Improvised Explosive Device Detection Using CNN With X-Ray Images

Chakkaphat Chamnanphan [1], Surapol Vorapatratorn [1,*], Khwunta Kirimasthong [1], Tossapon Boongoen [2], and Natthakan Iam-On [2]

[1] Center of Excellence in AI & Emerging Technologies, School of IT, Mae Fah Luang University, Chiangrai, Thailand;
Email: 6251301001@lamduan.mfu.ac.th (C.C.), khwunta.kir@mfu.ac.th (K.K.)
[2] Department of Computer Science, Aberystwyth University, Aberystwyth, Ceredigion, UK;
Email: tob45@aber.ac.uk (T.B.), nai7@aber.ac.uk (N.I.-O.)
*Correspondence: surapol.vor@mfu.ac.th (S.V.)

***Abstract*—The concept of a smart city and its associated services have been extensively explored in terms of innovation development and the application of technological concepts. One of the significant concerns in promoting smart living is the security of personal lives and assets, which are at risk from organized crime and acts of terrorism. A considerable amount of attention is paid to preventing bomb attacks in public places, especially the detection of an Improvised Explosive Device (IED). This research focuses on developing an analysis model that can accurately classify instances of x-ray images of baggage or objects as containing IEDs or not. The model provides an alternative to conventional techniques that fail to detect concealed or hidden devices. For this specific project, sample images are generated by experts to cover a range of cases encountered in operations during the past decade. These images are then used to develop a deep learning model, employing several data augmentation methods to overcome the issue of a limited number of training samples. As compared to a related work that exploits neural networks, the proposed model usually achieves higher accuracy rates for unseen samples, with the best accuracy rate being 0.985. Furthermore, an empirical study is conducted to determine the optimal size of the training set that exhibits good predictive performance. The study reveals that a large training set, apart from using a lot of resources, may not yield the best results as it may indicate overfitting.**

***Keywords*—Improvised Explosive Device (IED), detection, x-ray image, classification, Convolutional Neural Network (CNN), augmentation**

I. Introduction

With the recent surge of development and technological applications to smart cities, different sensor and artificial intelligence-led systems have been introduced to monitor and analyze behavioral patterns [1]. Based on the survey by Allam and Newman [2], previous works in this active area belong to various categories such as smart people [3], smart environment [4], and public safety [5]. In addition to road and transport safety, the last group also highly overlaps with the security domain, especially those related to organized crime and terrorism. As a result, there are several approaches proposed in the literature to address this threat to both public and individual safety [6].

For instance, some researchers focus on disclosing groups and communication trends from online and other relevant resources [7]. Some others direct their attention to cybersecurity, which has emerged as a significant theater of terrorist acts [8]. Besides, a few more concentrate on identifying and detecting Chemical, Biological, and Explosive (CBE) substances [9]. Specific to an Improvised Explosive Device (IED) that has been very popular for hiding a bombing attack, a computational method that supports accurate and automated detection is highly demanded [10]. IED terrorist attacks commonly take place in densely populated and busy locations such as airports, official and commercial buildings. To avoid positive identification by typical chemical-led detectors, IEDs are usually concealed within a package or baggage [11]. Therefore, an alternative method is required to recognize those units based on shape, density, and composition, which can be captured by an imaging sensor like an x-ray [12]. This trend has motivated the current research, which aims to investigate the use of Convolutional Neural Networks (CNNs) to develop a recognition capability for an accurate IED detection system. Details of the background, the proposed research work and the contribution are summarized below.

A. Background and Motivation

Terrorist attack incidents in Thailand's three southern provinces pose a significant challenge to national security, endangering the lives and property of government officials and residents in the area. Moreover, it results in economic losses and a decline in investment, leading to adverse effects on tourism, one of the primary sources of national income [13]. Although the government and the Ministry of Defense have worked together continuously to address the issue, their efforts have only been partially effective, with periodic bombings in public places and ambushes of government officials still being witnessed rather often.

As one of the preventive measures, x-ray explosive detection systems that can operate alongside ordnance-

Manuscript received November 18, 2022; revised January 29 2023; accepted March 21, 2023; published July 19, 2023.

disposal robots [14] can serve as a critical decision-making aid to prevent mishaps and minimize losses. This research work is part of a collaborative project between the Defense Technology Institute (DTI), Mae Fah Luang University, and the Department of Ordnance, Royal Thai Air Force (RTAF), with the objective to reduce the need for importing high-value technology from international companies. Note that the current study reinforces the success of previous projects through the same collaboration on topics of face recognition [15, 16] and suspect vehicle detection [17].

B. Proposed Work

The main objective of this study is to develop a deep learning model, specifically a Convolutional Neural Network (CNN) [18], using a set of x-ray images of baggage and containers that may contain IEDs. The model is intended to be used to provide a detection capability for the explosive-recovery robot system [14]. The scope of this work is limited to images captured by a portable x-ray camera installed on the robot system, which produces lower-quality images than those generated by a commercial unit commonly used in sensitive locations such as airports. In particular, a collection of image samples is collected from representative scenarios specified by experts. These involve different settings of both visible and concealed IEDs, which are seen in many incidents over the past decade. The problem under investigation is a binary image classification [19], where a new image is to be classified as it contains IEDs or not.

C. Contribution

The contribution made by this research is three folds. Firstly, it proposes an original approach of using CNN to develop a classification model for IED detection in the context of ordnance-disposal robots, which will be deployed to operational units in southern Thailand. The 2-D images used in this study are collected from actual x-ray sensors and cover real cases from the past, leading to an original application of CNN in homeland security and anti-terrorism. Secondly, this work presents a unique application of data augmentation techniques to x-ray images of IEDs during the data preprocessing stage. This is considered necessary given the limited amount and quality of those images acquired from domain experts. As such, the paper delivers original findings with respect to the relationship between the size/method of oversampling and predictive performance. Finally, this work provides a parameter analysis of the data augmentation process that would be useful for future extensions or applications of the proposed approach.

The remainder of this paper is structured as follows. Section II presents related works of IED and detection techniques, with a focus on image-based approaches. Then, Section III provides details of the proposed method and materials employed in this research. After that, Section IV presents the evaluation results and associated discussion, including those related to the data augmentation process and parameter analysis. Finally, Section V concludes the paper with future research directions.

II. Related Work

Since the 1990s, there has been a rapid increase in terrorist attacks worldwide, with explosive incidents being a common theme [20]. Given the tight control of military explosives in most countries, Improvised Explosive Devices (IEDs) have become the major tool for terrorism, as they are a lot easier to obtain [21]. IEDs are typically concealed and hidden in passengers' baggage, especially for air transportation, to avoid detection [12]. While many detection methods, such as mass spectrometry [22], ion mobility spectrometry [11], Raman spectrometry [23], and fluorescence sensors [24], are effective for common explosives, they are unfortunately less useful for the detection of IEDs. Consequently, since the event of September 11, 2001, research works have focused on exploring visual inspection of x-ray images, which can support visual search and decision-making accurately [25]. The current research direction is consistent with those related to the application of Artificial Intelligence (AI) techniques to the visual-based inspection of IEDs. These are given in Table I. For a comprehensive review encompassing baggage inspection, please refer to [26].

TABLE I. Summarization of Related Works on X-Ray Visual Inspection of IEDs

Relate work & Reference	Context of x-ray visual inspection	Exploited AI techniques
Bag-of-visual-word feature [27]	2-D image of baggage	Classification: SVM & SIFT [28]
Shape-based modeling [29-30]	2-D image of baggage	Classification: fuzzy KNN [31]
Robust bag-of-visual-word [32]	2-D image of baggage	Classification: SVM & SURF [33]
Spare texture descriptor [34]	2-D image of baggage	Classification: SVM [35]
3-D object classification [36]	2-D image of baggage	Classification: Clustering & RF [37]
Transfer learning [38]	2-D image of baggage	Object detection: CNN
Visual object detection [39, 40]	2-D image of baggage	Object detection: R-CNN
Active vision approach [41]	2-D image of baggage	Object detection: Q-learning
Attention based detection [42]	2-D image of baggage	Object detection: CNN
Object-wise anomaly detection [43]	2-D image of baggage	Object detection: Dual-CNN
Change detection [44]	2-D x-ray ground image	Object detection: CNN

Note: abbreviations used herein are Listed as follows: SVM (Support Vector Machine), SIFT (Scale-Invariant Feature Transform), KNN (K-Nearest Neighbors), SURF (Speeded-Up Robust Feature), RF (Random Forest), and R-CNN (Region-Based CNN), respectively.

At large, the recognition of objects in x-ray images presents a challenging problem in the fields of computer vision and data science. This is largely due to occlusions between the object of interest and others, self-occlusions, clutter in the background, and variations in viewpoints. As a result, manual feature engineering approaches such as those implemented by [27–37] may lead to the problem of overfitting to training data, i.e., there is no guarantee that a learned model is able to achieve good predictive performance with unseen images. To address this

challenge, several investigations have utilized deep learning models to automatically determine informative features [38–44]. Previous reports have suggested that a simple Convolutional Neural Network (CNN) can be effective for classification tasks with 2-D x-ray images, despite being applied in different contexts such as 3-D imaging or object detection [38, 42]. In fact, several works have shown that CNN models with standard layer sets can be competitive with more complex counterparts [45, 46]. However, the potential of a model is only realized when both the number of samples and their diversity are sufficient, as demonstrated by the development of large datasets [47] such as ImageNet [48] and GoogleNet [49], each containing over 10 million images for model training. Unfortunately, publicly available datasets of x-ray images are rare, and generating such a dataset is often difficult and expensive. This is also the case for the current project where the size of the dataset is quite limited.

In order to address the challenge of data availability, recent research studies have proposed various algorithms to optimize the quality of a target model learned from a limited sample set. One of these involves modifying structural settings and algorithmic processes within a network. For example, the introduction of dropout layers [50] has gained significant attention as a means of preventing overfitting. This approach reduces the number of parameters to update, thereby addressing the issue of insufficient training images. Another approach is the concept of transfer-learning [51], in which a model trained in one domain is re-trained with samples specific to a new one.

In addition to those mentioned above, data augmentation techniques are also commonly exploited to overcome the curse of limited data collection. They aim to increase the number of image samples by making slight changes to original ones. This involves operations such as flipping [52], rotation [53], and color space shifting [54]. The approach has demonstrated high utility in practice for the task of x-ray image classification [55]. For a comprehensive review of these methods, a reader may refer to the works of [56, 57].

III. Method and Material

In this section, we elaborate on the methodology employed in our current research. Specifically, we describe the various stages involved, including data collection, data preprocessing, model development, and evaluation, and provide detailed information regarding the proposed models. The purpose of this section is to provide context for the subsequent section, where we present our results and discuss their implications.

A. Data Collection and Preprocessing

As previously mentioned, the image dataset *X* used for this study was created by experts in the field following the approach outlined in [58]. The resulting dataset contains three subsets: $X = \{X_{train}, X_{valdate}, X_{unseen}\}$ that are used in the conventional flow of model development and evaluation. The first subset X_{train} is employed for training the classification model, while the second subset $X_{validate}$ is utilized as a hold-out set to find an optimal set of parameters. Finally, the subset X_{unseen} is used to evaluate the predictive quality of the learned and optimized model. Note that each of samples in *X* is labeled as either an explosive or non-explosive image. Examples of these two images categorise or classes are illustrated in Figs. 1 and 2, respectively. Moreover, further subset-specific distributions are given below.

- Train dataset (X_{train}): originally consists of 39 explosive and 27 non-explosive images,
- Validation dataset ($X_{validate}$): has 200 explosive and another 200 non-explosive 200 images, and
- Unseen dataset (X_{unseen}): is composed of 375 explosive and 375 non-explosive images.

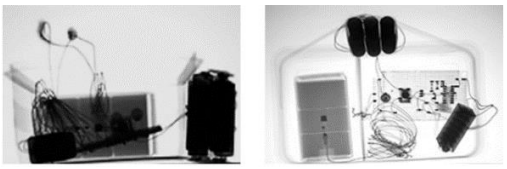

Figure 1. Examples of explosive images.

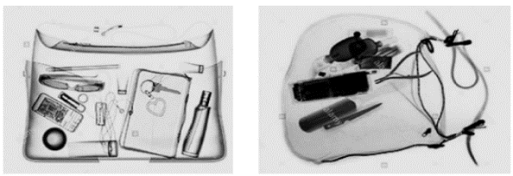

Figure 2. Examples of non-explosive images.

Prior the development of a classification model, a series of preparatory steps are conducted. Initially, those collected image samples in *X* undergo a filtering process that aims to reduce background noise. Subsequently, data augmentation techniques, including flipping [52], rotation [53], and brightness adjustment [54], are employed to increase the size of the training set beyond that of the original X_{train}, which consisted of only 66 images. This augmentation procedure results in the creation of various training-set variations, not only to address the issue of a limited training data but also to examine the correlation between the training-set size and the model predictive performance. Notably, to avoid the issue of imbalanced classes, the augmented training sets comprise an equal number of explosive and non-explosive images. As a result, five different augmented sets of training data are produced as follows, with details of augmentation techniques used in this generation in Tables II and III.

- Augmented set 500 ($X_{train-500}$): consists of 250 explosives and 250 non-explosive images,
- Augmented set 1000 ($X_{train-1000}$): consists of 500 explosives and 500 non-explosive images,
- Augmented set 2000 ($X_{train-2000}$): consists of 1000 explosive and 1000 non-explosive images,
- Augmented set 3000 ($X_{train-3000}$): consists of 1500 explosive and 1500 non-explosive images, and
- Augmented set 4000 ($X_{train-4000}$): consists of 2000 explosive and 2000 non-explosive images.

TABLE II. AUGMENTED TYPES AND NUMBERS OF EXPLOSIVE X-RAY IMAGES

	500 images	1000 images	2000 images	3000 images	4000 images
Exp	39	39	39	39	39
Exp-l	39	39	39	39	39
Exp-d	39	39	39	39	39
Exp-HP	39	39	39	39	39
Exp-HP-l	39	39	39	39	39
Exp-HP-d	39	39	39	39	39
Exp-R	4	45	129	211	295
Exp-l-R	2	44	127	211	294
Exp-d-R	2	44	127	211	294
Exp-HP-R	4	45	129	211	295
Exp-HP-l-R	2	44	127	211	294
Exp-HP-d-R	2	44	127	211	294
Sum	**250**	**500**	**1000**	**1500**	**2000**

Note: Abbreviations Used Herein Are listed as Follows: Exp (Explosive), Exp-l (Explosive Lightness), Exp-d (Explosive Darkness), Exp-HP (Explosive Horizontal Flipping), Exp-HP-l (Explosive Horizontal Flipping-lightness), Exp-HP-d (Explosive Horizontal Flipping-darkness), Exp-R (Explosive Rotation), Exp-l-R (Explosive Lightness Rotation), Exp-d-R (Explosive Darkness Rotation), Exp-HP-R (Explosive Horizontal Flipping Rotation), Exp-HP-l-R (Explosive Horizontal Flipping-lightness Rotation), Exp-HP-d-R (Explosive Horizontal Flipping-darkness Rotation).

TABLE III. AUGMENTED TYPES AND NUMBERS OF NONEXPLOSIVE X-RAY IMAGES

	500 images	1000 images	2000 images	3000 images	4000 images
Non-Exp	27	27	27	27	27
Non-Exp-l	27	27	27	27	27
Non-Exp-d	27	27	27	27	27
Non-Exp-HP	27	27	27	27	27
Non-Exp-HP-l	27	27	27	27	27
Non-Exp-HP-d	27	27	27	27	27
Non-Exp-R	16	57	141	223	307
Non-Exp-l-R	14	56	139	223	306
Non-Exp-d-R	14	56	139	223	306
Non-Exp-HP-R	16	57	141	223	307
Non-Exp-HP-l-R	14	56	139	223	306
Non-Exp-HP-d-R	14	56	139	223	306
Sum	**250**	**500**	**1000**	**1500**	**2000**

Note: Abbreviations Used Herein are Similar to Those Appear in Table II, with the Additional Entry of Non-Exp (Nonexplosive).

Upon the completion of initial experiments (on those five augmented training sets) that will be reported in the next section, a question arises as to whether a set containing between 2000 and 3000 images could yield a higher or more proximal accuracy compared to those five. This is motivated by the observation of high rates with the two sets of 2000 and 3000 samples. In order to address this inquiry, three supplementary training datasets comprising 2250, 2500, and 2750 images respectively, have been incorporated into the empirical study. See Tables IV and V for further detail.

- Augmented set 2250 ($X_{train-2250}$): consists of 1125 explosive and 1125 non-explosive images,
- Augmented set 2500 ($X_{train-2500}$): consists of 1250 explosive and 1250 non-explosive images, and
- Augmented set 2750 ($X_{train-2750}$): consists of 1375 explosive and 1375 non-explosive images.

TABLE IV. AUGMENTED TYPES AND NUMBERS OF EXPLOSIVE X-RAY IMAGES IN THE TRAINING DATASETS OF 2250, 2500 AND 2750 IMAGES

	2250 images	2500 images	2750 images
Exp	39	39	39
Exp-l	39	39	39
Exp-d	39	39	39
Exp-HP	39	39	39
Exp-HP-l	39	39	39
Exp-HP-d	39	39	39
Exp-R	150	170	191
Exp-l-R	148	169	190
Exp-d-R	148	169	190
Exp-HP-R	149	170	191
Exp-HP-l-R	148	169	190
Exp-HP-d-R	148	169	190
Sum	**1125**	**1250**	**1375**

TABLE V. AUGMENTED TYPES AND NUMBERS OF NONEXPLOSIVE X-RAY IMAGES IN THE TRAINING DATASETS OF 2250, 2500 AND 2750 IMAGES

	2250 images	2500 images	2750 images
Non-Exp	27	27	27
Non-Exp-l	27	27	27
Non-Exp-d	27	27	27
Non-Exp-HP	27	27	27
Non-Exp-HP-l	27	27	27
Non-Exp-HP-d	27	27	27
Non-Exp-R	162	182	203
Non-Exp-l-R	160	181	202
Non-Exp-d-R	160	181	202
Non-Exp-HP-R	161	182	202
Non-Exp-HP-l-R	160	181	202
Non-Exp-HP-d-R	160	181	202
Sum	**1125**	**1250**	**1375**

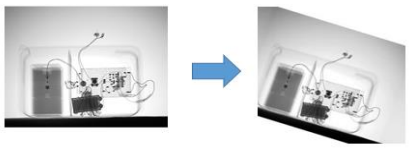

Figure 3. Example of rotating an image by 20 degrees.

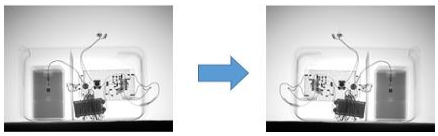

Figure 4. Example of a horizontal flipping.

Figure 5. Examples of brightness adjustment: 25% lightness and 50% darkness 25%.

In a data analytics research, data augmentation techniques have become widely used to mitigate the challenges posed by limited datasets. It is observed in many CNN-driven studies that the machine learning program usually treats an image as a matrix structure of numerical values that present a degree of brightness as an individual and a pattern as a group. Employing those three techniques of augmentation produces a series of new images with modified structures or value ranges. In

particular, the first involves an image rotation, which alters its orientation through a rotation degree. For this study, the right-hand rotation method is employed. An illustrative sample of this procedure is illustrated in Fig. 3, with a rotation degree of 20. In contrast to the previous method, flipping as the second augmented operation is similar to reflecting the image itself in one dimension, which can be either a horizontal or vertical flip. Fig. 4 demonstrates an example of a horizontal flip. Furthermore, Fig. 5 demonstrates two brightness operations employed in this work [59–61]. Note that the implementation of these three techniques and the learning model discussed in the following section are made available via GitHub (https://github.com/KwinLook/ADEX-Project.git).

B. Model Development and Evaluation

In line with the successful utilization of CNN for object detection in x-ray images [35, 39], this study employs a common model for the classification of input images as an explosive or a non-explosive one. Essentially, CNN is a special type of Artificial Neural Network (ANN). ANNs generally consist of an input layer, a hidden layer or layers, and an output layer, where each layer is comprised of a number of computational neurons, except for the input layer. For an image sample, each element is represented by a vector $X = [x_1, x_2, \ldots, x_n]$ where x_n denotes the value of the nth pixel taken by a specific neuron. Hence, the number of input neurons equals the resolution of a visual data matrix. Similarly, the size of the output layer is dependent on the number of target classes define for a classification problem. For instance, a binary classification problem requires two output neurons representing classes 0 and 1. One of these maintains the probability of an input instance belonging to a specific class. Each neuron within a layer is connected to all neurons in the adjacent layers, constructing a fully connected network with a set of connection-specific weights. The output of one layer serves as the input of the next layer. For each computational neuron, the weighted sum of the input vector x is passed through an activation function, which introduces nonlinear factors into the network to meet some of the problem specifications to be solved: $f(W^T + b)$ where W denotes the weight vector, and b denotes the bias. Two commonly used activation functions are Rectified Linear Unit (ReLU): $f(x) = max(0, x)$ and Sigmoid: $f(x) = (1 + e^{-x})^{-1}$.

The CNN model includes one or more layers added on top of the fully connected network that alternate with integration layers, as illustrated conceptually by the layer-wise framework in Fig. 6. Each convolutional layer comprises many property detectors, also called filters or kernels that have been designed to capture properties of each input image. These filters are essentially a matrix of weights that scan through an image; thus the distortion is achieved by computing the dot product of the matrix and the corresponding part of the image at all positions. With this, each result is regarded as an element in a new matrix known as a feature map. It is typically followed by an aggregation layer, where the collected features are divided into regions of the same size, such as 2 × 2, and one input value is subsampled from each region according to the maximum, average, or even a random value. Consequently, the subsampling function reduces the size of the input representation progressively, thereby decreasing the network complexity and accelerating computation. This also makes the network more robust to minor changes, distortions, and translations [62].

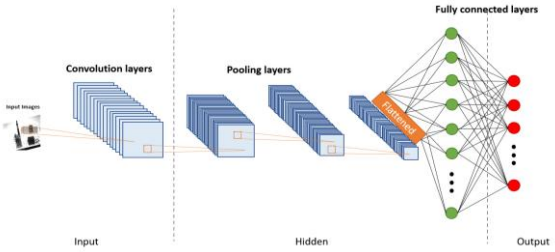

Figure 6. The layer-wise architecture of a common CNN.

In our work, a maximum integration strategy is employed to retain the most crucial characteristics of the input. Additionally, a dropout layer can be inserted between convolutional layers, which is a technique used to prevent overfitting by randomly ignoring some nodes from the network with respect to each training sample [63]. To adapt the predefined CNN model to a given training data, optimization is a crucial step. Several optimization tools are commonly used, including SGD, Adagrad, Adadelta, RMSprop, and Adam. In general, these methods extend a basic optimizer with batch size, momentum, adaptive learning rate, and other features [64, 65]. Fig. 7 provides a detailed description of the specific CNN model to be exploited in this empirical study. In particular, 3 by 3 filters are exploited in the coevolutionary layers (i.e., Conv2D), with ReLU being the default activation function [66]. Also, Adam and CrossEntropy are choices of the optimizer and the loss function, respectively. Other major parameter settings are the learning rate of 0.001, the batch size and epoch of and 35 and 100. Note that dropout layers are included in the first four sections of the proposed CNN network (with dropping ratios of 0.75, 0.5, 0.25 and 0.25), such that the chance of overfitting can be reduced, especially at the early stage in this network-driven learning.

Having finalized the network design, Fig. 8 depicts the process of model development and evaluation. The first step involves the preparation and augmentation of the training data, followed by the development of the CNN model, which is initially evaluated using the validation set to determine the hyper-parameter settings. To achieve this, patterns of accuracy and loss are examined with respect to both the training and validation sets. Once the desired model is established, it is evaluated against an unseen test set to conclude. The evaluation process is based on a confusion matrix, which represents a measure of comparison between the true classes and the predictions made by the model. Specifically, in the binary classification study presented in this research (as shown in Fig. 9), the confusion matrix comprises four measures: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [67]. The hardware utilized for training the CNN model is based solely on the Google

Colab service, which constitutes a component of the Google Cloud platform. In particular, all settings are configured in accordance with the following specification.

- Hardware accelerator: GPU,
- NVIDIA-SMI: 525.85.12,
- CUDA Version: 12.0,
- Runtime shape: High RAM (90 GB), and
- Disk Space: SSD (170 GB).

Layer (type)	Output Shape	Param #
********** Pooling 1 **********		
Conv2D	(None, 148, 148, 128)	3584
MaxPooling2D	(None, 74, 74, 128)	0
Dropout	(None, 74, 74, 128)	0
********** Pooling 2 **********		
Conv2D	(None, 72, 72, 100)	115300
MaxPooling2D	(None, 36, 36, 100)	0
Dropout	(None, 36, 36, 100)	0
********** Pooling 3 **********		
Conv2D	(None, 34, 34, 64)	57664
MaxPooling2D	(None, 17, 17, 64)	0
Dropout	(None, 17, 17, 64)	0
********** Pooling 4 **********		
Conv2D	(None, 15, 15, 32)	18464
MaxPooling2D	(None, 7, 7, 32)	0
Dropout	(None, 7, 7, 32)	0
********** Pooling 5 **********		
Conv2D	(None, 5, 5, 16)	4624
MaxPooling2D	(None, 2, 2, 16)	0
Dense	(None, 2, 2, 16)	272
Flatten	(None, 64)	0
Dense	(None, 2)	130
SoftMax	(None, 2)	6

Figure 7. The architecture of the CNN model and its layer-specific parameter settings.

IV. Results and Discussion

The predictive performance of the CNN model is evaluated using measures such as accuracy, precision, recall, and F1, which are derived from the confusion matrix. The accuracy, being the foremost measure of interest, is calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Furthermore, it is possible to define class-specific recall metrics, which can be expressed by:

$$recall_1 = \frac{TP}{TP + FN}, recall_0 = \frac{TN}{TN + FP}$$

Similarly, class-specific precision measures can also be estimated using the following equations.

$$precision_1 = \frac{TP}{TP+FP}, precision_0 = \frac{TN}{TN+FN}$$

With the knowledge of both precision and recall measures, it is possible to determine the class specific F1 values through the following expressions.

$$F1_1 = \frac{2\ x\ precision_1\ x\ recall_1}{precision_1 + recall_1}, F1_0 = \frac{2\ x\ precision_0\ x\ recall_0}{precision_0 + recall_0}$$

The initial part of this section entails the outcomes related to accuracy and loss of the CNN model throughout training and validation stages. Fig. 10 illustrates these metrics for the CNN model that was developed using the initial training set X_{train} (consists of only 66 images, including 39 explosive and 27 non-explosive ones). The accuracy of the resulting model exhibits the anticipated trend of gradual incline with the increase in the number of epochs (from 1 to 100). This was also apparent in the case of the model loss, which decreases rapidly as the model learned more from input images. However, both validation accuracy and loss were found to be worse than those of their training counterparts, indicating the poor performance on new cases that were not included in the training set. Besides, the accuracy and loss curves in the figure depict high fluctuations, suggesting that additional training samples would be required to stabilize the underlying learning process [55]. Having completed this with the original training set, the same experiment is repeated with those five augmented sets, namely $X_{train-500}$, $X_{train-1000}$, $X_{train-2000}$, $X_{train-3000}$ and $X_{train-4000}$, respectively.

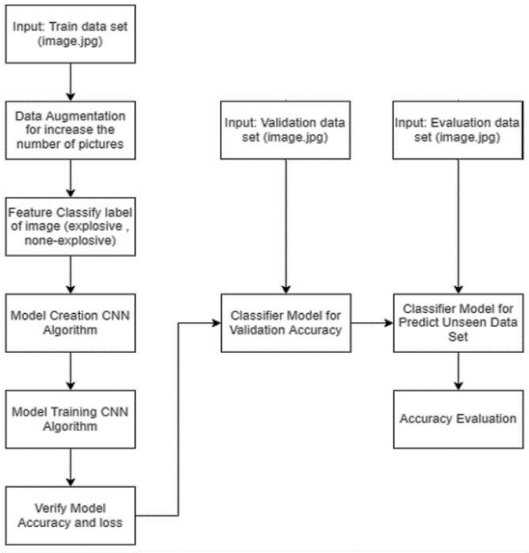

Figure 8. Overview of the model development and evaluation.

	Actual Values: Actually Positive (1) : explosive	Actual Values: Actually Negative (0) : none
Predicted Positive (1): explosive	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0):none	False Negatives (FNs)	True Negative (TNs)

Figure 9. Illustration of a confusion matrix for this binary classification problem.

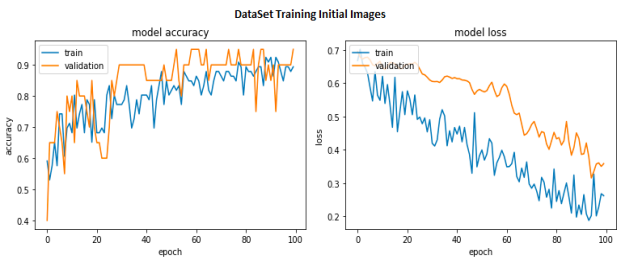

Figure 10. Comparison of accuracy & loss from initial training.

Figs. 11–13 illustrate the trends in model accuracy and loss using augmented train sets of 500, 1000, and 2000 samples. Compared to the previous experiment, these newly observed patterns of training accuracy and loss have become more stable, while the validation measures have slightly improved from those presented in Fig. 10.

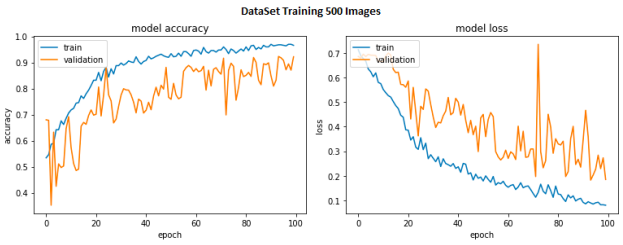

Figure 11. Comparison of accuracy & loss from set $X_{train-500}$.

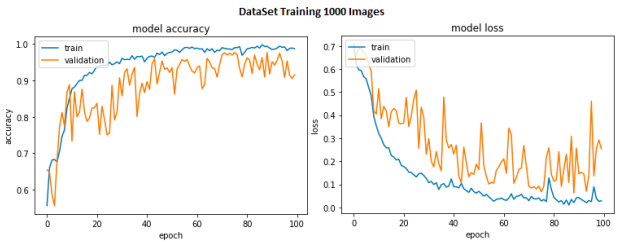

Figure 12. Comparison of accuracy & loss from set $X_{train-1000}$.

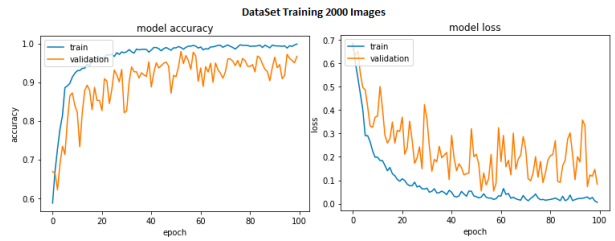

Figure 13. Comparison of accuracy & loss from set $X_{train-2000}$.

Figure 14. Comparison of accuracy & loss from set $X_{train-3000}$.

Figure 15. Comparison of accuracy & loss from set $X_{train-4000}$.

Figs. 14 and 15 depict that validation accuracy and loss have aligned with their respective training measures, indicating the stability of learned models. However, this new discovery has been achieved by increasing the size of the train set to 3000 and 4000 images, respectively. Despite the trends exhibited in these figures, there is a risk of overfitting as the model is trained with up to 4000 instances. To address this concern, all the five models developed from augmented train sets are evaluated against the unseen dataset, and the corresponding results reported in Table VI are obtained after 100 epochs.

Based on the results shown in Table VI, it can be observed that among the five CNN models developed using augmented training sets, $X_{train-3000}$ and $X_{train-2000}$ exhibit the most optimal predictive performance. Specifically, the former achieves the highest accuracy of 0.985, while the latter has a slightly lower but still favorable accuracy of 0.982. These two models also demonstrate similar trends with respect to other quality metrics such as F1 and Recall, which are specific to class 1. However, it is worth noting that the model trained with $X_{train-4000}$ achieves the best precision value of 0.986, indicating a lower number of false positives than the other models. Nevertheless, this model's recall capability is relatively weaker, making it less effective in recognizing explosive images, which is the main focus of this study, as evidenced by its higher FN value.

TABLE VI. CONFUSION MATRICES AND CORRESPONDING MEASURES OF PREDICTIVE PERFORMANCE WERE OBTAINED FROM FIVE AUGMENTED TRAIN SETS

Train set	TP	FP	TN	FN	Pc	Rc	F1	Accuracy
X-500	336	39	340	35	0.896	0.905	0.900	0.901
X-1000	357	18	360	15	0.952	0.959	0.955	0.956
X-2000	367	8	370	5	**0.978**	**0.986**	**0.982**	**0.982**
X-3000	367	8	372	3	**0.978**	**0.991**	**0.985**	**0.985**
X-4000	370	5	357	18	**0.986**	0.953	0.969	0.969

Note: Pc = Precision / Rc= Recall / F1 are Reported for Class 1 only here, and the Best Two Scores for Each Metric are Highlighted in Bold Font.

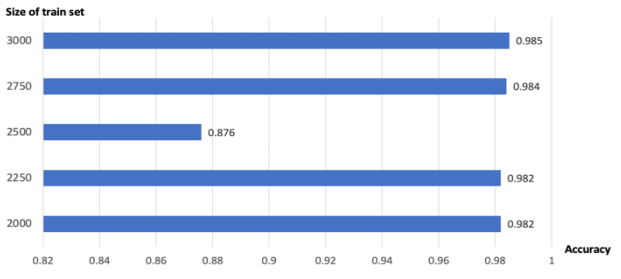

Figure 16. Comparison of accuracy measures from different sizes of train sets.

These results indicate that data augmentation has proven effective to substantially improve the CNN model's learning outcome. However, excessively large, augmented train sets may not be necessary to achieve the optimal assessment measures. Thus, further experiments were conducted to investigate the appropriate size of the train set, ranging from 2000 to 3000 samples, to minimize resource consumption while maintaining a high level of predictive quality. The new train sets generated are $X_{train-2250}$, $X_{train-2500}$, and $X_{train-2750}$. Fig. 16 illustrates the accuracy measures obtained by the CNN model trained with these new augmented sets. It suggests that the training process converges around a train set of 2750 to 3000 images, whereas a smaller number of images

might lead to an unstable outcome, especially in the case of the set with 2500 samples. Therefore, the aforementioned range is recommended for anyone willing to extend this work. It is of utmost significance to acknowledge that a larger training set can result in a wastage of resources during the training phase and often leads to overfitting, which ultimately diminishes the accuracy of predictions for new instances.

In addition to the results reported thus far, it is also important to match them to a benchmark method in the literature, thus leading to the next comparison between the CNN model built within this project and the conventional ANN: artificial neural network. Another experiment is conducted to train the ANN model of two hidden layers (using ReLu as the default activation function, SoftMax as the function in the output layer, SGD as the optimizer with the learning rate of 0.001) using the augmented training set of 3000 images and the same epoch setting of 100. Based on the comparative results shown in Table VII, it is clear that the ANN model tends to overfit with training samples, i.e., it may not be suitable to accommodate unseen data. To be exact, the ANN model cannot recall any of explosive images in the unseen set (i.e., Rc = 0.000), despite a very high measure of 0.858 has been obtained with the validation set. Note that the precision or Pc of this model on the unseen set is inherently zero as well provided that recall. In particular, the four components of its confusion matrix are: TP = 0, FP = 0, TN = 235, and FN = 515, respectively. Given these, it is justifiable for this study to employ a CNN model explicitly designed for image processing and offers superior efficacy in handling image classification tasks [68–70]. The difference witnessed here may be due to the way input features are represented and learned (or not learned). At first, the CNN makes use of kernels or filters to extract features in which pixels are spatially related, whilst each pixel is considered as a feature by the ANN. Of course, one of feature extraction methods can be used to generate a set of more discriminative features, but they are normally subjected to some assumptions that are not generalised across problems [71]. Also, these features are specified upfront, thus requiring an expertise to foresee an optimal setting. On the other hand, it is possible that the relevance of those filter-driven features can be learned in parallel as the CNN adjusts its algorithmic variables to minimize the difference between its prediction and the truth.

TABLE VII. SUMMARY OF PREDICTION PERFORMANCE MEASURES OBTAINED BY ANN AND CNN MODELS, USING THE AUGMENTED TRAINING SET WITH 3000 SAMPLES AND THE EPOCH OF 100

Model	Train set	Pc	Rc	F1	Accuracy
ANN-Validation data	X-3000	0.700	0.980	0.821	0.858
ANN-Unseen data	X-3000	0.000	0.000	0.000	0.314
CNN-Unseen data	X-3000	0.986	0.991	0.985	0.985

Based on the results presented and discussed in this section, it is reasonable to suggest that the CNN model developed herein performs well with the validation set and accurate for the unseen samples. Despite this, it can be widely agreed that the training step is usually resource demanding, with the implementation in this work has been on the platform of Google Colab. With this cloud-based service or a privately-owned computing stack, the task can be accomplished in the background, where the model can be re-learned with an updated set of training inputs. However, its application would be a lot quicker to produce a prediction of a new case. In practice, the learned model can be installed as another working module in the computing unit of a robot, where a newly captured x-ray image can be stored and fed to the model. This may provide a seamless flow of control from sensing to decision making at the point of operation. Nonetheless, it requires a careful design that allows the underlying model to be updated locally as a new version becomes available. This may lead to another complication given a large number of robots in operation. Yet, since they are deployed in a dangerous and hostile environment, their processing units and data maintained within may be lost altogether. To this end, another innovative approach has been introduced, where x-ray images captured by a robot are relayed to a computing station in the command post. Then, the underlying classification procedure is conducted in a much safer setting, where it can be realized as a locally installed program or a cloud-based service.

V. CONCLUSION AND FUTURE WORK

This article details a novel approach to the detection of explosives, developed as a collaborative study between Mae Fah Luang University and the Defense Technology Institute (Public Organization). The research utilizes a CNN model to classify images as either explosive or non-explosive, a crucial component in the identification of IEDs. This task is particularly challenging due to the limited number of images available, which necessitates the development of new methodologies to increase the image volume. To this end, data augmentation has been employed, incorporating three distinct processes of image rotation, flipping, and brightness adjustment. Following the acquisition of additional images, the researchers must ascertain the optimal number of images to be utilized during training. Excessively large training sets risk overlearning, leading to the consumption of excessive time and resources without an expected increase in accuracy. As such, a balance must be struck between maximizing accuracy and minimizing the required training time and resources. In the coming years, there will be a concerted effort to develop detection methods that are capable of targeting specific areas within an image rather than analyzing the entire image indiscriminately. Possible future works are summarized below.

To enhance the identification of explosive materials, one possible approach is to employ the use of OpenCV, a computer vision library, to improve the ability to focus on relevant areas of interest within the image before proceeding to the core analysis. This technology offers an opportunity to selectively highlight features that are relevant to the presence of explosives, thus improving the accuracy of detection.

Another potential avenue for future research is to employ transfer learning techniques to may enhance the current models' performance. By utilizing similar datasets,

such as x-ray images of various weapons, the model can be augmented, potentially enabling them to recognize features that are commonly presented across different types of explosive devices.

Lastly, it may be interesting to explore the technique of Generative Adversarial Networks or GANs to simulate the reconstruction of input images. GANs are capable of identifying distinctive features within a given image and generating a new image instance based on those learned features. It differs from the concept of data augmentation and may provide an alternative means of augmenting a limited dataset.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Chakkaphat Chamnanphan: Data curation, visualization, investigation, software, writing—original draft preparation. Surapol Vorapatratorn: methodology, validation, writing—reviewing and editing, supervision. Khwunta Kirimasthong: methodology, writing—reviewing and editing, supervision. Tossapon Boongoen: Conceptualization, methodology, writing—original draft preparation. Natthakan Iam-On: Validation, writing—reviewing and editing. All authors had approved the final version.

FUNDING

This research project is funded by Defense Technology Institute (DTI), the grant DTI-MFU2562/00: An automated detection of explosive units using x-ray images. Also, the publication is funded by Mae Fah Luang University.

ACKNOWLEDGMENT

This research is part of an MSc study at Mae Fah Luang University. The authors express their gratitude to the Defense Technology Institute (DTI) and the Department of Ordnance, Royal Thai Air Force (RTAF), for their generous contribution to the image dataset utilized in this research study.

REFERENCES

[1] Z. Allam and Z. A. Dhunny, "On big data, artificial intelligence and smart cities," *Cities*, vol. 89, pp. 80–91, June 2019.

[2] Z. Allam and P. Newman, "Redefining the smart city: Culture, metabolism and governance," *Smart Cities*, vol. 1, pp. 4–25, July 2018.

[3] R. Petrolo, V. Loscrì, and N. Mitton, "Towards a smart city based on cloud of things: A survey on the smart city vision and paradigms," *Transactions on Emerging Telecommunications Technologies*, vol. 28, p. e2931, Jan 2017.

[4] R. P. Dameri, "Searching for smart city definition: a comprehensive proposal," *International Journal of Computers and Technology*, vol. 11, pp. 2544–2551, Oct 2013.

[5] P. Neirotti, A. D. Marco, A. C. Cagliano, G. Mangano, and F. Scorrano, "Current trends in smart city initiatives: some stylised facts," *Cities*, vol. 38, pp. 25–36, June 2014.

[6] T. Boongoen, Q. Shen, and C. J. Price, "Disclosing false identity through hybrid link analysis," *Artificial Intelligence and Law*, vol. 18, no. 1, pp. 77–102, Feb. 2010.

[7] K. M. Carley, "Destabilization of covert networks," *Computational & Mathematical Organization Theory*, vol. 12, no. 1, 51–66, April 2006.

[8] E. M. Archer, "Crossing the Rubicon: Understanding cyber terrorism in the European context," *The European Legacy*, vol. 19, no. 5, pp. 606–621, July 2014.

[9] H. J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, *et al*, "Using Raman spectroscopy to characterize biological materials," *Nature Protocols*, vol. 11, no. 4, pp. 664–687, Mar 2016.

[10] Z. Bielecki, J. Janucki, A. Kawalec, J. Mikołajczyk, N. Pałka, *et al*, "Sensors and systems for the detection of explosive devices: An overview," *Metrology and Measurement Systems*, vol. 19, no. 1, pp. 3–28, Mar 2012.

[11] Y. Salinas, R. Martinez-Manez, M. D. Marcos, F. Sancenon, A. M. Costero, *et al*, "Optical chemo sensors and reagents to detect explosives," *Chemical Society Reviews*, vol. 41, no. 3, pp. 1261–1296, 2012.

[12] N. Hättenschwiler, M. Mendes, and A. Schwaninger, "Detecting bombs in x-ray images of hold baggage: 2D versus 3D imaging," *Human Factors*, vol. 61, no. 2, pp. 305–321, Mar 2019.

[13] N. Jenne and J. Chang, "Hegemonic distortions: The securitisation of the insurgency in Thailand's deep south," *TRaNS: Trans - Regional and -National Studies of Southeast Asia*, vol. 7, no. 2, pp. 209–232, Jan 2019.

[14] A. Tunwannarux and S. Tunwannarux, "The explosive ordnance disposal robot: CEO mission EOD," in *Proc. 10th WSEAS International Conference on Automatic Control, Modelling & Simulation*, 2008, pp. 433–438.

[15] T. Boongoen, N. Iam-On, and B. Undara, "Improving face detection with bi-level classification model," *NKRAFA Journal of Science and Technology*, vol. 12, pp. 52–63, Dec 2016.

[16] N. Iam-On and T. Boongoen, "Improving face classification with multiple-clustering induced feature reduction," in *Proc. 49th Annual International Carnahan Conference on Security Technology*, 2015, pp. 241–246.

[17] U. Thongsatapornwatana, W. Lilakiatsakun, and T. Boongoen, "Improvement of intelligent framework for suspect vehicle detection system," in *Proc. 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 2017, pp. 419–422.

[18] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Computer Science Review*, vol. 40, 100379, May 2021.

[19] M. Ramprasath, M. V. Anand, and S. Hariharan, "Image classification using convolutional neural networks," *International Journal of Pure and Applied Mathematics*, vol. 119, no. 17, pp. 1307–1319, 2018.

[20] E. R. Williams and R. N. Zare, "Detection of concealed explosive," *Science*, vol. 248, no. 4962, pp. 1471–1472, Jun 1990.

[21] L. Jun, T. Si, and Z. Zhang, "Mussel-inspired immobilization of silver nanoparticles toward sponge for rapid swabbing extraction and SERS detection of trace inorganic explosives," *Talanta*, vol. 204, pp. 189–197, Nov 2019.

[22] D. R. Ifa, N. E. Manicke, A. L. Dill, and R. G. Cooks, "Latent fingerprint chemical imaging by mass spectrometry," *Science*, vol. 321, no. 5890, pp. 805–805, Aug 2008.

[23] Y. Ding, S. Wang, J. Li, and L. Chen, "Nanomaterial-based optical sensors for mercury ions," *TrAC Trends in Analytical Chemistry*, vol. 82, pp. 175–190, Sep 2016.

[24] M. H. Wong, J. P. Giraldo, S. Y. Kwak, V. B. Koman, R. Sinclair, *et al*, "Nitroaromatic detection and infrared communication from wild-type plants using plant nano bionics," *Nature Materials*, vol. 16, pp. 264–272, Feb. 2017.

[25] J. M. Wolfe and M. J. V. Wert, "Varying target prevalence reveals two, dissociable decision criteria in visual search," *Current Biology*, vol. 20, pp. 121–124, Jan 2010.

[26] D. Mery, D. Saavedra, and M. Prasad, "X-ray baggage inspection with computer vision: A survey," *IEEE Access*, vol. 8, pp. 145620–145633, Aug 2020.

[27] M. Bastan, M. R. Yousefi, and T. M Breuel, "Visual words on baggage x-ray images," in *Proc. 14th International Conference on Computer Analysis of Images and Patterns*, 2011, pp. 360–368.
[28] D. G. Lowe, "Distinctive image features from scale-invariant key points," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov 2004.
[29] D. Mery, "X-ray testing by computer vision," in *Proc. Conference on Computer Vision and Pattern Recognition*, 2013, pp. 360–367.
[30] M. Mansoor and R. Rajashankari, "Detection of concealed weapons in x-ray images using fuzzy K-NN," *International Journal of Computer Science, Engineering and Information Technology*, vol. 2, no. 2, pp. 187–196, April 2012.
[31] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in *Proc. 2003 Conference on the Move to Meaningful Internet*, 2003, pp. 986–996.
[32] D. Turcsany, A. Mouton, and T. Breckon, "Improving feature-based object recognition for x-ray baggage security screening using primed visual words," in *Proc. 2013 International Conference on Industrial Technology*, 2013, pp. 1140–1145.
[33] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, June 2008.
[34] M. Bastan, W. Byeon, and T. Breuel, "Object recognition in multi-view dual energy x-ray images," in *Proc. British Machine Vision Conference*, 2013, pp. 1–11.
[35] T. Evgeniou and M. Pontil, "Support vector machines: Theory and applications," *Advanced Course on Artificial Intelligence*, pp. 249–257, Jan 2001.
[36] A. Mouton, T. Breckon, G. Flitton, and N. Megherbi, "3D object classification in baggage computed tomography imagery using randomised clustering forests," in *Proc. 2014 International Conference on Image Processing*, 2014, pp. 5202–5206.
[37] I. Nedjar, M. E. H. Daho, N. Settouti, S. Mahmoudi, and M. A. Chikh, "Random forest based classification of medical x-ray images using a genetic algorithm for feature selection," *Journal of Mechanics in Medicine and Biology*, vol. 15, no. 02, pp. 15400251–15400258, Mar. 2015.
[38] S. Akcay, M. Kundegorski, M. Devereux, and T. Breckon, "Transfer learning using convolutional neural networks for object classification within x-ray baggage security imagery," in *Proc. 2016 International Conference on Image Processing*, 2016, pp. 1057–1061.
[39] 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 Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2203–2215, Sep. 2018.
[40] D. Mery, E. Svec, M. Arias, V. Riffo, J. M. Saavedra, *et al*, "Modern computer vision techniques for x-ray testing in baggage inspection," *IEEE Transactions on Systems, Man and Cybernetics (System)*, vol. 47, no. 4, pp. 682–692, April 2017.
[41] V. Riffo, S. Flores, and D. Mery, "Threat objects detection in x-ray images using an active vision approach," *Journal of Nondestructive Evaluation*, vol. 36, no. 3, p. 44, May 2017.
[42] M. Xu, H. Zhang, and J. Yang, "Prohibited item detection in airport x-ray security images via attention mechanism based CNN," in *Proc. Chinese Conference on Pattern Recognition and Computer Vision*, 2018, pp. 429–439.
[43] Y. Falinie, A. Gaus, N. Bhowmik, S. Akçay, P. M. Guillén-Garcia, *et al.*, "Evaluation of a dual convolutional neural network architecture for object-wise anomaly detection in cluttered x-ray security imagery," in *Proc. 2019 International Joint Conference on Neural Networks*, 2019, pp. 1–8.
[44] S. R. Klomp and D. van de Wouw, "Real-time small-object change detection from ground vehicles using a Siamese convolutional neural network," *Journal of Imaging Science and Technology*, vol. 63, no. 6, pp. 1–16, Nov. 2019.
[45] L. J. Ba and R. Caruana, "Do deep nets really need to be deep?" in *Proc. International Conference on Neural Information Processing Systems*, 2014, pp. 2654–2662.
[46] A. Mosca and G. D. Magoulas, "Customised ensemble methodologies for deep learning: Boosted residual networks and related approaches," *Neural Computing and Applications*, vol. 31, pp. 1713–1731, June 2019.
[47] M. M. Rahman, M. S. Islam, R. Sassi, and Aktaruzzaman, "Convolutional neural networks performance comparison for handwritten Bengali numerals recognition," *SN Applied Sciences*, vol. 1, p. 1660, Nov 2019.
[48] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolution neural networks," in *Proc. 2012 Advances in Neural Information Processing Systems*, 2012, pp. 1106–1114.
[49] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, *et al*., "Going deeper with convolutions," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–8.
[50] N. Srivastava, G. Hinton, A. Krizhevsky and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
[51] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. 2014 IEEE International Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1717–1724.
[52] A. Vyas, S. Yu, and J. Paik, "Fundamentals of digital image processing," *Multiscale Transforms with Application to Image Processing*, 2018, pp. 3–11.
[53] L. Sifre and S. Mallat, "Rotation, scaling and deformation invariant scattering for texture discrimination," in *Proc. 2013 IEEE International Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1233–1240.
[54] J. Premaladha and K. S. Ravichandran, "Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms," *Journal of Medical Systems*, vol. 40, no. 4, p. 96, Feb. 2016.
[55] 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, Mar. 2019.
[56] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, p. 60, 2019.
[57] N. E. Khalifa, M. Loey, and S. Mirjalili, "A comprehensive survey of recent trends in deep learning for digital images augmentation," *Artificial Intelligence Review*, vol. 55, pp. 2351–2377, Mar 2022.
[58] Y. Zhang, S. Wang, H. Zhao, Z. Guo, and D. Sun, "CT image classification based on convolutional neural network," *Neural Computing and Applications*, vol. 33, no. 14, pp. 8191–8200, July 2021.
[59] N. E. Khalifa, M. Loey, and S. Mirjalili, "A comprehensive survey of recent trends in deep learning for digital images augmentation," *Springer Link Artificial Intelligence Review*, vol. 55, pp. 2351–2377, Mar. 2022.
[60] M. D. Bloice, C. Stocker, and A. Holzinger, "Augmentor: An image augmentation library for machine learning," arXiv preprint, arXiv:1708.04680, Aug. 2017.
[61] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A comprehensive survey of image augmentation techniques for deep learning," arXiv preprint, arXiv: 2205.01491, Nov 2022.
[62] T. Liu, S. Fang, Y. Zhao, P. Wang, and J. Zhang. "Implementation of Training Convolutional Neural Networks," arXiv preprint, arXiv:1506.01195, Jun 2015.
[63] A. N. Gomez, I. Zhang, S. R. Kamalakara, D. Madaan, K. Swersky, Y. Gal, and G. E. Hinton, "Learning sparse networks using targeted dropout," arXiv preprint, arXiv:1905.13678, Sep. 2019.
[64] S. Vani and T. V. M. Rao. "An experimental approach towards the performance assessment of various optimizers on convolutional neural networks," in *Proc. IEEE 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2019, pp. 331–336.
[65] A. M. Taqi, A. Awad, F. Al-Azzo, and M. Milanova, "The impact of multi-optimizers and data augmentation on tensor flow convolutional neural network performance," in *Proc. IEEE 2018 Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2018, pp. 140–145.
[66] Y. Ito, "Representation of functions by superpositions of a step or sigmoid function and their applications to neural network theory," *Science Direct Access*, vol. 4, pp. 385–394,1991.
[67] V. M. Patro and M. R. Patra, "Augmenting weighted average with confusion matrix to enhance classification accuracy," *Society for Science and Education United Kingdom*, vol. 2, no. 4, pp. 77–91, 2014.
[68] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of Artificial Neural Network (ANN) modeling and its application in

pharmaceutical research," *J. Pharm. Biomed. Anal.*, vol. 22, no. 5, pp. 717–727, June 2000.

[69] G. Zhang, M. Y. Hu, B. E. Patuwo, and D. C. Indro, "Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis," *Eur. J. Oper. Res.*, vol. 116, no. 1, pp. 16–32, July 1999.

[70] C. Zanchettin and T. B. Ludermir, "A methodology to train and improve artificial neural networks' weights and connections," in *Proc. 2006 International Joint Conference on Neural Networks, IJCNN'06*, 2006, pp. 5267–5274.

[71] S. Amraee, M. Chinipardaz, and M. Charoosaei, "Analytical study of two feature extraction methods in comparison with deep learning methods for classification of small metal objects," *Vis. Comput. Ind. Biomed. Art*, vol. 5, no. 13, pp. 1–13, May 2022.