compared to the MobileNetv2 (Howard et al. 2017).

6.1.4 Choice of one-time training dataset

The encoder–decoder model within the proposed framework is trained only once, and in this one-time training, it learns to reconstruct the normal baggage content at run-time robustly.

Table 4 Reconstruction performance of the proposed encoder–decoder model in terms of MSE scores when trained on different datasets

Training Dataset	SIXray	GDXray	OPIXray	COMP
SIXray	72.928	16.291	21.125	21.941
GDXray	96.532	13.639	45.923	51.649
COMP	81.692	21.638	30.113	10.582

The abbreviation 'COMP' represents the COMPASS-XP dataset (Griffin et al. 2019)

As the model does not learn to recognize suspicious objects (during training), it faces a hard time reconstructing them at the inference stage, and this is highlighted within the disparity maps.

In this experiment, we determine the optimal choice of the dataset for training the encoder–decoder model. The reconstruction performance of the proposed model (in terms of MSE scores) is reported in Table 4. Here, we can see that using SIXray (Miao et al. 2019) dataset for training; the proposed model produces the best reconstruction performance across all four datasets at the inference stage. This is because SIXray (Miao et al. 2019), to the best of our knowledge, contains the maximum amount of scans depicting diverse ranging normal baggage content, allowing the encoder–decoder model to learn the unique feature representations within the baggage X-ray scans robustly. Training on GDXray (Mery et al. 2015) dataset does not produce a very good performance for two reasons: 1) It contains significantly lesser normal baggage scans (around 1,130 of them) for training purposes. 2) GDXray is a grayscale dataset, thus styling the colored baggage X-ray scans as grayscale would create more ambiguities between normal and abnormal baggage content, resulting in the noisier disparity maps. We did not use OPIXray (Wei et al. 2020) dataset for training purposes because it does not contain any normal baggage X-ray scans (Wei et al. 2020). Also, COMPASS-XP (Wei et al. 2020) dataset does not have complex X-ray scans (like other datasets), i.e., it contains single-item scans, and the model trained on these scans does not get much exposure towards learning diversified feature representations contained within other datasets.

6.2 Comparison with supervised frameworks

In this series of experiments, we compared the proposed framework's detection and recognition performance with the state-of-the-art fully supervised baggage threat detection frameworks. Here, our approach is semi-supervised (since we mounted the optional classification module with the proposed framework for recognizing the suspicious items' categories). The comparison is reported in Table 5

Table 5 Comparison of proposed approach (semi-supervised version) with existing fully supervised baggage threat detection frameworks in terms of mAP

Model	SIXray	GDXray	OPIXray	COMP
Proposed	0.7941	0.8591	<u>0.7483</u>	<u>0.5439</u>
TST	<u>0.9516</u>	**0.9672**	**0.7532**	**0.5842**
CST	**0.9595**	<u>0.9343</u>	–	–
DTSD	0.6457	0.9162	–	–
DOAM	–	–	0.7401	-
GBAD	0.7483	–	–	–
DOAM-O	–	–	0.7457	–

Bold indicates the best score while the second-best scores are underlined. '–' indicates that the metric is not computed. Moreover, the abbreviations are *COMP* COMPASS-XP (Griffin et al. 2019), *TST* trainable structure tensors (Hassan and Werghi 2020), *CST* cascaded structure tensors (Hassan et al. 2020a), *DTSD* dual-tensor shot detector (Hassan et al. 2020b), *DOAM* de-occlusion attention module (Wei et al. 2020) with single-shot detector (Liu et al. 2016), *GBAD* GAN based anomaly detection (Dumapgi et al. 2020) with ResNet-101 (He et al. 2016), *DOAM-O* oversampling de-occlusion attention module (Tao et al. 2021) with single-shot detector (Liu et al. 2016)

where we can see that although the proposed framework lags from some state-of-the-art methods, its performance is still quite appreciable, especially considering the fact that it is a semi-supervised approach, unlike its conventionally trained, fully supervised competitors. Also, it achieves the good detection performance (i.e., it only lags from the best performing framework by 17.23%, 11.17%, 0.65%, and 6.89% on SIXray, GDXray, OPIXray, and COMPASS-XP dataset, respectively, in terms of mAP scores). Furthermore, compared to recently introduced GBAD (Dumapgi et al. 2020), DTSD (Hassan et al. 2020b), and DOAM-O (Tao et al. 2021) approaches, the proposed framework, on SIXray and OPIXray datasets, achieve 5.76%, 18.68%, and 0.347% improvements, respectively which is quite significant.

6.3 Comparison with unsupervised frameworks

We also compared the performance of the proposed unsupervised framework with state-of-the-art methods such as GANomaly (Akçay et al. 2018a), and Skip-GANomaly (Akçay et al. 2018a). Here, the experimental protocol is to classify the abnormal vs. normal baggage X-ray scans (except for the OPIXray), where abnormal scans are those scans that contain one or more anomalous regions, and the normal scans only have the normal baggage content. For the OPIXray dataset, we followed the strategy of classifying the scans as having *folding knives*, *utility knives*, *straight knives*, *multi-tool knives*, and *scissors*, since this dataset does not contain any normal baggage X-ray scans (Wei et al. 2020). Moreover, for fairness, all the frameworks were trained on a single SIXray dataset as per the training protocol defined

Table 6 Comparison of proposed framework with state-of-the-art unsupervised baggage threat detection frameworks in terms of F1 score

Model	SIXray	GDXray	COMP	OPIXray
PF	**0.4831**	**0.7839**	**0.4119**	**0.6560**
GA	0.1227	0.4994	0.2405	0.3074
SG	<u>0.1576</u>	<u>0.5305</u>	<u>0.2232</u>	<u>0.3464</u>

Bold indicates the best scores while the second-best are underlined. Moreover, the abbreviations are: *PF* proposed framework, *GA* GANomaly (Akçay et al. 2018a), *SG* skip-GANomaly (Akçay et al. 2019), *COMP* COMPASS-XP (Griffin et al. 2019)

in Sect. 5.2. Similarly, they are also applied to the other datasets in a zero-shot manner (without any re-training or fine-tuning). The comparison is reported in Table 6 in terms of F1 scores where we can see that the proposed framework has 67.37% lead on SIXray dataset, 32.32% lead on GDXray dataset, and 45.81% lead on COMPASS-XP dataset. Furthermore, on the OPIXray dataset, the proposed framework is leading the second-best Skip-GANomaly (Akçay et al. 2019) by 47.19%.

6.4 Qualitative evaluations

Figure 6 reports the proposed framework's qualitative evaluations on all four X-ray datasets. We can see here that the proposed framework effectively recognizes the contraband items irrespective of the scanner specifications. However, for cluttered cases, the quality of extracted masks is somewhat limited. For example, see the mask of *gun* in (H). This is because the framework recognizes the anomalous regions from the disparity maps in an unsupervised manner (by clustering their color distributions w.r.t the pixels of the normal baggage content). Therefore, it cannot differentiate the anomalous items' pixel well if they have very high-intensity correlations with the normal pixels.

7 Discussion and conclusion

This paper presents a novel unsupervised anomaly instance segmentation framework to detect baggage threats from the X-ray scans without any supervision and extensive training procedures. The proposed framework is ideal for screening baggage threats in the real world, easing the security officers' load by avoiding the tedious re-training and ground truth marking process as required in the conventional baggage threat detection frameworks. The proposed framework recognizes the baggage threats as anomalies by exploiting the original and the reconstructed scans' disparities. For cluttered cases, the disparity maps are a bit limited in generating the anomalous items' masks accurately due to their lesser intensity differences
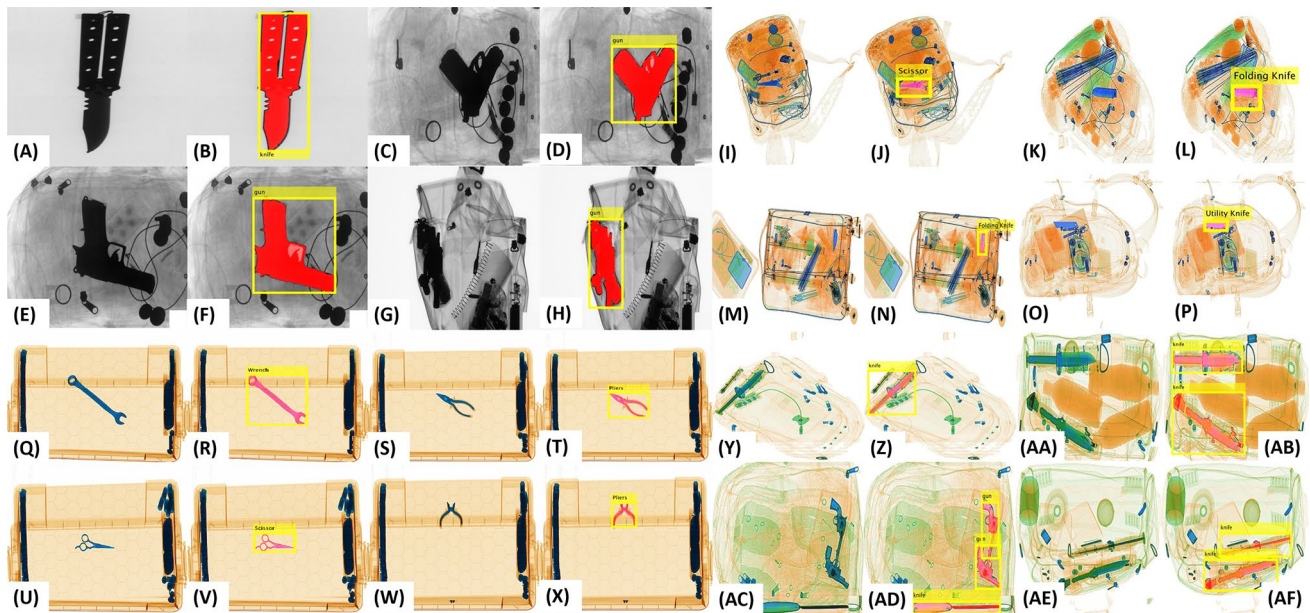
Fig. 6 Qualitative evaluation of proposed framework on four public X-ray datasets. **A–H** Show scans from the GDXray (Mery et al. 2015) dataset, **I–P** Show scans from the OPIXray (Wei et al. 2020) dataset, **Q–X** Show scans from the COMPASS-XP (Griffin et al. 2019) dataset, and **Y–AF** Show scans from the SIXray (Miao et al. 2019) dataset.

with the normal objects. In the future, we plan to improve this aspect by deploying a more adaptive attention mechanism that will highlight only the desired anomalous region within the candidate scan while suppressing the rest of the content. Also, we plan to test the proposed framework to detect 3D-printed and organic baggage threats from the security X-ray scans.

Author Contributions TH devised the idea, wrote the manuscript, and performed the experiments. SA also contributed to manuscript writing. MB co-supervised the research and reviewed the experiments. SK also reviewed the manuscript. NW supervised the complete research, contributed to manuscript writing, and reviewed the experiments.

Data Availability Statement The proposed framework has been thoroughly evaluated on four baggage X-ray datasets, and all of these four datasets are publicly available.

Declarations

References

Akçay S, Breckon T (2020) Towards automatic threat detection: a survey of advances of deep learning within X-ray security imaging. arXiv:200101293

Akçay S et al (2016) Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery. In: IEEE ICIP, pp 1057–1061

Akçay S, Atapour-Abarghouei A, Breckon TP (2018a) GANomaly: semi-supervised anomaly detection via adversarial training. In: Asian conference on computer vision. Springer, pp 622–637

Akçay S, Kundegorski ME, Willcocks CG, Breckon TP (2018b) 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

Akçay S, Atapour-Abarghouei A, Breckon TP (2019) Skip-GANomaly: skip connected and adversarially trained encoder-decoder anomaly detection. arXiv:190108954

An J et al (2019) Semantic segmentation for prohibited items in baggage inspection. In: International conference intelligence science and big data engineering. Visual data engineering, pp 495–505

Bastan M (2015) Multi-view object detection in dual-energy X-ray images. Mach Vis Appl 26:1045–1060

Bastan M et al (2011) Visual words on baggage X-ray images. In: International conference on computer analysis of images and patterns, pp 360–368

Cooley JW, Tukey JW (1965) An algorithm for the machine calculation of complex Fourier series. Math Comput 19(90):297–301

Council NR (1996) Airline passenger security screening: new technologies and implementation issues. The National Academics Press

Dumagpi JK, Jung WY, Jeong YJ (2020) A new GAN-based anomaly detection (GBAD) approach for multi-threat object classification on large-scale X-ray security images. IEICE Trans Inf Syst E103-D(2):454–458

Gaus YFA, Bhowmik N, Akçay S, Breckon T (2019a) Evaluating the transferability and adversarial discrimination of convolutional neural networks for threat object detection and classification within X-ray security imagery. In: 18th IEEE international conference on machine learning and applications (ICMLA)

Gaus YFA, Bhowmik N, Akçay S, Guillen-Garcia PM, Barker JW, Breckon TP (2019b) Evaluation of a dual convolutional neural

network architecture for object-wise anomaly detection in cluttered X-ray security imagery. In: 2019 international joint conference on neural networks (IJCNN), pp 1–8

Griffin LD, Caldwell M, Andrews JTA (2019) COMPASS-XP dataset. Computational Security Science Group, UCL

Hassan T, Werghi N (2020) Trainable structure tensors for autonomous baggage threat detection under extreme occlusion. In: Asian conference on computer vision (ACCV)

Hassan T, Bettayeb M, Akçay S, Khan S, Bennamoun M, Werghi N (2020a) Detecting prohibited items in X-ray images: a contour proposal learning approach. 27th IEEE international conference on image processing (ICIP)

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. MDPI Sens 20(22):1–25

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778

He K, Gkioxari G, Dollar P, Girshick R (2017) Mask R-CNN. In: Proceedings of the IEEE international conference on computer vision (ICCV), pp 2961–2969

Heitz G, Chechik G (2010) Object separation in X-ray image sets. In: International conference computer vision and pattern recognition, pp 2093–2100

Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) MobileNets: efficient convolutional neural networks for mobile vision applications, pp 1–9. arXiv:1704.04861

Huang G, Liu Z, Laurens VDM, Weinberger KQ (2017) Densely connected convolutional networks. In: IEEE international conference on computer vision and pattern recognition (CVPR)

Jaccard N, Rogers TW, Griffin LD (2014) Automated detection of cars in transmission X-ray images of freight containers. In: AVSS, pp 387–392

Jaccard N et al (2017) Detection of concealed cars in complex Cargo X-ray imagery using deep learning. J X-ray Sci Technol 25(3):323–339

Johnson J, Alahi A, Fei-Fei L (2016) Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision (ECCV)

Kingma DP, Ba J (2015) Adam: a method for stochastic optimization. In: Proceedings of the international conference on learning representations (ICLR)

Kundegorski M et al (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 (ICDP)

Liu W et al (2016) SSD: single shot multibox detector. In: European conference on computer vision (ECCV)

Lin TY et al (2017) Focal loss for dense object detection. In: IEEE international conference on computer vision and pattern recognition (CVPR)

Liu Z, Li J, Shu Y, Zhang D (2018) Detection and recognition of security detection object based on Yolo9000. In: 2018 5th international conference on systems and informatics (ICSAI), IEEE, pp 278–282

Megherbi N, Breckon TP, Flitton GT, Mouton A (2012) Fully automatic 3D threat image projection: application to densely cluttered 3D computed tomography baggage images. In: International conference on image processing theory, tools and applications

Mery D et al (2015) GDXray: the database of X-ray images for nondestructive testing. J Nondestr Eval 34(4):42

Mery D et al (2016) 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, Riffo V, Saavedra JM, Banerjee S (2017) Modern computer vision techniques for X-ray testing in baggage inspection. IEEE Trans Syst Man Cybern: Syst 47(4):682–692

Mery D, Saavedra D, Prasad M (2020) X-ray baggage inspection with computer vision: a survey. IEEE Access 8(19974224):145620–145633

Miao C et al (2019) SIXray: a large-scale security inspection X-ray benchmark for prohibited item discovery in overlapping images. In: IEEE CVPR, pp 2119–2128

Ren S et al (2016) Faster R-CNN: towards real-time object detection with region proposal networks. arXiv:150601497v3

Riffo V, Mery D (2015) Automated detection of threat objects using adapted implicit shape model. IEEE Trans Syst Man Cybern Syst 46(4):472–482

Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition

Sun Q, Liu Y, Chua TS, Schiele B (2019) Meta-transfer learning for few-shot learning. In: IEEE international conference on computer vision and pattern recognition (CVPR)

Szegedy C et al (2014) Going deeper with convolutions. arXiv:14094842v1

Tao R, Wei Y, Li H, Liu A, Ding Y, Qin H, Liu X (2021) Over-sampling de-occlusion attention network for prohibited items detection in noisy X-ray images, pp 1–13. arXiv:2103.00809

Turcsany D, Mouton A, Breckon TP (2013) Improving feature-based object recognition for X-ray baggage security screening using primed visual words. In: 2013 IEEE international conference on industrial technology (ICIT), IEEE, pp 1140–1145

Wang Q, Breckon TP (2020) Contraband materials detection within volumetric 3D computed tomography baggage security screening imagery. arXiv:201211753

Wang Q, Megherb N, Breckon TP (2020a) Multi-class 3D object detection within volumetric 3D computed tomography baggage security screening imagery. arXiv:200801218

Wang Q, Megherbi N, Breckon TP (2020b) A reference architecture for plausible threat image projection (TIP) within 3D X-ray computed tomography volumes. J X-ray Sci Technol 28(3):507–526

Wei Y, Tao R, Wu Z, Ma Y, Zhang L, Liu X (2020) Occluded prohibited items detection: an X-ray security inspection benchmark and de-occlusion attention module. In: Proceedings of the 28th ACM International Conference on Multimedia, pp 138–146

Xu M et al (2018) Prohibited item detection in airport X-ray security images via attention mechanism based CNN. In: Chinese conference on pattern recognition and computer vision (PRCV), pp 429–439

Yang Y, Soatto S (2020) FDA: Fourier domain adaptation for semantic segmentation. In: IEEE international conference on computer vision and pattern recognition (CVPR)

Zhang J et al (2014) Joint shape and texture based X-ray Cargo image classification. In: International conference on computer vision and pattern recognition (CVPR) workshops

Zhao Z et al (2018) A GAN-based image generation method for X-ray security prohibited items. In: Chinese conference on pattern recognition and computer vision (PRCV)