Savitribai Phule Pune University
Modern Education Society's College of Engineering, Pune
19, Bund Garden, V.K. Joag Path, Pune – 411001.

ACCREDITED BY NAAC WITH "A" GRADE (CGPA – 3.13)

DEPARTMENT OF COMPUTER ENGINEERING

A SEMINAR REPORT

ON

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

T.E. (COMPUTER)

SUBMITTED BY

Mr.Gaurav Ramesh Ghatge **Exam Seat No: T150324255**

UNDER THE GUIDANCE OF

Prof. Bhagyashri More

(Academic Year: 2020-2021)

Savitribai Phule Pune University
Modern Education Society's College of Engineering, Pune
19, Bund Garden, V.K. Joag Path, Pune – 411001.

ACCREDITED BY NAAC WITH "A" GRADE (CGPA – 3.13)

DEPARTMENT OF COMPUTER ENGINEERING

Certificate

This is to certify that seminar entitled

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

has been completed by

Mr.Gaurav Ramesh Ghatge (Prn No.S18111041)

of TE COMP S.S. in the Semester - II of academic year 2020-2021 in partial fulfillment of the Third Year of Bachelor degree in "Computer Engineering" as prescribed by the Savitribai Phule Pune University.

Prof. Bhagyashri More	**Dr.(Mrs.) N. F. Shaikh**
Seminar Guide	**H.O.D**

Place: MESCOE, Pune.
Date: 05/ 06/2020

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

Keywords:

- *Artificial intelligence*
- *Image segmentation*
- *U-Net*
- *Aviation security*
- *Detection algorithm*

ACKNOWLEDGEMENT

It gives me great pleasure and satisfaction in presenting this seminar on "O-Net:Dangerous Goods Detection in Aviation Security Based on U-Net".

I would like to express my deep sense of gratitude towards my seminar guide **Prof. Bhagyashri More** *or her support, continuous guidance and being so under-standing and helpful throughout the seminar.*

I have further-more to thank Computer Department HOD **Dr.(Mrs.) N. F.Shaikh** *to encourage me to go ahead and for continuous guidance. I also want to thank* **Prof.J.R.Pansare** *for all her assistance and guidance for preparing report.*

I would like to thank all those, who have directly or indirectly helped me forthe completion of the work during this mini project.

Student Name: Gaurav Ramesh Ghatge
TE Computer S.S.
PRN : S18111041

Contents

List of Figures

Chapter 1

INTRODUCTION

1.1 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. 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

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.

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

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.In general Convolutional Neural Networks gives decent results in easier image segmentation problems but it hasn't made any good progress on complex ones. That's where UNet comes in the picture. UNet was first designed especially for medical image segmentation. It showed such good results that it used in many other fields after.The main idea behind CNN is to learn the feature mapping of an image and exploit it to make more nuanced feature mapping. This works well in classification problems as the image is converted into a vector which used further for classification.

1.2 Motivation

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 . 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 .

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 paper , artificial neural network is been applied to fit such trends attempted to establish an algorithm based of U-Net structure for the automatic detection of dangerous goods from X-ray screening images.

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.

Chapter 2

Literature Survey

2.1 Existing Aviation Security Systems

An aviation security system employs technology to prevent illegal acts that endanger human life and property, jeopardise civil aviation operations' safety, or have a significant impact on the performance of aviation tasks. The security equipment used in aviation security systems is classed according to whether it scans "people" or "things."Hand-held metal detectors and walk-through metal detectors are two types of human search devices. A hand-held metal detector is a search equipment that uses an electromagnetic field to detect metal objects and is safe to use when looking for 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.

Object search devices include X-ray screening systems, whole-body scanners, explosive detecting systems, and explosive trace detectors. 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.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

2.2 Use of Image Recognition and Detection in Aviation Security Systems

X-ray transmittance is used to represent an X-ray image. White is used for places with no material or with a very low density, whereas saturated colours are used for areas with a high density."overlapping" phenomenon increases the screener's stress. In severe cases, dangerous goods go undetected and create a major problem in aviation security To solve the above mentioned problems, various algorithms are applied, such as detection based on X-ray transmittance information and the development of SIFT and SURF algorithms for extracting image features .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

The simultaneous execution of classification and localization of various objects is essential for target object detection.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 entity and aims to perform a dense prediction to classify every pixel in the image.

2.3 Baseline U-Net for Detection

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, Ladder-Net , and X-Net of one-stage U-Net improved performance by repeatedly using the U-Net structure.V-Net 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. Moreover, U-Net++ , R2U-Net, and MultiRes U-Net applied skip connection and a recurrent/residual structure to improve the utilization of information and performance.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.

Chapter 3

Details Of Technology

3.1 Artificial Intelligence

3.1.1 Overview Of Artificial Intelligence

It is "The science and engineering of making intelligent machines, especially intelligent computer programs".Artificial Intelligence is a way of making a computer, a computer-controlled robot, or a software think intelligently, in the similar manner the intelligent humans think.AI is accomplished by studying how human brain thinks, and how humans learn, decide, and work while trying to solve a problem, and then using the outcomes of this study as a basis of developing intelligent software and systems.Artificial intelligence is a science and technology based on disciplines such as Computer Science, Biology, Psychology, Linguistics, Mathematics, and Engineering.

A major thrust of AI is in the development of computer functions associated with human intelligence, such as reasoning, learning, and problem solving.Out of the following areas, one or multiple areas can contribute to build an intelligent system.Artificial intelligence's growing popularity in the 21st century is largely due to the advancements in the sub-field of machine learning. Machine learning develops systems that improve upon themselves, which is accomplished through the identification of algorithms. Some processes that machine learning optimizes include paperwork automation, forensic accounting, and algorithmic trading.

3.1.2 Overview Of Image recognition

As humans can easily distinguish places, objects, and people from images, but computers traditionally face a tough time comprehending these images. Thanks to the new image recognition technology, now we have specialized software and applications that can decipher visual information. We often use the terms "Computer vision" and "Image recognition" interchangeably, however, there is a slight difference between these two terms. Instructing computers to understand and interpret visual information, and take actions based on these insights is known as computer vision.Computer vision is a broad field that uses deep learning to perform tasks such as image processing, image classification, object detection, object segmentation, image colorization, image reconstruction, and image synthesis. On the other hand, image recognition is a subfield of computer vision that interprets images to assist the decision-making process.

3.1.3 Image Recognition Algorithms

Artificial Intelligence has transformed the image recognition features of applications. Some applications available on the market are intelligent and accurate to the extent that they can elucidate the entire scene of the picture. Researchers are hopeful that with the use of AI they will be able to design image recognition software that may have a better perception of images and videos than humans.

Image recognition comes under the banner of computer vision which involves visual search, semantic segmentation, and identification of objects from images. The bottom line of image recognition is to come up with an algorithm that takes an image as an input and interprets it while designating labels and classes to that image. Most of the image classification algorithms such as bag-of-words, support vector machines (SVM), face landmark estimation, and K-nearest neighbors (KNN), and logistic regression are used for image recognition also. Another algorithm Recurrent Neural Network (RNN) performs complicated image recognition tasks, for instance, writing descriptions of the image.

Type	Description
Convolutional Neural Network (CNN)	LeNet, AlexNet, ZFNet, VGGNet, GoogleLeNet, ResNet, Inception-ResNet, Inception v4
Detection Algorithms	R-CNN, Fast R-CNN, Faster R-CNN, YOLO v1, v2, v3
Segmentation Algorithms	Mask R-CNN, FCN, SegNet, Deep Lab v1, v2, v3, v3+, U-Net

Figure 3.1: Classification of Image Recognition and Detection in Artificial Intelligence

3.2 Convolutional Neural Network

3.2.1 Overview Of Convolutional Neural Network

A convolutional neural network (CNN) is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data.CNNs are powerful image processing, artificial intelligence (AI) that use deep learning to perform both generative and descriptive tasks, often using machine vison that includes image and video recognition, along with recommender systems and natural language processing.

A neural network is a system of hardware and/or software patterned after the operation of neurons in the human brain. Traditional neural networks are not ideal for image processing and must be fed images in reduced-resolution pieces. CNN have their "neurons" arranged more like those of the frontal lobe, the area responsible for processing visual stimuli in humans and other animals. The layers of neurons are arranged in such a way as to cover the entire visual field avoiding the piecemeal image processing problem of traditional neural networks.

3.2.2 How Convolutions Work

CNN image classifications takes an input image, process it and classify it under certain categories (Eg., Dog, Cat, Tiger, Lion). Computers sees an input image as array of pixels and it depends on the image resolution. Based on the image resolution, it will see h x w x d(h = Height, w = Width, d = Dimension). Eg., An image of 6 x 6 x 3 array of matrix of RGB (3 refers to RGB values) and an image of 4 x 4 x 1 array of matrix of grayscale image.Technically, deep learning CNN models to train and test, each input image will pass it through a series of convolution layers with filters (Kernals), Pooling, fully connected layers (FC) and apply Softmax function to classify an object with probabilistic values between 0 and 1.

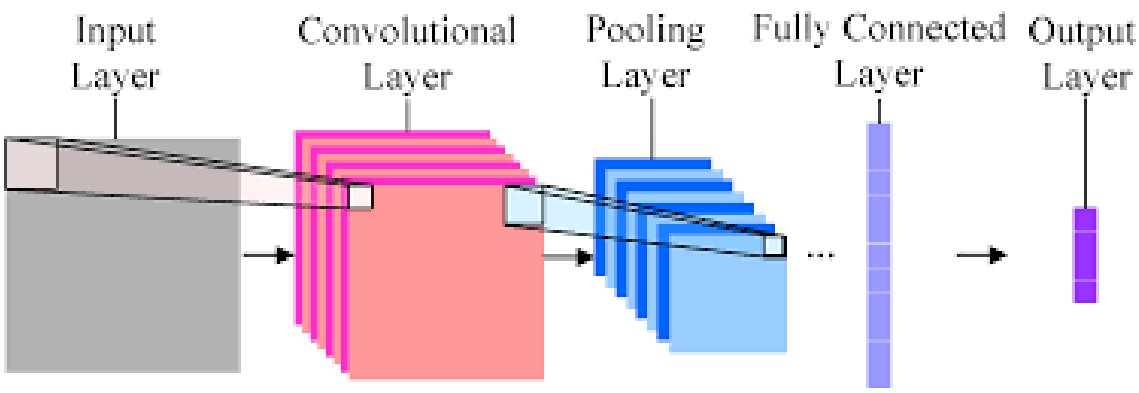


Figure 3.2: Elements interact in the following manner

3.2.3 Layers in a Convolutional Neural Network

A convolution neural network has multiple hidden layers that help in extracting information from an image. The four important layers in CNN are:

Convolution layer

This is the first step in the process of extracting valuable features from an image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values.

ReLU layer

ReLU stands for the rectified linear unit. Once the feature maps are extracted, the next step is to move them to a ReLU layer.ReLU performs an element-wise operation and sets all the negative pixels to 0. It introduces non-linearity to the network, and the generated output is a rectified feature map

Pooling layer

Pooling is a down-sampling operation that reduces the dimensionality of the feature map. The rectified feature map now goes through a pooling layer to generate a pooled feature map.The pooling layer uses various filters to identify different parts of the image like edges, corners, body, feathers, eyes, and beak.

Fully connected layer

Fully Connected Layer is simply, feed forward neural networks. Fully Connected Layers form the last few layers in the network.The input to the fully connected layer is the output from the final Pooling or Convolutional Layer, which is flattened and then fed into the fully connected layer.

3.3 O-Net 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.

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.

3.3.1 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. 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. 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

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. 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.

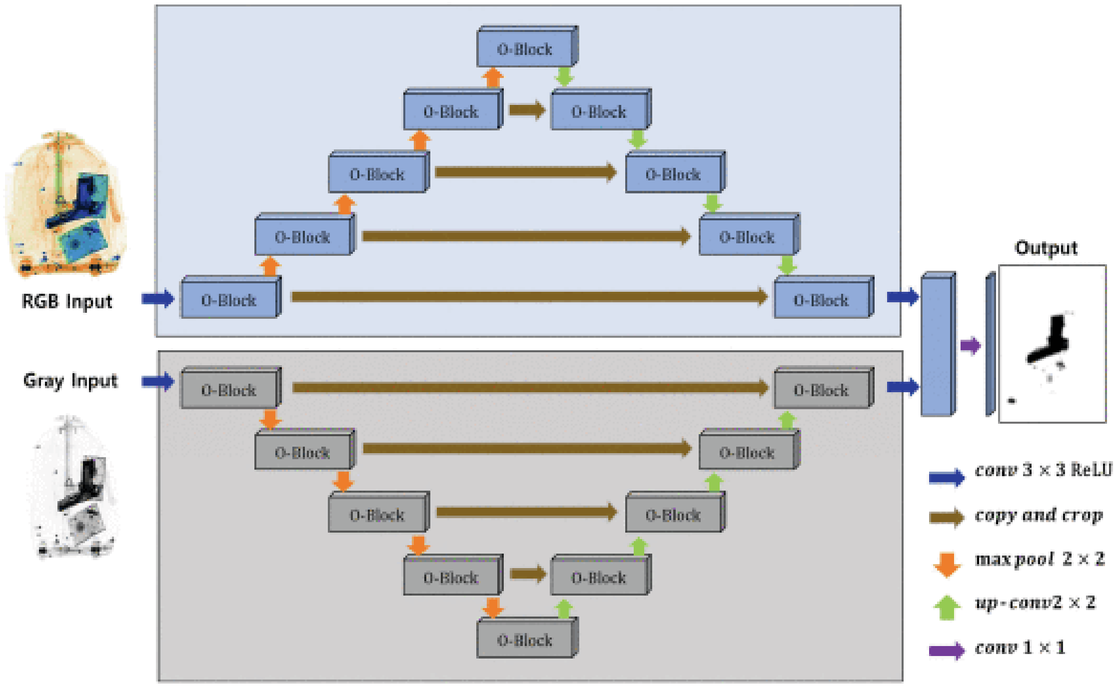


Figure 3.3: O-Net Structure: U-Net-based structure using an original X-ray image and a converted grayscale image as inputs.

3.3.2 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). 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.

2) Intersection Over Union (IOU)

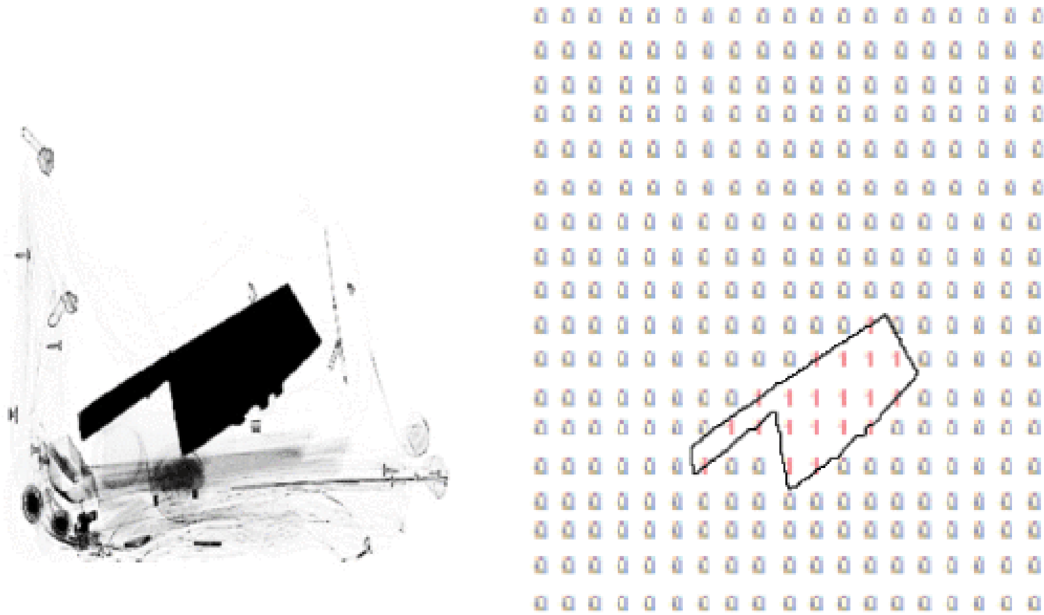
The evaluation index of the model evaluates the predicted value by pixel-wise Intersection-over-Union (IoU)

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.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.

4) Confidence Score

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.



(a) (b)

Figure 3.4: (a) Original ground truth in labeled knife and (b) output using one-hot encoding

Chapter 4

Analytical and experimental work

4.1 Overview of Experiment

4.1.1 Dataset Details

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.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.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.

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.

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.

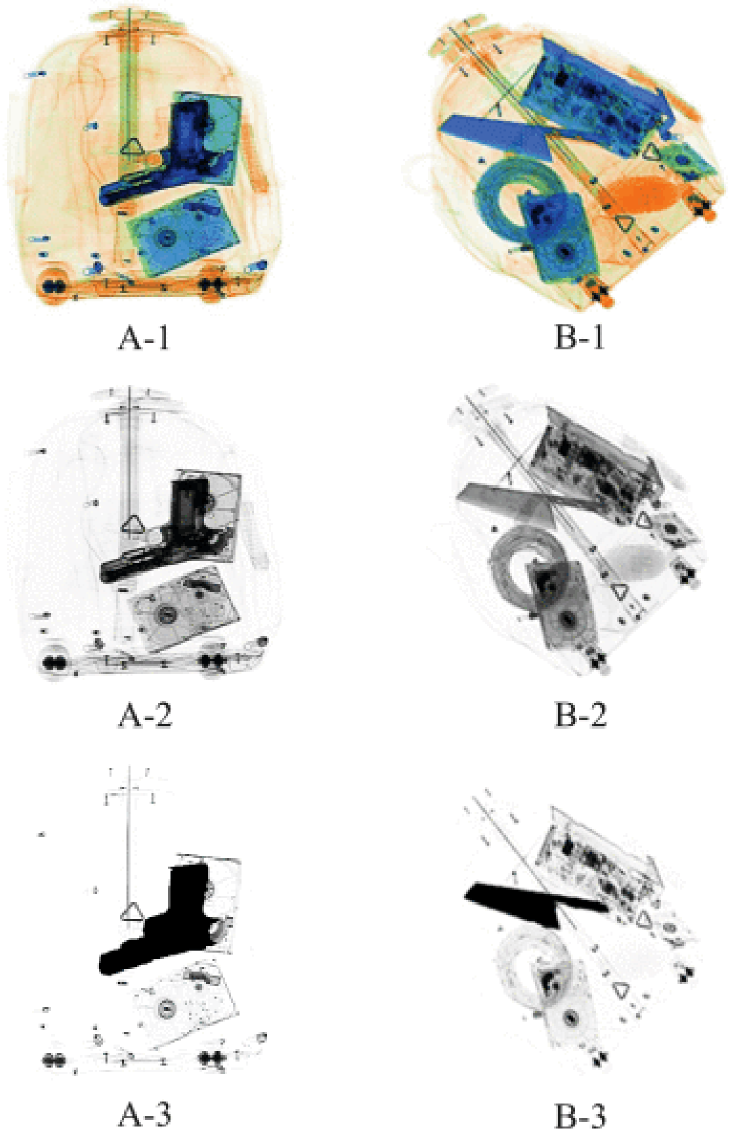


Figure 4.1: Comparison of (1) X-ray RGB image, (2) grayscale image, and (3) labeled image.

4.1.2 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.

4.2 Design of Experiment

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

4.2.1 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.

4.2.2 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.

4.2.3 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.

4.2.4 Comparative Analysis of Dangerous Goods Detection

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.

4.3 Results and Analysis

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.

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.

Figure 4.2: Description of the Computer Specifications and Parameter Settings

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 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%.

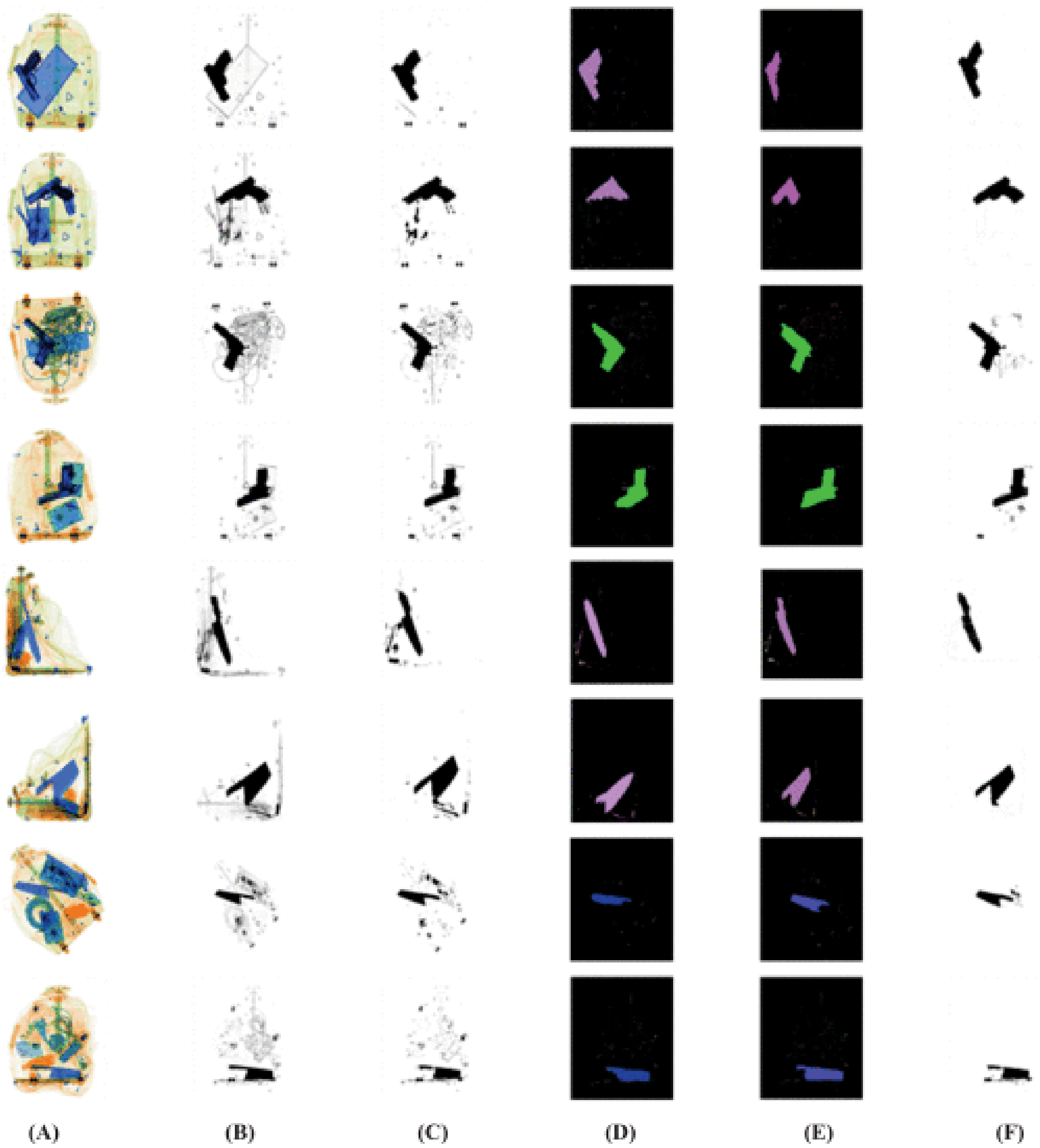


Figure 4.3: Comparison of other segmentation model images. (A): Original image, (B): Ground Truth, (C): U-Net, (D): FCN, (E): SegNet, and (F): O-Net.

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**

Figure 4.4: Comparative Analysis of Gun and Knife Detection

Chapter 5

CONCLUSION

We created a segmentation model in this study that can identify weapons and knives as risky luggage items during the aviation security process. The current detection methodology focused on the overall image, but dangerous goods identification in the airport security procedure requires looking beyond the overlapping objects found during X-ray screening.To overcome these restrictions, the segmentation model was used to create a new acceptable model for the aviation security process.

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] T. Hunter and F. Chau, "Islamist fundamentalist and separatist attacks against civil aviation since 11th september 2001," in *Aviation security challenges and solutions*, Avseco, 2011, pp. 35–54.

[2] V. K. Valloli and K. Mehta, "W-net: Reinforced u-net for density map estimation," *arXiv preprint arXiv:1903.11249*, 2019.

[3] P. Rajeshwari, P. Abhishek, P. Srikanth, and T. Vinod, "Object detection: An overview," *Int. J. Trend Sci. Res. Dev.(IJTSRD)*, vol. 3, no. 1, pp. 1663–1665, 2019.

[4] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *arXiv preprint arXiv:1506.01497*, 2015.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[7] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[8] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[9] 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 *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, IEEE, 2019, pp. 3636–3642.

[10] T. Nair, D. Precup, D. L. Arnold, and T. Arbel, "Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation," *Medical image analysis*, vol. 59, p. 101 557, 2020.

[11] Y. Xie and D. Richmond, "Pre-training on grayscale imagenet improves medical image classification," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018, pp. 0–0.