applied sciences

MDPI

Article

Evaluating GAN-Based Image Augmentation for Threat Detection in Large-Scale Xray Security Images

Joanna Kazzandra Dumagpi * and Yong-Jin Jeong

Department of Electronics and Communications Engineering, Kwangwoon University, Seoul 01897, Korea; yjjeong@kw.ac.kr

* Correspondence: jkmdumagpi@kw.ac.kr

Abstract: The inherent imbalance in the data distribution of X-ray security images is one of the most challenging aspects of computer vision algorithms applied in this domain. Most of the prior studies in this field have ignored this aspect, limiting their application in the practical setting. This paper investigates the effect of employing Generative Adversarial Networks (GAN)-based image augmentation, or image synthesis, in improving the performance of computer vision algorithms on an imbalanced X-ray dataset. We used Deep Convolutional GAN (DCGAN) to generate new X-ray images of threat objects and Cycle-GAN to translate camera images of threat objects to X-ray images. We synthesized new X-ray security images by combining threat objects with background X-ray images, which are used to augment the dataset. Then, we trained various Faster (Region Based Convolutional Neural Network) R-CNN models using different augmentation approaches and evaluated their performance on a large-scale practical X-ray image dataset. Experiment results show that image synthesis is an effective approach to combating the imbalance problem by significantly reducing the false-positive rate (FPR) by up to 15.3%. The FPR is further improved by up to 19.9% by combining image synthesis and conventional image augmentation. Meanwhile, a relatively high true positive rate (TPR) of about 94% was maintained regardless of the augmentation method used.

Keywords: GAN; image augmentation; image translation; threat detection; security; X-ray; deep learning

Citation: Dumagpi, J.K.; Jeong, Y.-J. Evaluating GAN-Based Image Augmentation for Threat Detection in Large-Scale Xray Security Images. *Appl. Sci.* **2021**, *11*, 36. https://dx.doi.org/10.3390/app11010036

Received: 30 November 2020
Accepted: 19 December 2020
Published: 23 December 2020

1. Introduction

X-ray imaging has been an established technology used in various security systems deployed in ports, borders, and certain establishments [1]. This technology offers a mechanism for nondestructive detection of concealed threats through the analysis of X-ray images [2], which is done by trained human inspectors. However, the consistent population growth of travel passengers and the increased shipment of goods across borders demands a corresponding increase in efficiency and accuracy from security systems [3]. Fortunately, computer vision algorithms that were primarily developed for photographic images have been adapted for X-ray security applications [4–7]. Convolutional Neural Network (CNN) is the current state-of-the-art approach to most computer vision problems, which has been through several architectural evolutions [8–10]. Such algorithms provide a means to automate tasks in X-ray security systems, which can be used to provide supplementary aid to human inspectors.

A detection algorithm, when used as an automation tool in security systems, must have a low false-positive rate (FPR), wherein positive means a threat is present in the image, in order to gain higher confidence from human inspectors. Otherwise, human inspectors will generally tend to ignore the predictions. Moreover, a detection algorithm with low FPR is necessary so as not to negatively impact the efficiency of the entire procedure by falsely sending benign baggage to manual inspections. Understandably, a detection algorithm is expected to have a high true-positive rate (TPR) if it is to be used in practical

applications where the safety of all parties involved are of utmost important [11]. CNN-based approaches have been shown to achieve up to 100% TPR and 0% FPR for recognizing the presence of threats on particular X-ray datasets [3]. However, it must be noted that the datasets used in this research, and in most related studies, are not reflective of the practical conditions in X-ray security systems. Unlike the huge datasets for camera images [12–14], prior studies used X-ray image datasets are significantly smaller in size [3,15,16], which may not accurately represent the images found in the practical setting. Specifically, these datasets failed to capture the imbalanced nature of the distribution between positive samples, images that contain at least one threat, and negative samples, images that do not contain any threat, among other characteristics. To address this issue, the first large-scale X-ray security image database was made publicly available and it was shown that CNN-based detection model fails to achieve state-of-the-art performance in the presence of the imbalance problem [17].

An imbalanced dataset is defined as a dataset where some classes severely out represent other classes in terms of population [18]. By this definition, X-ray image datasets are inherently imbalanced datasets since it is far less likely to encounter threat objects compared to normal objects during baggage screening. However, imbalanced data distribution is a major problem for conventional learning algorithms as such algorithms assume equal distribution between classes. During training, the imbalanced distribution biases the conventional algorithm to always predict the majority class, which is a serious problem since the misclassification cost of the minority class is significantly higher. On the other hand, removing the majority class during training biases the conventional algorithm to always predict the minority class, which causes the model to be extremely inefficient to be used in practical applications. A possible approach to the class imbalance problem is image augmentation, which is a group of techniques that enhances the size of an image dataset by creating new images [19]. Such techniques have been used to offset the effect of overfitting, especially when training models for highly-specific tasks where data is naturally scarce [20–22]. Conventional image augmentation techniques, which we refer to from hereafter as image transformation, include rotating, cropping, flipping, shearing, resizing, and applying any transformation to the original image. Recently, techniques based on Generative Adversarial Networks (GAN) [23], which we refer to from hereafter as image synthesis, have been used to augment image datasets by generating completely new and unseen images that are sampled from the training set distribution. In the context of imbalance, image augmentation of the minority class is considered an oversampling approach, wherein the goal is to either balance the dataset or allow the learning algorithm to see the minority class more frequently.

In this paper, we investigate the effects of using GAN-based image augmentation approaches to improve the performance of a threat detection model based on Faster-RCNN [24] on a large-scale and imbalanced X-ray security dataset, also referred to as a practical X-ray dataset. We adapt a Deep Convolutional GAN (DCGAN) model [25] to generate new X-ray images of threat objects and a Cycle-GAN model to translate X-ray images of threat objects from their camera image counterparts. We create new X-ray security images by overlapping the new X-ray images of threat objects to benign X-ray images of baggage scans. These new images are used to augment the training set. The main contributions of this paper are as follows: (a) the use of image augmentation specifically to address the class imbalance problem in a practical X-ray dataset, (b) the development of an image augmentation approach that takes into account the internal distribution of each threat object and their instances, (c) the exhaustive evaluation of image synthesis approaches against image transformation, and (d) the combination of image synthesis and image transformation approaches to enlarge the X-ray security image dataset. Experiment results indicate that image synthesis significantly improves the FPR by up to 15.3% on the dataset that closely resembles a practical X-ray dataset. The FPR is further improved by up to 19.9% on the same dataset by combining image synthesis and image transformation. Conversely, we observed that when the model already achieves a relatively high TPR, using

any image augmentation approach does not provide any further improvements. Overall, the best performance is achieved when image transformation is used simultaneously with any image synthesis approach, which altogether balances a high TPR and a low FPR.

This paper is organized as follows: Section 2 briefly reviews the related works. Section 3 provides a description of the properties of the imbalanced dataset. Section 4 discusses the augmentation approaches used in this study. Section 5 describes the details of the experiments, evaluation criteria, and results. Section 6 presents the conclusions of this study.

2. Related Works

In this section, we review prior studies that attempted to solve the class imbalance problem in large scale X-ray security images and the previous works that used GAN-based image augmentation to enlarge the training set.

A class-imbalance hierarchical refinement approach was proposed in [17] to improve the performance of a classification model on a large-scale and imbalanced dataset. The authors used CNN to approximate a function that is supposed to get rid of elements in the feature map that do not have a strong relationship with the target classes. They implemented this in each level of the convolution operation hierarchy and employed a loss function designed to reduce the noise caused by the negative samples. This approach achieved a 5.65% improvement when used on a ResNet-50 [8] backbone model. In a previous work [26], we adapted an anomaly detection approach to further improve the performance of a model for X-ray threat classification task on the same practical dataset. We trained three models in three separate stages. The first stage focused on maximizing the classification performance of a backbone CNN by training it only on an ideal dataset, i.e., a dataset with a balanced distribution of positive and negative classes. The second stage focused on training a Bidirectional GAN (Bi-GAN) [27] using only the features extracted by the backbone CNN from negative samples. In the final stage, an SVM model [28] is trained to classify whether the features generated by the Bi-GAN model belongs to a positive or a negative sample. These three models worked together at test time to achieve a 6.32% improvement compared to the plain a ResNet-50 [8] classifier.

The study conducted in [29] used an image augmentation approach based on GAN to improve the classification performance of a CNN model on X-ray images of threat objects. They trained a GAN model to generate new X-ray images of threat objects and used the images to augment the training set. Their approach achieved a marginal improvement compared to the model trained on the original training set. This could be attributed to the simplicity of the task, which is just a ten-class image classification problem, where most deep learning algorithms already achieve a very good performance. The work done in [30] also used GAN variants to generate new images of threat objects, which were used to create entirely new X-ray security images. They adapted a Self-Attention GAN (SAGAN) [31] model to generate new images of threat objects and a Cycle-GAN [32] model to translate camera images of threat objects to the X-ray image counterpart. Then, they augmented the X-ray security image training set by combining the generated and translated images of X-ray threats with normal X-ray images from baggage scans. Their method improved the detection performance of a Single-Shot Detector (SSD) [33] model on a seven-class problem by 5.6% compared to the model trained on the original training set. However, it must be noted that in both studies, the datasets used were not only collected in an ideal laboratory setting but are also extremely small and balanced, even after the augmentation. Hence, their approach may still not effectively scale to the real-world X-ray security systems. Furthermore, in the latter study, it was not clear how much contribution can be credited to the new X-ray security images that contain the translated X-ray images of threats, since it only comprised 4.5% of the already small training subset.

Prior studies that address the imbalance problem focused on improving the performance of a classification model. In this paper, we focus on evaluating the performance of detection models on imbalanced datasets since threat detection is a more practical task in

X-ray security systems. Moreover, while our augmentation approach bears similarities to the approach used in [30], we emphasize the differences in model architectures, training setups, and, specifically, the sampling procedure we used in image synthesis, which ensures that we do not introduce any bias in the augmentation process. Furthermore, we highlight that we do not just use image synthesis to augment the dataset, but more importantly, we evaluate its effectiveness as viable a solution to the imbalance problem in practical X-ray datasets.

3. The Imbalanced Dataset

In this section, we discuss the key properties of the imbalanced dataset used in this study and how these properties affect the way we synthesize new X-ray security images. We provide additional observations and insights on this dataset that were not presented in prior studies.

We used the largest publicly available database for X-ray security images known as the SIXray dataset [17]. This dataset is comprised of images collected in real-world security system setups, specifically subway stations. As a result, the collected dataset mirrors the highly imbalanced nature of the distribution between positive and negative samples. Only 8929 positive samples are collected, which is extremely small when compared to more than one million negative samples collected. The dataset is further divided into three subsets, namely SIXray10, SIXray100, and SIXray1000, each representing an imbalance ratio of 10%, 1%, and 0.1% between the positive samples and negative samples, respectively. Among these subsets, SIXray100 is said to be the one that closely mirrors the practical setting. A positive sample contains at least one or a combination of the following identified threats: gun, knife, wrench, pliers, scissors. Hence, a sample can be classified to belong to multiple classes. The distribution of classes in the entire dataset is shown in Figure 1, while the distributions of classes for each of the subsets are shown in Figure 2. [17] also explored the distribution of the samples in terms of aspect ratio and area of the image. They showed that even in these external properties, the distribution is heavily skewed. This could indicate that the size and type of baggage can also determine the presence of threats since both these properties reflect in the aspect ratio and area of the X-ray image scan.

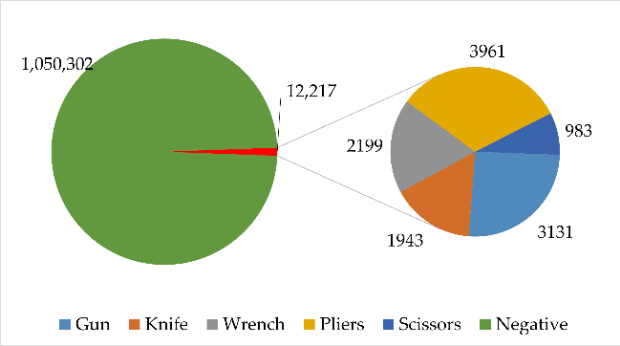


Figure 1. Class distribution of all the images in the SIXray dataset.

Moreover, we provide additional information about the dataset, specifically for the threat detection task. It is important to know the key characteristics of the training set so that we can preserve these properties even after the augmentation to avoid introducing any bias, which was not considered by the prior studies. Figure 3 shows the distribution of the threat count per image. We observed that it is more common to see only one or two threat objects present in one image, which is consistent with the idea that bringing more prohibited items increases the risk of being discovered. However, there are still small probabilities of more threat objects appearing in a single image. On the other hand, Figure 4 shows the distribution of instances per threat. Similarly, regardless of the kind of threat, we can generally expect to see either one or two instances of that threat in a

single image. The positive samples in the training and testing datasets for the SIXray10 and SIXray100 subsets are the same. The only difference between the two subsets is the negative samples in terms of the sample size and the images included. When synthesizing new X-ray security images, we made sure that our sampling procedures are consistent with the data distributions presented in this section.

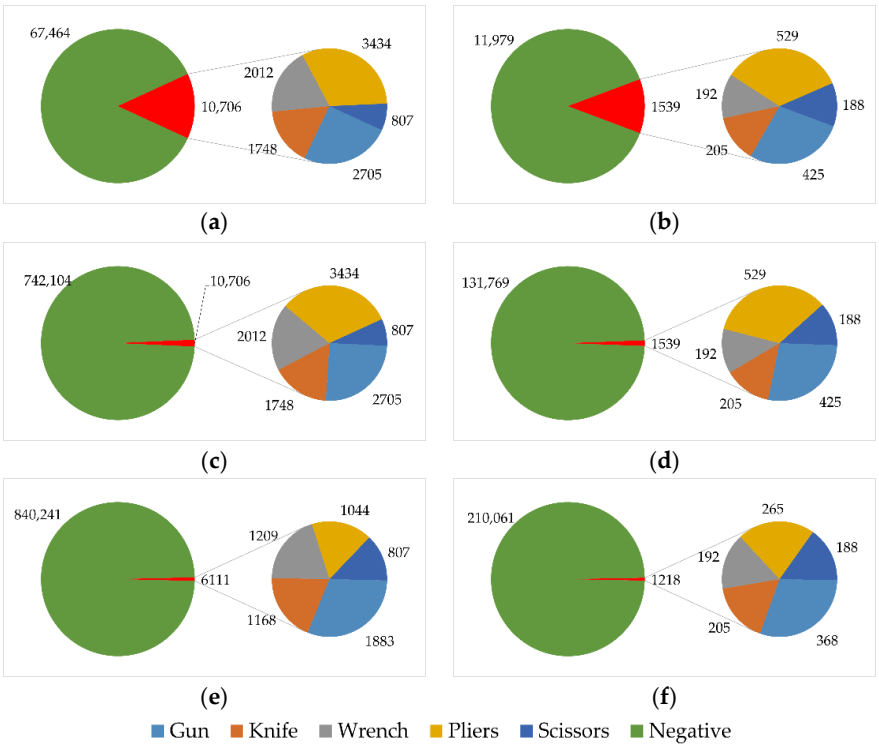


Figure 2. Class distribution of all the images in the following subsets: (**a**) SIXray10 training set. (**b**) SIXray10 test set. (**c**) SIXray100 training set. (**d**) SIXray100 test set. (**e**) SIXray1000 training set. (**f**) SIXray1000 test set.

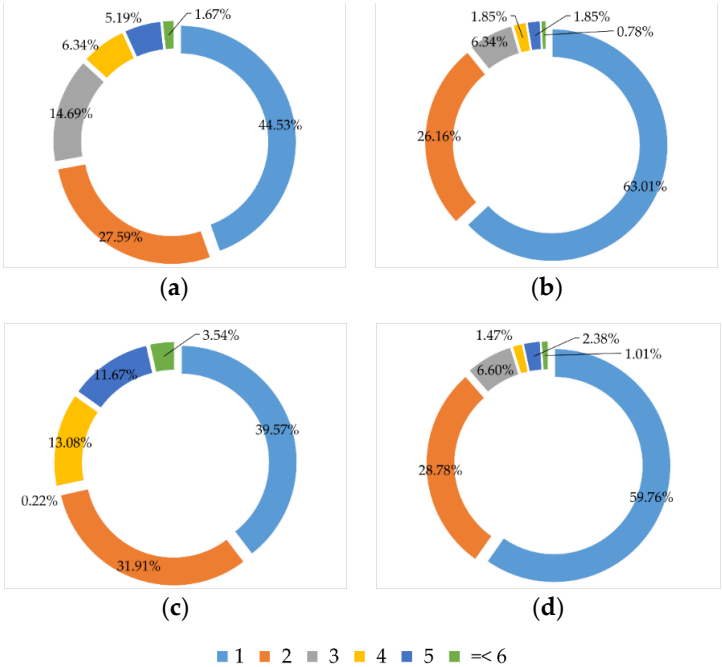


Figure 3. Distribution of threat count per image for (**a**) SIXray10/100 training set, (**b**) SIXray10/100 test set, (**c**) SIXray1000 training set, and (**d**) SIXray1000 test set.

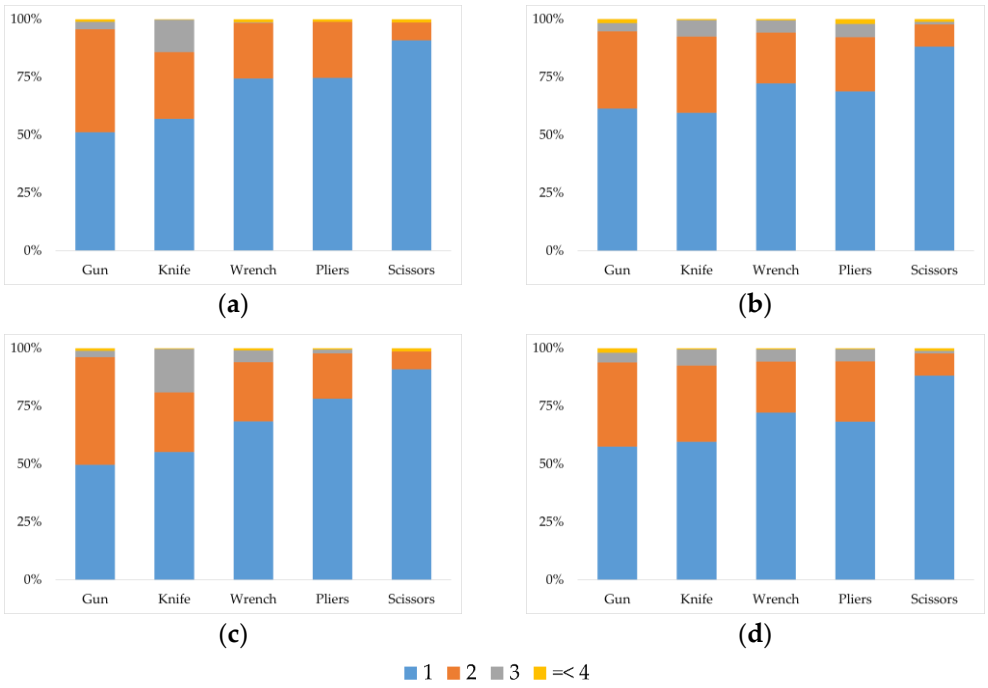


Figure 4. Distribution of instances per threat for (**a**) SIXray10/100 training set, (**b**) SIXray10/100 test set, (**c**) SIXray1000 training set, and (**d**) SIXray1000 test set.

4. Image Augmentation

In this section, we discuss the augmentation approaches used in this study. We group the approaches into two main categories, namely image transformation, and image synthesis, as shown in Figure 5. Image transformation is the conventional approach that creates new images from the positive samples by changing their properties using various manipulation techniques. Image synthesis is the recent approach that creates new images from both samples by extracting threat information from the positive samples and combining it with the information from the negative samples. The following subsections explain how these approaches are used to augment the dataset used to train the threat detection model.

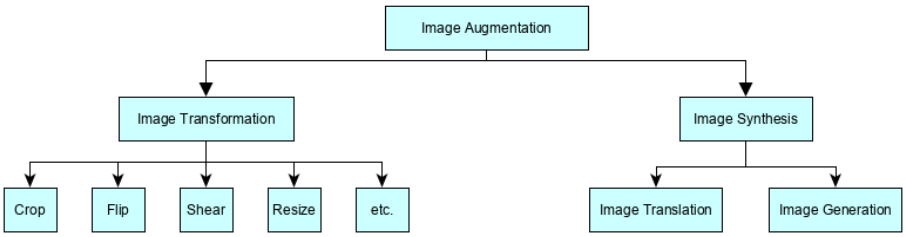
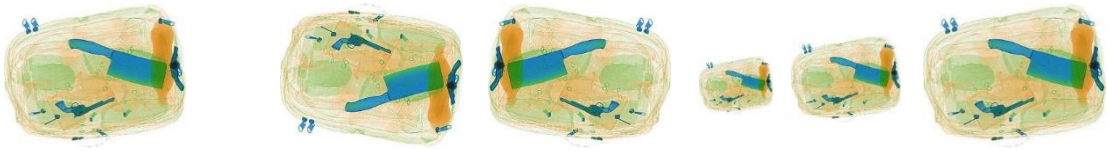


Figure 5. Image augmentation taxonomy.

4.1. Image Transformation

Despite the many possible ways that images can be manipulated, we only chose a few techniques that are relevant to realistic X-ray security system applications. For instance, we did not use color space transformation, adding random noise, or shearing because these would lead to images that are inconsistent in the practical context. X-ray images coming from a single machine should be in the same color space, should be of the same visual quality, and should not show any distortions. Instead, we used horizontal and vertical flipping so that the model can learn to detect threats that come in various orientations, and we used various scaling ratios so that the model can learn to detect threats that come

in various sizes, as shown in Figure 6. For this approach, we can only use images from the positive samples since only these images contain the target objects needed to be learned.



(**a**) (**b**) (**c**)

Figure 6. Image transformation. (**a**) Original image. (**b**) Flipped images. (**c**) Scaled images.

4.2. Image Generation

Currently, it is still not practical to directly use GANs to create X-ray security images since these models generally have a hard time recreating fine details in the images, such as threat objects. Instead, we first train a GAN to produce new X-ray images of threat objects. Then we combine these with images from the negative samples to create an entirely new X-ray security image, i.e., a positive sample. In this section, we describe the first approach used to create new X-ray images of threat, which is called image generation.

First, we sampled 1000 images from the positive samples and manually segmented the threat objects from these images, as shown in Figure 7. We extracted a total of 1783 threat objects, of which 451, 306, 424, 436, and 166 are guns, knives, wrenches, pliers, and scissors, respectively. We adapted a DCGAN as our image generation model and used all the extracted threat objects as the training input. DCGAN is one of the earliest methods that has been effective for generating images because it applies convolution layers, which preserves spatial information. Since then, more sophisticated GAN architectures, that are built on the foundations of DCGAN, have been proposed to generate better quality images. We chose to use DCGAN in this study to get the lower bound of the results. The architecture of our image generation model is shown in Figure 8. The discriminator network, $\boldsymbol{D}$, consists of six convolution blocks, accepts 256 $\times$ 256 images as inputs, and outputs a scalar value, which determines whether the input is determined to be real or fake. The generator network, $\boldsymbol{G}$, consists of six deconvolution blocks, which uses transposed convolutions [34], accepts a 100-dimensional random noise vector, $\boldsymbol{z}$, as input, and outputs a 256 $\times$ 256 X-ray image of a threat object. We used a separate DCGAN model for each threat object. The parameters of the discriminator and generator are updated in an alternating manner using the adversarial objective [23] defined in Equation (1).

$$\mathcal{L}_{adv}(G, D, x, z) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))], \quad (1)$$

wherein x is the input sampled from the dataset of real images. The discriminator tries to correctly classify the input image by minimizing the loss, while the generator tries to generate realistic X-ray images of threat by maximizing the loss. We trained the image generation models for 20,000 iterations with a batch size of 16 using Adam optimizer [35] with a learning rate of 0.0002, β_1 of 0.5 and β_2 of 0.999. Figure 9 shows some samples of the generated X-ray images of threat objects from each class. To evaluate the visual quality of the generated images, we used FID score [36], which calculates the Fréchet distance between two Gaussians fitted to feature representations of a pretrained Inception network. Lower FID scores indicate closer similarity in the characteristics of two datasets in terms of image quality and visual diversity. Table 1 shows the FID scores we got from comparing the generated X-ray images of threat objects to that of the real X-ray images of threat objects, which is about the expected FID score when using DCGANs [30].

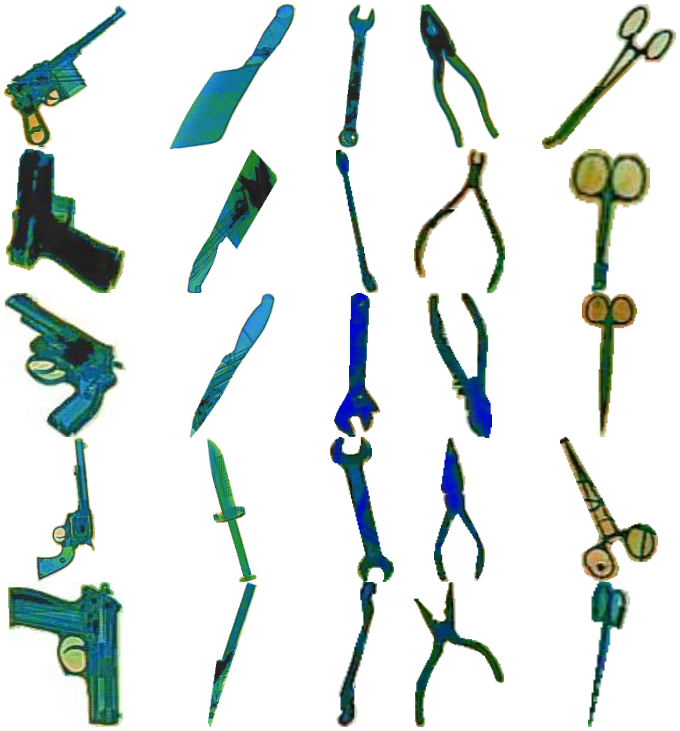


Figure 7. Samples of manually extracted threat objects.

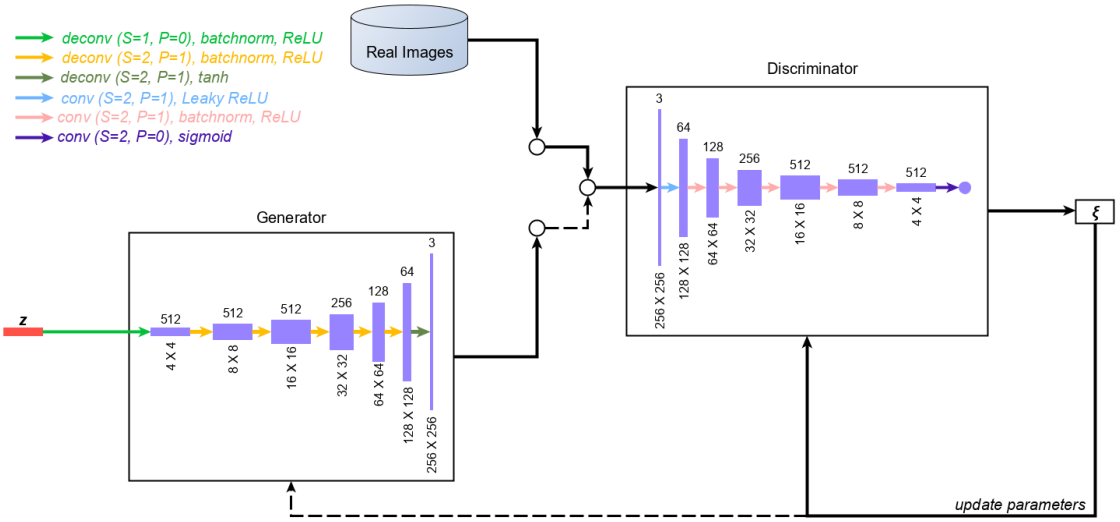


Figure 8. Image generation model. ξ is the cost function, which is the average of the losses per minibatch.

4.3. Image Translation

In this section, we describe our second approach used to create new X-ray security images of threats called image translation. Instead of using the extracted threat objects from the positive samples, we used camera images of threats objects as inputs to the image translation model. The goal of this model is to convert the camera images of threat objects to X-ray images. By doing this, we introduce more intra-class diversity for each threat object since there are far more camera images of threat objects than X-ray images. For instance, we observed that there only 9 types of gun present in the SIXray dataset, but we have collected more than 600 camera images of unique types of gun.

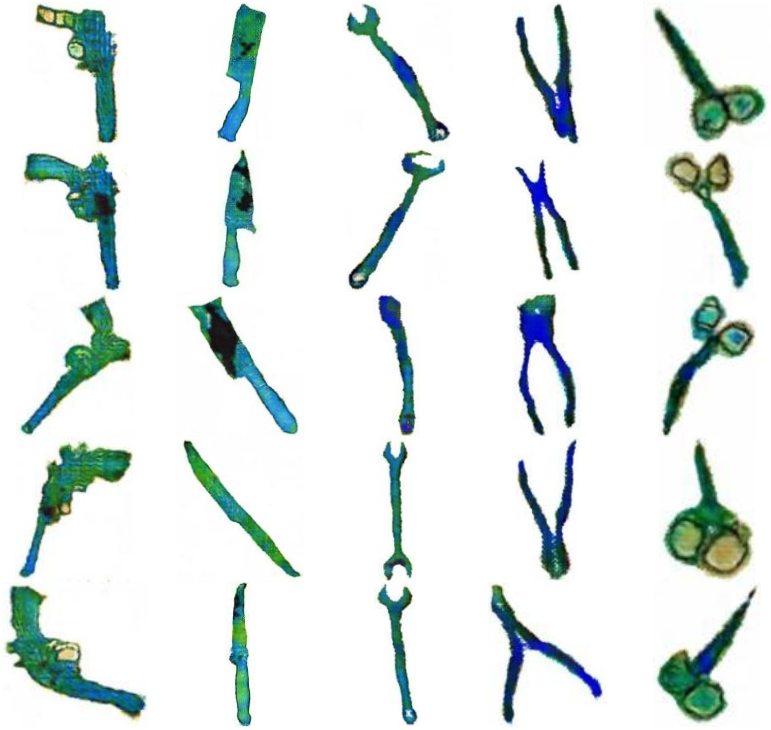


Figure 9. Samples of generated X-ray images of threat objects.

Table 1. FID Scores for each class of the generated threat objects.

Gun	Knife	Wrench	Pliers	Scissors
109	118	103	111	169

First, we scraped the internet for camera images of threats objects on a plain white background, as shown in Figure 10. We collected a total of 1670 camera images of threat objects, of which 662, 353, 234, 366, and 55 are guns, knives, wrenches, pliers, and scissors, respectively. We adapted a Cycle-GAN model as our image translation model. We chose Cycle-GAN for image translation to get the lower bound of the results since it is one earliest works in unpaired image-to-image translation. Moreover, prior translation models require paired training data, which is impossible to collect for this study where the X-ray images have been already collected. Cycle-GAN allows for the usage of camera images of threat objects even though they are not the exact counterpart of the X-ray images of threat objects. Figure 11 shows the architecture of the image translation model, which is made up of two sets of GANs with the same structure. The discriminators, $\boldsymbol{D_A}$ and $\boldsymbol{D_B}$, are each composed of four convolution blocks, accepts 256 $\times$ 256 images as inputs, and outputs a scalar value, which determines whether the input was determined real or fake. The generators, $\boldsymbol{G_A}$ and $\boldsymbol{G_B}$, are encoder-style networks whose inputs and outputs are 256 $\times$ 256 images. Each generator is composed of three convolutional blocks at the input side, nine ResNet blocks [8] in the middle, and three deconvolutional blocks, which uses transposed convolutions, at the output side. Similarly, we used a separate Cycle-GAN model for each threat object. $\boldsymbol{G_A}$ is tasked to translate camera images of threat objects to the equivalent X-ray images, and $\boldsymbol{D_A}$ is tasked to tell whether the input is a real or fake X-ray image of the threat object. $\boldsymbol{G_B}$ is tasked to translate X-ray images of threat objects to the equivalent camera images, and $\boldsymbol{D_B}$ is tasked to tell whether the input is a real or fake camera image of the threat object. Each of these GAN pairs is also trained alternately with the same adversarial loss defined in Equation (1), except that in this case, the input to the generator is no longer a random noise vector but images sampled from another domain. A cycle consistency loss, defined in Equation (2), is added to ensure that the generated

images are strictly the equivalent of the input images and not a random permutation of images in the target domain.

$$\mathcal{L}_{cyc}(G_A, G_B) = \mathbb{E}_{x \sim p_{data}(x)}[\|G_B(G_A(x)) - x\|_1] + \mathbb{E}_{y \sim p_{data}(y)}[\|G_A(G_B(y)) - y\|_1], \quad (2)$$

wherein x is the input sampled from the camera images of threat objects, and y is the input sampled from X-ray images of threat objects. This loss function implies that an image sampled from one domain, when translated to another domain and translated back to its original domain, should be the exact reconstruction of the original image. We trained the image translation model for 100 epochs with a batch size of 1 using Adam optimizer with a learning rate of 0.0002, β_1 of 0.5 and β_2 of 0.999. Figure 12 shows some samples of the generated X-ray images of threat objects from each class. Similarly, we show the FID scores comparing the translated X-ray images of threat objects to that of the real X-ray images of threat objects in Table 2. The FID scores are relatively smaller because translated images often have better quality than generated images since the inputs to the Cycle-GAN are camera images, which generally have significantly higher quality than the X-ray images.

Figure 10. Samples of camera images of threat objects.

Table 2. FID Scores for each class of the translated threat objects.

Gun	Knife	Wrench	Pliers	Scissors
92	106	86	62	130

4.4. Image Synthesis

In this section, we discuss our process of creating entirely new X-ray security images by combining the generated and translated X-ray images of threat objects with the images from the negative samples.

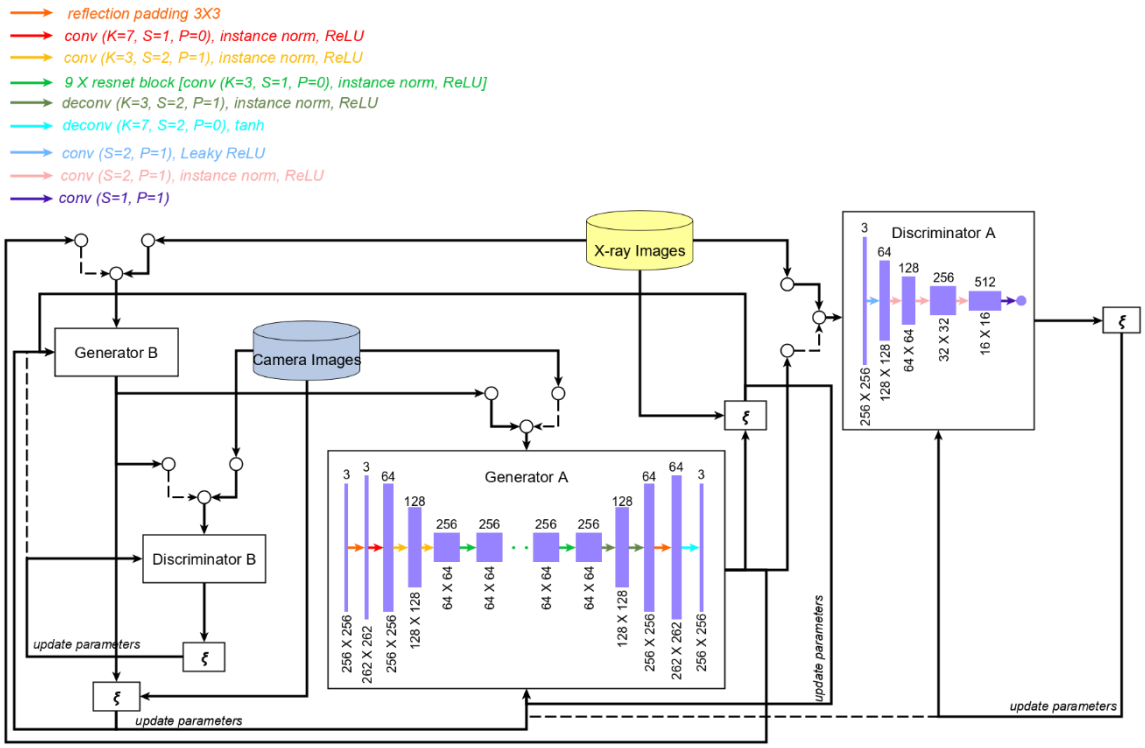


Figure 11. Image translation model. ξ is the cost function, which is the average of the losses per minibatch.

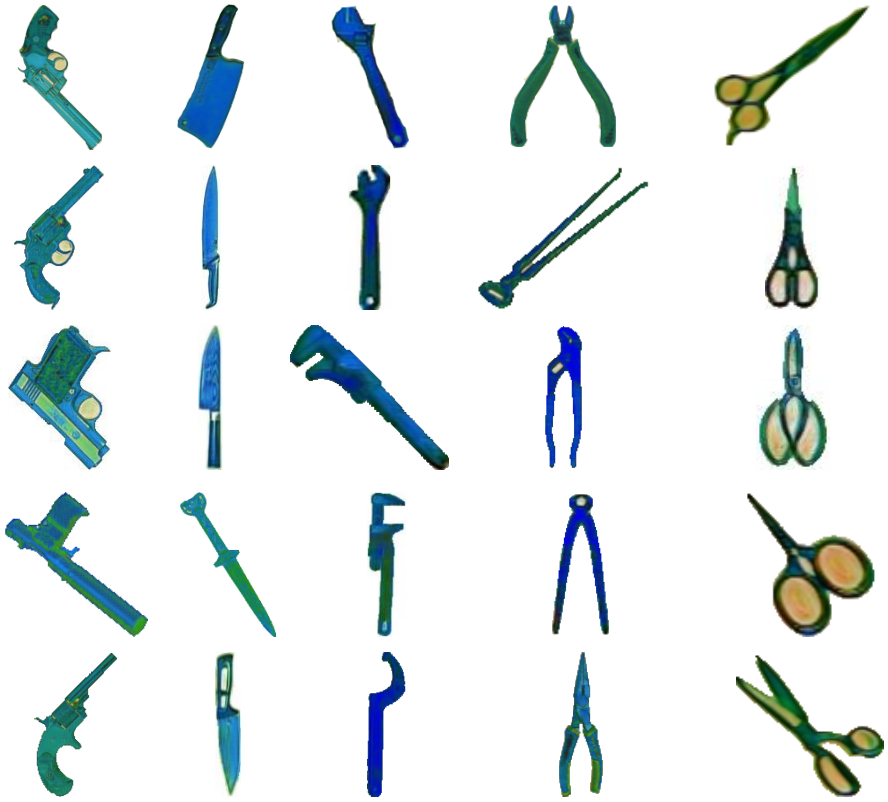


Figure 12. Sample of translated X-ray images of threat objects.

First, we randomly selected 10,000 images from the negative samples of the combined training sets of all subsets, as shown in Figure 13. Then we sampled from the distribution of threat count per image. Next, we selected the type of threat from the class distribution of threats then sampled from the distribution of instances for that threat. If the count of

instances is greater than the count of threats, we use the count of threats as the upper bound. On the other hand, if the count of instances is less than the count of threats, we need to select another type of threat from the class distribution of threats excluding the threats already selected. Then, we need to sample again from the distribution of instances for the newly selected threat until we complete the counts of threats per image. Knowing the exact count and type of threats, we then randomly select images from the dataset of generated or translated X-ray images of threat objects and blend it with an image from the 10,000 negative samples in a similar manner done in [30]. Another distinction in our approach is that we do not just randomly select locations in the image to place the threats, instead the location of the threats is randomly selected from the allowed areas in the image defined by the high-density regions found using the method described in [37]. Without this step, threats might appear outside of the baggage, which is unrealistic. We repeat this process until we have combined all 10,000 images of the negative samples with the generated and translated X-ray images of threat objects, respectively. Figures 14 and 15 show samples of the synthesized images.

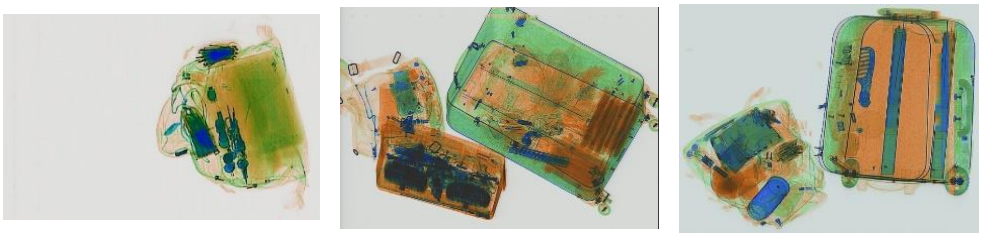


Figure 13. Sample of images from the negative samples.

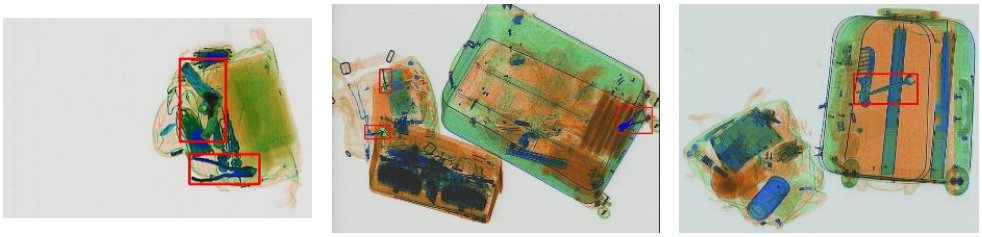


Figure 14. Sample of synthesized images using generated threat objects.

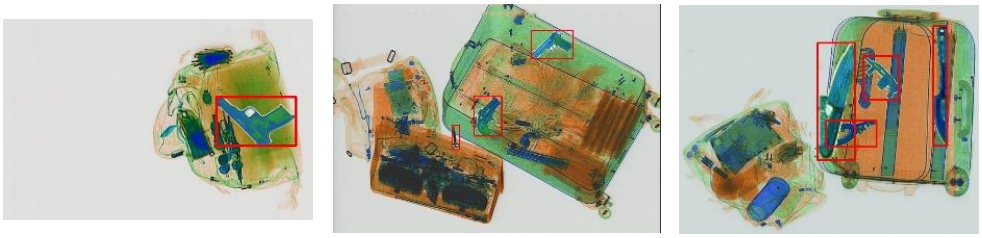


Figure 15. Sample of synthesized images using translated threat objects.

5. Experiment Results and Discussions

In this section, we discuss the metrics used for evaluation, the details of the experiment setup, and the analysis of the results.

5.1. Evaluation Metrics

We evaluate the threat detection model using mean Average Precision (AP) [38] and Log Average Miss Rate (LAMR) [39], defined in Equations (3) and (4), respectively. AP is the area under the curve (AUC) of the precision-recall curve calculated for each class of

threat objects, and the overall performance of the model is represented by the mean across all classes of threat objects.

$$AP = \sum_{n=0} (r_{n+1} - r_n)\, p_{interp}(r_{n+1}) \tag{3}$$

where r is the recall, p_{interp} is the interpolated precision given by $p_{interp}(r_{n+1}) = \max_{\tilde{r} \geq r_{n+1}} p(\tilde{r})$, wherein p is the precision at $\tilde{r}$, and n includes all the recall points. This metric is one of the most popular metrics used to evaluate object detection models. LAMR is the AUC of the miss rate-false positives per image (FPPI) curve calculated for each class of threat objects, and the overall performance of the model is represented by the mean across all classes of threat objects.

$$LAMR = exp^{[\frac{1}{n}\sum_{i=1}^{n} ln(a_i)]} \tag{4}$$

where a is a positive value corresponding to the miss rate at n evenly spaced FPPI points in log-space, between 10^{-2} and 10^{0}. This metric is often used in detection tasks where failure to detect the target must be avoided at high costs.

To draw further comparisons to the original work on the SIXray dataset, we also used the Localization Accuracy (LA) defined in Equation (5). A hit is counted if the intersection-over-union (IoU) of the detected and the ground truth bounding boxes is at least 50%, and a miss is counted, otherwise.

$$LA = \frac{\#\, of\ hits}{\#\, of\ hits + \#\, of\ miss} \tag{5}$$

Finally, we inspect the effect of augmentation at the image level by observing the TPR and FPR defined in Equations (6) and (7), respectively. A detection is considered True Positive (TP) if the model detected at least one threat object in a positive sample, False Positive (FP) if the model detected at least one threat object in a negative sample, False Negative (FN) if the model did not detect any threat object in a positive sample and True Negative (TN) if the model did not detect any threat in a negative sample.

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

$$FPR = \frac{FP}{FP + TN} \tag{7}$$

5.2. Experiment Setup

We adapted a Faster-RCNN [24] detection model for the task of detecting five types of threat objects in X-ray security images. We used ResNet50 [8], with pre-trained weights from ImageNet [12], as the backbone network. We trained the model for 30 epochs with a batch size of 8 using Stochastic Gradient Descent (SGD) with a learning rate of 0.000125 and a momentum of 0.9.

We trained eight detection models, which varies only with the augmentation approach used during training. The baseline approach (Base) used only the positive samples from the training set without any additional augmentation. The image translation approach (Transl) used the positive samples from the training set and 10,000 synthesized images containing the translated X-ray images of threat objects. The image generation approach (Gen) used the positive samples from the training set and 10,000 synthesized images containing the generated X-ray images of threat objects. The image transformation approach (Transf) applies random flipping and resizing to all images in the training set. Different variations of the combinations of these approaches make up the rest of the trained detection models. When augmenting the training set using the combined image synthesis approaches (Gen + Transl), we sampled 5000 synthesized images containing the translated

X-ray images of threat objects and 5000 synthesized images containing the generated X-ray images of threat objects.

5.3. Experiment Results

In this section, we present the results of the experiments, along with our interpretations. At test time, we used all the images in the test set of all SIXray subsets, which includes both positive and negative samples.

Tables 3–5 show the mean AP of the detection model for the SIXray10, SIXray100, and SIXray1000 subsets, respectively. The gun class has relatively higher AP compared to the other classes, likely because these threat objects tend to have more district features compared to the other classes. For instance, a knife or a wrench can easily be mistaken for a harmless metal rod that makes up the framework of the luggage. Additionally, guns make up a large part of the class distribution of threats objects. Consequently, scissors are the least represented class among all the threat objects, and it has the least AP across all subsets. The overall model performance indicates that using either image transformation or any of the image synthesis approaches brought almost the same performance gain across all subsets. Thus, image transformation and image synthesis seem provide about the same quality of additional information to the threat detection model. However, it can also be observed that combining image synthesis and image transformation provides the most performance gain, which implies that the additional information from either approach boosts the performance of the detection model in different way. Furthermore, it appears that the specific approach to image synthesis does not heavily impact performance improvement. The kind of information provided by each approach and how these different information combines to further enhance the overall performance of the threat detection model is revealed by analyzing the results from other metrics. Meanwhile, as the level of imbalance increases, we also achieve increased performance gain with all the approaches, which indicates that augmentation is an effective method that can be used with other methods to combat the class-imbalance problem.

Table 3. Mean Average Precision for SIXray10.

Approach	Gun	Knife	Wrench	Pliers	Scissors	Mean
Base	81	48	48	60	40	55.4
Base + Transf	83	55	53	66	48	61.0
Transl	83	52	49	62	59	61.0
Gen	83	61	49	62	47	60.4
Gen + Transl	84	60	50	64	49	61.4
Transl + Transf	83	64	52	66	60	65.0
Gen + Transf	84	65	57	67	57	66.0
Gen + Transl + Transf	88	65	57	67	59	**67.2**

Table 4. Mean Average Precision for SIXray100.

Approach	Gun	Knife	Wrench	Pliers	Scissors	Mean
Base	67	21	22	31	9	30.0
Base + Transf	74	36	24	35	16	37.0
Transl	78	38	22	31	16	37.0
Gen	77	44	22	32	13	37.6
Gen + Transl	78	41	22	33	15	37.8
Transl + Transf	81	47	27	39	23	43.4
Gen + Transf	82	53	29	39	21	**44.8**
Gen + Transl + Transf	84	49	28	39	21	44.2

Table 5. Mean Average Precision for SIXray1000.

Approach	Gun	Knife	Wrench	Pliers	Scissors	Mean
Base	50	11	9	10	4	16.8
Base + Transf	67	23	14	15	10	25.8
Transl	67	31	13	14	10	27.0
Gen	72	40	14	12	9	29.4
Gen + Transl	70	31	12	12	11	27.2
Transl + Transf	76	36	16	14	17	31.8
Gen + Transf	77	44	17	15	17	**34.0**
Gen + Transl + Transf	79	42	16	15	14	33.2

Tables 6–8 show the LAMR of the detection model for the SIXray10, SIXray100, and SIXray1000 subsets, respectively. For this metric, lower values indicate better performance. We obtain similar observations as we had with the mean AP metric. However, with LAMR, the performance gain we obtain as the level of imbalance increases is different. There is a performance drop from 12.5% to 10% from SIXray10 to SIXray100 and no further improvement from SIXray100 to SIXray1000. The reason for this is that LAMR focuses more on the detection of threat objects in the positive samples. Unlike AP, it is not heavily affected by the imbalance in the dataset caused by the negative samples. Nevertheless, the results show that using both image synthesis and image transformation still gives the best performance in this metric. Thus, making the model more reliable in terms of detecting threat objects when they are present in the image.

Next, we looked at the localization accuracy of the threat detection model trained on the augmented dataset compared to the previous approaches, shown in Tables 9–11 for the SIXray10, SIXray100, and SIXray1000 subsets, respectively. The baseline approach, which is a class activation map (CAM) [40] attached to a classification model, and the proposed approach in [17] were both significantly underperforming when compared to the Faster-RCNN trained on both original and augmented dataset. This is because Faster-RCNN is explicitly tuned to the object detection task, while the first two approaches derived the bounding box predictions indirectly by tracing the gradients of the most discriminant features in the image. This metric only evaluates the detection rate of the model for each instance in the positive sample. Thus, the test set for the SIXray10 and SIXray100 subsets are the same. Since object detection models are only trained on the positive samples, which are also the same images for both SIXray10 and SIXray100 training sets, we observe the same performance from the models in both subsets. In contrast, classification models are heavily affected by the imbalance in both training and testing, as reflected by their performance on this metric, which progressively decreases as the level of imbalance increases. Nevertheless, using both augmentation techniques still proved to give the best performance in this metric on all levels of imbalance.

Table 6. Log Average Miss Rate (%) for SIXray10.

Approach	Gun	Knife	Wrench	Pliers	Scissors	Mean
Base	31	68	71	61	75	61.2
Base + Transf	28	57	68	53	67	54.6
Transl	23	59	71	58	66	55.4
Gen	23	55	71	59	72	56.0
Gen + Transl	21	55	72	56	69	54.6
Transl + Transf	19	47	68	54	59	49.4
Gen + Transf	19	45	64	53	63	48.8
Gen + Transl + Transf	15	49	64	53	62	**48.6**

Table 7. Log Average Miss Rate (%) for SIXray100.

Approach	Gun	Knife	Wrench	Pliers	Scissors	Mean
Base	49	84	85	80	84	76.4
Base + Transf	41	76	83	77	81	71.6
Transl	34	77	85	80	84	72.0
Gen	34	73	85	80	87	71.8
Gen + Transl	34	75	84	79	85	71.4
Transl + Transf	27	71	85	77	82	68.4
Gen + Transf	24	65	83	77	84	66.6
Gen + Transl + Transf	22	67	82	77	82	**66.0**

Table 8. Log Average Miss Rate (%) for SIXray1000.

Approach	Gun	Knife	Wrench	Pliers	Scissors	Mean
Base	64	89	87	91	84	83.0
Base + Transf	49	84	87	88	83	78.2
Transl	51	82	88	90	85	79.2
Gen	44	76	89	91	76	75.2
Gen + Transl	46	83	88	90	86	78.6
Transl + Transf	40	79	87	89	83	75.6
Gen + Transf	34	74	86	89	81	**72.8**
Gen + Transl + Transf	34	74	88	88	84	73.6

Table 9. Localization Accuracy (%) for SIXray10.

Approach	Gun	Knife	Wrench	Pliers	Scissors	Mean
ResNet50 [8]	64	57	50	69	17	51.4
ResNet50+CHR [17]	69	59	54	77	16	54.9
Faster-RCNN [24]	86	72	68	77	79	76.4
Faster-RCNN+AUG [1]	86	74	73	79	78	**78.0**

[1] AUG means a combination of image transformation and image synthesis as image augmentation.

Table 10. Localization Accuracy (%) for SIXray100.

Approach	Gun	Knife	Wrench	Pliers	Scissors	Mean
ResNet50 [8]	48	53	28	40	2	34.1
ResNet50+CHR [17]	58	49	41	50	15	42.7
Faster-RCNN [24]	86	72	68	77	79	76.4
Faster-RCNN+AUG [1]	86	74	73	79	78	**78.0**

[1] AUG means a combination of image transformation and image synthesis as image augmentation.

Table 11. Localization Accuracy (%) for SIXray1000.

Approach	Gun	Knife	Wrench	Pliers	Scissors	Mean
ResNet50 [8]	42	49	2	20	3	26.7
ResNet50+CHR [17]	61	37	22	21	14	31.0
Faster-RCNN [24]	82	63	71	48	80	68.8
Faster-RCNN+AUG [1]	87	62	68	60	79	**71.2**

[1] AUG means a combination of image transformation and image synthesis as image augmentation.

Finally, the image-level performance of the detection model is summarized in Tables 12 and 13. As an automation tool in X-ray security systems, the best approach is the one in which the TPR is high, and the FPR is low. These metrics give us more insight into the individual impact of image transformation and image synthesis in improving the performance of the detection model. Since we cannot calculate the FPR in an object detection task due to not knowing the exact quantity of true negatives, we instead calculate the FPR and TPR at the image level.

Table 12. True Positive Rate (%) across subsets.

Approach	SIXray10/100	SIXray1000
Base	94.5	91.2
Base + Transf	**96.6**	89.3
Transl	94.6	90.7
Gen	93.7	86.9
Gen + Transl	95.0	91.1
Transl + Transf	92.8	**92.2**
Gen + Transf	92.8	91.0
Gen + Transl + Transf	94.4	90.9

Table 13. False Positive Rate (%) across subsets.

Approach	SIXray10	SIXray100	SIXray1000
Base	26.3	25.8	29.8
Base + Transf	26.9	26.6	21.0
Transl	12.9	12.6	9.7
Gen	10.2	10.5	**5.7**
Gen + Transl	12.5	12.5	10.0
Transl + Transf	6.6	6.6	7.8
Gen + Transf	**5.9**	**5.9**	7.7
Gen + Transl + Transf	7.8	7.7	6.8

The TPR is positively influenced by image transformation because this approach focuses on strengthening the model's understanding of the threats that are present in the training set by creating new images of different perspectives that these threats might appear. Since the threats and the context under which they appear, which are learned from the training set, are expected to appear in the test set, the model is better equipped at detecting them causing a performance boost. On the other hand, image synthesis approaches are shown to have little to no effect on the TPR because the information provided by these approaches does not have a strong tie to the threats that are present in the test set. Image synthesis approaches provide information about threat objects that are not present in the test set since the generated and translated threat objects are completely new. Not only are the threat objects different, but the context in which they appear, i.e., the type of baggage and other objects in the background, are also completely new and does not appear in the test set as well. In some cases, image synthesis might even degrade the TPR by overwhelming the detector with information that does not appear in the test set. On the contrary, the strength of image synthesis approaches can be largely attributed to the improvement in the FPR, where image synthesis approaches significantly outperform the image transformation approach. It is likely that the baseline model tries to fit previously unseen objects in the background to any of the threat object classes it has learned. Threat detection models are only trained on images that contains threats. Therefore, they are more likely to associate normal objects from negative samples as threats. Since image transformation only uses the information from the positive samples, it did not prevent the model from assuming previously unseen normal objects are threat objects causing the

FPR to not improve or even worsen. When we used images from negative samples as the background and blended them with the generated or translated X-ray images of threat objects, the model was able to better differentiate the threat objects against several examples of normal objects that usually only appears in the images found from the negative samples causing significant improvement in the FPR. By observing the TPR and FPR, we can further infer that the information provided by any of the image synthesis approaches is the same since all of them brings about the same improvement. Hence, one can choose to use any of the image synthesis approaches and expect to get the same results.

Some variations in the results shown for the SIXray1000 dataset can be attributed to its relatively smaller training set, which is in relation to its unrealistically extreme case of imbalance. For instance, Table 12 demonstrates that the TPR is substantially improved by using image transformation in SIXray10 and SIXray100 subsets. However, in the SIXray1000 subset, image transformation slightly degraded the TPR. The two smaller subsets have about 7500 positive samples in the training set, whereas there are only about 3500 positive samples in the training set for the SIXray1000 subset, which may not have been able to visually represent the types of threats present in the test set. Thus, when image transformation is added during training, the model was limited to detect the kinds of threats that are learnt from the training set but were not represented in the test set, causing the slight degradation in the TPR performance. Moreover, the improvement of the FPR in the SIXray1000, as shown in Table 13, can also be explained by the smaller training set, wherein the model was less confident in associating normal objects to threat objects given its limited knowledge. Regardless, we still observe that image augmentation approaches generally have the same effect to this largest subset.

Image transformation improves the model performance by ensuring the detection of threat objects when they are present in the image, but it does not help the detector to avoid frequently misclassifying normal objects as threats. Image synthesis improves the model performance by significantly reducing the association of normal objects to threat objects when there are none in the image, but it can also negatively affect the ability of the model to detect threat objects if the information from the synthesized images dominates over the information from real images. These are the important trade-offs to keep in mind when using image augmentation in a practical dataset. Consequently, when image transformation and any image synthesis approach are used together (such as in Transl + Transf, Gen + Transf, and Gen + Transl + Transf), we obtain the best overall detection performance by essentially combining the individual strengths of these approaches. The combined image synthesis and image transformation approaches balances a high TPR and low FPR across all subsets, which has been supported by the results we obtained from the previous metrics.

6. Conclusions

In this paper, we evaluated the impact of GAN-based image augmentation, also called image synthesis, in the detection performance on a practical X-ray security image dataset. We described the details of our image synthesis approach, which takes into consideration the internal distribution of threats in the existing training set such that we do not introduce bias during the learning phase.

We found that image synthesis significantly improves the FPR by up to 15.3% on the SIXray100 subset, which closely resembles a practical X-ray security dataset. This performance gain was due to the model gaining more information about the differences between normal and threat objects from the synthesized images. The FPR is further improved by up to 19.9% on the SIXray100 subset by combining image synthesis with image transformation, which enlarged the dataset even more and generated new perspectives of both the normal and threat objects. On the other hand, we observed that using any image augmentation approach does not significantly impact the TPR, especially when the model already achieves relatively high TPR. Overall, the best performance is achieved when image transformation is used simultaneously with any image synthesis approach.

In conclusion, GAN-based image augmentation is an effective approach to combat the class-imbalance problem. When used as a supplementary automation tool, the significant reduction of the FPR enables human screeners to trust more on the predictions of the detection algorithm, which in turn increases the efficiency of the security procedures. Inadvertently, this improvement did not come at the expense of the TPR, which remained relatively high, making threat detection models more suitable in practical applications. The potential benefits can be further maximized by using GAN-based augmentation together with other approaches.

Future research in this field includes developing an automatic and more realistic threat image projection models that also considers the differences in the orientation, color scheme, and other properties of threat objects with respect to the background, which could significantly improve not only the FPR but also TPR; hence, the overall performance of a detection model.

Author Contributions: Conceptualization, J.K.D. and Y.-J.J.; methodology, J.K.D. and Y.-J.J.; software, J.K.D.; validation, Y.-J.J.; formal analysis, J.K.D. and Y.-J.J.; investigation, J.K.D. and Y.-J.J.; resources, J.K.D.; data curation, J.K.D.; writing—original draft preparation, J.K.D.; writing—review and editing, J.K.D. and Y.-J.J.; visualization, J.K.D.; supervision, Y.-J.J.; project administration, Y.-J.J.; funding acquisition, Y.-J.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Kwangwoon University and by the MISP Korea, under the National Program for Excellence in SW (2017-0-00096) supervised by IITP.

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

References

1. Zentai, G. X-ray imaging for homeland security. *IEEE Int. Workshop Imaging Syst. Tech.* **2008**, *3*, 13. [CrossRef]
2. Mery, D. *Computer Vision for X-ray Testing*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 1–12.
3. Akcay, S.; Kundegorski, M.E.; Willcocks, C.G.; Breckon, T.P. Using Deep Convolutional Neural Network Architectures for Object Classification and Detection Within X-ray Baggage Security Imagery. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2203–2215. [CrossRef]
4. Mery, D.; Svec, E.; Arias, M. Object Recognition in X-ray Testing Using Adaptive Sparse Representations. *J. Nondestruct. Eval.* **2016**, *35*, 1–9. [CrossRef]
5. Mery, D.; Riffo, V. Automated Detection of Threat Objects Using Adapted Implicit Shape Model. *IEEE Trans. Syst. ManCybern. Syst.* **2016**, *46*, 472–482.
6. Akcay, S.; Kundegorski, M.E.; Devereux, M.; Breckon, T.P. Transfer Learning Using Convolutional Neural Networks for Object Classification within X-ray Baggage Imagery. In Proceedings of the IEEE International Conference on Image Processing, Phoenix, AZ, USA, 25–28 September 2016.
7. Gaus, Y.F.A.; Bhowmik, N.; Akcay, S.; Breckon, T. Evaluating the Transferability and Adversarial Discrimination of Convolutional Neural Networks for Threat Object Detection and Classification within X-ray. In Proceedings of the IEEE International Conference on Machine Learning and Applications, Boca Raton, FL, USA, 16–19 December 2019.
8. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference for Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
9. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
10. Huang, G.; Liu, Z.; von der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
11. Hättenschwiler, N.; Sterchi, Y.; Mendes, M.; Schwaninger, A. Automation in airport security X-ray screening of cabin baggage: Examining benefits and possible implementations of automated explosives detection. *Appl. Ergon.* **2018**, *72*, 58–68. [CrossRef]
12. Deng, J.; Dong, W.; Socher, R.; Li, J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale heirarchal image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009.
13. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Doll´ar, 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; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer: Cham, Switzerland, 2014.
14. Kuznetsova, A.; Rom, H.; Alldrin, N.; Uijlings, J.; Krasin, I.; Pont-Tuset, J.; Kamali, S.; Popov, S.; Malloci, M.; Kolesnikov, A.; et al. The Open Images Dataset V4. *Int. J. Comput. Vis.* **2020**, *128*, 1956–1981. [CrossRef]
15. Mery, D.; Riffo, 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*, 1–12. [CrossRef]

16. Caldwell, M.; Griffin, L.D. Limits on transfer learning from photographic image data to X-ray threat detection. *J. X-ray Sci. Technol.* **2019**, *27*, 1007–1020. [CrossRef]
17. 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 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
18. He, H.; Garcia, E.A. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284.
19. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
20. Dawar, N.; Ostadabbas, S.; Kehtarnavaz, N. Data Augmentation in Deep Learning-Based Fusion of Depth and Inertial Sensing for Action Recognition. *IEEE Sens. Lett.* **2019**, *3*, 1–4. [CrossRef]
21. Pham, T.D. Geostatistical Simulation of Medical Images for Data Augmentation in Deep Learning. *IEEE Access* **2019**, *7*, 68752–68763. [CrossRef]
22. Tang, Z.; Chen, K.; Pan, M.; Wang, M.; Song, Z. An Augmentation Strategy for Medical Image Processing Based on Statistical Shape Model and 3D Thin Plate Spline for Deep Learning. *IEEE Access* **2019**, *7*, 133111–133121. [CrossRef]
23. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. *arXiv* **2014**, arXiv:1406.2661v1.
24. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **2015**, arXiv:1506.01497. [CrossRef] [PubMed]
25. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv* **2015**, arXiv:1511.06434.
26. Dumagpi, J.K.; Jung, W.-Y.; Jeong, Y.-J. A New GAN-Based Anomaly Detection (GBAD) Approach for Multi-Threat Object Classification on Large-Scale X-ray Security Images. *IEICE Trans. Inf. Syst.* **2020**, 454–458. [CrossRef]
27. Donahue, J.; Krähenbühl, P.; Darrell, T. Adversarial Feature Learning. *arXiv* **2017**, arXiv:1605.09782v7.
28. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
29. Yang, J.; Zhao, Z.; Zhang, H.; Shi, Y. Data Augmentation for X-ray Prohibited Item Images Using Generative Adversarial Networks. *IEEE Access* **2019**, *7*, 28894–28902. [CrossRef]
30. 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. [CrossRef]
31. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-Attention Generative Adversarial Networks. *arXiv* **2018**, arXiv:1805.08318v2.
32. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *arXiv* **2017**, arXiv:1703.10593v7.
33. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. *arXiv* **2015**, arXiv:1512.02325.
34. Dumoulin, V.; Visin, F. A guide to convolution arithmetic for deep learning. *arXiv* **2016**, arXiv:1603.07285.
35. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
36. Lucic, M.; Kurach, K.; Michalski, M.; Gelly, S.; Bousquet, O. Are GANs Created Equal? A Large-Scale Study. *arXiv* **2017**, arXiv:1711.10337.
37. Dumagpi, J.K.; Jung, W.-Y.; Jeong, Y.-J. KNN-Based Automatic Cropping for Improved Threat Object Recognition in X-ray Security Images. *J. IKEEE* **2019**, *23*, 1134–1139.
38. Everingham, M.; Gool, L.V.; Williams, C.K.I.; Winn, J.; Zisserman, A. The PASCAL Visual Object Classification (VOC) Challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
39. Dolla'r, P.; Wojek, C.; Schiele, B.; Perona, P. Pedestrian Detection: An Evaluation of the State of the Art. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 743–761. [CrossRef] [PubMed]
40. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning Deep Features for Discriminative Localization. *arXiv* **2015**, arXiv:1512.04150.