НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2024-24-1-51-61

An improved performance of RetinaNet model for hand-gun detection in custom dataset and real time surveillance video

Pyone Pyone Khin[1]✉, Nay Min Htaik[2]

[1,2] Mandalay Technological University, Mandalay, 05072, Myanmar

[1] pyonekhin.ppk@gmail.com✉, https://orcid.org/0009-0002-0512-6414
[2] nayminhtaik@gmail.com, https://orcid.org/0009-0009-8295-6914

Abstract
The prevalence of armed robberies has become a significant concern in today's world, necessitating the development of effective detection systems. While various detection devices exist in the market, they do not possess the capability to automatically detect and alarm the presence of guns during robbery activities. In order to address this issue, a deep learning-based approach using gun detection using RetinaNet model is proposed. The objective is to accurately detect guns and subsequently alert either the police station or the bank owner. RetinaNet, the core of the system, comprises three main components: the Residual Neural Network (ResNet), the Feature Pyramid Network (FPN), and the Fully Convolutional Networks (FCN). These components work together to enable real-time detection of guns without the need for human intervention. Proposed implementation uses a custom robbery detection dataset that consists of gun, no-gun and robbery activity classes. By evaluating the performance of the proposed model on our custom dataset, it is evident that the ResNet50 backbone architecture yields outperforms for the accuracy in robbery detection that reached in 0.92 of Mean Average Precision (mAP). The model effectiveness lies in its ability to accurately identify the presence of guns during robbery activities.

Keywords
robbery activities, RetinaNet, ResNet50, FPN, FCN

For citation: Khin P.P., Htaik N.M. An improved performance of RetinaNet model for hand-gun detection in custom dataset and real time surveillance video. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 51–61. doi: 10.17586/2226-1494-2024-24-1-51-61

УДК 004.032/26

Улучшенная производительность модели RetinaNet для обнаружения огнестрельного оружия в пользовательском наборе данных и видеонаблюдения в реальном времени

Пьоне Пьоне Кхин[1]✉, Най Мин Хтайк[2]

[1,2] Мандалайский технологический университет, Мандалай, 05072, Мья́нма

[1] pyonekhin.ppk@gmail.com✉, https://orcid.org/0009-0002-0512-6414
[2] nayminhtaik@gmail.com, https://orcid.org/0009-0009-8295-6914

Аннотация
Распространенность вооруженных ограблений стала серьезной проблемой в современном мире, что требует разработки эффективных систем обнаружения. Существующие разнообразные устройства обнаружения не обладают способностью автоматически выявлять и предупреждать о наличии оружия во время осуществления вооруженных ограблений. Для решения этой проблемы предлагается подход, основанный на глубоком обучении, с использованием модели RetinaNet. В результате его применения возможно точное обнаружение оружия и дальнейшее предупреждение об ограблении полицейского участка или владельца банка. Ядро модели RetinaNet состоит из трех основных компонентов: остаточной сети (Residual Neural Network, ResNet), функциональной пирамидальной сети (Feature Pyramid Net, FPN) и полностью сверточной сети (Fully Convolutional Networks, FCN). Эти компоненты работают вместе, обеспечивая обнаружение оружия в режиме реального времени без

вмешательства человека. Предлагаемая реализация использует специальный набор данных для обнаружения грабежей, который состоит из классов активности с применением огнестрельного оружия, без оружия и грабежей. Оценка производительности предлагаемой модели на разработанном специальном наборе данных показал, что магистральная архитектура ResNet50 превосходит точность обнаружения ограблений, достигая меры оценки качества ранжирования (Mean Average Precision, mAP) 0,92. Эффективность модели заключается в ее способности точно определять наличие оружия во время ограбления.

Ключевые слова

вооруженное ограбление, RetinaNet, ResNet50, FPN, FCN

Ссылка для цитирования: Кхин П.П., Хтайк Н.М. Улучшенная производительность модели RetinaNet для обнаружения огнестрельного оружия в пользовательском наборе данных и видеонаблюдения в реальном времени // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 51–61 (на англ. яз.). doi: 10.17586/2226-1494-2024-24-1-51-61

Introduction

Robberies have indeed been a persistent and common criminal activity, causing loss and anxiety for the public. A robbery prevention system that is simple to use and mostly free of false alerts is required to prevent the rising rate of robberies throughout the world [1]. To address these requirements, a system is being developed that can detect robberies in stores or surveillance areas and promptly alert the police or the owner, enabling them to take appropriate action. The proposed system utilizes deep learning to detect robbery activities by focusing on gun detection. By employing deep learning techniques, the system can accurately identify the presence of a firearm and generate an alert message to notify the police station or the owner. The system will have a simple user interface and operate in real-time ensuring the protection of individuals and their valuable possessions.

In today's dynamic and evolving environment, ensuring security against robberies is a major concern. People's precious possessions require enhanced security measures. While many banks, commercial shops, and surveillance areas already have security cameras in place but need an anti-robbery system that can look after the safety of their property even when they are not there [2]. The proposed system specifically targets the security of many organizations in Myanmar. Installing a comprehensive security system typically incurs additional expenses. Although various security devices are available for crime detection and prevention, many organizations often have limited resources to invest in intelligent devices. Hence, the primary objective of the proposed system is to detect robbery activities in banks using deep learning techniques, particularly gun detection, and promptly alert the police station or bank owner through an alert message. The contributions of this paper are as follows:

— The deep learning-based robbery activities detection system that can provide the promising Mean Average Precision (mAP) results is proposed.
— Real time robbery activities detection system is proposed which can detect robbery activities quickly and accurately.
— Standard dataset (self-annotating from scratch) including small arms, robbery activity (kneeling and hand up), and no gun (cell phone, thermal gun and metal detector) that consist of seven challenges and can actually use in real world is created available for robbery activities.

Literature Reviews

Human supervision remains an essential component in surveillance systems, ensuring effective monitoring and response. However, recent advancements in computer vision have emerged as a pivotal trend in video surveillance offering significant efficiency gains. In paper [3], the authors developed theft detection and tracking system using Closed Circuit Television (CCTV) images that employs image processing techniques to detect theft incidents and track the movement of thieves, without the need for additional sensors. The primary focus of this paper lies in object detection, specifically identifying and monitoring individuals involved in theft activities. By leveraging real-time analysis of human movement in CCTV footage, this system enables security personnel to be promptly notified about suspicious individuals engaged in burglary.

The authors introduced a fully automated computer-based system designed for identifying handguns and rifles which are fundamental armaments of concern [4]. Recent advancements at deep learning and transfer learning had described significant improvement in the detection and recognition of object. Our approach involved implementing the You Only Look Once (YOLO) v3 object detection model, trained on a customized dataset specifically tailored to our task. The validation of training results indicated that YOLOv3 outperforms both YOLOv2 and traditional Convolutional Neural Network (CNN) models. The approach applied this paper did not require intensive GPU or high computational resources, as transfer learning techniques for training the model was applied. This paper protected human life and reduced instances of manslaughter or mass killings. Furthermore, this system has the potential to be deployed in high-end surveillance and security robots, enabling the detection of weapons or unsafe assets and mitigating the risks of assault or harm to human life.

The paper [5] focused on the application of three CNN approaches for automatic handguns detection at video surveillance images. This paper explored the potential reduction of false positive detections with the incorporation of pose information regarding with how the handguns are held in the training dataset. The system evaluation showed that RetinaNet fine-tuned by the unfrozen Residual Neural Network, ResNet50, backbone provided the highest Average Precision (AP) by 0.9636 and recall by 0.9723. YOLOv3 exhibited consistent improvement of around 2 % as the explicit consideration of pose information during

training, distinguishing it from the other architectures evaluated in this paper.

The real-time object detection system for automatic weapon detection was proposed in video surveillance systems [6]. This framework introduced an early weapon detection technique using state-of-the-art, real-time object detection systems like YOLO and Single Shot Multi-Box Detector (SSD). Furthermore, the importance of minimizing false alarms was focused for ensuring the applicability of the model in real-life conditions. The developed model was well-suited to indoor surveillance cameras deployed in various circumstances, containing banks, supermarkets, malls, gas stations, and other similar environments. The system implementation showed that it can serve as a preemptive system for deterring potential robberies with the integration of outdoor surveillance cameras.

In paper [7], the authors developed a Hybrid Weapon detection system. In phase I, image processing techniques were employed, resulting in an average accuracy of 0.9464 after several dataset partitions. In phase II, the line integration method was applied, yielding a normalized electric field of $2.25 \cdot 10^{-9}$ V/m, indicating the electromagnetic waves expected to have passed through the metal object (weapon). This low value, relative to that of air (448 °C), demonstrates the minimal electromagnetic wave transmission through the metal, thereby facilitating the identification of the object. Furthermore, the fuzzy logic system exhibited an 83 % decision rate suggesting that it can reliably issue commands in weapon detection scenarios.

The authors described the real-time visual detection of handguns in videos [8]. This paper utilized the YOLOv3 model and performed the comparison of the false positive and false negative rates with the Faster Region-based CNN (RCNN) model. For the improvement in the evaluation, a dataset was compromised by handguns from various angles; it was created and combined with the ImageNet dataset. This combined dataset was then trained utilizing the YOLOv3 model. Four different videos were utilized for the validation in the performance of YOLOv3 in comparison with Faster RCNN. The results demonstrated the improved performance in handguns detection across various scenes, encompassing different rotations, scales, and shapes. The system implementation described that YOLOv3 can perform as a viable element to Faster RCNN supporting faster speed, with accuracy and suitability for real-time applications.

For the reduction of false positives and false negatives, a binary classification approach was introduced, with the pistol class designated as the reference class [9]. The relevant confusion objects were introduced for refining the detection operation. As there is lack of standard dataset for real-time environments, the creation of custom dataset was done with the collection of weapon photos through in-house camera, the manual collection of images by the internet, the deduction of data through YouTube CCTV videos, applying GitHub repositories, data from the University of Granada, and the Internet Movies Firearms Database (IMFDB) imfdb.org. Two methods were utilized: sliding window/classification and region proposal/object detection. Various approaches, including VGG16, Inception-V3, Inception-ResNetV2, SSDMobileNetV1, Faster RCNN Inception-ResNetV2 (FRIRv2), YOLOv3, and YOLOv4, were employed. The evaluation parameters as precision and recall are utilized in the system evaluation. Among all the algorithms tested, YOLOv4 outperforms other approaches with F1-score by 0.91 and a mAP by 0.9173.

In paper [10], the comparative analysis of two state-of-the-art models, YOLOv3 and YOLOv4, for weapons detection was performed. A weapons dataset was created for training including images through Google Images and supplemented with various assets. The images are manually annotated in different formats considering that YOLO requires annotation files in text format, while other models require XML format. Both models were trained on a large dataset of weapons, and their results are tested for comparative analysis. The paper demonstrated that YOLOv4 outperforms YOLOv3 in terms of processing time and sensitivity, while the precision metric is used to compare the two models.

The paper [11] described the automatic gun detection system using the Faster RCNN model. The CNN architecture was employed as a feature extractor in Faster RCNN by the experiments with Inception-ResNetV2, ResNet50, VGG16, and MobileNetV2. The system evaluation of those proposed architectures was performed in the comparison with YOLOv2. The results showed that the promising performance was achieved by Faster RCNN with Inception-ResNetV2 as the feature extractor. However, YOLOv2 provided the shortest training and testing time followed by VGG16, MobileNetV2, ResNet50, and Inception-ResNetV2 which achieved the longest training and testing time.

Theoretical Background

This section gives some theoretical background of the paper such as RetinaNet, different backbones (ResNet50, VGG19 and VGG16).

RetinaNet

RetinaNet is indeed a one-stage object detection model that addresses the challenges of imbalanced data and objects of different sizes. It achieves this through the use of a Feature Pyramid Network (FPN) and a specialized loss function called Focal Loss. By incorporating these components, RetinaNet can effectively detect objects in an image. The Focal Loss function plays a crucial role in RetinaNet architecture [12]. It tackles the issue of imbalanced data by assigning higher weights to difficult examples which are objects that the model struggles to detect. This weighting mechanism allows RetinaNet to focus more on challenging instances, improving its ability to handle imbalanced datasets. RetinaNet comprises three main subnetworks: a backbone network, an FPN, and a Fully Convolutional Network (FCN) for classification and regression. The backbone network, such as ResNet50, VGG19, or VGG16, serves as the feature extractor. It processes the input image and extracts relevant features.

Residual Neural Network

The ResNet architecture introduced a revolutionary bottom-up pathway that consists of residual blocks to improve training deep neural networks. The ResNet

consists of many convolution modules, each has many convolution layers. The output of each convolution module is used in the top-down pathway. ResNet is used as backbone architecture to extract multi-scale features in RetinaNet [13]. The output feature maps with various resolutions from backbone are sent to FPN.

Feature Pyramid Network

The FPN plays a central role in the architecture of RetinaNet. It is responsible for addressing the challenge of detecting objects at different scales by combining multi-scale features through top-down and lateral connections. The process begins with an input image that is passed through the backbone network. The FPN takes these feature maps produced by the backbone network as input. It consists of two main components: the top-down pathway and the lateral connections. In the top-down pathway, the feature maps from the topmost level of the backbone network are progressively upsampled to a higher resolution while reducing their spatial dimensions. This upsampling process can be accomplished using techniques like nearest-neighbor upsampling or transposed convolutions. Simultaneously, the lateral connections enable the FPN to combine high-level semantic information from the upsampled maps with low-level fine-grained details from the backbone feature maps. By repeating the process of upsampling and merging through the top-down and lateral connections, the FPN creates a feature pyramid with different scales. The feature pyramid captures a range of spatial resolutions, allowing RetinaNet to effectively detect objects of varying sizes. These final feature output combinations from the FPN are then used for subsequent tasks, such as object classification and bounding box regression.

Fully Convolutional Network

In RetinaNet architecture, the FCN component is responsible for performing both the classification and box regression tasks. The classification subnetwork within the FCN is designed to handle all the classification tasks. For instance, it determines whether a gun is present or not in the given image. The classification subnetwork takes the feature maps generated by the preceding stages of the network, such as the FPN, and performs classification on these features. It assigns a class label to each anchor or region of interest indicating the presence or absence of a gun. On the other hand, the box regression task is performed by the box subnetwork within the FCN. This subnetwork is responsible for refining the localization of the gun in the image and recording the bounding box coordinates. It takes the same set of feature maps as the classification subnetwork and predicts the offset or adjustment required for the anchor boxes or regions of interest to accurately localize the gun. By regressing the box coordinates, the box subnetwork refines the initial estimates and produces more precise bounding box predictions for the detected objects. Together, the classification and box regression subnetworks in the FCN collaborate to identify the presence of guns in the image and accurately localize them by predicting the bounding box coordinates.

ResNet50

ResNet50 is a type of CNN which has a 50-layered architecture. ResNet is a principal neural network that is utilized as the backbone in computer vision fields. The pretrained network can be loaded more than a million images through the database of ImageNet. This network possesses the input image size of 224 × 224. The ResNet50 architecture contains 5 stages. In the first stage, it consists of convolution and pooling. In the rest stages, each step includes a convolution and identity block [14].

VGG19

VGG19 is a CNN architecture that was used to win the 2014 ILSVR (ImageNet) competition; it is a variation of the VGG model with 19 layers. The VGG19 sets the input image size to 224 × 224. The rectified linear unit (ReLU) activation function is applied.

VGG16

VGG16 is a CNN architecture that was applied for achieving the winner at ILSVR (ImageNet) competition at 2014. In VGG16 there are thirteen convolutional layers, five Max Pooling layers, and three Dense layers which sum up to 21 layers but it has only sixteen weight layers, i.e., learnable parameters layer [15].

Proposed System Architecture

We propose a new method for addressing the problems with the current research because it will help to reduce the unnecessary utilization of hardware and reduce the cost of the project. Our proposed system will reduce the classification errors with utilizing a variety of methods, such as dataset creation, model training, classification, and detection.

Dataset Creation

The process of data collection is an essential initial step in constructing a gun detection model. To address the challenges associated with the object detection, such as viewpoint variation, deformation, lighting conditions, cluttered or textured backgrounds, and more, appropriate data needs to be collected. In this system, our own standard dataset that consists of seven challenges and can actually use in real world to detect the robbery activities is created. This dataset consists of robbery photos from our own camera, manually collected images from internet, data extracted from YouTube CCTV videos. The robbery detection dataset is divided into the following three classes: Gun, No-Gun, Robbery activity.

Robbery Detection Dataset Categories for Gun Class

Gun class consists of pistol and rifle categories as shown in Fig. 1.

Robbery Detection Dataset Categories for No-Gun Class

To reduce the number of false positives and negatives, No-Gun class that can most be confused with Gun class is added. No-Gun class consists of thermal gun, cell phone and metal detector as depicted in Fig. 2.

Robbery Detection Dataset Categories for Robbery Activity Class

In order to increase overall accuracy and precision, robbery activity class is added in our annotated dataset too. Robbery activity class consists of kneeling and hand-up categories as shown in Fig. 3.

Fig. 1. Sample dataset for Gun class (including pistols and rifles) with various sources, resolutions and conditions

Fig. 2. Sample dataset for No-Gun class (including thermal gun (*a*), cell phone (*b*) and metal detector (*c*))

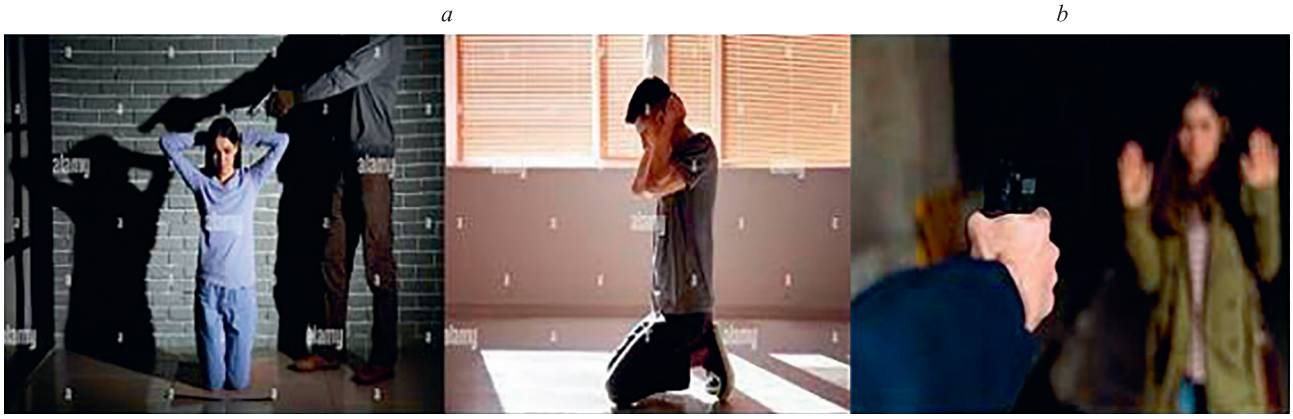

Fig. 3. Sample dataset for Robbery activity class (kneeling (*a*) and hand-up (*b*))

The step by step procedures of dataset creation are the followings.

Step 1: Collect the images from our own camera, internet, and YouTube CCTV videos.

Step 2: Annotate each image using the LabelMe annotation tool in Anaconda3 2021 in order to create ground truth labels with bounding boxes and corresponding class labels.

Step 3: Perform image preprocessing and augmentation.

Step 4: Construct our own custom dataset.

The flowchart of dataset preparation is also depicted in Fig. 4. Firstly, raw data images are collected. Later, the LabelMe annotation tool was utilized to annotate each image as shown in Fig. 5, *a*, and annotated images with JSON files are achieved as shown in Fig. 5, *b*.

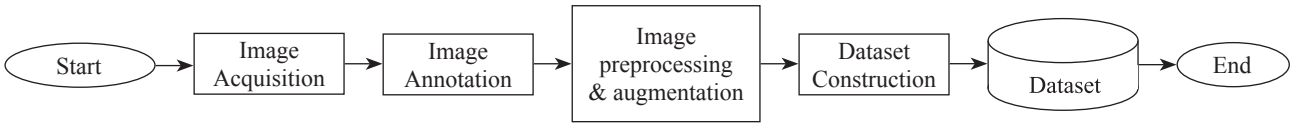

Fig. 4. Flowchart of dataset creation

Fig. 5. Image annotation and labeling (*a*), annotated images with JSON files (*b*)

After annotating images, image preprocessing and augmentation are performed by using Roboflow as illustrated in Fig. 6. In the image preprocessing, auto orientation and resizing images to 416 × 416 pixels were performed to improve the model performance. In data augmentation, random rotation between –5° and +5°, random shear of between –10° and +10° horizontally and vertically, and adding salt and pepper noise to 3 % of pixels were also employed to enhance the performance of deep neural networks. To facilitate data manipulation, the JSON files containing the annotations for each image were converted into a CSV file by using Roboflow. Finally, these images were constructed as the standard dataset. During the system evaluation, 80 % of the images were used for training the model while the remaining 20 % were used for validation and testing.

Training

Then, the RetinaNet model is trained with the different backbones (ResNet50, VGG16, VGG19). It starts with defining a problem, finding the required dataset, applying pre-processing methods, and then finally training and evaluating the dataset. The training process for different three models is shown in Fig. 7.

Classification and Detection

Features are extracted from input images using different backbones (ResNet50, VGG16 and VGG19). First, input image is achieved from real-time camera. The image passes through the detection model. The ImageNet collection is used to train the ResNet50 model. The first layer input image is a 416 × 416 RGB image. For RetinaNet with a backbone (ResNet50, VGG16 and VGG19), 3 batch-sizes require 139 MB of RAM. The flow chart of detection and classification processes with different models is also shown in Fig. 7. After the model training is finished, its AP, recall, and mAP are validated using a test image dataset at this stage. In the detection and classification stage, the value of confidence and Intersection-Over-Union (IoU) threshold is 0.5. The system then shows the output of the model. An alarm message is received by the organization if the system detects a gun. The alarm message will be checked by the police station or the owner of the organization, and they will act against it.

Fig. 6. Image pre-processing and augmentation

System Evaluation

This model was trained on a 64-bit desktop computer with NVIDIA GeForce RTX 2080 graphic card (Compute Capability = 7.5). TensorFlow, Keras, and OpenCV are used to train a gun detection system. Using the gun detection method, a model is trained and tested on a desktop computer achieving accuracy of 0.92 during training and testing. The complete process to evaluate the proposed system is shown in the following Fig. 8. Firstly, the input video frames from CCTV camera are collected. Then, images are pre-processed to improve the quality of operation. This pre-processing step generally includes resizing frame images in order to achieve more accurate deep learning model. After the pre-processing process, the training model is built using Transfer Learning. Then, the prediction of the input resized images is performed. After that, the output for label and boundary box with confidence score is displayed if the confidence value is greater than the threshold value. In this system, the confidence value is considered with the IoU score and the classification score. Finally, the alert message is sent to the user. The step by step procedures to analyze the proposed system are the followings.

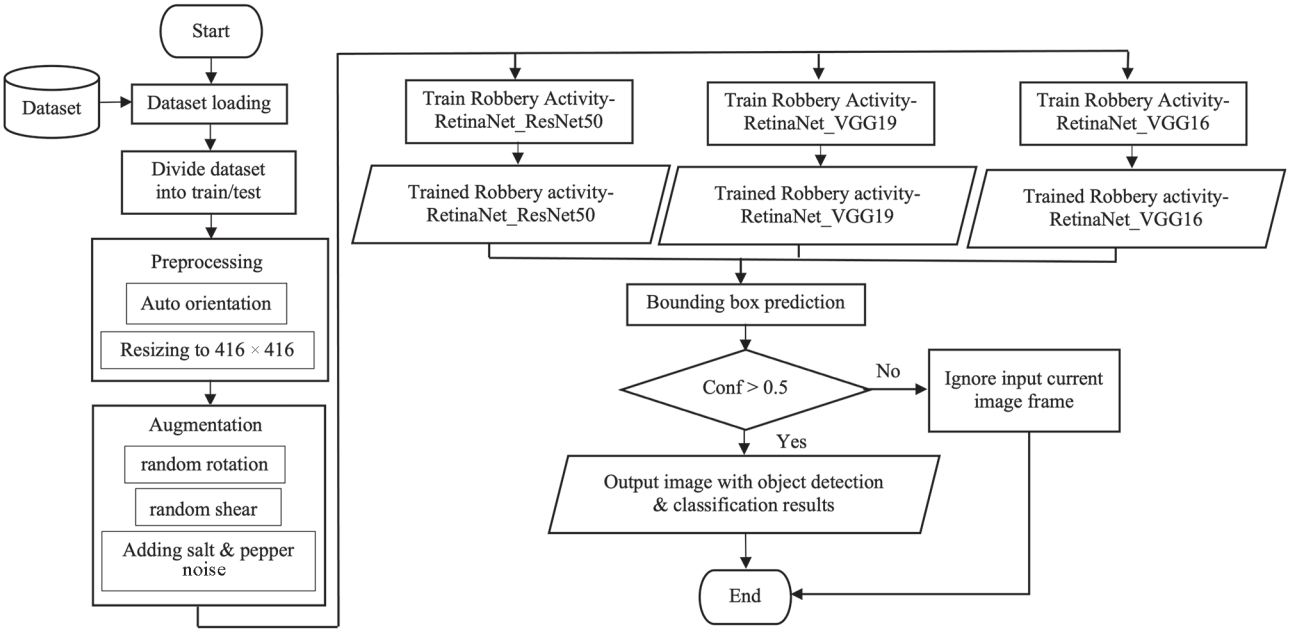

Fig. 7. Training and detection processes with three models

Step 1: Get input CCTV frame or image.

Step 2: Perform preprocessing step to resize the input image.

Step 3: Predict input resized image using the pre-trained model.

Step 4: If confidence value is greater than threshold value, then print output label and bounding box with confidence score and send an alert.

The key metrics of performance measures (AP, Recall, mAP, and Training Time) are evaluated for this proposed system analysis. This system can accurately detect the guns with low resolution and light. RetinaNet is known

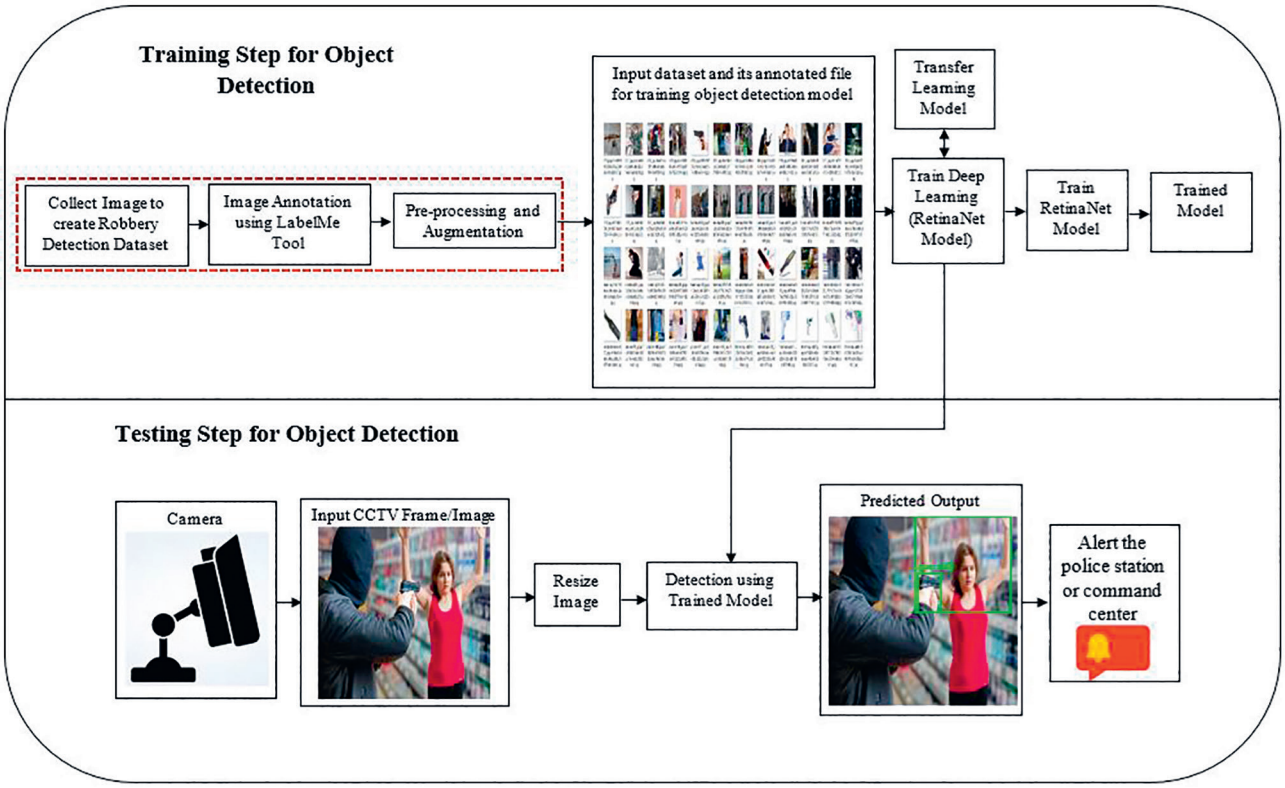

Fig. 8. Proposed system design

Table 1. Splitting Our Custom Dataset

Dataset	Pistol	Rifle	Kneeling	Hands up	No gun_phone	No gun_thermal gun	No gun_metal detector	Total images
Training	1602	1602	1861	1866	1378	1831	1860	12,000
Validation	193	193	242	234	188	228	222	1500
Testing	193	193	242	234	188	228	222	1500

Table 2. AP Results of Three Models for Each Class

Model	AP (gun)	AP (robbery activity)	AP (no gun thermal gun)	AP (no gun_metal detector	AP (no gun_phone)
RetinaNet_ResNet50	0.79	0.98	0.98	0.94	0.92
RetinaNet_VGG19	0.53	0.71	0.87	0.67	0.35
RetinaNet_VGG16	0.42	0.62	0.77	0.50	0.32

for its strong object detection performance. In a hand-gun detection system, RetinaNet capacity to detect objects across various angles and orientations is particularly beneficial. It can detect handguns in both well and poor environments. This adaptability is crucial for ensuring that handguns are identified regardless of the lighting conditions in surveillance or security scenarios. It tends to have a low false positive rate minimizing the likelihood of false alarms in a handgun detection system. This is critical for avoiding unnecessary security interventions. It directly predicts object bounding boxes and class labels without the need for a separate region proposal network. This simplicity can lead to faster inference times and simplified system architecture.

Analysis and Discussion

The system develops the gun detection system using RetinaNet_ResNet50, RetinaNet_VGG19, and RetinaNet_VGG16 models. The self-annotated dataset includes 15,000 images in total: 12,000 images for training, 1500 images for validation and 1500 images for testing as shown in Table 1. To demonstrate system evaluation using three different models, 40 epochs are used because the value of mAP started to stabilize at approximately 0.92. The comparison of training time for three models is described in Fig. 9. The AP results for each class on our annotated dataset with RetinaNet_ResNet50, RetinaNet_VGG19, and RetinaNet_VGG16 models are depicted in Table 2. The recall and mAP on our own custom dataset using three different models are also shown in Table 3. According to the evaluation results, the selection of ResNet50 can be considered the best for target detection although the RetinaNet_ResNet50 takes a little more training time than the other two models.

Table 3. Recall and mAP Results of Three Models

Model	Recall	mAP
RetinaNet_ResNet50	0.74	0.92
RetinaNet_VGG19	0.62	0.63
RetinaNet_VGG16	0.58	0.53

Detection Results: Gun Class in CCTV Records and Robbery Videos

The detection results of Gun class with CCTV records and robbery videos for RetinaNet_ResNet50 model are also illustrate in Fig. 10.

Detection Results: Robbery Activity, Thermal Gun, Cell Phone and Metal Detector Classes

The detection results of robbery activity, thermal gun, cell phone and metal detector classes for RetinaNet_ResNet50 model are also illustrated in Fig. 11. According to the detection results of Fig. 11, it can be seen that our model accurately detects multiple classes in an image.

Misdetections

The misdetection results of RetinaNet_ResNet50 model are also depicted in Fig. 12. Regarding to the detection results, it is seen that other things not being guns are mistakenly detected as guns by our model. So, we will maintain our focus on the continued reduction of false positives recognizing the persistent need for improvement. Additionally, we may explore the possibility of expanding the number of classes or objects in our work.

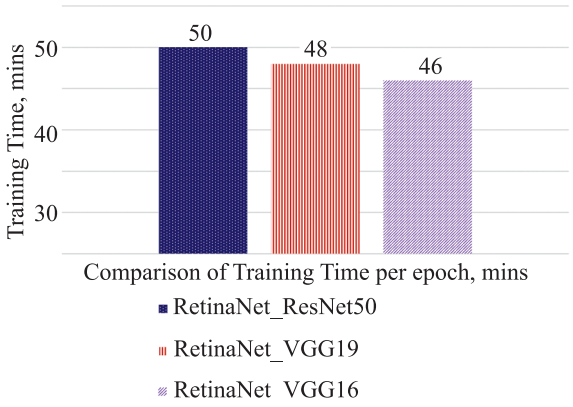

Fig. 9. Training time results of three models

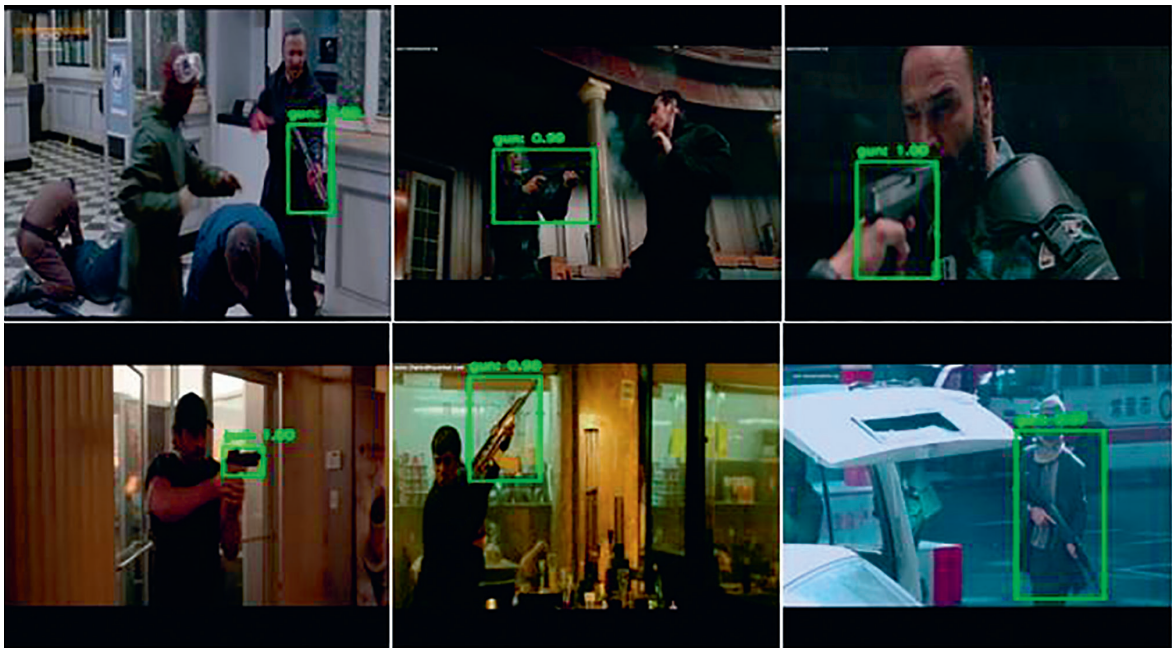

Fig. 10. Detection results of Gun class with robbery videos

Fig. 11. Detection results of robbery activity, thermal gun, cell phone and metal detector classes

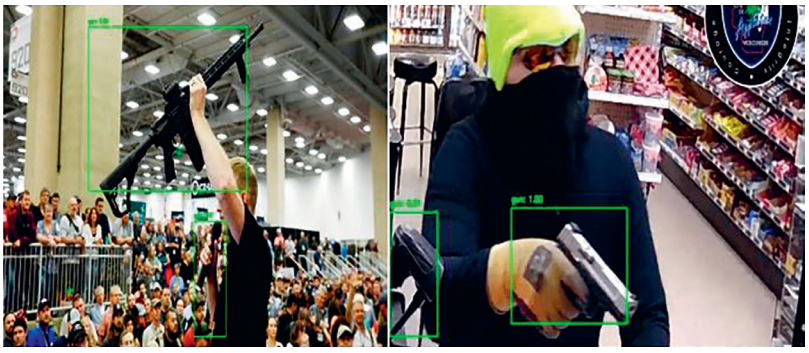

Fig. 12. Misdetections: False positives

Conclusion

The main objective of the research conducted was to develop an accurate gun detection system aimed at addressing security concerns and reducing or preventing robbery incidents. The focus was on creating an anti-robbery device that utilizes gun detection capabilities. To achieve this, RetinaNet_ResNet50, RetinaNet_VGG19, and RetinaNet_VGG16 were simulated for the purpose of gun detection. These models were trained and evaluated using an annotated dataset specifically designed for this research. The performance of the models is compared using AP, recall and mAP on the self-annotated dataset. According to the AP results, the RetinaNet_ResNet50 model has shown outstanding AP for each class in our dataset than the other models. In this evaluation, RetinaNet_ResNet50 model achieves the highest detection accuracy and outperforms accurately in detecting gun, robbery activity, thermal gun, metal detector and phone. As mentioned by the evaluation, the RetinaNet_ResNet50 model achieves the best recall and mAP values among three models. It gave 0.92 mAP and 0.74 (recall) on all types of images. Overall, it may be said that RetinaNet_ResNet50 exhibited the highest detection accuracy among the three models. So, RetinaNet_ResNet50 model was deemed the most suitable choice for target detection, considering its superior performance in terms of detection accuracy.

Based on the results of this study, we will further improve our developed dataset for detecting robbery cases precisely. Regarding the classification approaches, other deep learning models, such as EfficientDet and YOLO, would be used for detecting guns. Also, it is recommended the presented results in the paper to be compared with other deep learning models of ensemble classifiers to achieve more accurate detection model. We suggest that researchers should use ensemble method by combining object detection models to detect robbery cases precisely and enhance the mAP score of the robbery detection surveillance system.

References

1. Dever J., da Vitoria Lobo N., Shah M. Automatic visual recognition of armed robbery. *Proc. of the 16th International Conference on Pattern Recognition. V. 1*, 2002, pp. 451–455. https://doi.org/10.1109/ICPR.2002.1044755
2. Ahmed S., Bhatti M.T., Khan M.G., Lövström B., Shahid M. Development and optimization of deep learning models for weapon detection in surveillance videos. *Applied Sciences*, 2022, vol. 12, no. 12, pp. 5772. https://doi.org/10.3390/app12125772
3. Kakadiya R., Lemos R., Mangalan S., Pillai M. AI based automatic robbery/theft detection using smart surveillance in banks. *Proc. of the 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2019, pp. 201–204. https://doi.org/10.1109/iceca.2019.8822186
4. Narejo S., Pandey B., Vargas D.E., Rodriguez C., Anjum M.R. Weapon detection using YOLO V3 for smart surveillance system. *Mathematical Problems in Engineering*, 2021, vol. 2021, pp. 9975700. https://doi.org/10.1155/2021/9975700
5. Salido J., Lomas V., Ruiz-Santaquiteria J., Deniz O. Automatic handgun detection with deep learning in video surveillance images. *Applied Sciences*, 2021, vol. 11, no. 13, pp. 6085. https://doi.org/10.3390/app11136085
6. Zahrawi M., Shaalan K. Improving video surveillance systems in banks using deep learning techniques. *Scientific Reports*, 2023, vol. 13, pp. 7911. https://doi.org/10.1038/s41598-023-35190-9
7. Ineneji C., Kusaf M. Hybrid weapon detection algorithm, using material test and fuzzy logic system. *Computers & Electrical Engineering*, 2019, vol. 78, pp. 437–448. https://doi.org/10.1016/j.compeleceng.2019.08.005
8. Warsi A., Abdullah M., Husen M.N., Yahya M., Jawaid N. Gun detection system using YOLOv3. *Proc. of the IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*, 2019, pp. 1–4. https://doi.org/10.1109/ICSIMA47653.2019.9057329
9. Bhatti M.T., Khan M.G., Aslam M., Fiaz M.J. Weapon detection in real-time CCTV videos using deep learning. *IEEE Access*, 2021, vol. 9, pp. 34366–4382. https://doi.org/10.1109/ACCESS.2021.3059170
10. Hashmi T.S.S., Haq N.U., Fraz M.M., Shahzad M. Application of deep learning for weapons detection in surveillance videos. *Proc. of the 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, 2021. https://doi.org/10.1109/ICoDT252288.2021.9441523
11. Alaqil R.M., Alsuhaibani J.A., Alhumaidi B.A., Alnasser R.A., Alotaibi R.D., Benhidour H. Automatic gun detection from images using Faster R-CNN. *Proc. of the IEEE International Conference of Smart Systems and Emerging Technologies (SMARTTECH)*, 2020,

Литература

1. Dever J., da Vitoria Lobo N., Shah M. Automatic visual recognition of armed robbery // Proc. of the 16th International Conference on Pattern Recognition. V. 1. 2002. P. 451–455. https://doi.org/10.1109/ICPR.2002.1044755
2. Ahmed S., Bhatti M.T., Khan M.G., Lövström B., Shahid M. Development and optimization of deep learning models for weapon detection in surveillance videos // Applied Sciences. 2022. V. 12. N 12. P. 5772. https://doi.org/10.3390/app12125772
3. Kakadiya R., Lemos R., Mangalan S., Pillai M. AI based automatic robbery/theft detection using smart surveillance in banks // Proc. of the 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA). 2019. P. 201–204. https://doi.org/10.1109/iceca.2019.8822186
4. Narejo S., Pandey B., Vargas D.E., Rodriguez C., Anjum M.R. Weapon detection using YOLO V3 for smart surveillance system // Mathematical Problems in Engineering. 2021. V. 2021. P. 9975700. https://doi.org/10.1155/2021/9975700
5. Salido J., Lomas V., Ruiz-Santaquiteria J., Deniz O. Automatic handgun detection with deep learning in video surveillance images // Applied Sciences. 2021. V. 11. N 13. P. 6085. https://doi.org/10.3390/app11136085
6. Zahrawi M., Shaalan K. Improving video surveillance systems in banks using deep learning techniques // Scientific Reports. 2023. V. 13. P. 7911. https://doi.org/10.1038/s41598-023-35190-9
7. Ineneji C., Kusaf M. Hybrid weapon detection algorithm, using material test and fuzzy logic system // Computers & Electrical Engineering. 2019. V. 78. P. 437–448. https://doi.org/10.1016/j.compeleceng.2019.08.005
8. Warsi A., Abdullah M., Husen M.N., Yahya M., Jawaid N. Gun detection system using YOLOv3 // Proc. of the IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA). 2019. P. 1–4. https://doi.org/10.1109/ICSIMA47653.2019.9057329
9. Bhatti M.T., Khan M.G., Aslam M., Fiaz M.J. Weapon detection in real-time CCTV videos using deep learning // IEEE Access. 2021. V. 9. P. 34366–4382. https://doi.org/10.1109/ACCESS.2021.3059170
10. Hashmi T.S.S., Haq N.U., Fraz M.M., Shahzad M. Application of deep learning for weapons detection in surveillance videos // Proc. of the 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2). 2021. https://doi.org/10.1109/ICoDT252288.2021.9441523
11. Alaqil R.M., Alsuhaibani J.A., Alhumaidi B.A., Alnasser R.A., Alotaibi R.D., Benhidour H. Automatic gun detection from images using Faster R-CNN // Proc. of the IEEE International Conference of Smart Systems and Emerging Technologies (SMARTTECH). 2020. P. 149–154. https://doi.org/10.1109/SMART-TECH49988.2020.00045

pp. 149–154. https://doi.org/10.1109/SMART-TECH49988.2020.00045
12. Yang M., Xiao X., Liu Z., Sun L., Guo W., Cui L., Sun D., Zhang P., Yang G. Deep RetinaNet for dynamic left ventricle detection in multiview echocardiography classification. *Scientific Programming*, 2020, vol. 2020, pp. 7025403. https://doi.org/10.1155/2020/7025403
13. Bhabad D., Kadam S., Malode T., Shinde G., Bage D. Object detection for night vision using deep learning algorithms. *International Journal of Computer Trends and Technology*, 2023, vol. 71, no. 2, pp. 87–92. https://doi.org/10.14445/22312803/ijctt-v71i2p113
14. Rani E.E., Baulkani S. Construction of deep learning model using ResNet 50 for schizophrenia prediction from rsFMRI images. *Research Square*, 2022. https://doi.org/10.21203/rs.3.rs-2106170/v1
15. Guan Q., Wang Y., Ping B., Li D., Du J., Qin Y., Lu H., Wan X., Xiang J. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: A pilot study. *Journal of Cancer*, 2019, vol. 10, no. 20, pp. 4876–4882. https://doi.org/10.7150/jca.28769

12. Yang M., Xiao X., Liu Z., Sun L., Guo W., Cui L., Sun D., Zhang P., Yang G. Deep RetinaNet for dynamic left ventricle detection in multiview echocardiography classification // Scientific Programming. 2020. V. 2020. P. 7025403. https://doi.org/10.1155/2020/7025403
13. Bhabad D., Kadam S., Malode T., Shinde G., Bage D. Object detection for night vision using deep learning algorithms // International Journal of Computer Trends and Technology. 2023. V. 71. N 2. P. 87–92. https://doi.org/10.14445/22312803/ijctt-v71i2p113
14. Rani E.E., Baulkani S. Construction of deep learning model using ResNet 50 for schizophrenia prediction from rsFMRI images // Research Square. 2022. https://doi.org/10.21203/rs.3.rs-2106170/v1
15. Guan Q., Wang Y., Ping B., Li D., Du J., Qin Y., Lu H., Wan X., Xiang J. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: A pilot study // Journal of Cancer. 2019. V. 10. N 20. P. 4876–4882. https://doi.org/10.7150/jca.28769

Authors

Pyone Pyone Khin — Master, PhD Student, Lecturer, Mandalay Technological University, Mandalay, 05072, Myanmar, https://orcid.org/0009-0002-0512-6414, pyonekhin.ppk@gmail.com

Nay Min Htaik — PhD, Professor, Mandalay Technological University, Mandalay, 05072, Myanmar, https://orcid.org/0009-0009-8295-6914, nayminhtaik@gmail.com

Авторы

Кхин Пьоне Пьоне — магистр, аспирант, преподаватель, Мандалайский технологический университет, Мандалай, 05072, Мья́нма, https://orcid.org/0009-0002-0512-6414, pyonekhin.ppk@gmail.com

Хтайк Най Мин — PhD, профессор, Мандалайский технологический университет, Мандалай, 05072, Мья́нма, https://orcid.org/0009-0009-8295-6914, nayminhtaik@gmail.com

Received 08.08.2023
Approved after reviewing 17.11.2023
Accepted 13.01.2024

Статья поступила в редакцию 08.08.2023
Одобрена после рецензирования 17.11.2023
Принята к печати 13.01.2024