ARTICLE

Research on Enhanced Contraband Dataset ACXray Based on ETL

Xueping Song[1,*], Jianming Yang[1], Shuyu Zhang[1] and Jicun Zhang[1,2,*]

[1]School of Mechanical Engineering, Dalian Jiaotong University, Dalian, 116028, China

[2]Neusoft Reach Automotive Technology (Dalian) Co., Ltd., Dalian, 116085, China

*Corresponding Authors: Xueping Song. Email: sxp0724@djtu.edu.cn; Jicun Zhang. Email: zhang_jc@neusoft.com

ABSTRACT

To address the shortage of public datasets for customs X-ray images of contraband and the difficulties in deploying trained models in engineering applications, a method has been proposed that employs the Extract-Transform-Load (ETL) approach to create an X-ray dataset of contraband items. Initially, X-ray scatter image data is collected and cleaned. Using Kafka message queues and the Elasticsearch (ES) distributed search engine, the data is transmitted in real-time to cloud servers. Subsequently, contraband data is annotated using a combination of neural networks and manual methods to improve annotation efficiency and implemented mean hash algorithm for quick image retrieval. The method of integrating targets with backgrounds has enhanced the X-ray contraband image data, increasing the number of positive samples. Finally, an Airport Customs X-ray dataset (ACXray) compatible with customs business scenarios has been constructed, featuring an increased number of positive contraband samples. Experimental tests using three datasets to train the Mask Region-based Convolutional Neural Network (Mask R-CNN) algorithm and tested on 400 real customs images revealed that the recognition accuracy of algorithms trained with Security Inspection X-ray (SIXray) and Occluded Prohibited Items X-ray (OPIXray) decreased by 16.3% and 15.1%, respectively, while the ACXray dataset trained algorithm's accuracy was almost unaffected. This indicates that the ACXray dataset-trained algorithm possesses strong generalization capabilities and is more suitable for customs detection scenarios.

KEYWORDS

X-ray contraband; ETL; data enhancement; dataset

1 Introduction

With the rapid development of the economy and foreign trade, the volume of customs clearance business has significantly increased. Consequently, the requirements for customs logistics supervision have become increasingly stringent [1]. Currently, customs still heavily relies on a large number of inspectors to manually inspect customs contraband X-ray images. The manual inspection method is both inefficient and costly. When various items are stored irregularly and their placement is obscured by obstructions, inspectors are easy to false detection and missing detections. Faced with complex regulatory environment and high customs clearance pressures, it is necessary to replace manual visual recognition with automatic detection to achieve intelligent recognition of contraband images [2].

With the development of graphics processing units (GPUs) and neural network technology, deep learning image recognition techniques have a wide range of applications in various fields. At customs supervision sites, deep learning demonstrates a relatively strong self-learning ability, which significantly reduces costs, improves efficiency, and enhances the accuracy of detecting contraband. Gaus et al. [3] employed convolutional neural networks to detect guns in luggage. Because of the adaptability of deep learning algorithms and X-ray images for contraband detection, it has become the main research algorithm of X-ray contraband detection [4,5]. When addressing the problem of X-ray image classification with deep convolutional neural networks, Akçay et al. [6] emphasized the importance of a training dataset with strong generalization ability to enhance the neural network's convolution accuracy. Similarly, Krizhevsky et al. [7] trained a large deep convolutional neural network and pointed out that when using machine learning methods, a larger dataset can improve the performance of machine learning algorithms. All researchers have shown that a large and stable contraband dataset is the prerequisite to ensure the accuracy of X-ray detection.

Deep learning object detection model training primarily relies on optical image datasets. While open optical image data is easily accessible, X-ray images can only be obtained using specialized equipment. Currently, the X-ray datasets used for contraband detection include Grima Data X-ray (GDXray), SIXray, China's High-quality X-ray (HiXray) dataset and OPIXray dataset.

Mery et al. [8] developed the GDXray dataset, which comprises 8,150 grayscale X-ray images featuring clear and easily distinguishable target outlines, simple scenes, and minimal object overlap and occlusion. However, the human eye is less sensitive to grayscale than to color images, which reduces the efficiency of customs inspectors in identifying low-density dangerous goods. Consequently, the GDXray dataset is not very effective in customs detection applications and is unsuitable for research into contraband detection algorithms. Miao et al. [9] created the SIXray dataset, a large collection of over one million X-ray images from realistic scenes, making it an ideal resource for training segmentation models. However, with less than 1% of the images being positive samples of contraband, the dataset has a significant imbalance between positive and negative samples. Additionally, SIXray focuses on five types of metal contraband and lacks samples of other materials, limiting the accuracy of algorithms trained with this dataset in detecting non-metal contraband. Tao et al. [10] constructed the HiXray dataset, containing 45,364 real-life X-ray images of 8 common types of prohibited items at airports. HiXray labeling is done manually by professionals, and it is costly and inefficient. When datasets need to be updated or expanded, manually labeling new sample images increases the cost, limiting the scalability and application scope. Wei et al. [11] established the OPIXray dataset, which includes 5 types of tools across 8,885 images. The limited number of X-ray contraband images in the OPIXray dataset presents challenges in training machine-learning-based automatic threat detection (ATD) algorithms due to data insufficiency.

Due to the different models of customs X-ray machines, various types of unstructured data are generated during inspections. This unstructured data needs to be standardized before it can be utilized effectively. Patel et al. [12] found that ETL is highly effective for converting unstructured data into structured data, facilitating real-time integration of large-scale data and the establishment of a data warehouse. The ETL [13–16] process involves extracting data from various data sources, transforming it through data cleansing to meet the storage requirements of the data warehouse, and then loading it into the warehouse. Selecting appropriate data cleansing tools can enhance cleaning efficiency and reduce errors [17]. Common data cleansing tools include OpenRefine [18], MapReduce [19], and Talend. Sreemathy et al. [20] proposed that Talend is suitable for data integration within ETL processes. Talend converts heterogeneous data into homogeneous data for easier analysis and stores all integrated data in the data warehouse for subsequent cleaning and filtering. Considering the need for consistency

with the data cleaning functions required in this study, Talend has been selected as the data cleaning tool. It enables the conversion of original non-standard data to standard data and uniformly transfers the pictures and structured data of different formats generated by X-ray machines. The transmission process should support various data transmission protocols to ensure the stability of the transmission [21]. Protocols used include HyperText Transfer Protocol (HTTP) [22], File Transfer Protocol (FTP) [23], Secure File Transfer Protocol (SFTP) [24], Message Queuing Telemetry Transport (MQTT) [25], and other data transmission protocols to meet the needs of different collection terminals.

Data annotation is an important part of dataset construction. Jaiswal et al. [26] stated that in supervised learning, algorithms for recognizing contraband need to be trained with labeled image samples. Andrews et al. [27] found that existing automated security image analysis methods focus on detecting specific categories of threats, and this detection method may not detect previously unseen contraband. When new categories of contraband are encountered, the dataset must be manually labeled again, leading to inefficient detection and an inability to update the dataset in real time. In the process of dataset construction, image enhancement is an important means to expand the number of sample datasets. Ghosh et al. [28] pointed out that image stitching technology can increase the number of positive samples in the dataset, but traditional image stitching technology struggles to fuse images with different backgrounds, which affects the effectiveness of image enhancement.

Supervised learning requires training and testing a model with a labeled dataset. Evaluation metrics are used to verify the performance of the algorithmic model. Supervised learning is mainly divided into classification problems and regression problems [29]. Algorithms for regression problems predict continuous targets, such as a person's height or weight, and the predictions are numerical. The commonly used evaluation metrics are Mean-Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE). The evaluation metrics for classification problems are Accuracy, Precision, Recall. Contraband detection belongs to classification problem [30]. The prediction result is the category of the target (contraband, non-contraband).

The contraband images in the current open source dataset were manually collected from customs X-ray machines and manually annotated by professionals. This process not only incurs high human costs but also limits the ability to update the dataset, preventing real-time updating and maintenance. As a result, it is challenging to cope with the complicated and dynamic situations encountered in current customs inspections. When the model is trained using these open source datasets or a small number of homemade datasets, the types and amounts of contraband are less than those likely to be encountered in real scenarios. This limitation makes it difficult for models trained on open source datasets to be effectively applied in the field of customs. It is necessary to construct a large-scale and high-quality X-ray contraband dataset, such as ACXray, which is crucial for improving the detection of contraband items in X-ray images.

Summarizing the work of predecessors, the main problems solved in this paper are as follows:

1) Various X-ray machines are distributed across different areas and come in a wide variety of models. Typically, a single-machine operation mode is adopted without any networking mechanism. Each X-ray machine produces a large number of images daily, which are stored locally using cyclic overlay, with only a short storage duration. Due to the lack of unified and standardized storage and management, all kinds of unstructured data cannot be effectively utilized or formed into valuable X-ray contraband sample assets.

2) Data annotation involves correctly framing objects to be labeled within an image. These labeled images are then fed into a machine learning algorithm to train the model. The quality of the dataset directly determines the effectiveness of the final model. The more comprehensive

and diverse the data, the better the training model. Automatic labeling by neural network algorithms can increase labeling speed, but achieving both accuracy and speed concurrently through neural network pre-labeling is challenging.

3) Currently, academic research primarily focuses on model training using open datasets or a limited number of self-made datasets. However, the types and quantities of contraband included in these training datasets are significantly less than those likely encountered in real-world scenarios. Consequently, models trained in this way are challenging to deploy and apply in practical engineering contexts. Moreover, different application scenarios require varied sample data for deep learning, and no single dataset is suitable for all types of X-ray machine contraband detection. Training multi-layer deep networks with a small number of X-ray contraband images tends to result in overfitting. Therefore, constructing a large, comprehensive dataset of X-ray contraband images for training in customs contraband detection is an urgent problem that needs to be addressed.

The main contributions of this paper are as follows:

1) A method for constructing a customs contraband sample dataset using ETL has been proposed. Image data from X-ray machines distributed across different areas can be collected and aggregated through an intelligent terminal. By cleansing the collected data, non-standard original data is converted to standard data, which can then be transmitted and stored in real-time using big data technology.

2) A method combining neural networks + manual annotation for labeling contraband data is proposed. Using a neural network algorithm for automatic labeling can enhance labeling speed. However, for data types that are difficult to distinguish, human correction of labels is necessary. Effective identification of positive samples is achieved through multiple rounds of training, thereby improving the labeling efficiency of contraband data.

3) To address the issues of a small number of contraband items in images from various X-ray machines, significant differences in placement angles, and variations in imaging colors, a method is proposed to enhance the x-ray contraband data by fusion of target and background. Utilizing the concept of instance segmentation, specific regions of contraband are identified, labeled, and cropped. These cropped images are then combined with other contraband images using image processing technology to generate a greater variety and quantity of contraband samples.

The process of building the ACXray Contraband dataset is shown in Fig. 1.
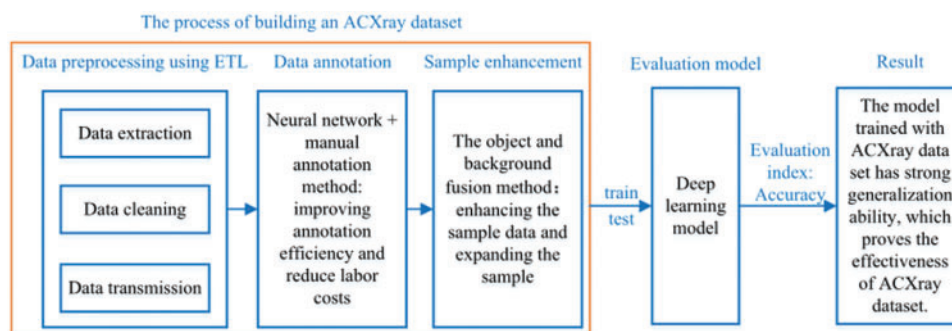

Figure 1: The process of building the ACXray contraband dataset

The organization of this paper is as follows: In Section 2, the collection of unstructured data using ETL is described, and the collected data is cleaned. Kafka message queues and the ElasticSearch distributed search engine are utilized to transfer the data to cloud servers in real-time. Section 3 proposes the method of using neural network + manual annotation to label contraband data, which improves the efficiency of contraband data labeling. Section 4 introduces the enhancement of X-ray image contraband data through object and background fusion methods to solve the problem of limited positive samples of contraband. Section 5 presents the impact of the ACXray dataset on the accuracy of contraband identification when training algorithms such as Mask R-CNN. Section 6 is discussion with some limitation. At the end of the paper, some important conclusions are drawn.

2 Data Preprocessing Using ETL

According to the characteristics of customs business, a large-scale and high-quality X-ray contraband dataset called ACXray is constructed specifically for airport customs. The ACXray dataset includes five common contraband items: knives, liquids, guns, cell phones, and power banks. The logical architecture of ETL processing of X-ray image data is shown in Fig. 2.
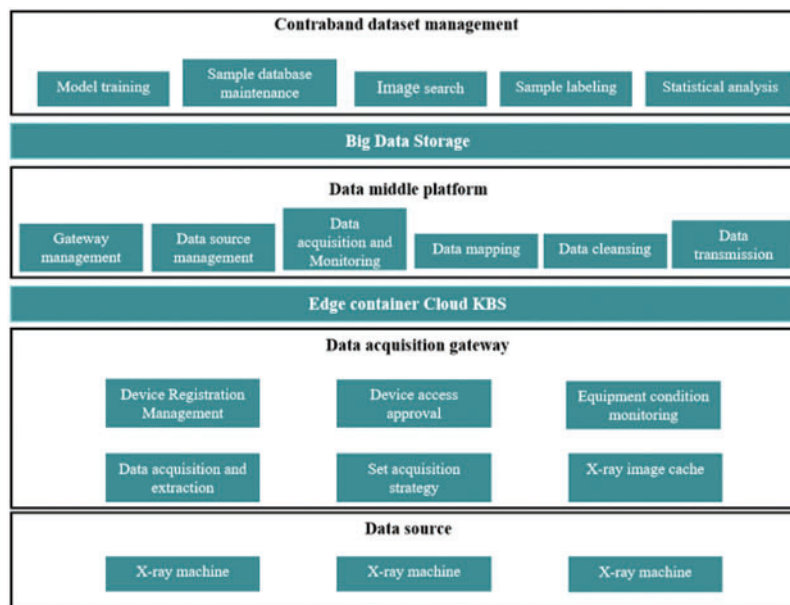

Figure 2: Data processing logical architecture diagram based on ETL

Data source: Different brands and types of X-ray machines produce different X-ray image data formats.

Data acquisition gateway: Connect to a variety of different X-ray machine data sources. Formulate acquisition strategy for structured data and unstructured data, extract data, set data cache area, and manage acquisition equipment.

Data middle platform: The hierarchical and horizontal decoupling of data is realized. The X-ray image data transmitted in real time by data acquisition gateway is cleaned and converted, and then aggregated into big data storage.

Contraband dataset management: Annotating the X-ray image data of the data middle platform, the contraband sample database can be continuously updated and maintained, image search, statistical analysis and other application.

2.1 Data Extraction

Data sources include X-ray picture files in different formats stored in various X-ray machines, which are extracted through data acquisition gateways.

Data extraction is the core stage of the data acquisition process, involving the process of collecting data from the edge data source to the target endpoint. Common data extraction methods include two approaches: Full extraction and incremental extraction. Full extraction involves extracting all data from the specified data source to achieve full coverage of data. Incremental extraction involves extracting a portion of data each time and capturing the changing data. Common incremental data extraction scenarios include: Incremental data extraction based on timestamp fields is completed by using incremental fields in data; When the data table does not have any incremental identification, the incremental extraction of snapshot analysis after the last data cache is completed by snapshot comparison. To reduce the network load and improve data acquisition efficiency, incremental data extraction is adopted in this paper, focusing solely on collecting newly generated X-ray image data.

The X-ray data includes: Data ID, data generation time, the customs outlet name, X-ray machine number, capture terminal network management device ID, X-ray image size, pixel value, X-ray image file, and data transmission log, etc. Among them, X-ray images and data transmission logs are unstructured data, while others are structured data. In the process of data collection, time series data is packaged by the data acquisition gateway and sent to the message queue as a packet for transmission.

2.2 Data Cleaning

Due to the lack of X-ray machine data standards, invalid data and garbage data may be generated, which affects the efficiency and accuracy of data acquisition.

Data cleaning carries out verification, filtering, removing duplicates, error correction, format conversion, consistency checks, and other processing for different sources and different types of data to ensure and improve data quality. Automatic data cleaning functions include cleaning missing values, cleaning format content, removing unreasonable values, and eliminating unnecessary fields. If the system cannot automatically process the data, it provides a hint, and the data is then processed manually. Data association involves linking the cleaned, standardized data with fundamental and knowledge data. This process not only improves but also supplements the foundational and knowledge data, thereby enhancing both their quality and value density. Data comparison encompasses structured data comparison, keyword comparison, and similar tasks.

One of the key tasks of data cleaning is to formulate cleaning rules. For each type of data quality problems, the corresponding cleaning rules and cleaning algorithms are proposed. By implementing these rules and algorithms, "dirty data" is transformed into "clean data" that meets the requirements of users or applications. The object description method is used to describe the cleaning rules.

The cleaning rules take the following form:

{

Type of cleaning rules (Input errors | Missing field values | Field values are not in the value range | Single field contains multiple field values | Reference field values missing or error | Similar duplicate records | Record sets do not meet the association constraints | Exception);

Error judgment conditions;

Name of the record set to be cleaned;

Name of the field to be cleaned;

Record keywords to be cleaned (If not empty, this rule is valid only for that record, otherwise it is valid for the entire record set);

Cleaning strategy (Ignore | Throw | Automatic conversion);

Script management (Some automatic processing scripts predefined by the program; User-defined processing scripts; User-defined processing script after an error is thrown);

}

2.3 Kafka+Flink Real-Time Data Transmission

Due to the high timeliness requirements of the contraband dataset, a technology architecture based on Hadoop is adopted, utilizing the combination of Kafka and Flink to build the computing engine and achieve real-time data ingestion. Kafka distributed message queue technology supports high throughput message publishing and subscription. The Kafka cluster is deployed as the system's message bus on two machines. Machine configuration: Central Processing Unit (CPU): E5-2630 V4 $\times$ 2, Memory: 16 GB Dynamic Double Rate 4th-generation Synchronous (DDR4) $\times$ 2. Flink is used as a framework for stateful computation of data streams and as a distributed processing engine. It can operate in all common cluster environments and perform computations at in-memory speed and at any scale. Flink is extensively applied in distributed real-time computing scenarios, where storage and computation are spread across multiple servers and subsequently aggregated. The processing speed can reach the millisecond level, meeting the real-time requirements.

2.4 Internet of Things (IoT) X-Ray Image Data Collection

The system employs a self-developed intelligent collection box terminal to address the issue of varying interfaces and X-ray image format differences among different X-ray machine manufacturers. The system configuration of the intelligent collection box is shown in Table 1.

Table 1: The system configuration of the intelligent collection box

Name	Configured parameters
Processor	Intel® Core™ I7-4712HQ, I5-4410E, I3-4110E, Celeron® 2000E
Chipset (CS)	Mobile Intel QM87 Express
Internal storage	2 GB DDR3 SO-DIMM

In addition to the basic function of data access, the intelligent data collection box also has the capability to convert data formats. As shown in Fig. 3, it can operate steadily for an extended period (365 days $\times$ 24 h) without manual intervention, ensuring real-time performance.

By connecting intelligent data collection boxes to 50 X-ray machines located in different areas, the system can perform real-time reading of X-ray images. These images are then transmitted to an ETL data processing system via a 4G router or a high-bandwidth dedicated network. Concurrently, the data center stores these X-ray images for the purpose of training a contraband database.
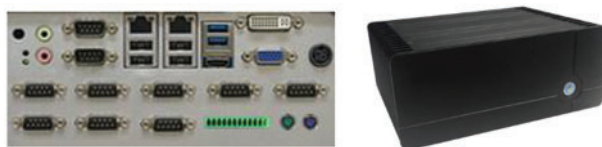
Figure 3: Intelligent data acquisition terminal

In the X-ray image data transmission process, considering the aspects of security and stability, a data transfer method based on the FTP is adopted. An FTP server is deployed on the front-end data collection box system, which is connected to the X-ray machine via an Ethernet interface. Utilizing the FTP transmission protocol, the original files generated by the X-ray machine scans are collected by the data collection agent (collection box). These files are then transmitted to the cloud-based data center through a 4 G router, as depicted in Fig. 4.
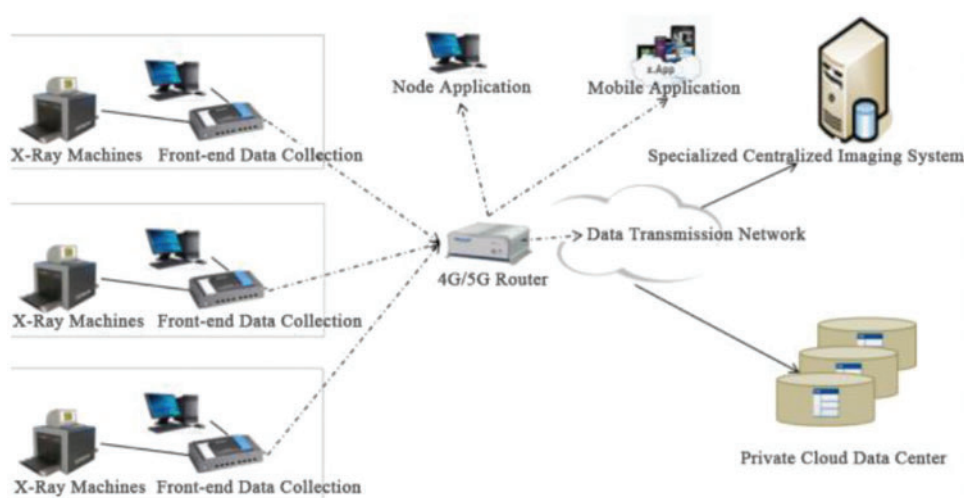
Figure 4: Topological structure diagram of X-ray machine internet of things

3 Combining Neural Networks and Manual Contraband Data Annotation

Image annotation is the process of accurately selecting the objects that need to be marked in the image, and then inputting the generated annotation data into the machine learning algorithm to train the model.

Data is crucial for training machine learning models, and the quality of the dataset directly determines the effectiveness of the final model. The more abundant and representative the data is, the more effective the training model will be, and the stronger the algorithm's robustness. Previously, manual annotation was the primary method used by annotators to label contraband items such as knives, liquid bottles, guns, smartphones, and power banks in images. However, the data workload required for X-ray image algorithms is substantial, resulting in slow manual annotation speed. While using neural network algorithms for automatic annotation can increase the speed of annotation, achieving both precision and speed simultaneously may not be possible with neural network pre-annotation. Therefore, this paper adopts a combined annotation strategy of neural networks and manual annotators, as shown in Fig. 5.
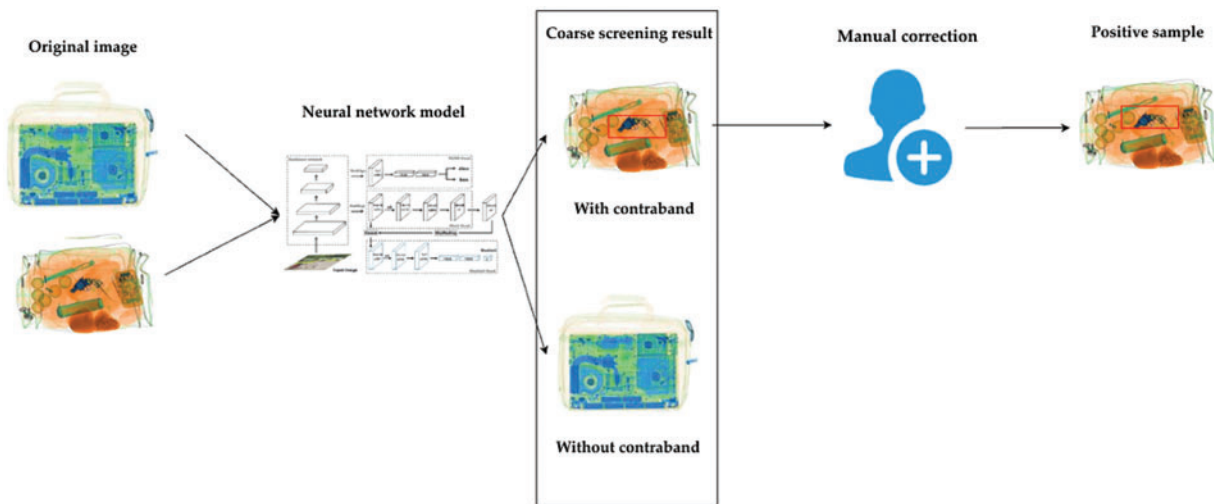
Figure 5: Schematic diagram of data annotation

Firstly, a part of the image data is manually selected and labeled to train the initial model, and the rest of the unlabeled data is predicted by the trained model. Secondly, the data that are difficult to distinguish in the model are selected. The labels of these "hard" data are artificially modified and added to the training set to fine-tune the training model again. Finally, after several training, the neural network model used for data annotation will have a high recognition accuracy and achieve the effect of automatically determining positive samples.

The neural network model comprises a total of 18 layers, which include 1 convolutional layer, 1 maximum pooling layer, and 4 sets of sub-modules. Each set of sub-modules consists of 4 convolutional layers of size 3×3. The input is a 416×416 pixel image. After passing through the initial convolutional layer, it transforms into an image of size $208 \times 208 \times 64$. Subsequently, it undergoes a maximum pooling layer, leading to an image of size $104 \times 104 \times 64$. The changes in the image after passing through the convolutional layers of the 4 sets of sub-modules are as follows: After the first set of 4 convolutions, the image becomes a size of $104 \times 104 \times 64$. After the second set of 4 convolutional layers, it becomes a size of $52 \times 52 \times 128$. After the third set of 4 convolutional layers, it becomes a size of $26 \times 26 \times 256$. Finally, after the fourth set of 4 convolutional layers, the image becomes a size of $13 \times 13 \times 512$, as shown in Table 2.

Using a combination of neural network automatic labeling + manual annotation can significantly improve efficiency and reduce labeling costs. As shown in Table 3, the comprehensive annotation efficiency of manual annotation is approximately 20 X-ray images per minute. The annotation efficiency for positive samples of contraband is 10 X-ray images per minute, while for negative samples, it is 30 X-ray images per minute. The combined efficiency of neural network + manual annotation method has reached 100 X-ray pictures per minute, which is 5 times faster than the manual annotation. The annotation efficiency for positive samples is 90 X-ray images per minute, which is 9 times faster than manual annotation alone.

Table 2: Neural network architecture

Number	Size	Strided convolutions	Image size after passing through the layer
1	3×3	1	$104 \times 104 \times 64$
2	3×3	1	$104 \times 104 \times 64$
3	3×3	1	$104 \times 104 \times 64$
4	3×3	1	$104 \times 104 \times 64$
5	3×3	2	$52 \times 52 \times 128$
6	3×3	1	$52 \times 52 \times 128$
7	3×3	1	$52 \times 52 \times 128$
8	3×3	1	$52 \times 52 \times 128$
9	3×3	2	$26 \times 26 \times 256$
10	3×3	1	$26 \times 26 \times 256$
11	3×3	1	$26 \times 26 \times 256$
12	3×3	1	$26 \times 26 \times 256$
13	3×3	2	$13 \times 13 \times 512$
14	3×3	1	$13 \times 13 \times 512$
15	3×3	1	$13 \times 13 \times 512$
16	3×3	1	$13 \times 13 \times 512$

Table 3: Efficient comparison between manual and semi-automatic

Method	Synthetical labeling efficiency	Positive sample labeling efficiency	Negative sample labeling efficiency
Manual annotation	20 sheets/min	10 sheets/min	30 sheets/min
Automatic annotation + Manual annotation	100 sheets/min	90 sheets/min	110 sheets/min

4 Sample Enhancement of Object and Background Fusion Method

Data enhancement significantly increases the amount of training data, thereby improving the model's generalization ability. Enhancing the original data can be viewed as introducing noise, which in turn enhances the model's robustness. Data enhancement can be categorized into two types: offline enhancement and online enhancement. Offline enhancement is the direct expansion of the dataset. Online enhancement involves applying various transformations such as rotation, translation, scaling, cropping, and flipping to batch data after it is obtained. This enhances the sample data during the training process, based on the premise that the information from the enhanced samples can be calculated using a formula. Many machine learning frameworks already support this kind of data enhancement and can use GPUs to optimize calculations.

Due to the small number of positive samples of X-ray contraband, geometric and color transformations are applied to existing samples through data enhancement to expand the positive sample size. Among the methods of sample data enhancement used in X-ray security image training, the method

that does not require re-annotation is suitable for online enhancement, while the method that requires re-annotation is suitable for offline enhancement.

When enhancing samples from X-ray security images, flipping, rotation, affine transformation, colour transformation, saturation, and contrast transformation are taken into account. Among them, the effect of flipping is more obvious in the process of sample enhancement.

Sample flip

In some scenarios, such as vehicle detection on the road, pedestrian detection, affected by gravity, the object is always a head-up, and vertical flip cannot be used for sample enhancement. Since the image of the X-ray security inspection machine is a top-down perspective, the orientation of the object can be in any direction, and the sample can be enhanced by using two flipping methods: Horizontal and vertical. Take the knife commonly used in customs contraband detection as an example, as shown in Fig. 6, samples of knives with different viewing angles can be obtained by image flipping.
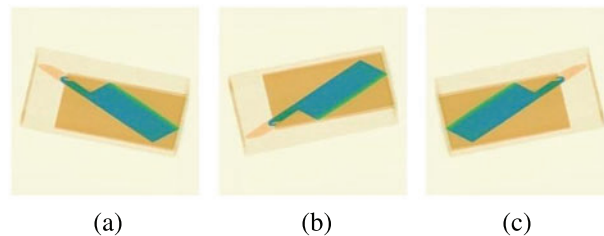

(a)   (b)   (c)

Figure 6: Image Flip: (a) Original image; (b) Flip vertical; (c) Flip horizontal

4.1 Sample Data Enhancement with Combination Method

One of the challenges in constructing an X-ray image contraband dataset is the scarcity of real contraband, which leads to an imbalance between positive and negative samples. Collecting real X-ray images of contraband is time-consuming and labor-intensive. To address this, contraband is extracted from existing images using image synthesis techniques. These contraband elements are then combined with ordinary X-ray images to create new X-ray image samples that include contraband, thus mitigating the issue of limited real contraband samples.

The traditional image stitching method involves cutting the labeled contraband object frame as a whole and stitching it into the object image to create new contraband samples. This method is easy to operate, requires no manual intervention, and can be performed online during model training. A major drawback of this method is that it retains the background around the object edges from the original image due to the direct cropping of the contraband's bounding box, as depicted in Fig. 7. Consequently, the background features around the target in the sample image obtained by stitching are essentially the same as those in the original image, and background discrepancies with the target image reduce the effectiveness of data enhancement. Given the variety of angles and backgrounds in X-ray image contrabands, there is significant potential for further data enhancement. Based on the concept of instance segmentation, specific areas of contraband are marked and cropped using the annotation methods of instance segmentation. Through image processing techniques, the cropped contraband is then merged with other non-contraband images using iconographic methods to create a new, more realistic sample, as illustrated in Fig. 8.
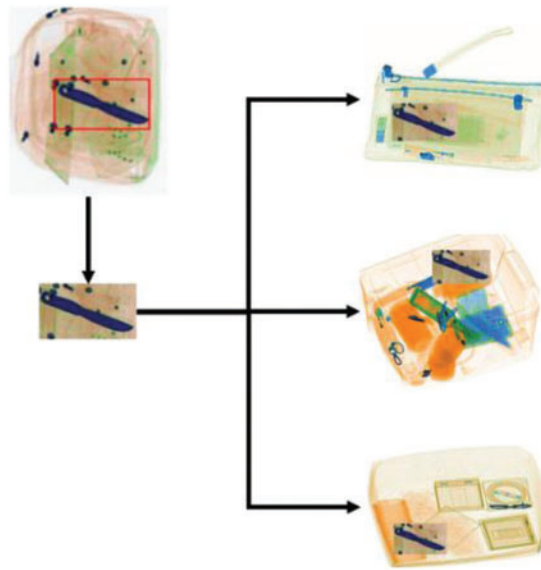
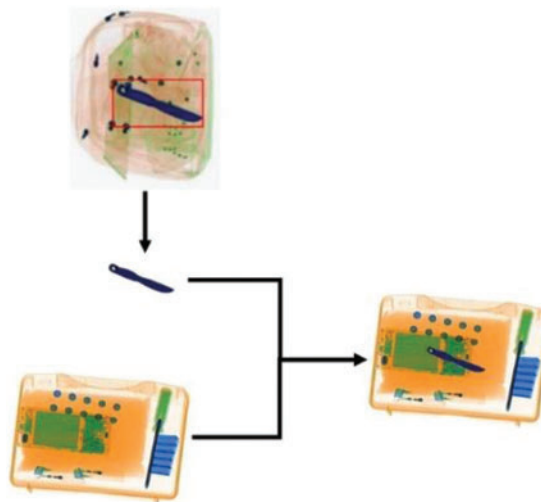**Figure 7:** Data enhancement of traditional image

Figure 8: Artificially fused images of contraband mosaic methods

Annotation instances by manual means often require a lot of labor costs. Therefore, the pre-trained instance segmentation model can be combined with the image method to obtain a crude annotation sample, as shown in Fig. 9.

These coarsely labeled instances are often incomplete or noisy regions, and the results need to be corrected manually. As shown in Fig. 10, the corrected instances of contraband are fused with X-ray images without contraband by combining methods such as rotation, flipping, and color transformation to generate new samples of contraband.
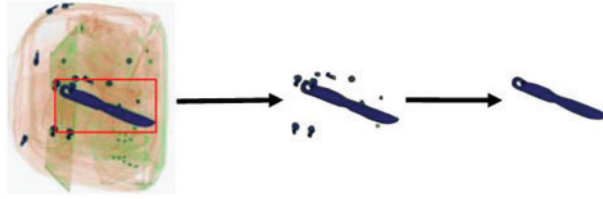
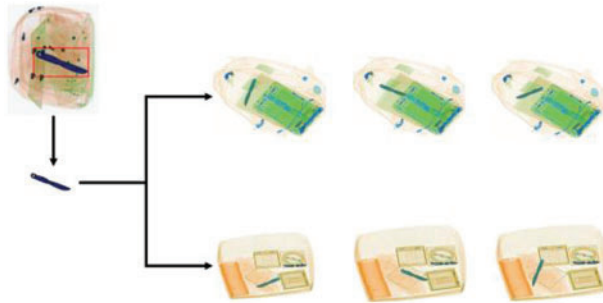**Figure 9:** Segmentation object extraction of contraband

Figure 10: Generating new samples of contraband by combination method

The cropped contraband is annotated, with its background area marked in white. Through a binary method, the position information matrices A^O and A^B are obtained. Let M be the set of all pixels contained within the object. Both A^O and A^B have the same dimensions as the image. P_{ij} represents the pixel point at coordinates (i,j) in the image.

$$A^O_{ij} = \begin{cases} 1, & p_{ij} \in M \\ 0, & p_{ij} \notin M \end{cases} \tag{1}$$

$$A^B_{ij} = \begin{cases} 0, & p_{ij} \in M \\ 1, & p_{ij} \notin M \end{cases} \tag{2}$$

Let the crop cut out contraband image matrix be P^O, the stitched area image matrix of the stitched image is P^B, and the A^O, A^B, P^O, P^B size is the same, then the spliced area image matrix P^F is:

$$P^F = A^O \times P^O + A^B \times P^B \tag{3}$$

Here, matrix multiplication refers to Hadamard product, which means multiplying the corresponding elements at each position. Let A^1, A^2 be matrices of size $m \times n$. The Hadamard product of A^1, A^2, denoted as A^3, has the same size of $m \times n$. The value of element A^3_{ij} in column j of row i be:

$$A^3_{ij} = A^1_{ij} \times A^2_{ij} \tag{4}$$

Among them, A^1_{ij}, A^2_{ij} are the elements of the i-th row and j-th column of A^1, A^2.

Suppose the element at the i-th row and j-th column of the fused image matrix P^F is denoted as P^F_{ij}. The elements at the i-th row and j-th column of matrices P^O and P^B are as follows:

$$P^O_{ij} = \left(r^O, g^O, b^O \right) \tag{5}$$

$$P^B_{ij} = \left(r^B, g^B, b^B \right) \tag{6}$$

Obviously, the elements of row i and column j of A^O and A^B satisfy:

$$A^O_{ij} = 1 - A^B_{ij} \tag{7}$$

$$P^F_{ij} = \begin{cases} \left(r^O, g^O, b^O\right), A^O_{ij} = 1 \\ \left(r^B, g^B, b^B\right), A^O_{ij} = 0 \end{cases} \tag{8}$$

The contraband data augmentation process for the combined method is shown in Fig. 11, and the process is as follows:

1) The contraband region is coarsely labeled by algorithm and the contraband targets with noise are extracted;
2) The coarsely labeled target contraband is corrected by manual correction, remove the background noise and form a new contraband instance;
3) By applying rotations, flips, and color transformations, image transformations are performed on the object contraband, generating more instances of contraband as required;
4) The position information matrix of the target contraband and non-contraband images is calculated, and matrix multiplication operation is performed using the Hadamard product method;
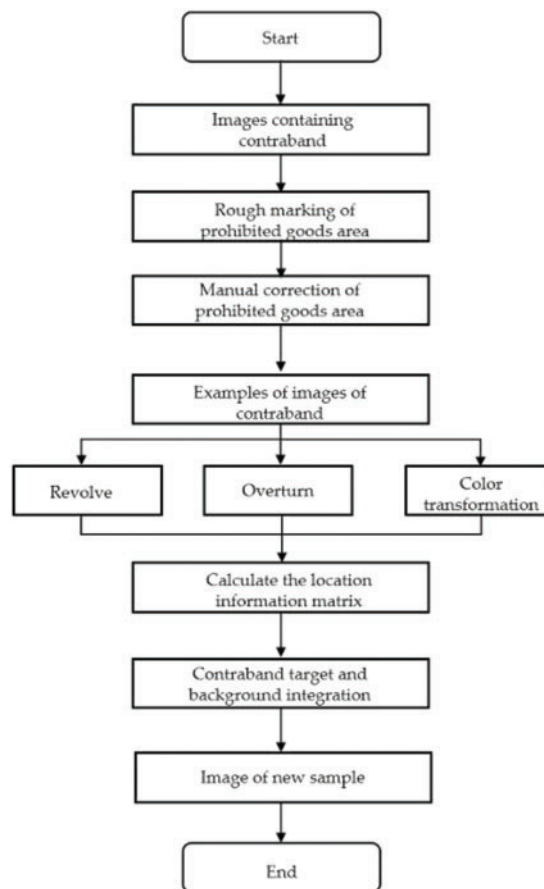5) Fusion of target contraband with non-contraband images to generate new contraband samples.

Figure 11: The process of enhancing contraband data with combined methods

The combination method can generate synthetic contraband images, and obtain contraband samples that are closer to the real data. It effectively solves the problem of less contraband in the real environment, and effectively expands the contraband type and quantity of the contraband sample library. The contraband samples generated by the combined method of image enhancement are used to train the deep learning model, and the accuracy of the model is greatly improved.

5 Experimental Results and Analysis

5.1 System Performance Testing

X-ray image acquisition, intelligent retrieval modules, and algorithm modules are all deployed in clusters. As the scale expands, the volume of X-ray image data access will continually increase. Each service node can smoothly scale horizontally by expanding hardware, thereby ensuring high availability of the service. In a testing environment, this system is capable of processing over 10 million X-ray image data entries per day.

System function testing employs the Cypress End to End (E2E) Web testing framework to conduct automatic User Interface (UI) testing on the application layer of the Web platform. Each test result is saved from the server and generates txt files to verify the functionality and operational effectiveness of the system. The test results demonstrate that the application functions operate normally, and the UI response time is less than 100 ms, which satisfies the actual requirements.

The system performance test is to monitor the response time of each module of the system while performing UI automation test. When new data is generated, the response time of each module is monitored to verify the performance of each module of the system. Table 4 shows the average response time of each module when adding data.

Table 4 : Average response time of each module when data is added

Data stream	Kafka message bus	ES search engine
100 records per second	5 ms	120 ms
200 records per second	7 ms	190 ms
500 records per second	12 ms	300 ms
1000 records per second	14 ms	410 ms
2000 records per second	17 ms	630 ms

Test results indicate that the response time of each function of the system is less than 1 s when the X-ray image data reaches 100 million. When incoming data rates increase to 2000 records per second, that is, more than 100 million new data per day, the Kafka message bus runs within 20 ms, the ES search engine runs at about 630 ms, and the X-ray image intelligent retrieval runs at more than 500 ms.

The transmission, storage and operation time of each module also reach the response of seconds. In terms of performance, it can achieve real-time analysis and real-time query, which can meet the needs of current users.

In the future, with the increase of customs X-ray machine equipment, the amount of new daily data will continue to increase. Kafka and ES nodes can be considered through hardware expansion, and the load capacity of the system can be improved through cluster deployment, so as to meet the real-time requirements of the system.

5.2 Using ETL to Collect Contraband Images

The X-ray image acquisition system based on ETL collects data from X-ray images located in various places. After labeling the contraband items, a contraband dataset named ACXray is formed, which is conforms to customs business. As shown in Table 5, the ACXray dataset contains a total of 34,900 images, of which 21,000 positive samples and 13,900 negative samples. The specific sample of ACXray dataset is shown in Fig. 12.

Table 5: Quantity of contraband in ACXray dataset

Positive (58,470)				
Knife	Liquid	Gun	Phone	Portable battery
8300	21,200	3050	13,578	12,342
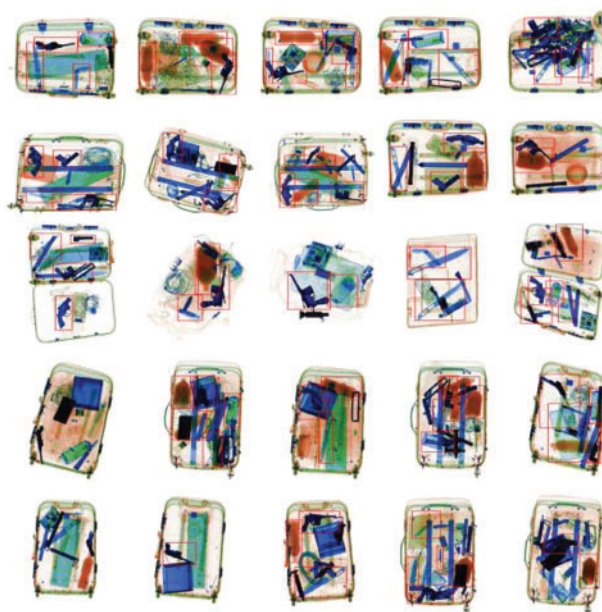
Figure 12: Sample of ACXray dataset

5.3 ACXray Dataset Verification

In order to verify the effectiveness of the ACXray dataset for contraband detection using deep learning algorithms, four deep learning algorithms are selected: Mask R-CNN, Faster Region-based Convolutional Neural Network (Faster R-CNN), You Only Look Once v3 (YOLOv3), and Single Shot MultiBox Detector 513 (SSD513). The open-source datasets SIXray, OPIXray and ACXray proposed in this chapter are selected for training. Since the GDXray dataset is composed of gray images, the imaging mode is quite different from other X-ray datasets, and there is no comparability, so GDXray is not selected as the comparative evaluation object in this experiment. Contraband detection is a classification problem. The object recognition experiment is carried out on the test dataset, and

Accuracy was used as the evaluation index:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{9}$$

In formula (9), Accuracy is defined as the proportion of the correct predicting ones in all samples; True Positives (TP) represents positive samples are detected as positive samples; False Positives (FP) denotes negative samples are detected as positive samples; True Negatives (TN) denotes negative samples are detected as negative samples; False Negatives (FN) represents positive samples are detected as negative samples. The object recognition experiment is carried out on the test dataset of contraband, 80% of the data was used as the training dataset, and the remaining 20% was used as the test dataset. In addition, 400 real X-ray images of customs operations are selected as a test set to test the algorithm. The detection accuracy of X-ray contraband images is shown in Table 6.

Table 6: Accuracy comparison using different datasets

Algorithm	Training dataset	SIXray	OPIXray	ACXray
Mask R-CNN	Training dataset and test set from the same source	71.7%	66.3%	70.3%
	Test data: 400 customs X-ray images	55.4%	51.2%	70.2%
Faster R-CNN	Training dataset and test set from the same source	56.6%	66.3%	45.7%
	Test data: 400 customs X-ray images	31.2%	55.2%	45.2%
YOLOv3	Training dataset and test set from the same source	46.3%	64.2%	39.2%
	Test data: 400 customs X-ray images	32.5%	50.1%	39.6%
SSD	Training dataset and test set from the same source	42.5%	56.1%	38.5%
	Test data: 400 customs X-ray images	20.1%	42.5%	38.8%

When the training and test sets are the same data source, the four algorithms perform well in both the SIXray and OPIXray datasets. The Mask R-CNN algorithm is trained and tested using the SIXray dataset, achieving an identification accuracy of 71.7%, which is the highest among the three dataset test results. The Faster R-CNN algorithm is trained and tested using the OPIXray dataset and achieved a recognition accuracy of 66.3%.

When 400 customs X-ray images are used as the test set for testing, the recognition accuracy of the Mask R-CNN algorithm trained by the SIXray training set is 55.4%, which decreased by 16.3%. After training on the OPIXray training set, the recognition accuracy of the Mask R-CNN algorithm is 51.2%, with a decrease of 15.1%. However, after training on the ACXray training set, the recognition accuracy of the Mask R-CNN algorithm is 70.3%, which is a decrease of 0.1% in accuracy. The results show that the recognition accuracy of the Mask R-CNN algorithm trained by the ACXray training set does not decrease under different test data. It shows that the model trained by ACXray dataset has strong generalization ability. Therefore, ACXray is better suited for customs X-ray contraband detection.

5.4 Image Data Enhancement Verification

The enhanced image data is used to train the model and verify the effect of the enhanced data on the detection accuracy. The effect of X-ray image data enhancement is shown in Fig. 13. After image enhancement, the edges of the package and its contents become clearer than in the original image, making it easier to identify contraband. Image darkening intensifies the color of the package edges and the items inside, facilitating the distinction between different types of items. Furthermore, when the colored edges are enhanced, the item edges become more pronounced, and the outline of the item is more distinct.
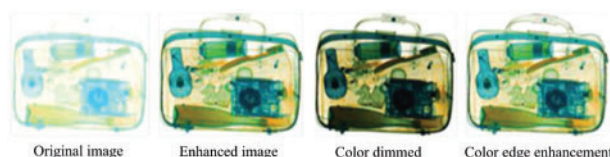
Figure 13: Effect of X-ray image data augmentation

Through color transformation data enhancement, the feature of one X-ray image wrapped item is extended to four parcels, effectively expanding the dataset.

The three datasets of SIXray, HiXray, and ACXray are combined and trained using Mask R-CNN to compare the detection effects of data enhancement and no data enhancement, as shown in Table 7.

Table 7: Comparison of the accuracy of Mask R-CNN algorithm after data enhancement training

Algorithm	Knife	Liquid	Gun	Phone	Portable battery
Mask R-CNN	65.5%	84.1%	91.3%	84.0%	65.4%
Mask R-CNN (Data enhancement)	67.7%	89.5%	92.8%	87.9%	66.7%
Improved accuracy	2.2%	5.4%	1.5%	3.9%	1.3%

As shown in Table 7 and Fig. 14, the accuracy of Mask R-CNN algorithm is significantly different in identifying different types of contraband samples. When the algorithm model trained with the enhanced dataset identifies 5 kinds of contraband, the accuracy is higher than that of the algorithm model trained without the enhanced dataset. Through the model trained by the data-enhanced dataset, the recognition accuracy of the knife increased from 65.5% to 67.7%, an increase of 2.2%. The recognition accuracy of liquid increased from 84.1% to 89.5%, an increase of 5.4%. Additionally, the recognition accuracy of the gun increased from 91.3% to 92.8%, resulting in a 1.5% improvement. Moreover, the recognition accuracy of mobile phones increased from 84.0% to 87.9%, representing a 3.9% improvement. Furthermore, the recognition accuracy of the power bank increased from 65.4% to 66.7%, showing a 1.3% increase. Overall, the generalization ability and recognition accuracy of the model have been improved.
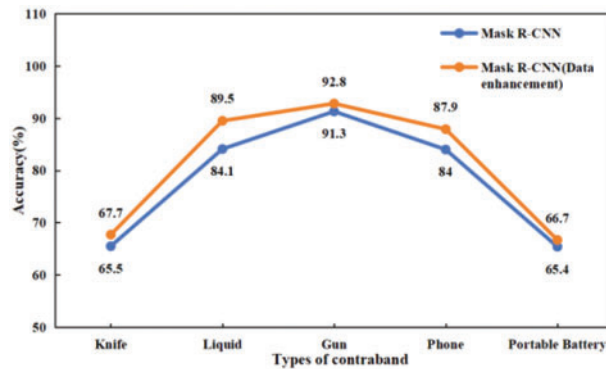
Figure 14: The accuracy comparison of Mask R-CNN algorithm before and after data enhancement

6 Discussion

The ACXray dataset based on ETL is more suitable for customs business scenarios than other open datasets and has better application results. Data source is formed through data acquisition based on task scheduling and data cleaning based on rules. The data source is transferred to the cloud for storage, forming a networking mechanism, so that different data scattered in different places can be collected in real time. This helps to expand the number of positive samples and increase the type and number of samples. The object and background fusion method is used to enhance the X-ray contraband data, which improves the quality of X-ray image data and reduces the false positive rate in commercial detection.

The contraband images in the existing open source dataset usually contain a lot of noise. If Convolutional Neural Networks (CNN) model is trained using the existing open source dataset, the model cannot achieve satisfactory performance in the actual customs detection scenario. Dataset construction and model development is necessary to updating. In this paper, Mask R-CNN algorithm is using to train and test the dataset. Since the current deep learning algorithm is updated quickly, the customs must continue to update the algorithm to maintain its advanced nature and achieve better detection results. Because the algorithm trained by ACXray dataset has strong generalization, it can adapt to the constant updating iteration of the algorithms.

This paper solves the current problem that the datasets in various fields of customs are not common, but the ACXray dataset also has some limitations: It is only applicable to customs contraband at present, and contraband data conforming to other scenarios are yet to be developed. It is still necessary to continuously expand the variety of contraband samples to meet the needs of actual customs detection.

Difficulties in the use of ACXray dataset are: the huge scale of data will lead to the problem of insufficient storage and computing; the acquisition and storage of front-end data as well as the structured processing of image data will affect the quality of the dataset; with the increase in the number of accessed customs X-ray machine equipment, it is necessary to improve the system's loading capacity in order to satisfy the system's real-time requirements.

For companies, algorithmic models trained with existing open source datasets tend to have low recognition rates when they recognize the real images of customs. The algorithm model trained by ACXray dataset shows stronger generalization ability and higher recognition accuracy in the

identification of real customs contraband. The difficulty of deploying the model directly to real engineering applications is solved by using ACXray dataset.

The ACXray dataset can be updated in real time using ETL. The annotation method of neural network + manual annotation can significantly improve the annotation efficiency. Through these methods, the frequency of dataset update is effectively improved, thereby reducing labor costs and saving expenses. These methods can reduce labor costs and improve the frequency of dataset update, bringing greater economic benefits to enterprises. However, there are difficulties in the process of companies using ACXray datasets: The dataset contraband categories need to be regularly updated and expanded; since the ACXray dataset is an important part of the automated image system, it is even more costly to have private cloud data storage center when applying the automated image system.

7 Conclusion

A dataset designed for customs contraband detection has been developed. The ETL method was used to address the issue of X-ray images in the customs security inspection field being dispersed across various locations, and the unstructured data lacking unified standards for storage and management. The method of blending targets with backgrounds was applied to enhance contraband data, increasing the variety and quantity of contraband items and addressing the shortage of positive contraband samples. Deep learning algorithms were used for testing, and the results indicate that the ACXray dataset possesses strong generalization capabilities in customs X-ray image recognition. This highlights the effectiveness of the ACXray dataset in customs contraband detection. However, the ACXray dataset still has some limitations. In future research, the number and types of samples in the dataset will continue to expand to meet customs needs. By exploring new optimization algorithms and improving image detection architectures, the ACXray dataset can be better utilized to enhance the accuracy of contraband detection.

Author Contributions: Study conception and design: Xueping Song, Jicun Zhang; data collection: Jianming Yang, Shuyu Zhang; analysis and interpretation of results: Jianming Yang; draft manuscript preparation: Jianming Yang, Shuyu Zhang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Due to the commercial nature of this research, so supporting data is not available.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] E. Drobot, A. Klevleeva, P. Afonin, and S. Gamidullaev, "Risk management in customs control," *Econ. Reg.*, vol. 13, no. 2, pp. 550–558, 2017. doi: 10.17059/2017-2-19.

[2]   P. Kuppusamy, M. Sanjay, P. V. Deepashree, and C. Iwendi, "Traffic sign recognition for autonomous vehicle using optimized yolov7 and convolutional block attention module," *Comput. Mater. Contin.*, vol. 77, no. 1, pp. 445–466, Oct. 2023. doi: 10.32604/cmc.2023.042675.

[3]   Y. F. A. Gaus, N. Bhowmik, and T. P. Breckon, "On the use of deep learning for the detection of firearms in X-ray baggage security imagery," in *Proc. HST*, Woburn, MA, USA, Nov. 5–6, 2019, pp. 5–6. doi: 10.1109/HST47167.2019.9032917.

[4]   Y. T. Zhang, H. G. Zhang, T. F. Zhao, and J. F. Yang, "Automatic detection of prohibited items with small size in X-ray images," *Optoelectron. Lett.*, vol. 16, no. 4, pp. 313–317, Jul. 2020. doi: 10.1007/s11801-020-9118-x.

[5]   A. Presenti, Z. Liang, L. F. A. Pereira, J. Sijbers, and J. D. Beenhouwer, "Automatic anomaly detection from x-ray images based on autoencoders," *Nondestructive Test. Eval.*, vol. 37, no. 5, pp. 552–565, Jun. 2022. doi: 10.1080/10589759.2022.2074415.

[6]   S. Akçay, M. E. Kundegorski, M. Devereux, and T. P. Breckon, "Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery," in *Proc. ICIP*, Phoenix, AZ, USA, Sep. 2016, pp. 1057–1061. doi: 10.1109/ICIP.2016.7532519.

[7]   A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 24, 2017. doi: 10.1145/3065386.

[8]   D. Mery *et al.*, "GDXray: The database of X-ray images for nondestructive testing," *J. Nondestruct Eval.*, vol. 34, no. 42, pp. 84–90, Nov. 13, 2015. doi: 10.1007/s10921-015-0315-7.

[9]   C. Miao *et al.*, "SIXray: A large-scale security inspection X-ray benchmark for prohibited item discovery in overlapping images," in *Proc. CVPR*, Long Beach, CA, USA, Jun. 15–20, 2019, pp. 2114–2123. doi: 10.1109/CVPR.2019.00222.

[10] R. Tao *et al.*, "Towards real-world X-ray security inspection: A high-quality benchmark and lateral inhibition module for prohibited items detection," in *Proc. ICCV*, Montreal, QC, Canada, Oct. 2021, pp. 10–17. doi: 10.1109/ICCV48922.2021.01074.

[11] Y. L. Wei, R. S. Tao, Z. J. Wu, Y. Q. Ma, L. B. Zhang, and X. L. Li, "Occluded prohibited items detection: An X-ray security inspection benchmark and de-occlusion attention module," in *Proc. 28th ACM Int. Conf. Multimed. Assoc. Comput. Mach.*, New York, USA, Oct. 12, 2020, vol. 12, pp. 138–146. doi: 10.1145/3394171.3413828.

[12] M. Patel and D. B. Patel, "Progressive growth of ETL tools: A literature review of past to equip future," in *Proc. AISC*, Singapore, Springer, Oct. 2, 2020, vol. 1187, pp. 389–398. doi: 10.1007/978-981-15-6014-9_45.

[13] J. Sreemathy, R. Brindha, M. S. Nagalakshmi, N. Suvekha, N. K. Ragul and M. Praveennandha, "Overview of ETL tools and talend-data integration," in *Proc. ICACCS*, Coimbatore, India, Mar. 19–20, 2021, pp. 1650–1654. doi: 10.1109/ICACCS51430.2021.9441984.

[14] A. Wojciechowski and R. Wrembe, "Research problems of the ETL technology," *Found. Comput. Decis. Sci.*, vol. 35, pp. 283–305, Jan. 2010. https://api.semanticscholar.org/CorpusID:60086434.

[15] A. Nambiar and M. Divyansh, "An overview of data warehouse and data lake in modern enterprise data management," *Big Data Cogn. Comput.*, vol. 6, no. 4, pp. 132, 2022. doi: 10.3390/bdcc6040132.

[16] V. Raman and J. M. Hellerstein, "Potter's wheel: An interactive data cleaning system," in *Proc. 27th VLDB Conf.*, Roma, Italy, Sep. 30, 2001, vol. 1, no. 4, pp. 381–390.

[17] M. Z. Wang and Z. C. Li, "Research status and prospect of data extraction and cleaning technology in large environment," in *Proc. ICDSBA*, Cham, Springer, Mar. 28, 2018, pp. 293–299. doi: 10.1007/978-3-319-72745-5_32.

[18] T. F. Kusumasari and Fitria, "Data profiling for data quality improvement with OpenRefine," in *Proc. ICITSI*, Bandung, Indonesia, Oct. 24–27, 2016, pp. 1–6. doi: 10.1109/ICITSI.2016.7858197.

[19] T. M. Tuan, T. T. Ngan, and N. T. Trung, "Object detection in remote sensing images using picture fuzzy clustering and MapReduce," *Comput. Syst. Sci. Eng.*, vol. 43, no. 3, pp. 1241–1253, May 9, 2022. doi: 10.32604/csse.2022.024265.

[20] J. Sreemathy, V. I.Joseph, S. Nisha, I. C.Prabha, and R. M. G.Priya, "Data integration in ETL using TALEND," in *Proc. ICACCS*, Coimbatore, India, Mar. 6–7, 2020. doi: 10.1109/ICACCS48705.2020.9074186.

[21] Y. X. Liu *et al.*, "APMD: A fast data transmission protocol with reliability guarantee for pervasive sensing data communication," *Pervasive Mob. Comput.*, vol. 41, no. 9, pp. 413–435, Oct. 2017. doi: 10.1016/j.pmcj.2017.03.012.

[22] A. Y. Cheng, H. S. Shen, and X. Y. Lin, "Implementation of vehicle logistics information transmission based on HTTP communication protocol," *Appl. Mech. Mater.*, vol. 8, no. 2, pp. 122–129, 2014. doi: 10.1504/IJWMC.2015.068627.

[23] D. Springall, Z. Durumeric, and J. A. Halderman, "FTP: The forgotten cloud," in *Proc. DSN*, Toulouse, France, Jun. 2016, pp. 503–513. doi: 10.1109/DSN.2016.52.

[24] J. K. Rout, S. K. Bhoi, and S. K. Panda, "SFTP: A secure and fault-tolerant paradigm against blackhole attack in MANET," *Int. J. Comput. Appl.*, vol. 64, no. 4, pp. 27–32, Feb. 2014. doi: 10.5120/10623-5343.

[25] D. Alhejaili and O. H. Alhazmi, "Lightweight algorithm for MQTT protocol to enhance power consumption in healthcare environment," *J. Internet of Things*, vol. 4, no. 1, pp. 21–33, May 16, 2022. doi: 10.32604/jiot.2022.019893.

[26] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, "A survey on contrastive self-supervised learning," *Technologies*, vol. 9, no. 2, pp. 2, Dec. 28, 2021. doi: 10.3390/technologies9010002.

[27] J. T. A. Andrews, N. Jaccard, T. W. Rogers, and L. D. Griffin, "Representation-learning for anomaly detection in complex X-ray cargo imagery," in *Proc. ADIX II. Int.*, Anaheim, CA, USA, SPIE, May 1, 2017, vol. 10187, pp. 46–56. doi: 10.1117/12.2261101.

[28] D. Ghosh and N. Kaabouch, "A survey on image mosaicing techniques," *J. Vis. Commun. Image Represent.*, vol. 34, no. 3, pp. 1–11, Jan. 2016. doi: 10.1016/j.jvcir.2015.10.014.

[29] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised classification algorithms in machine learning: A survey and review," in *Emerg. Technol. Model. Graph.: Proc. IEM Graph.*, Singapore, Springer, Jul. 17, 2019, pp. 99–111. doi: 10.1007/978-981-13-7403-6_11.

[30] Y. Zhang, W. Xu, S. Yang, Y. Xu, and X. Yu, "Improved YOLOX detection algorithm for contraband in X-ray images," *Appl. Opt.*, vol. 61, no. 21, pp. 6297–6310, Jul. 15, 2022. doi: 10.1364/AO.461627.