Towards Real-world X-ray Security Inspection: A High-Quality Benchmark And Lateral Inhibition Module For Prohibited Items Detection

Renshuai Tao[1,2]   Yanlu Wei[1]   Xiangjian Jiang[1]   Hainan Li[1]
Haotong Qin[1]   Jiakai Wang[1]   Yuqing Ma[1]   Libo Zhang[3]   Xianglong Liu[1*]
[1]State Key Laboratory of Software Development Environment, Beihang University
[2]iFLYTEK Research   [3]Institute of Software Chinese Academy of Sciences
{rstao, weiyanlu, hainan, qinhaotong, jk_buaa_scse, mayuqing, xlliu}@buaa.edu.cn
silencejiang12138@gmail.com, libo@iscas.ac.cn

Abstract

*Prohibited items detection in X-ray images often plays an important role in protecting public safety, which often deals with color-monotonous and luster-insufficient objects, resulting in unsatisfactory performance. Till now, there have been rare studies touching this topic due to the lack of specialized high-quality datasets. In this work, we first present a **Hi**gh-quality **X-ray** (HiXray) security inspection image dataset, which contains **102,928** common prohibited items of **8** categories. It is the largest dataset of high quality for prohibited items detection, gathered from the real-world airport security inspection and annotated by professional security inspectors. Besides, for accurate prohibited item detection, we further propose the **L**ateral **I**nhibition **M**odule (LIM) inspired by the fact that humans recognize these items by ignoring irrelevant information and focusing on identifiable characteristics, especially when objects are overlapped with each other. Specifically, LIM, the elaborately designed flexible additional module, suppresses the noisy information flowing maximumly by the Bidirectional Propagation (BP) module and activates the most identifiable charismatic, boundary, from four directions by Boundary Activation (BA) module. We evaluate our method extensively on HiXray and OPIXray and the results demonstrate that it outperforms SOTA detection methods.[1]*

1. Introduction

As the density if the crowd density increases in public transportation hubs, security inspection has become more and more important in protecting public safety. X-ray scanners, which are adopted usually to scan the luggage and

---

Figure 1. Nature and X-ray images of the 8 categories of common prohibited items in HiXray. These prohibited items usually contain lithium battery, liquid, lighter, *etc*.

generate the complex X-ray images, play an important role in security inspection scenario. However, security inspectors struggle to accurately detect the prohibited items after a long time highly-concentrating work, which may cause severe danger to the public. Therefore, it is imperative to develop a rapid, accurate and automatic detection method.

Fortunately, the innovation of deep learning [27, 28, 38, 18, 43, 16], especially the convolutional neural network, makes it possible to accomplish this goal by transferring it into object detection task in computer vision [15, 26, 42, 11]. However, different from traditional detec-

tion tasks, in this scenario, items within a luggage are randomly overlapped where most areas of objects are occluded resulting in heavy noise in X-ray images. Thus, this characteristic leads to strong requirement of high-quality datasets and models with satisfactory performance for this task.

Regarding dataset, to the best of our knowledge, there are only three released X-ray benchmarks, namely GDXray [23], SIXray [25] and OPIXray [40]. Both GDXray and SIXray are constructed for classification task and the images of OPIXray are synthetic. Besides, the categories and quantities of labeled instances in the three datasets are far from meeting the requirements in real-world applications. We make a detailed comparison in Table 1. Regarding models, traditional CNN-based models [41, 7, 29] trained through common detection datasets fail to achieve satisfactory performance in this scenario because that different from natural images [5, 33] with simple visual information, X-ray images [36, 24, 22] are characterized by the lacking of strong identification properties and containing heavy noises. This urgently requires researchers to make breakthroughs in both datasets and models.

To address the above drawbacks, in this work, we contribute the largest high-quality dataset for prohibited items detection in X-ray images, named **Hi**gh-quality **X-ray** (HiXray) dataset, which contains 102,928 common labeled instances of 8 categories, such as lithium battery, liquid, *etc*. All of these images are gathered from real-world daily security inspections in an international airport. Thus, The categories, quantities and locations of prohibited items are in line with the data distribution in real-world scenarios. Besides, each instance is manually annotated by professional inspectors from the international airport, guaranteeing the accurate annotations. In addition, our HiXray dataset can serve the evaluation of various detection tasks including small, occluded object detection, *etc*.

For accurate prohibited items detection, we present the **L**ateral **I**nhibition **M**odule (LIM), which is inspired by the fact that humans recognize these items by ignoring irrelevant information and focusing on identifiable characteristics, especially when objects are overlapped with each other. LIM consists of two core sub-modules, namely Bidirectional Propagation (BP) and Boundary Activation (BA). BP filters the noise information to suppress the influence from the neighbor regions to the object regions and BA activates the boundary information as the identification property, respectively. Specifically, BP eliminates noises adaptively through the bidirectional information flowing across layers and BA captures the boundary from four directions inside each layer and aggregates them into a whole outline.

HiXray dataset and LIM model provide a new and reasonable evaluation benchmark for the community, and helps make a wider breadth of real-world applications. The main contributions of this work are as follows:

- We present the largest high-quality dataset named HiXray for X-ray prohibited items detection, providing a new and reasonable evaluation benchmark for the community. We hope that contributing this dataset can promote the development of this issue.

- We propose the LIM model, which exploits the lateral inhibition mechanism to improve the detecting ability for accurate prohibited items detection, inspired by the intimate relationship between deep neural networks and biological neural networks.

- We evaluate LIM on the HiXray and OPIXray datasets and the results show that LIM can not only be versatile to SOTA detection methods but also improve the performance of them.

2. Related Work

Prohibited Items Detection in X-ray Images. X-ray imaging offers powerful ability in many tasks such as medical image analysis [9, 4, 21] and security inspection [25, 12]. As a matter of fact, obtaining X-ray images is difficult, so rare studies touch security inspection in computer vision due to the lack of specialized high-quality datasets.

Several recent efforts [1, 23, 2, 25, 19, 40] have been devoted to constructing such datasets. A released benchmark, GDXray [23] contains 19,407 gray-scale images, part of which contain three categories of prohibited items including gun, shuriken and razor blade. SIXray [25] is a large-scale X-ray dataset which is about 100 times larger than the GDXray dataset but the positive samples are less than 1% to mimic a similar testing environment and the labels are annotated for classification. Recently, [40] proposed the OPIXray dataset, containing 8,885 X-ray images of 5 categories of cutters. The images of OPIXray dataset are artificial synthetic. Other relevant works [1, 2, 19] have not made their data available to download.

Object Detection. In computer vision area, object detection is one of important tasks, which underpins a few instance-level recognition tasks and many downstream applications. Here we review some works that is the closest to ours. Most of the CNN-based methods can be summarized into two general approaches: one-stage detectors and two-stage detectors. Recently one-stage methods have gained much attention over two-stage approaches due to their simpler design and competitive performance. SSD [20] discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales. YOLO [30, 31, 32, 3, 14] is the collection of a series of well-known methods, which values both real-time and accuracy among one-stage detection algorithms. Moreover, FCOS [35] proposes a fully convolutional one-stage object detector to solve object detection in a per-pixel prediction fashion, analogue to other dense prediction problems.

Dataset	Year	Category	N_p	Annotation			Color	Task	Data Source
				Bounding Box	Number	Professional			
GDXray [23]	2015	3	8,150	✓	8,150	✗	Gray-scale	Detection	Unknown
SIXray [25]	2019	6	8,929	✗	✗	✗	RGB	Classification	Subway Station
OPIXray [40]	2020	5	8,885	✓	8,885	✓	RGB	Detection	Artificial Synthesis
HiXray	**2021**	**8**	**45,364**	✓	**102,928**	✓	RGB	Detection	Airport

Table 1. Comparison of existing open-source X-ray datasets. N_p refers to the number of images containing prohibited items. In our HiXray dataset, some images contain more than one prohibited item and every prohibited item is located with a bounding-box annotation, causing the number of annotation is greater than N_p.

3. HiXray Dataset

As Table 1 illustrates, the existing datasets are less than satisfactory and thus fail to meet the requirements in real-world applications. In this work, we construct a new high-quality dataset for X-ray prohibited items detection. Then we introduce the construction principles, data properties and potential tasks of the proposed HiXray dataset.

3.1. Construction Principles

We construct the HiXray dataset in accordance with the following five principles:

Realistic Source. Considering realistic source can make the data more meaningful for research, we gather the images of the HiXray dataset from daily security inspections in an international airport to ensure the authenticity of data.

Data Privacy. We strictly follow the standard de-privacy procedure by deleting private information (name, place, *etc*.), ensuring that nobody can connect the luggage with owners through the images to guarantee the privacy.

Extensive Diversity. HiXray contains the 8 categories of prohibited items such as lithium battery, liquid, lighter, etc., all of which are frequently seen in daily life.

Professional Annotation. Objects in X-ray images are difficult to be recognized for people without professional training. In HiXray, each instance is manually localized with a box-level annotation by professional security inspectors of the airport, who are very skillful in daily work.

Quality Control. We followed the similar quality control procedure of annotation as the famous Pascal VOC [6]. All inspectors followed the same annotation guidelines including what to annotate, how to annotate bounding, how to treat occlusion, *etc*. Besides, the accuracy of each annotation was checked by another inspector, including checking for omitted objects to ensure exhaustive labelling.

3.2. Data Details

Instances per category. HiXray contains 45,364 X-ray images, 8 categories of 102,928 common prohibited items. The statistics are shown in Table 2.

Category	PO1	PO2	WA	LA	MP	TA	CO	NL	Total
Training	9,919	6,216	2,471	8,046	43,204	3,921	7,969	706	82,452
Testing	2,502	1,572	621	1,996	10,631	997	1,980	177	20,476
Total	12,421	7,788	3,092	10,042	53,835	4,918	9,949	883	102,928

Table 2. The statistics of category distribution of HiXray dataset, where PO1, PO2, WA, LA, MP, TA, CO and NL denote "Portable Charger 1 (lithium-ion prismatic cell)", "Portable Charger 2 (lithium-ion cylindrical cell)", "Water", "Laptop", "Mobile Phone", "Tablet", "Cosmetic" and "Nonmetallic Lighter".

Instances per image. On average there are 2.27 instances per image in HiXray dataset. In comparison SIXray has 1.37 (in positive samples), both OPIXray and GDXray have 1 instance per image on average. Obviously, the larger average number of instances per image brings more contextual information, which is more valuable. The statistics is shown in Table 3.

N_i	1	2	3	4	5	6	7	8	9	10
Training	12,726	10,905	6,860	3,286	1,521	602	254	91	35	11
Testing	3,227	2,722	1,705	810	354	145	54	41	8	2
Total	15,953	13,627	8,565	4,096	1,875	747	308	132	43	13

Table 3. The quantity distribution of images containing different numbers of prohibited items. Note that N_i refers to the number of prohibited items in each image.

Division of training and testing. The dataset is partitioned into a training set and a testing set, where the ratio is about 4 : 1. The statistics of category distribution of training set and testing set are also shown in Table 2.

Color Information. Different X-ray machine models may have some differences in color imaging, and we adopt one of the most classic color imaging strategy. The colors of objects under X-ray are mainly determined by their chemical composition, which is introduced in detail in Table 4.

Data Quality. All images are stored in JPG format with a 1200*900 resolution, averagely. The maximum resolution of samples can reach 2000*1040.

Color	Material	Typical examples
Orange	Organic Substances	Plastics, Clothes
Blue	Inorganic Substances	Irons, Coppers
Green	Mixtures	Edge of phones

Table 4. The color information of different objects under X-ray.

3.3. Potential Tasks

Our HiXray dataset can further serve the evaluation of various detection tasks including small object detection, occluded object detection, *etc*.

Small Object Detection. Security inspectors often struggle to find small prohibited items in baggage or suitcase. In our HiXray dataset, there are many small prohibited items. According to the definition of small by SPIE, the size of small object is usually no more than 0.12% of entire image size. We thus define the small object as the object whose ground-truth bounding box accounts for less than 0.1% of entire image, while the large object is defined as the object whose ground-truth bounding box takes up more than 0.2% proportion in the entire image, and the rest is the medium. The images of "Portable Charger 2" and "Mobile Phone" can be divided into three subsets respectively. The categories distribution is illustrated in Table 5.

Category	Total	Large	Medium	Small
PO2	2,502	587	986	929
MP	10,631	3,547	4,248	2,836

Table 5. The category distribution of "Portable Charger 2" and "Mobile Phone" (PO2 and MB for short) when the two thresholds are set as 0.1% and 0.2%.

Occluded Object Detection. The items in baggage or suitcase are often overlapped with each other, causing the occlusion problem in X-ray prohibited items detection. [40] proposed the occluded prohibited items detection task in X-ray security inspection. The occlusion problem also exists in HiXray dataset with large-scale images with more categories and numbers. In order to study the impact brought by object occlusion levels, researchers can divide the HiXray dataset into three (or more) subsets according to different occlusion levels (illustrated in Figure 2).

4. The Lateral Inhibition Module

In neurobiology, lateral inhibition disables the spreading of action potentials from excited neurons to neighboring neurons in the lateral direction. We mimic this mechanism by designing a bidirectional propagation architecture to adaptively filter the noisy information generated by the neighboring regions of the prohibited items. Also, lat-

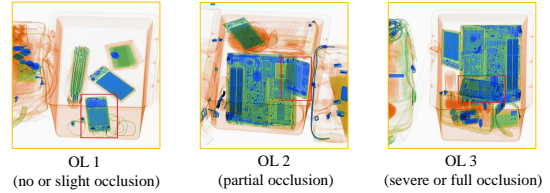| OL 1	OL 2	OL 3
(no or slight occlusion)	(partial occlusion)	(severe or full occlusion)

Figure 2. Three occlusion levels on prohibited items in the images of HiXray, where the numbers indicates occlusion levels.

eral inhibition creates contrast in stimulation that allows increased sensory perception, so we activate the boundary information by intensifying it from four directions inside each layer and aggregating them into a whole shape.

Therefore, inspired by the mechanism that lateral suppression by neighboring neurons in the same layer making the network more efficient, we propose the Lateral Inhibition Module (LIM). In this section, we will first introduce the network architecture in Section 4.1 and further the two core sub-modules, namely Bidirectional Propagation (BP) and Boundary Activation (BA), in Section 4.3 and Section 4.2, respectively.

4.1. Network Architecture

Figure 3 illustrates the architecture of our LIM. It takes a single-scale image of an arbitrary size as input, and outputs proportionally sized feature maps at multiple levels. Similar to FPN [17] and some other varieties like PANet [39], this process is independent of the backbone architectures.

Specifically, suppose there are N training images $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$ and L convolutional layers in the backbone network. One sample $\mathbf{x} \in \mathbf{X}$ is fed into the backbone network and computed feed-forwardly, which computes a feature hierarchy consisting of feature maps at several scales with a scaling step of 2.

Suppose $\mathcal{F}(\cdot)$ as a composite function of three consecutive operations: Batch Normalization (BN) [13], followed by a rectified linear unit (ReLU) [8] and a 3×3 Convolution (Conv). $\mathcal{F}^l(\mathbf{x})$ is the feature map generated by the l-th layer of the network. **Firstly**, in the left part of BP, the noisy information is adaptively filtered because their propagation is reduced from high-level to low-level feature maps in the top-down pathway. **Secondly**, the output feature maps are fed into BA. BA refines the feature maps to activate the boundary information by enhance it from four directions and outputs the refined feature maps. **Thirdly**, similar to the left, the right part of BP reduces the propagation of noisy information from low-level to high-level feature maps through the bottom-up pathway. **Finally**, the feature maps outputted by each layer of the right of BP combine those of the corresponding layer from the backbone
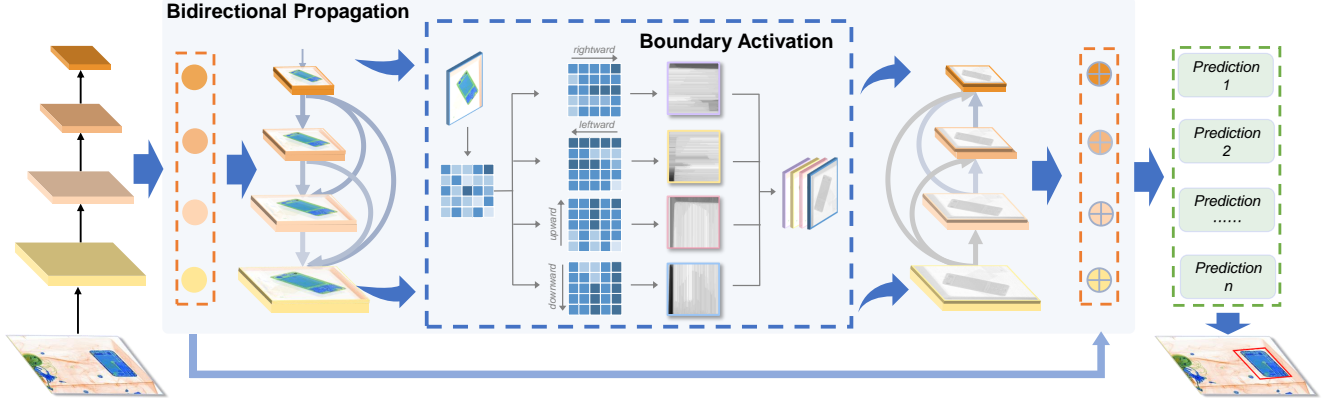
Figure 3. The network structure of Lateral Inhibition Module (LIM). Bidirectional Propagation filters noisy information to suppress the influence from neighbor regions to object regions and Boundary Activation activates the boundary as the identification property, respectively.

Algorithm 1: The Procedure of LIM.

Input: The feature map set
$$\mathbf{F} = \{\mathcal{F}^1(\mathbf{x}), \cdots, \mathcal{F}^L(\mathbf{x})\}.$$
Output: The refined feature map set
$$\mathbf{C} = \{\mathbf{C}^1, \cdots, \mathbf{C}^L\}.$$
for *all* $l = 1, 2, \ldots, L$ **do**
    Compute $\mathbf{A}^l$ based on Eq. (1);
    for *direction in fourDirections* **do**
        if *direction is horizontal* **then**
            // Avoid column loop for faster speed;
            Rotate feature map;
        for *row $\leftarrow$ 1 to heightOfMap* **do**
            Compute $\mathbf{B}^l_{ijc}$ based on Eq. (4);
    Generate $\mathbf{B}^l$ by concatenating all $\mathbf{B}^l_{ijc}$;
    Compute $\mathbf{C}^l_{\text{t}}$ based on Eq. (2);
    Compute $\mathbf{C}^l$ based on Eq. (3);
Obtain the feature map set $\mathbf{C} = \{\mathbf{C}^1, \cdots, \mathbf{C}^L\}$.

---

network. And the combined feature maps is conveyed to the following prediction layers. Algorithm 1 summarizes the whole process (We add explanations about acceleration operation in code implementation in Algorithm 1) and the details of modules are described in the following sections.

4.2. Bidirectional Propagation

To disable the spreading of noisy information of neighboring regions, we mimic this mechanism by designing the bidirectional propagation architecture. Moreover, we add a dense mechanism to enhance the ability of BP to choose proper information to propagate.

As shown in Figure 3, for the dense top-down pathway on the left of BP, up-sampling spatially coarser but semantically stronger feature maps from higher pyramid levels hal-

lucinates higher resolution features. These feature maps are enhanced by the corresponding feature maps from the convolutional layers via lateral connections. Each lateral connection merges feature maps of the same spatial size from the convolutional layer and the top-down pathway. The feature map of low convolutional layer is of lower-level semantics, but its activation is more accurately localized as it was sub-sampled fewer times. Further, we construct the dense connections to ensure maximum the filter.

Specifically, to preserve the feed-forward nature, $\mathcal{F}^l(\mathbf{x})$ obtains additional inputs from the feature maps $\mathcal{F}^{l+1}(\mathbf{x})$, $\cdots$, $\mathcal{F}^L(\mathbf{x})$ of all preceding layers and passes on its own feature-maps to the feature maps $\mathcal{F}^{l-1}(\mathbf{x}), \cdots, \mathcal{F}^1(\mathbf{x})$ of all subsequent layers. Figure 3 illustrates this layout schematically. We define $\mathcal{U}^m(\cdot)$ as the up-sampling operation (2^m times) and $\mathcal{V}(\cdot)$ as a 1×1 convolutional layer to reduce channel dimensions. The process is formulated as follows:

$$\mathbf{A}^l = \mathcal{V}\left(\mathcal{F}^l(\mathbf{x})\right) + \sum_{m=1}^{L-l} \mathcal{U}^m\left(\mathbf{A}^{l+m}\right), \qquad (1)$$

where $\mathbf{A}^l$ refers to the feature map outputted by the m-th layer of the left part of BP.

Regarding the right part of BP, as the Figure 3 illustrated, suppose the input feature map $\mathbf{B}^l$ refers to the feature map whose boundary has been activated in Eq. (4) (Boundary Activation will be introduced in the following section). Similar to the previous definition, $\mathcal{D}^m(\cdot)$ is the down-sampling operation (2^m times). This process can be formulated as follows:

$$\mathbf{C}^l_{\text{t}} = \mathcal{V}\left(\mathbf{B}^l\right) + \sum_{m=1}^{l-1} \mathcal{D}^m\left(\mathbf{C}^{l-m}_{\text{t}}\right), \qquad (2)$$

$$\mathbf{C}^l = \mathbf{C}^l_{\text{t}} + \mathcal{F}^l(\mathbf{x}), \qquad (3)$$
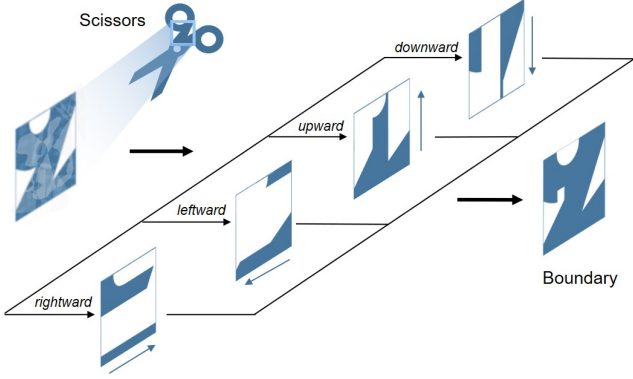
Figure 4. The schematic diagram of Boundary Activation.

where $\mathbf{C}_t^l$ refers to the output of l-th layer of the bottom-up pathway and $\mathbf{C}^l$ refers to the feature map generated of the l-th layer of BP. Finally, we convey $\mathbf{C}^l$, outputted by LIM, to the following prediction layers.

4.3. Boundary Activation

To mimic the mechanism that lateral inhibition creates contrast in stimulation that allows increased sensory perception, we activate the boundary information by intensifying it from four directions inside the feature maps outputted by each layer and aggregating them into a whole shape. The schematic diagram is shown in Figure 4.

As is shown in Figure 4, the key to capturing the boundary of object is to determine whether a position is a boundary-point. Motivated by the schematic diagram, we design the module Boundary Activation to perceive the sudden changes of boundary and its surroundings. Suppose we want to capture the left boundary of the object in the feature map $\mathbf{A}^l \in \mathbb{R}^{H \times W \times C}$ (the output of left part of Bidirectional Propagation). $\mathbf{A}_c^l$ donates the c-th channel of $\mathbf{A}^l$. Further, $\mathbf{A}_{ijc}^l$ refers to the location (i, j) of the feature map $\mathbf{A}_c^l$. To determine whether there is a sudden change between a position and the left of the point, the right-most point $\mathbf{A}_{iWc}^l$ traverses to the left. The process of perceiving the left boundary can be formulated as Eq. (4).

$$\mathbf{B}_{ijc}^l = \begin{cases} \mathbf{A}_{iWc}^l & \text{if } j = W, \\ \max\left\{\mathbf{A}_{ijc}^l, \mathbf{A}_{i(j+1)c}^l, \ldots, \mathbf{A}_{iWc}^l\right\} & \text{otherwise,} \end{cases} \quad (4)$$

where the $\mathbf{B}_{ijc}^l$ refers to the location (i, j) of c-th channel of the feature map $\mathbf{B}^l$ after Boundary Activation.

5. Experiments

In this section, we conduct comprehensive experiments on HiXray and OPIXray dataset to evaluate the effectiveness of LIM. To the best of our knowledge, HiXray and OPIXray [40] are the only two datasets currently available for X-ray prohibited items detection (RGB).

First, we verify the effectiveness of LIM by comparing the base and LIM-integrated classic or SOTA detection methods (SSD [20], FCOS [35] and YOLOv5 [14]). We evaluate all the base detection methods and the LIM-integrated methods on HiXray and OPIXray datasets. Second, we evaluate the superiority of our LIM over other feature pyramid mechanisms by comparing two famous methods FPN [17] and PANet [39] on the HiXray dataset. Third, we perform an ablation study to thoroughly evaluate each part of LIM. Finally, we conduct the visualization experiment to demonstrate the performance improvement.

5.1. Experiment Setting Details

LIM: LIM is implemented by PyTorch for its high flexibility and powerful automatic differentiation mechanism. The LIM-integrated model refers to the model that we implement this mechanism inside (Section 5.2). Both FPN and PANet contain feature pyramid mechanism similar to LIM, but they are not plug-in model. Therefore, we refer to their published code and re-implemented the mechanism deployed in SSD (Section 5.3). Unless specified, we use the following implementation details.

Backbone Networks: The backbone networks of SSD, FCOS and YOLOv5 are VGG16 [34], ResNet50 [10] and CSPNet [37] respectively. For each backbone network, we modify the corresponding network architecture to implement LIM mechanism.

Parameters: All experiments of LIM and baselines are optimized by the SGD optimizer and the initial learning rate is set to 0.0001. The momentum and weight decay are set to 0.9 and 0.0005 respectively. The batch size is set to 32 with shuffle strategy while training. We evaluate the mean Average Precision (mAP) of the object detection to measure the performance of all models fairly. Besides, the IOU threshold measuring the accuracy of the predicted bounding box against the ground-truth is set to 0.5.

5.2. Comparing with SOTA Detection Methods

We verify the effectiveness of LIM by implementing this mechanism to several detection approaches, including traditional SSD, the latest FCOS and YOLOv5. We integrate LIM to the three detection approaches and compare the LIM-integrated methods to the original baselines. In addition, we integrate the latest detection method for security inspection DOAM (in the work of OPIXray dataset) into the three detection approaches above and compare the results with our LIM. The experimental results on HiXray dataset and OPIXray dataset are shown in Table 6.

Table 6 demonstrates that in HiXray dataset, the LIM-integrated network improves the average performance by 1.7%, 1.6% and 1.5% over the original base models SSD,

Method	HiXray Dataset (**Ours**)									OPIXray Dataset [40]					
	AVG	PO1	PO2	WA	LA	MP	TA	CO	NL	**AVG**	FO	ST	SC	UT	MU
SSD [20]	71.4	87.3	81.0	83.0	97.6	93.5	92.2	36.1	0.01	70.9	76.9	35.0	93.4	65.9	83.3
SSD+DOAM [40]	72.1	88.6	82.9	83.6	97.5	94.1	92.1	38.2	0.01	74.0	81.4	41.5	95.1	68.2	**83.8**
SSD+LIM	**73.1**	**89.1**	**84.3**	**84.0**	**97.7**	**94.5**	**92.4**	**42.3**	**0.1**	**74.6**	**81.4**	**42.4**	**95.9**	**71.2**	82.1
FCOS [35]	75.7	88.6	86.4	86.8	89.9	88.9	88.9	63.0	13.3	82.0	86.4	68.5	90.2	78.4	86.6
FCOS+DOAM [40]	76.2	88.6	87.5	87.8	89.9	89.7	88.8	63.5	12.7	82.4	86.5	68.6	90.2	78.8	**87.7**
FCOS+LIM	**77.3**	**88.9**	**88.2**	**88.3**	**90.0**	**89.8**	**89.2**	**69.8**	**14.4**	**83.1**	**86.6**	**71.9**	**90.3**	**79.9**	86.8
YOLOv5 [14]	81.7	95.5	94.5	92.8	97.9	98.0	94.9	63.7	16.3	87.8	93.4	67.9	98.1	85.4	94.1
YOLOv5+DOAM [40]	82.2	95.9	94.7	93.7	98.1	98.1	95.8	65.0	16.1	88.0	93.3	69.3	97.9	84.4	**95.0**
YOLOv5+LIM	**83.2**	**96.1**	**95.1**	**93.9**	**98.2**	**98.3**	**96.4**	**65.8**	**21.3**	**90.6**	**94.8**	**77.6**	**98.2**	**88.9**	93.8

Table 6. Comparisons of common detection approaches SSD, FCOS and YOLOv5 (for simplicity, we use the lightest YOLOv5s model in the YOLOv5 experiment), and the latest related model DOAM on the HiXray dataset and OPIXray dataset, where the definition of the categories (PO1, PO2, *etc.*) can be found in Table 2. FO, ST, SC, UT and MU donate "Folding Knife", "Straight Knife", "Scissor", "Utility Knife" and "Multi-tool Knife" in the OPIXray dataset, respectively.

FCOS and YOLOv5 respectively. Besides, LIM outperforms DOAM by 1%, 1.1% and 1% with the base model SSD, FCOS and YOLOv5 respectively. In OPIXray dataset, the LIM-integrated network improves the mean performance by 3.7%, 1.1% and 2.8% over the original models SSD, FCOS and YOLOv5 respectively. Besides, LIM outperforms DOAM by 0.6%, 0.7% and 2.6% with the base models SSD, FCOS and YOLOv5 respectively.

Note that the performances are particularly low in two classes (CO and NL) for all models in Table 6, it is mainly because that, compared to other categories, NL and CO are far more difficult to recognize. For NL, it is very small in size and composed of a small piece of iron and a plastic body. The plastic shown under X-ray appears orange, which almost blends in with the background. For CO, the main reason is that there are big differences in the shapes of cosmetics, such as round and square, which are easy to be confused with other kinds of items.

5.3. Comparing with Feature Pyramid Mechanisms

LIM can be regarded as another feature pyramid method with a novel dense connection mechanism with specific feature enhancement. Therefore, We compare our LIM with the classical feature pyramid mechanism FPN and the variety PANet in different base models. Note that there is the same feature pyramid mechanism of FPN in FCOS and a variety of PANet mechanism in YOLOv5, so we replace the feature pyramid mechanism with our LIM in FCOS and YOLOv5 to verify that our mechanism works better in the base model FCOS and YOLOv5 (the same as Section 5.2). The experimental results are shown in Table 7.

LIM improves by both 1.1% to FPN and PANet in the base SSD model, 1.1% to FPN in the base FCOS model and 1.5% to the variety of PANet in the base YOLOv5 model. Further, we observe from the Table 7 that LIM improves significantly than FPN on the categories like "Portable Charger 1" (0.7%), "Portable Charger 2" (1.9%) and "Wa-

Method	**AVG**	PO1	PO2	WA	LA	MP	TA	CO	NL
SSD [20]	71.4	87.3	81.0	83.0	97.6	93.5	92.2	36.1	0.01
+FPN [17]	72.0	87.4	81.5	83.2	97.9	93.9	92.2	40.3	0.02
+PANet [39]	72.0	88.3	83.2	82.8	**97.9**	93.8	**92.6**	37.3	0.01
+LIM	**73.1**	**89.1**	**84.3**	**84.0**	97.7	**94.5**	92.4	**42.3**	**0.1**

Table 7. Comparing with feature pyramid mechanisms in the base SSD model, where the feature pyramid mechanisms and our LIM are implemented inside the base model respectively.

ter" (1.1%). The visual information like boundary of the three categories is more abundant in their X-ray images, demonstrating the effectiveness of Boundary Activation in our LIM and verifies the novel dense connection mechanism with specific feature enhancement.

5.4. Ablation Study

In this section, we conduct several ablation studies to in-depth investigate our method. We first analyze the effectiveness of the Dense mechanism by implementing the Single-directional Propagation (the left part of the Boundary Propagation) inside the base model. Then, we evaluate the performance of Boundary Activation alone that no boundary information aggregation inside the feature map. Moreover, we add the Boundary Activation module. The experimental results are shown in Table 8.

In Table 8, we can observe that the performance of the network with only Single-directional Propagation improves by 0.7% than the base model, verifying the effectiveness of our dense mechanism. After applying the propagation toward another direction, the performance improves by 1.2% than the base model and 0.5% than the Single-directional Propagation, which demonstrates the effectiveness of our bidirectional mechanism. Further, Table 8 shows that after the integration of our Boundary Activation module, the performance improves 1.7% than the base model and 0.5%

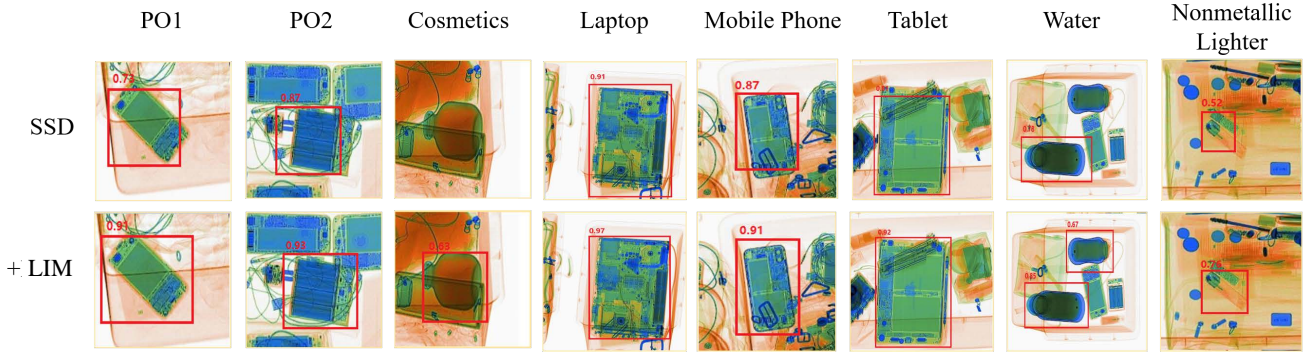| | PO1 | PO2 | Cosmetics | Laptop | Mobile Phone | Tablet | Water | Nonmetallic Lighter |

Figure 5. Visualization of the performance of both the baseline SSD and the LIM-integrated model. We select one pair of images of each category. The ability to identify and locate prohibited items of the LIM-integrated model outperforms the baseline SSD obviously.

Method	**AVG**	PO1	PO2	WA	LA	MP	TA	CO	NL
SSD [20]	71.4	87.3	81.0	83.0	97.6	93.5	92.2	36.1	0.01
+SP	72.1	87.9	82.3	83.8	97.9	92.4	92.6	38.8	0.63
+BP	72.6	88.1	83.4	83.9	**97.8**	93.8	**92.8**	40.3	0.03
+BP+BA	**73.1**	**89.1**	**84.3**	**84.1**	97.7	**94.5**	92.4	**42.3**	**0.1**

Table 8. Results of ablation study. Note that in this table, SP refers to the single-directional dense propagation (the left part of Boundary Propagation). BP refers to the Boundary Propagation and BA refers to the Boundary Activation.

Figure 6. Visualization of the effectiveness of our LIM.

than Boundary Propagation alone, indicating the effectiveness of boundary information aggregation inside the feature map. In conclusion, ablation studies have verified the validity of each part of our LIM model.

5.5. Visualization

In this section, we visualize the accuracy of recognition and localization in Figure 5 and the effectiveness of LIM and traditional boundary-enhanced methods in Figure 6.

Figure 5 shows that the LIM-integrated model has a significant improvement over the baseline. In columns 1, 2, 5, 6 and 8, the detection boundaries of prohibited items by the base SSD model are not precise enough and the LIM-integrated model performs better obviously. In column 3, the cosmetic escapes from the detection of the base SSD model but is caught by LIM-integrated model with the confidence of 91%. In column 7, the base SSD model only detects one prohibited item while there are two, but both of them are detected by LIM. Figure 6 illustrated the effectiveness of our LIM and traditional boundary-enhanced methods, including DOAM [40], EEMEFN [44], *etc.*

6. Conclusion

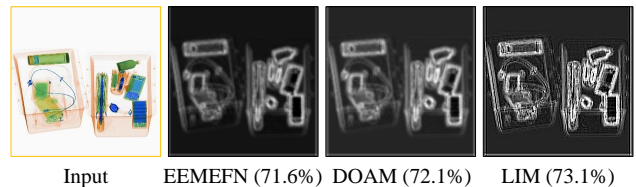In this paper, we investigate prohibited items detection in X-ray security inspection, which plays an important role in protecting public safety. However, this track has not been widely studied due to the lack of specialized public datasets. To facilitate research in this field, we construct and release a dataset with high-quality X-ray images for prohibited items detection, namely HiXray, including 8 categories of 102,928 common prohibited items. All images are gathered from the real-world scenario and manually annotated by professional inspectors. Besides, we propose the Lateral Inhibition Module (LIM) to address the problem that the items to be detected are usually overlapped with the stacked objects during X-ray imaging. Inspired by the lateral suppression mechanism of neurobiology, LIM eliminates the influence of noisy neighboring regions on the object regions of interest and activates the boundary of items by intensifying it. We comprehensively evaluate LIM on the HiXray and OPIXray dataset and the results demonstrate that LIM can improve the performance of SOTA detection methods. We hope that contributing this high-quality data set and LIM model can promote the rapid development of prohibited items detection in X-ray security inspection.

References

[1] Samet Akcay and Toby P Breckon. An evaluation of region based object detection strategies within x-ray baggage security imagery. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 1337–1341. IEEE, 2017. 2

[2] Samet Akcay, Mikolaj E Kundegorski, Chris G Willcocks, and Toby P Breckon. Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security imagery. *IEEE transactions on information forensics and security*, 13(9):2203–2215, 2018. 2

[3] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020. 2

[4] Arjun Chaudhary, Abhishek Hazra, and Prakash Chaudhary. Diagnosis of chest diseases in x-ray images using deep convolutional neural network. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pages 1–6. IEEE, 2019. 2

[5] Zhanzhan Cheng, Fan Bai, Yunlu Xu, Gang Zheng, Shiliang Pu, and Shuigeng Zhou. Focusing attention: Towards accurate text recognition in natural images. In *Proceedings of the IEEE international conference on computer vision*, pages 5076–5084, 2017. 2

[6] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010. 3

[7] Keren Fu, Qijun Zhao, and Irene Yu-Hua Gu. Refinet: a deep segmentation assisted refinement network for salient object detection. *IEEE Transactions on Multimedia*, 21(2):457–469, 2018. 2

[8] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323, 2011. 4

[9] Shuai Guo, Songyuan Tang, Jianjun Zhu, Jingfan Fan, Danni Ai, Hong Song, Ping Liang, and Jian Yang. 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*, pages 55–59, 2019. 2

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 6

[11] Binh-Son Hua, Minh-Khoi Tran, and Sai-Kit Yeung. Pointwise convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 984–993, 2018. 1

[12] Shengling Huang, Xin Wang, Yifan Chen, Jie Xu, Tian Tang, and Baozhong Mu. Modeling and quantitative analysis of x-ray transmission and backscatter imaging aimed at security inspection. *Optics express*, 27(2):337–349, 2019. 2

[13] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015. 4

[14] Glenn Jocher, Alex Stoken, Jirka Borovec, NanoCode012, ChristopherSTAN, Liu Changyu, Laughing, tkianai, Adam Hogan, lorenzomammana, yxNONG, AlexWang1900, Laurentiu Diaconu, Marc, wanghaoyang0106, ml5ah, Doug, Francisco Ingham, Frederik, Guilhen, Hatovix, Jake Poznanski, Jiacong Fang, Lijun Yu, changyu98, Mingyu Wang, Naman Gupta, Osama Akhtar, PetrDvoracek, and Prashant Rai. ultralytics/yolov5: v3.1 - Bug Fixes and Performance Improvements, Oct. 2020. 2, 6, 7

[15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. 1

[16] Hainan Li, Renshuai Tao, Jun Li, Haotong Qin, Yifu Ding, Shuo Wang, and Xianglong Liu. Multi-pretext attention network for few-shot learning with self-supervision. In *2021 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2021. 1

[17] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017. 4, 6, 7

[18] Aishan Liu, Jiakai Wang, Xianglong Liu, Chongzhi Zhang, Bowen Cao, and Hang Yu. Patch attack for automatic checkout. *CoRR*, abs/2005.09257, 2020. 1

[19] Jinyi Liu, Xiaxu Leng, and Ying Liu. Deep convolutional neural network based object detector for x-ray baggage security imagery. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 1757–1761. IEEE, 2019. 2

[20] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016. 2, 6, 7, 8

[21] Jianjie Lu and Kai-yu Tong. Towards to reasonable decision basis in automatic bone x-ray image classification: A weakly-supervised approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 9985–9986, 2019. 2

[22] Domingo Mery and Aggelos K Katsaggelos. A logarithmic x-ray imaging model for baggage inspection: Simulation and object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 57–65, 2017. 2

[23] Domingo Mery, Vladimir Riffo, Uwe Zscherpel, German Mondragón, Iván Lillo, Irene Zuccar, Hans Lobel, and Miguel Carrasco. Gdxray: The database of x-ray images for nondestructive testing. *Journal of Nondestructive Evaluation*, 34(4):42, 2015. 2, 3

[24] Domingo Mery, Erick Svec, Marco Arias, Vladimir Riffo, Jose M Saavedra, and Sandipan Banerjee. Modern computer vision techniques for x-ray testing in baggage inspection. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(4):682–692, 2016. 2

[25] Caijing Miao, Lingxi Xie, Fang Wan, Chi Su, Hongye Liu, Jianbin Jiao, and Qixiang Ye. Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in

overlapping images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2119–2128, 2019. 2, 3

[26] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3883–3891, 2017. 1

[27] Haotong Qin, Zhongang Cai, Mingyuan Zhang, Yifu Ding, Haiyu Zhao, Shuai Yi, Xianglong Liu, and Hao Su. Bipoint-net: Binary neural network for point clouds, 2020. 1

[28] Haotong Qin, Ruihao Gong, Xianglong Liu, Mingzhu Shen, Ziran Wei, Fengwei Yu, and Jingkuan Song. Forward and backward information retention for accurate binary neural networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1

[29] Haotong Qin, Ruihao Gong, Xianglong Liu, Mingzhu Shen, Ziran Wei, Fengwei Yu, and Jingkuan Song. Forward and backward information retention for accurate binary neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2250–2259, 2020. 2

[30] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016. 2

[31] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017. 2

[32] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 2

[33] Baoguang Shi, Xiang Bai, and Serge Belongie. Detecting oriented text in natural images by linking segments. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2550–2558, 2017. 2

[34] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 6

[35] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos: Fully convolutional one-stage object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 9627–9636, 2019. 2, 6, 7

[36] Ivan Uroukov and Robert Speller. A preliminary approach to intelligent x-ray imaging for baggage inspection at airports. *Signal Processing Research*, 4:1–11, 2015. 2

[37] Chien-Yao Wang, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh, and I-Hau Yeh. Cspnet: A new backbone that can enhance learning capability of cnn. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 390–391, 2020. 6

[38] Jiakai Wang, Aishan Liu, Zixin Yin, Shunchang Liu, Shiyu Tang, and Xianglong Liu. Dual attention suppression attack: Generate adversarial camouflage in physical world. *CoRR*, abs/2103.01050, 2021. 1

[39] Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, and Jiashi Feng. Panet: Few-shot image semantic segmentation with prototype alignment. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9197–9206, 2019. 4, 6, 7

[40] Yanlu Wei, Renshuai Tao, Zhangjie Wu, Yuqing Ma, Libo Zhang, and Xianglong Liu. 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*, page 138–146, 2020. 2, 3, 4, 6, 7, 8

[41] Huaxin Xiao, Jiashi Feng, Yunchao Wei, Maojun Zhang, and Shuicheng Yan. Deep salient object detection with dense connections and distraction diagnosis. *IEEE Transactions on Multimedia*, 20(12):3239–3251, 2018. 2

[42] Tom Young, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria. Recent trends in deep learning based natural language processing. *ieee Computational intelligenCe magazine*, 13(3):55–75, 2018. 1

[43] Jiliang Zhang, Binhang Qi, Zheng Qin, and Gang Qu. Hcic: Hardware-assisted control-flow integrity checking. *IEEE Internet of Things Journal*, 6(1):458–471, 2018. 1

[44] Minfeng Zhu, Pingbo Pan, Wei Chen, and Yi Yang. Eemefn: Low-light image enhancement via edge-enhanced multi-exposure fusion network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13106–13113, 2020. 8