Scientific Research Publishing

Journal of Computer and Communications, 2023, 11, 208-225
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Application of Dual-Energy X-Ray Image Detection of Dangerous Goods Based on YOLOv7

Baosheng Liu[1], Fei Wang[1], Ming Gao[2], Lei Zhao[1]

[1]School of Computer Science and Technology, Shandong University of Technology, Zibo, China
[2]Shanghai Wuying Technology Co., Ltd., Shanghai, China
Email: 854530534@qq.com

How to cite this paper: Liu, B.S., Wang, F., Gao, M. and Zhao, L. (2023) Application of Dual-Energy X-Ray Image Detection of Dangerous Goods Based on YOLOv7. *Journal of Computer and Communications*, **11**, 208-225.
https://doi.org/10.4236/jcc.2023.117013

Received: July 1, 2023
Accepted: July 28, 2023
Published: July 31, 2023

Abstract

X-ray security equipment is currently a more commonly used dangerous goods detection tool, due to the increasing security work tasks, the use of target detection technology to assist security personnel to carry out work has become an inevitable trend. With the development of deep learning, object detection technology is becoming more and more mature, and object detection framework based on convolutional neural networks has been widely used in industrial, medical and military fields. In order to improve the efficiency of security staff, reduce the risk of dangerous goods missed detection. Based on the data collected in X-ray security equipment, this paper uses a method of inserting dangerous goods into an empty package to balance all kinds of dangerous goods data and expand the data set. The high-low energy images are combined using the high-low energy feature fusion method. Finally, the dangerous goods target detection technology based on the YOLOv7 model is used for model training. After the introduction of the above method, the detection accuracy is improved by 6% compared with the direct use of the original data set for detection, and the speed is 93FPS, which can meet the requirements of the online security system, greatly improve the work efficiency of security personnel, and eliminate the security risks caused by missed detection.

Keywords

X-Ray, Dangerous Goods Detection, High and Low Energy Image Fusion, Accuracy, Real-Time Detection

1. Introduction

In order to protect people's personal and property safety when taking public transport, safety inspection has become a necessary means to ensure the safety of people's lives and property, so it is necessary to prevent people from carrying dangerous goods on board or entering public places. X-ray security technology is the core component of security equipment. X-ray source emits radiation to irradiate the inspected items, and detector receives the X-ray passing through the object and converts it into an X-ray image with different gray values. Security inspectors usually judge whether dangerous goods are contained by looking at the shape of the object in the X-ray image. This method has been widely used in major transportation hubs, public places, tourist attractions, and so on. The X-ray image quality is affected by the density, composition and spatial placement of the object. In the process of security check, due to the arbitrariness of the placement angle and the uncertainty of the types of dangerous goods, dangerous goods may be ignored due to shielding, which is not conducive to the judgment of security personnel; coupled with the high-intensity work of security inspectors and other reasons, there will often be missed detection and false detection, causing security risks or reducing work efficiency, in addition, due to the high mobility of China's population and large passenger flow, the demand for security inspection technology and requirements are also increasing.

With the development of deep learning, object detection technology based on convolutional neural networks has been widely used in the industrial field. Target detection technology is the further development of classification technology, it can not only predict the target category, but also give the location information and confidence of the target. At present, the popular object detection framework is based on convolutional neural networks. Convolutional neural networks are mainly composed of convolutional layer, pooling layer and fully connected layer. The convolutional neural network reduces the complexity and training difficulty of the network model by using three strategies: local sensitivity field, weight sharing and downsampling. It is not affected by affine transformations, such as translation, scaling and rotation of images, and has a strong feature extraction ability. Current target detection technologies are mainly divided into two categories: two-stage and one-stage. Two-stage mainly includes R-CNN [1], Fast R-CNN [2] and Faster R-CNN [3]. Such technologies need to use heuristic methods or convolutional neural networks to generate pre-selection frames. Then do classification and regression on the pre-selection box. One-stage mainly includes SSD [4] and YOLO [5] series, such technologies extract features and predict the location and category information of the target directly over the network. Two-stage target detection technology needs to carry out multiple operation detection, which has a large amount of calculation, high precision but slow speed, and can not meet the real-time requirements of the security inspection system. One-stage object detection technology speeds up the detection speed and reduces the detection accuracy.

In order to meet the speed requirements of the security inspection system, this paper uses the one-stage target detection framework based on YOLOv7 [6]. At the same time, to meet the accuracy requirements, this paper uses a dangerous goods data set expansion method using Threat Image Projection (TIP) technology to solve the problem of insufficient real packages containing dangerous goods data [7]. In addition, a high- and low-energy image fusion method is developed to enhance the image and make full use of the two energy X-ray images of the dual-energy X-ray security detector. The experimental results show that satisfactory results are obtained on the data set collected by Shanghai Wuying Technology Co., Ltd. and meet the real-time requirements of the security check system, which can be integrated into the security check system.

2. Research Status

2.1. Dangerous Goods Detection Algorithm in Traditional X-Ray Images

Before the target detection network model based on convolutional neural networks is widely used, the research on X-ray dangerous goods target detection is limited. Mikolaj E. Kundegorski1, and Samet Akcay1 [8] used various feature point descriptors as visual vocabulary variants in the Bag of Visual Words (BoVW) representation method, and with the support of support vector machines and random forest classification, Object detection in X-ray baggage screening using a series of feature point detectors and descriptors. The model based on a bag of visual words can only grasp the local information of the image, but cannot make full use of the global information of the image.

Wang *et al.* [9] proposed a detection method based on Scale Invariant Feature Transform (SIFT) [10] and Implicit Shape Model (ISM); the SIFT algorithm is used to extract the key points of the target, and the ISM model of the target is constructed. In the detection process, the extracted SIFT descriptor of the target is matched with the visual descriptor in the ISM model, and the voting mechanism is used to determine whether the target is a hazardous material.

2.2. X-Ray Image Dangerous Goods Detection Method Based on Deep Learning

After the Convolutional Neural Network (CNN) was proposed, it has been widely used in various fields and has become an indispensable part of various object detection models in computer vision. Krizhevsky *et al.* [11] proposed an image classification method based on convolutional neural networks, which achieved record-breaking results in image classification at that time and was far ahead of the second place in the ILSVRC-2012 competition. The proposed neural network has made a breakthrough in the method based on convolutional neural networks in the field of computer vision.

Akcay [12] first applied convolutional neural networks to the field of X-ray images, and discussed the applicability and effectiveness of the traditional sliding window-based convolutional neural network detection pipeline and area-based object detection technology in the problem of object detection in X-ray security images. Based on this, Akcay *et al.* [13] proposed the use of deep convolutional neural networks and transfer learning [14] to solve the problem of image classification and detection in X-ray luggage security images. For classification problems, CNN is compared with the traditional BoVW method based on hand-crafted features, and a Support Vector Machine classifier is used to train the traditional hand-crafted features. Various CNNS were investigated to understand the impact of network complexity on the overall performance. In addition to classification, object detection strategies have been investigated to further improve the performance on cluttered datasets where classification techniques fail.

Lu *et al.* [15] proposed a detection algorithm of dangerous goods in security check packages based on improved YOLOv3 [16], which reduced the original prediction of three bounding boxes per grid in Y0L0v3 to two bounding boxes. K-means clustering was used to calculate the prior box according to the data set, and data enhancement method was adopted, multi-scale input training strategy was adopted. The detection speed and accuracy are improved to some extent.

Wu *et al.* [17] proposed to improve the detection method of X-ray security dangerous goods by combining atrous convolution and transfer learning to improve YOLOv4 [18]. By increasing the receptive field, multi-scale context information is aggregated, the initial candidate box is obtained by K-means clustering algorithm, and the learning rate is optimized by cosine annealing to accelerate model training, which can effectively reduce the false detection rate of dangerous goods. And it improves the detection ability of small targets.

Liu *et al.* [19] proposed an object detection method for X-ray images. Firstly, a color-based foreground-background segmentation method was proposed to contour the detected object, and then Faster-RCNN, an object detection framework based on deep convolutional neural networks, was used to achieve a mAP of 77%.

Zhang *et al.* [20] proposed an improved SSD [4] algorithm and its application in subway security inspection. The convolution operation of each scale feature in the SSD algorithm is unchanged in size, and the corresponding features before and after convolution are fused in lightweight network to generate a new pyramid feature layer, and the detection unit based on the residual module is added to avoid increasing the capacity and computational complexity of the network model. This paper aims to solve the problem of easy missed detection and low detection accuracy when detecting small targets.

Han [21] proposed a detection and tracking algorithm for dangerous goods in X-ray images based on deep learning. On the one hand, a deep learning detection network based on improved single-shot multi-box detection method is designed to improve the detection accuracy. On the other hand, a tracker based on the detection results is implemented, and real-time detection and tracking is achieved through the cooperation between the tracker and the detector.

In various studies, the object detection algorithm based on YOLO [5] can reduce the amount of calculation and speed up the training of the model under the condition of ensuring accuracy. Because the X-ray security inspection system requires real-time performance, the YOLO object detection model has become the mainstream method in this field.

3. Method

The experiment is mainly divided into the following steps: 1) performing Threat Image Projection on the empty package, inserting the separately collected dangerous goods into the empty package to create fake dangerous goods data; 2) performing high- and low-energy image fusion operation on the data, combining high- and low-energy images into one image. 3) YOLOv7 object detection model is used for training and prediction.

3.1. Threat Image Projection

Collecting the data of parcels containing dangerous goods is a time-consuming and labor-intensive work. The dangerous goods in each collected parcel are artificially placed, and will not contain all the positions and angles of dangerous goods in real life. In order to reduce the labor cost and time cost of collecting hazardous materials parcel data, and improve the complexity and variability of hazardous materials placement. In this paper, a Threat Image Projection (TIP) [7] method is used, which can insert different dangerous goods into different luggage packages at various angles and positions. Therefore, while expanding the training data set, TIP times can be increased for individual dangerous goods with fewer samples to solve the problem of sample imbalance. After TIP insertion, the number of dangerous goods in each parcel is balanced.

TIP is a baggage screening technique used to train security agents and automatic threat recognition algorithms. Dangerous goods are placed in the tray separately for collection, and the acquired image is processed by simple threshold to obtain clean images of dangerous goods from the background. Then the affine transformation is applied to the dangerous goods image, and random rotation or scaling is carried out, so that dangerous goods of different angles and different sizes can be fully considered. Carry out threshold processing and morphological operations on the collected baggage package A that does not contain dangerous goods, obtain the image of the baggage package area B, bring the minimum external rectangular coordinates of the dangerous goods data into the baggage package area, limit the insertion range, and ensure that the inserted dangerous goods are located in the entire baggage package area. At the selected effective position $M\left(i, j\right)$, the dangerous goods image D is superimposed on the baggage package image B to generate a composite composite image C.

In order to ensure the reliability of the synthesized TIP image C, two parameters are introduced in image fusion. The parameter α controls the transparency of the original image A ($\alpha = 0.9$). Another parameter is the dangerous goods pixel threshold T, which ensures the consistency of the original image$A$ and the target image B in image contrast. The purpose of using the dangerous goods pixel threshold T is to remove high-value pixels inserted into the dangerous goods image so that the inserted dangerous goods are not visually too bright compared to the target area of the empty package overlay. The dangerous goods image threshold T can be empirically calculated by the following formula:

$$T = \min\left(\exp\left(x^5\right) - 0.5, 0.95\right) \tag{1}$$

where x is the normalized average intensity of the inserted area in B, and the calculation formula is as follows:

$$x = \frac{\sum_{i,j} B(i,j) * M(i,j)}{\sum_{i,j} 255 * M(i,j)} \in [0,1] \tag{2}$$

Image composition can be expressed as follows:

$$C(i,j) = \begin{cases} (1-\alpha)B(i,j) + \alpha A(i',j'), & M(i,j) = 1 \text{ and } A(i',j') < T * 255, \\ B(i,j), & \text{otherwise} \end{cases} \tag{3}$$

In the formula, $A(i,j)$ represents the pixel value of row i and column j of image A, and the other values are the same; Since the T value calculated by Equation (1) is in the range of 0.5 - 0.95, any pixel in image A higher than $T * 255$ will be ignored during image synthesis.

As shown in **Figure 1**, 1) the truly collected dangerous goods package data, and 2) the TIP method is used to insert the separately collected dangerous goods into the package that does not contain dangerous goods to generate the true composite image containing dangerous goods. In this method, dangerous goods are randomly inserted into the image at several angles to make up for the problem of expensive collection of dangerous goods, and avoid the problem of expensive collection of certain dangerous goods data (such as drugs, explosives).

3.2. High- and Low-Energy Image Fusion

In this paper, a method of high- and low-energy image fusion is proposed to enhance the fine granularity of X-ray images, make the outline of X-ray images clearer, and help the network model to learn the characteristics of dangerous goods in X-ray images. The dual energy X-ray image is composed of a high-energy image and a low-energy image. The ray is first illuminated on the low-energy detector, and the detector obtains the low-energy signal value. A copper sheet is used between the low-energy detector and the high-energy detector to filter out

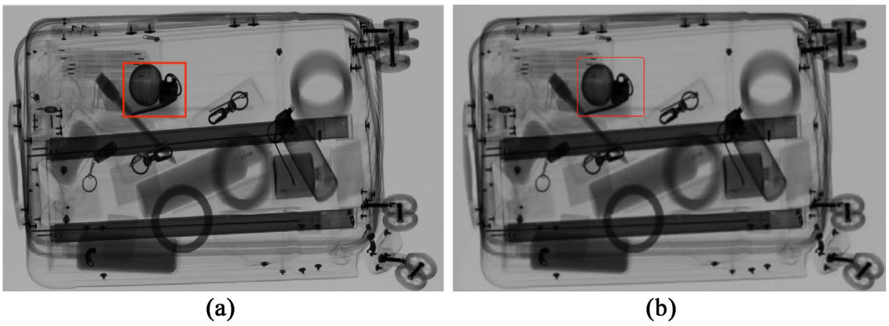

Figure 1. (a) Real dangerous goods data; (b) Synthetic dangerous goods data.

the low-energy rays and irradiate the high-energy rays on the high-energy detector, resulting in high-energy images. The gray value of the low-energy image is small, and the gray value of the high-energy image is large. The objects with thick internal structure are displayed more clearly in the high-energy image, and the objects with thin internal structure are displayed more clearly in the low-energy image.

The value of each pixel after the fusion of high- and low-energy images is calculated as follows:

$$factor = \frac{Low(x,y)}{Max(Low)} \tag{4}$$

$$V_{i,j} = Low(i,j) * factor + High(i,j) * (1 - factor) \tag{5}$$

where *Low* is the low-energy image as shown in Figure 2(a), *Max* (*Low*) is the maximum value in the low-energy image, and *factor* is the ratio of the current pixel value and the maximum value in the low-energy image. *High* denotes the high-energy image as shown in Figure 2(b), *i* denotes the *i*th row of the image, and *j* denotes the *j*th column of the image. The gray value of the low-energy image is small, and the penetration effect is not good, so the gray value is assigned a large weight coefficient. The high-energy image has a large gray value and good penetration effect, which assigns a small weight coefficient to the gray value. Using the above formula, the high- and low-energy images were fused to obtain an image with richer feature information and clearer contour.

Since the flat panel detector acquired X-ray images with high dynamic range, in order to display the high dynamic image to a common display device, a hue mapping algorithm based on multi-scale local edge preserving filter was used to convert the low-low-energy fusion image into a low-dynamic range image [22], and the low-low-energy fusion image was passed through the LEP filter to obtain the base layer image representing the approximate information. Then the gray value of the corresponding position of the base layer image is different, and the

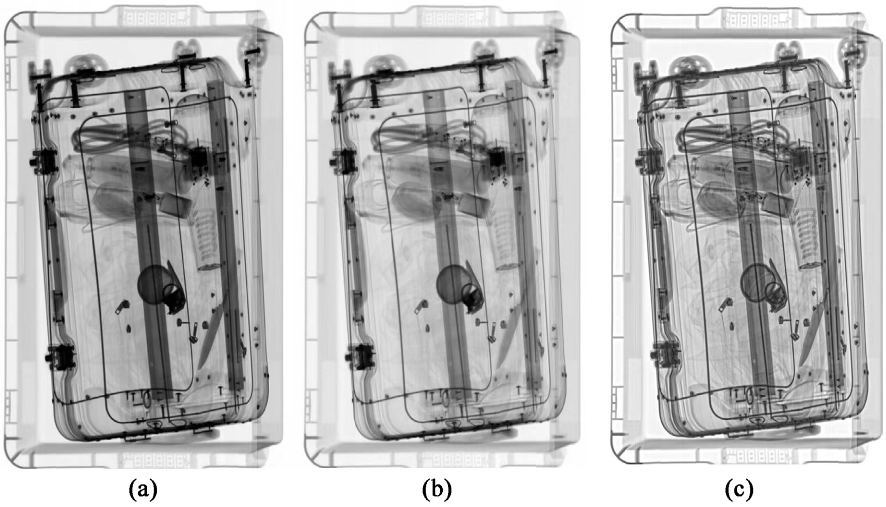

Figure 2. (a) Low-energy map; (b) High-energy map; (c) Image fusion.

detail layer image representing the fine edge is obtained. After two similar decomposition of the base layer image, the original image is decomposed into one base layer and three detail layer images. The detail information of each detail layer image is enhanced and fused with the base layer image, and the contrast of the image is improved by histogram equalization to obtain a low dynamic range image that retains the rich details in the original image as shown in **Figure 2(c)**.

3.3. Object Detection Network Model Structure

The dangerous goods package data set is obtained through the above two methods, and the YOLOv7 [6] model is used for training. YOLOv7 is a one stage target detection algorithm. The research shows that it is better than the previous version in accuracy and speed. Since X-ray dangerous goods detection needs to consider both accuracy and speed, YOLOv7 is selected as the model for the detection of this dangerous goods dataset in this paper. YOLOv7 is mainly composed of four parts: Input, backbone, neck, and head. The following are introduced separately.

After TIP dangerous goods insertion and high- and low-energy image fusion, the data set is input into the network model. There are several characteristics of the data collected by X-ray security equipment. 1) The size of the image is fixed, and the size of each item will not change after several times of collection. 2) The whole X-ray image is grayscale, without color and texture information, only the internal structure information of the material is retained. 3) X-ray images are different from optical images. X-ray images are not affected by illumination, so there is no significant color change in the image. Therefore, the input side is preprocessed by Mosaic data enhancement, adaptive image scaling and other preprocessing, randomly using four X-ray images for random scaling, and then random splicing, which greatly enriches the detection data set, especially the random scaling increases many small targets. It can make the model learn some potentially valuable information, so as to improve the generalization ability of the model, deal with more complex application scenarios, and make the network more robust.

The Backbone part is shown in **Figure 3**. After preprocessing, the data is sent to the backbone network for feature extraction. The backbone network is mainly composed of CBS module, E-ELAN module and MP module. CBS is composed of Conv convolution layer, BN batch normalization layer and Silu loss function. CBS module is superimposed by convolution kernels and step sizes of different scales to reduce the number of channels while extracting shallow features. E-ELAN module is composed of CBS modules with different convolution kernel sizes and step sizes, which is divided into two branches. The first branch changes the number of channels after a 1 × 1 convolution, and the other branch performs feature extraction after a 1 × 1 convolution and three CBS modules of 3 × 3 convolution. Finally, the output features of the two branches are superimposed.

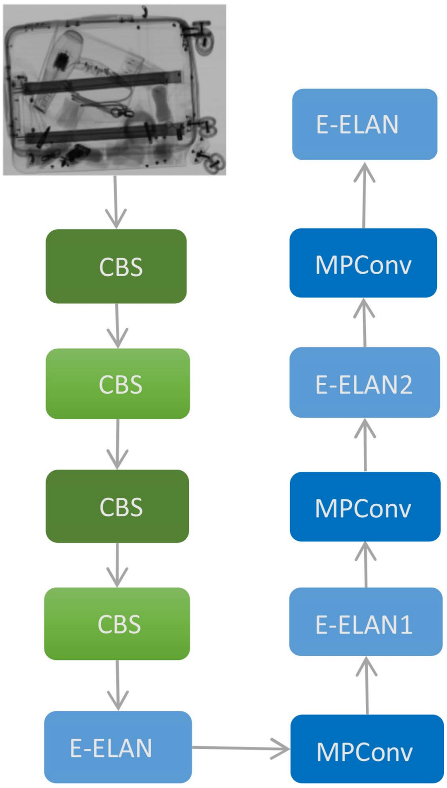

Figure 3. This diagram is the backbone of the network structure Backbone. The CBS = Convolution + Batch Normalization + SiLU, where (3, 2) means that the convolution kernel size is 3 and the step size is 2. E-ELAN is composed of multiple CBS, the input and output feature size remains unchanged. The MPConv consists of a maximum pooling operation and a CBS module, the main purpose of which is to perform down sampling operations.

The MP module performs the Max pooling operation to perform the downsampling operation. The backbone network can learn more features and is more robust.

The neck part is shown in **Figure 4**. The Neck part mainly adds the SPPCSPC module and up sampling layer, and uses the nearest neighbor interpolation method to enlarge the image size to the size of the image after the ELAN module in the backbone network, and the two perform feature fusion. At the same time, it shortened the gradient path to construct new E-ELAN module on the basis of E-ELAN module to obtain more feature information in the second branch. On this basis, it introduced MP module to perform maximum pooling operation again on the feature fusion image, and obtained three different sizes of feature maps.

The head part is shown in **Figure 5**. The feature maps of three different sizes are processed by the REP module with different convolution kernel sizes to realize feature extraction and smooth features. Finally, the position information and category information of the target are regressed by convolution, batch normalization, and sigmoid function. Since the data set has a total of 38 categories, each position predicts 3 anchors, and each anchor contains center point coordinates

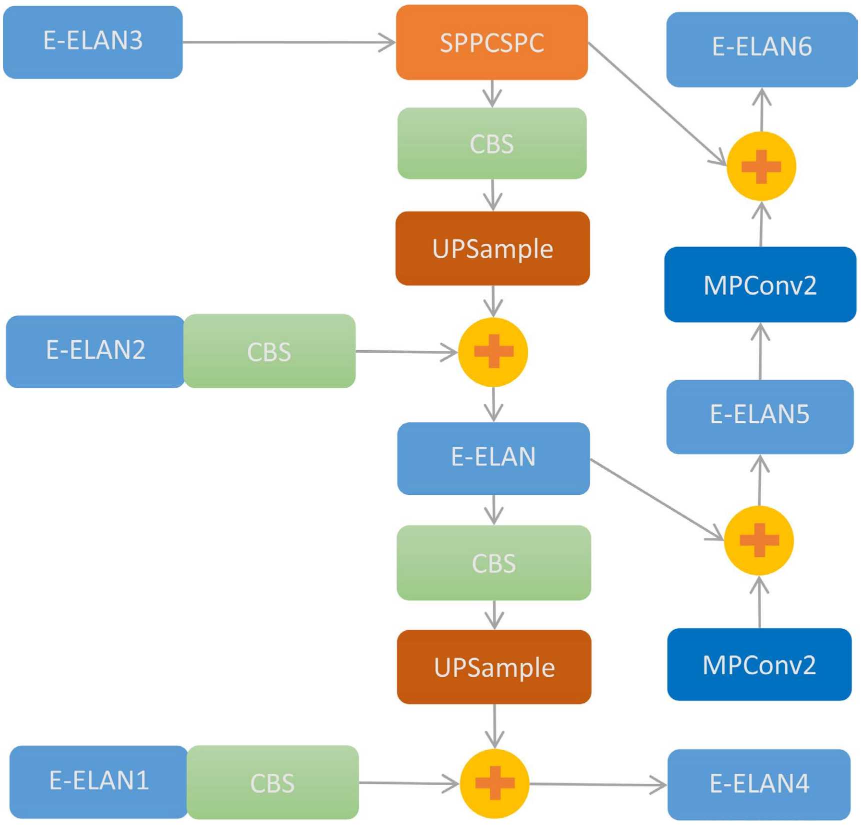

Figure 4. The E-ELAN (1, 2, 3) modules correspond to the modules in the backbone network. SPPCSPC module with CBS as submodule is used to increase the receptive field. The UPSample module is an up sampling operation. MPConv2 differs from MPConv in the number of channels.

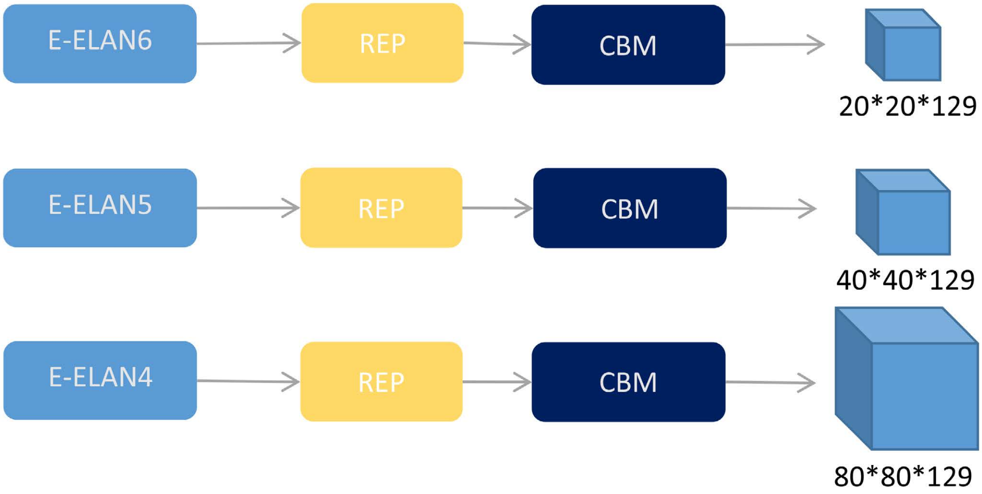

Figure 5. The figure is the head part of the network structure, with E-ELAN (4, 5, 6) corresponding to the modules in the neck. The REP module is composed of convolution layer and batch normalization for feature extraction. The CBM module differs from CBS in that it uses the sigmoid activation function.

(x, y), width and height (w, h), confidence and 38 category information, the number of channels of the output feature map is 3*(5 + 38) = 129.

4. Experiment and Results Analysis

4.1. Experimental Environment

The experimental environment is Window10, memory is 16G.The experimental environment is CPU Intel Sliver 4210, GPU is 2080Ti, pytorch1.10 is used as a training framework for deep learning models, and the CUDA version is 12.0.

4.2. Experimental Data

The data set was collected using ICT6040, a security inspection equipment independently developed by Shanghai Wuying Technology Co., LTD. The dataset contains all categories in SIXray and OPIXray, breaking down firearms, knives, explosives, hand tools, and 38 dangerous goods categories in its data. Because it is not easy to collect dangerous goods data, after the TIP method proposed in this paper is inserted, the number of different types reaches a balance. The dataset consists of 23,780 pictures of real data and 10,220 pictures of synthetic data, with a total of 34,000 pictures. LabelImg annotation tool was used to annotate the data set in xml format. After annotation, it was converted into data label format required by YOLO, and the data set was divided into training set, validation machine and test set. Training verification set and test set, training set and verification set were all in accordance with 9:1. 27,540 images are used as the training set, 3060 images are used as the verification set, and 3600 images are used as the test set.

4.3. Experimental Parameter Setting

Depending on hardware environment, the input image size was uniformly scaled to 640 × 640, the training batch size was set to 8, the training period was set to 300, the initial learning rate was set to 0.001, and the optimizer used Adam.

4.4. Experimental Evaluation Metrics

- Evaluation Metrics

In the experiment, accuracy rate, recall rate, F1-score, mAP0.5, mAP0.95 and FPS were used as evaluation indexes to evaluate the performance of the model. Where TP indicates that it is actually a positive sample and is predicted to be a positive sample; FP indicates that it is actually a positive sample and predicted to be a negative sample; FN indicates that the actual sample is actually negative and is predicted to be positive; FS indicates the inference speed of the model.

Precision represents the proportion of the number of real positive samples in the predicted positive sample, and represents the accuracy rate. The calculation formula is as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

Recall represents the proportion of the predicted positive samples in all the real positive samples, indicating the recall ratio. The calculation formula is as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

The F1-score, also known as the balanced F Score, takes into account both accuracy and recall, and is the harmonic average of accuracy and recall. The calculation formula is as follows:

$$\text{F1} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \tag{8}$$

AP is the area surrounded by the PR curve and coordinate axis of each category under setting the IOU size. mAP indicates the average value of aps in all categories. mAP0.5 indicates that the threshold is IOU = 0.5, and mAP0.95 indicates the average value of aps with a interval of 0.05 from 0.5 to 0.95 for each category. The mAP shows the accuracy of the network model. Its formula is defined as:

$$\text{AP} = \int_0^1 P(r)\,\text{d}r \tag{9}$$

$$\text{mAP} = \frac{1}{n}\sum_{i=1}^{n} \text{AP}_i \tag{10}$$

- Evaluation Metrics Results Analysis

Figure 6 shows the accuracy curve, Figure 7 shows the recall rate curve, and Figure 8 shows the F1-score curve. It can be seen from the figure that when the

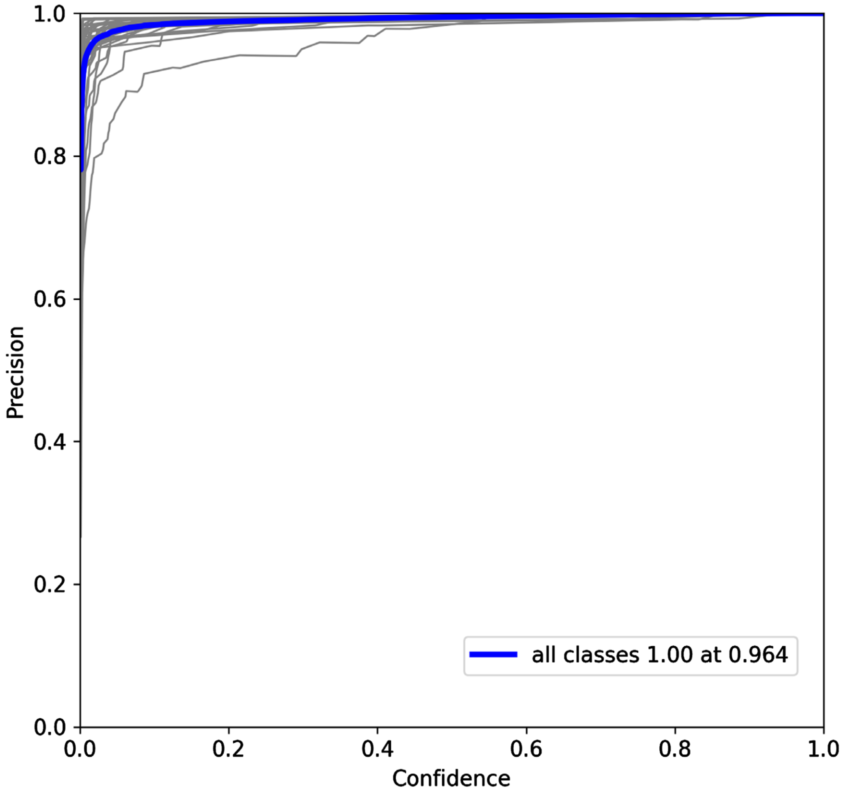

Figure 6. Accuracy curve.

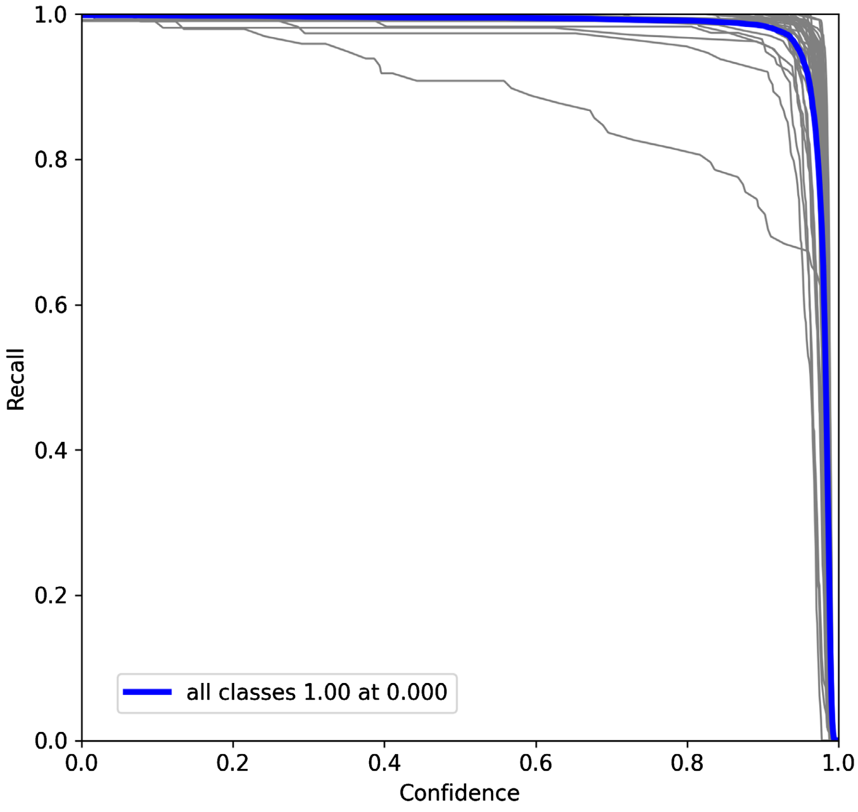

Figure 7. Recall curve.

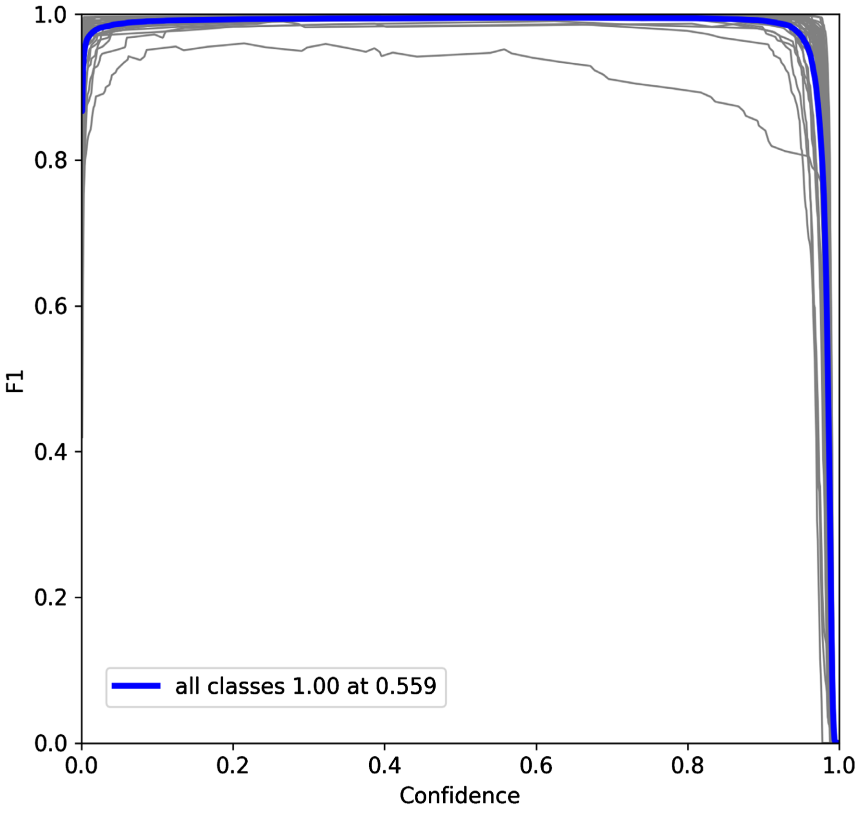

Figure 8. F1-score curve.

confidence is greater than 0.4, the accuracy is close to 1, that is, the greater the confidence is, the greater the probability of predicting the real positive sample to the positive sample in the test set. When the confidence level is less than 0.6, the recall rate is close to 1, that is, the greater the probability of predicting all the real positive samples in the test set. In order to comprehensively measure the accuracy rate and recall rate, the F1-score evaluation index is introduced to reconcile the accuracy rate and recall rate. As shown in Figure 3 above, when the confidence level is 0.4 - 0.6, the recall rate can be adjusted. When F1-score is close to 1, the model performs better.

4.5. Comparison Experiment

Under the condition that the operating environment of the system and the initialization parameters of the model are the same, four groups of comparison tests are carried out. As shown in Table 1, the first group directly used YOLOv7 to train the data set, and both mAP0.5 and mAP0.95 reached more than 90%. The second group of experiments was trained after the TIP method was introduced. It can be seen from the results that when the number of parameters of the model increased, its mAP also increased by 6.6%. In the third group of experiments, after the introduction of high-low energy image fusion method for training, the number of model parameters decreased compared with the TIP method only. When the threshold value was 0.5, the Map value was slightly lower than that of the TIP method. The threshold value was 0.5 to 0.95, and the step size was 0.05 incrementally higher than that of the TIP method only. When the threshold is set higher, the model can perform better. The fourth group of experiments combined the above two methods for training, and the number of parameters of its model decreased significantly, and mAP0.5 and mAP0.95 increased by 0.06 and 0.057 respectively compared with the two methods. Although the number of images per second can be processed has decreased, it has met the requirements of real-time detection system.

4.6. Test Section

Based on the above results, the test set was used to conduct tests based on YOLOv7 + TIP + Map to evaluate the quality of the model. The experiment was carried out using the package data containing bullets, grenades, kitchen knives, vernier calipers, carving knives, screwdrivers, wrenches, and razors, the results

Table 1.Contrast ablation experiment.

Model	Parameters/M	mAP@0.5	mAP@0.95	FPS/S
YOLOv7	114.5	0.933	0.928	102
YOLOv7 + TIP	125.9	0.979	0.966	84
YOLOv7 + Map	107.6	0.968	0.974	99
YOLOv7 + TIP + Map	71.7	0.993	0.985	93

of which are shown in Table 2.

In all dangerous goods identification, the average accuracy of all categories with a threshold of 0.5 is above 0.99, because its texture is clear and the feature information is easy to learn, which should be caused by the simple data collected and the background is not particularly complex. Among them, the recognition rate of hand grenade, kitchen knife and vernier caliper is the best, because the hand grenade has rich structural information, and its shape has high consistency in X-ray images at any Angle and position, so it is easy to identify the target. Compared with other dangerous goods, the structure of kitchen knife and vernier caliper is relatively simple, and the projected texture information is easier to learn. The reason for the analysis is that under different angles, the imaging morphology changes greatly. For example, the imaging of the knife face and the knife back are completely inconsistent, and the shape of the knife in the X-ray image is very different. For the model, it is difficult to identify the target, so the recognition accuracy is not high, but other categories are above 99%. Its detection accuracy has met the demand.

4.7. Prediction Section

Inference is performed using the best parameter profile trained by the model. As shown in Figure 9, even if the grenade overlaps with other items, it can still be accurately identified due to its own unique structural information. For inorganic objects, such as knives, kitchen knives, and axes, the pixel value after imaging in the X-ray image is much smaller than that of other substances, and after the fusion of high-low energy images, the outline of such objects is more obvious. This helps the network model to learn the characteristic information of such objects and to recognize them.

5. Conclusion

In this paper, TIP [7] and high- and low-energy image fusion methods are introduced to the collected X-ray dangerous goods data set to expand the data set,

Table 2. Test experimental results.

Dangerous goods	Precision	Recall	mAP@0.5	mAP@0.95
bullet	0.986	0.979	0.997	0.969
Hand grenade	0.997	1.0	1.0	0.992
Kitchen knife	0.996	0.991	0.996	0.995
Vernier calipers	0.996	1.0	0.999	0.995
Carving knife	0.965	0.973	0.995	0.928
Screwdriver	0.997	0.992	0.996	0.96
Wrench	0.99	0.995	0.995	0.986
Razor blade	0.986	0.983	0.995	0.961

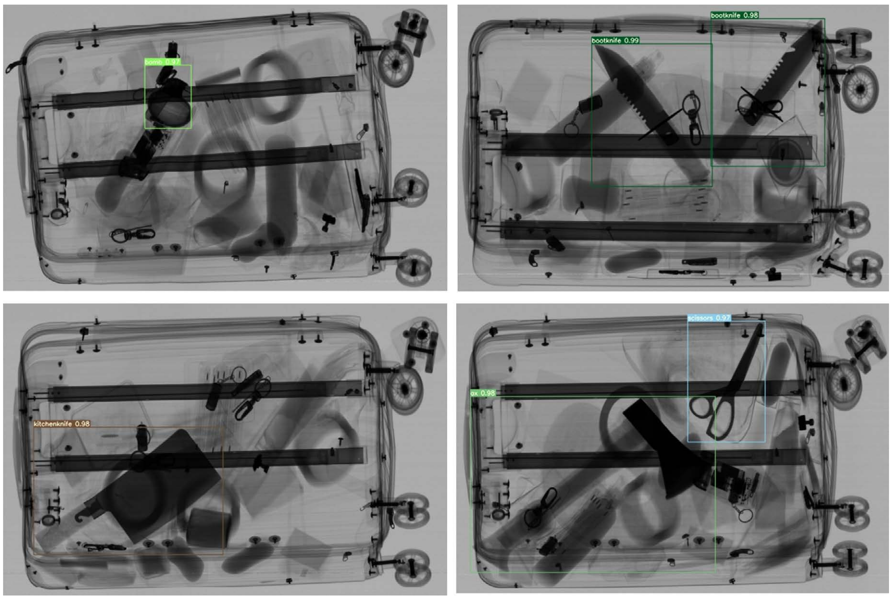

Figure 9. The predicted results of dangerous goods.

and feature enhancement is carried out. The YOLOv7 model, which is superior to the previous target detector in both accuracy and speed, is used for training, and its training accuracy rate is more than 99%. Compared with direct training without TIP and high- and low-energy fusion, the accuracy is improved by 6%, and the FPS is 93. The experimental results show that the method of using TIP and high-low energy fusion and then training based on YOLOv7 can meet the real-time detection requirements of X-ray security equipment, and the detection accuracy is much higher than the industry detection accuracy standards. It can be integrated into X-ray security equipment to assist security staff to carry out work, improve work efficiency and reduce security risks. Since the categories of dangerous goods in real life are far more than those collected in this experiment, the model may ignore some unknown dangerous goods in practical application. In the next step, more dangerous goods categories will be collected for training, and the model will be trained with open data sets to improve the robustness of the model and enable the model to identify unlabeled dangerous goods. In addition, although the detection speed of the model has met the requirements of the real-time detection system, there is still a lot of room for improvement. The next step will be to comprehensively improve the performance of the model by optimizing the structure of the network model, and reducing the number of parameters and calculation amount of the model under the premise of ensuring accuracy.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Girshick, R., Donahue, J., Darrell, T., *et al.* (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *IEEE Conference on Computer*

Vision and Pattern Recognition, Columbus, 23-28 June 2014, 580-587. https://doi.org/10.1109/CVPR.2014.81

[2] Girshick, R. (2015) Fast R-CNN. *Proceedings of* 2015 *IEEE International Conference on Computer Vision* (*ICCV*), Santiago, 7-13 December 2015, 1440-1448. https://doi.org/10.1109/ICCV.2015.169

[3] Ren, S., He, K., Girshick, R., *et al.* (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

[4] Liu, W., Anguelov, D., Erhan, D., *et al.* (2016) SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M., Eds., *European Conference on Computer Vision*, Vol. 9905, Springer, Cham, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[5] Redmon, J., Divvala, S., Girshick, R., *et al.* (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/CVPR.2016.91

[6] Wang, C.Y., Bochkovskiy, A. and Liao, H.Y.M. (2022) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. ArXiv: 2207.02696.

[7] Bhowmik, N., Wang, Q., Gaus, Y.F.A., *et al.* (2019) The Good, the Bad and the Ugly: Evaluating Convolutional Neural Networks for Prohibited Item Detection Using Real and Synthetically Composited X-Ray Imagery. ArXiv: 1909.11508.

[8] Kundegorski, M.E., *et al.* (2016) On Using Feature Descriptors as Visual Words for Object Detection within X-Ray Baggage Security Screening. *International Conference on Imaging for Crime Detection and Prevention*, Madrid, 23-25 November 2016, 1-6.

[9] Wang, H.J. and Hui, J. (2018) Based on SIFT Features and Hazardous ISM of X-Ray Image Detection Method. *Computer Measurement and Control*, **26**, 31-33.

[10] Lowe, D.G. (2014) Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, **60**, 91-110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[11] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) Imagenet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, Lake Tahoe, 3-6 December 2012, 1097-1105.

[12] Akcay, S., *et al.* (2016) Transfer Learning Using Convolutional Neural Networks for Object Classification within X-Ray Baggage Security Imagery. 2016 *IEEE International Conference on Image Processing*, Phoenix, 25-28 September 2016, 1057-1061. https://doi.org/10.1109/ICIP.2016.7532519

[13] Akcay, S. and Breckon, T. (2017) An Evaluation of Region Based Object Detection Strategies within X-Ray Baggage Security Imagery. 2017 *IEEE International Conference on Image Processing*, Beijing, 17-20 September 2017, 1337-1341. https://doi.org/10.1109/ICIP.2017.8296499

[14] Oquab, M., Bottou, L., Laptev, I., *et al.* (2014) Learning and Transferring Mid-Level Image Representations Using Convolutional Neural Networks. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 1717-1724. https://doi.org/10.1109/CVPR.2014.222

[15] Lu, G.-Y. and Gu, Z.-H. (2021) Improved YOLOv3 Detection Algorithm for Dangerous Goods in Security Inspection Packages. *Computer Application and Software*,

38, 197-204.

[16] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement. ArXiv: 1804.02767.

[17] Wu, H.B., Wei, X.Y., Liu, M.H., Wang, A.L., Liu, H. and Yuzhi, I. (2021) X-Ray Security Dangerous Goods Detection Based on Dilated Convolution and Transfer Learning Improved YOLOv4. *Chinese Optics*, **14**, 1417-1425.

[18] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M. (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection. ArXiv: 2004.10934.

[19] Liu, J., Leng, X. and Liu, Y. (2019) Deep Convolutional Neural Network Based Object Detector for X-Ray Baggage Security Imagery. 2019 *IEEE* 31*st International Conference on Tools with Artificial Intelligence* (*ICTAI*), Portland, 4-6 November 2019, 1757-1761. https://doi.org/10.1109/ICTAI.2019.00262

[20] Zhang, Z., Li, M.Z., Li, H.F. and Ma, J.Q. (2021) Improve SSD Algorithm and Its Application in Subway Security. *Computer Engineering*, **47**, 314-320.

[21] Han, N. (2018) Research on Detection and Tracking Algorithm of X-Ray Image Dangerous Goods Based on Deep Learning. Lanzhou University, Lanzhou.

[22] Que, L.S., Wang, M.Q., Zhang, J.S., *et al*. (2019) Edge Keep Filtering Based on Multi-Scale Local X-Ray Image Tone Mapping Algorithm. *Science*, *Technology and Engineering*, **19**, 217-221.