Engineering Applications of Artificial Intelligence 130 (2024) 107639

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Enhancing security in X-ray baggage scans: A contour-driven learning approach for abnormality classification and instance segmentation

Abdelfatah Ahmed [a,*,1], Divya Velayudhan [a,*,1], Taimur Hassan [b], Mohammed Bennamoun [c], Ernesto Damiani [a], Naoufel Werghi [a]

[a] KUCARS and C2PS, Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi, United Arab Emirates
[b] Department of Electrical, Computer and Biomedical Engineering, Abu Dhabi University, United Arab Emirates
[c] Computer Science and Software Engineering, University of Western Australia, Australia

ARTICLE INFO

Keywords:
Affinity Loss
Baggage X-ray Scans
Contour-driven learned model
Instance Segmentation
Abnormality classification
Data imbalance

ABSTRACT

The task of automatically identifying hazardous items in luggage through X-ray scans is highly crucial, yet immensely complex. This method presents an improvement over the traditional, labor-intensive, time-consuming, and error-prone manual evaluation of X-ray images. In the present day, an enormous volume of luggage is continually passing through airports, seaports, and land ports, making these advancements exceedingly vital. Nevertheless, numerous existing solutions grapple with the issue of data imbalance, as the frequency of dangerous items is relatively low, adversely affecting the efficiency of networks. In this study, we introduce a solution that merges a contour-driven learned model with a unique loss function known as Balanced Affinity Loss. This function equitably distributes the focus of training models towards less represented items. We tested the proposed system using three public luggage X-ray datasets, where it exceeded the performance of the most advanced methods by 34.1%, 8.64%, and 10.04% in terms of intersection-over-union for instance segmentation. Likewise, the introduced system registered improvements of 9.11% and 3.66% for abnormality classification.

1. Introduction

Detecting illegal and smuggled items in baggage is essential to ensure the safety of travelers and host countries and to lower crime rates. Manual monitoring of baggage X-ray scans at checkpoints has been the norm; however, the process is highly cumbersome and requires constant attention from experienced personnel, making it prone to human error and poor performance due to stress and fatigue. Many previous attempts have been proposed utilizing various tools in image analysis and machine learning (Jaccard et al., 2016; Mery et al., 2017). Numerous researchers have proposed autonomous frameworks capable of identifying suspicious items in baggage. The current state-of-the-art addresses computerized X-ray scan analysis employing image enhancement (Lu and Conners, 2006), classification (Yang et al., 2019), detection (Liang et al., 2018), and segmentation (Hassan et al., 2020a) paradigms. While early attempts focus on using image analysis techniques coupled with traditional machine learning (Zhang et al., 2014). Recent studies, based on advanced deep learning-based solutions, boosted state of the art to some extent. However, much of the proposed solutions suffer from the problem of data imbalance; statistically, a very low percentage of screened baggage contains threat items. For instance, out of thousands of bags scanned daily, only a handful might contain prohibited items. This imbalance leads to challenges in training where models might miss the rare, but critical, threats in favor of non-threat items. In addition, the frequency of illegal categories has a long-tail distribution, exhibiting majority representation for some and scarcity for others. Little studies paid attention in addressing the issue of class imbalance in the realm of threat item detection using X-ray images. To bridge this gap, we introduce an innovative compound Balanced Affinity loss function that leverages effective class samples. This approach aims to simultaneously enhance the distance between different categories while minimizing variations within the same category. This means that within each category, the variability of representations is minimized, ensuring that similar items are recognized as such, regardless of minor differences. Conversely, the approach maximizes the distinctiveness between categories, ensuring that items from different groups are easily distinguishable, even in the face of occasional overlap in features. By doing so, the model becomes capable of acquiring robust and discriminative features from novel samples. Moreover, most existing

* Corresponding authors.
E-mail addresses: 100059689@ku.ac.ae (A. Ahmed), 100058254@ku.ac.ae (D. Velayudhan).
[1] The corresponding authors contributed equally.

https://doi.org/10.1016/j.engappai.2023.107639
Received 10 June 2023; Received in revised form 30 October 2023; Accepted 28 November 2023
Available online 11 December 2023
0952-1976/

methodologies for threat item detection are primarily designed for color images (Fang et al., 2023; Kolte et al., 2022), disregarding the unique characteristics of X-ray scans. Unlike grayscale or colored images, X-ray scans lack textural details. For instance, a knife or gun, when seen in X-ray scans, might not show its material but rather just its shape. This makes X-ray images particularly challenging, as different items can have similar shapes but entirely different functions and threat levels. While there have been significant strides in detecting threats using X-ray images, key challenges remain unaddressed. One major challenge is the data imbalance. Specifically, X-ray scans reveal a vast number of non-threat items, while actual threats are rare, leading to a skewed dataset. This disparity in representation makes accurate threat detection more challenging. Additionally, most research has predominantly focused on color images, neglecting the nuances and specificities of X-ray scans, like their lack of rich textural details. Given these gaps, there is a pressing need for a method tailored to effectively handle the data imbalance in X-ray threat detection and to address the unique characteristics of these images.

To bridge this gap, we propose an alternative approach where we employ contour-driven representation as the input for our threat detection system, as opposed to region-based techniques. This decision stems from the insight that shape information serves as the most reliable cue in X-ray scans. For example, suspicious items created through 3D-printing, such as guns, manifest solely as outlines in X-ray scans. This is exemplified by the fact that a 3D-printed plastic gun might not have the same texture or material as a metal one but would have a similar outline that can be detected. Therefore, we propose utilizing the contour information extracted from X-ray images to detect the presence of threat items. The proposed loss function has been found to outperform its counterparts by efficiently clustering the samples through max-margin learning based on effective sample representation, as evident from Fig. 1. It can be observed that the classes cannot be clustered easily in the case of cross-entropy since the learning is biased towards the majority class. In contrast, our proposed balanced affinity loss exhibits the capability to effectively separate classes by learning equi-spaced clusters of uniform shapes for both majority and minority classes, utilizing effective samples.

This paper proposes a novel framework for detecting threat and non-threat items in the X-ray image. We suggest employing contour map representation extracted from the X-ray scan in two models: (1) A contour-driven instance segmentation for detecting and extracting individual threat items; (2) A binary classification model that classifies the X-ray scan as either containing or non-containing threat. In addition to the class imbalance challenge, we also address the fine-grained classification aspect characterizing contour-based and instance segmentation problems, which require our framework to possess a high discriminative capacity. Hence, we have developed a novel loss function that combines max-margin learning and effective sample representation to address the challenges of class imbalance and fine-grained classification. Our proposed loss function, known as the balanced affinity loss, aims to simultaneously maximize the distance between different categories and minimize variations within the same category. This is achieved by leveraging effective class samples, allowing the model to acquire robust and discriminative features from novel samples. By utilizing the balanced affinity loss during training, the frameworks exhibit improved capability in distinguishing pixels belonging to contours of the same class, while also ensuring enhanced generalization across diverse scanners.

To summarize, the main contributions of the paper are:

- A contour-map-driven framework for the threat detection and identification in X-ray scan images.
- Cost-sensitive learning maximizes the margin between highly imbalanced classes in baggage threat recognition.
- An algorithmic mitigation to data imbalance, utilizing a novel balanced affinity loss function and applying it to train threat contour-driven instance segmentation and classification models.
- A benchmarking of the data imbalance problem on three publicly accessible baggage X-ray datasets provides a comprehensive comparison with other class imbalance techniques and state-of-the-art threat detection methods.

The subsequent sections of this paper are structured as follows: Section 2 presents an extensive literature review encompassing the relevant studies in the field. In Section 3, we provide detailed information about our proposed framework. The experimental setup utilized in our study is described in Section 4. The presentation and analysis of the experimental results are presented in Section 5. Finally, in Section Section 6, we conclude the paper with remarks and discuss potential directions for future research.

2. Literature review

In this section, we provide an enumeration of conventional methods used in baggage threat detection, followed by an overview of deep learning-based approaches. Furthermore, we explore various techniques aimed at addressing the issue of class imbalance within the context of baggage threat detection. For a comprehensive survey on this topic, we recommend referring to the works of Akcay and Breckon (2020), Mery et al. (2017) and Johnson and Khoshgoftaar (2019).

2.1. Classic methods

Machine Learning (ML), a subset of artificial intelligence that enables systems to learn from data patterns and make decisions, has been foundational in traditional approaches for baggage threat detection. Conventional Machine Learning (ML) techniques have been extensively explored for item detection in baggage X-ray scans. For example, Jaccard et al. (2014) utilized a random forest classifier to identify cars based on their intensity, structural, and symmetrical characteristics observed in cargo transmission X-ray images. Similarly, Bastan et al. (Baştan, 2015) developed a learning system that employed dual-energy levels to compute low-textured key points in multi-view X-ray imaging. They also presented an X-ray classification framework that integrated SIRF, SURF, Harris, and FAST descriptors for extracting key points. These key points were subsequently utilized as input to a Bag of Words (BoW) model (Bastan et al., 2011). Riffo et al. (Riffo and Mery, 2015) proposed an Adaptive Implicit Shape Model (AISM) specifically designed for detecting various threat items in the GDXray dataset. Additionally, Mery et al. (2015) introduced a novel approach called Adaptive Sparse Representation in the field of object recognition. Similarly, Turcsany et al. (2013) proposed a framework for recognizing baggage items in X-ray scans by utilizing the SURF descriptor to extract features, which were subsequently processed through a BoW model.

2.2. Deep-learning approaches

Deep learning, an advanced branch of machine learning inspired by neural networks in the brain, has brought revolutionary changes to the domain. One significant contribution from this subfield is the use of transfer learning techniques. Transfer learning, where models pretrained for one task are repurposed for another, has found its place in threat detection, often coupled with object detection (Liu et al., 2018; Xu et al., 2018; Gaus et al., 2019a; Hassan et al., 2020b) and segmentation strategies (Gaus et al., 2019b; An et al., 2019). More recently, there has been a growing interest in using anomaly detection methods (Akcay et al., 2018; Bhowmik et al., 2019) for the detection of contraband items. Another intriguing aspect of modern deep learning research in this context is the use of attention mechanisms. These mechanisms enable models to selectively focus on specific parts of input data, mimicking human attention and ensuring more accurate predictions. Hassan et al., for instance, have made significant contributions with frameworks such as Cascaded Structure Tensors (CST) (Hassan

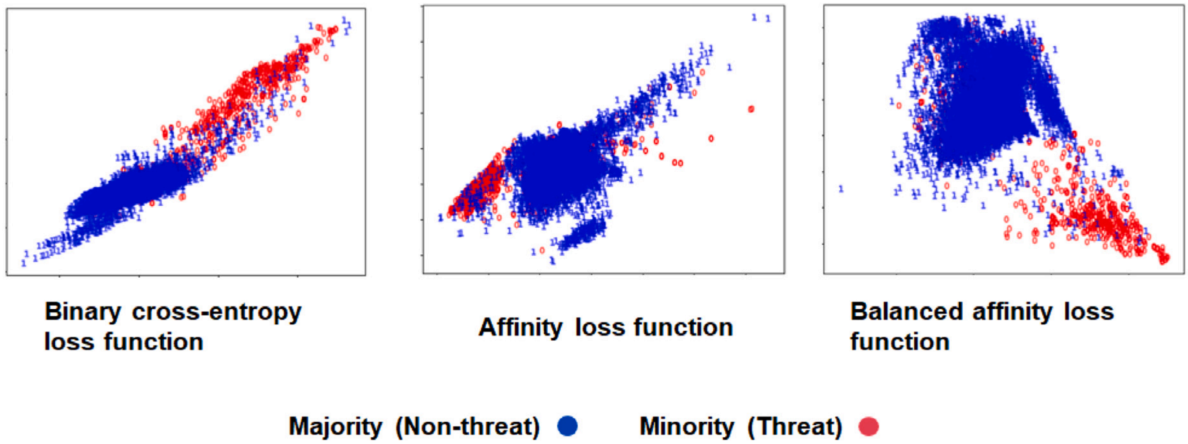


Fig. 1. The extraction of 2D latent features from a major (non-threat) and a minor (threat) class in the SIXray10 dataset (Miao et al., 2019) using three distinct loss functions: novel balanced affinity, binary cross-entropy, and affinity loss.

et al., 2020b), Trainable Structure Tensors (TST) (Hassan et al., 2020a), Tensor Pooling Driven (TPD) (Hassan et al., 2021), and dual tensor-shot detection networks (TSD) (Hassan et al., 2020c). These frameworks effectively utilize the contour information of suspicious items to achieve accurate recognition and localization. Moreover, Wei et al. (2020) introduced an innovative framework called the De-occlusion Attention Module (DOAM), which serves as a versatile module that can be seamlessly integrated with object detectors to precisely locate threatened items in X-ray scans, even in the presence of occlusion. The effectiveness of DOAM has been demonstrated and validated using the publicly available Occluded Prohibited Items X-ray (OPIXray) dataset. Recent literature accentuates the pace of innovation in this domain. For instance, Ma et al. (2022) proposed a solution for overlapping threat items. They employed a dense attention-centric instance segmentation system which adeptly adjusts course contours through high-level semantic feature utilization. In a different direction, Sara et al. (Sara and Mandava, 2023) integrated spatial and channel attentions to magnify pertinent features, using the LinkNet architectural framework for enhanced threat segmentation. Occlusion challenges, a persistent hurdle, have also been tackled by leveraging the depth offered by volumetric CT images. For example, Wang et al. (Wang and Breckon, 2021) applied a blend of U-Net and PointNet++ (Qi et al., 2017) for 3D segmentation of prohibited items, diving deep into the realm of CT scans. Further enhancing this avenue, they also explored the capabilities of Faster R-CNN (Ren et al., 2015) and RetinaNet (Lin et al., 2017a) for 3D threat detection in concealed baggage (Wang et al., 2020). In addition to the aforementioned methodologies, insights can be drawn from related fields that employ deep learning for image segmentation. For instance, the work on the Cx22 dataset for cervical cytology images introduces a comprehensive approach to image segmentation, providing meticulously annotated cellular instances and proposing baseline methods for semantic and instance segmentation tasks (Liu et al., 2022). While this dataset is tailored for medical imaging, the rigorous annotation process and the validation of deep learning-based segmentation tasks offer valuable lessons for enhancing the accuracy and reliability of threat detection in baggage X-ray scans. The attention to detail in dataset creation and the exploration of segmentation methodologies in the paper underscore the importance of high-quality data and robust validation methods, which are crucial for the development of effective deep learning models in any domain.

2.3. Data imbalance

Despite the above wealth of research, one particular issue remains understudied, namely, the data-imbalance complexity posed by publicly available datasets on the developed models, rendering many of the biased and causing a high rate of false negatives. In a setting as important as threat detection (e.g. bombs and knifes detection), a high rate of false negatives can be catastrophic. The problem of data imbalance has received considerable attention in the fields of signal processing and machine learning, resulting in a wide range of approaches that can be classified into data-level and algorithmic-level methods. Data-level methods aim to modify the distribution of training data in order to achieve equally effective classifiers for both majority and minority classes. This is commonly achieved through techniques such as random under-sampling and over-sampling (Shen et al., 2016; Krawczyk et al., 2014a; Geifman and El-Yaniv, 2017). However, these approaches offer limited additional information and are also susceptible to overfitting (Huang et al., 2016). Algorithmic-level methods (Khan et al., 2017; Krawczyk et al., 2014b; Huang et al., 2017; Ghosh and Davis, 2018) avoid these concerns by assigning a heavier penalty on misclassifying the minority class. Besides, the algorithm-level approaches aim to adapt existing classifiers to strengthen learning concerning the minority class. These methods can efficiently address the class imbalance problem with the originally available dataset without modification. Many works suggested a suitable loss function among the algorithmic-based methods and showed some improvement at handling imbalanced data effects (Hayat et al., 2019; Lin et al., 2017b; Abraham and Khan, 2019; Taghanaki et al., 2019; Cui, 2019; Salehi et al., 2017). However, little has been done to address the class imbalance in threat detection using X-ray images. To the best of our knowledge, there are limited works available addressing the issue of class imbalance in this specific context. Notable contributions include a data-level method presented in previous works (Hassan et al., 2020b; Tao et al., 2021) and a recent algorithmic method proposed by Miao et al. (2019) and Ahmed et al. (2022). In their work (Hassan et al., 2020b), Hassan et al. proposed a framework for automated detection of suspicious items in X-ray baggage scans, employing a multi-orientation object-boundaries-driven approach. Their method relies on contour-based proposals generated offline, allowing for the derivation of equal-sized training sets. Miao et al. (2019) focused on the imbalanced nature of suspicious items in real-world scenarios and proposed a novel approach called the class-balanced hierarchical framework (CHR), which involves modifications to the loss function. CHR demonstrates the ability to address the class imbalance problem. However, it has only been evaluated on a single dataset (Miao et al., 2019). Despite the extensive research and advancements in threat detection using X-ray imagery, certain critical challenges persist, hindering the efficacy of existing solutions. A predominant issue is the imbalance in data distribution, with a plethora of benign objects compared to the scarce instances of threats, resulting in datasets that are inherently skewed. This imbalance poses significant challenges in

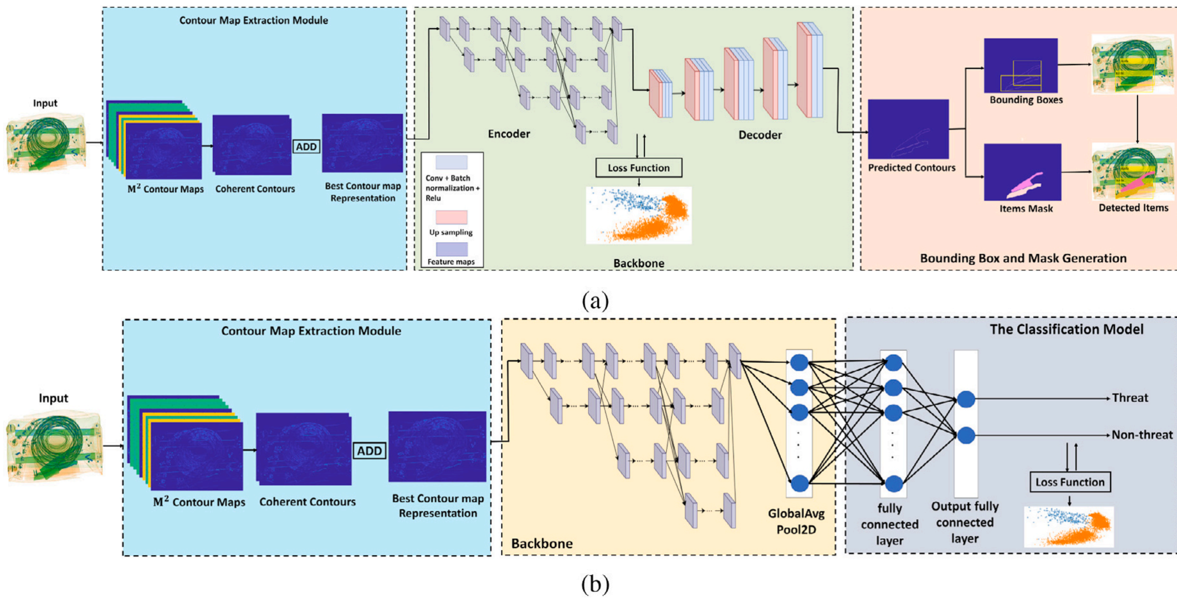


Fig. 2. Block diagram of (a) the proposed instance segmentation model, (b) and classification framework.

achieving accurate and reliable threat detection. Furthermore, the majority of existing studies have primarily concentrated on methodologies suited for color images, overlooking the unique attributes and challenges presented by X-ray imagery, such as the absence of rich textural information. Ahmed et al. addressed the challenge of data imbalances in threat detection through their notable work (Ahmed et al., 2022). They highlighted the difficulties arising from such imbalances in the aviation security domain. Their approach uniquely integrated multiple imbalanced learning strategies to enhance traditional deep learning models, making them more adept at distinguishing between regular and threat items. Despite the novel approaches advanced by various researchers in addressing data imbalances in threat detection, certain limitations persist. There is a recurring issue of these methods not being extensively validated across a wide range of datasets. This constraint limits the universal applicability and robustness of such solutions. Moreover, while adjustments to loss functions have been proposed, foundational segmentation model refinements are often overlooked, potentially affecting the overall performance. Addressing the challenges of data imbalance and the unique characteristics of X-ray imagery necessitates a novel approach. Our proposed methodology capitalizes on contour information extracted directly from X-ray images, aiming to reliably identify potential threats based on their shape. This is a departure from conventional region-based techniques, providing a more suitable solution for the specific challenges posed by X-ray threat detection. The effectiveness of our approach is further enhanced by our innovative loss function, which demonstrates superior performance in clustering samples through max-margin learning. This ensures robust sample representation and significantly improves the model's ability to detect threats in X-ray scans. Unlike traditional cross-entropy methods, which tend to be biased towards the majority class, our balanced affinity loss excels in class separation, achieving uniform and equi-spaced clustering for both majority and minority classes. In light of these advancements, our approach emerges as a robust solution, addressing the limitations of previous works and offering enhanced accuracy and precision in threat detection within X-ray images. This positions our contribution as pivotal in the domain of baggage threat detection.

3. Proposed framework

This work presents novel frameworks for automated recognition of contraband items in baggage, comprising contour-driven instance segmentation (refer to Fig. 2-a) and classification (refer to Fig. 2-b). These frameworks are trained using a unique compound loss function that specifically addresses the challenges posed by highly imbalanced datasets. The crux of our approach is rooted in the belief that the contours of items, often overlooked in traditional methodologies, hold vital clues for efficient threat detection. The aim is to achieve effective classification and detection of threatened items, which are typically associated with the minority class. This integration of deep learning methods with contour-driven insights positions our framework as a promising avenue for future advancements in the domain of baggage threat detection. The subsequent sections provide detailed explanations of the segmentation and classification frameworks, including their individual sub-modules.

3.1. Contour-driven instance segmentation model

This model (Fig. 2-a) is composed of three modules, (1) the Contour Map Extraction Module, (2) the instance contour-driven segmentation model, and (3) the Bounding box and mask Generation model.

3.1.1. Contour Map Extraction

This module is developed based on the principles of structured tensors (Kass et al., 1988), which are 2×2 symmetric positive-semidefinite matrices derived from the outer products of image gradients. These tensors capture the dominant orientations of changes in the local neighborhood of each pixel. In this context, the module utilizes a set of η orientations denoted as $\theta_1, \ldots, \theta_\eta$. The tensors are computed using the following equation:

$$“\mathcal{T}(i = 1 : \eta, j = i + 1 : \eta) = \phi * \nabla\theta_i \nabla_{\theta_j}(I)” \quad (1)$$

Here, ϕ represents a Gaussian smoothing function, and $\nabla_{\theta_i}(I)$ represents the image gradient in the θ_i direction. By applying these calculations, a set of $\mathcal{T}_k$ tensors is derived, where k ranges from 1 to $\mathcal{N} = \eta(\eta + 1)/2$. the tensor corresponds to a matrix which is second-moment, serving as a representation of transition intensity maps. These intensity maps are then summed together, resulting in the generation of a “coherent contour map”. This map gives a detailed depiction of both contraband and common baggage objects. An example output of the coherent contour map generated by this module is illustrated in Fig. 3.

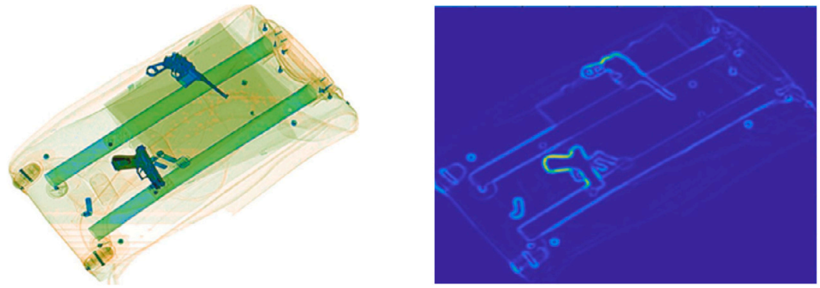


Fig. 3. An exemplar X-ray scan from the SIXray dataset and the coherent contour map derived from it.

3.1.2. Instance contour-driven segmentation

We introduce here a unique approach to instance segmentation. Unlike conventional methods that perform segmentation directly on images, we propose conducting segmentation on the coherent contour map generated earlier. Our model is trained to detect both the contours themselves and instances of contours corresponding to threat items. To facilitate this process, we have developed the contour-driven instance segmentation architecture, which consists of a High-Resolution Network (HRNet) and a decoder model. The novel Balanced Affinity Loss function is utilized during the training process of the segmentation network. (discussed in Section 3.3.3). This loss function not only enhances the segmentation performance but also addresses the challenge of class imbalance in the dataset.

3.1.3. High-Resolution Network

The image resolution plays a crucial role in extracting relevant information. Many existing advanced frameworks rely on cascaded sub-networks, connecting high-to-low resolution convolutions and then reverting to high resolution through inverse operations. This iterative process is repeated multiple times to improve model performance. However, performing undersampling and oversampling operations on the same feature map can lead to the loss of valuable information. To address this, we propose utilizing a model that can maintain high-resolution maps throughout while also leveraging low-resolution maps through parallel branches. For this purpose, we adopted the High-Resolution Network (HRNet), known for its ability to preserve high-resolution representations (Sun et al., 2019). The HRNet architecture begins with a high-resolution convolution stream and progressively adds high-to-low resolution convolution streams, connecting the multi-resolution streams in parallel. Additionally, the HRNet facilitates frequent information exchange across resolutions, resulting in a richer representation that is more spatially accurate (Yu et al., 2021).

3.1.4. Mask and bounding box generation

Following the segmentation of the contour maps of the threat items, post-processing is carried out to eliminate false positives and noisy segmented blobs. The segmented output comprises contours, both open and closed, that represent the suspicious items. To address the open contours, their endpoints are connected and utilized for mask generation using morphological reconstruction. Subsequently, bounding boxes are generated based on the created masks, as demonstrated in Fig. 4.

3.2. Classification framework

The classification framework (Fig. 2-b) task is to classify an input scan as either threat or non-threat. It encompasses the contour extraction module, and the HRNet modules described earlier. Here, we leverage again our original approach of using the coherent contour map rather than the crude X-ray scan. The HRNet, employed as a feature extractor in this model, is appended with an average pooling layer. The classification block comprises two fully connected layers to obtain the prediction. Also, we plugged the proposed loss function into this classification model.

3.3. Class imbalance remedy approach

To effectively tackle the challenges associated with fine-grained classification in the context of contour-based approach and instance segmentation, our system adopts a combination of cost-sensitive learning and Max-Margin learning techniques. For cost-sensitive learning, we utilize the Class-Balanced Loss method (Cui, 2019), which involves assigning weights to class samples to address the issue of class imbalance. However, instead of simply using inverse class frequency weighting, this approach incorporates a weighting scheme based on the "number of effective samples". This allows us to focus on the most informative subset within each class. In addition to cost-sensitive learning, we sought a method that can handle fine-grained classification while remaining robust against class imbalance. After careful evaluation, we identified the affinity loss function as a suitable choice for our approach. The affinity loss function aligns well with our requirements and helps to effectively address the challenges associated with fine-grained classification and class imbalance. In the following sections, we will provide a concise overview of these two methods, highlighting their relevance and significance within our approach.

3.3.1. Class-balanced loss

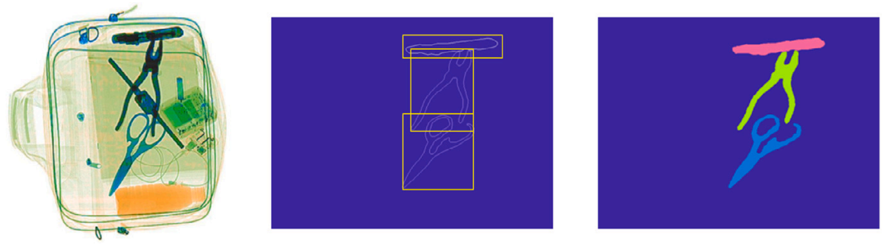
In the technique previously referenced (Cui, 2019), the authors introduced the concept of "Effective Number of Samples", a measure representing the smallest group within a class that encompasses the most relevant examples, while ignoring repetitive or minorly varied instances. To better understand this idea, let us think about a class containing pictures of a person's front and side profiles. Imagine we produce 24 pictures from each perspective, using simple data enhancement methods such as image flipping or noise introduction. Despite the total count of enhanced samples in this class reaching 50, the effective sample count is merely two. This is because the most instructive subset is any set of front and side view pictures, and surplus images do not offer substantial extra information. The class-balanced method (Cui, 2019) approximates the Effective Number of Samples (E_{n_y}) for a specific class (y) using the equation:

$$\text{“}E_{n_y} = \frac{1-\beta^{n_y}}{1-\beta}\text{”} \tag{2}$$

In this equation, n_y represents the number of samples in the ground-truth class y, and β is a parameter that controls the growth rate of the effective number as the sample count increases. Although β is typically expected to be specific to each class, the authors assume it to be the same for all classes for simplicity. The class-balanced loss ($\mathcal{CB}$) is weighted by E_{n_y} instead of the raw sample count, as seen in traditional methods. This weighting is expressed as:

$$\text{“}\mathcal{CB}(\mathbf{p}, y) = \frac{1}{E_{n_y}}\mathcal{L}(\mathbf{p}, y).\text{”} \tag{3}$$

In this equation, $\mathcal{CB}$ represents the class-balanced loss, $\mathcal{L}$ denotes the chosen loss function, and $\mathbf{p}$ represents the predicted probability for the class y. It is important to note that the class-balanced term ($\mathcal{CB}$) is independent of the specific choice of the loss function ($\mathcal{L}$) and can be applied in conjunction with any loss function.



(a) (b) (c)

Fig. 4. (a) An example X-ray scan from the SIXray dataset; (b) generated bounding boxes; (c) item masks.

3.3.2. Affinity loss

The Affinity loss function, presented in Hayat et al. (2019), fuses the tasks of classification and clustering using a max-margin method along with Gaussian Affinity. This differs from the Softmax loss function, which relies on the internal vector product $\langle w, f \rangle$ between class weights w and feature vectors f at the final fully connected layer. Instead, the Affinity loss function employs Gaussian similarity measures grounded on the Bergman divergence (Hayat et al., 2019). This divergence-based method offers a more thorough similarity evaluation, enabling a more efficient learning process. To tackle the issue of class imbalance and foster the distribution of intra-class centroids, the authors integrate diversity regularization and multi-centered partitioning terms into the loss function. These added elements help ensure a more evenly distributed set of instances within each class and improve the distinction between different classes, thus enhancing the model's discriminative ability. Let us consider X_i, y_i as the input–output pairs, with C representing the number of classes and N denoting the number of training samples. The feature space representation of the input samples is denoted as f_i for $i = 1 : N$, and the class weights are represented by w_j for $j = 1 : C$. The Affinity loss function is defined as follows:

$$“\mathcal{L}a = \mathcal{L}mm + R(w), \tag{4}$$

where

$$\mathcal{L}mm = \sum_j max(0, \lambda + d(f_i, w_j) - d(f_i, w_{y_i})), \ j \neq y_i, \tag{5}$$

and

$$R(w) = E[(\|w_j - w_k\|^2 - \mu)^2], j < k.” \tag{6}$$

In these equations, $i \in [1, N]$ and $j \in [1, C]$. The term $d(f_i, w_j)$ represents the similarity with the true class, $d\left(f_i, w_{y_i}\right)$ denotes the similarity of the sample with other classes, and λ represents the enforced margin. The Affinity loss function effectively combines classification and clustering principles, allowing for robust feature learning and improved performance in the presence of class imbalance. In addition, Eq. (6) introduces a 'diversity regularizer' to tackle the soft-max loss's challenge of not ensuring uniform classification regions. This regularizer promotes uniform dispersion of class centers (w) in the feature space by penalizing deviations from an average distance, μ.

3.3.3. Proposed loss function

To harness the advantages offered by the class-balanced loss and the affinity loss function, we propose a fusion of these two methodologies. This integration is achieved by combining Eqs. (3) and (4), resulting in the formulation of a novel loss function known as the Class Balanced Affinity (CBA) loss function:

$$\mathcal{CBA}(\mathbf{p}, y) = \frac{1}{E_{n_y}} \mathcal{L}_a(\mathbf{p}, y) \tag{7}$$

This novel loss function, the CBA loss, integrates the principles of class balancing from the class-balanced loss and the similarity-based clustering of the affinity loss. It addresses the challenges of class imbalance while promoting fine-grained classification. By combining the benefits of these two approaches, the CBA loss function aims to enhance the overall performance of the system in instance segmentation and threat detection tasks.

4. Experimental setup

In this section, we detail the experimental setup employed to evaluate our proposed framework. We begin with a comprehensive overview of the datasets utilized, namely SIXray, OPIXray, and COMPASS XP. Following this, we delve into the specific implementation details, describing the computational infrastructure, training parameters, and data partitioning strategy. Lastly, the performance metrics adopted to assess the efficacy of our approach.

4.1. The utilized datasets

Our framework was thoroughly evaluated on three well-known datasets: SIXray, OPIXray, and COMPASS XP. The subsequent sections will provide a comprehensive description of these datasets.

4.1.1. SIXray

The Security Inspection X-ray dataset, also known as SIXray (Miao et al., 2019), is a significant benchmark in the field of security inspection. It consists of over a million X-ray images, making it the largest dataset available. The dataset is designed to reflect real-world scenarios, with less than 1% of the scans containing threats. Within the SIXray dataset, there are 1,059,231 scans that encompass six major categories of prohibited items, namely guns, knives, pliers, wrenches, scissors, and hammers. Each scan in the dataset is accompanied by detailed bounding box annotations for accurate object localization. To facilitate analysis and evaluation, the dataset is further divided into three sub-classes: SIXray10, SIXray100, and SIXray1000. The division is based on the proportion of anomalies to normal baggage scans within each sub-class.

4.1.2. OPIXray

The Occluded Prohibited Items X-ray dataset, known as OPIXray, focuses on the detection of prohibited items in baggage scans. It consists of 8,885 scans containing five categories of prohibited items, namely folding knives, straight knives, utility knives, multi-tool knives, and scissors. The dataset is specifically designed to include varying levels of occlusion within the X-ray scans, which provides a realistic representation of the challenges faced in threat detection. To facilitate evaluation and analysis, the OPIXray dataset is divided into three sub-classes: OL1, OL2, and OL3, based on the extent of occlusion present in the scans. Additionally, the dataset provides box-level annotations, enabling the assessment of object detection frameworks for precise localization of the prohibited items.

4.1.3. COMPASS-XP

Compass-XP (Griffin et al., 2019) dataset is quite different from other datasets in the domain, with different representations (low-energy and high-energy X-ray, density, color, grayscale and RGB images) for the same baggage scan. This amounts to a total of 11,568 scans, with each representation comprising 1928 images. 258 scans of them contain threat items. This disproportion also infers a class imbalance problem. The dataset has been primarily designed for the evaluation of classification frameworks.

4.2. Implementation details

The proposed system was deployed on a computer system equipped with an Intel(R) Core(TM) i7-10750H CPU running at a base frequency of 2.60 GHz. The system configuration included 24 GB of RAM and a single NVIDIA GTX 1660Ti GPU, providing computational support for the training process. Through empirical analysis, several training protocols were examined to find an optimal configuration for our specific task. The primary goal was to identify settings that resulted in efficient learning while also avoiding overfitting. For the segmentation framework, the ADADELTA optimizer (Zeiler, 2012) was employed to optimize the training procedure. The training was conducted over a specific number of epochs, with a batch size of 16. The learning rate was set to 1.0, and a decay rate of 0.95 was applied to adapt the learning rate during training. For the classification task, we utilized the Adam optimizer (Kingma and Ba, 2017), selected after empirical evaluations showed its superior performance, especially in handling sparse gradients. Similar to the segmentation framework, the training was carried out over a predefined number of epochs, with a batch size of 32. The learning rate was set to 1.0, and a dropout rate of 0.1 was applied to prevent overfitting. Additionally, a decay rate of 0.95 was incorporated to adjust the learning rate as the training progressed. In terms of data splitting, the available dataset was divided into training and testing sets. Specifically, 80% of the samples were allocated for training, allowing the model to learn from the majority of the data. The remaining 20% of the samples were reserved for testing, enabling the evaluation of the trained model's performance on unseen data. Furthermore, within the training set, 10% of the samples were set aside for validation. This validation set served as a means to monitor and assess the model's performance during the training process, providing insights into its generalization capabilities and assisting in making informed decisions regarding hyperparameter tuning and model selection.[2]

4.3. Performance metrics

To evaluate the proposed approach and compare it with existing works, we employed standard metrics for instance segmentation and classification. These metrics include Intersection-over-Union, Dice Coefficient, accuracy, mean Average Precision, and F1-score.

5. Results

This section presents the results acquired through the proposed framework's evaluation on three publicly available datasets. Moreover, we report an ablation study through which we examined different loss functions and determined the optimal backbone for baggage threat detection along with the value of the parameter β. Furthermore, we report a detailed comparative study with state-of-the-art frameworks.

[2] The source code for the proposed framework can be accessed and downloaded from the following GitHub repository: https://github.com/taimurhassan/balanced_affinity.

5.1. Ablation study

The ablation study in this paper intends to examine different class imbalance loss functions and backbones for (1) The classification and (2) The contour-driven instance segmentation. The group of class imbalance loss functions we compared our proposed loss function (7) to, includes the Affinity Loss (Hayat et al., 2019), the Focal Loss (Lin et al., 2017b), the Tversky Loss (Salehi et al., 2017), the Focal Tversky Loss (Abraham and Khan, 2019), and the Combo Loss (Taghanaki et al., 2019). In addition, different values of the β have been studied in order to improve the performance of the proposed CBA loss function (7). Before presenting our results, it is crucial to note our choice of metrics. For segmentation, we opted for the robust μDC and μIOU scores, which gauge the alignment between predictions and ground truth. For classification, accuracy was prioritized for its clarity in interpretation, and the F1 score was used to counteract class imbalances. These metrics collectively reflect our model's performance across various datasets and configurations.

5.1.1. Selection of the β value

We conducted the first series of experiments to find the optimal $beta$ value. To that end, we evaluated the balanced affinity loss function with different beta values on three tasks, namely, the Contour-driven Instance Segmentation, abnormality classification, and multi-label classification over the SIXray, OPIXray, and COMPASS-XP datasets. We followed the protocol recommended in Cui (2019) for the range and the mode of β variation. From the obtained results, reported in Fig. 5, we found that $\beta = 0.99$ produces the highest performance (In terms of μIOU and accuracy for segmentation and classification, respectively). We adopted this value for β for the rest of the experiments.

5.1.2. Contour-driven instance segmentation

Table 1 presents the performance results achieved using different loss functions on the three datasets. It can be observed that for the SIXray10 dataset, the CBA loss function achieves an μDC score of 0.9782 and an μIOU score of 0.9425, surpassing the second-best loss function by 3.93% and 4.65%, respectively. On the SIXray100 dataset, the CBA loss function achieves the highest performance with scores of 0.9155 for μDC and 0.8966 for μIOU, outperforming the closest competitor by 0.42% and 2.19%, respectively. Moreover, on the SIXray1000 subset, the CBA loss function shows improvements of 3.8% and 2.3% on the two metrics, respectively. Similar trends are observed on COMPASS-XP and OPIXray datasets, where the CBA loss function consistently outperforms other loss functions across both metrics.

5.1.3. Classification

In this set of experiments, we assessed the performance of the classification model with respect to (1)Using the CBA loss function, as compared with the group of class imbalance loss functions, (2) The HRNet model as compared with a group of other backbone CNN models which include ResNet50 (He et al., 2016), EfficientNetB0 (Tan and Le, 2019), VGG16 (Simonyan and Zisserman, 2014), NASnet (Zoph et al., 2017) and InceptionV3 (Szegedy et al., 2016). We conducted the experiments with SIXray and COMPASS-XP datasets only because the OPIXray dataset does not contain non-threat items. The results of these experiments are grouped in Tables 2 and 3 showing, respectively, the accuracy and the F1 scores across the different configurations. From Table 2, The following observations can be drawn from these results: (1) For the two datasets, the classification model scores the best performance across all the other backbone models with the proposed CBA loss function except at InceptionV3 where the CBA loss produced the second-best score lagging the affinity loss slightly. (2) With the CBA loss function, the HRNet produces the best performance over all the other backbone models. From Table 3, we notice the following (1) For the Sixray dataset, the classification model scores the best performance across all the other backbone models with the proposed CBA loss

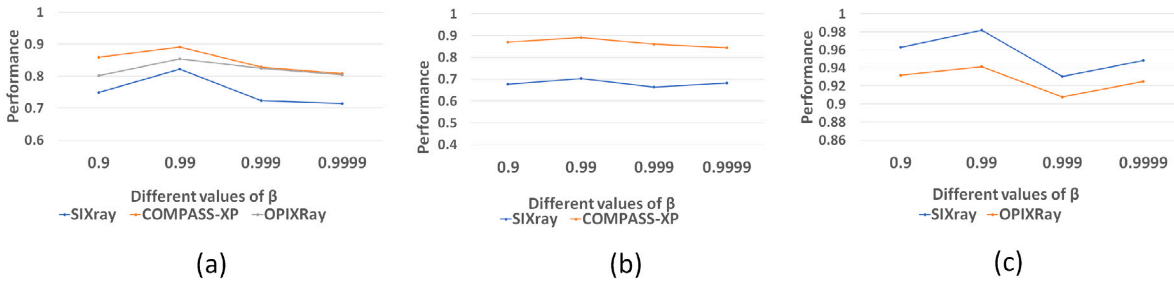


Fig. 5. Performance comparison of balanced affinity loss function for (a) instance segmentation; (b) abnormality classification; and (c) multi-label classification with different β values. Please zoom-in for best visualization.

Table 1
Performance comparison of different loss functions for instance segmentation. Values in bold denote the highest values for each metric.

Loss function	Subsets	μIoU	μDC
Proposed	**SIXray10**	**0.9425**	**0.9782**
	SIXray100	**0.8966**	**0.9155**
	SIXray1000	**0.8227**	**0.8772**
	OPIXray	**0.8547**	**0.9473**
	COMPASS-XP	**0.8911**	**0.9257**
Affinity	SIXray10	0.9006	0.9412
	SIXray100	0.8774	0.9117
	SIXray1000	0.7936	0.8572
	OPIXray	0.8405	0.9165
	COMPASS-XP	0.8600	0.9177
Focal	SIXray10	0.8914	0.9265
	SIXray100	0.8584	0.9009
	SIXray1000	0.7918	0.8221
	OPIXray	0.8071	0.8962
	COMPASS-XP	0.8536	0.9055
Tversky	SIXray10	0.8898	0.9226
	SIXray100	0.8643	0.9077
	SIXray1000	0.7935	0.8240
	OPIXray	0.8074	0.8966
	COMPASS-XP	0.8412	0.8905
Focal Tversky	SIXray10	0.8962	0.9239
	SIXray100	0.8897	0.9113
	SIXray1000	0.8025	0.8489
	OPIXray	0.8360	0.9088
	COMPASS-XP	0.8744	0.8925
Combo	SIXray10	0.8560	0.8814
	SIXray100	0.8374	0.8545
	SIXray1000	0.7198	0.7732
	OPIXray	0.8044	0.8827
	COMPASS-XP	0.7801	0.8501

function except for InceptionV3 at the SIXray1000 subset and VGG16 at SIXRay10 subset, where it comes second after the affinity loss. (2) For the COMPASS-XP dataset, the classification model scores the best performance across all the other backbone models with the proposed CBA loss function except for ResNet50 and InceptionV3, where the affinity loss shows a slightly better score. (3) The HRNet produces the best performance over all the other backbone models with the CBA loss function. The above observations concur towards confirming the HRNet model with the proposed CBA loss function as the best combination. At InceptionV3 where the CBA loss produced the second-best score lagging the affinity loss slightly. The classification model scores the best performance with the HRNet.

Apart from this, a second test was conducted to evaluate the performance of various imbalanced loss functions for classifying different threat items. The classification results using these imbalanced loss functions are presented in Table 4. The SIXray and OPIXray datasets were utilized for this test, with each dataset containing different threat items. The results in Table 3 demonstrate that the proposed loss function exhibited superior performance compared to the affinity loss function for the SIXray dataset. The proposed loss function achieved a mAP of 0.9820, surpassing the affinity loss function by 0.85%. Similarly, for the OPIXray dataset, the proposed balanced affinity loss function achieved a mAP of 0.9416, resulting in a 0.33% improvement compared to other loss functions.

5.2. Comparative studies

Contour-driven Instance Segmentation Results: We evaluated and compared the performance of the proposed segmentation framework, which utilizes the $\mathcal{L}_{cba}$ loss function, with existing threat detection systems. The results are summarized in Table 5. It can be observed that the proposed method outperforms the second-best scheme by 4.84% and 7.76% in terms of μDC and μIoU, respectively, on the SIXray10 dataset. On the SIXray100 dataset, it leads the state-of-the-art by 1.36% and 8.88% in terms of μDC and μIoU, respectively. Additionally, significant performance improvements of 15.36% and 34.1% in μDC and μIoU, respectively, were obtained on the imbalanced SIXray1000 dataset. Similarly, on the OPIXray dataset, the proposed system achieved improvements of 8.35% and 10.04% in μDC and μIoU, respectively. Moreover, on the COMPASS-XP dataset, it outperformed the state-of-the-art by 2.66% and 8.64% in terms of μDC and μIoU, respectively. These enhancements demonstrate the efficacy of the proposed framework in effectively detecting contraband data in scenarios with high class imbalance. It is worth mentioning that the proposed framework achieved the second-best performance on the entire SIXray dataset, outperforming TSD (Hassan et al., 2020c) and scoring slightly lower than CST (Hassan et al., 2020b) in terms of μDC and μIoU. Moreover, the qualitative evaluation of the proposed framework on SIXray, OPIXray, and COMPASS-XP dataset is presented in Fig. 6.

Classification Results: To assess the effectiveness of our proposed compound loss function, we have compared its performance with existing approaches. We have utilized the softmax layer as the baseline method for comparison in the abnormality classification, and this is because there are no state-of-the-art methods for this aspect. Moreover, Fig. 7 emphasizes the results of the proposed balanced affinity and softmax loss employing the HRNet model and based on SIXray and COMPASS-XP datasets. It can be observed that conjugating our proposed loss function with any network has a remarkable impact on improving classification accuracy and F1-score. For instance, for the highest imbalance level of the SIXray dataset, our proposed loss function has surpassed the softmax by 9.11% and 11.98% regarding accuracy and F1-score, respectively. Furthermore, based on COMPASS-XP, the balanced affinity loss outperforms the softmax by 3.66% and 3.53%. Also, it can be observed that our proposed loss function has shown its capability to leverage the performance regardless of the imbalance problem for all the different networks employed.

Apart from this, the multi-classification of the threat items has been compared with two state-of-the-art methods, namely, CHR (Miao et al., 2019), and SXMNet (Hu et al., 2020). Here, it should be noticed that the state-of-the-art approaches' protocols have been followed. Besides, Table 6 highlights the results of this comparison. As shown, we demonstrate that our method achieved the best results across almost

Table 2
The accuracy of abnormality classification is evaluated using various loss functions and network architectures. The highest accuracy values are indicated in bold, representing the superior performance across different configurations.

Loss function	Subsets	ResNet50	InceptionV3	VGG16	EfficientNet	NASNet	HRnet
Proposed	**SIXray10**	**0.9171**	**0.8938**	**0.9047**	**0.8972**	**0.9017**	**0.9172**
	SIXray100	**0.8280**	**0.7988**	**0.8293**	**0.8017**	**0.8057**	**0.8344**
	SIXray1000	**0.6974**	0.6432	**0.6957**	0.6500	**0.6533**	**0.7031**
	COMPASS-XP	**0.8754**	0.8445	**0.8656**	**0.8576**	**0.8619**	**0.8916**
Affinity Loss	SIXray10	0.9047	0.8797	0.9031	0.8621	0.8449	0.9093
	SIXray100	0.8232	0.7810	0.8125	0.7654	0.7501	0.8156
	SIXray1000	0.6885	**0.6705**	0.6890	**0.6571**	0.6439	0.6912
	COMPASS-XP	0.8729	**0.8599**	0.8567	0.8427	0.8258	0.8856
Focal Loss	SIXray10	0.8844	0.8859	0.8750	0.8780	0.8699	0.8891
	SIXray100	0.8046	0.7715	0.7972	0.7646	0.7576	0.8106
	SIXray1000	0.6783	0.6698	0.6761	0.6638	0.6577	0.6800
	COMPASS-XP	0.8609	0.8349	0.8619	0.8275	0.8199	0.8776
Tversky Loss	SIXray10	0.9005	0.8839	0.8974	0.8758	0.8796	0.8957
	SIXray100	0.8160	0.7915	0.8111	0.7842	0.7876	0.8088
	SIXray1000	0.6653	0.6609	0.6572	0.6548	0.6577	0.6785
	COMPASS-XP	0.8625	0.8446	0.8514	0.8368	0.8405	0.8764
Focal Tversky Loss	SIXray10	0.8969	0.8807	0.8879	0.8732	0.8773	0.8561
	SIXray100	0.8216	0.7823	0.7909	0.7756	0.7792	0.7989
	SIXray1000	0.6691	0.6684	0.6619	0.6627	0.6658	0.6802
	COMPASS-XP	0.8646	0.8540	0.8625	0.8467	0.8507	0.8795
Combo Loss	SIXray10	0.8984	0.8885	0.8896	0.8975	0.8664	0.8989
	SIXray100	0.8008	0.7943	0.8269	0.8023	0.7745	0.8047
	SIXray1000	0.6581	0.6534	0.6577	0.6600	0.6371	0.6650
	COMPASS-XP	0.8621	0.8525	0.8610	0.8611	0.8313	0.8748

Table 3
The F1 score performance on the SIXray and COMPASS-XP datasets is analyzed using different loss functions and network architectures for abnormality classification. The highest F1 score values are highlighted in bold, indicating the superior performance across the various configurations.

Loss function	Subsets	ResNet50	InceptionV3	VGG16	EfficientNet-B2	NASNET	HRnet
Proposed	**SIXray10**	**0.8881**	**0.8667**	0.8762	**0.8685**	**0.8755**	**0.8986**
	SIXray100	**0.8026**	**0.7720**	**0.8030**	**0.7760**	**0.7823**	**0.8162**
	SIXray1000	**0.6756**	0.6232	**0.6737**	0.6292	**0.6343**	**0.6926**
	COMPASS-XP	0.8497	0.8275	**0.8555**	**0.8302**	**0.8369**	**0.8714**
Affinity Loss	SIXray10	0.8771	0.8529	**0.8845**	0.8345	0.8204	0.8823
	SIXray100	0.7964	0.7561	0.7868	0.7409	0.7283	0.7918
	SIXray1000	0.6512	**0.6482**	0.6567	**0.6361**	0.6253	0.6593
	COMPASS-XP	**0.8549**	**0.8428**	0.8329	0.8157	0.8019	0.8669
Focal Loss	SIXray10	0.8569	0.8512	0.8321	0.8499	0.8447	0.8602
	SIXray100	0.7812	0.7455	0.7677	0.7402	0.7356	0.7942
	SIXray1000	0.6333	0.6312	0.6451	0.6426	0.6387	0.6690
	COMPASS-XP	0.8389	0.8097	0.8303	0.8010	0.7961	0.8413
Tversky Loss	SIXray10	0.8815	0.8561	0.8690	0.8478	0.8541	0.8877
	SIXray100	0.7925	0.7641	0.7631	0.7591	0.7648	0.7859
	SIXray1000	0.6512	0.6394	0.6384	0.6339	0.6386	0.6671
	COMPASS-XP	0.8353	0.8125	0.8245	0.8101	0.8161	0.8589
Focal Tversky Loss	SIXray10	0.8677	0.8561	0.8694	0.8452	0.8518	0.8381
	SIXray100	0.7999	0.7580	0.7714	0.7508	0.7567	0.7718
	SIXray1000	0.6415	0.6298	0.6316	0.6415	0.6465	0.6452
	COMPASS-XP	0.8341	0.8387	0.8417	0.8196	0.8260	0.8593
Combo Loss	SIXray10	0.8720	0.8510	0.8643	0.8688	0.8413	0.8796
	SIXray100	0.7895	**0.7751**	0.8019	0.7766	0.7521	0.7983
	SIXray1000	0.6344	0.6381	0.6305	0.6389	0.6187	0.6374
	COMPASS-XP	0.8304	0.8396	0.8315	0.8336	0.8072	0.8462

all categories of SIXray and OPIXray datasets. When comparing our proposed method with the existing approaches, our method reached the most significant improvement performance, i.e. 1.45% and 3.67% in terms of mAP for SIXray and OPIXray, respectively.

6. Conclusion

In this study, we introduce novel frameworks for contour-driven instance segmentation and classification, aiming to address the challenge of imbalanced distributions in contraband data within baggage X-ray imagery. These frameworks incorporate a unique loss function called the balanced affinity loss, which leverages the class balanced factor to enhance the affinity loss function's capacity. This enhancement enables the model to learn equi-spaced clusters of uniform size in the feature space, leading to improved class separability and reduced intra-class disparities. The integrated loss function is applied within single-staged contour-driven instance segmentation and classification frameworks, significantly enhancing their ability to detect and recognize cluttered, concealed, and highly imbalanced prohibited items in X-ray imagery. By harnessing the strengths of the balanced affinity

Fig. 6. Proposed framework qualitative results. The first, second, and third row represent scans from SIXray, OPIXray, and COMPASS-XP datasets, respectively.

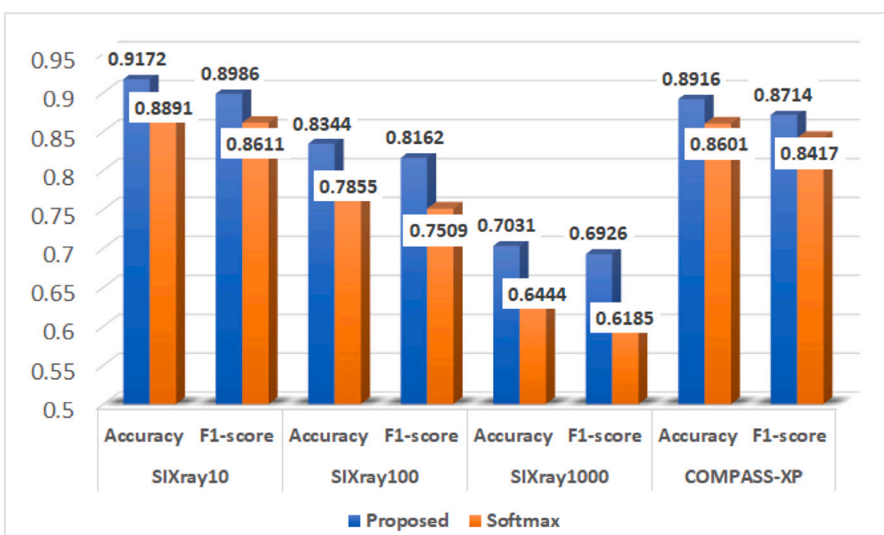


Fig. 7. Performance of the introduced balanced affinity loss compared with the softmax loss that is considered as the baseline method for abnormality classification.

loss, our frameworks demonstrate enhanced performance in handling complex scenarios commonly encountered in baggage threat detection tasks. In our extensive testing on three publicly available datasets, our frameworks outperformed state-of-the-art solutions by significant margins. Specifically, we achieved performance improvements of 34.1%, 8.64%, and 10.04% in terms of intersection-over-union for instance segmentation. For abnormality classification tasks, our proposed system reported gains of 9.11% and 3.66%. The power of the balanced affinity loss function shines through in our results, especially as our frameworks adeptly handle the intricacies and challenges commonly encountered in baggage threat detection tasks. As we look to the future, we are keen on further refining our approach, particularly focusing on the development

Table 4
Performance of classification task based on different loss functions for multi-label classification. Values in boldface denote the highest values for each metric in terms of mAP.

Dataset	Category	Affinity	Tversky	Focal Tversky	Focal	**Proposed**
SIXray	Gun	0.9796	0.9283	0.9561	0.8509	**0.9844**
	Knife	0.9766	0.9395	0.9632	0.9014	**0.9980**
	Pliers	**0.9866**	0.8995	0.9549	0.8127	0.9685
	Scissors	0.9642	0.9658	0.9672	0.9258	**0.9795**
	Wrench	0.9614	0.9144	0.9312	0.8391	**0.9795**
	Mean	0.9737	0.9295	0.9545	0.8660	**0.9820**
OPIXray	Folding	0.9584	0.9237	0.9369	0.7766	**0.9715**
	Straight	**0.9505**	0.9420	0.9460	0.8687	0.8009
	Scissor	0.8331	0.9404	0.9366	0.7969	**0.9897**
	Utility	0.9466	0.9267	0.9525	0.8068	**0.9719**
	Multi-tool	0.8631	0.8982	0.9204	0.8016	**0.9740**
	Mean	0.9103	0.9262	0.9385	0.8101	**0.9416**

Table 5
Performance contrast between our proposed model and other leading threat detection frameworks, best results are marked in bold.

Method	Dataset	μIoU	μDC
Proposed	**SIXray10**	**0.9425**	**0.9782**
	SIXray100	**0.8966**	**0.9155**
	SIXray1000	**0.8227**	**0.8772**
	SIXray*	**0.9521**	**0.9797**
	OPIXray	**0.8547**	**0.9473**
	COMPASS-XP	**0.8911**	**0.9257**
TST	SIXray10	0.8746	0.9330
	SIXray100	0.8235	0.9032
	SIXray1000	0.6135	0.7604
	OPIXray	0.7767	0.8743
	COMPASS-XP	0.8202	0.9014
CST	SIXray*	0.9689	0.9842
TSD	SIXray*	0.9238	0.9603

* All SIXray dataset samples are utilized.

Table 6
Performance of the suggested system compared with state-of-the-art for multi-label X-ray classification frameworks. Bold indicates the best scores in terms of mAP.

Dataset	Category	CHR	SXMNet	**Proposed**
SIXray	Gun	0.9829	**0.9884**	0.9844
	Knife	0.9487	0.9522	**0.9980**
	Pliers	0.9640	**0.9833**	0.9685
	Scissors	0.9175	0.9611	**0.9795**
	Wrench	0.8851	0.9538	**0.9795**
	Mean	0.9396	0.9678	**0.9820**
OPIXray	Folding	0.9462	0.9611	**0.9715**
	Straight	0.6742	0.7537	**0.8009**
	Scissor	0.9893	**0.9934**	0.9897
	Utility	0.8039	0.8039	**0.9719**
	Multi-tool	0.9728	0.9728	**0.9740**
	Mean	0.8773	0.9083	**0.9416**

of a practical framework that can estimate the effective number of samples, aiming to optimize the performance of the balanced affinity loss function even further.

Funding

The research conducted in this study is backed by a research funds from Khalifa University, Ref: CIRA-2021-052, and the Advanced Technology Research Center Program (ASPIRE), Ref: AARE20-279.

CRediT authorship contribution statement

Abdelfatah Ahmed: Conceptualization, Methodology, Software, Data curation, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Divya Velayudhan:** Conceptualization, Methodology, Software, Data curation, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Taimur Hassan:** Conceptualization, Methodology, Software, Data curation, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Mohammed Bennamoun:** Supervision, Writing – review & editing, Project administration. **Ernesto Damiani:** Supervision, Writing – review & editing, Project administration. **Naoufel Werghi:** Supervision, Writing – review & editing, Project administration.

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.

Data availability

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

References

Abraham, N., Khan, N.M., 2019. A novel focal tversky loss function with improved attention u-net for lesion segmentation. In: 2019 IEEE 16th International Symposium on Biomedical Imaging. ISBI 2019, IEEE, pp. 683–687.

Ahmed, A., Velayudhan, D., Hassan, T., Hassan, B., Dias, J., Werghi, N., 2022. Baggage threat detection under extreme class imbalance. In: 2022 2nd International Conference on Digital Futures and Transformative Technologies. ICoDT2, pp. 1–6. http://dx.doi.org/10.1109/ICoDT255437.2022.9787472.

Akcay, S., Atapour-Abarghouei, A., Breckon, T.P., 2018. Ganomaly: Semi-supervised anomaly detection via adversarial training. In: Asian Conference on Computer Vision. Springer, pp. 622–637.

Akcay, S., Breckon, T.P., 2020. Towards automatic threat detection: A survey of advances of deep learning within X-ray security imaging, CoRR abs/2001.01293. arXiv:2001.01293.

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

Baştan, M., 2015. Multi-view object detection in dual-energy X-ray images. Mach. Vis. Appl. 26 (7), 1045–1060.

Bastan, M., Yousefi, M., Breuel, T., 2011. Visual words on baggage X-Ray images. ISBN: 978-3-642-23671-6, pp. 360–368. http://dx.doi.org/10.1007/978-3-642-23672-3_44.

Bhowmik, N., Gaus, Y.F.A., Akçay, S., Barker, J.W., Breckon, T.P., 2019. On the impact of object and sub-component level segmentation strategies for supervised anomaly detection within x-ray security imagery. In: 2019 18th IEEE International Conference on Machine Learning and Applications. ICMLA, IEEE, pp. 986–991.

Cui, Y., 2019. Class-balanced loss based on effective number of samples. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9268–9277.

Fang, C., Liu, J., Han, P., Chen, M., Liao, D., 2023. FSVM: A few-shot threat detection method for X-ray security images. Sensors 23 (8), 4069.

Gaus, Y.F.A., Bhowmik, N., Akcay, S., Breckon, T.P., 2019a. Evaluating the transferability and adversarial discrimination of convolutional neural networks for threat object detection and classification within X-Ray security imagery. arXiv:1911.08966.

Gaus, Y.F.A., Bhowmik, N., Akçay, S., Guillen-Garcia, P.M., Barker, J.W., Breckon, T.P., 2019b. Evaluation of a dual convolutional neural network architecture for object-wise anomaly detection in cluttered X-ray security imagery. arXiv:1904.05304.

Geifman, Y., El-Yaniv, R., 2017. Deep active learning over the long tail. arXiv preprint arXiv:1711.00941.

Ghosh, P., Davis, L.S., 2018. Understanding center loss based network for image retrieval with few training data. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops.

Griffin, L., Caldwell, M., Andrews, J., 2019. COMPASS-XP Dataset. Computational Security Science Group, UCL.

Hassan, T., Akcay, S., Bennamoun, M., Khan, S., Werghi, N., 2020a. Trainable structure tensors for autonomous baggage threat detection under extreme occlusion. arXiv:2009.13158.

Hassan, T., Akçay, S., Bennamoun, M., Khan, S., Werghi, N., 2021. Tensor pooling-driven instance segmentation framework for baggage threat recognition. Neural Comput. Appl. 1–12.

Hassan, T., Bettayeb, M., Akçay, S., Khan, S., Bennamoun, M., Werghi, N., 2020b. Detecting prohibited items in X-Ray images: a contour proposal learning approach. In: 2020 IEEE International Conference on Image Processing. ICIP, pp. 2016–2020. http://dx.doi.org/10.1109/ICIP40778.2020.9190711.

Hassan, T., et al., 2020c. Meta-transfer learning driven tensor-shot detector for the autonomous localization and recognition of concealed baggage threats. Sensors 20 (22), 6450.

Hayat, M., et al., 2019. Gaussian affinity for max-margin class imbalanced learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6469–6479.

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.

Hu, B., Zhang, C., Wang, L., Zhang, Q., Liu, Y., 2020. Multi-label X-ray imagery classification via bottom-up attention and meta fusion. In: Proceedings of the Asian Conference on Computer Vision.

Huang, C., Li, Y., Loy, C.C., Tang, X., 2016. Learning deep representation for imbalanced classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5375–5384.

Huang, C., Loy, C.C., Tang, X., 2017. Discriminative sparse neighbor approximation for imbalanced learning. IEEE Trans. Neural Netw. Learn. Syst. 29 (5), 1503–1513.

Jaccard, N., Rogers, T.W., Griffin, L.D., 2014. Automated detection of cars in transmission X-ray images of freight containers. In: 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance. AVSS, pp. 387–392. http://dx.doi.org/10.1109/AVSS.2014.6918699.

Jaccard, N., Rogers, T.W., Morton, E.J., Griffin, L.D., 2016. Tackling the X-ray cargo inspection challenge using machine learning. In: Anomaly Detection and Imaging with X-Rays (ADIX), Vol. 9847. SPIE, pp. 131–143.

Johnson, J., Khoshgoftaar, T., 2019. Survey on deep learning with class imbalance. J. Big Data 6, 27. http://dx.doi.org/10.1186/s40537-019-0192-5.

Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: Active contour models. Int. J. Comput. Vis. 1 (4), 321–331.

Khan, S.H., Hayat, M., Bennamoun, M., Sohel, F.A., Togneri, R., 2017. Cost-sensitive learning of deep feature representations from imbalanced data. IEEE Trans. Neural Netw. Learn. Syst. 29 (8), 3573–3587.

Kingma, D.P., Ba, J., 2017. Adam: A method for stochastic optimization. arXiv:1412.6980.

Kolte, S., Bhowmik, N., Dhiraj, 2022. Threat object-based anomaly detection in X-ray images using GAN-based ensembles. Neural Comput. Appl. 1–16.

Krawczyk, B., Woźniak, M., Schaefer, G., 2014a. Cost-sensitive decision tree ensembles for effective imbalanced classification. Appl. Soft Comput. 14, 554–562.

Krawczyk, B., Woźniak, M., Schaefer, G., 2014b. Cost-sensitive decision tree ensembles for effective imbalanced classification. Appl. Soft Comput. 14, 554–562.

Liang, K.J., et al., 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) III, Vol. 10632. International Society for Optics and Photonics, 1063203.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P., 2017a. Focal loss for dense object detection. In: IEEE Conference on Computer Vision and Pattern Recognition.

Lin, T.-Y., et al., 2017b. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2980–2988.

Liu, G., Ding, Q., Luo, H., Sha, M., Li, X., Ju, M., 2022. Cx22: A new publicly available dataset for deep learning-based segmentation of cervical cytology images. Comput. Biol. Med. 150, 106194.

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.

Lu, Q., Conners, R.W., 2006. Using image processing methods to improve the explosive detection accuracy. IEEE Trans. Syst. Man Cybern C (Appl. Rev.) 36 (6), 750–760.

Ma, B., Jia, T., Su, M., Jia, X., Chen, D., Zhang, Y., 2022. Automated segmentation of prohibited items in X-ray baggage images using dense de-overlap attention snake. IEEE Trans. Multimed..

Mery, D., Svec, E., Arias, M., 2015. Object recognition in baggage inspection using adaptive sparse representations of X-ray images. In: Image and Video Technology. Springer, pp. 709–720.

Mery, D., Svec, E., Arias, M., Riffo, V., Saavedra, J.M., Banerjee, S., 2017. Modern computer vision techniques for X-Ray testing in baggage inspection. IEEE Trans. Syst. Man Cybern. Syst. 47 (4), 682–692. http://dx.doi.org/10.1109/TSMC.2016.2628381.

Miao, C., et al., 2019. Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2119–2128.

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In: International Conference on Neural Information Processing Systems. NIPS.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. Adv. Neural Inf. Process. Syst. 28.

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.

Salehi, S.S.M., et al., 2017. Tversky loss function for image segmentation using 3D fully convolutional deep networks. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp. 379–387.

Sara, D., Mandava, A.K., 2023. MC-CDPNet: Multi-channel correlated detail preserving network for X-Ray-based baggage screening. J. Nondestruct. Eval. 42 (2), 47.

Shen, L., Lin, Z., Huang, Q., 2016. Relay backpropagation for effective learning of deep convolutional neural networks. In: European Conference on Computer Vision. Springer, pp. 467–482.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Sun, K., Xiao, B., Liu, D., Wang, J., 2019. Deep high-resolution representation learning for human pose estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5693–5703.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2818–2826.

Taghanaki, S.A., et al., 2019. Combo loss: Handling input and output imbalance in multi-organ segmentation. Comput. Med. Imaging Graph. 75, 24–33.

Tan, M., Le, Q.V., 2019. EfficientNet: Rethinking model scaling for convolutional neural networks, CoRR abs/1905.11946. arXiv:1905.11946.

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. arXiv preprint arXiv:2103.00809.

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

Wang, Q., Bhowmik, N., Breckon, T.P., 2020. On the evaluation of prohibited item classification and detection in volumetric 3D computed tomography baggage security screening imagery. In: 2020 International Joint Conference on Neural Networks. IJCNN, pp. 1–8. http://dx.doi.org/10.1109/IJCNN48605.2020.9207389.

Wang, Q., Breckon, T.P., 2021. Contraband materials detection within volumetric 3D computed tomography baggage security screening imagery. In: 2021 20th IEEE International Conference on Machine Learning and Applications. ICMLA, pp. 75–82. http://dx.doi.org/10.1109/ICMLA52953.2021.00020.

Wei, Y., et al., 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., Zhang, H., Yang, J., 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, Springer, pp. 429–439.

Yang, J., et al., 2019. Data augmentation for X-ray prohibited item images using generative adversarial networks. IEEE Access 7, 28894–28902.

Yu, C., Xiao, B., Gao, C., Yuan, L., Zhang, L., Sang, N., Wang, J., 2021. Lite-HRNet: A lightweight high-resolution network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 10440–10450.

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

Zhang, J., et al., 2014. Joint shape and texture based X-ray cargo image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 266–273.

Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V., 2017. Learning transferable architectures for scalable image recognition, CoRR abs/1707.07012. arXiv:1707.07012.