IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 17, 2021, accepted September 23, 2021, date of publication September 29, 2021,
date of current version October 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3116255

Learning-Based Image Synthesis for Hazardous Object Detection in X-Ray Security Applications

HYO-YOUNG KIM, SUNG-JIN CHO, SEUNG-JIN BAEK, (Member, IEEE), SEUNG-WON JUNG, (Senior Member, IEEE), AND SUNG-JEA KO, (Fellow, IEEE)
School of Electrical Engineering, Korea University, Seoul 02841, South Korea

Corresponding author: Seung-Jin Baek (sjinbaek@korea.ac.kr)

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.2019-0-00268, Development of SW Technology for Recognition, Judgment and Path Control Algorithm Verification Simulation and Dataset Generation). This work was supported by a Korea University Grant.

ABSTRACT X-ray baggage inspection has been widely used for maintaining airport and transportation security. Towards automated inspection, recent deep learning-based methods have attempted to detect hazardous objects directly from X-ray images. Since it is challenging to collect a large number of training images from real-world environments, most previous learning-based methods rely on image synthesis for training data generation. However, these methods randomly combine foreground and background images, restricting the effectiveness of synthetic images for object detection. To solve this problem, in this paper, we propose a learning-based X-ray image synthesis method for object detection. Specifically, for each foreground object to be synthesized, we first estimate positions difficult to detect by the object detector. These positions and their corresponding confidence values are then used to construct a difficulty map, which is used for sampling the target foreground position for image synthesis. The performance analysis using various state-of-the-art object detectors shows that the proposed synthesis method can produce more useful training data compared with the conventional random synthesis method.

INDEX TERMS Deep learning, neural network, object detection, X-ray, inspection.

I. INTRODUCTION

Baggage inspection based on X-ray screening is an essential task for reducing the risk of crime and terrorist attacks and preventing the propagation of pests and diseases [1]. In general, the X-ray images are visually inspected by trained human inspectors to detect dangerous objects. Although it may take less than a second to investigate each piece of baggage, each inspector has to check a large amount of baggage over a long time. The possibility of human error is thus non-negligible, even with specialized training. Therefore, an automated X-ray baggage inspection system based on computer vision techniques, such as feature-based detection methods [2]–[5], is needed to detect hazardous objects robustly.

Recently, motivated by the remarkable success of convolutional neural networks (CNNs) in solving computer vision problems, learning-based automated X-ray inspection methods have been proposed [6]–[9]. To ensure performance in such learning-based approaches, a large dataset of X-ray images and their corresponding annotations is essentially required. Several publicly available datasets can be used for object detection in X-ray images. Mery *et al.* [3] presented the GDX-ray database that contains five object categories: castings, welds, baggage, natural objects, and settings. Miao *et al.* [10] introduced a larger size dataset called SIXray, which contains diverse types of hazardous objects in baggage with cluttered background items. However, the number of positive samples, *i.e.*, images with hazardous objects, is much less than the number of negative samples (12,277 versus 1,050,302 samples). Although this class imbalance may reflect the real-world application environments, it also makes network training difficult.

To overcome such lack of training images, many recent methods paid attention to learning from synthetic data [8], [9], [11]–[13]. Considering that X-ray imaging can be modeled using the absorption law that characterizes the intensity distribution of X-rays through matter [14], Mery and Katsaggelos [15] introduced a solid mathematical model for the synthesis of threat objects to the background baggage.

The associate editor coordinating the review of this manuscript and approving it for publication was Mauro Gaggero.

Following this method, Jain *et al.* [11] synthesized X-ray images during training for data augmentation and demonstrated the effectiveness of the synthetic data using several standard object detection models, such as YOLOv2 [16] and Faster R-CNN [17]. Yang *et al.* [12] further introduced a generative adversarial network (GAN)-based approach for generating realistic hazardous objects. Zhu *et al.* [13] applied a similar data augmentation framework to Jain's method [11] and demonstrated that the accuracy of the SSD model [18] can be increased by 5.6% in terms of the mean average precision (mAP) when the model is trained using the augmented dataset. Saavedra *et al.* [9] combined PGGAN [19] and the X-ray image synthesis technique [15].

The X-ray and natural images show a clear difference when multiple objects are overlapped with each other. As shown in Figs. 1(a) and (b), because X-ray is penetrable, both front and rear objects are visible in X-ray images [10]. However, Figs. 1(c) and (d) show that the occluded regions of the rear objects are generally not visible in natural images. Due to this difference, the X-ray images have an advantage in that the target object may be synthesized at any desired position, not at a limited or random location. However, the existing X-ray image synthesis methods that overlay foreground objects at arbitrary locations regardless of the background content cannot fully take advantage of synthesized images for object detection. To this end, in this paper, we propose a novel learning-based X-ray image synthesis method.

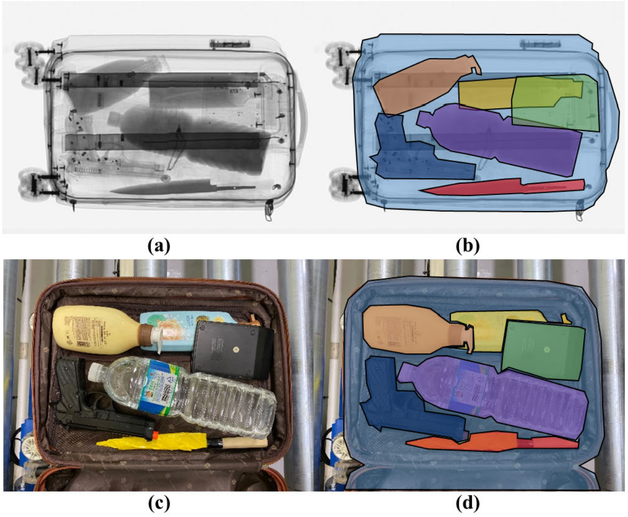


FIGURE 1. Sample images of (a) X-ray image and (b) its annotated segmented image (ground-truth), and sample images of (c) natural image and (d) its annotated segmented image (ground-truth).

In our proposed method, an object detection network is first trained using the X-ray images synthesized with hazardous objects at random positions. The hazardous objects are then synthesized at hard-to-detect locations estimated by the object detector during the learning process. By this simple but effective way, we can generate hard samples that can contribute to further boost the object detection performance. The experimental results obtained by various detection networks demonstrate the superiority of the proposed synthesis method.

The rest of this paper is organized as follows. The related works are reviewed in Section II. The proposed method is detailed in Section III. Experimental results are provided in Section IV. Finally, our conclusion is given in Section V.

In summary, this paper presents two major contributions. (i) We propose a difficulty map that represents the locations at which the detector is difficult to find the objects without any additional network. (ii) Using the difficulty map, we introduce a data synthesis technique that produces hard-to-detect samples to train the detector effectively.

II. RELATED WORK

Before the explanation of the proposed method, in this section, we briefly review its related techniques including CNN-based object detection, X-Ray computer vision algorithms, X-Ray image synthesis, and data augmentation.

A. CNN-BASED OBJECT DETECTION

CNN-based methods have been very successful in the recognition and localization of objects. According to the design principle, these methods can be classified into two-stage methods [17], [20] and single-stage methods [18], [21], [22].

The two-stage methods first identify candidate bounding boxes using a deep network and then refine the candidates using another sub-network. To this end, Ren *et al.* [17] introduced the region proposal network (RPN), which performs efficiently by sharing full-image convolutional features with a subsequent detection network. Lin *et al.* [20] proposed a feature pyramid network (FPN) which combines low-resolution and high-resolution features via top-down paths and lateral connections. This feature pyramid contains rich semantics from all levels and can be built from a single-scale input image, thereby exhibiting effectiveness in terms of representational power, speed, and memory.

The single-stage methods detect objects via a single network inference. As a pioneering work, YOLO [21] used a unified detection network that predicts bounding boxes and classifies objects at the same time from an entire image. The computational efficiency and robustness of YOLO and its advanced versions [16], [23] have been demonstrated thoroughly. Liu *et al.* [18] designed a reduced VGG network architecture that extracts features from multi-layers, enabling the network to handle objects with various scales effectively. Lin *et al.* [22] adopted ResNet as a basic feature extractor and used a focal loss to address the class imbalance problem caused by the biased foreground-background ratio.

B. X-RAY COMPUTER VISION ALGORITHMS

In the area of baggage inspection, some computer vision algorithms based on a single view of a single energy have been reported. Riffo and Mery [2] proposed automated detection algorithm based on visual codebooks. Mery *et al.* [3] used adaptive sparse representations [24] to detect objects, with less constrained conditions including some contrast variability, pose, intra-class variability, size of the image and focal distance. On the other hand, in the analysis of single dual-energy images, Baştan *et al.* [4] presented a bag of visual words (BoVW) model with several hand-crafted feature representations. Additionally, there are some methods based on a single energy multi-view, using active vision [25], [26]. Support vector machine (SVM) classifiers and visual dictionaries are proposed in dual-energy multi-views X-ray [5], [27].

Recently, several methods based on deep convolutional neural networks have been proposed. Akçay *et al.* [28] suggested CNN-based object classification method using transfer learning in order to overcome the limited amount of training data, and provided performance comparison among CNN-based object detection algorithms for X-ray baggage security imagery [6]. Gu *et al.* [8] proposed automatic X-ray object detection using feature enhancement module. Saavedra *et al.* [9] introduced GAN strategy in data augmentation for the threat object detection.

C. X-RAY IMAGE SYNTHESIS

Many studies assume that X-ray image formation obeys the Beer-Lambert law. Based on this assumption, at image location (x, y), the pixel intensity of the X-ray image $I(x, y)$ is defined as

$$I(x, y) = I_0 \exp\left(-\int \mu(x, y, z)\, dz\right), \tag{1}$$

where I_0 is the beam intensity, z represents the depth coordinate, and μ is the effective attenuation coefficient of the objects in the scene [29].

Based on this image formation model, Rogers *et al.* [30] introduced a data synthesis technique, called TIP, which generates synthesized threat images that have no significant differences compared with real threat images. More specifically, they synthesize images by multiplying the foreground mask $F(x, y)$ and background mask $B(x, y)$ as follows:

$$I(x, y) = I_0 F(x, y) B(x, y), \tag{2}$$

where

$$\begin{aligned} F(x, y) &= \exp\left(-\int \mu_F(x, y, z) dz\right), \\ B(x, y) &= \exp\left(-\int \mu_B(x, y, z) dz\right). \end{aligned} \tag{3}$$

μ_F and μ_B represent the effective attenuation coefficients of the foreground and background masks, respectively. It is worth noting that when N foreground masks are overlapped in the image, $F(x, y)$ in (2) can be replaced with $\prod_{i=1}^{N} F^i(x, y)$, where F^i indicates the i-th foreground mask, as shown for $N = 3$ in Fig. 2.

D. DATA AUGMENTATION

To increase generalization performance and attenuate overfitting problem simultaneously, functional solutions such as dropout regularization [31], batch normalization [32], and transfer learning [33] have been developed. In contrast to such techniques, data augmentation approaches focus on training datasets, which is the root cause of the overfitting problem.

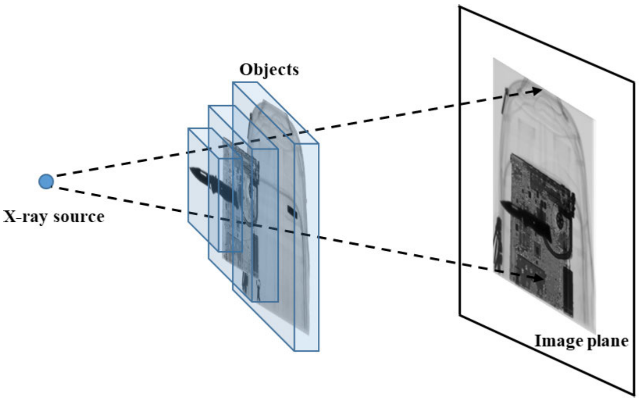


FIGURE 2. Example of the irradiation of the three different objects.

In general, the data augmentation is conducted by simple transformations such as horizontal flipping, color space augmentations, and random cropping [34]. Moreno-Barea *et al.* [35] proposed noise injection as an additional data augmentation, demonstrating that adding noise to images for nine datasets in UCI repository could help CNN learn more robust features. Kang *et al.* [36] devised PatchShuffle Regularization (PSR), which is a kernel filter that randomly swaps pixel values in $n \times n$ sliding windows. Experiments on different filter sizes and probabilities of shuffling the pixels at each step, the authors demonstrated the effectiveness of PSR by achieving a 5.66% error rate on CIFAR-10 compared with an error rate of 6.33%. Inspired by the dropout regularization mechanism, Zhong *et al.* [37] developed a random erasing method that performs dropout in the input data space rather than in the feature space to prevent overfitting problems effectively.

As described above, although the data augmentation can be applied to images in the input space, it can also be applied to feature space. Konno and Iwazume manipulated the modularity of neural networks after training, improving the performance on CIFAR-100 from 66% to 73% accuracy. Xie *et al.* [38] presented DisturbLabel (DL), which is an adversarial training technique that randomly replaces labels at each iteration. On the MNIST dataset with LeNet CNN architecture, DL produced 0.32% error rate compared with a baseline error rate of 0.39%.

The first GAN architecture proposed by Ian Goodfellow [31] is a framework for generative modeling through adversarial training. Such a network architecture can be applied to data augmentation tasks by generating new training data that results in better-performing classification models. Researches to apply GAN to data augmentation and report the resulting classification performance have been conducted in the field of biomedical image analysis [39]. Maayan *et al.* [40] tested the effectiveness of generating liver lesion medical images using DCGAN. On top of classical

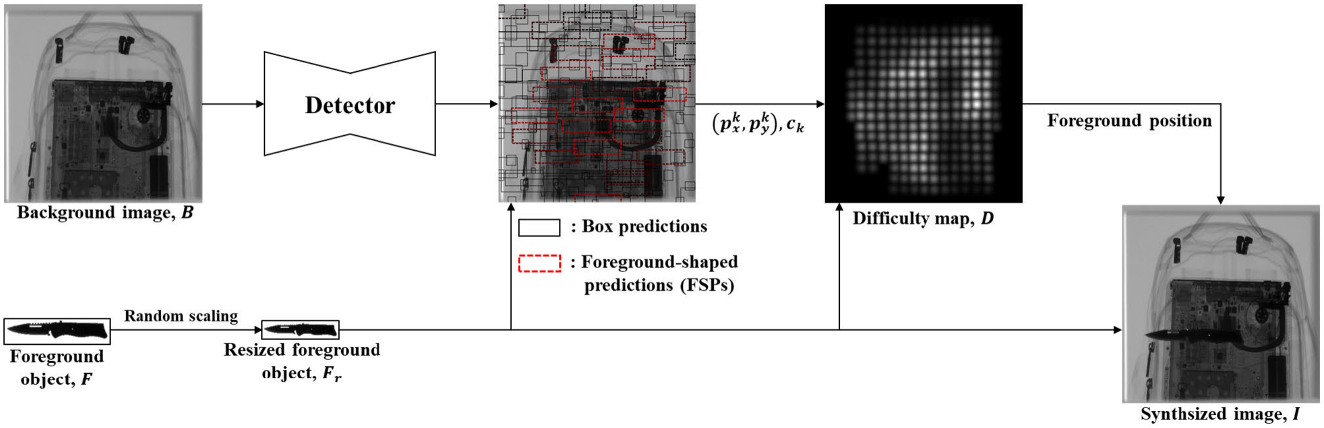


FIGURE 3. The flowchart of the proposed difficulty map-based X-ray image synthesis. The intensity of box predictions is proportional to the confidence of the prediction.

augmentations to attain 78.6% sensitivity and 88.4% specificity, the authors employed additional DCGAN-generated samples, finally achieving the performance of 85.7% sensitivity and 92.4% specificity. In the literature of X-ray security inspection, Yang *et al.* [12] proposed a GAN-based data augmentation method to generate the images of prohibited items, and Zhu *et al.* [13] improved SAGAN [41] to generate the realistic prohibited item images.

III. PROPOSED METHOD

Fig. 3 illustrates the proposed X-ray image synthesis framework. In this section, we first define a difficulty map and describe how the difficulty map is used to sample target foreground positions for the generation of hard training samples.

A. DIFFICULTY MAP EXTRACTION

Regardless of the difference between single-stage and two-stage approaches, most deep learning-based object detection networks produce locations of objects and their corresponding confidence values [18], [20], [22]. Therefore, if we feed the background image to an object detector, we can obtain foreground positions that can confuse the detector when evaluated after the image synthesis. We thus attempt to use this degree of confusion as valuable information for determining the target position of foreground objects.

We first feed the background image to the object detector and obtain the box predictions with confidence estimates. Note that the detection network outputs the position of the box predictions, whether there is a target object in the image or not. Let (p_x^k, p_y^k) and c_k denote the center position and its corresponding confidence value of the k-th box prediction, respectively. Given the randomly scaled foreground object we want to synthesize, we collect the boxes with 50% or higher intersection over union (IoU) among multiple box candidates. We use these remained boxes, referred to as foreground-shaped predictions (FSPs), and their confidence estimates to define our difficulty map.

Let D denote the difficulty map, which is defined as follows:

$$D(x, y) = \sum_k c_k \cdot exp\left(-\frac{(x - p_x^k)^2 + (y - p_y^k)^2}{2\sigma_k^2}\right), \quad (4)$$

where x and y represent the pixel coordinates and σ_k is the standard deviation of the Gaussian function. To avoid having difficulty values very close to zero, we set σ_k as the center distance between the k-th FSP and its closest FSP. In this manner, difficulty values slowly decay between distant FSPs, which is advantageous for our probabilistic sampling of foreground positions. As illustrated in Fig. 4, the difference between the object detection process and the difficulty map extraction process is that object detection uses box predictions from higher scores and non-maximum suppression to sort out final results, whereas difficulty map extraction uses all box predictions from FSPs without non-maximum suppression. However, the difficulty map can be obtained using the detector network the same as that used in the detection process. Therefore, there are no additional network and loss functions to extract the difficulty map.

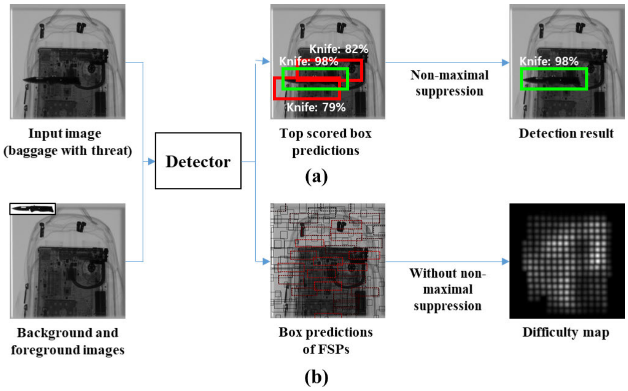


FIGURE 4. Comparison of (a) object detection process and (b) difficulty map extraction process, using the same detector.

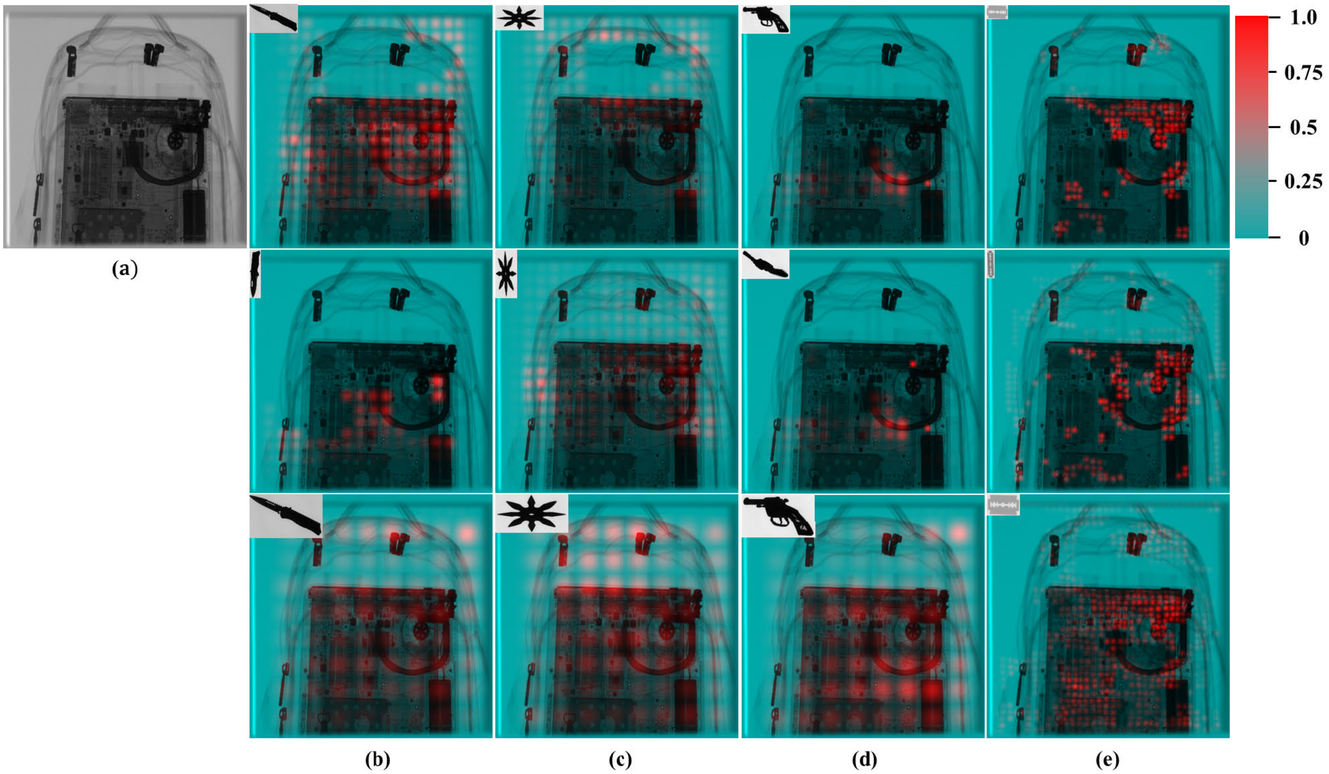


FIGURE 5. (a) The background image and its difficulty maps corresponding to (b) knife, (c) shuriken, (d) razor, and (e) gun using SSD trained with the proposed method. Each image in the top and middle rows represents the differences in the difficulty map depending on the size of the object. Each image in the bottom row shows how the difficulty map varies depending on the different angles of the object. The difficulty maps are normalized in the range [0], [1] and overlapped with the background image for better visualization.

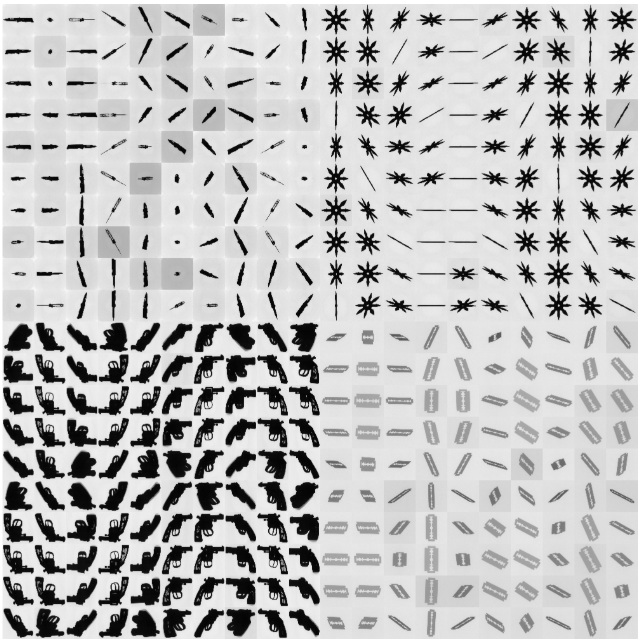


FIGURE 6. Foreground image samples from the GDX-ray database.

Fig. 5 shows the difficulty maps for four foreground objects, shown in Fig. 6, which are the results of our case study for hazardous object detection. Note that Fig. 5 is a difficulty map extracted using SSD as an example, and the difficulty map differs if the detection network changes. In the first and second rows of Figs. 5(b), (c), and (e), it can be seen that different angles of objects with the same size produce the changes on the difficulty map. When the objects have both the same size and aspect ratio with different angles, considerably similar difficulty maps were generated, as shown in the first and second rows of Fig. 5(d). Moreover, the first and last rows in Fig. 5 represent the changes of difficulty map according to the size of the object. From the result that values of difficulty maps for larger objects have relatively uniform distributions, it is confirmed that the larger the foreground object is, the easier it is to be detected by machine learning algorithms, likewise human. Fig. 5(e) shows the difficulty map of the razor blade, which is largely influenced by the background due to its smaller and thinner characteristics than other objects. Using these difficulty maps, the proposed X-ray image synthesis is performed.

B. IMAGE SYNTHESIS USING THE DIFFICULTY MAP

After the normalization of the difficulty map D to have the sum to be one, we obtain $\overline{D}$, which can be treated as a probability map. We then sample the target foreground position

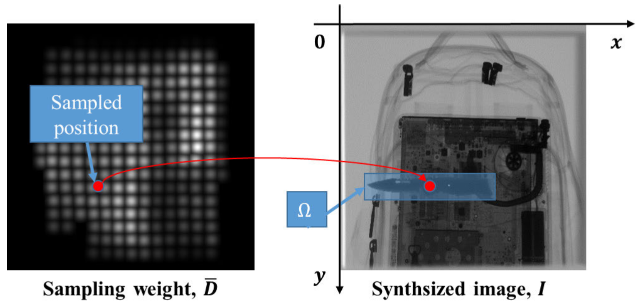


FIGURE 7. Example of sampling the target foreground position and image synthesis in the proposed method.

using $\overline{D}$.[1] Fig. 7 shows the process of sampling the target foreground position. Note that the previous X-ray synthesis [9], [11]–[13] also performs probabilistic sampling but using the 2D uniform distribution. On the contrary, we sample the positions where the object detector may get confused. In other words, our method can generate hard training samples that can boost the performance of the object detector. The proposed X-ray image synthesis is performed during training of the object detector as online data augmentation. Therefore, the same background image can be used multiple times for the same object with different scales as well as the other objects.

Given the foreground image F and the background image B, X-ray image synthesis can be performed as [8], [9], [42] according to the Beer-Lambert's law [29]. Specifically, the synthesized X-ray image I is obtained as follows:

$$I(x,y) = \begin{cases} F_r(x,y)B(x,y), & \text{if } (x,y) \in \Omega, \\ B(x,y), & \text{otherwise}, \end{cases} \tag{5}$$

where F_r denotes the randomly-scaled version of F, and Ω is a set of pixels in the F_r. I, F_r, and B have normalized values in the range [0], [1]. Because the location and class of objects are known in the image synthesis process, the ground truth can also be obtained to train the detector. The synthesized image and ground truth pairs are used to learn the detection model which extracts the difficulty maps.

IV. EXPERIMENTAL RESULTS

In this section, we present the superiority of the proposed X-ray image synthesis method by applying it to various object detection networks including SSD [18], RefineDet [43], PFPNet [44], and RFBNet [45], and comparing it with existing random synthesis methods.

A. DATASET DESCRIPTION AND EXPERIMENTAL SETUP

Our experiments have been conducted using the GDX-ray database [46]. The database contains not only 200 test images for X-ray threat detection, but also 48 background images, along with 576, 144, 200, and 100 foreground images of a knife, shuriken, gun, and razor, respectively, that are suitable for X-ray image synthesis. Using these images, we generated training data by synthesizing X-ray baggage images using the existing random position synthesis method [9] and the proposed difficulty map-based method, respectively. The total number of each training data is 30k, which is equivalent to the number of training iterations. For generating training data using the conventional synthesis method, we followed the authors' procedure using the source code provided [9]. All detection networks employed in our experiments were trained for 24k iterations with a learning rate of 1e-4, followed by 6k iterations with a learning rate of 1e-5. We used the Adam optimizer [47], and the batch size was set to 8. Our whole training process was conducted using a single NVIDIA TITAN X GPU.

B. PERFORMANCE EVALUATION

We evaluated the object detection performance using the average precision (AP) and mean AP (mAP). Table 1 shows a performance comparison of the synthesis methods on 200 real-world test images of the GDX-ray database. It can be seen that the proposed method improved the mAP scores by 3.2%, 5.0%, 3.0%, and 5.2% for SSD, RefineDet, PFPNet, and RFBNet, respectively.

TABLE 1. Average precision (AP) and mean average precision (mAP) on the test set.

Model	Synthesis	Knife	Shuriken	Gun	Razor	mAP
SSD	Random	0.167	0.718	0.999	0.655	0.635
	Proposed	0.237	0.770	0.998	0.646	0.663
RefineDet	Random	0.386	0.757	0.999	0.695	0.709
	Proposed	0.600	0.772	0.995	0.669	0.759
PFPNet	Random	0.538	0.712	0.999	0.716	0.741
	Proposed	0.607	0.770	0.998	0.709	0.771
RFBNet	Random	0.062	0.624	0.998	0.587	0.568
	Proposed	0.085	0.742	0.993	0.660	0.620

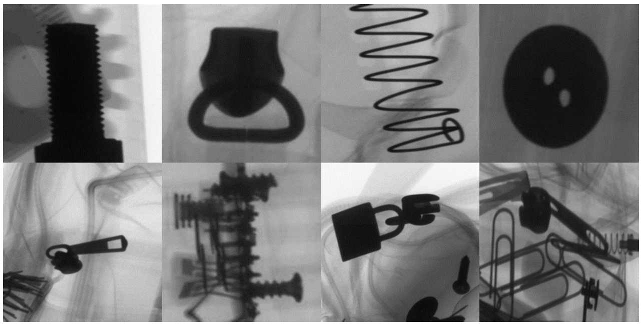


FIGURE 8. Non-hazardous image samples from the GDX-ray database.

To demonstrate the effectiveness of the proposed method more clearly, the performance comparison needs to be performed on more challenging images. To this end, we generated additional challenging test images by synthesizing more foreground objects (both hazardous and non-hazardous objects) that are largely overlapped with the existing objects in the original test images.[2] The non-hazardous foreground object samples are illustrated in Fig. 8. The experimental

[1] https://docs.python.org/3/library/random.html#random.choices

[2] https://github.com/hykim0/Xray_synthesis

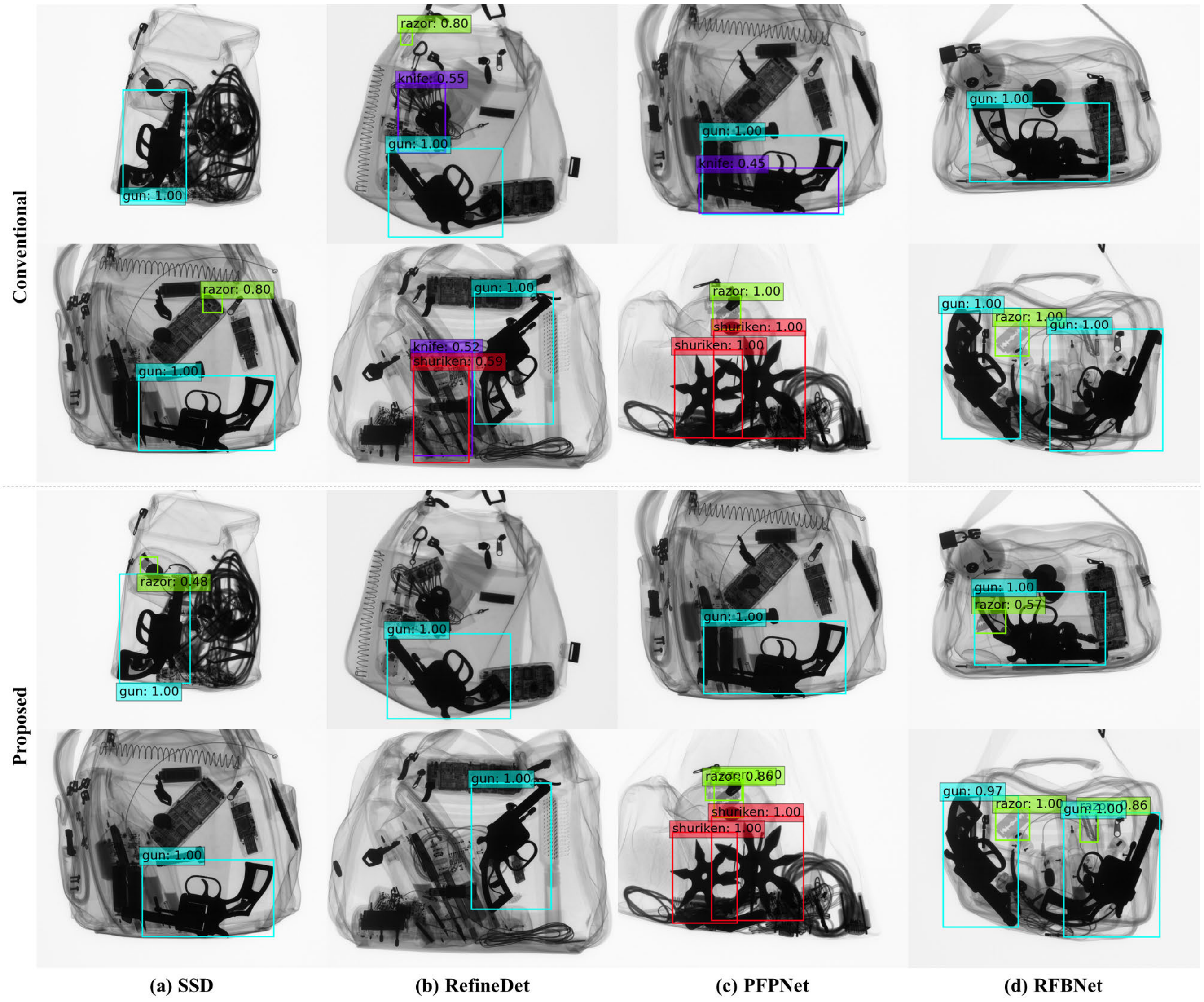


FIGURE 9. Examples of the real-world test results. Each colored box represents the detection result with the confidence score higher than 0.4.

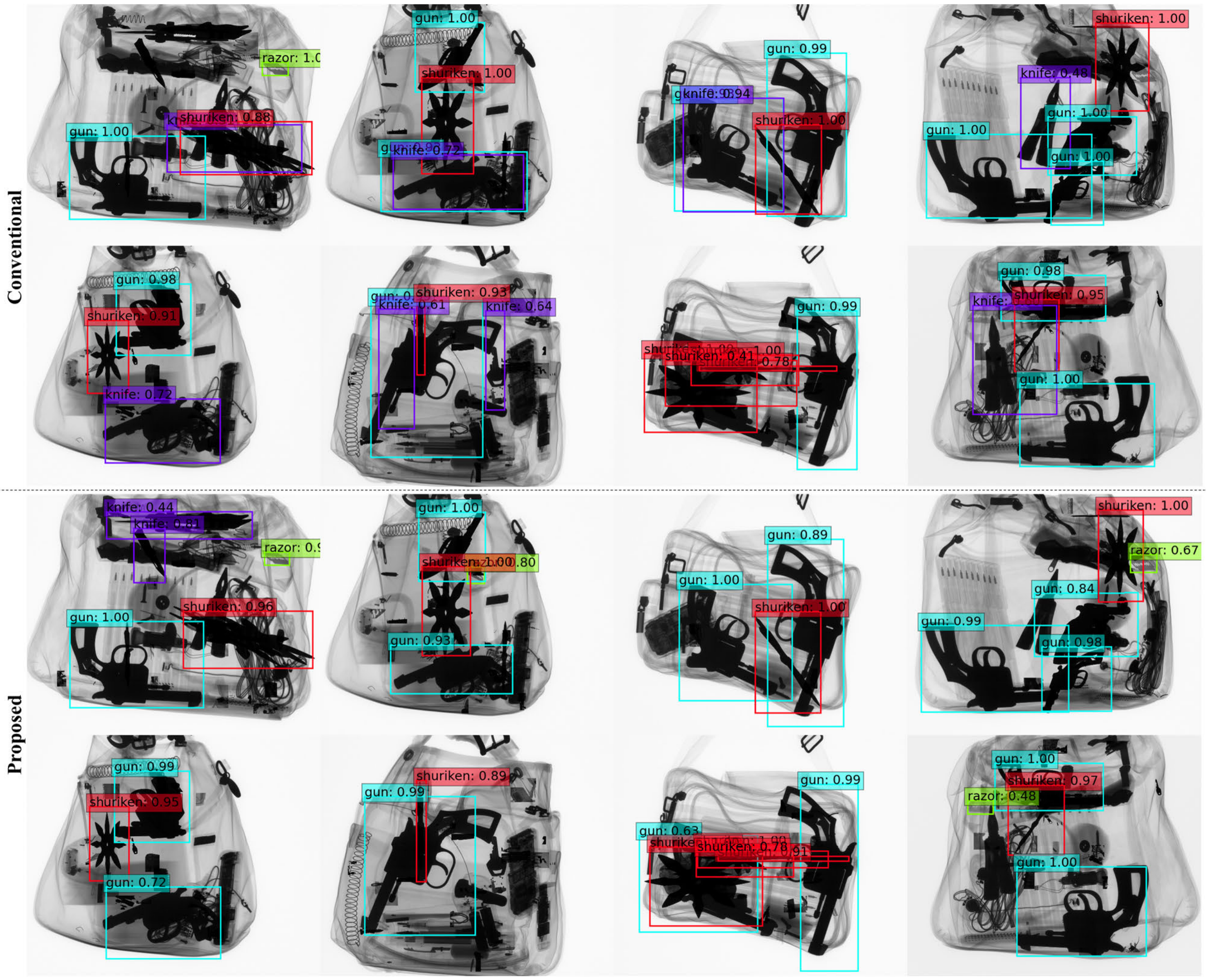
TABLE 2. Average precision (AP) and mean average precision (mAP) on the synthesized test set.

Model	Synthesis	Knife	Shuriken	Gun	Razor	mAP
SSD	Random	0.397	0.737	0.894	0.397	0.606
	Proposed	0.434	0.741	0.884	0.394	0.613
RefineDet	Random	0.637	0.816	0.921	0.441	0.704
	Proposed	0.650	0.855	0.901	0.470	0.719
PFPNet	Random	0.663	0.779	0.882	0.461	0.696
	Proposed	0.660	0.853	0.926	0.447	0.722
RFBNet	Random	0.440	0.738	0.897	0.228	0.576
	Proposed	0.488	0.811	0.883	0.254	0.609

results on this synthesized dataset are shown in Table 2. Note that the overall performance decreased due to the difficulty of object detection in cluttered scenes. However, the proposed method enabled more solid and consistent performance improvements for all tested object detection networks.

Fig. 9 shows several object detection results obtained using the conventional and proposed synthesis methods on the real-world test images. As shown in Figs. 9(a) and (d), the conventional method failed in detecting small occluded objects. On the contrary, such objects can be correctly detected by applying our synthesis method. Furthermore, the proposed method reduced false alarms as shown in Figs. 9(b) and (c).

The results of each method on the synthesized test images are illustrated in Fig. 10. Figs. 10(a), (b), and (d) show that the conventional method failed in detecting occluded objects, which can be correctly detected by applying our proposed method. Moreover, although the conventional method caused false alarms by a more complicated test set, the proposed method provided accurate detection results.



(a) SSD (b) RefineDet (c) PFPNet (d) RFBNet

FIGURE 10. Examples of the test results on the synthesized test set. Each colored box represents the detection result with the confidence score higher than 0.4.

V. CONCLUSION

A novel learning-based image synthesis method was proposed to train object detection networks for X-ray security applications. The proposed method extracts the difficulty map, which is used for sampling the target foreground position for image synthesis during the training process. By synthesizing foreground objects at hard-to-detect locations, more challenging training samples can be generated, yielding improved object detection performance. The experimental results show that the proposed method improves the performance of various object detection networks compared to the previous standard of random image synthesis.

REFERENCES

[1] J. L. Glover, L. T. Hudson, and N. G. Paulter, "Improved threat identification using tonemapping of high-dynamic-range X-ray images," *J. Test. Eval.*, vol. 46, no. 4, pp. 1462–1467, May 2018.

[2] V. Riffo and D. Mery, "Automated detection of threat objects using adapted implicit shape model," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 4, pp. 472–482, Apr. 2016.

[3] D. Mery, E. Svec, and M. Arias, "Object recognition in baggage inspection using adaptive sparse representations of X-ray images," in *Proc. Image Video Technol.* Cham, Switzerland: Springer, 2015, pp. 709–720.

[4] M. Baştan, M. R. Yousefi, and T. M. Breuel, "Visual words on baggage X-ray images," in *Proc. Int. Conf. Comput. Anal. Images Patterns*. Berlin, Germany: Springer, 2011, pp. 360–368.

[5] T. Franzel, U. Schmidt, and S. Roth, "Object detection in multi-view X-ray images," in *Pattern Recognition*. Berlin, Germany: Springer, 2012, pp. 144–154.

[6] S. Akcay, M. E. Kundegorski, C. G. Willcocks, and T. P. Breckon, "Using deep convolutional neural network architectures for object classification and detection within X-ray baggage security imagery," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 9, pp. 2203–2215, Sep. 2018.

[7] L. D. Griffin, M. Caldwell, J. T. Andrews, and H. Bohler, "'Unexpected item in the bagging area': Anomaly detection in X-ray security images," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 6, pp. 1539–1553, Jun. 2019.

[8] B. Gu, R. Ge, Y. Chen, L. Luo, and G. Coatrieux, "Automatic and robust object detection in X-ray baggage inspection using deep convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 68, no. 10, pp. 10248–10257, Oct. 2021.

[9] D. Saavedra, S. Banerjee, and D. Mery, "Detection of threat objects in baggage inspection with X-ray images using deep learning," *Neural. Comput. Appl.*, vol. 33, pp. 7803–7819, Jul. 2021.

[10] C. Miao, L. Xie, F. Wan, C. Su, H. Liu, J. Jiao, and Q. Ye, "SIXray: A large-scale security inspection X-ray benchmark for prohibited item discovery in overlapping images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 2119–2128.

[11] D. Jain and D. Kumar, "An evaluation of deep learning based object detection strategies for threat object detection in baggage security imagery," *Pattern Recognit. Lett.*, vol. 120, pp. 112–119, Apr. 2019.

[12] J. Yang, Z. Zhao, H. Zhang, and Y. Shi, "Data augmentation for X-ray prohibited item images using generative adversarial networks," *IEEE Access*, vol. 7, pp. 28894–28902, 2019.

[13] Y. Zhu, Y. Zhang, H. Zhang, J. Yang, and Z. Zhao, "Data augmentation of X-ray images in baggage inspection based on generative adversarial networks," *IEEE Access*, vol. 8, pp. 86536–86544, 2020.

[14] H. E. Martz, C. M. Logan, D. J. Schneberk, and P. J. Shull, *X-Ray Imaging: Fundamentals, Industrial Techniques and Applications*. Boca Raton, FL, USA: CRC Press, 2016.

[15] D. Mery and A. K. Katsaggelos, "A logarithmic X-ray imaging model for baggage inspection: Simulation and object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2017, pp. 57–65.

[16] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 7263–7271.

[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2016.

[18] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 21–37.

[19] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–26.

[20] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2117–2125.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.

[22] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2980–2988.

[23] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: http://arxiv.org/abs/1804.02767

[24] J. Liu, Y. Hu, J. Yang, Y. Chen, H. Shu, L. Luo, Q. Feng, Z. Gui, and G. Coatrieux, "3D feature constrained reconstruction for low-dose CT imaging," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 5, pp. 1232–1247, May 2018.

[25] V. Riffo and D. Mery, "Active X-ray testing of complex objects," *Insight, Non-Destructive Test. Condition Monit.*, vol. 54, no. 1, pp. 28–35, Jan. 2012.

[26] V. Riffo, S. Flores, and D. Mery, "Threat objects detection in X-ray images using an active vision approach," *J. Nondestruct. Eval.*, vol. 36, no. 3, pp. 1–13, Sep. 2017.

[27] M. Baştan, "Multi-view object detection in dual-energy X-ray images," *Mach. Vis. Appl.*, vol. 26, nos. 7–8, pp. 1045–1060, Nov. 2015.

[28] S. Akçay, M. E. Kundegorski, M. Devereux, and T. P. Breckon, "Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2016, pp. 1057–1061.

[29] H. H. Barrett and W. Swindell, *Radiological Imaging: The Theory of Image Formation, Detection, and Processing*. New York, NY, USA: Academic, 1996.

[30] T. W. Rogers, N. Jaccard, E. D. Protonotarios, J. Ollier, E. J. Morton, and L. D. Griffin, "Threat image projection (TIP) into X-ray images of cargo containers for training humans and machines," in *Proc. IEEE Int. Carnahan Conf. Secur. Technol.*, Oct. 2016, pp. 1–7.

[31] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[32] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[33] L. Shao, F. Zhu, and X. Li, "Transfer learning for visual categorization: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 5, pp. 1019–1034, May 2014.

[34] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.

[35] F. J. Moreno-Barea, F. Strazzera, J. M. Jerez, D. Urda, and L. Franco, "Forward noise adjustment scheme for data augmentation," in *Proc. IEEE Symp. Ser. Comput. Intell.*, Nov. 2018, pp. 728–734.

[36] G. Kang, X. Dong, L. Zheng, and Y. Yang, "PatchShuffle regularization," 2017, *arXiv:1707.07103*. [Online]. Available: http://arxiv.org/abs/1707.07103

[37] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 13001–13008.

[38] L. Xie, J. Wang, Z. Wei, M. Wang, and Q. Tian, "DisturbLabel: Regularizing CNN on the loss layer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 4753–4762.

[39] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101552.

[40] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018.

[41] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 7354–7363.

[42] H.-Y. Kim, S. Park, Y.-G. Shin, S.-W. Jung, and S.-J. Ko, "Detail restoration and tone mapping networks for X-ray security inspection," *IEEE Access*, vol. 8, pp. 197473–197483, 2020.

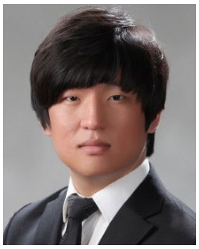
[43] S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li, "Single-shot refinement neural network for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4203–4212.

[44] S.-W. Kim, H.-K. Kook, J.-Y. Sun, M.-C. Kang, and S.-J. Ko, "Parallel feature pyramid network for object detection," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 234–250.

[45] L. Deng, M. Yang, T. Li, Y. He, and C. Wang, "RFBNet: Deep multimodal networks with residual fusion blocks for RGB-D semantic segmentation," 2019, *arXiv:1907.00135*. [Online]. Available: http://arxiv.org/abs/1907.00135

[46] D. Mery, V. Riffo, U. Zscherpel, G. Mondragón, I. Lillo, I. Zuccar, H. Lobel, and M. Carrasco, "GDXray: The database of X-ray images for nondestructive testing," *J. Nondestruct. Eval.*, vol. 34, no. 4, p. 42, Nov. 2015.

[47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980



HYO-YOUNG KIM received the B.S. degree in electrical engineering from Korea University, in 2013, where he is currently pursuing the Ph.D. degree. His current research interests include image processing, computer vision, and artificial intelligence.

SUNG-JIN CHO received the B.S. degree in electronics engineering from Korea University, in 2018. He entered the Computer Vision and Image Processing Laboratory, Department of Electronic Engineering, Korea University, in 2018. His research interests include image processing and computer vision.

SEUNG-JIN BAEK (Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Korea University, Seoul, South Korea, in 2007 and 2013, respectively. He joined the Digital Media and Communications Research and Development Center, Samsung Electronics Company Ltd., Suwon, South Korea, in 2013, where he was a Senior Engineer, from 2014 to 2015. From 2015 to 2020, he was a Staff Engineer with the Visual Display Business Division, Samsung Electronics Company Ltd., Suwon. He is currently a Research Professor with the Research Institute of Information and Communication Technology, Korea University. His current research interests include image processing and computer vision.

SEUNG-WON JUNG (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Korea University, Seoul, South Korea, in 2005 and 2011, respectively. From 2011 to 2012, he was a Research Professor with the Research Institute of Information and Communication Technology, Korea University. From 2012 to 2014, he was a Research Scientist with Samsung Advanced Institute of Technology, Yongin, South Korea. From 2014 to 2020, he was an Assistant Professor with the Department of Multimedia Engineering, Dongguk University, Seoul. In 2020, he joined the Department of Electrical Engineering, Korea University, where he is currently an Associate Professor. He has published over 60 peer-reviewed articles in international journals. His current research interests include image processing and computer vision. He received Hae-Dong Young Scholar Award from the Institute of Electronics and Information Engineers, in 2019.

SUNG-JEA KO (Fellow, IEEE) received the B.S. degree in electronic engineering from Korea University, in 1980, and the M.S. and Ph.D. degrees in electrical and computer engineering from The State University of New York at Buffalo, in 1986 and 1988, respectively.

From 1988 to 1992, he was an Assistant Professor with the Department of Electrical and Computer Engineering, University of Michigan-Dearborn. In 1992, he joined the Department of Electronic Engineering, Korea University, where he is currently a Professor. He has published over 210 international journal articles and holds over 60 registered patents in fields, such as video signal processing and computer vision.

Dr. Ko is a member of the National Academy of Engineering of Korea. He was a recipient of the Best Paper Award from IEEE Asia–Pacific Conference on Circuits and Systems, in 1996; Hae-Dong Best Paper Award from the Institute of Electronics and Information Engineers (IEIE), in 1997; the 1999 LG Research Award; the Research Excellence Award from Korea University, in 2004; the Technical Achievement Award from IEEE Consumer Electronics (CE) Society, in 2012; the 15-Year Service Award from TPC of ICCE, in 2014; the Chester Sall Award (First Place Transaction Paper Award) from IEEE CE Society, in 2017; and the Science and Technology Achievement Medal from the Korean Government, in 2020. He served as the General Chairperson for ITC-CSCC 2012 and IEICE 2013. He was the President of IEIE, in 2013; the Vice President of IEEE CE Society, from 2013 to 2016; and the Distinguished Lecturer of IEEE, from 2015 to 2017. He is also an Editorial Board Member of IEEE TRANSACTIONS ON CONSUMER ELECTRONICS.