

Article

Novel Learning Framework with Generative AI X-Ray Images for Deep Neural Network-Based X-Ray Security Inspection of Prohibited Items Detection with You Only Look Once

Dongsik Kim and Jinho Kang * 

School of Electronic Engineering, Gyeongsang National University, Jinju-si 52828, Republic of Korea; ehdtlr9724@gnu.ac.kr

* Correspondence: jinhokang@gnu.ac.kr

Abstract: As the rapid expansion of future mobility systems increases, along with the demand for fast and accurate X-ray security inspections, deep neural network (DNN)-based systems have gained significant attention for detecting prohibited items by constructing high-quality datasets and enhancing detection performance. While Generative AI has been widely explored across various fields, its application in DNN-based X-ray security inspection remains largely underexplored. The accessibility of commercial Generative AI raises safety concerns about the creation of new prohibited items, highlighting the need to integrate synthetic X-ray images into DNN training to improve detection performance, adapt to emerging threats, and investigate its impact on object detection. To address this, we propose a novel machine learning framework that enhances DNN-based X-ray security inspection by integrating real-world X-ray images with Generative AI images utilizing a commercial text-to-image model, improving dataset diversity and detection accuracy. Our proposed framework provides an effective solution to mitigate potential security threats posed by Generative AI, significantly improving the reliability of DNN-based X-ray security inspection systems, as verified through comprehensive evaluations.

Keywords: X-ray security inspection; prohibited items; machine learning; deep neural network; object detection; generative AI; copy-paste augmentation; novel framework; image generation; YOLO



Academic Editor: George A. Tsihrintzis

Received: 14 February 2025

Revised: 21 March 2025

Accepted: 25 March 2025

Published: 28 March 2025

Citation: Kim, D.; Kang, J. Novel Learning Framework with Generative AI X-Ray Images for Deep Neural Network-Based X-Ray Security Inspection of Prohibited Items Detection with You Only Look Once. *Electronics* **2025**, *14*, 1351. <https://doi.org/10.3390/electronics14071351>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Security inspections are essential for ensuring passenger safety in public transportation, including airplanes, ships, and future mobility systems [1–3]. Currently, X-ray detection systems rely on manual inspection by security personnel to identify prohibited items. However, this task can cause significant fatigue and stress, potentially affecting the accuracy of prohibited item detection. Additionally, personnel-based systems are incompatible with future autonomous mobility systems. Recently, deep neural network (DNN)-based accurate and fast automatic systems used to detect prohibited items in X-ray security inspection have received great attention in recent years.

Specifically, object detection techniques within computer vision that can individually identify prohibited items on images have been widely studied for X-ray security inspection [4–8]. However, X-ray images, produced using high-energy radiation, often appear as overlapping objects due to their transmission properties, making detection in security inspections more challenging. These overlap issues can occur between objects and the back-

ground or among multiple objects [4]. In addition, it is known that the manual collection and annotation of X-ray images is labor-intensive and costly [9].

To resolve these challenges, constructing high-quality datasets and DNN models have been widely studied to improve detection performance in X-ray security inspection. Benchmark datasets have been extensively studied for X-ray security inspection in real-world scenarios, such as GDXray [10], SIXray [11], OPIXray [4], HIXray [5], CLCXray [6], PIDray [7], LSIray [12], DvXray [13], and LDXray [14]. These datasets have been used to improve the detection performance of X-ray prohibited items by addressing issues such as lack of data, positive data imbalances, overlap between objects, and image building from multiple directions. In addition, adding or improving additional modules in DNN models was studied to enhance the detection performance based on benchmark datasets [4–6,8,11–17]. In addition, the framework for data augmentation of X-ray security inspection images with generative adversarial networks (GANs) was proposed to improve detection performance [9,18].

On the other hand, with OpenAI's launch of ChatGPT in 2020, Generative AI (Gen AI) has received great worldwide attention in various fields. It autonomously generates new data such as text, images, video, etc., thereby automating tasks traditionally performed by humans [19,20]. With the availability of various commercial Generative AIs (Gen AIs) such as ChatGPT, Bard, DALL·E, and so on, an environment has been created where anyone can easily generate text or high-quality images simply by providing prompt input [19–24]. In this regard, research on the utility and impact of Gen AIs, including several concerns of their exploiting capabilities, has been widely studied in various applications [25,26]. Specifically, images created by Gen AI have become so sophisticated that distinguishing them from real images has become increasingly challenging [27]. Accordingly, the potential risks associated with Gen AI technologies, such as deepfakes, have become more prominent [26]. However, to the best of our knowledge, research on the utility and impact of Gen AIs in machine learning-based X-ray security inspection systems has not been studied, despite its tremendous utility and concerns.

In particular, the widespread availability of commercial Gen AIs, which can generate highly realistic images with just a few lines of input, raises concerns about the potential creation of new prohibited items, posing significant safety threats. Since the performance of machine learning-based X-ray security inspection systems relies heavily on the training dataset [28], there is a concern that systems trained solely on traditional benchmark datasets may fail to detect newly created prohibited items generated by commercial Gen AIs. This limitation could lead to serious security risks in public transportation and future mobility systems. Hence, to accurately detect prohibited items, including newly emerging ones, synthetic X-ray images generated by commercial Gen AIs should be incorporated into the training of DNNs for X-ray security inspection systems to investigate their impact on the object detection performance of the prohibited items.

In this paper, we propose a novel machine learning framework that integrates real-world X-ray images with Gen AI X-ray images to enhance dataset diversity and improve the detection of prohibited items in DNN-based X-ray security inspection systems. To support our framework, we establish a newly developed Gen AI X-ray image dataset using a commercial text-to-image model, DALL·E 3, providing a valuable resource for training and evaluating DNN models. To improve model robustness and generalization, we leverage copy-paste augmentation, which enhances object overlap, occlusions, and cluttered environments. We systematically evaluate the impact of Gen AI images on detection accuracy by training the model with varying numbers of real-world and synthetic Gen AI X-ray images, assessing its ability to generalize and identify potential limitations. Our evaluations demonstrate that YOLOv8, when trained solely on real-world X-ray images,

struggles to detect prohibited items in Gen AI X-ray images, highlighting the necessity of synthetic data for improved generalization. In contrast, the proposed framework, which incorporates copy-paste-augmented Gen AI X-ray images, significantly enhances detection performance without compromising real-world accuracy in terms of Precision, Recall, and mAP50-95.

2. Proposed Approach

2.1. Framework on Learning with Generative AI X-Ray Images

Our proposed framework is illustrated in Figure 1. We devised a novel approach that enables the automatic security inspection system to learn to detect well-known prohibited items from real-world X-ray image datasets, as well as prohibited items from synthetic X-ray images generated by commercial Generative AI (Gen AI). This approach addresses the limitations of traditional X-ray datasets by enhancing dataset diversity. By reflecting real-world environments and use cases, text-based prompts are used to generate synthetic X-ray images.

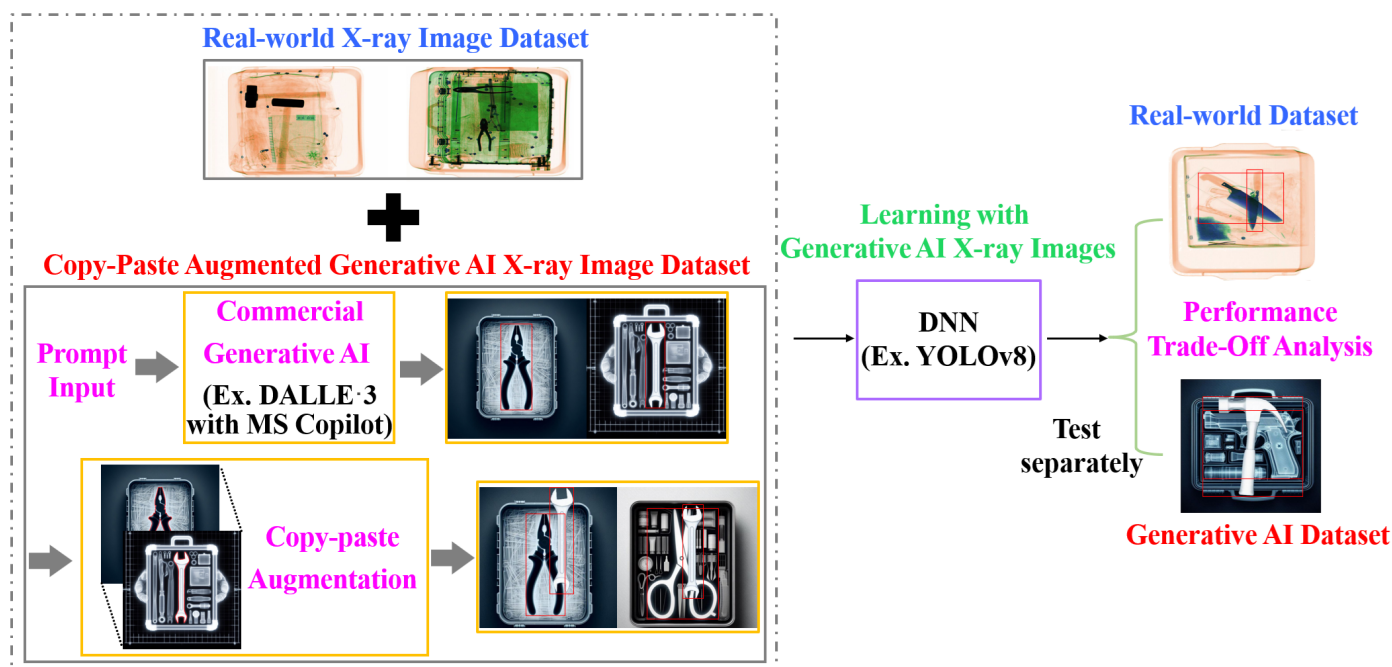


Figure 1. Illustration of our proposed framework.

In addition, to develop a more robust and accurate system while preventing the commercial Gen AI dataset from consisting of simplistic scenarios with a lack of object overlap, we leverage the copy-paste augmentation technique [29], in which individual objects extracted from Gen AI X-ray images are superimposed onto real X-ray detection systems. This augmentation technique exposes the DNN model to a wider range of object placements, occlusions, and cluttered environments, enhancing its ability to detect concealed or partially visible prohibited items. By integrating copy-paste-augmented Gen AI X-ray images into training, our framework enables the ability to improve the model's generalization ability, enhancing detection performance on Gen AI X-ray images while maintaining accuracy in real-world scenarios. The augmented dataset enables the model to learn a broader spectrum of object interactions, occlusions, and spatial complexities, ultimately improving detection accuracy across both domains. The process of generating copy-paste-augmented Gen AI X-ray images is described in the next subsection.

The DNN model is trained using an enriched hybrid dataset that integrates real-world X-ray images and Gen AI X-ray images, enabling the model to generalize across both domains [30]. In this paper, we employed YOLOv8 as our chosen DNN model [31]. YOLOv8 incorporates an efficient architecture that utilizes the self-attention mechanism and deformable convolution, reducing memory requirements while simultaneously providing high detection accuracy and fast processing speed [31,32]. Additionally, YOLOv8 supports built-in data augmentation functions such as scaling, flipping, and cropping, enhancing user convenience and contributing to its widespread adoption across various industries [12,33–35]. Although YOLOv10, the latest version of the YOLO model, has been recently released with advancements and optimizations [36], we chose to utilize YOLOv8 due to its deployment-friendly nature and suitability for industrial applications requiring fast detection.

To evaluate its effectiveness, we conducted a performance tradeoff analysis focusing on the following aspects:

- The impact of Gen AI images on detection accuracy.
- The model's ability to generalize across real-world and synthetic Gen AI X-ray images.

To systematically analyze these tradeoffs, we trained the DNN model with varying numbers of real-world and Gen AI X-ray images in the training dataset. The trained models were then evaluated separately using distinct test datasets. This approach enabled us to determine whether synthetic data enhances model accuracy or introduces limitations, ultimately refining the effectiveness of Gen AI-enhanced datasets. This approach allowed us to determine whether synthetic data improves model accuracy or introduces limitations, ultimately enhancing the effectiveness of Gen AI-enhanced datasets.

2.2. Real-World X-Ray Image Dataset for Prohibited Items

In this paper, we utilized “X-ray multi-object detection data” as the real-world X-ray images of prohibited items, which were provided by AI Hub [37]. In this subsection, we briefly introduce this dataset. AI Hub, supported by MSIT (Ministry of Science and ICT) and NIA (National Information Society Agency) of South Korea, is a platform that provides the infrastructure needed for the development of AI technologies and services, along with various datasets applicable. The “X-ray multi-object detection data” broadly categorizes various types of items in X-ray images into three categories: (1) “Prohibited Items”, (2) “Information Storage Devices”, and (3) “General Items”, where the dataset consists of a total of 541,260 images across 317 classes. The details on the selected dataset that we utilized in the training and evaluation for the prohibited items are described in [37].

Figure 2 shows examples of real-world X-ray images that we utilized. The original dataset of prohibited items was classified into six categories (Gun, Knife, Wrench, Pliers, Scissors, and Hammer) based on the SIXRAY dataset [11] for practical X-ray security inspection. Various subcategories were merged, and unrelated or mislabeled data were excluded. The final dataset consisted of 51,210 real-world X-ray images, which were divided into Train, Validation, and Test sets.

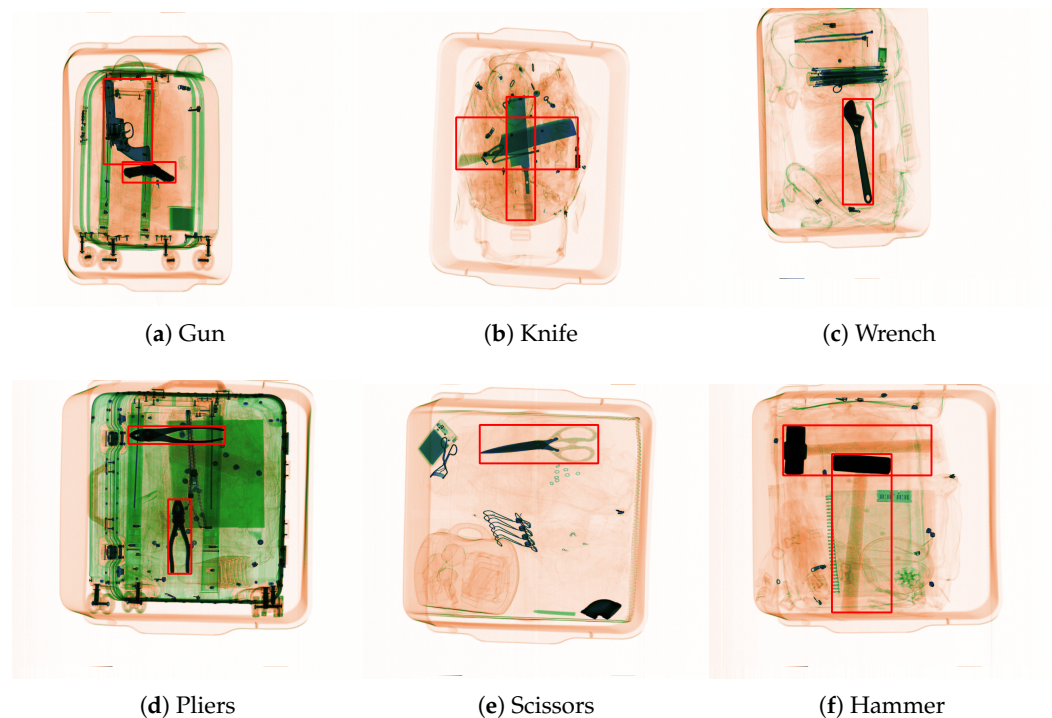


Figure 2. Examples of real-world X-ray images for prohibited items where the red boxes indicate the prohibited items that have been labeled.

2.3. Copy-Paste-Augmented Generative AI X-Ray Image Dataset for Prohibited Items

In order to leverage the Gen AI X-ray image dataset for prohibited items, we produced X-ray images for the prohibited items created by the commercial Gen AI. Among the several text-to-image Gen AI models, we adopted the DALL·E 3 [38], which is developed by Open AI, thanks to its accessibility and availability in ChatGPT4 [20] as well as Microsoft Copilot [39]. To create the images, we input the text prompt into the uncustomized original DALL·E 3 through Microsoft Copilot (we leveraged Microsoft Copilot due to its utility, where several images corresponding to the prompt are generated at once), without any modifications or additional fine-tuning, as follows:

“X-ray image of a box containing a *Prohibited Item*. This image is in the style typically seen by airport security personnel.”

In the above prompt, a *Prohibited Item* means one of Gun, Knife, Wrench, Pliers, Scissors, or Hammer so that it would generate the images with the same classes as the real-world X-ray images. Then, we manually labeled each generated image by utilizing the “Labelling” program, which supports labeling with YOLO and Pascal VOC formats [40], since the generated images lack information on object location, quantity, and type.

Figure 3 shows examples of Gen AI X-ray images for prohibited items created from DALL·E 3. Most images contain one prohibited item, but some images also contain several prohibited items. In addition, due to creative generation on some images, we excluded the generated images that were unclear to identify objects. To match the default input image size of YOLOv8, the size of each Gen AI image was adjusted 640×640 from the initial size of 1024×1024 .



Figure 3. Examples of Gen AI X-ray images for prohibited items that we created, where the red boxes indicate the prohibited items that have been labeled.

Next, we produced augmented images based on the created Gen AI X-ray images using a separate copy-paste Python (version 3.12.2) script [41] instead of using YOLO's built-in augmentation features. This study focuses on images generated by commercial Gen AI. To address the limitations of existing Gen AI X-ray images, which often feature simple compositions and lack object overlap, we applied the copy-paste augmentation technique while preserving other image properties such as color and size. This approach enhances individual images by incorporating additional object information and introducing object occlusion, thereby creating more realistic training data.

Figure 4 illustrates examples of the copy-paste-augmented Gen AI X-ray images for prohibited items that we created. For copy-paste augmentation with instance segmentation labels, we exploited a commonly used labeling tool that supports labeling tasks for various image segmentation and augmentation techniques. In this study, augmentation was primarily performed using the copy-paste technique. However, exploring additional performance improvements by integrating it with various augmentation techniques remains as an on-going topic of research.

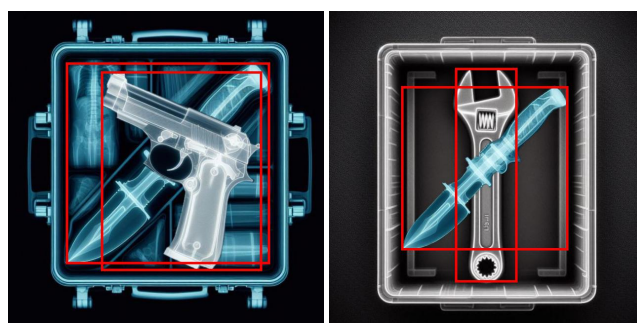


Figure 4. Examples of copy-paste-augmented Gen AI images for prohibited items that we created, where the red boxes indicate the prohibited items that have been labeled.

3. Results and Discussion

In this section, our goal is to evaluate and analyze the performance of our proposed framework by training YOLOv8 while varying the number of real-world X-ray training images and copy-paste-augmented Gen AI X-ray images.

3.1. Performance Metrics and Setup

To evaluate our proposed framework, we considered the performance metrics that are widely adopted in object detection based on the confusion matrix of Table 1 as follows [42]:

Table 1. Confusion matrix.

Actual class	Predicted Class	
	Positive	Negative
	TP	FN
Positive		
Negative	FP	TF

- Precision: Proportion of predicted positive classes that are actually positive, which is defined as $\frac{TP}{TP+FP}$.
- Recall: Proportion of actual positive instances that the model correctly identifies as positive, which is defined as $\frac{TP}{TP+FN}$.
- mAP50-95: A metric that averages the mean Average Precision (mAP) calculated at Intersection over Union (IoU) thresholds ranging from 0.5 to 0.95 in increments of 0.05. This metric includes higher IoU criteria compared to mAP50, thus requiring more accurate bounding boxes.

The environment that we adopted for evaluation is described in Table 2. We utilized YOLOv8 for the DNN model thanks to its lightweight, fast, and high-performance characteristics [31]. The pretrained weights from the COCO dataset were applied to our YOLOv8 model to improve its generalization ability in object detection, which has been widely adopted for benchmarking object detection models [43].

Table 2. Environment for evaluation.

Component	Specification
OS	Windows 11
CPU	Intel i9 139006 (Santa Clara, CA, USA)
RAM	64 GB
Graphics	NVIDIA RTX 4080 SUPER (Santa Clara, CA, USA)
VRAM	16 GB
DNN Model	YOLOv8.2.42

The hyperparameters considered for model training are described as follows. The model was trained for 20 epochs using the AdamW optimizer with an initial learning rate of 0.01, a momentum of 0.937, a weight decay of 0.0005, and a batch size of 16. In addition, the input image size was set to 640×640 . We considered the default hyperparameter values for YOLOv8 to maintain a consistent experimental environment without the influence of tuning on performance differences. Although it is known that enhancing the YOLOv8 structure and tuning its hyperparameters can improve performance, this study focused on performance analysis based on the inclusion of Gen AI images. Therefore, the default configuration was retained to enable a clearer comparison. Extending the proposed framework to incorporate

an improved YOLOv8 by optimizing its structure and hyperparameter configurations remains an ongoing area of research.

To evaluate our proposed framework through performance tradeoff analysis, we trained six separate YOLOv8 models, each with varying numbers of real-world X-ray images and Gen AI X-ray images. Table 3 shows the total number of images for each set, where non-copy-paste-augmented Gen AI (NC-Gen AI) X-ray images mean Gen AI images without copy-paste augmentation technique such as the images in Figure 3. For performance comparison based on the copy-paste augmentation technique, Models 1–3 were trained using NC-Gen AI X-ray images, while Models 4–6 were trained using copy-paste-augmented Gen AI (C-Gen AI) X-ray images such as the images in Figure 4.

Table 3. Total number of X-ray images.

Number of Images	Train	Valid	Test
Real-World	35,850	7680	7680
Non-Copy-Paste Augmented Gen AI	300	60	60
Copy-Paste Augmented Gen AI	300	60	60

More specifically, Table 4 provides the number of training images for Models 1–3 and Models 4–6, respectively, where $\alpha \in [0, 300]$ represents the number of Gen AI X-ray images used in the training set. For example, Model 1 was trained with 1000 real-world X-ray images and α NC-Gen AI X-ray images, while Model 4 was trained with 1000 real-world X-ray images and α C-Gen AI X-ray images. During the evaluations, α increased from 0 to 300 so that each model was trained with an increasing number of Gen AI X-ray images. In this case, a uniform distribution was used to sample the number of images corresponding to the value of α among all training datasets containing 300 Gen AI X-ray images. Also, for Models 1 and 4 and Models 2 and 5, 1000 and 10,000 images, respectively, were also randomly sampled from a real-world training dataset containing 35,850 images.

Table 4. The number of training images for Models 1–6.

YOLOv8	Real-World	NC-Gen AI/C-Gen AI
Model 1/4	1000	α / α
Model 2/5	10,000	α / α
Model 3/6	35,850 (All)	α / α

For a performance tradeoff analysis, each trained model was separately evaluated on a real-world X-ray image test set, an NC-Gen AI X-ray image test set, and a C-Gen AI image test set. This approach aimed to assess the model's existing detection capability while determining its effectiveness in identifying new threat elements in complex environments. Consequently, the impact of the Gen AI dataset on model performance was quantitatively analyzed. Each performance metric was calculated as the average of five experiments conducted with different sampled images for a fixed number of images, i.e., 1000 or 10,000 sampled real-world images and α sampled Gen AI images.

3.2. Evaluation and Performance Analysis

In this subsection, Figures 5–7 illustrate the performance of each model in terms of Precision, Recall, and mAP50-95 with respect to the number of training Gen AI X-ray images, i.e., α , evaluated respectively by (a) a real-world X-ray image test set, (b) an NC-

Gen AI X-ray image test set, and (c) a C-Gen AI X-ray image test set. In addition, the detailed values associated with each figure are provided in Tables 5–7, respectively.

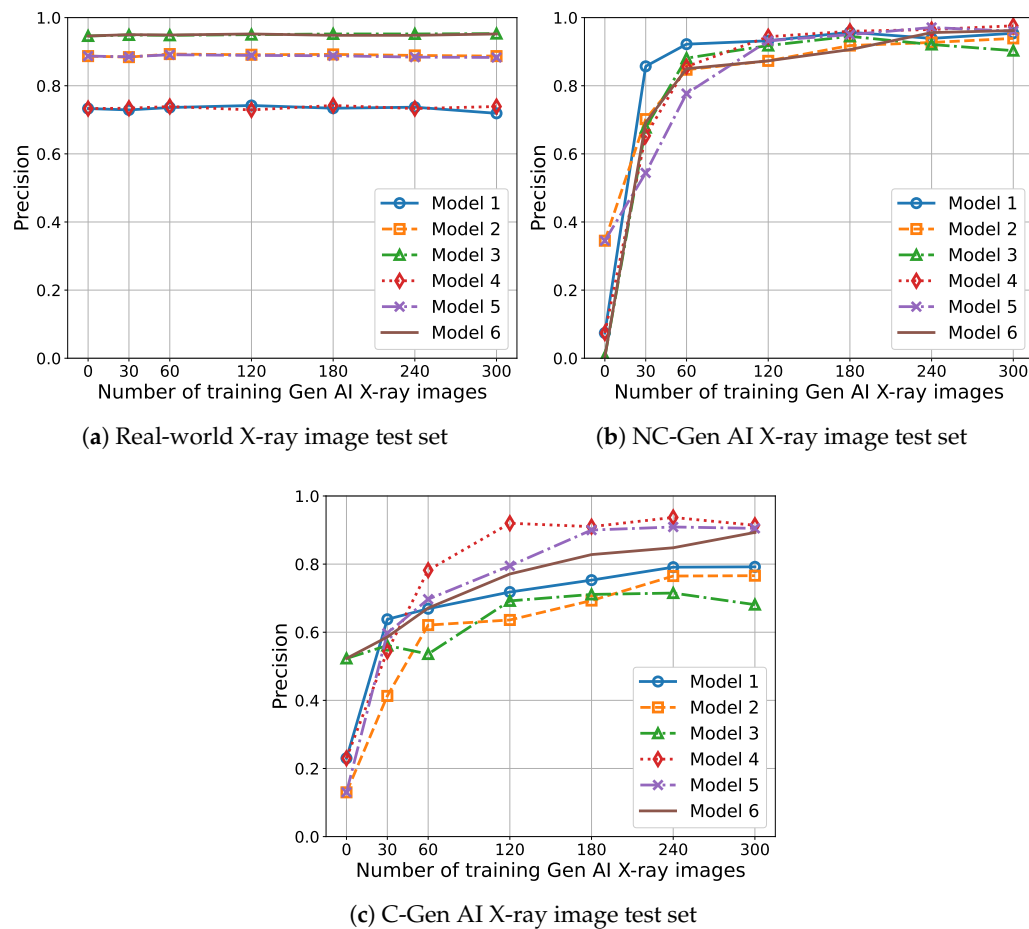


Figure 5. Precision with respect to the number of training Gen AI X-ray images evaluated by (a) real-world X-ray image test set, (b) NC-Gen AI X-ray image test set, (c) C-Gen AI X-ray image test set.

Table 5. Precision with respect to the number of training Gen AI X-ray images evaluated by real-world X-ray image test set, NC-Gen AI X-ray image test set, and C-Gen AI X-ray image test set.

Number of Training Gen AI X-Ray Images		0	30	60	120	180	240	300
Real-world X-ray image test	Model 1	0.733	0.729	0.736	0.742	0.734	0.737	0.719
	Model 2	0.887	0.884	0.893	0.891	0.892	0.889	0.887
	Model 3	0.946	0.949	0.948	0.950	0.952	0.952	0.953
	Model 4	0.733	0.734	0.739	0.729	0.742	0.733	0.739
	Model 5	0.887	0.885	0.891	0.889	0.888	0.884	0.883
	Model 6	0.946	0.950	0.949	0.952	0.948	0.948	0.952
NC-Gen AI X-ray image test	Model 1	0.074	0.857	0.922	0.932	0.957	0.939	0.954
	Model 2	0.345	0.702	0.847	0.873	0.918	0.927	0.939
	Model 3	0.004	0.677	0.881	0.918	0.945	0.921	0.903
	Model 4	0.074	0.653	0.856	0.944	0.960	0.965	0.976
	Model 5	0.345	0.544	0.777	0.932	0.950	0.971	0.960
	Model 6	0.004	0.692	0.850	0.873	0.906	0.956	0.962
C-Gen AI X-ray image test	Model 1	0.230	0.638	0.669	0.718	0.753	0.791	0.792
	Model 2	0.130	0.413	0.621	0.636	0.693	0.765	0.766
	Model 3	0.523	0.561	0.536	0.692	0.711	0.715	0.681
	Model 4	0.230	0.544	0.782	0.920	0.910	0.937	0.914
	Model 5	0.130	0.597	0.697	0.795	0.900	0.909	0.905
	Model 6	0.523	0.586	0.671	0.771	0.828	0.848	0.893

First, Figure 5a–c represent the Precision performance, and the corresponding detailed values are provided in Table 5. In Figure 5a, the Precision remained consistently high for all models, regardless of α . Specifically, the copy-paste augmentation or not did not affect the Precision performance for real-world images. As expected, Models 3 and 6 exhibited the highest performance, while Models 1 and 4 showed the lowest performance. Figure 5b shows that the Precision considerably increased as α increased for all models. The models trained without the Gen AI X-ray images (i.e., $\alpha = 0$) yielded Precision values smaller than 0.4, while the models trained with 120 images achieved Precision values higher than 0.8. Models 4–6 slightly outperformed Models 1–3 and achieved near-perfect Precision values with 300 C-Gen AI images. Figure 5c demonstrates the Precision performance evaluated for the C-Gen AI test set. Similarly, increasing α led to improved performance. On the other hand, Models 4–6 considerably outperformed Models 1–3 for all Gen AI image levels and achieved a 0.9 Precision value with 300 C-Gen AI training images.

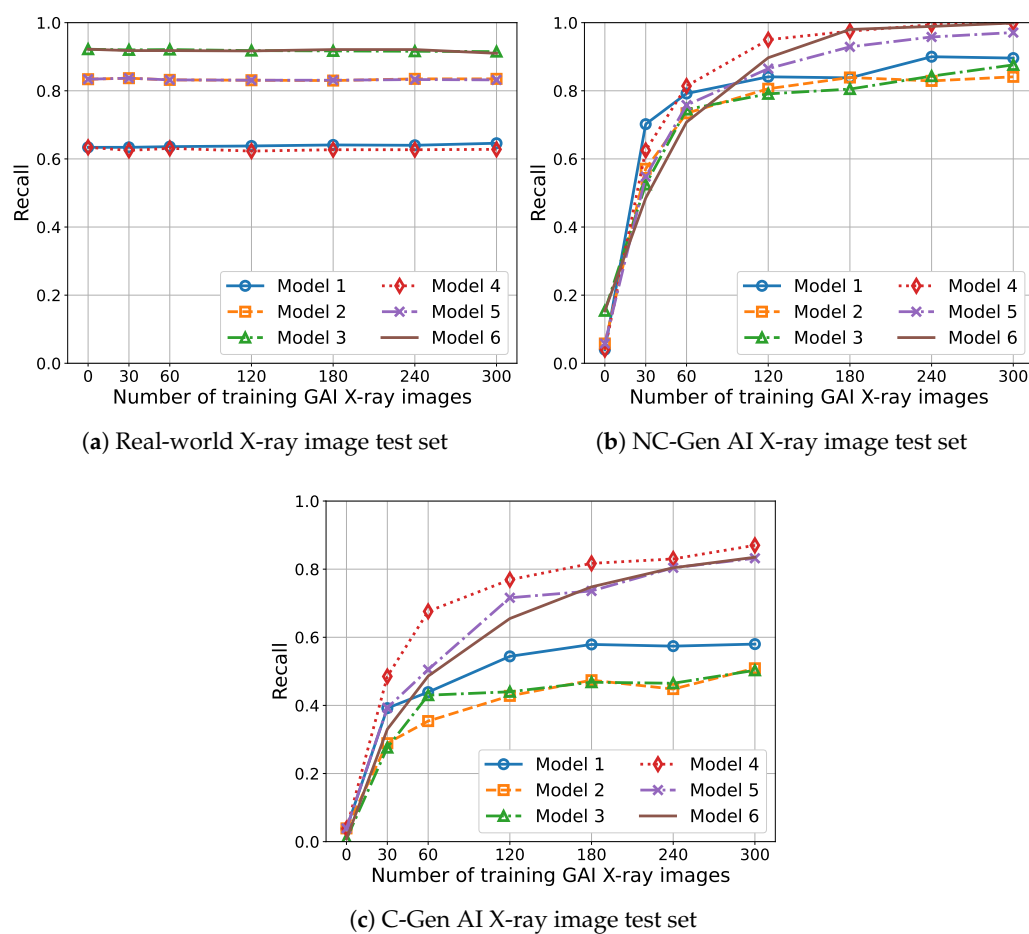


Figure 6. Recall with respect to the number of training Gen AI X-ray images evaluated by (a) real-world X-ray image test set, (b) NC-Gen AI X-ray image test set, (c) C-Gen AI X-ray image test set.

Next, Figure 6a–c show the Recall performance, and the corresponding detailed values are provided in Table 6. In Figure 6a, the Recall remained consistently high for all models regardless of α , and the copy-paste augmentation or not had little to no effect on the Recall performance for real-world images, where Models 3 and 6 exhibited the highest performance. Figure 6b,c exhibit that the Recall considerably increased as α increased on both the NC-Gen AI and C-Gen AI test sets. In particular, the models trained without the Gen AI X-ray images (i.e., $\alpha = 0$) yielded Recall values smaller than 0.2. Moreover, Models 4–6 outperformed Models 1–3 on both test sets after 120 images, with a larger performance gap observed in the C-Gen AI test set. Models 4–6 achieved near-perfect Recall values on

the NC-Gen AI test set and approximately 0.85 in their Recall values on the C-Gen AI test set when trained with 300 C-Gen AI images. In addition, Model 4 demonstrated the highest performance on the C-Gen AI test set.

Table 6. Recall with respect to the number of training Gen AI X-ray images evaluated by real-world X-ray image test set, NC-Gen AI X-ray image test set, and C-Gen AI X-ray image test set.

Number of Training Gen AI X-Ray Images		0	30	60	120	180	240	300
Real-world X-ray image test	Model 1	0.634	0.634	0.636	0.638	0.641	0.640	0.646
	Model 2	0.834	0.837	0.832	0.831	0.830	0.835	0.835
	Model 3	0.922	0.920	0.921	0.918	0.917	0.916	0.915
	Model 4	0.634	0.625	0.631	0.623	0.627	0.627	0.628
	Model 5	0.834	0.837	0.832	0.831	0.831	0.833	0.832
	Model 6	0.922	0.918	0.918	0.917	0.921	0.921	0.910
NC-Gen AI X-ray image test	Model 1	0.040	0.702	0.792	0.841	0.838	0.900	0.896
	Model 2	0.058	0.571	0.734	0.806	0.839	0.829	0.841
	Model 3	0.154	0.525	0.745	0.791	0.805	0.843	0.876
	Model 4	0.040	0.625	0.814	0.950	0.975	0.994	0.999
	Model 5	0.058	0.548	0.758	0.865	0.929	0.958	0.971
	Model 6	0.154	0.485	0.708	0.897	0.980	0.989	0.999
C-Gen AI X-ray image test	Model 1	0.038	0.392	0.439	0.544	0.579	0.574	0.580
	Model 2	0.039	0.289	0.354	0.428	0.474	0.448	0.509
	Model 3	0.008	0.276	0.430	0.440	0.468	0.465	0.503
	Model 4	0.038	0.485	0.676	0.769	0.817	0.830	0.870
	Model 5	0.039	0.393	0.505	0.716	0.736	0.804	0.832
	Model 6	0.008	0.330	0.486	0.655	0.748	0.804	0.835

Table 7. mAP50-95 with respect to the number of training Gen AI X-ray images evaluated by real-world X-ray image test set, NC-Gen AI X-ray image test set, and C-Gen AI X-ray image test set.

Number of Training Gen AI X-Ray Images		0	30	60	120	180	240	300
Real-world X-ray image test	Model 1	0.539	0.539	0.542	0.547	0.548	0.549	0.550
	Model 2	0.769	0.770	0.769	0.769	0.770	0.770	0.770
	Model 3	0.861	0.860	0.861	0.860	0.861	0.861	0.861
	Model 4	0.539	0.537	0.538	0.534	0.540	0.535	0.538
	Model 5	0.769	0.770	0.769	0.767	0.767	0.767	0.767
	Model 6	0.861	0.860	0.860	0.859	0.861	0.860	0.860
NC-Gen AI X-ray image test	Model 1	0.001	0.511	0.700	0.823	0.859	0.885	0.888
	Model 2	0.002	0.399	0.586	0.663	0.712	0.791	0.810
	Model 3	0.001	0.436	0.716	0.799	0.804	0.826	0.837
	Model 4	0.001	0.449	0.650	0.849	0.918	0.934	0.956
	Model 5	0.002	0.318	0.552	0.709	0.775	0.872	0.898
	Model 6	0.001	0.417	0.673	0.806	0.876	0.898	0.934
C-Gen AI X-ray image test	Model 1	0.003	0.217	0.293	0.419	0.430	0.449	0.478
	Model 2	0.004	0.134	0.201	0.247	0.288	0.337	0.358
	Model 3	0.001	0.146	0.252	0.321	0.324	0.354	0.339
	Model 4	0.003	0.263	0.439	0.637	0.704	0.759	0.783
	Model 5	0.004	0.178	0.296	0.452	0.544	0.637	0.687
	Model 6	0.001	0.207	0.353	0.501	0.602	0.659	0.699

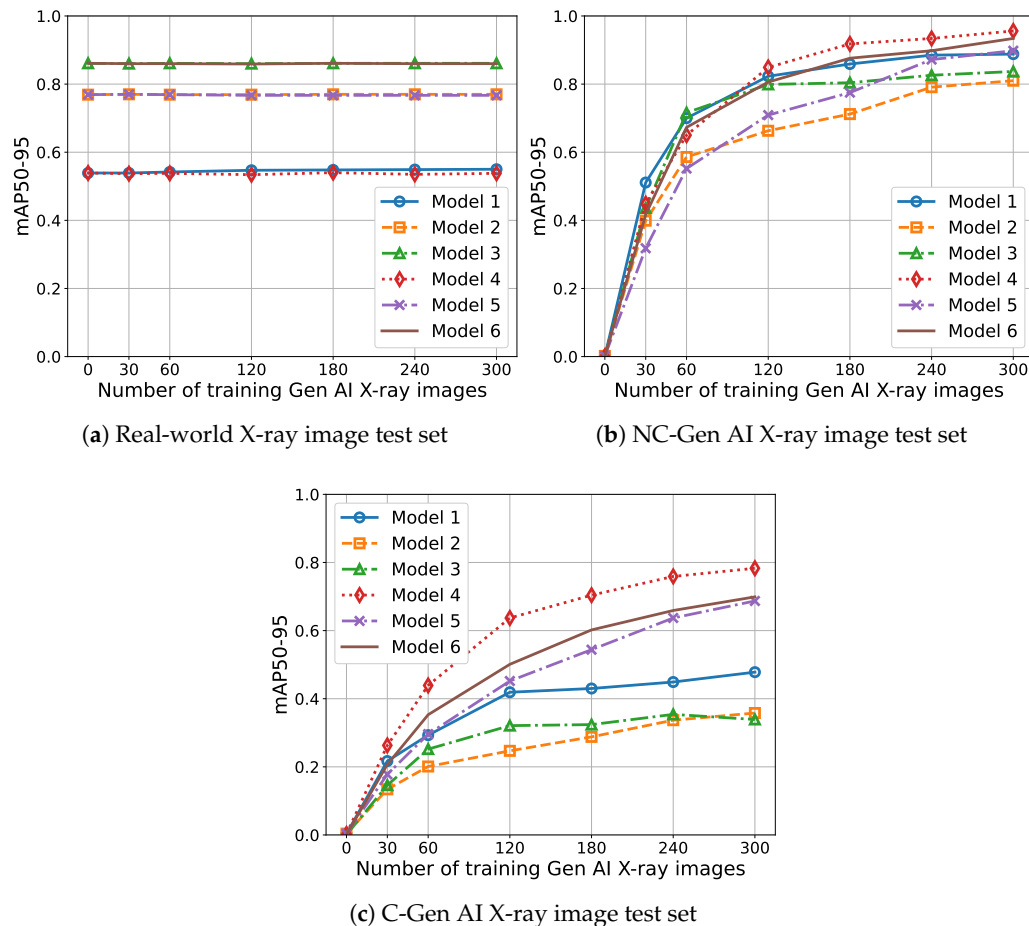


Figure 7. mAP50-95 with respect to the number of training Gen AI X-ray images evaluated by (a) real-world X-ray image test set, (b) NC-Gen AI X-ray image test set, (c) C-Gen AI X-ray image test set.

Figure 7a–c show the mAP50-95 performance, and the corresponding detailed values are provided in Table 7. Figure 7a demonstrates the similar performance to the Precision and Recall results with respect to α . Also, Figure 7b,c demonstrate that the mAP50-95 significantly increased as α increased on both the NC-Gen AI and C-Gen AI test sets. Models 4–6 outperformed Models 1–3 on both test sets, with a larger performance gap observed in the C-Gen AI test set. Specifically, Models 4 and 6 achieved 0.9 in their mAP50-95 values on both NC-Gen AI test sets with 300 C-Gen AI training images. On the other hand, Model 4 demonstrated the highest mAP50-95 performance on the C-Gen AI test set and achieved near 0.8 in its mAP50-95 value with 300 C-Gen AI training images, while Model 6 yielded near 0.7 value.

3.3. Discussion

In this subsection, we analyze and discuss the performance of each model individually from the experimental results. First, Model 1 consistently exhibited the lowest performance in the real-world X-ray image test set, regardless of the increase in Gen AI X-ray images. However, in the NC-Gen AI X-ray image test set, its performance improved as the number of Gen AI images increased, exceeding the average. In the C-Gen AI X-ray image test set, Model 1 achieved the highest performance among Models 1–3 as the number of Gen AI images increased, but the performance gap between Model 1 and Models 4–6 widened progressively.

Model 2 maintained stable performance in the real-world X-ray image test set, even as the number of Gen AI X-ray images increased. In the NC-Gen AI X-ray test set, it demonstrated excellent Precision as the number of Gen AI images increased, but the Recall and mAP50-95 values remained relatively low. In the C-Gen AI X-ray image test set, Model 2 exhibited a significant performance gap compared to Models 4–6. Among Models 1–3, it showed strong performance in terms of Precision, but the Recall and mAP50-95 values remained low.

Model 3 achieved the highest performance in the real-world X-ray image test set, regardless of the increase in Gen AI X-ray images. In the NC-Gen AI X-ray test set, the Precision was strong, but the Recall and mAP50-95 were relatively poor. In the C-Gen AI X-ray image test set, Model 3 recorded the lowest performance overall.

Model 4 exhibited a performance trend similar to Model 1 in the real-world X-ray image test set. However, in both the NC-Gen AI and C-Gen AI X-ray image test sets, Model 4 achieved the highest performance as the number of Gen AI images increased.

Model 5 showed performance nearly identical to Model 2 in the real-world X-ray image test set. It performed well in the NC-Gen AI X-ray image test set and exhibited high performance across all metrics in the C-Gen AI X-ray image test set. However, its mAP50-95 score was approximately 0.1 lower than that of Model 4.

Finally, Model 6 demonstrated performance similar to Model 3, achieving the highest scores in the real-world X-ray image test set. It also performed well in the NC-Gen AI X-ray test set. In the C-Gen AI X-ray image test set, it achieved high performance, but its mAP50-95 score was approximately 0.1 lower than that of Model 4, similar to Model 5.

The results conclude that YOLOv8, when trained solely on real-world X-ray images, exhibits limited detection ability on Gen AI images. However, incorporating Gen AI images into training enhances performance on synthetic data while maintaining real-world accuracy. Additionally, copy-paste augmentation further improves the overall detection capabilities.

4. Conclusions

This paper studied a novel machine learning framework for detecting prohibited items in DNN-based X-ray security inspection systems. As Generative AI technology advances, allowing anyone to create desired images using natural language without requiring specialized AI knowledge, security systems trained solely on real-world images may fail to detect newly generated prohibited items. This limitation could pose a security vulnerability and potential risk. To address this, a new Gen AI X-ray image dataset was created using a commercial text-to-image Gen AI model and was used to train and evaluate YOLOv8. Experimental results indicate that when the YOLOv8 model was trained solely on real-world X-ray images, it failed to detect prohibited items in Gen AI X-ray images, suggesting the necessity of supplementary training using Generative AI data. In contrast, models incorporating copy-paste Gen AI X-ray images significantly improved detection performance without compromising real-world detection accuracy. These findings suggest that leveraging commercial Gen AI can enhance the accuracy of DNN-based X-ray security inspection systems. Furthermore, incorporating just a small number of Gen AI images into the training set can provide a powerful solution to mitigate the security risks associated with Gen AI.

To verify whether the generated X-ray images truly reflect real-world X-ray images, our ongoing research focuses on implementing Human Expert Verification to enhance the quality and reliability of the generated AI X-ray image dataset [44–46]. This approach will strengthen our proposed framework, making it more robust and accurate. Moreover, our findings demonstrate that training with Generative AI data can enhance performance. However, they often introduce bias, generate unrealistic objects, and may lead to model

collapse, creating a tradeoff [44–46]. To address these challenges, future research will focus on developing strategies to maximize performance while minimizing drawbacks. This includes constructing additional Generative AI datasets and adjusting the ratio of real to generated data. Furthermore, expert validation of the generated images will be conducted to ensure data quality and reliability, as well as to explore strategies for mitigating potential model bias.

Author Contributions: Conceptualization, D.K. and J.K.; Methodology, D.K. and J.K.; Software, D.K.; Validation, D.K. and J.K.; Formal analysis, D.K. and J.K.; Investigation, D.K. and J.K.; Resources, J.K.; Writing—original draft preparation, D.K.; writing—review and editing, J.K. Visualization, D.K. and J.K.; Supervision J.K.; Project administration, J.K.; Funding acquisition, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the NRF(National Research Foundation of Korea) grant funded by the Korea government (Ministry of Science and ICT) (RS-2023-00214142), and in part by the IITP(Institute of Information & Communications Technology Planning & Evaluation)-ICAN(ICT Challenge and Advanced Network of HRD) grant funded by the Korea government (Ministry of Science and ICT) (IITP-2025-RS-2022-00156409).

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author. For the real-world X-ray image dataset, this paper used datasets from ‘The Open AI Dataset Project (AI Hub, S. Korea)’, where all data information can be accessed through ‘AI Hub (<https://www.aihub.or.kr>, accessed on 31 January 2024)’.

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

References

1. Akçay, S.; Kundegorski, M.E.; Devereux, M.; Breckon, T.P. Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 1057–1061.
2. Saavedra, D.; Banerjee, S.; Mery, D. Detection of threat objects in baggage inspection with X-ray images using deep learning. *Neural Comput. Appl.* **2021**, *33*, 7803–7819.
3. Lee, J.N.; Cho, H.C. Development of artificial intelligence system for dangerous object recognition in X-ray baggage images. *Trans. Korean Inst. Electr. Eng.* **2020**, *69*, 1067–1072.
4. Wei, Y.; Tao, R.; Wu, Z.; Ma, Y.; Zhang, L.; Liu, X. Occluded prohibited items detection: An X-ray security inspection benchmark and de-occlusion attention module. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 138–146.
5. Tao, R.; Wei, Y.; Jiang, X.; Li, H.; Qin, H.; Wang, J.; Ma, Y.; Zhang, L.; Liu, X. Towards real-world X-ray security inspection: A high-quality benchmark and lateral inhibition module for prohibited items detection. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 10–17 October 2021; pp. 10923–10932.
6. Zhao, C.; Zhu, L.; Dou, S.; Deng, W.; Wang, L. Detecting overlapped objects in X-ray security imagery by a label-aware mechanism. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 998–1009.
7. Zhang, L.; Jiang, L.; Ji, R.; Fan, H. Pidray: A large-scale X-ray benchmark for real-world prohibited item detection *Int. J. Comput. Vis.* **2023**, *131*, 3170–3192.
8. Kim, E.; Lee, J.; Jo, H.; Na, K.; Moon, E.; Gweon, G.; Yoo, B.; Kyung, Y. SHOMY: Detection of Small Hazardous Objects using the You Only Look Once Algorithm. *KSII Trans. Internet Inf. Syst. (TIIS)* **2022**, *16*, 2688–2703.
9. Liu, J.; Lin, T.H. A framework for the synthesis of X-ray security inspection images based on generative adversarial networks. *IEEE Access* **2023**, *11*, 63751–63760.
10. Mery, D.; Rizzo, V.; Zscherpel, U.; Mondragón, G.; Lillo, I.; Zuccar, I.; Lobel, H.; Carrasco, M. GDXray: The database of X-ray images for nondestructive testing. *J. Nondestruct. Eval.* **2015**, *34*, 42.
11. Miao, C.; Xie, L.; Wan, F.; Su, C.; Liu, H.; Jiao, J.; Ye, Q. Sixray: A large-scale security inspection X-ray benchmark for prohibited item discovery in overlapping images. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2119–2128.

12. Han, L.; Ma, C.; Liu, Y.; Jia, J.; Sun, J. SC-YOLOv8: A security check model for the inspection of prohibited items in X-ray images. *Electronics* **2023**, *12*, 4208. [CrossRef]
13. Ma, B.; Jia, T.; Li, M.; Wu, S.; Wang, H.; Chen, D. Towards dual-view X-ray baggage inspection: A large-scale benchmark and adaptive hierarchical cross refinement for prohibited item discovery *IEEE Trans. Inf. Forensics Secur.* **2024**, *19*, 3866–3878.
14. Tao, R.; Wang, H.; Guo, Y.; Chen, H.; Zhang, L.; Liu, X.; Wei, Y.; Zhao, Y. Dual-view X-ray detection: Can AI detect prohibited items from dual-view X-ray images like humans? *arXiv* **2024**, arXiv:2411.18082.
15. Li, Y.; Zhang, C.; Sun, S.; Yang, G. X-ray detection of prohibited item method based on dual attention mechanism. *Electronics* **2023**, *12*, 3934. [CrossRef]
16. Jing, B.; Duan, P.; Chen, L.; Du, Y. EM-YOLO: An X-ray prohibited-item-detection method based on edge and material information fusion. *Sensors* **2023**, *23*, 8555. [CrossRef] [PubMed]
17. Zhang, H.; Zhao, Z.; Yang, J. Attention-based prohibited item detection in X-ray images during security checking. *IET Image Process.* **2024**, *18*, 1119–1131. [CrossRef]
18. Zhu, Y.; Zhang, Y.; Zhang, H.; Yang, J.; Zhao, Z. Data augmentation of X-ray images in baggage inspection based on generative adversarial networks. *IEEE Access* **2020**, *8*, 86536–86544.
19. OpenAI. ChatGPT. 2024. Available online: <https://chat.openai.com> (accessed on 4 April 2024).
20. Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F.L.; Almeida, D.; Altenschmidt, J.; Altman, S.; Anadkat, S.; et al. GPT-4 technical report. *arXiv* **2023**, arXiv:2303.08774.
21. Anil, R.; Dai, A.M.; Firat, O.; Johnson, M.; Lepikhin, D.; Passos, A.; Shakeri, S.; Taropa, E.; Bailey, P.; Chen, Z.; et al. Palm 2 technical report. *arXiv* **2023**, arXiv:2305.10403.
22. Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Roziere, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. Llama: Open and efficient foundation language models. *arXiv* **2023**, arXiv:2302.13971.
23. Ramesh, A.; Pavlov, M.; Goh, G.; Gray, S.; Voss, C.; Radford, A.; Chen, M.; Sutskever, I. Zero-shot text-to-image generation. In Proceedings of the 38th International Conference on Machine Learning, Virtual Event, 18–24 July 2021; pp. 8821–8831.
24. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-resolution image synthesis with latent diffusion models. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 10684–10695.
25. Andreoni, M.; Lunardi, W.T.; Lawton, G.; Thakkar, S. Enhancing autonomous system security and resilience with generative AI: A comprehensive survey. *IEEE Access* **2024**, *12*, 109470–109493. [CrossRef]
26. Golda, A.; Mekonen, K.; Pandey, A.; Singh, A.; Hassija, V.; Chamola, V.; Sikdar, B. Privacy and security concerns in generative AI: A comprehensive survey. *IEEE Access* **2024**, *12*, 48126–48144. [CrossRef]
27. Bird, J.J.; Lotfi, A. CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images. *IEEE Access* **2024**, *12*, 15642–15650.
28. Lee, Y.; Kang, J. Performance Analysis by the Number of Learning Images on Anti-Drone Object Detection System with YOLO. *J. Korean Inst. Commun. Inf. Sci.* **2024**, *49*, 356–360.
29. Ghiasi, G.; Cui, Y.; Srinivas, A.; Qian, R.; Lin, T.-Y.; Cubuk, E.D.; Le, Q.V.; Zoph, B. Simple copy-paste is a strong data augmentation method for instance segmentation. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 2918–2928.
30. Terven, J.; Córdova-Esparza, D.M.; Romero-González, J.A. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [CrossRef]
31. Jocher, G.; Chaurasia, A.; Qiu, J. Ultralytics YOLO. 2024. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 31 January 2024).
32. Sohan, M.; SaiRam, T.; RamiReddy, V.C. A review on yolov8 and its advancements. In Proceedings of the International Conference on Data Intelligence and Cognitive Informatics, Tirunelveli, India, 18–20 November 2024; pp. 529–45.
33. Wang, A.; Yuan, P.; Wu, H.; Iwahori, Y.; Liu, Y. Improved YOLOv8 for Dangerous Goods Detection in X-ray Security Images. *Electronics* **2024**, *13*, 3238. [CrossRef]
34. Fan, J.; Haji Salam, M.S.B.; Rong, X.; Han, Y.; Yang, J.; Zhang, J. Peach Fruit Thinning Image Detection Based on Improved YOLOv8 and Data Enhancement Techniques. *IEEE Access* **2024**, *12*, 191199–191218. [CrossRef]
35. Zhang, L.; Wu, X.; Liu, Z.; Yu, P.; Yang, M. ESD-YOLOv8: An Efficient Solar Cell Fault Detection Model Based on YOLOv8. *IEEE Access* **2024**, *12*, 138801–138815.
36. Mao, M.; Lee, A.; Hong, M. Efficient Fabric Classification and Object Detection Using YOLOv10. *Electronics* **2024**, *13*, 3840. [CrossRef]
37. AI Hub. Available online: <https://aihub.or.kr> (accessed on 31 January 2024).
38. OpenAI. DALL·E 3 System Card. 2023. Available online: <https://openai.com/research/dall-e-3-system-card> (accessed on 4 April 2024).
39. Microsoft Copilot. Copilot. 2023. Available online: <https://copilot.microsoft.com/> (accessed on 4 April 2024).

40. Tzutalin. LabelImg. 2015. Available online: <https://github.com/HumanSignal/labelImg> (accessed on 4 April 2024).
41. Conrad, R. Copy-Paste-Aug. 2020. Available online: <https://github.com/conradry/copy-paste-aug> (accessed on 5 December 2024).
42. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
43. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
44. Thielen, N.; Rachinger, B.; Schröder, F.; Preitschaft, A.; Meier, S.; Seidel, R. Comparative Study on Different Methods to Generate Synthetic Data for the Classification of THT Solder Joints. In Proceedings of the 1st International Conference on Production Technologies and Systems for E-Mobility (EPTS), Bamberg, Germany, 5–6 June 2024; pp. 1–6.
45. Singh, K.; Navaratnam, T.; Holmer, J.; Schaub-Meyer, S.; Roth, S. Is synthetic data all we need? benchmarking the robustness of models trained with synthetic images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 2505–2515.
46. Paproki, A.; Salvado, O.; Fookes, C. Synthetic data for deep learning in computer vision & medical imaging: A means to reduce data bias. *ACM Comput. Surv.* **2024**, *56*, 271.

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