SHOMY: Detection of Small Hazardous Objects using the You Only Look Once Algorithm

Eunchan Kim[1†], Jinyoung Lee[2†], Hyunjik Jo[2†], Kwangtek Na[3], Eunsook Moon[2], Gahgene Gweon[1], Byungjoon Yoo[1], and Yeunwoong Kyung[4*]

[1] Department of Intelligence and Information, Seoul National University
Seoul 08826, Republic of Korea
[e-mail: eunchan@snu.ac.kr]
[2] Department of Artificial Intelligence, Yonsei University
Seoul 03722, Republic of Korea
[3] Department of Electrical and Computer Engineering, Inha University
Incheon 22212, Republic of Korea
[4] School of Computer Engineering, Hanshin University
Osan 18101, Republic of Korea
[e-mail: ywkyung@hs.ac.kr]
*Corresponding author: Yeunwoong Kyung
†These authors contributed equally to this work

Received January 24, 2022; revised May 10, 2022; accepted June 4, 2022; published August 31, 2022

Abstract

Research on the advanced detection of harmful objects in airport cargo for passenger safety against terrorism has increased recently. However, because associated studies are primarily focused on the detection of relatively large objects, research on the detection of small objects is lacking, and the detection performance for small objects has remained considerably low. Here, we verified the limitations of existing research on object detection and developed a new model called the Small Hazardous Object detection enhanced and reconstructed Model based on the You Only Look Once version 5 (YOLOv5) algorithm to overcome these limitations. We also examined the performance of the proposed model through different experiments based on YOLOv5, a recently launched object detection model. The detection performance of our model was found to be enhanced by 0.3 in terms of the mean average precision (mAP) index and 1.1 in terms of mAP (.5:.95) with respect to the YOLOv5 model. The proposed model is especially useful for the detection of small objects of different types in overlapping environments where objects of different sizes are densely packed. The contributions of the study are reconstructed layers for the Small Hazardous Object detection enhanced and reconstructed Model based on YOLOv5 and the non-requirement of data preprocessing for immediate industrial application without any performance degradation.

Keywords: Computer vision, detection of hazardous items, small-object detection, YOLO, air transport, security industries.

A preliminary version of this paper was presented at ICONI 2021, and was selected as an outstanding paper.
This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (No.2020R1G1A1100493).

http://doi.org/10.3837/tiis.2022.08.012 **ISSN : 1976-7277**

1. Introduction

Over several decades, efforts have been underway to improve the security levels of passenger and cargo transport [1, 2]. With the increasing risk of air transport-related terrorism, especially after the September 11 attacks in 2001 [3], the detection of prohibited objects, including explosives, via X-ray scans of personal baggage has become even more important [2, 4, 5]. Formerly, object detection via X-ray imaging depended heavily on human vision, and therefore had high fallibility and limitations in discrimination [6, 7]. Nowadays, cabin baggage inspection systems have gradually advanced using computer vision technology. Furthermore, with the development of artificial intelligence (AI), a sophisticated baggage inspection system has been introduced to improve the detection performed by security personnel. With this new technology, the accuracy and speed of baggage inspection have been gradually enhanced [8, 9].

However, despite recent advancements, some published studies on cabin baggage inspection have revealed different problems in automated object detection via X-ray scans. This technology is either limited to supporting security personnel in the field [6–8] or is slow as it performs two or more object detection processes [10, 11]. Furthermore, the technology often fails to detect relatively small hazardous objects because it has been devised with a focus on recognizing large objects linked to explosions and killings [12, 13]. Therefore, we to focused on resolving this shortcoming to recognize and detect small hazardous objects.

Herein, we first verify the limitations of existing object detection and detection models through a literature review of baggage inspection in Section 2. Although it is common to oversample or manipulate target labels to enhance performance [14–16], the rather low performance of these methods when implemented on actual systems is continuously highlighted. In Section 3, we present our model development process that does not require any data transformation, and our new model that employs this process—called the Small Hazardous Object detection enhanced and reconstructed Model based on YOLOv5 (SHOMY). In Section 4, we compare SHOMY and the existing YOLOv5 model, and in Section 5 state the implications of the study and discuss future research directions. The main contributions of this study are 1) reconstructed layers for SHOMY based on YOLOv5, 2) elimination of data preprocessing for immediate industrial application without performance degradation, and 3) small-object detection in overlapping environments regardless of the field.

2. Literature Review

2.1 Development of Object Detection Technology

Object detection is the process of showing an object's location for verification within an image, and automatically categorizing its type [8, 9]. Object detection algorithms can be categorized into two types: two-stage detector algorithms, which perform region proposal and object classification separately, and one-stage detector algorithms, which perform these two processes concurrently [10]. The former can be transformed into the latter with greatly enhanced speeds and lower calculation costs. The most representative algorithms of the two-stage detector type, listed in order of increasing speed, are the region-based convolutional neural network (R-CNN), fast R-CNN, and faster R-CNN. In detail, two-stage detector algorithms propose a region of interest where the object for detection might be located, extract the object features, and perform learning for the marking and categorization of the bounding box of the object. Contrastingly, one-stage detectors can be categorized into You Only Look

Once (YOLO) and single-shot multibox detector (SSD) algorithms. As can be inferred from its name, a one-stage detector performs proposal and categorization of the bounding box simultaneously, which saves time and reduces the cost in calculation and inference.

For two-stage detectors, the calculation speed must be enhanced because calculation and inference are performed in two stages at the beginning of object detection. However, although the development of one-stage object detection has resulted in enhanced speed, its relatively low performance is a downside. Therefore, researchers have tried to improve its performance. To detect small objects using the YOLOv2 algorithm, the early altered model of YOLO, the receptive field was expanded to apply an altered characteristic extraction model (backbone) concatenating a convolution layer, which can include more region information, and a general convolution layer to achieve enhanced performance compared to previous models. Case studies of one-stage detectors based on the YOLO model have reported that they have object detection performance on par with those of two-stage detectors [11]. However, the ability of one-stage detectors to accurately detect some objects, particularly small objects, remains limited [14–16].

We considered the characteristics of real-time detection of target objects through aeronautical X-ray scans and inferred that slow two-stage detectors are unfit for baggage search and detection. Model development and research on YOLO [12, 13], the fastest network among one-stage detectors, were then performed. We modified YOLOv5, the latest YOLO model, and developed SHOMY, a new model with enhanced performance.

2.1.1 Research on Small-Object Detection

YOLO is relatively weak at detecting small objects owing to its high detection speed [12], and therefore studies are continuously being conducted in the aeronautical field to enhance the detection of small objects that are difficult to identify with the naked eye. Liu et al. [13] developed a model that detects automobiles and people from a video of the ground filmed by an unmanned airplane. A generative adversarial network has been utilized for data augmentation to develop a network that produces a high-resolution video from a low-resolution satellite-filmed video [14]. Object detection and model learning were performed simultaneously, leading to enhanced detection of small objects. However, despite these efforts, prior research on the detection of small objects failed to maintain the performance or suggest enhancements for datasets that include objects of different sizes.

Owing to the difficulties in analyzing large datasets and in securing proper data labeling, even prior research on X-ray object detection failed to result in uniform performances for the detection of small, medium, and large objects. Therefore, studies on the topic are scarce. Lee and Cho [17] used the airport baggage X-ray data disclosed and provided by the Korea National Information Society Agency, which is the same dataset used in this study. They reported the detection of prohibited objects in baggage based on baggage images taken by X-ray scanners. The Xception algorithm, a lightening model in which input data are received to reduce the number of channels through a 1×1 convolution product and are made to go through a 3×3 convolution product for individual output channel, was used. A detection and categorization model for 12 prohibited objects was then developed, which demonstrated a high performance based on the F-1 score. An experiment to detect and categorize large hazardous items and single items from a single image was then conducted. The model exhibited a generally good performance in the application of the latest exponentially developed image detection algorithm. However, only a fraction of the numerous types of items that must be detected in the field were used, and thus there were limitations in applying the model to an actual environment in which objects of two or more classes must be detected. However, a

dataset of X-ray images of six prohibited items was used, and performance comparison tests were conducted against models such as YOLOv2, R-CNN, and region-based fully CNN (R-FCN) to determine the most outstanding model [18]. This study focused on large objects such as laptops and cameras and, thus, it failed to evaluate the detection performance of small objects and related situations. Similarly, the use of YOLOv3 solely in the detection of large, harmful, prohibited objects, such as razor blades, knives, and guns, among others was researched [19].

To overcome the limitations of prior research, we conducted object detection wherein a total of 38 objects of different sizes applicable to an actual baggage search and videos were included in the detection efforts. Furthermore, the application of YOLOv5, the latest altered model of YOLO, as a one-stage detector was proposed. Although it is common to manipulate data to enhance the detection performance on small objects, as in [15, 16], data preprocessing would require considerable additional calculation costs and may have an adverse effect on learning. Therefore, we chose to approach the research question only with model tuning and without learning data alteration or manipulation, such as increasing the data and resolution of the input data, oversampling, copying and pasting small objects, and tiling. This ensures that the object detection performance is maintained in actual systems.

3. Methodology

This section examines the overall structure and outline of YOLOv5 and introduces the newly developed SHOMY model for enhanced small-object detection. Despite its high speed, the ability of the YOLO series to detect small objects is rather weak [12]. The architecture of SHOMY, which adds a neck layer to further expand the feature map and apply a new methodology to the detection layer, is described. The overall descriptions of baggage image data and sizes of objects, used for result comparison, are also defined.

3.1 YOLOv5 Model

YOLO realizes a one-stage detector model by defining object detection problems as regression problems. Available YOLO versions range from v1 to v5; among these, YOLOv5 was used in this study. Its object detection structure comprises a backbone layer for extracting traits, neck layer for aiding the detection of objects in different scales, and head layer for detecting objects.

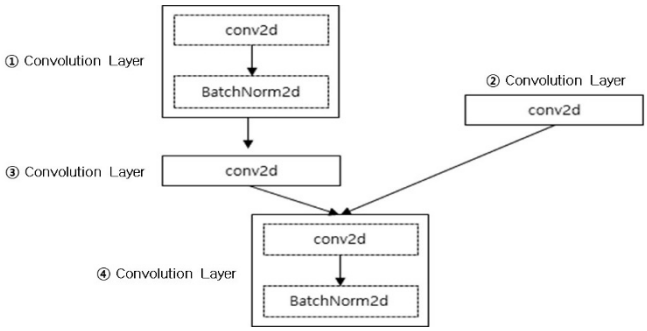

Fig. 1. Structure of cross-stage partial network (CSPNet).

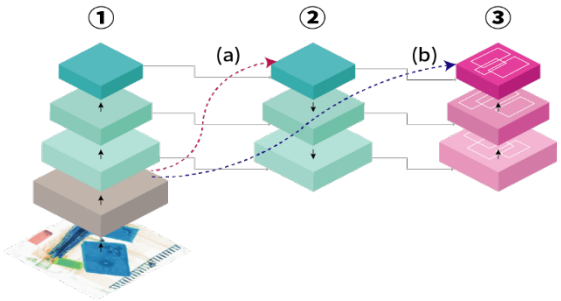

Fig. 2. Architecture of path aggregation network (PANet).

The first stage, referred to as a backbone layer–cross-state partial network, is for extracting the traits of images. In this stage, a cross-stage partial network (CSPNet), which is a model with an alleviated amount of calculation for use in a low-function computer environment or for real-time image detection, is used. Whereas the general CNN model requires large numbers of calculations with duplicate gradient problems, CSPNet integrates the feature map at the beginning and at the end to innovatively reduce the number of calculations.

The CSPNet utilized by YOLOv5 is structured as shown in **Fig. 1**. The output value in the early base layer is divided into convolution ① and convolution ②. Subsequently, the value gained through convolution layers ① and ③, and that gained through convolution layer ② are merged. The value also passes through convolution layer ④, and consequently, the output value of the base layer is connected directly to the final convolution layer, serving as the gradient shortcut.

The second stage, referred to as a neck layer–path aggregation network, responds to different scales of objects. Through modeling that utilizes a path aggregation network (PANet) [21] as its backbone, the feature pyramid network (FPN) resolves problems with information from the first layer not being reflected properly in the final prediction.

In **Fig. 2**, layer ① is the FPN backbone and comprises considerably large networks such as ResNet-50. Therefore, low-level feature information must go through numerous layers to be conveyed to (a) a high level, wherein loss of information is inevitable. PANet is designed to fully convey the information even if the information is passed through the shortcut of (b) and through multiple convolution layers, with the addition of layer ③.

The last stage, or the head layer, is used for predicting the possible locations of objects and utilizes the complete intersection over union (CIoU) loss proposed in [22]. CIoU has a high learning speed compared to those of IoU, generalized IoU (GIoU), and distance IoU (DIoU), which are often used for object detection loss rates and are useful in detecting small objects.

$$\text{CIoU Loss} = S(B, B^{gt}) + D(B, B^{gt}) + V(B, B^{gt}) \tag{1}$$

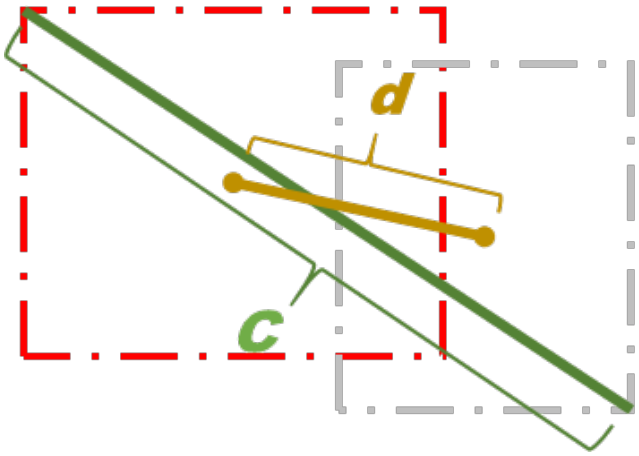

Fig. 3. Definition of *D* in CIoU.

As shown in (1), CIoU comprises the surface area (*S*), distance (*D*), and aspect ratio (*V*) of two boxes (*B*: box coordinates predicted by the model, B^{gt}: coordinates of the actual box (ground truth)).

$$S = 1 - IoU \tag{2}$$

S, defined in (2), is the loss of the overlapping area of the two boxes.

$$D = \frac{\rho^2(p,p^{gt})}{c^2} \tag{3}$$

The distance *D* of the box is calculated based on the diagonal distance (*c*) and the distance of the central points (*d*), as shown in **Fig. 3**. In (3), *c* is the diagonal distance, and ρ is the distance between central points. Therefore, if the two objects are close, *D* approaches zero, and the loss decreases.

$$V = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2 \tag{4}$$

Finally, *V* calculates the difference in the proportions of the widths (2) and heights (*h*) of the two boxes, as in (4), to determine whether the two boxes have similar forms. Ultimately, CIoU induces learning by maximizing the overlapping area between the two boxes, minimizing the distance between them, and maintaining a similar form simultaneously.

$$\begin{aligned}
&\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}[(x_i-\hat{x}_i)^2 + (y_i-\hat{y}_i)^2] \\
&+\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}[(\sqrt{w_i}-\sqrt{\hat{w}_i})^2 + (\sqrt{h_i}-\sqrt{\hat{h}_i})^2] \\
&+\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}(C_i-\hat{C}_i)^2 \\
&+\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{noobj}(C_i-\hat{C}_i)^2 \\
&+\sum_{i=0}^{S^2}1_{i}^{obj}\sum_{c\in classes}(p_i(c)-\hat{p}_i(c))^2
\end{aligned} \tag{5}$$

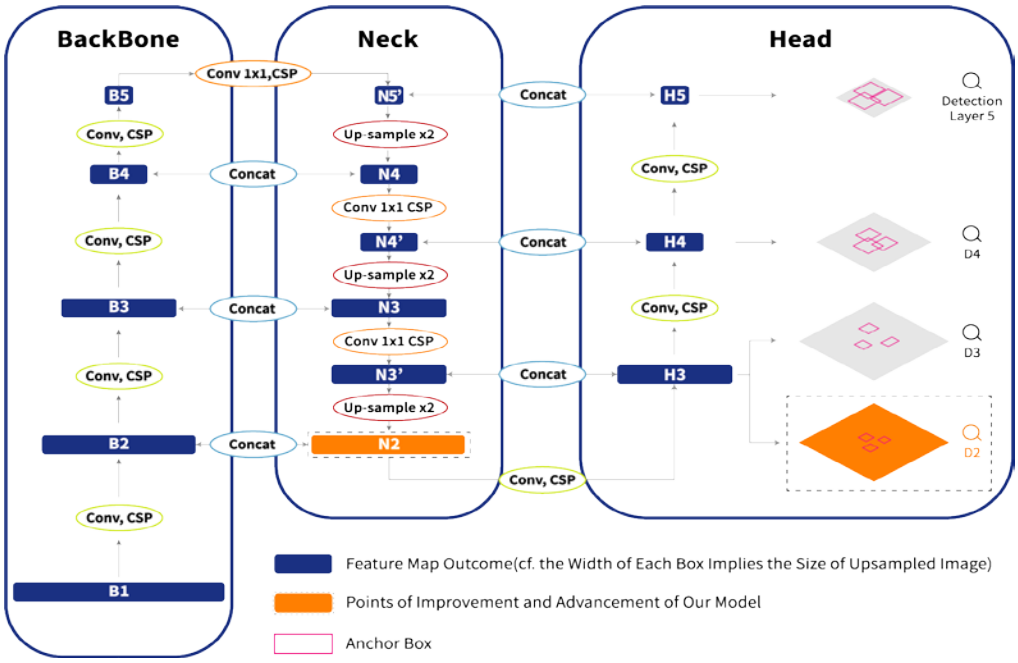

Fig. 4. Architecture of SHOMY.

YOLO defines loss by separating cases wherein there are objects on the grid cell and wherein there are no objects on the grid cell. If (5) were to be labeled (a)–(e) from the top, (a) would be the mean squared error (MSE) for the central point coordinates (x,y) of the object when there is an object in the grid cell. Part (b) would be the MSE for the height (h) and width (w) of the object when there is an object in the grid cell. Here, root is used to decrease the scale difference between large and small objects. Parts (c) and (d) would be the confidence scores for when there is an object in the grid cell and for when there is no object in the grid cell, respectively. This is defined as $P(object) \times CIoU$. Finally, (e) would be the loss value of the conditional probability for the class when there is an object in the grid cell.

3.2 SHOMY: Performance Enhancement Model for Small-Object Detection

As reported in prior studies, X-ray baggage detection differs between large and small objects. In addition, whereas the detection performance of large objects is now adequate to a certain degree, the detection rates for small objects, such as USB flash drives, bullets, and lighters, are relatively low. Scaled-YOLOv4 [23] adds detection layers to improve the probability of acquiring enhanced results compared with those of the basic model, for increasing the detection rates for such small objects. The backbone network of YOLOv5 finds the location and spatial information of objects, whereas the neck network finds the semantic information. Based on the characteristics of Scaled-YOLOv4 and YOLOv5, this study added a neck layer, N2, which produces a massive feature map, as shown in **Fig. 4**. This further expands the last feature map (upsample) and secures as much semantic information as possible.

The added N2 is connected to B2 of the same scale in the backbone to help the network find the traits and spatial information of small objects. Meanwhile, the head is the final stage of object detection. As discussed, a stage dedicated to detecting small objects was added [23].

As shown in **Fig. 4**, the detection layer D2, which utilizes an anchor box for small objects, was added to H3 of the head. The sizes of the boxes (5 × 7, 8 × 15, 17 × 12) are half of those

of three anchor boxes combined (10 × 13, 16 × 30, 33 × 23). As a result, regression learning is performed with a smaller anchor box in a space expanded even further to enable the detection of small objects such as USB flash drives (10 × 17).

By adding N2 to the neck, a large feature map is produced, whereas by adding D2 to the produced feature map and utilizing a smaller anchor box, the detection of small objects is improved. The results of the proposed model are provided in Section 4.2.

3.3 Computing Resources

The experiments were conducted using an Nvidia RTX 3080 GPU, Intel i7-10700 2 CPU, and 64 GB of RAM.

3.4 Definition of Dataset and Object Size

As mentioned, we utilized the image data of prohibited objects obtained by a scanner from Rapiscan and provided by the National Information Society Agency, South Korea. To build a system that detects target objects using X-ray for security searches at airports, ports, train stations, private companies, and public offices, a dataset having a considerable amount of data is necessary.

The dataset provided 38 discernable target objects. Based on Microsoft Common Objects in Context (MS COCO), small, medium, and large objects were defined according to their surface areas, as outlined in **Table 1**. They were categorized into 5 small, 26 medium, and 7 large objects. Among the small objects, USB flash drives are sensitive storage media linked closely to the leakage of confidential information, whereas bullets, nail clippers, batteries, and lighters are objects that must be detected for flight safety [24–26].

Table 1. Definition of object size (small, medium, large) and classification of target objects based on MS COCO standard.

Size	MS COCO Definition (Pixel)		Classification of X-ray Target Objects (38 EA)
	Min.	Max.	
small	1 × 1	32 × 32	USB, bullet, nail clippers, battery, lighter (5 EA)
medium	32 × 32	96 × 96	throwing knife, match, electronic cigarettes, electronic cigarettes (liquid), awl, thinner, SSD, screwdriver, Zippo oil, liquid, aerosol, knife, portable gas, supplementary battery, smart phone, HDD, alcohol, scissors, spanner, handcuffs, gun parts, solid fuel, pliers, chisel, gun, firecracker (26 EA)
large	96 × 96	∞ × ∞	hammer, tablet PC, laptop, saw, axe, bat, metal pipe (7 EA)

Four types of data are provided according to their provision methods, as shown in **Table 2**. Because multiple target items for detection and general items are included together in actual images in the field, data types of Multiple & Categories and Multiple & Others were selected here. The former is a type of dataset that includes multiple target items for detection, other hazardous objects, and other general objects. The total number of distinct items for detection was 38, with the dataset having a total of 54,949 data points. Of these, 41,211 (75 %) were used for training, whereas the remaining 13,738 (25 %) were used for validation. Data were randomly extracted and produced from each item at an equal rate of 25 %.

Table 2. Types of X-ray baggage datasets.

Type	Details
Single Default	1 target object only
Single & Others	1 target object + other non-targeted objects
Multiple & Categories	multiple target objects + other target objects
Multiple & Others	multiple target objects + other non-targeted objects

3.5 Measurement Indexes

The basic indexes for measuring model performance in object detection are precision and recall. The former refers to the proportion of actual true values among the experimental values identified to be true, whereas the latter refers to the proportion of values identified to be true among the actual true values. They are defined in (6):

$$\begin{aligned} \text{Precision} &= \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \\ \text{Recall} &= \frac{\text{True Positive}}{\text{True Positive} + \text{False } Negative} \end{aligned} \tag{6}$$

Precision and recall are complementary. When recall is enhanced in object detection, precision is bound to decrease, and it is important to allow the model to learn such that the values of the two indexes are high. Therefore, the mean average precision (mAP) index, an average index for detection that accounts for both precision and recall, is generally used. Generally, mAP is measured based on IoU (intersection over union) = 0.5, which refers to the overlapping ratio of the predicted area of the box in which an object is placed to the actual area of the box (ground-truth bounding box), as shown in **Fig. 5**. In other words, the IoU is the area of overlap divided by the area of union. When the IoU is 0.5 or higher, mAP categorizes the predicted value as true [27].

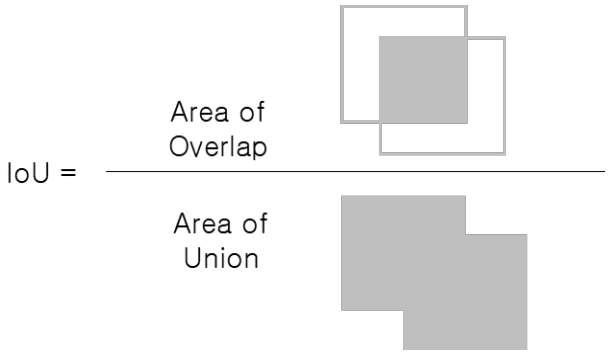

Fig. 5. Calculation of IoU.

Another index measured with mAP is mAP (.5:.95), which is the average value measured based on increasing 0.5 to 0.95 by intervals of 0.05 and is used as a more rigorous object detection performance index compared to mAP [27]. This study focuses on how detection performance on small objects is enhanced in terms of the two indexes.

4. Experimental Results

This section presents the experimental results of enhancing the performance of detecting small objects from X-ray baggage videos. In particular, the experimental results of YOLOv5 and SHOMY are compared and analyzed with the definition of object size and classification of target objects described in **Table 1**.

4.1 Application Results of YOLOv5 Basic Model

Examining the results of the YOLOv5 basic model revealed the average overall performance to be mAP = 98.7 and mAP (.5:.95) = 87.8, as shown in **Table 3**. The detection performance on small objects was verified to be low compared with those on medium and large objects. In particular, the drop in the small-object detection performance is notably large in terms of mAP (.5:.95) with rigorous IoU settings. On the contrary, unlike in the YOLOv5 basic model, which performs learning based on an anchor box with a pre-designated size, an auto-anchor box uses the K-means algorithm to automatically produce an optimal anchor box in the box coordinate distribution of the dataset class. However, although learning to automatically produce 3–5 auto-anchor boxes was performed, the performance was similar to or slightly lower than that of the basic model. Although analysis for the underlying causes may be necessary in the future, it is assumed that there was a failure to produce optimal groupings from the 38-object box coordinates.

Table 3. Experimental results of YOLOv5 default and SHOMY with auto-anchor settings.

	mAP%				mAP(.5:.95)%			
	All	**S**	M	L	**All**	**S**	M	L
YOLOv5 Default	**98.7**	**96.4**	99.2	98.8	**87.8**	**76.8**	88.6	92.6
/w Auto Anchor 3	**98.7**	**96.2**	99.2	98.8	**87.7**	**76.7**	88.7	91.5
/w Auto Anchor 4	**98.6**	**96.0**	99.2	98.8	**87.4**	**76.0**	88.3	92.2
/w Auto Anchor 5	**98.6**	**95.8**	99.2	98.8	**87.4**	**75.6**	88.5	91.8
SHOMY (N2+D2)	**99.0**	**98.3**	99.2	98.7	**88.9**	**81.6**	89.5	92.1
/w Auto Anchor 3	**98.7**	**96.7**	99.1	98.8	**88.1**	**77.8**	89.3	90.9
/w Auto Anchor 4	**98.7**	**96.7**	99.1	98.9	**88.3**	**77.6**	89.3	92.3
/w Auto Anchor 5	**98.7**	**96.8**	99.2	98.8	**88.5**	**78.0**	86.6	92.7

(S: Small, M: Medium, L: Large; : Yolov5, : SHOMY).

For each of the small items, the YOLOv5 basic model demonstrated lower performance for smaller sizes, as shown in **Table 4**. Section 4.2 discusses the ways in which the SHOMY model enhances object detection.

Table 4. Results of small-object detection for YOLOv5 default vs. SHOMY.

	USB drive (10 × 17)	Bullet (15 × 25)	Nail Clippers (18 × 34)	Battery (20 × 40)	Lighter (21 × 43)
YOLOv5 mAP%	92.0	96.3	99.5	95.6	98.5
YOLOv5 mAP(.5:.95)%	64.6	74.0	85.1	75.1	85.1
SHOMY mAP% (vs. YOLOv5 Default)	96.7 (▲4.7)	97.4 (▲1.1)	99.7 (▲0.1)	98.7 (▲3.1)	99.1 (▲0.6)
SHOMY mAP(.5:.95)% (vs. YOLOv5 Default)	72.7 (▲8.1)	79.3 (▲5.3)	87.1 (▲2.0)	81.3 (▲6.2)	87.5 (▲2.4)

(: Yolov5, : SHOMY).

Table 5. Experimental results of the SHOMY model.

	mAP%				mAP(.5:.95)%			
	All	**S**	M	L	**All**	**S**	M	L
SHOMY (N2+D2)	**99.0**	**98.3**	99.2	98.7	**88.9**	**81.6**	89.5	92.1
SHOMY (N2)	**98.8**	**96.9**	99.2	98.8	**88.1**	**77.3**	88.9	92.8
YOLOv5 Default	**98.7**	**96.4**	99.2	98.8	**87.8**	**76.8**	88.6	92.6

(S: Small, M: Medium, L: Large).

4.2 Results of Small-Object Detection Enhancement

Examining the performance of SHOMY revealed the average overall performance to be mAP = 99.9 and mAP (.5:.95) = 88.9 (**Table 5**), which is an enhancement of 0.3 and 1.1, respectively, compared with those of YOLOv5. In particular, enhancements of 1.9 and 4.8 were observed for small objects. Simultaneously, the detection performance of medium-sized and large objects was maintained without significant reductions or, in some cases, even increased slightly. Even models without the detection layer D2 exhibited slight increases in performance compared with that of YOLOv5 but remained weak compared to SHOMY. This result verified that the addition of detection layer D2, dedicated to the detection of small objects, contributed greatly to enhancing the overall performance. Additionally, in learning, there were weight losses for each of the detection layers D2, D3, D4, and D5, as shown in **Fig. 4**. A weight of 4 was allotted to D2, whereas those of 1, 0.3, and 0.1 were allotted to layers D3, D4, and D5, respectively, which are dedicated to detecting medium-sized and large objects. The purpose of attributing large losses to errors in the detection of small objects is to ensure that the models have strong learning abilities in detecting small objects.

The USB drive, the smallest object analyzed in this study, had a size of 10 × 17. When the basic anchor box value was used, the size of the box was set higher than that of the object, eventually leading to a decreased learning ability. Anchor boxes optimized to the object sizes of the used data were therefore identified to be necessary. As a result, we used new anchor

boxes with dimensions of 5 × 7, 8 × 15, and 17 × 12, or half the size of the smallest anchor box of the YOLOv5 basic model.

Table 3 shows that the performance of the newly developed SHOMY(N2+D2) model with the new anchor setting value was superior to that of the SHOMY model with the auto-anchor boxes. The decline in detection performance on small objects was especially significant when the auto-anchor settings were used. This result shows that in learning a dataset of small objects, manually setting anchor boxes according to the characteristics of these objects would be more effective. As shown in **Table 4**, SHOMY exhibited an enhanced performance for small-object detection compared with that of the YOLOv5 basic model. Notably, the mAP and mAP(.5:.95) indexes for the USB drive, the smallest object analyzed in this study, increased by 4.7 and 8.1 to 96.7 and 72.7, respectively. **Fig. 6** is an example of an inference by SHOMY showing its high performance in detecting objects of different sizes in an X-ray baggage image featuring multiple different objects. In this study, SHOMY was demonstrated to be capable of detecting small objects such as USB drives, lighters, and bullets.

Fig. 6. Example of inference images by SHOMY model.

5. Conclusion

The object detection field is currently facing important challenges because existing studies are primarily focused exclusively on the detection of medium- and large-sized objects, resulting in low detection performance for small objects. However, the potential hazards of certain small baggage items cannot be ignored, as highlighted by incidents such as the relatively recent smartphone explosions and continuous industrial espionage cases using ultrasmall storage media, which have been reported in prior research. We therefore aimed to enhance the detection performance on small objects while maintaining the current detection performance levels of medium-sized and large objects.

To this end, we developed the SHOMY model based on YOLOv5 for enhancing the detection performance on small objects in X-ray baggage videos. A comparative analysis was performed between the performances of the two models with different auto-anchors to verify the superior small-object detection performance of the newly developed model. To enhance the detection of small objects, a neck layer, N2, which further expands the last feature map by one stage (2×), was added. This feature map was then added to the early feature map of the backbone, to ensure the effectiveness of spatial and semantic information. A new detection

layer, D2, was also added to the N2-produced feature map, and an anchor box of very small size was created such that learning would specialize on small objects.

Consequently, the small-object detection performance was enhanced compared with that of previous studies and the YOLOv5 model and a suitable detection performance of medium-sized and large objects maintained. This has great academic significance in that this was not previously achieved in the field of object detection. In addition, prior studies were frequently performed using synthetic target data and oversampling. Thus, when the models were implemented on actual systems, the performance was greatly reduced compared with the reported performance. Therefore, the model was reconstructed only through tuning and the enhancement of the architecture without data manipulation, while producing an improved performance, which is of great significance.

The necessity to detect small hazardous items in baggage checks at transport hubs such as airports, ports, and train stations is directly linked to the lives and safety of passengers and is expected to become more critical in the future. Terrorism is increasing, the methods involved are continuously becoming more complex, and the types of items used are becoming more diverse, leading to more hazards. In addition to the small items analyzed in this study, many other small and ultrasmall items that may be used for terrorism must be rapidly identified for detection and researched continuously. In addition to the model enhancement methods proposed herein, continuous efforts to overcome the limitations and achieve improved results, such as research on bold layer composition, the diversification of feature map information utilization, and the improvement of detection efficiency, are extremely necessary.

We believe that SHOMY will contribute further to research dedicated to overcoming the limitations of one-stage object detectors. It is also expected that air transport and security industries, which require detection of various small objects in environments packed with objects of different sizes, can benefit from the development of an embedded system or application of an experimental method utilizing the results of this study.

Acknowledgement

The authors are grateful to the Korea National Information Society Agency (NIA) for providing the data for the study. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (No.2020R1G1A1100493). This Study was supported by the Institute of Management Research at Seoul National University.

References

[1] A. Abadie and J. Gardeazabal, “Terrorism and the world economy,” *Eur. Econ. Rev.*, vol. 52, no. 1, pp. 1–27, 2008. Article (CrossRef Link)

[2] B. Philip, *Violence in the Skies: A History of Aircraft Hijacking and Bombing*, West Sussex, Summersdale Publishers Ltd., 2016.

[3] National Consortium for the Study of Terrorism and Responses to Terrorism (START), University of Maryland. The Global Terrorism Database (GTD). [Online]. Available: https://www.start.umd.edu/gtd/.

[4] A. K. Novakoff, “FAA bulk technology overview for explosives detection,” *Applications of Signal and Image Processing in Explosives Detection Systems*, vol. 1824, pp. 2–12, 1993. Article (CrossRef Link)

[5] S. Singh and M. Singh, “Explosives detection systems (EDS) for aviation security,” *Signal Processing*, vol. 83, no. 1, pp. 31–55, 2003. Article (CrossRef Link)

[6] D. Gillen and W. G. Morrison, "Aviation security: costing, pricing, finance and performance," *J Air Transp Manag*, vol. 48, pp. 1–12, 2015. Article (CrossRef Link)
[7] Y. Sterchi and A. Schwaninger, "A first simulation on optimizing eds for cabin baggage screening regarding throughput," in *Proc. of the 49th IEEE International Carnahan Conference on Security Technology*, Taipei, Taiwan, pp. 55–60, 2015. Article (CrossRef Link)
[8] N. Hättenschwilericole, Y. Sterchi, M. Mendes, and A. Schwaninger, "Automation in airport security x-ray screening of cabin baggage: examining benefits and possible implementations of automated explosives detection," *Appl. Ergon.*, vol. 72, pp. 58–68, 2018. Article (CrossRef Link)
[9] H. S. Eom, "Softonnet Inc. Introduces Artificial Intelligence X-Ray Security Search Automatic Reading System at Incheon Airport," [Online]. Available: https://www.boannews.com/media/view.asp?idx=93264&direct=mobile.
[10] Á. Morera, Á. Sánchez, A. B. Moreno, Á. D. Sappa, and J. F. Vélez, "SSD vs. YOLO for detection of outdoor urban advertising panels under multiple variabilities," *Sensors*, vol. 20, no. 16, pp. 4587, 2020. Article (CrossRef Link)
[11] K. Tong, Y. Wu, and F. Zhou, "Recent advances in small object detection based on deep learning: a review," *Image Vis Comput*, vol. 97, 2020, Art. no. 103910. Article (CrossRef Link)
[12] P. Adarsh, P. Rathi, and M. Kumar, "YOLO v3-Tiny: Object detection and recognition using one stage improved model," in *Proc. of 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, pp. 687–694, 2020. Article (CrossRef Link)
[13] M. Liu, X. Wang, A. Zhou, X. Fu, Y. Ma, and C. Piao, "UAV-YOLO: small object detection on unmanned aerial vehicle perspective," *Sensors*, vol. 2, no. 8, 2020, Art. no. 2238. Article (CrossRef Link)
[14] J. Rabbi, N. Ray, M. Schubert, S. Chowdhury, and D. Chao, "Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network," *Remote Sens.*, vol. 12, no. 9, 2020, Art. no. 1432. Article (CrossRef Link)
[15] M. Kisantal, Z. Wojna, J. Murawski, J. Naruniec, and K. Cho, "Augmentation for small object detection," *arXiv:1902.07296*, 2019. Article (CrossRef Link)
[16] F. O. Unel, B. O. Ozkalayci. and C. Cigla, "The power of tiling for small object detection," in *Proc. of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, California, USA, 2019. Article (CrossRef Link)
[17] J. N. Lee and H. J. Cho, "Development of artificial intelligence system for dangerous object recognition in x-ray baggage images," *Trans. Korean Inst. Electr. Eng.*, vol. 69, no. 7, pp. 1067–1072, 2020. Article (CrossRef Link)
[18] S. Akcay, M. E. Kundegorski, C. G. Willcocks, and T. P. Breckon, "Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security Imagery," *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 9, pp. 2203–2215, 2018. Article (CrossRef Link)
[19] D. Saavedra, S. Banerjee, and D. Mery, "Detection of threat objects in baggage inspection with x-ray images using deep learning," *Neural Comput. Appl.*, vol. 33, pp. 7803–7819, 2021. Article (CrossRef Link)
[20] C. Wang, H. M. Liao, Y. Wu, P. Chen, J. Hsieh, and I. Yeh, "CSPNet: a new backbone that can enhance learning capability of CNN," in *Proc. of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Washington, USA, pp. 390–391, 2020. Article (CrossRef Link)
[21] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Utah, USA, pp. 8759–8768, 2018. Article (CrossRef Link)
[22] Z. Zheng, P. Wang, D. Ren, W. Liu, R. Ye, Q. Hu, and W. Zuo, "Enhancing geometric factors in model learning and inference for object detection and instance segmentation," *arXiv:2005.03572*, 2020. Article (CrossRef Link)
[23] C. Wang, A. Bochkovskiy, and H. M. Liao, "Scaled-YOLOv4: scaling cross stage partial network," *arXiv:2011.08036*, 2020. Article (CrossRef Link)

[24] S. M. Kim, "8 Types of Aviation Security Equipment," [Online]. Available: https://www.boannews.com/media/view.asp?idx=80922
[25] N. A. Andriyanov, A. K. Volkov, A. K. Volkov, and A. A. Gladkikh, "Research of recognition accuracy of dangerous and safe x-ray baggage images using neural network transfer learning," in *Proc. of IOP Conference Series: Materials Science and Engineering*, vol. 1061, 2021, Art. no. 012002. Article (CrossRef Link)
[26] Y. Wei and X. Liu, "Dangerous goods detection based on transfer learning in x-ray images," *Neural Comput. Appl.*, vol. 32, no. 12, pp. 8711–8724, 2020. Article (CrossRef Link)
[27] A. O. Vuola, S. U. Akram, and J. Kannala, "Mask-RCNN and U-net ensembled for nuclei segmentation," in *Proc. of 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Venice, Italy, pp. 208–212, 2019. Article (CrossRef Link)

Eunchan Kim received the B.A. degree in economics from the University of Minnesota, Twin Cities in 2012 and the M.S. degree in management information systems from Seoul National University in 2017. He is currently pursuing his Ph.D. degree at the Department of Intelligence and Information, Seoul National University in parallel with his career as a Senior Researcher with Hanwha Group, Republic of Korea. His research interests include artificial intelligence, data science, information systems, digitalization, and financial studies.

Jinyoung Lee received the B.S. degree in informatics from University of Washington–Seattle in 2008 and the M.S. degree in artificial intelligence at Yonsei University in 2022. He is currently working as a Data Scientist with Hyundai Group. His research interests include deep learning, data analysis, customer relationship management (CRM), and digital transformation.

Hyunjik Jo received the M.S. degree in artificial intelligence at Yonsei University. He is currently working as a Research Engineer with LG AI Research. His research interests include computer vision, natural language processing, large scale pre-train model.

Kwangtek Na received the B.Sc. degree in civil engineering and the M.S. degree in computer science and engineering from Inha University, South Korea, in 2013 and 2017, respectively, where he is currently pursuing the Ph.D. degree at the Department of Electrical and Computer Engineering. He is currently researching machine learning with the Hanwha Group. His research interests include statistical machine learning, reinforcement learning, and recommender systems.

Eunsook Moon received the M.S. degree in artificial intelligence at Yonsei University in 2022. She is currently working as an IT Technician with Hyundai Motors. Her research interests include machine and deep learning, computer vision, and natural language processing.

Gahgene Gweon is an Associate Professor with the Graduate School of Convergence Science and Technology, Seoul National University. She received the B.A. degree in computer science and economics from University of California, Berkeley. She also holds M.S. and Ph.D. degrees in human–computer interaction from Carnegie Mellon University. Her research interests include natural language processing, human–computer interaction, learning science, and multimedia educational technology.

Byungjoon Yoo is a Professor with the College of Business Administration, Seoul National University. Prior to joining Seoul National University, he worked at Korea University and Hong Kong University of Science and Technology. His research interests include B2B e-commerce, online auctions, and pricing strategies of digital goods such as software products and online games. He has published on these topics in journals such as Management Science, Journal of Management Information Systems, Journal of Marketing, and Decision Support Systems. He has consulting experience with Korea Stock Exchange, Korea Association of Game Industry, and other companies in which he measured the impact of information systems and online transactions, and recommended ways to use information systems strategically.

Yeunwoong Kyung received B.S. and Ph.D. degrees from Korea University, Seoul, Korea, in 2011 and 2016, respectively, both in School of Electrical Engineering. He was a staff engineer at advanced CP Lab., Mobile Communications Business, in Samsung Electronics. He is currently an Assistant Professor with the School of Computer Engineering, Hanshin University, Osan, South Korea. His current research interests include mobility management, mobile cloud computing, SDN/NFV, and IoT.