electronics

MDPI

Article

SC-YOLOv8: A Security Check Model for the Inspection of Prohibited Items in X-ray Images

Li Han, Chunhai Ma, Yan Liu *, Junyang Jia and Jiaxing Sun

School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China; hanli@zzuli.edu.cn (L.H.); 332207050698@email.zzuli.edu.cn (C.M.); 332207050657@email.zzuli.edu.cn (J.J.); 332207010619@email.zzuli.edu.cn (J.S.)
* Correspondence: lyanzju@zzuli.edu.cn; Tel.: +86-136-6386-9878

Abstract: X-ray package security check systems are widely used in public places, but they face difficulties in accurately detecting prohibited items due to the stacking and diversity of shapes of the objects inside the luggage, posing a threat to personal safety in public places. The existing methods for X-ray image object detection suffer from low accuracy and poor generalization, mainly due to the lack of large-scale and high-quality datasets. To address this gap, a novel large-scale X-ray image dataset for object detection, LSIray, is provided, consisting of high-quality X-ray images of luggage and objects of 21 types and sizes. LSIray covers some common categories that were neglected in previous research. The dataset provides more realistic and rich data resources for X-ray image object detection. To address the problem of poor security inspection, an improved model based on YOLOv8 is proposed, named SC- YOLOv8, consisting of two new modules: CSPnet Deformable Convolution Network Module (C2F_DCN) and Spatial Pyramid Multi-Head Attention Module (SPMA). C2F_DCN uses deformable convolution, which can adaptively adjust the position and shape of the receptive field to accommodate the diversity of targets. SPMA adopts the spatial pyramid head attention mechanism, which can utilize feature information from different scales and perspectives to enhance the representation ability of targets. The proposed method is evaluated through extensive experiments using the LSIray dataset and comparisons with the existing methods. The results show that the method surpasses the state-of-the-art methods on various indicators. Experimenting using the LSIray dataset and the OPIXray dataset, our SC-YOlOv8 model achieves 82.7% and 89.2% detection accuracies, compared to the YOLOv8 model, which is an improvement of 1.4% and 1.2%, respectively. The work not only provides valuable data resources, but also offers a novel and effective solution for the X-ray image security check problem.

Keywords: security check; prohibited items; YOLOv8

check for updates

Citation: Han, L.; Ma, C.; Liu, Y.; Jia, J.; Sun, J. SC-YOLOv8: A Security Check Model for the Inspection of Prohibited Items in X-ray Images. *Electronics* **2023**, *12*, 4208. https://doi.org/10.3390/electronics12204208

Academic Editor: Hyunjin Park

Received: 22 September 2023
Revised: 8 October 2023
Accepted: 9 October 2023
Published: 11 October 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The X-ray package security check system is a common security measure in public places, such as airports, subways and railway stations. It can scan the objects in the luggage and detect prohibited items, such as knives, bullets, guns and explosives [1–5]. However, due to the complexity of X-ray images and the phenomenon of object occlusion, manual inspection often struggles to accurately identify potentially dangerous items and suffers from a low stability and accuracy, which poses a considerable risk to public safety. Therefore, developing a method for detecting prohibited items that are in closed luggage and that are fast, accurate and automated to effectively assist inspectors in X-ray image analysis is an extremely important task.

In recent years, with the development of deep learning [6] and computer vision techniques [7], several researchers have tried to apply these techniques to the problem of security checks in X-ray images and have made some progress. However, the existing datasets and methods still suffer from many shortcomings, such as small datasets, low quality, incomplete categories, neglected but common categories (such as lighters, powerbanks

and liquid bottles) and poor detection performance due to multiple objects of different shapes being stacked chaotically at different angles in a closed suitcase or when the suitcase appears to be reversed.

To address these issues, a new dataset named Large-Scale X-ray Image Dataset for Object Detection (LSIray) is proposed. LSIray contains many high-quality X-ray images covering 21 categories and sizes of baggage and objects. In terms of categories, it is more comprehensive and includes categories that have been overlooked in previous studies, such as lighters, powerbanks and liquid bottles. The dataset provides a more realistic and richer data resource on the issue of object detection in security checks. To address the poor detection problem, an improved model based on an attention mechanism and deformable convolutional fusion of YOLOv8 is proposed. The deformable convolution module can adaptively adjust the receptive field position and shape to accommodate targets with different shapes and angles, while the attention mechanism module allows the model to utilize different levels and aspects of feature information simultaneously, thus increasing the robustness and accuracy of object detection in complex scenarios. The following is a summary of this paper's main contributions:

(1) A new dataset (LSIray) is provided, which contains high-quality X-ray images of 21 types and sizes of baggage and objects, covering some common categories that have been neglected in previous studies.
(2) A new CSPnet Deformable Convolution Network (C2F_DCN) module based on deformable convolution has been introduced to the model. It can adaptively change the position and shape of the receptive field to accommodate targets with different shapes and angles, improving the localization capability of the model.
(3) To enhance the robustness and accuracy of target recognition in complicated scenarios, a Spatial Pyramid Multi-Head Attention module (SPMA) is introduced to the model. This module enables the model to simultaneously utilize different levels and aspects of feature information.
(4) Numerous experiments on the LSIray dataset are conducted, and the method is contrasted with other methods, showing that this method has significant advantages in various metrics.

2. Related Work

2.1. X-ray Security Image Datasets and Benchmarks

X-ray plays a crucial role in various domains, including medical imaging analysis [8–10] and security inspection [11,12]. In fact, it is challenging to obtain X-ray images, leading to a lack of specialized high-quality datasets for security inspection in computer vision. This limitation hinders the progress of the research in this field. To address this issue and enhance the development of prohibited item detection methods, some datasets have been created in previous studies. As shown in Table 1, the public release of the Dbf6 Dataset [13], Dbf3 Dataset [14] and SASC Dataset [15] is still pending. However, there are four datasets that have already been published, namely, GDXray [16], SIXray [11], OPIXray [17] and CLCXray [18].

To study the multi-class detection problem, the DBf6 dataset has images of six types of objects, namely, firearms, firearm parts, knives, ceramic knives, cameras and laptops. The challenge of the DBf3 dataset is complex backgrounds and contains three types of prohibited items: firearms, firearm components and knives. The SASC dataset contains only two types of prohibited items, scissors and aerosols, and the background is relatively simple. SIXray contains six categories of control devices, in which the number of positive samples is much smaller than the number of negative samples, which is closer to the real-world situation. In order to study the effect of sample imbalances on the results, SIXray also provides three subsets, namely, SIX10, SIX100 and SIX1000, where the number after "SIX" represents the ratio of negative samples to positive samples. CLCXray contains 12 categories, including five types of tools and seven types of containers. The OPIxray dataset focuses on cutting tools and has five categories. The

test set provides three subsets, OL1, OL2 and OL3, corresponding to the cases of slight or no occlusion, partial occlusion, and severe or complete occlusion, respectively.

Table 1. Our LSIray dataset compared to the existing X-ray benchmarks. Classification, object detection, and instance segmentation are each represented by C, O, and I.

Dataset	Year	Classes	Baggage Image	Annotated	Type	Application	Availability
GDXray [16]	2015	3	8150	1552	Real	C + O	√
Dbf6 [13]	2017	6	11,627	11,627	Real	C + O	×
Dbf3 [14]	2018	3	7603	7603	Real	C + O	×
SASC [15]	2018	2	3250	3250	Unknown	C + O	×
SIXray10 [11]			88,372				
SIXray100 [11]	2019	6	882,802	8929	Real	C + O	√
SIXray1000 [11]			1,054,231				
OPIXray [17]	2020	5	882,802	8885	Synthetic	C + O	√
CLCXray [18]	2022	12	1,054,231	9565	Simulated and Real	C + O	√
LSIray (Ours)	2023	21	8885		Real	C + O + I	√

2.2. Object Detection

Two-stage object detection algorithms aim to generate region proposals from the input image first and then perform classification and regression on each region proposal to obtain the final object category and location. Region proposals are generated by RCNN [19] using selective search; then, cropping, scaling, feature extraction and SVM [20] classification are performed on each region proposal; and finally, the detection results are fine-tuned with bounding box regression. Great progress has been made by RCNN algorithm, but it is slow in computation and complex in training process. To solve these problems, a spatial pyramid pooling layer was added by SPPNet [21] on the basis of RCNN, avoiding repeated cropping and scaling operations, which was 20 times faster than the RCNN algorithm and also improved the accuracy. Region proposal generation and feature extraction were combined into one step by Fast RCNN [22]; improving on SPPNet, the SVM classifier was discarded and the SoftMax classifier was used instead, and a multi-task loss function was designed. As a result, Fast RCNN is 200 times faster than the SPPNet algorithm. Selective search [23] was replaced by a Region Proposal Network by Faster RCNN [24]; it achieves end-to-end learning of region proposal generation, convolutional features are shared and an alternating training strategy is adopted. The development of two-stage object detection did not stop there, and Faster RCNN was improved and extended by many algorithms, such as R-FCN [25], introducing position-sensitive score maps, making the detection head more lightweight, dividing each region proposal into k × k sub-regions and performing position-sensitive classification and regression. FPN (Feature Pyramid Network) [26] uses multi-scale features to detect different sizes of objects and performs RPN and detection head on each pyramid layer. Mask RCNN [27] introduces a segmentation branch, outputting the mask of the object, replacing ROI pooling layer with ROI align layer and solving the pixel misalignment problem. Cascade RCNN [28] introduces the cascade structure, dividing the detection process into three stages; each stage has an independent detection head, and the region proposal regression threshold gradually increases for each stage.

One-stage object detection algorithms directly predict the category and location of objects from the input image, without generating region proposals. The SSD [29] algorithm was the first end-to-end real-time deep learning object detection algorithm that used multi-scale feature maps to predict different sizes of objects and the anchor mechanism to predefine bounding boxes of different shapes and ratios. Good speed and accuracy are achieved by SSD, but it is not good enough for small object detection. As it can be seen in the experimental section, the accuracy of the SSD algorithm in the object detection of small objects (mAP_s) is the lowest of all the algorithms. To solve these problems, a series of improved algorithms were proposed by subsequent researchers. The YOLO series [30–33] was the first algorithm that treated object detection as a single regression problem. The

input image is divided into grids by the YOLO algorithm, and a bounding box and a category are predicted for each grid. Multiple versions of updates are also provided by the YOLO algorithm. RetinaNet algorithm was proposed by T. Lin et al., which is the first algorithm that uses the Focal Loss [34] function to solve the problem of positive and negative sample imbalance. Multi-scale features are fused by the RetinaNet algorithm using FPN, and bounding boxes and categories are predicted using the anchor mechanism.

X-ray security inspection is a challenging scenario because the prohibited items in the images often have occlusion and overlap problems. To solve these problems, many works have been conducted by previous researchers. For example, Wei et al. [17] proposed a de-occlusion module to deal with the occlusion of prohibited items; however, DOAM makes the algorithm pay more attention to highly occluded objects, and the performance is poor for non-occluded ones. Miao et al. [11] proposed a hierarchical refinement method that divides the image into three layers—foreground, prohibited items and background—to solve the overlap problem. Tao et al. [35] proposed a lateral inhibition module (LIM) similar to the attention mechanism, which ignores useless information and focuses on useful information to deal with the problem of objects overlapping with each other; however, the performance is poor for small-target prohibited items. Zhao et al. [18] introduced a label-aware approach (LA) for addressing the issue of overlap, leveraging label correlations to enhance prediction results. Liu et al. [36] introduced spatial attention into YOLOv7, which extracts the dependency of low-level feature maps to improve the detection accuracy. In addition, there are few studies that focus on multiple objects of different shapes that are randomly stacked at different angles in a closed suitcase; especially when the suitcase is inverted, the detection performance will decrease significantly. Therefore, we propose a novel YOLOv8 algorithm that integrates deformable convolution and a multi-head attention mechanism to enhance the detection of various prohibited items under complex scenarios.

3. The LSIray Dataset

The performance of deep learning models is considerably impacted by the dataset's quality, particularly when it comes to the demanding and crucial task of object detection in security checks. A high-quality dataset should have sufficient quantity, diversity and accuracy of annotations, which can reflect the real-world scenarios and challenges of object detection in security inspections and thus provide a fair and reliable benchmark for model evaluation. This section provides an overview of the LSIray dataset that we constructed, including information on how the data were collected and their statistical characteristics.

3.1. Data Collection

To construct a more comprehensive and diverse dataset, this paper collected multiple existing datasets from different sources and fused them into a new dataset, called LSIray. The LSIray dataset was collected in different scenarios, such as airports and subway stations. The dataset covers a wide range of categories that are relevant for security inspections using X-ray images, such as knives, guns and liquids, which are common in other public datasets. Moreover, the dataset also includes some novel categories that are not found in other datasets, such as powerbanks and glass bottles. These categories are very common and important in real-world security inspection scenarios; so, the dataset can better reflect the complexity and diversity of the real world. Compared to other datasets, the dataset also has a higher quality and quantity. It has more images, more accurate annotations, a more balanced distribution and richer attributes.

3.2. Data Statistics

LSIray, an X-ray image dataset for prohibited item detection, consists of 37,106 images with 21 categories of prohibited items, and is the largest and most comprehensive dataset of its kind to date. The categories include Blade, Scissor, Knife, Dagger, SwissArmyKnife, PlasticBottle, Cans, VacuumCup, GlassBottle, CartonDrinks, Tin, SprayCans, Hammer, Wrench, Gun, HandCuffs, Baton, Pliers, Powerbank, Lighter and Bullet. These items are

frequently forbidden or restricted in public areas, such as subway stations and airports. Each category contains at least 200 images distributed across training, validation and test sets in a ratio of 8:1:1 (29,686 images for the training set and 3710 images for the validation and test sets each). Detailed annotation information is provided for each image, including the item's category, location, size, orientation, material and occlusion level. Figure 1 shows the image and annotation quantities for each category. The dataset is characterized by high quality, quantity, difficulty and challenge due to the complex interactions and occlusion relationships among the items as well as the large variations in shape, size and material of the items.

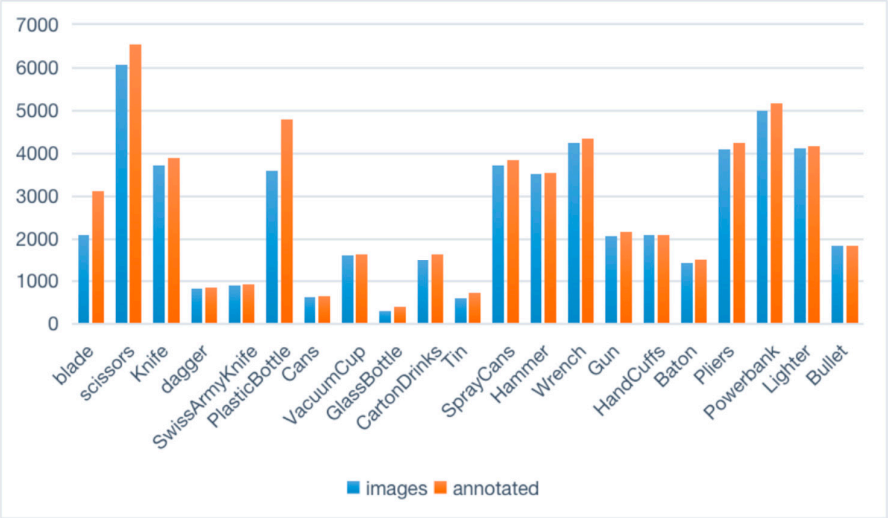

Figure 1. The quantity of images and annotations available for each category in the LSIray dataset.

4. The Proposed SC-YOLOv8 Algorithm

YOLOv8 is a state-of-the-art one-stage object detector that integrates the best practices of previous YOLO versions and other popular methods. However, it still faces some challenges in detecting prohibited items in security checks, which include complex scenes with various shapes, sizes, materials, poses and occlusions of objects. Based on the data analysis and experimental results, the following reasons for the inaccurate detection were identified: (a) Objects in the luggage have different shapes, and the targets of multiple scales and angles are mixed together, when the luggage may be reversed. When they are piled, becoming dense and chaotic baggage, normal convolutional networks cannot adapt to the different shapes. (b) Compared with normal sizes, when baggage is dense, small items are more difficult to recognize and locate in an image due to the fact they are more likely to be partially blocked by larger objects and to overlap with other objects.

To address the above problems, an improved YOLOv8 detection algorithm is proposed, as shown in Figure 2, which utilizes deformable convolution and Self-Attention Pyramid Fusion to enhance the representation and understanding of objects in X-ray images. Firstly, the C2F_DCN module is proposed, a module for the fusion of high-level and low-level features, where deformable convolution [37] is used instead of standard convolution. It can split the input channel into two parts, then repeat the Bottleneck_DCN module on one of the parts and finally stitch all the outputs together. This allows the separation and fusion of features and improves the efficiency and performance of the network. The convolution kernel's sampling position on the feature map can be adaptively adjusted by the network in response to variations in the shape and pose of the input object. The benefit of this is that it can preserve more detailed information of the object and reduce the risk of information loss and blurring. Then, the final layer of the feature map undergoes a multi-head self-attention operation to calculate the similarity between each pixel and other pixels and assign different attention weights based on the similarity. This captures more

dependencies and relationships between pixels and improves the understanding of objects by the network.

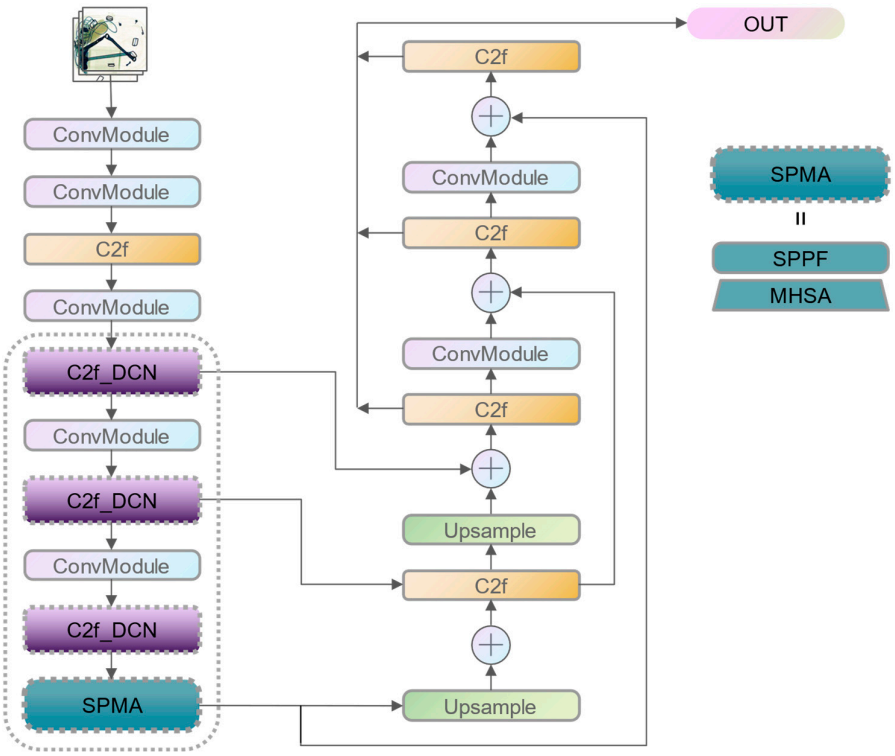

Figure 2. Improved YOLOv8 network framework.

4.1. Adaptation of Modules with Variable Shapes

The CSPnet Deformable Convolution Network (C2F_DCN) module is proposed, which is a novel feature fusion module that can adapt to the shape and pose changes in the objects in security checks. As shown in Figure 3, the C2F_DCN module consists of four sub-modules: the Deformable Bottleneck with SiLu (DBSModule), the Split, the Darknet-Bottleneck and the Concat. The C2F_DCN module splits the input feature map into two parts along the channel dimension. One part goes through the DBSModule, which applies a deformable convolution, a batch normalization and a SiLu activation function to extract low-level features. The other part goes through the DarknetBottleneck, which contains several deformable bottleneck blocks to extract high-level features. The final output feature map is generated by concatenating the DBSModule and DarknetBottleneck outputs along the channel dimension. DarknetBottleneck has two kinds of structures, as shown in Figure 4. In DarknetBottleneck, deformable convolution also replaces standard convolution. However, not all standard convolutions are replaced, and deformable convolutions are used only when the size of the convolution kernel is 3×3. If the size of the convolution kernel is 1×1 or another value, standard convolution is still used. The reasons are due to: (a) A 3×3 convolution kernel is one of the more common and effective sizes that improve a network's depth and expressiveness while keeping its receptive field and number of parameters unchanged. Deformable convolution can further enhance the flexibility and adaptability of the 3×3 convolution kernel, allowing it to capture more detailed information about the objects. (b) If the size of the convolution kernel is too small or too large, using deformable convolution may have negative effects. A 1×1 convolution kernel can only operate on a single pixel and has no room for spatial variation; thus, using deformable convolution is pointless. If the size of the convolution kernel is 5×5 or larger, using deformable convolution may increase the network's complexity and computational cost and may cause overfitting or gradient vanishing problems. In summary, using deformable

convolution only when the convolution kernel is 3×3 is a more appropriate choice as it balances the network's complexity, stability, flexibility and robustness.

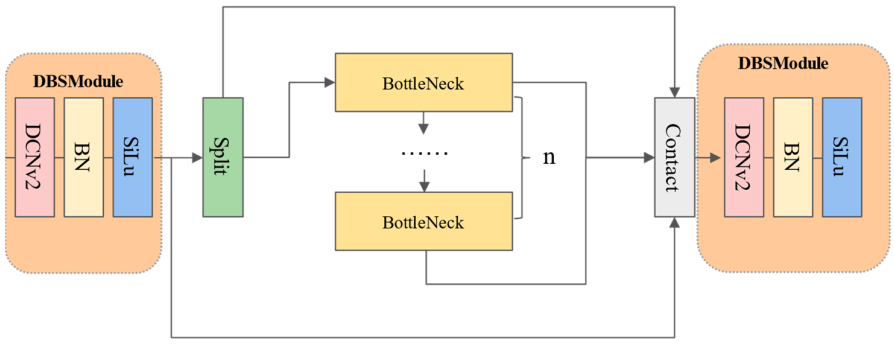

Figure 3. The structure of the C2f_DCN module, which consists of four sub-modules: DBSModule, Split, DarknetBottleneck and Contact.

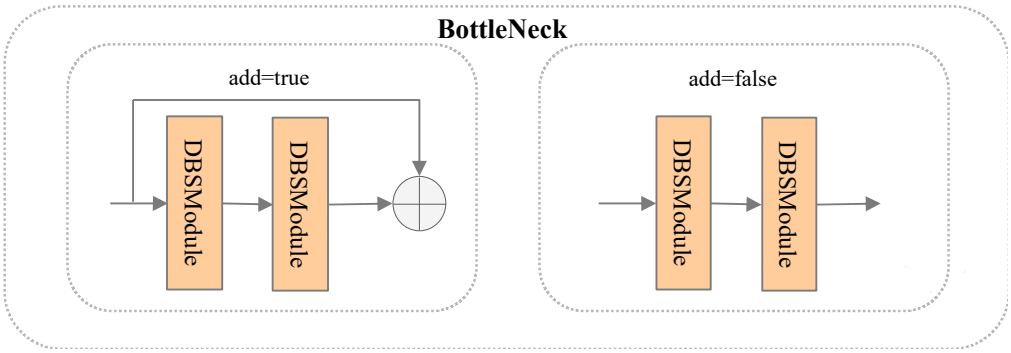

Figure 4. The Bottleneck can be divided into two distinct structures, A and B, based on the true or false value of "add". $\oplus$ means contact.

4.2. Spatial Pyramid Multi-Head Attention Module

A novel module, named Spatial Pyramid Multi-Head Attention (SPMA), is proposed, which aims to enhance the diversity and complexity of features in the backbone network of YOLOv8 and to adapt to the targets of different shapes and angles in X-ray images. The SPMA module combines the Multi-Head Self-Attention (MHSA) mechanism, which can capture the long-range dependencies and relationships between pixels, and the Spatial Pyramid Pooling with Fusion (SPPF) module, which has the ability to extract multi-scale features from different levels of the feature pyramid. The SPMA module consists of two sub-modules: the SPPF sub-module and MHSA sub-module, as shown in Figure 2. The SPPF sub-module applies spatial pyramid pooling to different scales of the feature maps and fuses them together to obtain a more rich and diverse feature representation. The MHSA sub-module applies the multi-head self-attention mechanism on the concatenated feature map, thus utilizing feature information from different levels and aspects and increasing the diversity and complexity of target detection. First, four scales are divided from the input feature map, and max pooling is performed on each scale. Next, concatenation along the channel dimension is performed after upsampling the pooled feature maps to match the size of the input feature map. The concatenated feature map is next subjected to a 1×1 convolution layer in order to minimize the channel count and fuse features from various scales. The MHSA sub-module is illustrated in Figure 5, where three separate 1×1 convolution layers are used on an input feature map to generate query (Q), key (K) and value (V) matrices. These matrices are then divided into N heads for applying the scaled dot-product self-attention (SDPA) mechanism individually within each head. Finally, these N heads are re-concatenated, followed by a 1×1 convolution layer, which results in obtaining the final output feature map.

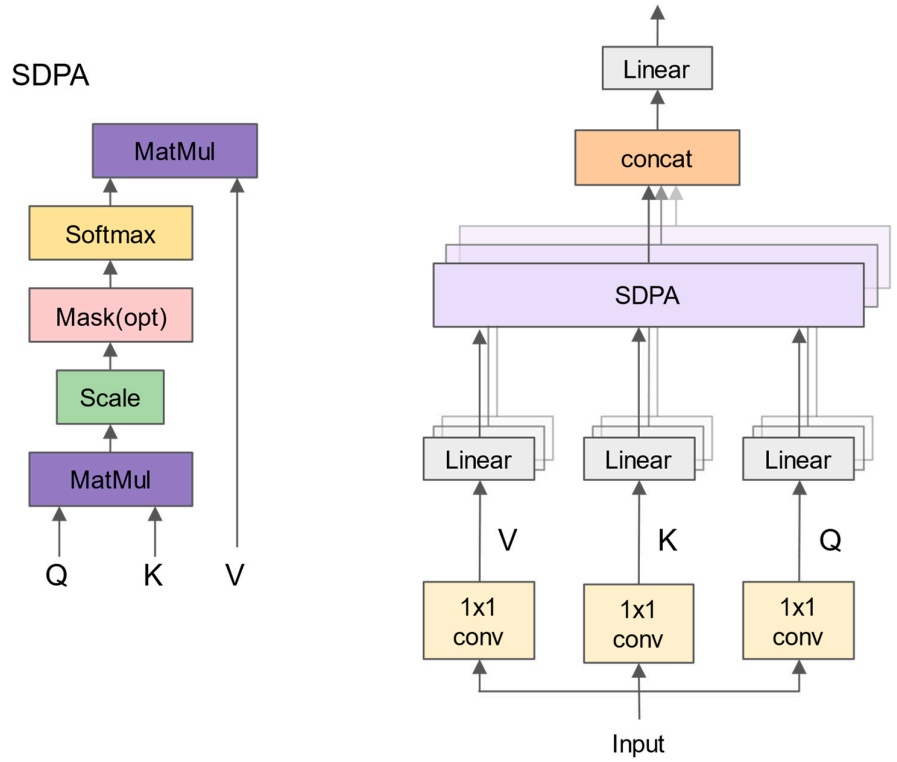

Figure 5. The MHSA sub-module of the SPMA module.

5. Experiments

Extensive experiments using the LSIray dataset and the public dataset OPIXray were conducted with the common methods used in security detection. Then, the effectiveness of the modules using our method was demonstrated by ablation studies.

5.1. Details

The method proposed is based on the PyTorch deep learning framework, a Python-based platform widely used and supported by the computer vision community. To execute the algorithm, a PyTorch deep learning environment was established on an Ubuntu OS, i.e., CUDA v11.7, cuDNN v7.6.4 and PyTorch 2.0.1. Our method was implemented using the MMDetection toolkit, an open-source object detection platform based on PyTorch, and experiments were conducted on a machine equipped with four NVIDIA TITAN RTX GPUs that have a high performance and 32 GB memory capacity. To ensure a fair comparison, all the compared methods were trained on the same dataset (LSIray or OPIXray) training set, and then evaluated on the respective dataset test set.

5.2. Evaluation Metrics

According to the evaluation metric of MS COCO [38], as shown in Table 2, the mean average precision (mAP) represents the average accuracy calculated across 10 IoU thresholds ranging from 0.5 to 0.95 with an interval of 0.05, which reflects the model's accuracy and robustness in locating and classifying targets. A higher mAP indicates that the model is more likely to correctly detect different categories of targets in the image and have a greater overlap between the predicted bounding box and the actual bounding box, which is the main challenge metric. The formula for precision is:

$$\text{precision} = \frac{TP}{TP + FP} \tag{1}$$

where True Positive (TP) denotes the number of true examples and False Positive (FP) denotes the number of false positive examples. A higher precision value indicates that

the model is better at accurately detecting the target in the image. The mAP50 represents the mean Average Precision calculated at an IoU threshold of 0.5, meaning that, for a predicted bounding box to be considered correct, it should have at least 50% overlap with the ground truth bounding box. Similarly, mAP75 represents the mean Average Precision calculated at an IoU threshold of 0.75, requiring a minimum overlap of 75%. The mAPs correspond to small objects with an area less than or equal to 322 pixels, while mAPm refers to medium-sized objects with an area between 322 and 962 pixels. Lastly, mAPl represents large objects with an area greater than or equal to 962 pixels.

Table 2. Evaluation results of the different methods on the LSIray dataset.

Model	mAP	mAP_50	mAP_75	mAP_s	mAP_m	mAP_l
ATSS_lA [18]	0.724	0.875	0.832	0.819	0.35	0.755
YOLOv3 [30]	0.699	0.893	0.823	0.702	0.348	0.668
SSD [29]	0.62	0.879	0.737	0.625	0.318	0.647
faster-rcnn [24]	0.739	0.915	0.856	0.736	0.414	0.702
Dynamic_rcnn [39]	0.742	0.924	0.862	0.778	0.425	0.733
Fcos [40]	0.727	0.905	0.826	0.734	0.358	0.691
YOLOv5	0.647	0.875	0.764	0.667	0.330	0.686
YOLOv8	0.813	0.933	0.895	0.809	0.490	0.775
SC-YOLOv8 (ours)	0.827	0.939	0.906	0.848	0.493	0.819

5.3. Experimental Result Analysis

The effectiveness of the proposed SC-YOLOv8 was verified by comparing it with the previous methods for detecting prohibited items and several popular object detection models on the LSIray dataset. The evaluation metrics included mAP, mAP_50, mAP_75, mAP_s, mAP_m and mAP_l. These metrics were chosen because they can comprehensively reflect the performance and robustness of the model. As presented in Table 2, the SC-YOLOv8 model was significantly outperformed by other models in all evaluation metrics, especially in mAP, mAP_s and mAP_l, for which the model was 1.4%, 3.9% and 4.4% higher than the second-ranked YOLOv8, respectively.

To demonstrate the visual results of SC-YOLOv8, several pictures from the test results that contain different types and quantities of prohibited items were randomly selected. The detection results of YOLOv5, the original YOLOv8 algorithm and the improved YOLOv8 algorithm in different scenes are shown in Figure 6. The first column in the figure is the Ground Truth image, the second column is the YOLOv5 detection result, the third column is the original YOLOv8 detection result and the fourth column is the improved YOLOv8 detection result in this paper.

Compared with the SC-YOLOv8 algorithm, YOLOv5 as well as the original YOLOv8 algorithm has leakage and false detection. In the first and second row images, there is a small target. The lighter and scissors are correctly detected by the improved YOLOv8 algorithm, but are not detected by YOLOv5 and the original YOLOv8. In the third and fourth rows of images, there are multiple prohibited items overlapped. In the third row of images, a false detection is detected in the original YOLOv8 algorithm results, recognizing one plier and one wrench overlapped as two wrenches and one plier, whereas YOLOv5 and the improved YOLOv8 do not demonstrate this error. In the fourth row of images, similarly, false detections are shown by YOLOv5 and the original YOLOv8, while the improved YOLOv8 algorithm agrees with the GT due to the fact that the proposed DBSModule and Attention Mechanism Module are well adapted to irregularly shaped objects, such as pliers and wrenches. In the last row, YOLOv5 and the original YOLOv8 have missed detections and wrong detections; the overlapped scissors and knife with a side shape are not detected in the YOLOv5 detection results, while the improved YOLOv8 matches the GT, again showing that the improved algorithm is better suited for a variety of shapes of prohibited items and for security inspections.

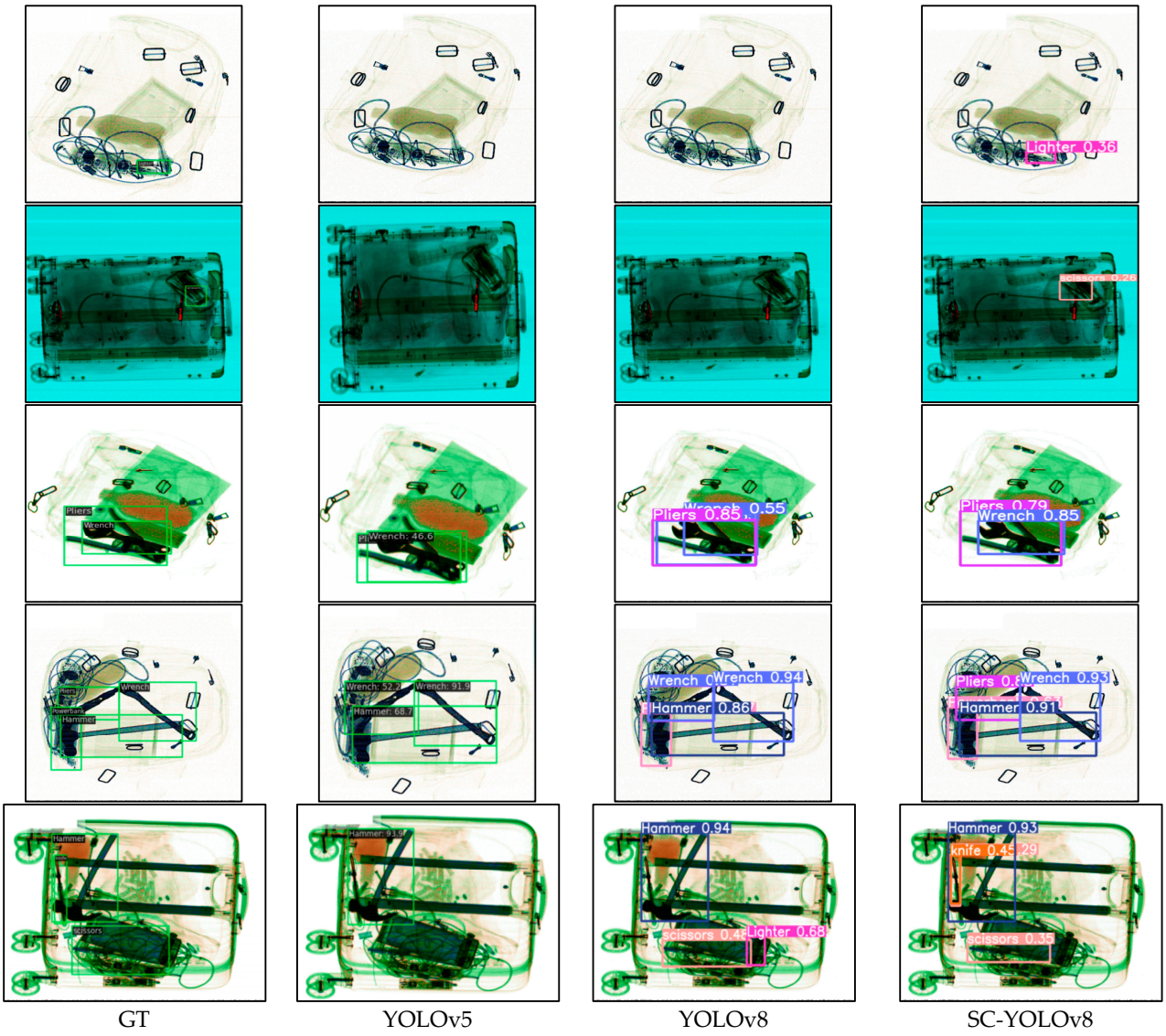

Figure 6. Comparison results of YOLOv8 and improved YOLOv8.

From the experimental results, the improved YOLOv8 target detection algorithm achieved a better performance than the existing model, improving the accuracy of target detection and having a practical application value.

5.4. Evaluation of Different Categories

The improved model (SC-YOLOv8) and the original model (YOLOv8) were compared in different categories, as shown in Table 3. It can be observed that the improved model has a certain degree of improvement in all categories, regardless of Box(P), Box(R), mAP50 or mAP50-95. This indicates that the improved model can detect the objects in the images more accurately and comprehensively than the original model, covering different difficulties and scales of objects. In particular, there is a significant improvement in some categories, such as SwissArmyKnife, Dagger, Hammer, Handcuffs and Tin. These categories may be difficult to recognize or locate by the original model, because they may be small or slender. The improved model adopts the DBSModule module, which can adapt to various shapes, such as slender, round and transparent. This can improve the model's ability to capture the features and positions of these categories. It makes it possible to improve the recognition rate, compared to the original algorithm, by 0.8%, 0.3% and 3% for difficult-to-detect objects, such as slender dagger blades, ringed handcuffs and transparent glass bottle, respectively.

Table 3. The comparison of Box(P), Box(R), mAP50 and mAP50-95 between SC-YOLOv8 and YOLOv8 across different categories.

Class	Box(P)	Box(R)	mAP50	mAP50-95
	SC-YOLOv8/YOLOv8	SC-YOLOv8/YOLOv8	SC-YOLOv8/YOLOv8	SC-YOLOv8/YOLOv8
All	0.954/0.945	0.917/0.91	0.943/0.94	0.842/0.83
Blade	0.988/0.982	0.984/0.984	0.994/0.991	0.803/0.792
Scissors	0.977/0.979	0.968/0.946	0.986/0.984	0.857/0.844
Knife	0.953/0.968	0.906/0.885	0.946/0.942	0.843/0.838
Dagger	0.993/0.985	0.989/0.989	0.99/0.989	0.94/0.924
SwissArmyKnife	0.998/0.993	1/1	0.995/0.995	0.842/0.818
PlasticBottle	0.909/0.926	0.907/0.922	0.962/0.963	0.829/0.821
Cans	0.986/0.97	0.793/0.784	0.895/0.867	0.812/0.782
VacuumCup	0.935/0.954	0.994/0.965	0.985/0.984	0.899/0.886
GlassBottle	0.525/0.495	0.231/0.269	0.341/0.325	0.289/0.287
CartonDrinks	0.982/0.989	0.971/0.983	0.987/0.992	0.864/0.854
Tin	0.949/0.921	0.921/0.916	0.961/0.957	0.849/0.839
SprayCans	0.958/0.956	0.925/0.922	0.964/0.961	0.875/0.869
Hammer	0.996/0.987	0.989/0.984	0.995/0.995	0.962/0.961
Wrench	0.988/0.978	0.978/0.973	0.985/0.983	0.938/0.923
Gun	0.938/0.936	0.984/0.953	0.991/0.984	0.86/0.853
HandCuffs	0.989/0.985	1/1	0.995/0.995	0.943/0.94
Baton	0.964/0.964	0.977/0.966	0.987/0.986	0.91/0.882
Pliers	0.991/0.991	0.993/0.986	0.995/0.995	0.931/0.92
Powerbank	0.944/0.936	0.959/0.936	0.983/0.977	0.851/0.845
Lighter	0.944/0.965	0.819/0.789	0.894/0.888	0.727/0.718
Bullet	0.955/0.955	0.961/0.961	0.975/0.975	0.853/0.853

By comparing the improved model and the original model across different categories, the following conclusions can be drawn: The improved model demonstrates a significant improvement on the object detection task, especially on some categories that are difficult to recognize or locate. This is mainly attributed to the introduction of the C2F_DCN module and the attention mechanism module, which can adapt to various shapes, thus improving the model's ability to capture the features and positions of objects. This is a valuable and meaningful improvement, which can provide new insights and contributions to the object detection field.

5.5. Experiments on the OPIXray Dataset

To validate the effectiveness of our method, we conducted a comparative experiment using the OPIXray dataset, comparing it with four state-of-the-art methods. The summarized results in Table 4 demonstrate that our method achieves the highest mAP across all five categories of cutting implements: 91.1%, 82.8%, 97.3%, 85.7% and 89.3%. Furthermore, the average mAP of our method was 89.24%, representing a significant improvement of 15.41% compared to the lowest-ranked S-YOLOv4 model and an increase of 1.22% compared to the second-place YOLOv8 model. These results highlight the evident superiority and robustness of our approach for security inspection tasks.

Table 4. Comparison of the mAP of different methods using the OPIXray dataset.

Method	Folding_Knife	Straight_Knife	Scissor	Utility_Knife	Tool_Knife	AVG
S-YOLOv4 [41]	81.44	41.07	94.70	68.25	83.67	73.83
DOAM [17]	81.37	41.50	95.12	68.21	83.83	74.01
MLM [42]	83.04	48.73	96.54	73.19	87.03	77.70
YOLOv8	88.5	82.8	96.3	82.7	89.8	88.02
Ours	91.1	82.8	97.3	85.7	89.3	89.24

5.6. Ablation Experiment

In this section, we performed ablation experiments to assess the impact and importance of the enhanced modules on YOLOv8 performance, using a selective module addition approach. Detection average precision (AP) and recall (AR) were used as the main evaluation metrics. The proposed method was compared with the baseline YOLOv8 and other related works using the dataset LSIray.

In Table 5, the performance improvement after the introduction of the C2F_DCN module and the MHSA module on the baseline YOLOv8 is demonstrated. From the table, it can be observed that the C2F_DCN module offers a 1% increase in detection mAP and 0.4% increase in detection AR, indicating that the C2F_DCN module can effectively adjust the position and shape of the receptive field to adapt to targets with different shapes and angles, and enhances the effect of multilayer feature fusion. On the other hand, the MHSA module leads to a further increase of 0.4% in detection mAP and 0.3% in detection AR, indicating that the MHSA module can effectively fuse different layers and aspects of feature information, and enhance the diversity and complexity of target detection. These results show that our improvements to YOLOv8 are effective and necessary, that these improvement modules have a complementary and collaborative relationship with each other and that removing any one of them would deteriorate the detection performance.

Table 5. Ablation studies using the LSIray dataset.

C2F_DCN	MHSA	mAP@0.5:0.95	AR
		0.813	0.798
√		0.823	0.802
	√	0.822	0.799
√	√	0.827	0.805

6. Conclusions

This paper studied the problem of image object detection of prohibited items during security checks using deep learning and computer vision techniques. In order to solve the problem of a lack of a large-scale prohibited item dataset, a new dataset, named LSIray, was created, which comprises a large number of high-quality X-ray images, involving 21 different types and sizes of luggage and objects, which includes categories ignored by common datasets. An improved YOLOv8 model based on attention mechanism and deformable convolution fusion was proposed to address the problem of poor accuracy. Extensive experiments were conducted using the dataset, which was compared with other related works. The experimental results show that the created dataset has a high quality, comprehensive categories, realistic scenes and other characteristics, providing more valuable and challenging data support for X-ray image object detection problems. Our improved method achieved 82.7% and 89.2% accuracy on the LSIray dataset and the OPIxray dataset, respectively. The relative improvement over the original YOLOv5 algorithm was 1.4% and 1.2%, respectively. The developed model has a high detection accuracy, strong adaptability to different shapes and angles of targets and other characteristics, providing a more effective and robust solution for X-ray image security inspection problems.

Author Contributions: Conceptualization, L.H. and Y.L.; methodology, Y.L. and C.M.; software, C.M.; validation, C.M., J.J. and J.S.; formal analysis, Y.L.; investigation, L.H.; resources, L.H.; data curation, J.J. and J.S.; writing—original draft preparation, C.M.; writing—review and editing, Y.L.; visualization, C.M.; supervision, L.H.; project administration, Y.L.; funding acquisition, L.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Scientific and Technological project in Henan Province, grant number 222102210015, and the Young Scientist project, grant number 225200810098.

Data Availability Statement: Dataset available at https://github.com/MACHUNHAI/LSIray (accessed on 9 October 2023).

Acknowledgments: The authors would like to appreciate the reviewers of Electronics for their criticism and suggestions on this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Akcay, S.; Breckon, T. Towards automatic threat detection: A survey of advances of deep learning within X-ray security imaging. *arXiv* **2020**, arXiv:2001.01293. [CrossRef]
2. Mery, D.; Saavedra, D.; Prasad, M. X-ray Baggage Inspection with Computer Vision: A Survey. *IEEE Access* **2020**, *8*, 145620–145633. [CrossRef]
3. Wei, Y.; Liu, X.; Liu, Y. Research on the application of high-efficiency detectors into the detection of prohibited item in X-ray images. *Appl. Intell.* **2021**, *52*, 4807–4823. [CrossRef]
4. Rafiei, M.; Raitoharju, J.; Iosifidis, A. Computer Vision on X-ray Data in Industrial Production and Security Applications: A Comprehensive Survey. *IEEE Access* **2023**, *11*, 2445–2477. [CrossRef]
5. Kolte, S.; Bhowmik, N. Dhiraj Threat Object-based anomaly detection in X-ray images using GAN-based ensembles. *Neural Comput. Appl.* **2022**, 1–16. [CrossRef]
6. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
7. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. [CrossRef] [PubMed]
8. Guo, S.; Tang, S.; Zhu, J.; Fan, J.; Ai, D.; Song, H.; Liang, P.; Yang, J. Improved U-net for guidewire tip segmentation in X-ray fluoroscopy images. In Proceedings of the 2019 3rd International Conference on Advances in Image Processing, Chengdu, China, 8–10 November 2019.
9. Chaudhary, A.; Hazra, A.; Chaudhary, P. Diagnosis of chest diseases in X-ray images using deep convolutional neural network. In Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 6–8 July 2019.
10. Lu, J.; Tong, K.-Y. Towards to Reasonable Decision Basis in Automatic Bone X-ray Image Classification: A Weakly-Supervised Approach. *Proc. Conf. AAAI Artif. Intell.* **2019**, *33*, 9985–9986. [CrossRef]
11. Miao, C.; Xie, L.; Wan, F.; Su, C.; Liu, H.; Jiao, J.; Ye, Q. SIXray: A Large-Scale Security Inspection X-ray Benchmark for Prohibited Item Discovery in Overlapping Images. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
12. Huang, S.; Wang, X.; Chen, Y.; Xu, J.; Tang, T.; Mu, B. Modeling and quantitative analysis of X-ray transmission and backscatter imaging aimed at security inspection. *Opt. Express* **2019**, *27*, 337–349. [CrossRef] [PubMed]
13. Akcay, S.; Breckon, T.P. An evaluation of region based object detection strategies within x-ray baggage security imagery. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017. [CrossRef]
14. Akcay, S.; Kundegorski, M.E.; Willcocks, C.G.; Breckon, T.P. Using Deep Convolutional Neural Network Architectures for Object Classification and Detection Within X-ray Baggage Security Imagery. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2203–2215. [CrossRef]
15. Liu, Z.; Li, J.; Shu, Y.; Zhang, D. Detection and Recognition of Security Detection Object Based on Yolo9000. In Proceedings of the 2018 5th International Conference on Systems and Informatics (ICSAI), Nanjing, China, 23 September 2018. [CrossRef]
16. Mery, D.; Riffo, V.; Zscherpel, U.; Mondragón, G.; Lillo, I.; Zuccar, I.; Lobel, H.; Carrasco, M. GDXray: The Database of X-ray Images for Nondestructive Testing. *J. Nondestruct. Evaluation* **2015**, *34*, 42. [CrossRef]
17. Wei, Y.; Tao, R.; Wu, Z.; Ma, Y.; Zhang, L.; Liu, X. 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, Seattle, WA, USA, 12–16 October 2020. [CrossRef]
18. Zhao, C.; Zhu, L.; Dou, S.; Deng, W.; Wang, L. Detecting Overlapped Objects in X-ray Security Imagery by a Label-Aware Mechanism. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 998–1009. [CrossRef]
19. Lu, X.; Li, B.; Yue, Y.; Li, Q.; Yan, J. Grid r-cnn. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
20. Chandra, M.A.; Bedi, S.S. Survey on SVM and their application in image classification. *Int. J. Inf. Technol.* **2018**, *13*, 1–11. [CrossRef]
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef] [PubMed]
22. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015.
23. Uijlings, J.R.R.; van de Sande, K.E.A.; Gevers, T.; Smeulders, A.W.M. Selective Search for Object Recognition. *Int. J. Comput. Vis.* **2013**, *104*, 154–171. [CrossRef]
24. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems 28 (NIPS 2015), Montreal, QC, Canada, 7–12 December 2015.
25. Dai, J.; Li, Y.; He, K.; Sun, J. R-fcn: Object detection via region-based fully convolutional networks. In Proceedings of the Advances in Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; p. 29.

26. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, 21–26 July 2017.
27. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
28. Cai, Z.; Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
29. Liu, W.; Fu, C.-Y.; Berg, A.-C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016.
30. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
31. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
32. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
33. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023. [CrossRef]
34. Lin, T.-Y.; Goyal, P.; Girshick, R.; Kaiming, P.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
35. Tao, R.; Wei, Y.; Jiang, X.; Li, H.; Qin, H.; Wang, J.; Ma, Y.; Zhang, L.; Liu, X. Towards real-world X-ray security inspection: A high-quality benchmark and lateral inhibition module for pro-hibited items detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021.
36. Yuan, J.; Zhang, N.; Xie, Y.; Gao, X. Detection of Prohibited Items Based upon X-ray Images and Improved YOLOv7. *J. Phys. Conf. Series.* **2022**, *2390*, 012114. [CrossRef]
37. Dai, J.; Qi, X.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
38. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Proceedings, Part V 13. Springer: Berlin/Heidelberg, Germany, 2014.
39. Zhang, H.; Chang, H.; Ma, B.; Wang, N.; Chen, X. Dynamic R-CNN: Towards High Quality Object Detection via Dynamic Training. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020. [CrossRef]
40. Tian, Z.; Shen, C.; Chen, H.; He, T. Fcos: Fully convolutional one-stage object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019.
41. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. Scaled-yolov4: Scaling cross stage partial network. In Proceedings of the IEEE/cvf Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021.
42. Liu, D.; Tian, Y.; Xu, Z.; Jian, G. Handling occlusion in prohibited item detection from X-ray images. *Neural Comput. Appl.* **2022**, *34*, 20285–20298. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.