Received October 22, 2020, accepted November 3, 2020, date of publication November 16, 2020,
date of current version November 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3037719

O-Net: Dangerous Goods Detection in Aviation Security Based on U-Net

WOONG KIM[1], (Member, IEEE), SUNGCHAN JUN[1], SUMIN KANG[2], AND CHULUNG LEE[3]
[1]Department of Industrial Management Engineering, Korea University, Seoul 02841, South Korea
[2]Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109, USA
[3]School of Industrial Management Engineering, Korea University, Seoul 02841, South Korea

Corresponding author: Chulung Lee (leecu@korea.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (Ministry of Science and ICT) under Grant NRF-2020R1F1A1076812.

ABSTRACT Aviation security X-ray equipment currently searches objects through primary screening, in which the screener has to re-search a baggage/person to detect the target object from overlapping objects. The advancements of computer vision and deep learning technology can be applied to improve the accuracy of identifying the most dangerous goods, guns and knives, from X-ray images of baggage. Artificial intelligence-based aviation security X-rays can facilitate the high-speed detection of target objects while reducing the overall security search duration and load on the screener. Moreover, the overlapping phenomenon was improved by using raw RGB images from X-rays and simultaneously converting the images into grayscale for input. An O-Net structure was designed through various learning rates and dense/depth-wise experiments as an improvement based on U-Net. Two encoders and two decoders were used to incorporate various types of images in processing and maximize the output performance of the neural network, respectively. In addition, we proposed U-Net segmentation to detect target objects more clearly than the You Only Look Once (YOLO) of Bounding-box (Bbox) type through the concept of a ''confidence score''. Consequently, the comparative analysis of basic segmentation models such as Fully Convolutional Networks (FCN), U-Net, and Segmentation-networks (SegNet) based on the major performance indicators of segmentation-pixel accuracy and mean-intersection over union (m-IoU)-revealed that O-Net improved the average pixel accuracy by 5.8%, 2.26%, and 5.01% and the m-IoU was improved by 43.1%, 9.84%, and 23.31%, respectively. Moreover, the accuracy of O-Net was 6.56% higher than that of U-Net, indicating the superiority of the O-Net architecture.

INDEX TERMS Artificial intelligence security system, aviation security, detection algorithm, image segmentation, U-Net, X-ray detection.

I. INTRODUCTION

The aviation industry is steadily growing owing to the increasing number of passengers and the volume of air cargo transportation trade; therefore, the global aviation industry continues to profit [1]. To accommodate the growing demand for a wide range of passengers, the number of air-ways between cities has increased rapidly, and the airline has a record occupancy rate of 81.9% in the passenger sector [2]. Airlines and airports are committed to providing safe services for passengers from their point of departure to their destination along with the safe transportation of luggage. Thus, aviation safety was mainly focused on operational safety. However, after the awareness of the enormous human and economic losses caused by aircraft terrorism in the September 11 attacks, countries around the world, began to strengthen their aviation security against dangerous goods in the plane or cargo transportation [3].

Since then, airports worldwide have put safety first and are focusing on aviation security as preparation for emergencies such as aircraft terrorism, aircraft hijackings, and aircraft explosions. Among the preparations for countermeasures against aircraft terrorism, X-ray screening systems, have been employed to reinforce transport security by limiting dangerous goods in carry-on baggage and ensuring safe air-cargo transportation. However, as the aviation security process is a continuously operating 24×7 system, the screeners' fatigue level increases; moreover, the X-ray overlapping

The associate editor coordinating the review of this manuscript and approving it for publication was Victor Hugo Albuquerque.

phenomenon exacerbates fatigue. In addition, an increase in the fatigue level of the screener acts as a factor of increasing false alarms of dangerous goods [4]. The results of thorough baggage inspection have reduced overseas travel and the key airline factors such as costs of changes in airport entry and exit and losses led by changes in flight schedules have increased the relevance of security in the aviation industry [5], [6]. The aviation security screening personnel trained in air-baggage sorting work closely monitor X-ray images to classify dangerous goods; the deployed machine can automatically detect explosive images or dangerous goods [7], [8]. Although the security control system is said to have developed to a professional level, the real error rate of identification, discrimination, and classification of dangerous goods increases the work stress and fatigue of aviation security personnel owing to numerous immigration and intelligent terrorist approaches [9]–[11]. Artificial intelligence-based X-ray scanning, which includes scientific system design, can quickly and accurately identify dangerous goods in the baggage beyond the limits of human abilities for ensuring the safety of aviation security.

Aviation security image search demonstrates the following technologies: biomedical image scanning technology, video analytics technology, and real-time image processing technology. The technology using biometrics and images constitutes a scanning technology that can be navigated to all areas of the body while conducting simple security checks for passengers [12]. Although this shows a high performance of the object search method, it has the disadvantage of lowering passenger satisfaction through the invasion of personal privacy [13], [14]. Furthermore, the introduction of an image analysis technology utilizing OpenCV Library, several algorithms such as image conversion, pattern recognition, and noise control presents a general image processing technology and supports real-time image processing that can be applied on various platforms. Image search technology in the field of computer vision plays a significant role in the aviation security industry.

As mentioned above, with the development of computing technology, aviation security systems are being transformed into intelligent security systems. In this research, we applied an artificial neural network to fit such trends attempted to establish an algorithm for the automatic detection of dangerous goods from X-ray screening images.

II. LITERATURE REVIEW

A. TYPE OF AVIATION SECURITY SYSTEMS

An aviation security system uses technology to prevent illegal activities that pose a danger to the safety of human life and property, risk the maintenance of safety in civil aviation operations or have a serious influence on the performance of aviation tasks. The types of security equipment used in aviation security systems are classified based on scanning ''people'' or ''objects''. An aviation security equipment searches the airport passengers, carry-on baggage, checked baggage, and cargo to detect dangerous goods or dangerous substances. These search equipment block any unwanted situation arising out of the malevolent use of such dangerous goods and prevent occurrences of accidents in airports and aircrafts. Human searches devices are classified into hand-held metal detectors [15] and walk-through metal detectors [16]. Hand-held metal detector is a search device that detects a metal objects by using an electromagnetic field and is safe for finding metal objects hidden on the body. Similarly, walk-through metal detector is a search device that detects a metal object by using an electromagnetic field to find such objects hidden by passengers. Therefore, these machines will accurately identify objects and inform the screener of what the person is carrying.

Object search devices include X-ray screening systems [15], whole-body scanners [16], explosive detecting systems [17], and explosive trace detectors [18]. X-ray screening equipment is a search device that uses an X-ray system to search a target and display the contents of the search on a monitor. A whole-body scanner detects dangerous objects such as weapons and explosives that are difficult to detect with a metal detector without touching the body and displays them on a monitor. The explosive detecting system is a device that inhales the chemical substances hidden on the investigation target and uses chemical-ion analysis to detect explosives. Similarly, an explosive trace detector (ETD) is a device that accurately detects and identifies particles of explosives contained in carry-on or checked baggage and air cargo and informs the screener of the objects contained therein.

Therefore, the application of these devices for aviation security systems depends on the type and purpose of the search object. Although most of the equipment detects the presence of dangerous goods, X-ray search equipment can capture an image of the contents present inside baggage; the screener can then directly check for the existence of dangerous goods. As the process of human identification causes mistakes, there is a high probability of false positives. Therefore, in order to strengthen the aviation security process, we intend to apply an image recognition algorithm that detects dangerous objects via X-ray images.

B. IMAGE RECOGNITION AND DETECTION

An X-ray image is represented based on X-ray transmittance. Areas without material or areas of very low density are displayed in white, and areas of high density are displayed in saturated colors [19]. For overlapping objects, X-rays pass through all the objects and are diagrammed based on the degree of transmission, so all the information is expressed in the X-ray image even in overlapping conditions. As a result, X-ray video images have the problems of view difficulty, complexity, and superposition [20]–[22], and the ''overlapping'' phenomenon increases the screener's stress. In severe cases, dangerous goods go undetected and create a major problem in aviation security [9]. To solve the abovementioned problems, various algorithms are applied, such as detection

TABLE 1. In the aviation security system, the types of aviation security equipment.

	Type	Description
People search	Hand-held Metal Detector [15]	Detects a metal object by using an electromagnetic field and is safe for finding such objects hidden on the body.
	Walk-through Metal Detector [16]	Detects metal objects hidden by passengers using an electromagnetic field.
Objects search	X-ray Screening System [15]	An X-ray system is used to search a target, and the content of the search is displayed on a monitor.
	Whole Body Scanner [16]	Detects dangerous objects such as weapons and explosives that are difficult to detect with a metal-detector without touching the body and displays them on a monitor.
	Explosive Detecting System [17]	Inhales chemical substances buried in the investigation target and uses chemical-ion analysis to detect explosives.
	Explosive Trace Detector [18]	Detects particles of explosives contained in carry-on or checked baggage and air cargo. It accurately identifies and informs the screener of the objects contained in the baggage and cargo.

based on X-ray transmittance information [23], [24] and the development of SIFT and SURF algorithms for extracting image features [25], [26]. However, the problem pertaining to the overlapping phenomenon is still not resolved. To overcome it, we can apply artificial intelligence-based image recognition technology to significantly increases the object detection performance of such security systems [27].

The performance of artificial intelligence has dramatically improved with the advancement of computing speed and technology. In the image classification of the convolutional neural network (CNN) structure, a traditional LeNet model that recognizes hand-written numbers was proposed that compensated for the weak points of a topology change or noise immunity in the existing fully-connected neural network (FCNN) to perform a more accurate image recognition [28]. Thereafter, as the computer environment developed, a GPU specialized in parallel computation was able to perform a large amounts of computation at high speed, and AlexNet was developed [29]. In the LeNet and AlexNet structures, the ZFNet method, which adjusted the kernel layer size and the initial part of the layer in the Stride, showed higher classification accuracy [30], and then VGGNet, which has good performance for recognizing large-scale image data, was then developed by deepening the network and increasing the number of layers [31]. GoogleLeNet was developed using an inception module, which merged the results from the convolution and pooling layers executed in parallel [32]. Similarly, ResNet was developed by minimizing the residual as the network became deeper to reduce the learning error rate [33]. In addition, Inception-ResNet and Inception-v4, which added ResNet to GoogleLeNet, were developed to present a model structure with high speed and better performance [34]. As shown in image classification, the upgraded algorithm model achieves continuous performance improvement with an increase in the number of CNN-based convolution layers.

The simultaneous execution of classification and localization of various objects is essential for target object detection. First, R-CNN [35] uses ''Selective Search'' to find numerous objects in an image by region proposal or bounding-box (Bbox) and finally classifies the image with a support vector machine (SVM). Fast R-CNN [36] solved the three main disadvantages of R-CNN, i.e., the linear regression for Bbox, SVM for classification, and execution of CNN for every Bbox. Subsequently, Faster R-CNN [37], which improved the slow computation speed by including the region proposal network (RPN) method in Fast R-CNN, was developed from the R-CNN model series showing the basics of target detection in a two-stage detector method. Furthermore, Mask R-CNN [38], a type of instance segmentation, extracts more accurate pixel positions and speeds by adding a binary mask network that masks each pixel corresponding to an object. However, as this has a weak point in real-time object detection, a one-stage detector method of the YOLO model series was developed to compensate for Mask R-CNN. The YOLO model consists of three types—YOLOv1 [39] divides the image to be predicted into grid cells and predicts it as an object for each cell to display as an anchor box and simultaneously classify as a Bbox; YOLOv2 [40] was developed through a change in neural network model structure and stabilization of boundary boxes; subsequently, YOLOv3 [41] was presented by developing a more extensive dataset and a deeper network structure. As a technique for recognizing an image, an object detection method for locating an object using a Bbox as a target object and a segmentation method for recognizing the appearance of the target object in unit of pixels.

In particular, the image segmentation method can be divided into semantic and instance segmentations. Representatively, there are fully convolutional networks (FCN), SegNet, DeepLab, and U-Net models. First, semantic segmentation treats multiple objects of the same class as a single

TABLE 2. Classification of Image recognition and detection in artificial intelligence.

Type	Description	Reference
Convolutional Neural Network (CNN)	LeNet, AlexNet, ZFNet, VGGNet, GoogleLeNet, ResNet, Inception-ResNet, Inception v4	[28][29][30][31][32] [33][34]
Detection Algorithms	R-CNN, Fast R-CNN, Faster R-CNN, YOLO v1, v2, v3	[35][36][37][39][40][41]
Segmentation Algorithms	Mask R-CNN, FCN, SegNet, Deep Lab v1, v2, v3, v3+, U-Net	[38][42][43][44][45][46][47]

entity and aims to perform a dense prediction to classify every pixel in the image. FCN was developed through transfer learning from VGG16 to preserve the locational information in an image by changing the fully connected layer—the last layer—to a 1×1 convolution layer [42]. SegNet remembers the process of max-pooling in an encoder process without a skip connection process and is used in the up-sampling at the decoder to increase the location information [43]. DeepLab uses atrous convolution with various expansion ratios in parallel to capture more features, and high-performance V1, V2, V3, and V3+ have been developed by adding pooling techniques such as atrous spatial pyramid pooling, ResNet, and depth-wise separable convolution [43], [44], [44]–[46]. The basic U-Net was developed by constructing a neural network that can easily find a specific cell to be searched for in a transmitted light microscopy image [47]. The contents are summarized in Table 2.

C. BASELINE U-NET FOR SEMANTIC SEGMENTATION

A detection model delivers high accuracy in detecting an object from a general image, but the performance of the detection model is degraded when an object is detected in pixel units in an X-ray image. On the contrary, the segmentation model detects objects in pixel units, and U-Net is an artificial neural network that is typically used in medical X-ray images. U-Net describes a relationship between neighboring pixels, and it is possible to capture a context that identifies an image by viewing one of its parts as a contracting path and to perform a more accurate localization by combining the feature map and context through the expansive path. As it is a logically designed structure, various baseline U-Net architectures were developed based on the structure. W-Net of two-stage U-Net [48], Ladder-Net [49], and X-Net of one-stage U-Net [50] improved performance by repeatedly using the U-Net structure. V-Net [51] utilized the concepts of U-Net where 2-D image data was available, but the structure can be expanded to incorporate 3-D image data as well. Similarly, X-Net [52] and RAU-Net [53] improved the performance of the U-Net model for applications in medical data of various sizes by using a feature similarity module (FSM) block and an augmented attention module (AAM) block of an inception structure. Moreover, U-Net++ [54], R2U-Net [55], and MultiRes U-Net [56] applied skip connection and a recurrent/residual structure to improve the utilization of information and performance. Correspondingly, in order to improve the performance of the resulting image, U-Net for Pan-sharpening [57] and Feature-level U-Net [58] used the Pan-sharpening algorithm to enhance the quality of input data. In addition, information loss can be minimized through BRU-Net by applying input image data to every down-sampling process [59]. TPUAR-Net [60], which used a single-input image data by merging four different images among MRI data, improved the performance by using residual U-Net in parallel. Multispectral U-Net [61], modified U-Net [62], and dense multi-path U-Net [63] design multiple encoder structures utilize multiple input image data, and provide features of various input data. In contrast, dual U-Net [64] and W-Net of reinforced U-Net [65] used two decoder structures to improve their output performance, and 3-D MRI image of U-Net [66] also modified the decoder structure. However, the results based on the type of uncertainty were schematically illustrated together. Therefore, there have been studies based on changing the structure of U-Net in various ways, such as the modification or addition of an input image-process step.

In this study, we aimed to develop an algorithm that can detect dangerous goods, guns and knives, even if there were overlapping phenomena in the X-ray images. Alternatively, we studied the characteristics of X-ray images by designing two U-Nets with a parallel structure using two input images.

III. METHODOLOGY

This section elucidates the performance index to evaluate the O-Net architecture and network, which were developed based on a dangerous goods detection algorithm using X-ray screening images.

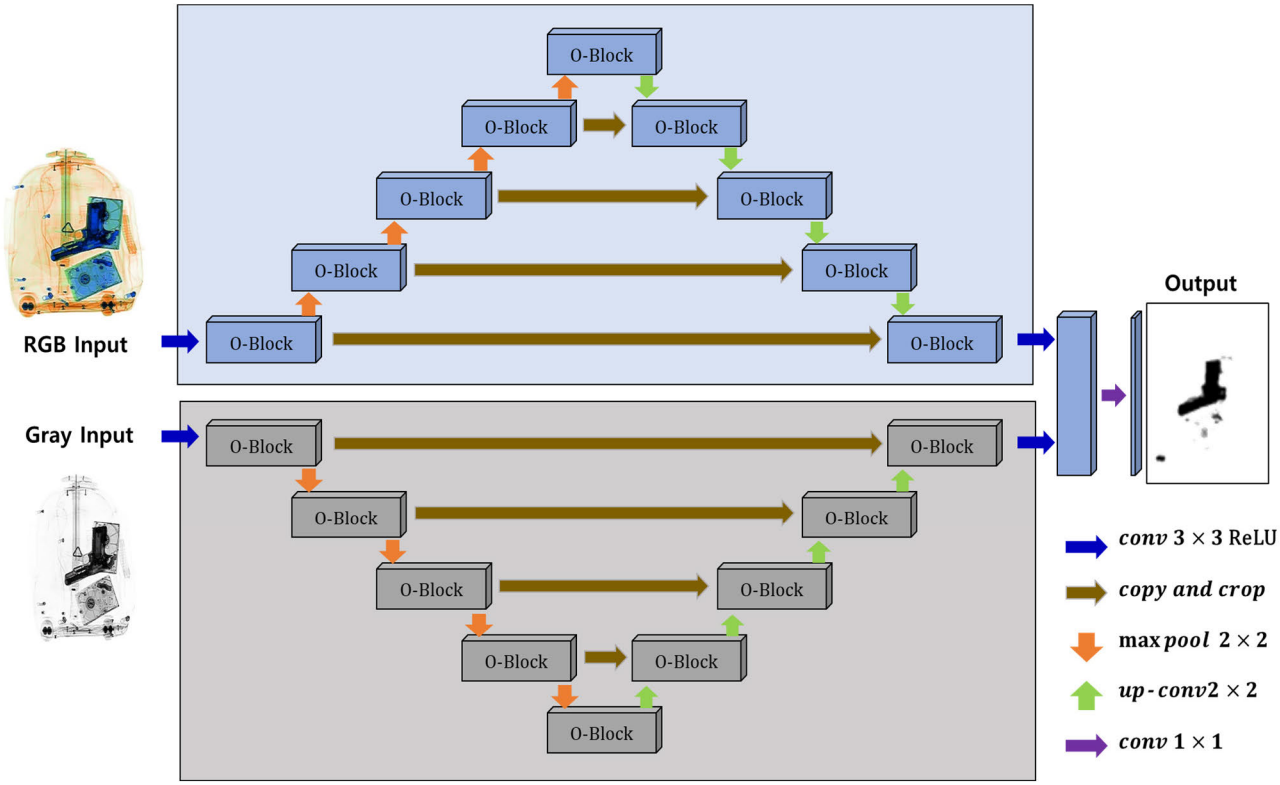

FIGURE 1. O-Net Structure: U-Net-based structure using an original X-ray image and a converted grayscale image as inputs.

The detection of dangerous goods from aviation security X-ray images becomes difficult when multiple objects are overlapping inside the baggage. Therefore, we used two input images to resolve this problem. One image was used as an input in the existing U-Net, whereas the image feature extraction was maximized by using two images as the input in the O-Net with the addition of a grayscale image. Moreover, as the extent of learning a target object through verification increased, the O-Net showed a higher pixel accuracy even for overlapping objects owing to the construction of an additional neural network.

A. O-NET ARCHITECTURE

The O-Net network is generally composed of a FCN based on the U-Net network of semantic segmentation; an encoder–decoder structure of image segmentation is depicted in Fig. 1. The existing U-Net has a structure in which a single-channel image format with a maximum input image size of 572×572 is fed to the network structure and delivered as a single image through image segmentation. In contrast, the structure of the O-Net uses an n×m size of random images, where three-channel and single-channel format images are processed to the network fixed to an input image size of 256×256, and the output is obtained as an image through image segmentation. As there are two input images, two encoders and two decoders form the contracting and expanding paths, respectively. The encoder extracts the feature map through 3×3 convolution kernel filter operations twice on the input image, and then repeatedly passes through the max-pooling type of sub-sampling process to lower the pixel unit of the feature map. Only the robust features representing the entire image are left. In addition, the computational redundancy was reduced when max-pooling by using a feature channel and two strides for each convolution. In the decoder, two iterations of 3×3 convolution kernel filters per convolution (10 operations in total) and four iterations of an encoder convolution that has undergone max-pooling were copied and cropped to the convolution per decoder. In this method, by repeated processing of the up-pixel unit is again raised in the previous encoder, so the input and the output images can be restored in the same dimension. This process maintains the purpose of semantic segmentation through class prediction for all pixels by restoring the size of coarse feature maps to the size of the original image. Twenty-three convolution layers each from input images 1 and 2 (19 Conv 3×3 ReLu + 4 Up-Conv 2×2) along with the addition of 1×1 Conv—47 convolution layers in total—were used to improve the pixel accuracy.

Fig. 2 depicts the numerous layers present in the O-block, which are multi-channel feature maps within the block, and is represented as a convolution performed by depth-wise individual convolutions where the number of parameters can be greatly reduced. The most important step in the network was to copy and crop the box-computed 3×3 convolution kernel on the multi-channel feature map in the encoder part to prevent the loss of border pixels in each convolution process, thereby concatenating it in the up-convolution decoder part.

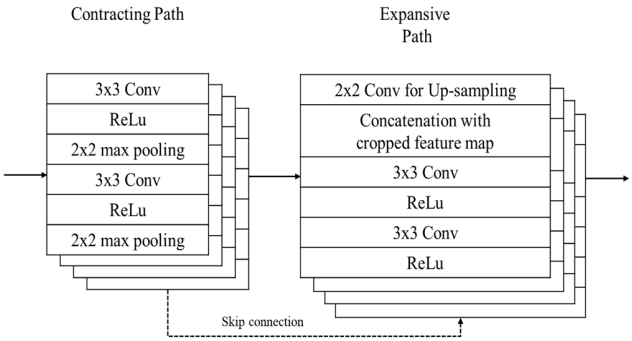

FIGURE 2. The detail in O-block.

TABLE 3. Confusion matrix.

		Actual	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

This skip connection between the encoder and decoder not only provided accurate localization of spatial information but also compensated for the disappearance of detailed pixel information by reducing the image size and then increasing it again.

The color and gray-scale images were used as the first and second images of the input value, respectively. The two images were trained on the neural network, respectively, and the segmentation map representing the predicted class of each pixel was represented as the output image.

B. PERFORMANCE MEASURE

Segmentation typically proceeds performance evaluation through the performance indices of *pixel accuracy* and *m-IoU*. The performance indices are defined based on the confusion matrix, which is comprised of true positive (TP), false positive (FP), false negative (FN), and true negative (TN). Each indicator has been explained and expressed in Table 3 and Fig. 3 for clarity. TP predicts the true answer to be true, and FP predicts the false answer to be true, FN predicts a correct answer as false, and TN predicts a false answer as false.

1) PIXEL ACCURACY

The *pixel accuracy* represents the number of successful pixels of prediction among all the classes of pixels, i.e., it indicates how close the system output is to the truth. This is expressed as (1):

$$Pixel\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

2) INTERSECTION OVER UNION (IOU)

The evaluation index of the model evaluates the predicted value by pixel-wise Intersection-over-Union (IoU).

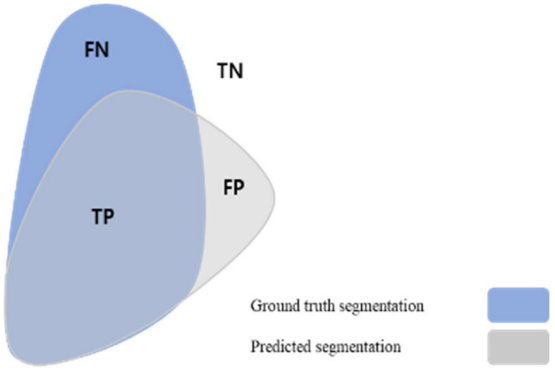

FIGURE 3. A visualization of the confusion matrix described in Table 3, where FN, TP, FP, and TN stand for false negative, true positive, false positive, and true negative, respectively.

(a) (b)

FIGURE 4. Between original ground truth and the 0–1 mask selection color. (a) Original ground truth in labeled knife and (b) output using one-hot encoding (0: Background/Unknown, 1: Knife).

The formula for IoU is expressed by (2):

$$IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i} \tag{2}$$

TP is defined as the number of accurately predicted guns and knives, i.e., the target detection objects. FP is defined as the number of pixels that were incorrectly predicted as a target detection object. FN is defined as the number of target detection object pixels that were incorrectly predicted as other pixels. The IoU, also known as the Jaccard index definition, can define (2) as $TP + FP + FN = Ground\ truth \cup Prediction$ and $TP = Ground\ truth \cap Prediction$. The *m-IoU* expresses IoU as an arithmetic mean of a number of test images, and it is expressed as (3):

$$mIoU = \frac{1}{n} \sum_{i=1}^{n} IoU_i \tag{3}$$

The following Fig. 4 shows the visual image characteristics of the IoU. A segmentation map of the output was created by forming an output channel for each class with one-hot encoding based on the target detection object set to 1 and the background image and unidentifiable area outside the target detection object set to 0.

3) PRECISION AND RECALL

Precision and recall are measured independently of other classes because they are evaluation scales measured for a particular class. First, *precision* is the ratio of what the model classifies as true to what is actually true, indicating the consistency of the system outputs; it is expressed by (4).

The *recall* is the ratio of what is predicted to be true by the model among the true values, indicating how well it detects objects without missing them; it is expressed by (5).

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

4) CONFIDENCE SCORE

Although there is a difference between the accuracy index of detection-based YOLOv3 and segmentation-based U-Net respectively based on object detection and pixel units, the performances of object detection of the two neural networks were compared based on confidence comparative analysis.

The probability that the corresponding model has an object in the corresponding image (or box) and the probability that the object is the predicted object can be expressed as the *confidence score*. The confidence score is obtained by multiplying the object of probability and IoU as (6). The object of probability is precision in the segmentation algorithm.

Therefore, the confidence score was computed using U-Net and O-Net, which are based on image segmentation. Moreover, the detection was judged with a threshold of 0.7, the same as that of YOLOv3.

$$Confidence\ Score = \Pr(object) \times IoU_{pred}^{truth} \tag{6}$$

IV. EXPERIMENTAL RESULTS

A. DATASET

The experiment of this study was conducted on an image dataset jointly produced by a large hub airport in Northeast Asia and an international hub airport in Asia. In addition, we verified Realize, Comprehensive, and Randomize to ensure the correctness of the data. The collected dataset contained image information for not only the dangerous goods of concern (guns and knives) but also for the baggage of ordinary passengers. The images in the dataset were captured using HI-SCAN 6040i X-ray equipment and HI-SCAN 6040-2is HR X-ray equipment using Heimann X-ray technology from Smiths Detection GmbH (Germany). In addition, learning and experiments were conducted with as few as 20 images and as many as 1,500 images. The aviation security process dataset of our study used 2,000 RGB image data, which is Comprehensive with a relatively high amount of data as compared to other studies. Based on the datasets, the experiment was conducted by composing a training set of 700 images and a validation set of 300 images, a total of 1,000 images for each of the ''gun'' and ''knife'' datasets.

Last, it was necessary to prevent over-learning only certain patterns and ensure correct learning by demonstrating

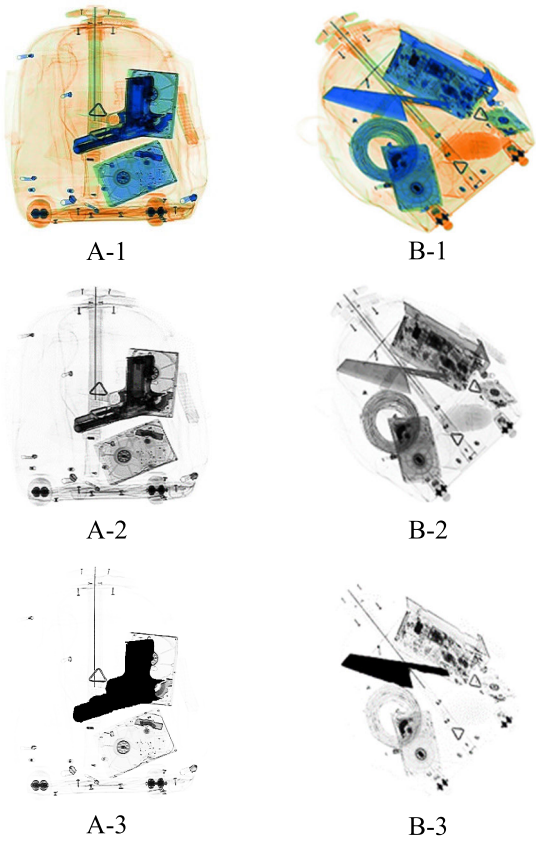

FIGURE 5. Comparison of (1) X-ray RGB image, (2) grayscale image, and (3) labeled image.

Randomize. Therefore, we included various perspectives and random positions in the process to construct an image dataset that is difficult for the X-ray to identify.

In this way, the collected dataset demonstrated all the factors of Realize, Comprehensive, and Randomize and can be termed as a reliable dataset. In addition, it can be regarded as a standard data set with high reliability because the standard data set was secured under national research costs and manpower. Based on the aviation security data set, this study uses a grayscale image that maximizes the image feature extraction to clearly detect the target object by grasping the degree of overlap of each product material and color according to the raw RGB image data from X-ray transmission [67]. Each of the guns and knives in the baggage was collected as the image dataset shown in Fig. 5, and their dataset labeling was performed as follows.

1) GUN

An image dataset for ''gun'' was collected for three samples in a portable bag, and the dataset labeling was performed as follows. The target guns were in the form of air pistols that can easily fit inside a carrying bag and have the same size and structure as that of actual pistols. Actually, the air pistols had the standardized structure of a gun except for its trigger portion. There were three types of guns: Beretta M9A1, Colt 1911A1, and Beretta M92.

2) KNIFE

For knives, the image dataset was collected from three samples in a portable bag, and the ''knife'' dataset labeling was performed as follows. The knives were targeted in the form of real miniature knives that can be put inside an actual carrying bag. They were classified into types according to their use and were divided into general kitchen knives, Chinese cleavers, and butter knives. The handle portion was also labeled to maintain the shape's uniformity.

B. EARLY-STOPPING POINT

The experiment was performed under the conditions of epoch = 100 and batch size = 8. In the experiment, too many epochs can cause overfitting and too few epochs can cause under-fitting. The timely determination of threshold is the key to early-stopping and learning is generally terminated when the performance in the hold-out validation set no longer increases. In addition, the learning is stopped if the error continues to increase compared to the previous epoch. Therefore, the number of epochs needs to be determined to set the standard of error as *patience*. Early-stopping can reduce unnecessary learning due to errors and significantly reduce the total learning time of large image datasets.

The early-stopping algorithm was applied in two stages to select an early-stopping point, as described in Table 4. Algorithm 1 identified an epoch candidate group that became an early-stopping point. The early-stopping patience was increased from 1 for comparisons with the previous accuracy and loss values to construct the epoch candidate set, N_j. Algorithm 2 locates an early-stopping point in the epoch candidate set N_j, and it derives the largest epoch with the early-stopping patience. In case of two or more epochs of the same patience, the smaller epoch was selected to set an early-stopping point and prevent overfitting.

In this experiment, the early-stopping point was derived by applying a total epoch = 10 and an early-stopping patience = 30.

C. DESIGN OF EXPERIMENTS

The design of the experiments comprised the following four components for the application of O-Net on aviation security, specifically for guns and knives:

- Comparison between detection-based YOLOv3 and segmentation-based U-Net
- Design of O-Net structure
- Comparison of other segmentation models
- Comparative analysis of dangerous goods detection

The suitability of image segmentation and superiority of O-Net was attested to through the experiments.

1) COMPARISON BETWEEN DETECTION-BASED YOLOv3 AND SEGMENTATION-BASED U-NET

YOLOv3 is a representative target object detection model that shows excellent performance with breakneck speed in the real-time region for multi-object detection. In contrast, U-Net is best known as a high-performance image segmentation model. Thus, the YOLOv3 and U-Net models were adopted to run the comparative analysis based on the confidence score performance indicators to assess the suitability of these models for aviation security processes.

TABLE 4. Algorithm for implementation of optimization early-stopping to solve O-Net.

Algorithm 1: Finding a candidate set, N_j

Require: Baseline early-stopping epoch of set N_j, where is early-stopping patience

```
Initialize set N_j → { }
Initialize i = 0, epoch
Initialize j = 0, early-stopping patience
while i ≤ Total Epoch do
    while j ≤ early-stopping patience do
        If Acc_i > Acc_{i+j} and Loss_i < Loss_{i+j} do
            Set N_j → N_j + {i }
        j++
    i++
    j → 0
```

Algorithm 2: Finding the optimal epoch

Require: Determine Epoch E

```
Initialize set E → { }
Initialize set N_j, Result Algorithm 1
Initialize j = early-stopping patience
while N_j = ∅ do
    j−−
E = min{N_j}
```

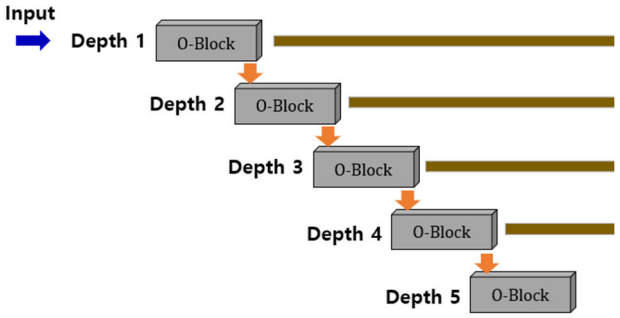

FIGURE 6. Five depths composed of 5 O-blocks in O-Net.

2) DESIGN OF O-NET STRUCTURE

The accurate learning rates and network structure that maximized the accuracy were derived from the implementation of and numerical experiments conducted on O-Net. The reasons for this experiment are as follows. The proper learning with optimal weight does not occur owing to overshooting and divergence of the weight value when the rate value is high in the learning rate optimization for the hyper parameter. Contrastingly, the weight values may converge for low rate values, but too many iterations need to be performed; thus, it takes too long to learn correctly. To find the optimum learning rate, the independent integer value was set within the

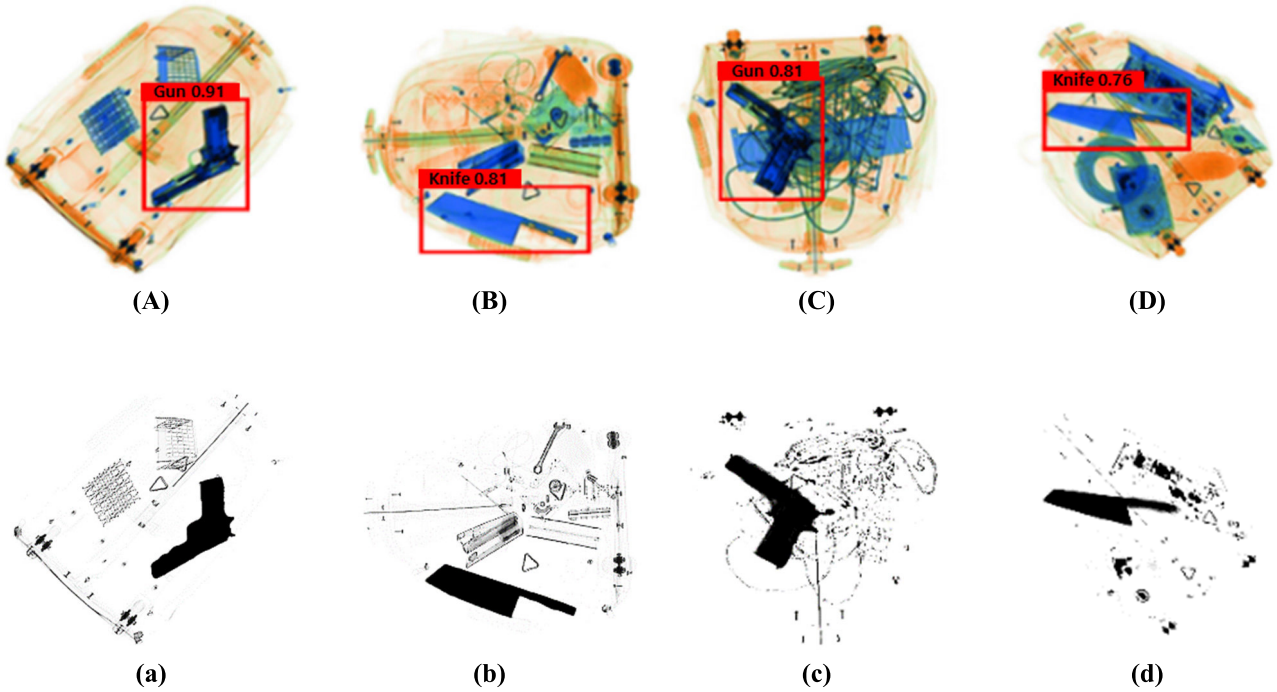

FIGURE 7. Comparison of images of non-overlapped (A, a / B, b) and overlapped(C, c / D, d) objects using YOLOv3 and U-Net. Upper case: YOLOv3 output image, lower case: U-Net output image.

experimental range. The network structure can be designed in various structures, as shown in Fig. 6, through changes in dense and depth. As can be expected, the learning speed and accuracy are affected by the depth and number of parameters in the structure. Therefore, the optimal O-Net structure was developed through experiments with learning rates and dense/depth changes.

3) COMPARISON TO OTHER SEGMENTATION MODELS

In contrast to the earlier experiments that established a superior model to U-Net, this experiment shows how the images of other semantic segmentation architectures, such as FCN and SegNet, were added to and compared with the ground truth to verify whether each image is classifiable with m-IoU and IoU values. This shall further prove the superiority of O-Net.

4) COMPARATIVE ANALYSIS OF DANGEROUS GOODS DETECTION

The detection of dangerous goods in aviation security is an important issue, and this study determines the detection of dangerous goods using image segmentation models, such as U-Net and O-Net, through the previously defined confidence score. The threshold of detection for detecting dangerous goods based on the confidence score was set at 0.7 or higher.

This experiment consisted of 300 test datasets and 200 datasets without dangerous substances. The validation dataset consisted of 500 gun and 500 knife datasets to derive a confusion matrix for results.

Thus, the FCN, SegNet, U-Net, and O-Net (the developed model) models were used to detect dangerous goods and transform the detection into binaries based on the confidence score and run comparative analyses on performance indices such as accuracy, precision, and recall.

D. RESULTS AND ANALYSIS

The experimental environment settings are as shown in Table 5, and the results are as follows.

TABLE 5. Description of the computer specifications and parameter settings.

Configurations	Description
Computer Specifications	Intel(R) Core(TM) i7-9750H CPU 2.6GHz; 16GB RAM; NVIDIA GeForce RTX 2060 6GB GPU; Windows 10 Home 64-bit. In addition, the programs used were Python 3.7.6, CUDA 10.0, Keras 2.2.4, and TensorFlow 1.13.1.
Parameter Settings	The image size was set to 256×256, and the activation function used rectified linear unit (ReLu), which is an improvement of the existing linear function sigmoid. The loss function is described as binary cross entropy that determines whether a detection object is present, the learning rate was studied by applying the Adam algorithm, and the 'he_normal' of He initialization having a random value was used for the weight initialization.

1) COMPARISON BETWEEN DETECTION-BASED YOLOv3 AND SEGMENTATION-BASED U-NET

The detection-based YOLOv3 and segmentation-based U-Net are compared in Fig. 7. In YOLOv3, when the target object did not overlap with other objects, the detection

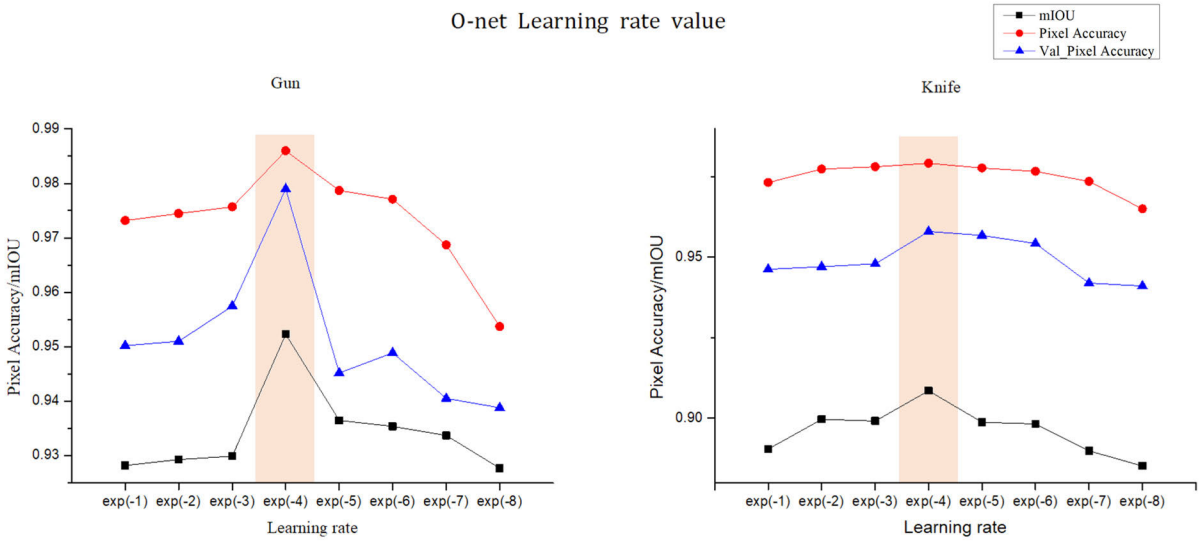

FIGURE 8. O-Net learning rate value; left: Gun, right: Knife.

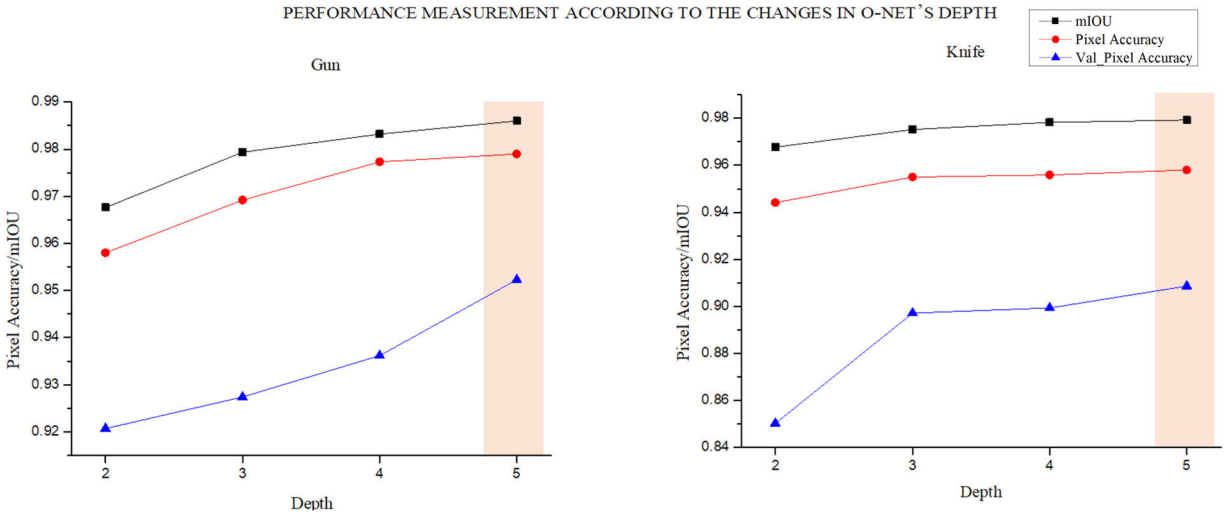

FIGURE 9. Performance measurement according to the changes in depth of O-Net; left: Gun, right: knife.

probability for the gun was 0.91 and 0.94, and for the knife was 0.81 and 0.79. On the contrary, when the target object overlapped with other objects, the detection probability was reduced to 0.81 and 0.84 for the gun, and 0.76 and 0.74 for the knife, indicating an error range of 5 to 10%.

Thus, the object detection accuracy of YOLOv3 was reduced when the target object to be searched overlapped with another object, as signified by the confidence score criteria. This limitation on single-object detection accuracy can be resolved by employing U-Net in the primary object-search process.

Table 6 presents the comparative results of YOLOv3 and U-Net based on the confidence score of the two classes used in this study: guns and knives. Here, it can be seen that the confidence score differed by 20 to 23%.

Referring to Fig. 7, the segmentation U-Net method suggests that security personnel or X-ray baggage detectors can visually transmit the target object detection more clearly during the object-search process. The development of O-Net satisfied a relatively greater number of learning volumes and parameter conditions to verify its excellence.

2) DESIGN OF THE O-NET STRUCTURE

O-Net is a structure consisting of a combination of U-Net-based encoder–decoder parts and performs a comparative analysis of two classes, guns and knives, by comparing the learning rates.

As a result of the experiment depicted in Fig. 8, the U-Net and O-Net structures were the most globally optimal in terms of accuracy, val_accuracy, loss, val_loss, and val_iou at a learning rate of e^{-4}. val_loss was the most optimal at a learning rate of e^{-5} when using the O-Net structure for knives because he_normal, which is the initial value of the weight of the ReLu activation function, was set at random.

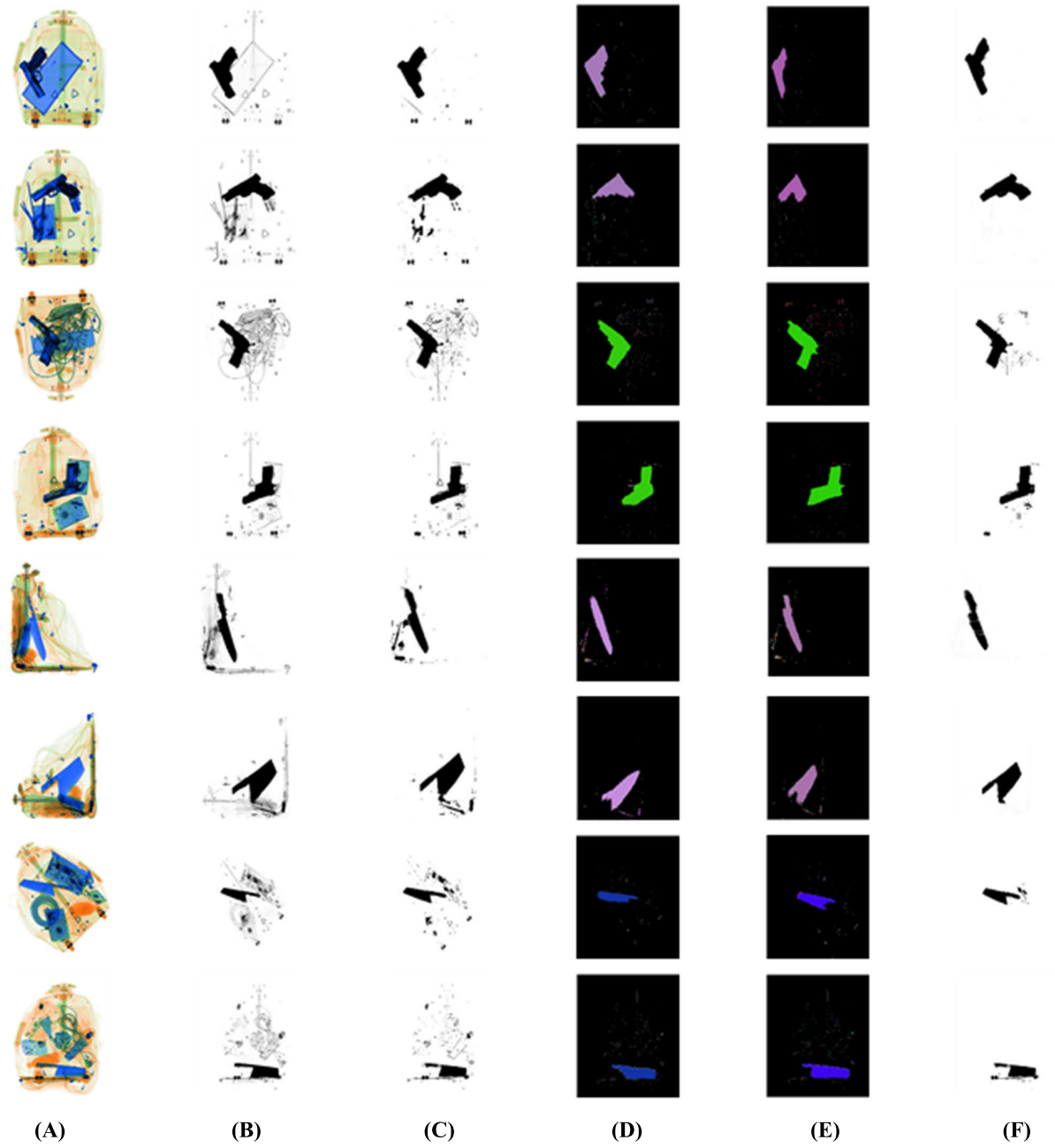

FIGURE 10. Comparison of other segmentation model images. (A): Original image, (B): Ground Truth, (C): U-Net, (D): FCN, (E): SegNet, and (F): O-Net.

TABLE 6. Comparative analysis of YOLOv3 and U-Net's confidence scores.

Class	Model	
	YOLOv3	U-Net
Gun	0.73	0.93
Knife	0.68	0.91

The stability experiment based on depth degree of learning was similar to Fig. 6 and is shown in Fig. 9. The results are indicative of the fixing of the local optimal learning rate at e^{-4} in the previous experiment.

Owing to the specificity of copy and crop in the O-Net neural network, the depth needs to start with a minimum value of 2; as density increases, experiments with a density of up to 5 are possible with the existing complexity of parameters and computations. For depths exceeding 6, the number of parameters turned too large to be computed. In other words, the increase in the performance index with the increase in dense/depth of the neural network was confirmed. Upon considering the m-IoU over accuracy of the experimental results,

TABLE 7. Segmentation model experiment results on gun and knife.

Gun				Knife			
Base Model	Pixel Accuracy	Loss	m-IOU	Base Model	Pixel Accuracy	Loss	m-IOU
FCN	0.9389	0.0213	0.5005	FCN	0.9102	0.0172	0.4984
SegNet	0.9463	**0.0125**	0.7612	SegNet	0.9187	**0.0038**	0.6334
U-Net	0.9678	0.0172	0.8389	U-Net	0.9522	0.0072	0.8251
O-Net (our model)	**0.9860**	0.0165	**0.9523**	O-Net (our model)	**0.9792**	0.0054	**0.9086**

TABLE 8. Comparative analysis of gun and knife detection.

Gun				Knife			
Base Model	Accuracy	Precision	Recall	Base Model	Accuracy	Precision	Recall
U-Net	0.9080	0.9578	0.8853	U-Net	0.8652	0.9223	0.8456
O-Net (current model)	**0.9692**	**0.9802**	**0.9671**	O-Net (current model)	**0.9352**	**0.9466**	**0.9462**

the gun detection accuracy was improved by 0.0316 between a minimum of 0.9207 at depth 2 and a maximum of 0.9523 at depth 5. Similarly, knife detection accuracy was improved by 0.0583 between a minimum of 0.8503 at depth 2 and a maximum of 0.9086 at depth 5.

3) COMPARISON TO OTHER SEGMENTATION MODELS

Table 7 compares four segmentation models based on the two classes of gun and knife. For the proposed O-Net model, the experiments on guns showed the best figures among the four models with 98.60% pixel accuracy, 95.23% m-IoU; the results for the knife showed 97.92% pixel accuracy and 90.86% m-IoU. In comparison to FCN, O-Net exhibited a 4.71% higher pixel accuracy and a 45.18% higher m-IoU for the gun, and 6.90% higher pixel accuracy and 41.02% higher m-IoU for the knife. The figures representing the output results of the models are shown in Fig. 10.

4) COMPARATIVE ANALYSIS OF DANGEROUS GOODS DETECTION

However, since the m-IoUs of FCN and SegNet were less than 0.7 in the previous experiment, the confidence score could not exceed 0.7, even if the precision value were multiplied. Thus, the experiment was limited to U-Net and O-Net, and the comparative analysis has been presented in Table 8.

The analysis of the detection results for guns show that all the O-Net performance indices were improved compared to the U-Net. In summary, the accuracy was increased by about 6%, and the recall, which confirms the degree of dangerous goods detection, was also improved by roughly 8%.

Moreover, O-Net delivered higher performance than U-Net in terms of knife detection. The accuracy was improved by approximately 7%, and the recall was improved by roughly 10%. Therefore, the proposed O-Net architecture was verified to have a very high detection rate of guns and knives with good accuracy.

V. CONCLUSION

In this study, we developed a segmentation model that can identify guns and knives as dangerous baggage items in the aviation security process. The existing detection model focused on the general image, but dangerous goods detection in the aviation security process has to look beyond the overlapping objects detected under X-ray screening. To overcome these limitations, a new suitable model for the aviation security process was developed using the segmentation model.

The proposed model, O-Net, was developed based on the U-Net structure of segmentation by simultaneously using two inputs—a general X-ray RGB image and an image converted to grayscale—to solve the overlapping-objects problem in X-ray images. In addition, semantic segmentation removes all unnecessary background areas other than the target area to increase the detection accuracy of the target object that is relevant for aviation security processes, thus reducing human error. The optimal structure for the O-Net network design was derived through various learning rates and dense- and depth-wise experiments to improve the performance. Three basic semantic segmentation algorithms—FCN, U-Net, and SegNet—were comparatively analyzed in terms of performance indicators of segmentation such as pixel accuracy and m-IoU. On average, pixel accuracy and m-IoU using O-Net was improved by 5.8%, 2.26%, and 5.01%, respectively, and the m-IoU was improved by 43.1%, 9.84%, and 23.31%, respectively. Moreover, the accuracy of O-Net was 6.56% higher than U-Net, indicating the superiority of O-Net.

REFERENCES

[1] *WATS World Air Transport Statistics 2019*, Int. Air Transp. Assoc. Corp., Geneva, Switzerland, 2019.

[2] IATA. *In Numbers: World Air Transport Statistics 2019*. Accessed: Oct. 30, 2019. [Online]. Available: https://www.airlines.iata.org/data/in-numbers-world-air-transport-statistics-2019

[3] *The 9/11 Commission Report: The Final Report of the National Commission on Terrorist Attacks Upon the United States*, Barnes & Noble Publishing, New York, NY, USA, 2004.

[4] M. Johnston, M. McNeil, K. D. Giudice, and B. Kudrick, "Using a game to evaluate passenger screener fatigue and sleepiness at airport screening checkpoints," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 58, no. 1. Los Angeles, CA, USA: SAGE Publications, 2014, pp. 2290–2294.

[5] G. Blalock, V. Kadiyali, and D. H. Simon, "The impact of post-9/11 airport security measures on the demand for air travel," *J. Law Econ.*, vol. 50, no. 4, pp. 731–755, 2007.

[6] T. Hunter and F. Chau, "Islamist fundamentalist and separatist attacks against civil aviation since 11th Sep. 2001," in *Aviation Security: Challenges and Solutions*. Hong Kong: Avseco, 2011, pp. 35–54.

[7] N. Hattenschwiler, S. Michel, M. Kuhn, S. Ritzmann, and A. Schwaninger, "A first exploratory study on the relevance of everyday object knowledge and training for increasing efficiency in airport security X-ray screening," in *Proc. Int. Carnahan Conf. Secur. Technol. (ICCST)*, Sep. 2015, pp. 25–30.

[8] S. Michel and A. Schwaninger, "Human-machine interaction in X-ray screening," in *Proc. 43rd Annu. Int. Carnahan Conf. Secur. Technol.*, Oct. 2009, pp. 13–19.

[9] J. Skorupski and P. Uchroński, "A human being as a part of the security control system at the airport," *Procedia Eng.*, vol. 134, pp. 291–300, Jan. 2016.

[10] T.-C. Wang and L.-H. Chuang, "Psychological and physiological fatigue variation and fatigue factors in aircraft line maintenance crews," *Int. J. Ind. Ergonom.*, vol. 44, no. 1, pp. 107–113, Jan. 2014.

[11] E. K. Chung, Y. Jung, and Y. W. Sohn, "A moderated mediation model of job stress, job satisfaction, and turnover intention for airport security screeners," *Saf. Sci.*, vol. 98, pp. 89–97, Oct. 2017.

[12] A. K. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," *EURASIP J. Adv. Signal Process.*, vol. 2008, pp. 1–17, 2008.

[13] A. Adler, "Images can be regenerated from quantized biometric match score data," in *Proc. Can. Conf. Elect. Comput. Eng.*, vol. 1, 2004, pp. 469–472.

[14] C. Morosan, "An empirical examination of U.S. travelers' intentions to use biometric e-gates in airports," *J. Air Transp. Manage.*, vol. 55, pp. 120–128, Aug. 2016.

[15] International Air Transport Association. (2019). *IOSA Standards Manual*. [Online]. Available: https://www.iata.org/en/iata-repository/publications/iosa-audit-documentation/iosa-standards-manual-ism-ed-132/

[16] [Online]. Available: https://www.iata.org/contentassets/517a926779c3491582366cc58fce1eb4/india-facilitation-and-security-paper_jan-2019.pdf

[17] Y. Sterchi and A. Schwaninger, "A first simulation on optimizing EDS for cabin baggage screening regarding throughput," in *Proc. Int. Carnahan Conf. Secur. Technol. (ICCST)*, Sep. 2015, pp. 55–60.

[18] M. Vahčič, D. Anderson, M. R. Osés, G. Rarata, and G. Diaconu, "Development of inert, polymer-bonded simulants for explosives detection systems based on transmission X-ray," *Molecules*, vol. 24, no. 23, p. 4330, Nov. 2019.

[19] K. D. Krug, W. F. Aitkenhead, R. F. Eilbert, J. H. Stillson, and J. A. Stein, "Detecting explosives or other contraband by employing transmitted and scattered X-rays," U.S. Patent 5 600 700, Feb. 4, 1997.

[20] N. Donnelly, A. Muhl-Richardson, H. Godwin, and K. Cave, "Using eye movements to understand how security screeners search for threats in X-ray baggage," *Vision*, vol. 3, no. 2, p. 24, Jun. 2019.

[21] A. Schwaninger, D. Hardmeier, and F. Hofer, "Aviation security screeners visual abilities & visual knowledge measurement," *IEEE Aerosp. Electron. Syst. Mag.*, pp. 29–35, Jun. 2005.

[22] A. Bolfing, T. Halbherr, and A. Schwaninger, "How image based factors and human factors contribute to threat detection performance in X-ray aviation security screening," in *Proc. Symp. Austrian HCI Usability Eng. Group*. Berlin, Germany: Springer, 2008, pp. 419–438.

[23] G. Harding, "X-ray scatter tomography for explosives detection," *Radiat. Phys. Chem.*, vol. 71, nos. 3–4, pp. 869–881, Oct. 2004.

[24] C. Cui, S. M. Jorgensen, D. R. Eaker, and E. L. Ritman, "Direct three-dimensional coherently scattered X-ray microtomography," *Med. Phys.*, vol. 37, no. 12, pp. 6317–6322, Nov. 2010.

[25] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[26] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[27] P. Rajeshwari, P. Abhishek, P. Srikanth, and T. Vinod, "Object detection: An overview," *Int. J. Trend Sci. Res. Develop.*, vol. 3, no. 3, pp. 1663–1665, Apr. 2019.

[28] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[30] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 818–833.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[34] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.

[35] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[36] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[37] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[38] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2961–2969.

[39] 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. (CVPR)*, Jun. 2016, pp. 779–788.

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

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

[42] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[43] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[44] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," 2014, *arXiv:1412.7062*. [Online]. Available: http://arxiv.org/abs/1412.7062

[45] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: http://arxiv.org/abs/1706.05587

[46] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.

[47] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[48] K.-H. Uhm, S.-W. Kim, S.-W. Ji, S.-J. Cho, J.-P. Hong, and S.-J. Ko, "W-Net: Two-stage U-Net with misaligned data for raw-to-RGB mapping," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3636–3642.

[49] J. Zhuang, "LadderNet: Multi-path networks based on U-Net for medical image segmentation," 2018, *arXiv:1810.07810*. [Online]. Available: http://arxiv.org/abs/1810.07810

[50] J. Bullock, C. Cuesta-Lázaro, and A. Quera-Bofarull, "XNet: A convolutional neural network (CNN) implementation for medical X-ray image segmentation suitable for small datasets," *Proc. SPIE*, vol. 10953, Mar. 2019, Art. no. 109531Z.
[51] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.
[52] K. Qi, H. Yang, C. Li, Z. Liu, M. Wang, Q. Liu, and S. Wang "X-Net: Brain stroke lesion segmentation based on depthwise separable convolution and long-range dependencies," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Oct. 2019, pp. 247–255.
[53] Z.-L. Ni, G.-B. Bian, X.-H. Zhou, Z.-G. Hou, X.-L. Xie, C. Wang, Y.-J. Zhou, R.-Q. Li, and Z. Li "RAUNet: Residual attention U-Net for semantic segmentation of cataract surgical instruments," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, Dec. 2019, pp. 139–149.
[54] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 3–11.
[55] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*. [Online]. Available: http://arxiv.org/abs/1802.06955
[56] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020.
[57] W. Yao, Z. Zeng, C. Lian, and H. Tang, "Pixel-wise regression using U-Net and its application on pansharpening," *Neurocomputing*, vol. 312, pp. 364–371, Oct. 2018.
[58] W. Wiratama, J. Lee, and D. Sim, "Change detection on multi-spectral images based on feature-level U-Net," *IEEE Access*, vol. 8, pp. 12279–12289, 2020.
[59] S. Apostolopoulos, S. De Zanet, C. Ciller, S. Wolf, and R. Sznitman, "Pathological OCT retinal layer segmentation using branch residual u-shape networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Sep. 2017, pp. 294–301.
[60] M. K. Abd-Ellah, A. A. M. Khalaf, A. I. Awad, and H. F. A. Hamed, "TPUAR-Net: Two parallel U-Net with asymmetric residual-based deep convolutional neural network for brain tumor segmentation," in *Proc. Int. Conf. Image Anal. Recognit.* Cham, Switzerland: Springer, Aug. 2019, pp. 106–116.
[61] G. Prathap and I. Afanasyev, "Deep learning approach for building detection in satellite multispectral imagery," in *Proc. Int. Conf. Intell. Syst. (IS)*, Sep. 2018, pp. 461–465.
[62] V. Khryashchev, R. Larionov, A. Ostrovskaya, and A. Semenov, "Modification of U-Net neural network in the task of multichannel satellite images segmentation," in *Proc. IEEE East-West Design Test Symp. (EWDTS)*, Sep. 2019, pp. 1–4.
[63] J. Dolz, I. B. Ayed, and C. Desrosiers, "Dense multi-path U-Net for ischemic stroke lesion segmentation in multiple image modalities," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, Sep. 2018, pp. 271–282.
[64] X. Li, Y. Wang, Q. Tang, Z. Fan, and J. Yu, "Dual U-Net for the segmentation of overlapping glioma nuclei," *IEEE Access*, vol. 7, pp. 84040–84052, 2019.
[65] V. K. Valloli and K. Mehta, "W-Net: Reinforced U-Net for density map estimation," 2019, *arXiv:1903.11249*. [Online]. Available: http://arxiv.org/abs/1903.11249
[66] T. Nair, D. Precup, D. L. Arnold, and T. Arbel, "Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation," *Med. Image Anal.*, vol. 59, Jan. 2020, Art. no. 101557.
[67] Y. Xie and D. Richmond, "Pre-training on grayscale ImageNet improves medical image classification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 1–9.

WOONG KIM (Member, IEEE) received the M.S. degree in industrial management engineering from Korea University, Seoul, South Korea, in 2016, where he is currently pursuing the Ph.D. degree in industrial engineering.

Since 2015, he has been a Teacher with the Sungil Information High School. His research interests include artificial intelligence, machine learning, logistics and transportation information system, and supply chain management (SCM).

SUNGCHAN JUN received the B.S. degree in industrial management engineering from Korea University, Seoul, South Korea, in 2018, where he is currently pursuing the M.S. degree in industrial engineering with a focus in industrial artificial intelligence.

His research interests include artificial intelligence, logistics transportation, and patent analysis.

SUMIN KANG was born in Busan, Republic of Korea, in 1998. He is currently pursuing the B.S.E. degree in aerospace engineering with a minor in materials science and engineering with the University of Michigan, Ann Arbor, MI, USA.

His research interests include patent analysis to predict prospective technologies and AI technology using deep learning. He has been a member of the American Institute of Aeronautics and Astronautics (AIAA) and the American Society for Engineering Education (ASEE), since 2018.

CHULUNG LEE received the B.S. and M.S. degrees in industrial engineering from Seoul National University, Seoul, South Korea, in 1992 and 1994, respectively, and the Ph.D. degree in industrial engineering from Pennsylvania State University, State College, PA, USA, in 2000.

From December 2000 to August 2005, he was an Assistant Professor with the Department of Industrial and Systems Engineering, National University of Singapore, Singapore. Since 2005, he has been a Professor with the School of Industrial Management Engineering, Korea University and since 2015, he has also served as a Head Professor with the Department of Master of Intellectual Property, Seoul, South Korea. His research interests include supply chain management (SCM), logistics and transportation information system, logistics technology innovation, technology management, and intellectual property.